JOURNAL OF QUATERNARY SCIENCE (2023) 1–11 ISSN 0267-8179. DOI: 10.1002/jqs.3526

JQS Journal of Quaternary Science QRA Quaternary Research Association

Shifting potential tree species distributions from the Last Glacial Maximum to the Mid-Holocene in North America, with a correlation assessment

BRICE B. HANBERRY*

USDA Forest Service, Rocky Mountain Research Station, Rapid City, SD, USA

Received 26 May 2022; Revised 29 March 2023; Accepted 10 April 2023

ABSTRACT: Pollen reconstructions of tree genera in North America since glaciation are available, but species distributions predicted for paleoclimate based on tree inventories may inform knowledge gaps. Here I examined the distributions of 25 species or species groups from 20 000 years ago (ka) to 5 ka to give potential paleoecological ranges of boreal and temperate tree species. I also assessed the effects of correlated climate variables on species distribution models for current and past climate, which were modeled with the non-linear random forests classifier. Climate change alone was enough to create unique, species-specific paths that did not run directly north. At 20 ka, black spruce (*Picea mariana*) occurred as far north as the ice sheet and most boreal species generally may have extended as far south as 32°N, or the northern southeastern United States. Temperate eastern species may have extended as far north as 39°N in low-elevation locations and temperate eastern species displayed both continuous and clustered distributions across the southeastern United States. Rate of movement was 5.5 km per century between 20 and 14 ka, 13.3 km per century between 14 and 10 ka, and 3.25 km per century between 10 and 5 ka. Species retreated southward between 7 and 5 ka. Regarding correlation, models with all variables had greater accuracy than models with the two most important variables, which had greater accuracy than models with two variables of intermediate importance, demonstrating both reduced accuracy with omission of relevant variables and isolation of important variables that improve accuracy in models with correlated climate variables. Models from different climates identified the same two most important variables and ranked the remaining variables similarly, revealing robustness in the models over time. Distribution models agreed with pollen reconstructions regarding timing and rate of change, while generating detailed species-specific information about movement trajectories and velocities, latitudinal extents, and distribution continuities.

KEYWORDS: boreal; correlation; migration; paleoclimate; random forests; rates

Introduction

Species responses to climate change represent a critical research question due to anthropogenic radiative forcing. The last time that carbon dioxide levels were above 400 parts per million was perhaps about 4.5 million years ago (Pagani et al., 2010). Although future temperatures and species responses are uncertain, species have experienced climate warming under past climate. Tree genera responses to climate change are recorded in fossil pollen data, which provide the primary source of information about tree genera responses to climate warming during the past 20 000 years (Williams et al., 2004).

Tree species of modern forests persisted through glacial and interglacial periods by shifting with changing temperatures. Although tracking of ice sheets is the best-documented response by tree species to climate change, range shifts probably involved adaptation through differential survival, growth and reproduction under new conditions (Davis & Shaw 2001). Continental ice sheets reached their maximum extent 26.5–19 ka during the Last Glacial Maximum (Clark et al., 2009). Glacier recession was slow until warming after 15 ka followed by fluctuations until reaching relative stability about 10 ka (Dyke, 2005; Ordonez & Williams 2013). Lastly, modern climate developed 7–5 ka after cooling from the mid-Holocene thermal maximum (Dyke, 2005; Ordonez & Williams 2013).

During this interval of slower climate change, modern locations of tree distributions consequently stabilized, reflected by changes in the relative abundance of species within ranges rather than replacement by more southern tree species (Williams et al., 2004; Hanberry et al., 2013). Williams et al. (2004) demonstrated that tree pollen distribution and composition were relatively stable during the period 21–17 ka and from 7 ka to about 600 years ago, but pollen changed rapidly during 16–8 ka and again starting about 600 years ago after Euro-American settlement in northern and eastern North America. Equally, Blois et al. (2013) modeled tree pollen dissimilarity in eastern North America since 21 ka, finding rapid compositional turnover from 14 to 10 ka, and particularly during 11–10 ka, composition becoming similar to that at present by 10 ka across much of eastern North America. Likewise, Mottl et al. (2021) determined fast rates of tree pollen change between 15 and 10 ka that decreased over time across North America. However, Mottl et al. (2021)

*Correspondence: Brice B. Hanberry, as above.
E-mail: brice.hanberry@usda.gov

detected rapid rates of pollen change over the past few thousand years; they stated that a reasonable inference was anthropogenic activities primarily drove recent pollen change rather than climate change.

Many boreal species now occupy new ranges compared to occupied extents during the Last Glacial Maximum, whereas temperate species still maintain a portion of their glacial extents (Davis & Shaw 2001). During advancement poleward, boreal forests (consisting of black and white spruce, *Picea glauca*; balsam fir, *Abies balsamea*; tamarack, *Larix laricina*; balsam poplar, *Populus balsamifera*; quaking aspen, *Populus tremuloides*; white birch, *Betula papyrifera*; jack pine, *Pinus banksiana*) moved 1490 km, at mean rates of 102 m per year, with maximum rates reaching 202 m per year (Dyke, 2005). For example, Ordonez and Williams (2013) determined that *Abies* ranges moved at rates of 188 m per year at their northern edges during about 12–10 ka. Temperate forest species (e.g., oaks, *Quercus* spp.; American beech, *Fagus grandifolia*) also shifted at similar rates of 100–200 m per year (Dyke, 2005). However, after about 8 ka, the northern limit of temperate forests retreated 150 km southwards at an average rate of 19 m per year (Dyke, 2005).

Pollen and plant macrofossil records have supplied information about paleoecological vegetation and informed paleoclimate reconstructions (Webb et al., 1998; Bartlein et al., 2011; Dawson et al., 2016). Nonetheless, pollen records are an incomplete and biased representation of tree distributions, with taxonomic resolution that is typically at the genus level, and include issues with chronology (Dawson et al., 2016; Kujawa et al., 2016; Moreno-Amat et al., 2017; Chaput, 2018). An alternative approach of species distribution modeling based on past climate and modern comprehensive tree surveys may fill gaps that remain between localized pollen collection sites (Jackson et al., 2000). For example, hindcasted species distribution models can provide potential paleoecological ranges of boreal and temperate tree species, unlike pollen-based reconstructions that are limited to genera, and help answer questions such as how far south boreal tree species potentially were displaced and whether boreal species and temperate broadleaf species were confined to the Lower Mississippi Valley (Jackson et al., 2000). Here, I used inventory information from Canada and the United States (an extent of about 20 million km^2; Figure 1) for 25 tree species that represent current ecosystems of North America to model distributions under current climate and then predict distributions under paleoclimate. Because of widespread tree species distribution changes following altered land use and disturbance associated with Euro-American settlement and also mismatch between climate space and realized distributions due to non-climate factors (Williams et al., 2004; Dawson et al., 2016; Hanberry 2021, 2022), I modeled based on coarser ecological spatial units to account for uncertainty. To my knowledge, tree distribution modeling to supplement paleoecological data has not been applied in North America, at least using downscaled paleoclimate, although this research is an extension of modeled tree distributions under historical climate during Euro-American settlement (Hanberry 2022). Climate variables are derivatives of temperature and precipitation and inherently correlated due to a stable spatiotemporal structure, which is beneficial in terms of both having a reliable climate, rather than random weather, and maintaining long-term associations with species. Nonetheless, concerns exist about the robustness of species distribution models using correlated data and predictions to different times (Dormann et al., 2013). Additionally, prediction of tree species distributions to thousands of years in the past is uncommon. Therefore, I assessed the accuracy of species distribution models based on multiple highly correlated climate variables compared to models from two pairs of relatively uncorrelated climate variables and assessed stability by ranking the importance of climate variables from models of climate throughout time.

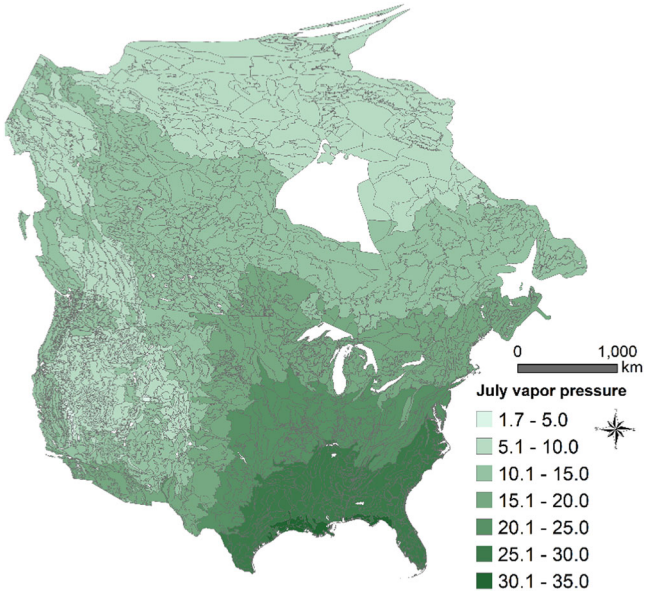

Figure 1. Study extent in Canada and the United States, with spatial units for modeling (outlined), and the climate variable of July vapor pressure displayed. [Color figure can be viewed at wileyonlinelibrary.com]

Methods

Data and modeling

Tree species data were from USDA Forest Service Forest Inventory and Analysis surveys (FIA, 2021) and Canada's National Forest Inventory (Beaudoin et al., 2017). Modeling and prediction did not include Mexico, and although climate may have been suitable in some locations of Mexico, desert, steppe and tropical climates also occurred. I considered species present if they comprised ≥0.5% of all species in 2180 sample units of ecological spatial units (Figure 1) of subsections (Keys et al., 2007) or districts (Government of Canada 2022). I focused on eight boreal species and 17 species primarily representative of ecoregions that were compressed in the US (Table 1). To have a minimum sample size of close to 100 ecological units, I combined western oak species (*Quercus agrifolia*, *Q. chrysolepis*, *Q. kelloggii*, *Q. wislizeni*) and southeastern pine species (*Pinus palustris* and *Pi. echinata*). Cottonwood comprised eastern cottonwood (*Populus deltoides*) and the plains subspecies (*Po. deltoides* subsp. *monilifera*).

Lorenz et al. (2016a, b) provided decadal mean climate from 21 ka to the present for North America after downscaling and debiasing the Community Climate System Model (CCSM3) circulation model simulation (Liu et al., 2009). I selected climate data from 200-year intervals centered on 5, 7, 10, 13, 14 and 20 ka, as well as current day (1960–1989, representing climate under which current trees probably established). Spatial resolution was 0.5°, matching coarse-scale analysis.

I extracted 12 climate variables, of which six were thermal. Temperature variables were growing degree days above 0 and 5 °C and mean minimum and maximum temperature annually and for the coldest and warmest months. Total annual precipitation, the coefficient of variation for precipitation, a moisture index and the aridity index of total annual precipitation

Table 1. Species, sensitivity and specificity for species distribution models based on current climate (1960–1989) in Canada and the USA.

Species	Scientific name	Sensitivity	Specificity
American beech	*Fagus grandifolia*	1.00	0.92
Balsam fir	*Abies balsamea*	0.99	0.95
Balsam poplar	*Populus balsamifera*	0.91	0.90
Black spruce	*Picea mariana*	0.99	0.96
Bur oak	*Quercus macrocarpa*	1.00	0.94
Douglas-fir	*Pseudotsuga menziesii*	0.97	0.88
Eastern cottonwood	*Populus deltoides*	0.90	0.90
Engelmann spruce	*Picea engelmannii*	0.98	0.97
Gambel oak	*Quercus gambelii*	0.96	0.79
Jack pine	*Pinus banksiana*	0.99	0.92
Paper birch	*Betula papyrifera*	0.96	0.93
Southern pines	*Pinus palustris, Pinus echinata*	1.00	1.00
Ponderosa pine	*Pinus ponderosa*	0.95	0.84
Post oak	*Quercus stellata*	0.97	0.97
Quaking aspen	*Populus tremuloides*	0.88	0.90
Red pine	*Pinus resinosa*	0.94	0.84
Red spruce	*Picea rubens*	0.97	1.00
Rocky Mountain juniper	*Juniperus scopulorum*	0.98	0.82
Subalpine fir	*Abies lasiocarpa*	0.91	0.89
Tamarack	*Larix laricina*	0.93	0.93
Two-needle pinyon pine	*Pinus edulis*	0.94	0.83
Utah juniper	*Juniperus osteosperma*	0.98	0.96
Western oaks	*Q. agrifolia, Q. chrysolepis, Q. kelloggii, Q. wislizeni*	1.00	1.00
White oak	*Q. alba*	1.00	0.99
White spruce	*Picea glauca*	0.95	0.89

to potential evapotranspiration indicated changing dryness. The ratio of December and January precipitation to annual precipitation and July vapor pressure (Figure 1) specifically constrained eastern and western tree species. I calculated mean values for each of the spatial units.

For modeling species distributions (i.e. presence or absence by spatial unit) under current climate (mean values by spatial unit), I applied random forests and extreme gradient boosting classifiers (Kuhn 2008; R Core Team 2021). Random forests and extreme gradient boosting classifiers are non-linear ensemble methods, which aggregate results of many decision trees or rule-based models to output the most optimal result, helping to minimize the influence of error. Each algorithm will have different approaches for modeling; the random forests classifier runs models in parallel (i.e. bagging) and averages results to reduce variance (i.e. overfitting), whereas the extreme gradient boosting runs models in sequence (i.e. boosting) to increase the weight of better models and reduce bias although extreme gradient boosting has a modification to prevent overfitting of conventional gradient boosting (Khaledian & Miller 2020; Zhang et al., 2021).

Models were trained with 10-fold cross-validation and then accuracy for models occurred on separate testing data (25% for this modeling), with withheld known classes, to determine how well the classifier assigned classes of presence or absence using predictor variables. The true positive rate or sensitivity is a measure of the number of predicted observations that were accurately assigned to a class, while the true negative rate or specificity is a measure of the number of predicted observations where the response variable of interest did not occur. For this modeling, prevalence, or the number of present samples to pseudoabsent samples, was equal. Then, I predicted species distributions for paleoclimate, mapped predicted distributions and calculated the distance between centroids of distributions over time. To synthesize change over time, I displayed maps of species combinations.

Correlation

Concern about multicollinearity stems from large standard errors of estimates of the coefficients for linear regression models (Dormann et al., 2013). Including collinear variables in the regression only inflates standard errors, while leaving coefficient estimates unbiased, making regression results more conservative (Lindner et al., 2020). Regardless, the non-linear modeling classifiers for this modeling are not able to determine coefficient values.

Multicollinearity is not a problem for predictions by linear or non-linear models (Goldberger & Goldberger 1991; Vaughan & Berry 2005; Shmueli 2010; Vatcheva et al., 2016; Williams et al., 2013; Lindner et al., 2020). Furthermore, ensemble models, which determine non-linear interactions among variables, are normally able to distinguish the influence of redundant, highly correlated predictors, compared to linear models (Tomaschek et al., 2018; Kuhn & Johnson 2019). Conceptually, tree-based modeling can distinguish the influence of correlated variables because only a subset of the variables is considered for each decision tree and each rule split of every tree is constructed with a threshold from a single variable.

Climate variables are correlated because they are derivatives of temperature and precipitation, which have a spatial structure. The structure of climate, rather than randomness, maintains species associations with climate over time, as demonstrated in pollen reconstructions. Indeed, greater collinearity of variables results in greater consequences of omitting relevant predictors (Lindner et al., 2020). In contrast, following a standard guideline to remove correlated variables at a 0.7 threshold before modeling (Dormann et al., 2013) will result in information loss and may result in incorrect omission of relevant variables to constrain or develop the model, generating bias.

Although ensemble models can isolate correlated strong predictors rather than splitting the explanatory value into intermediate importance, concern remains that correlated variables may interfere with measuring the influence of variables (Dormann et al., 2013). Enduring collinearity among climate variables creates a reliable basis for predictions, but collinearity will probably be a problem for spurious predictors when a model is trained on data from one region or time and predicted to another, due to a different structure of collinearity (Dormann et al., 2013). To address these issues, I reported correlation of paired variables with the Pearson's correlation coefficient. To evaluate the robustness of models with correlated variables, I modeled current species distributions four times with random forests under current climate with (i) all variables and also with (ii) two variables, consisting of a precipitation and temperature variable pair that was identified as the most important in the all variable models, and (iii) two variables that were identified as middling importance to compare accuracy. That is, if models with all variables were not able to distinguish the influence of redundant, highly correlated predictors then the predictors ranked as intermediate may produce greater accuracy than the most important variables, and models with all variable models may have least accuracy. To evaluate the robustness of models to hindcast predictions, I also modeled current species distributions four times with random forests and climate at 5, 7, 10 and 13 ka,

which occurred after loss of the ice sheet at least for most of the US. Ice sheets are not compatible with the majority of plant species distributions and are likely to affect some climate variables (Vizcaíno et al., 2008). I provided ranking of all variables for four models under each climate to assess consistency in the important values. This is to test whether models with correlated variables remained similar under different climates.

Results

Models

Models were accurate, with a mean true positive rate (or sensitivity) of 0.96, ranging from 0.88 to 1.00, and a mean true negative rate (or specificity) of 0.92, ranging from 0.82 to 1.00, for the 25 species or species groups (Table 1). The random forests and extreme gradient boosting models were similar. Because random forests models were generally slightly more accurate for true positive rate, I selected models from that classifier with the exception of eastern cottonwood, subalpine fir (*Abies lasiocarpa*), and Utah juniper (*Juniperus osteosperma*).

The centroids of distributions and distributions showed overall poleward advancement between 20 and 5 ka (Figures 2 and 3; note for Figure 3 that species distributions overlapped and the key displays the order of precedence, with more southerly species generally covering the full extent of more northerly species). Boreal species, in particular, expanded in absolute area with poleward advancement, although relative to the distribution at 20 ka, eastern cottonwood, bur oak (*Quercus macrocarpa*), post oak (*Quercus stellata*), jack pine and southern pines increased >4.5-fold (for all species, mean change in range area between 20 and 5 ka was 2.8). The potential ranges of black and white spruce, quaking aspen, Engelmann spruce (*Picea engelmannii*) and subalpine fir extended as far north as Alaska at 20 ka, due to less extreme temperatures along western North America. However, I did not adjust areal extent to account for changing land area due to varying sea levels and displacement by ice sheets, given that glacial positioning appeared to have been relatively dynamic in western Canada during this interval and scattered populations may have occurred.

Rate of movement for species, based on the distance between distribution centroids, was generally greatest (mean of 13.3 km per century, ranging from 3 to 33 km per century) during the period 14–10 ka, followed by 20–14 ka (mean of 5.5 km per century, ranging from 2 to 12 km per century) and 10–5 ka (mean of 3.25 km per century, ranging from 0.4 to 7.5 km per

Figure 2. Centroids of potential distributions at 20 and 5 ka.

Published 2023. This article is a U.S. Government work and is in the public domain in the USA. *Journal of Quaternary Science* published by John Wiley & Sons Ltd on behalf of Quaternary Research Association.

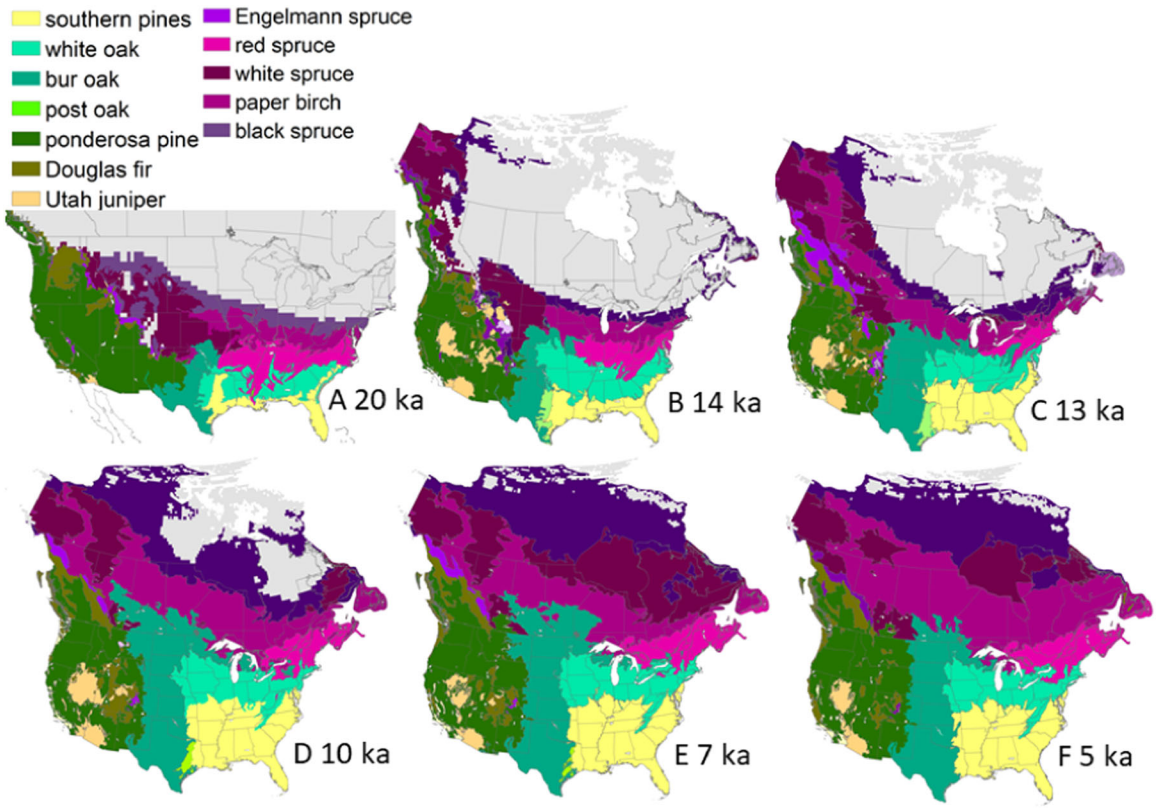

Figure 3. Distributions of wide-ranging species at 20 and 5 ka. Species distributions overlapped, so to display the southern species, the more southerly species generally covered the full extent of more northerly species. The key shows the order of precedence. General location of the ice sheet is indicated (gray). [Color figure can be viewed at wileyonlinelibrary.com]

century; Figure 4). Centroid movement did not occur only in a direct northward path. That is, the summed distance of the time interval divided into 20–14 ka, 14–10 ka and 10–5 ka was greater than the distance between centroids at 20 and 5 ka. Indeed, an overall southern reversion between 10 and 5 ka occurred for eight species (Rocky Mountain juniper, *Juniperus scopulorum*, 222 km southward in latitude; bur oak, 192 km; eastern cottonwood, 84 km; post oak, 63 km; ponderosa pine, *Pinus ponderosa*, 40 km; balsam poplar, 33 km; Utah juniper, 26 km; western oaks, 13 km). Furthermore, all but four species (subalpine fir; two-needle pinyon pine, *Pinus edulis*; red spruce, *Picea rubens*; and tamarack) had potential distributions that retreated south between 7 and 5 ka.

At 20 ka, potential species distributions of black and white spruce extended into the southernmost US states, but only red spruce reached the Gulf Coast (Figure 5A). Similarly, potential distributions of quaking aspen (Figure 5B); paper birch (*Betula papyrifera*), tamarack and balsam fir (Figure 5C); and jack and red pine (*Pinus resinosa*; Figure 5D) did not extend to the Gulf Coast. American beech extended to the Gulf Coast (Figure 5F), whereas white oak (Quercus alba) extended halfway into Florida (Figure 5E). Post oak, bur oak and southern pines covered all of Florida and part of southern Texas (Figure 5E, F). Eastern cottonwood had a patchy distribution but was present in parts of southern Florida and southern Texas (Figure 5E). Many western tree species extended to the Mexico border (Figure 6).

Evaluation of robustness

In addition to high model accuracy and examination of mapped models, I assessed models for robustness. Climate variables are highly correlated, providing consistent spatial structure over time. The six temperature variables had correlations ranging from 0.79 to 0.996. The foremost important variable identified by the all-variable model was July water vapor pressure, which had r values ranging from 0.72 to 0.79 with the two growing degree-day variables and annual minimum temperature. Summer vapor pressure was the most important model variable for 11 species. Despite formation of the ice sheet, which probably affects climate, correlation among vapor pressure values for the seven time intervals ranged from 0.78 to 0.95. Maximum temperature was the most important temperature variable for three species, and was also the secondmost important model variable for six species. Correlation among maximum temperature values for the seven time intervals ranged from 0.72 to 0.86. Correlation between vapor pressure and maximum temperature was relatively low (r = 0.53–0.72) during all time intervals. Correlation among two intermediate (ranks of six and eight) predictors, of precipitation and maximum monthly temperature, was low (r = 0.21) under current climate, and all time intervals as well.

Modeling with correlated climate variables increased accuracy, and the same two variables were identified consistently in time as the two most important variables, despite or actually because of correlation (Tables 2 and 3). Modeling with multiple correlated climate variables had greater accuracy (mean sensitivity of 0.95 and mean specificity of 0.92) than two-variable models. Additionally, accuracy decreased from the two-variable model identified, by the all-variable model, as most important (of vapor pressure and maximum temperature, with mean sensitivity of 0.91 and mean specificity of 0.87) to the two-variable model (of precipitation and maximum monthly temperature, with mean sensitivity of 0.86 and mean specificity of 0.81) identified as intermediate in importance. Furthermore, modeling current species observations with past climate at 5, 7, 10 and 13 ka

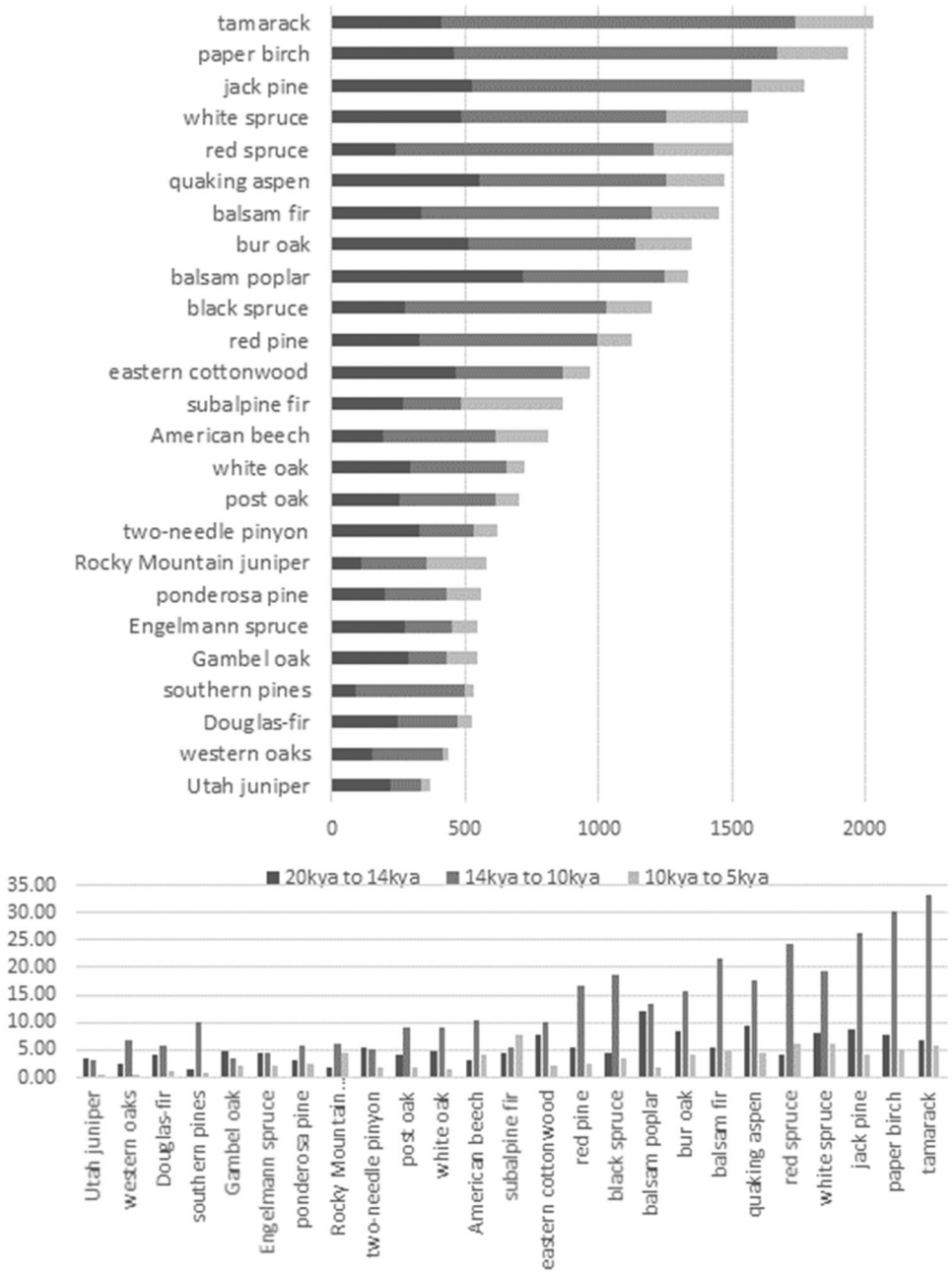

Figure 4. Distance (km) and rate of change (km per century) between centroids of potential distributions at 20 and 14 ka, 14 and 10 ka, and 10 and 5 ha.

(i.e. climate without glaciation into the US) resulted in models that identified the same two most important precipitation and temperature variables under different climates, with similar general ranking of all variables.

Discussion

Alignment with pollen reconstructions and information from species-specific distributions

Predictions of tree response to paleoclimate based on models of current tree distributions can help inform paleovegetation reconstructions from fossil pollen and plant macrofossils. Mapped syntheses of fossil pollen data have demonstrated that responses to past climate change were genus-specific ('individualistic'; Davis & Shaw 2001). Tree species ranges are directionally distinctive, in part because rooted individuals cannot migrate (Davis & Shaw 2001; Williams et al., 2004). While modeled species distributions overall appeared to be convergent with reconstructions based on pollen (Williams et al., 2004; Dyke, 2005; Blois et al., 2011), as discussed in detail below, these species distribution models were able to demonstrate uniquely species-specific responses rather than genus-specific responses to climate change. Predicted species distributions also provided additional species-specific details about potential paleoecological northern and southern extents, movement rates, and continuity of distributions that cannot be resolved with pollen data by genera, which contain both boreal and temperate species.

As documented by pollen reconstructions, tree species distribution models were compressed due to cooling and glaciation. Boreal and temperate species overlapped when ranges were compressed, and a transition zone between latitudes 34 and 33°N marked the ecotone between boreal and temperate communities (Delcourt & Delcourt 1987), but the extent of geographic displacement may be unknown (Jackson et al., 2000). The black spruce distribution model occurred as far north as the ice sheets;

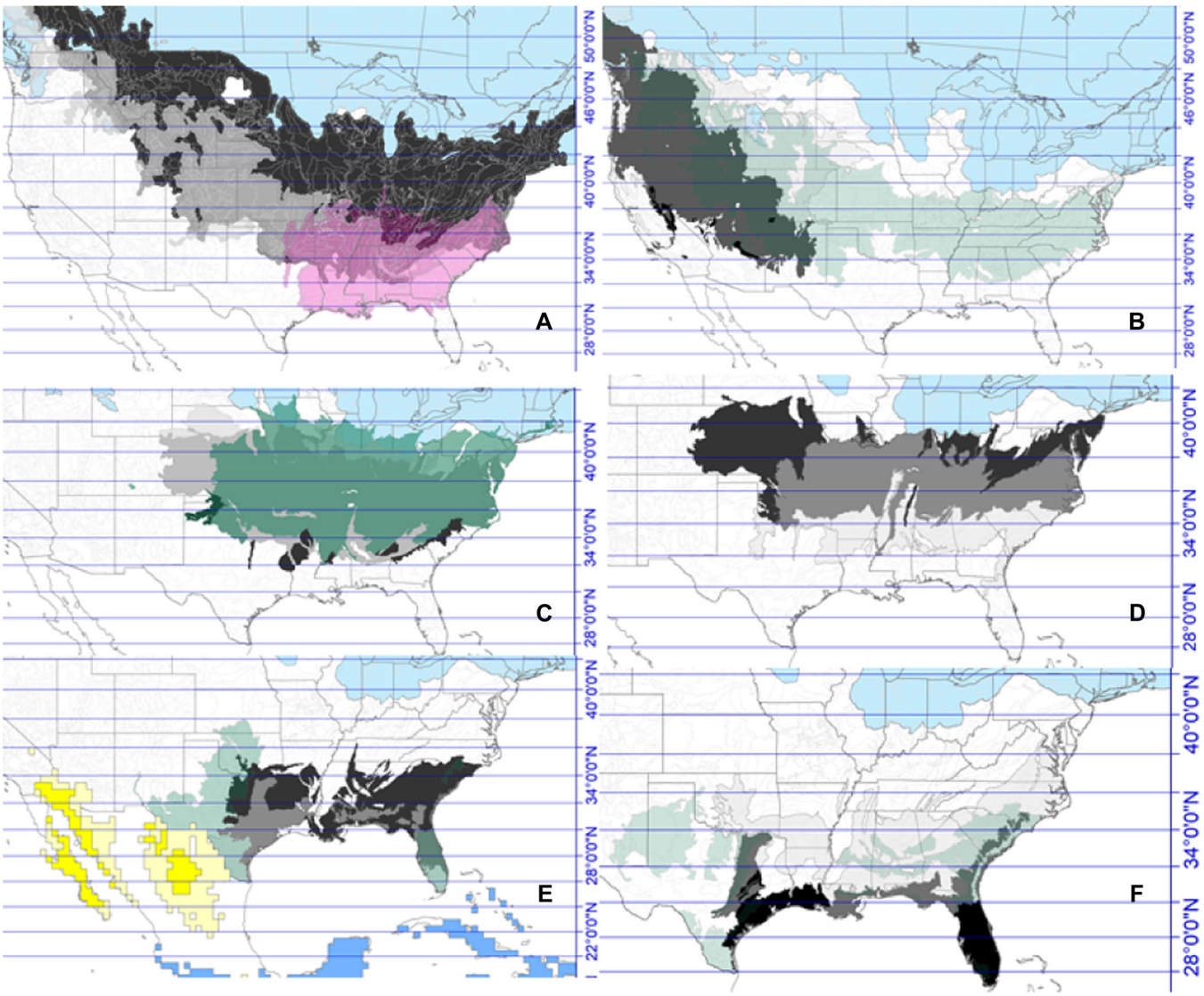

Figure 5. Combined potential distributions of continental and eastern species to provide context for comparison, with the ice sheet (light blue) at 20 ka. Species colored in gray and color (either pink or green) will be darker when overlapping in range and a lighter, more transparent color when not overlapping. Black spruce (black), white spruce (gray) and red spruce (pink; A); Engelmann spruce (black), subalpine fir (gray) and quaking aspen (green; B); paper birch (black), tamarack (gray) and balsam fir (green; C); jack pine (black) and red pine (gray; D); white oak (black), post oak (gray) and bur oak (green), with an illustration of desert (dark yellow), steppe (light yellow) and tropical (blue) classes in Mexico (E); and southern pines (black), American beech (gray) and eastern cottonwood (green; F). [Color figure can be viewed at wileyonlinelibrary.com]

thus, temperature was high enough for black spruce to be present in any ice sheet gaps or discontinuities. According to these modeled distributions, most boreal species may have extended as far south as 32°N, answering the question of how far south white and black spruce could occur based on climate alone (Jackson et al., 2000). In agreement with Jackson et al. (2000), even though jack pine only extended as far south as 34°N, red pine reached 32°N, with an extension past 30°N in Florida. Even though boreal species distributions arced away from the southernmost extents in a curving trajectory, the arcs were gradual, resulting in coverage across the northern southeastern US, instead of confinement to pollen study locations in the Lower Mississippi Alluvial Valley (Jackson et al., 2000).

Although a few spruce cones assigned to white spruce have been located south of the modeled white spruce tree distributions (Watts 1980), these cones were probably from the extinct *Picea critchfieldii*, which had a southern rather than boreal distribution (Jackson et al., 2000; Grimm & Jacobson 2003). Red spruce tree distributions extended into eastern Texas and northern Florida, conforming with *Picea* pollen detected in both Florida (Watts et al., 1992) and Texas (Cordova & Johnson 2019). However, Cordova and Johnson (2019) stated that *Picea* presence in Texas was questionable and *Picea* pollen was probably due to long-distance transportation by strong winds.

Regarding temperate species distributions at 20 ka, a diverse number of oak and temperate species occur, but a few species may have extended to about 39°N, at least in low-elevation locations. Glaciation ranged from 39°N to 42°N, with tundra commonly south of the glaciers and then *Picea–Abies–Pinus* (Jackson et al., 2000). At 20 ka, *Quercus* and *Ostrya* pollen were present at 37°N (Liu et al., 2013) and temperate tree species were present at 38°N (Baxstrom 2019). The modeled oak species and American beech generally extended as far north as 35°N, with bur oak reaching 38°N in the interior and beech reaching up to 39°N along the low-elevation Atlantic Coast. White oak and beech had continuous distributions across the southeastern US, while post oak distributions were nearly continuous, and bur oak and eastern cottonwood displayed patchy, scattered distributions, corresponding to the spectrum of possibilities suggested by Jackson et al. (2000). The models indicated that southern pine species could have extended from Florida to Texas (Noss 2013).

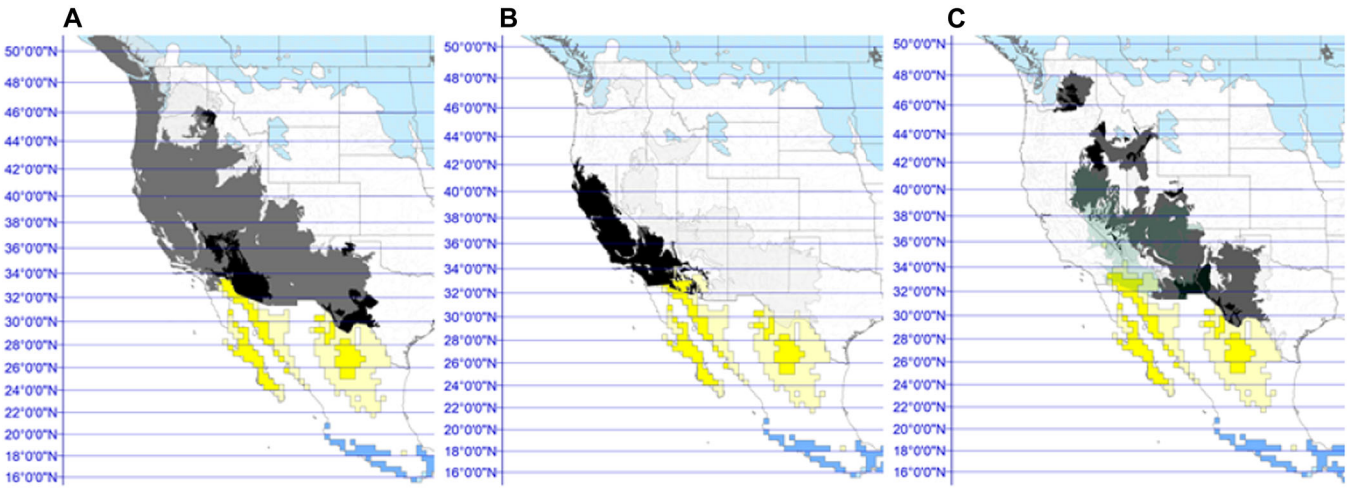

Figure 6. Combined potential distributions of western species to provide context for comparison, with illustration of desert (dark yellow), steppe (light yellow) and tropical (blue) classes in Mexico and the ice sheet (light blue) at 20 ka. Species colored in gray and green color will be darker when overlapping in range and a lighter, more transparent color when not overlapping. Ponderosa pine (black) and Douglas-fir (gray; A); western oaks (black) and Gambel oak (gray; B); and two-needle pinyon pine (black), Rocky Mountain juniper (gray) and Utah juniper (green; C). [Color figure can be viewed at wileyonlinelibrary.com]

Table 2. Accuracy of models using all variables compared to two variables identified as the most important by the all-variable model and two variables identified as intermediate in importance, for four model runs, current climate (1960–1989), and 25 tree species in Canada and the USA.

Variables	Sensitivity	Specificity
All variables	0.9613	0.9153
All variables	0.9607	0.9153
All variables	0.9511	0.9254
All variables	0.9462	0.9178
Mean	0.9548	0.9185
Two foremost variables	0.9101	0.8703
Two foremost variables	0.9076	0.8745
Two foremost variables	0.9211	0.8728
Two foremost variables	0.9057	0.8600
Mean	0.9111	0.8694
Two middle variables	0.8693	0.8151
Two middle variables	0.8484	0.8014
Two middle variables	0.8452	0.8143
Two middle variables	0.8698	0.8101
Mean	0.8582	0.8103

Table 3. Ranked sums of importance values for climate variables under current climate (1960–1989) and four past climates for four models of 25 tree species in Canada and the USA.

Variable	Current	5 ka	7 ka	10 ka	13 ka	Mean
Vapor pressure	1	1	1	1	1	1
Maximum temperature	2.25	2	2	2	2	2.05
Aridity index	2.75	4	3.75	4.5	3.75	3.75
Monthly minimum temperature	4.5	4	4.5	3	4.5	4.1
Moisture index	4.5	4	3.75	5.5	3.75	4.3
Precipitation	8	6.5	6	5	6	6.3
Minimum temperature	8.5	7.25	7.75	7	7.75	7.65
Monthly maximum temperature	6.25	8.5	8.75	8.25	8.75	8.1
Winter precipitation	9.25	8.5	8	10.25	8	8.8
Growing degree days 0	8.5	11.5	10.75	9.25	10.75	10.15
Precipitation variation	12	10.25	10.5	10.5	10.5	10.75
Growing degree days 5	10.5	10.5	11.25	11.75	11.25	11.05

Likewise, rates of distribution shifts were comparable to rates reconstructed for pollen by genera, ranging from 20 to 200 m per year, depending on the species, during the period 20–5 ka. In agreement with greatest areal expansion, boreal species shifted the greatest distance. In contrast, western species shifted the least, perhaps due to less influence from ice sheets and cooling and greater influence of orography. Fastest rates synchronized with fastest glacier recession during 14–7 ka, while rates slowed after 7 ka, similar to slowing rates after 8 ka determined from pollen reconstructions (Dyke, 2005; Ordonez & Williams 2013). However, species distributions between 10 and 5 ka included northern advancement by 7 ka and southern retreat by 5 ka; therefore, distances and rates would increase if centroid points for more intervals were included. The leading edge and trailing edge reversal between 7 and 5 ka exhibited by species distributions, whether boreal or temperate, accorded with the retreat of the northern limit of temperate forests southwards, as climate cooled after reaching a mid-Holocene maximum (Dyke, 2005).

Concurring with pollen data, the tree species distribution models showed poleward expansion with replacement of boreal species by temperate species under a warming climate. Climate for boreal species was no longer suitable in most of the US by 5 ka when modeled by climate alone, similar to current distributions, albeit it may take a century or more for species to first respond to climate (Williams et al., 2002). It is not clear if current distributions of boreal species reflect climate constraints of physiological ranges, as maximum temperature was an influential variable for boreal species, or reveal ecological gradients in competitive ability and fitness (Loehle 1998). A warmer climate may potentially be fatal to seedlings but may produce only phenological mismatches and reduced growth and reproduction in established adult trees rather than direct mortality (Aitken et al., 2008).

Published 2023. This article is a U.S. Government work and is in the public domain in the USA. *Journal of Quaternary Science* published by John Wiley & Sons Ltd on behalf of Quaternary Research Association.

Caveats and correlation

Although current tree distributions are comprehensively sampled to the species level with limited bias, unlike pollen data, species distributions represent complex responses to factors beyond climate, and no-analog climates may produce novel species responses that are different from observed records under current climate (Williams et al., 2014). Therefore, the typical caveats apply regarding modeled climate and species distribution models, along with additional uncertainty due to novelties of past vegetation and influences of human activities. The simulated global temperature progression matches paleoclimate reconstructions from Greenland and Antarctic ice cores (Liu et al., 2009). While modeled climate data captured major trends of paleoclimate reconstructions, climate may not be accurate and downscaling may add additional errors. Potential tree distributions based on climate and observed records are not the same as realized distributions, which are reduced by non-climate factors, such as other species, barriers to dispersal or disturbances (Seliger et al., 2021). Conversely, species may have greater climate tolerances than identified from observed records because when non-climate constraints are added or removed, species distributions may expand (Schweger et al., 2011; Kujawa et al., 2016; Goring & Williams 2017; Hanberry 2021). Furthermore, during the rapid climate change from 16 to 11 ka, realized climate space based on observed pollen records represented shifting subsets of the climate conditions over time (Ordonez 2013). Samples of current realized distributions were coarse, encompassing entire ecological units rather than point locations, to help match the climate space and also account for the uncertainty due to current range changes of historically important tree species due to land-use change (Kujawa et al., 2016; Goring & Williams 2017; Hanberry 2022). Species distribution models are correlational and climate variables may replicate tree distributions rather than physiological limitations. For example, American beech now is an eastern species but had a continental distribution before the glacial period (Tubbs & Houston 1990).

Model robustness is indicated by high accuracies of models and alignment of predictions with the pollen research. Because of correlation, (i) models improved in accuracy with the information of additional variables, (ii) models identified variables that contributed more to accuracy, and (iii) models remained similar under different climates, at least when appropriately correlated climate variables are modeled with the non-linear random forests regressor. Comparisons with models of correlated variables during past climates demonstrated the ability of models to distinguish the influence of redundant, highly correlated predictors. Multicollinearity is typically not an issue in terms of isolation of relevant predictors for tree-based models, which evaluate only a subset of variables in each decision tree and select only one variable at each rule split in the tree. Moreover, if a concern is hindcasting to the past, derivatives of temperature and precipitation were correlated in the past and will be correlated in the future because climate variables reflect order in the environment. Indeed, this strong spatiotemporal structure makes climate more dependable than other predictors for forecasting or hindcasting, due to relatively long-standing species associations with climate over time, which are supported by pollen reconstructions, as opposed to novel or spurious associations of species with other predictors (Chatterjee et al., 2000).

Conclusions

Here, I have modeled current tree species distributions and current climate to generate potential distributions based on paleoclimate. These species distribution models were able to demonstrate unique species-specific distributions and movement trajectories in response to past climate change. Species distribution models concurred with pollen reconstructions in terms of the timing, rate and general response of change while also generating species-specific outcomes regarding the southern extents of boreal species and northern extents of temperate species and distribution continuities. The correlation assessment showed that the non-linear random forests classifier could differentiate important correlated climate variables, which remained consistent over time and increased accuracy compared to omission of climate variables. Although these models may require consideration and refinement based on other evidence, species-specific models may provide valuable supplemental information that cannot be derived from pollen data.

Acknowledgements. I thank the reviewers. This research was supported by the USDA Forest Service, Rocky Mountain Research Station. The findings and conclusions in this publication are those of the author and should not be construed to represent any official USDA or US Government determination or policy. The author has no conflicts of interest.

Data availability statement

Modeled tree distributions are available at https://doi.org/10.5281/zenodo.7278358. Publicly available datasets were analyzed in this study. These data were derived from the following resources available in the public domain: Canada National Forest Inventory, https://doi.org/10.23687/ec9e2659-1c29-4ddb-87a2-6aced147a990. FIA DataMart, https://apps.fs.usda.gov/fia/datamart/datamart.html, Dryad Dataset, https://doi.org/10.5061/dryad.1597g.

Author contributions—BBH performed all authorship tasks.

Abbreviations. °C, degrees celsius; ka, years ago; km, kilometer; m, meter.

References

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

Bartlein, P.J., Harrison, S.P., Brewer, S., Connor, S., Davis, B.A.S., Gajewski, K. et al. (2011) Pollen-based continental climate reconstructions at 6 and 21 ka: a global synthesis. *Climate Dynamics*, 37, 775–802.

Baxstrom, K.W. (2019). Climate and vegetation change in late Pleistocene Central Appalachia: Evidence from stalagmites and lake cores. Thesis, Ohio University, Athens, Ohio.

Beaudoin, A., Bernier, P.Y., Villemaire, P., Guindon, L. & Guo, X.J. (2017). Species composition, forest properties and land cover types across Canada's forests at 250m resolution for 2001 and 2011. Natural Resources Canada, Canadian Forest Service, Laurentian Forestry Centre, Quebec, Canada. Available at: https://doi.org/10.23687/ec9e2659-1c29-4ddb-87a2-6aced147a990 [Accessed on 17 February 2021].

Blois, J.L., Williams, J.W., Grimm, E.C., Jackson, S.T. & Graham, R.W. (2011) A methodological framework for assessing and reducing temporal uncertainty in paleovegetation mapping from late-Quaternary pollen records. *Quaternary Science Reviews*, 30, 1926–1939.

Blois, J.L., Williams, J.W., Fitzpatrick, M.C., Ferrier, S., Veloz, S.D., He, F. et al. (2013) Modeling the climatic drivers of spatial patterns in vegetation composition since the Last Glacial Maximum. *Ecography*, 36, 460–473.

Chatterjee, S., Hadi, A.S. & Price, B. (2000) *Regression Analysis by Example*, Third ed. John Wiley and Sons. ISBN 978-0-471-31946-7. Available at: https://archive.org/details/regressionanalys0000chat_q4i3/page/234/mode/2up?q=multicollinearity. [Accessed 19 March 2023].

Chaput, M. 2018. Environmental change and population history of North America from the late Pleistocene to the Anthropocene. Dissertation, University of Ottawa, Ottawa, Ontario.

Clark, P.U., Dyke, A.S., Shakun, J.D., Carlson, A.E., Clark, J., Wohlfarth, B. et al. (2009) The last glacial maximum. *Science*, 325, 710–714.

Cordova, C.E. & Johnson, W.C. (2019) An 18 ka to present pollen-and phytolith-based vegetation reconstruction from Hall's Cave, south-central Texas, USA. *Quaternary Research*, 92, 497–518.

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

Dawson, A., Paciorek, C.J., McLachlan, J.S., Goring, S., Williams, J.W. & Jackson, S.T. (2016) Quantifying pollen-vegetation relationships to reconstruct ancient forests using 19th-century forest composition and pollen data. *Quaternary Science Reviews*, 137, 156–175.

Delcourt, P.A. & Delcourt, H.R. (1987) Late-Quaternary dynamics of temperate forests: Applications of paleoecology to issues of global environmental change. *Quaternary Science Reviews*, 6, 129–146.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G. et al. (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36, 27–46.

Dyke, A.S. (2005) Paleoenvironments. *Géographie physique et Quaternaire*, 59, 211–262.

Forest Inventory and Analysis DataMart [FIA]. (2021) Available online: https://apps.fs.usda.gov/fia/datamart/datamart.html [Accessed on 30 January 2021].

Goldberger, A.S. & Goldberger, A.S.G. (1991) *A course in econometrics*. Harvard University Press. Available at: https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnxlY29ub21ldHJpY3NpdGFtfGd4OjY5NmZiMDEwMzMwOTQyZTk [Accessed 19 March 2023].

Goring, S.J. & Williams, J.W. (2017) Effect of historical land-use and climate change on tree-climate relationships in the upper Midwestern United States. *Ecology Letters*, 20, 461–470.

Government of Canada. (2022). Terrestrial Ecodistricts of Canada. Available at: https://open.canada.ca/data/en/dataset/ [Accessed on 17 February 2021].

Grimm, E.C., & Jacobson, Jr., G.L. (2003) Late-Quaternary vegetation history of the eastern United States. *Developments in Quaternary Sciences*, 1, 381–402.

Hanberry, B.B. (2021) Timing of tree density increases, influence of climate change, and a land use proxy for tree density increases in the eastern United States. *Land*, 10, 1121.

Hanberry, B.B. (2022) Climate envelopes do not reflect tree dynamics after Euro-American settlement in eastern North America. *Land*, 11, 1536.

Hanberry, B.B., Palik, B.J. & He, H.S. (2013) Winning and losing tree species of reassembly in Minnesota's mixed and broadleaf forests. *PLoS One*, 8, e61709.

Jackson, S.T., Webb, R.S., Anderson, K.H., Overpeck, J.T., Webb, III, T., Williams, J.W. et al. (2000) Vegetation and environment in eastern North America during the Last Glacial Maximum. *Quaternary Science Reviews*, 19, 489–508.

Keys, J.E., Cleland, D.T. & McNab, W.H. (2007). Delineation, peer review, and refinement of subregions of the conterminous United States; Gen. Tech. Report WO-76A; Department of Agriculture, Forest Service: Washington, DC, USA.

Khaledian, Y. & Miller, B.A. (2020) Selecting appropriate machine learning methods for digital soil mapping. *Applied Mathematical Modelling*, 81, 401–418.

Kuhn, M. (2008) Building predictive models in R using the caret package. *Journal of Statistical Software*, 28, 1–26.

Kuhn, M. & Johnson, K. (2019) *Feature engineering and selection: A practical approach for predictive models*. CRC Press. Available at: https://bookdown.org/max/FES/feature-selection-simulation.html [Accessed 6 July 2022].

Kujawa, E.R., Goring, S., Dawson, A., Calcote, R., Grimm, E.C., Hotchkiss, S.C. et al. (2016) The effects of anthropogenic land cover change on pollen-vegetation relationships in the American Midwest. *Anthropocene*, 15, 60–71.

Lindner, T., Puck, J. & Verbeke, A. (2020) Misconceptions about multicollinearity in international business research: Identification, consequences, and remedies. *Journal of International Business Studies*, 51, 283–298.

Liu, Z., Otto-Bliesner, B.L., He, F., Brady, E.C., Tomas, R., Clark, P.U. et al. (2009) Transient simulation of last deglaciation with a new mechanism for Bølling-Allerød warming. *Science*, 325, 310–304.

Liu, Y., Andersen, J.J., Williams, J.W. & Jackson, S.T. (2013) Vegetation history in central Kentucky and Tennessee (USA) during the last glacial and deglacial periods. *Quaternary Research*, 79, 189–198.

Loehle, C. (1998) Height growth rate tradeoffs determine northern and southern range limits for trees. *Journal of Biogeography*, 25, 735–742.

Lorenz, D.J., Nieto-Lugilde, D., Blois, J.L., Fitzpatrick, M.C. & Williams, J.W. (2016a) Downscaled and debiased climate simulations for North America from 21,000 years ago to 2100AD. *Scientific Data*, 3, 160048.

Lorenz, D.J., Nieto-Lugilde, D., Blois, J.L., Fitzpatrick, M.C. & Williams, J.W. (2016b). Data from: Downscaled and debiased climate simulations for North America from 21,000 years ago to 2100 AD. Available at: https://doi.org/10.5061/dryad.1597g [Accessed on 1 December 2020].

Moreno-Amat, E., Rubiales, J.M., Morales-Molino, C. & García-Amorena, I. (2017) Incorporating plant fossil data into species distribution models is not straightforward: pitfalls and possible solutions. *Quaternary Science Reviews*, 170, 56–68.

Mottl, O., Flantua, S.G.A., Bhatta, K.P., Felde, V.A., Giesecke, T., Goring, S. et al. (2021) Global acceleration in rates of vegetation change over the past 18,000 years. *Science*, 372, 860–864.

Noss, R.F. (2013) *Forgotten grasslands of the South*. Washington, D.C.: Island Press. p. 320.

Ordonez, A. (2013) Realized climatic niche of North American plant taxa lagged behind climate during the end of the Pleistocene. *American Journal of Botany*, 100, 1255–1265.

Ordonez, A. & Williams, J.W. (2013) Climatic and biotic velocities for woody taxa distributions over the last 16,000 years in eastern North America. *Ecology Letters*, 16, 773–781.

Pagani, M., Liu, Z., LaRiviere, J. & Ravelo, A.C. (2010) High Earth-system climate sensitivity determined from Pliocene carbon dioxide concentrations. *Nature Geoscience*, 3, 27–30.

R Core Team. (2021). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria.

Schweger, C., Froese, D., White, J.M. & Westgate, J.A. (2011) Pre-glacial and interglacial pollen records over the last 3 Ma from northwest Canada: Why do Holocene forests differ from those of previous interglaciations? *Quaternary Science Reviews*, 30, 2124–2133.

Shmueli, G. (2010) To explain or to predict? *Statistical Science*, 25, 289–310.

Seliger, B.J., McGill, B.J., Svenning, J.C. & Gill, J.L. (2021) Widespread underfilling of the potential ranges of North American trees. *Journal of Biogeography*, 48, 359–371.

Tomaschek, F., Hendrix, P. & Baayen, R.H. (2018) Strategies for addressing collinearity in multivariate linguistic data. *Journal of phonetics*, 71, 249–267.

Tubbs, C.H. & Houston, D.R. (1990) American Beech. In: Burns Russell, M. & Honkala, B.arbaraH., editors. *Silvics of North America: 2. Hardwoods. Agriculture Handbook 654*. Washington, DC: U.S. Department of Agriculture, Forest Service. p. 877. Available at: https://www.srs.fs.usda.gov/pubs/misc/ag_654/volume_2/fagus/grandifolia.htm [Accessed on 21 May 2021].

Vatcheva, K.P., Lee, M., McCormick, J.B. & Rahbar, M.H. (2016) Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology (Sunnyvale Calif.)*, 6(2), 227. https://doi.org/10.4172/2161-1165.1000227

Vaughan, T.S. & Berry, K.E. (2005) Using Monte Carlo techniques to demonstrate the meaning and implications of multicollinearity. *Journal of Statistics Education*, 13, 1.

Vizcaíno, M., Mikolajewicz, U., Gröger, M., Maier-Reimer, E., Schurgers, G. & Winguth, A.M.E. (2008) Long-term ice sheet–climate interactions under anthropogenic greenhouse forcing simulated with a complex Earth System Model. *Climate Dynamics*, 31, 665–690.

Watts, W.A. (1980) The late Quaternary vegetation history of the southeastern United States. *Annual Review of Ecology and Systematics*, 11, 387–409.

Watts, W.A., Hansen, B.C.S. & Grimm, E.C. (1992) Camel Lake: A 40 000-yr record of vegetational and forest history from northwest Florida. *Ecology*, 73, 1056–1066.

Webb, III, T., Anderson, K.H., Bartlein, P.J. & Webb, R.S. (1998) Late quaternary climate change in eastern North America. *Quaternary Science Reviews*, 17, 587–606.

Williams, J.W., Post*, D.M., Cwynar, L.C., Lotter, A.F. & Levesque, A.J. (2002) Rapid and widespread vegetation responses to past climate change in the North Atlantic region. *Geology*, 30, 971–974.

Williams, J.W., Shuman, B.N., Webb, III, T., Bartlein, P.J. & Leduc, P.L. (2004) Late-Quaternary vegetation dynamics in North America: Scaling from taxa to biomes. *Ecological Monographs*, 74, 309–334.

Williams, M.N., Grajales, C.A.G. & Kurkiewicz, D. (2013) Assumptions of multiple regression: Correcting two misconceptions. *Practical Assessment Research and Evaluation*, 18, 11.

Zhang, W., Wu, C., Zhong, H., Li, Y. & Wang, L. (2021) Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. *Geoscience Frontiers*, 12, 469–477.