Ecological Informatics 75 (2023) 102110

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Effectiveness of land use and disturbance measures, compared to climate, soil, and topography, for modeling relative abundances of tree species in the eastern United States

Brice B. Hanberry

USDA Forest Service, Rocky Mountain Research Station, Rapid City, SD 57702, USA

ARTICLE INFO

Keywords:
Agriculture
Drivers of change
Fire
Species abundance model

ABSTRACT

Environmental factors control species distributions and abundances, but effectiveness of land use and disturbance variables for modeling species generally is unknown compared to climate, soil, and topography variables. Therefore, I used predictor variables from categories of 1) land use and disturbance, 2) climate, and 3) soil, topography, and wind speed to model the relative abundances (i.e., percentage of all trees) of 65 common tree species in the eastern United States, with a contrast to presence-absence models of species distributions. First, I modeled variables within each category to identify the five most important variables. Then, I combined variables from each category to isolate most important variables, based on five model combinations of input variables from each category, ranging from one (i.e., three total) to five (i.e., 15 total) variables. From the five models of combined categories for each tree species, I identified the model with the greatest R^2 value. Overall, climate variables were most important for tree species models with one and two input variables from each category, but land use and disturbance variables were most important for models with three to five input variables from each category. Although a range of R^2 values occurred by species and number of input model variables, 32 species had best models with greatest R^2 values of 0.50 to 0.81. For all best species models, the most important variables were temperature of the warmest quarter, historical fire return interval for all fires, agricultural area during years 1850 to 1997, and precipitation of the driest month. Current land cover classes, which are accessible and the most commonly modeled land use variables, were not important for modeling tree species abundances or distributions. Climate variables were most important for modeling species distributions. Results support the concept that while climate sets soft boundaries on distributions, relative abundances within distributions are affected by other filters. Future modeling may establish other important land use and disturbance variables, or refinements within the important variables of historical fire return interval and agricultural area over time, advancing integration of both land use and climate variables into studies.

1. Introduction

Species have a unique set of life history traits in terms of survival, growth, and reproduction response to influential factors in the environment, such as climate and disturbance. Species do not contain all traits and cannot invest unlimited resources to all components at once; thus, trade-offs occur based on constraints. In response to climate, cold tolerance and growth rate traits generally are inversely related (Loehle, 1998). Species of colder climate need to survive cold exposure and freeze damage. In contrast, species in locations with long growing seasons and greater productivity typically can devote resources to maximizing growth. Nonetheless, they contend with greater competition to acquire resources for growth and other biological interactions, such as from insects and disease with increased over-winter survival and accelerated growth from a greater number of growing degree days (Deutsch et al., 2008). In response to disturbance, species adapted to frequent low severity disturbance have traits to survive under chronic stress, often through early belowground growth, whereas species adapted to relatively frequent severe disturbance devote resources to rapid growth at least until quickly reaching reproduction age, at which point, they may switch from fast growth to prolific production (Grime, 1977; Hanberry, 2019). Species under similar climate or disturbance regimes will share similar traits in response to those filters, but differences among species allow coexistence and fine spatiotemporal

E-mail address: brice.hanberry@usda.gov.

https://doi.org/10.1016/j.ecoinf.2023.102110
Received 20 September 2022; Received in revised form 19 April 2023; Accepted 21 April 2023
Available online 24 April 2023

partitioning under variation of the regimes or when multiple filters occur, either in parallel or compounded in series (Gagnon et al., 2021).

Abundance and distribution of species reflect the legacy of different climate and disturbance filters on the ability of species to compete for resources, through successful survival, growth, and reproduction. Densities and ranges of tree species have changed during recent centuries in the United States, following the front of Euro-American settlement (Motzkin et al., 1999). Recent disruptions in disturbance through land use change, including surface fire exclusion, correspond with tree density and distribution changes based on timing, magnitude and direction of change, and mechanism (Hanberry, 2021). Hundreds of tree species occur in the eastern United States, and after no longer needing traits to tolerate frequent surface fire as a filter, other filters become more influential. Species that respond to relatively frequent severe disturbances are likely to be successful due to chronic severe disturbance from land use in the eastern U.S., particularly in the frequently harvested southeastern U.S., where loblolly pine (*Pinus taeda*) plantations are widespread (Hanberry, 2020; Pan et al., 2011). The outcome of land use and disturbance change is unstable tree distributions (García-Valdés et al., 2013), with increased abundance of historically rare native tree species at the expense of foundational ecosystems comprised predominantly of surface fire-dependent oak or pine in the eastern U.S. (75% pine or pine-oak in the southeastern U.S. and about 50% oak in the central eastern U.S.; Hanberry et al., 2020). Historical accounts and tree surveys since year 1635 provide evidence that rapid tree growth and expansion occurred trailing the front of Euro-American land use and disturbance change (Davidson, 1840; Sargent, 1884; Leiberg, 1902; Day, 1953). Many of the tree species changes involved loss of spatial partitioning within regions, in wetlands or elevational bands, with consequent tree density increases in uplands or in lower elevations where surface fire historically was frequent. However, tree species also have expanded and increased outside of forested regions, such as *Acer saccharinum*, *Acer negundo*, *Celtis occidentalis*, *Fraxinus pennsylvanica*, *Gleditsia triacanthos*, *Juglans nigra*, *Juniperus virginiana*, *Maclura pomifera*, *Morus rubra*, and *Prunus serotina* that have expanded and increased in the grasslands of the central U.S., which was maintained by fire (Hanberry, 2021; Hanberry et al., 2014).

Unlike land use and disturbance change, climate generally has remained within the historical range of variation throughout the eastern and central U.S, albeit on a trajectory to novel temperature and precipitation patterns (Hanberry et al., 2020; Partridge et al., 2018; Rojas et al., 2019). Temperature generally has not warmed during the 1900s in the eastern U.S., due to a cooling trend that initiated during the late 1950s ('the warming hole'; Partridge et al., 2018). Even in the northeastern U.S., a region where precipitation has increased (Fig. 1; PRISM Climate Group, 2021), a shift in trends to a wetter climate only occurred during 2002 (Huang et al., 2017). Drought severity index values, which measure moisture conditions available to trees based on water balance, during the 1900s were not outside the range of historical variation during the past 1000 years for the eastern U.S., based on change point detection (Hanberry et al., 2020).

Soil and topography affect tree species distributions by influencing competition among species and also by amplifying or dampening effects of climate and disturbance though interactions. Topography in particular modulates site conditions (Mod et al., 2016). Nevertheless, unlike lethal direct effects of cold or fire, many tree species can grow in most soils and topographic sites. Tree species associations with soil and topography are very malleable and in many cases, associations that currently appear strong have changed since Euro-American settlement (Hanberry et al., 2013, 2014). Species have moved up- and downslope in elevation (e.g., firs, spruces), from lowlands to uplands (diverse eastern tree species), and from dry to wet sites (e.g., junipers; Hanberry et al., 2014). For example, in Minnesota, tamarack (*Larix laricina*), pines, and oaks have become restricted to sites with either wetter or sandier and drier soils due to increased aspen (*Populus tremuloides*) and fire-sensitive tree species in mesic sites (Hanberry et al., 2013). Wind speed is a component of atmospheric modeling but is similar to soil and topography in interacting with climate and disturbance, albeit wind also can act as a disturbance (e.g., Hanberry and Noss, 2022).

Whether or not climate is directly influential on species or acts indirectly through species interactions, climate variables create an ideal match to species distributions but not abundances (Canham and Thomas, 2010; Chambers et al., 2013). Climate occurs continuously in space and time, forming nearly systematic grids to model distributions, unlike other types of variables. Climate variables are well-modeled from instrumental records and easily available. For example, Worldclim 2, a dataset which supplies bioclimatic variables, has 8000 citations, with thousands of new citations annually (Fick and Hijmans, 2017). Temperature- and water-related variables were included in 177 of 200 plant species distribution studies during years 2010–2015 (Mod et al., 2016). Similarly, climate variables were used in 88% of 190 species distribution studies during 2012–2016 (Fourcade et al., 2018). More than half of the plant species distribution studies modeled with solely temperature and water plus one additional variable (Mod et al., 2016), whereas 85% of

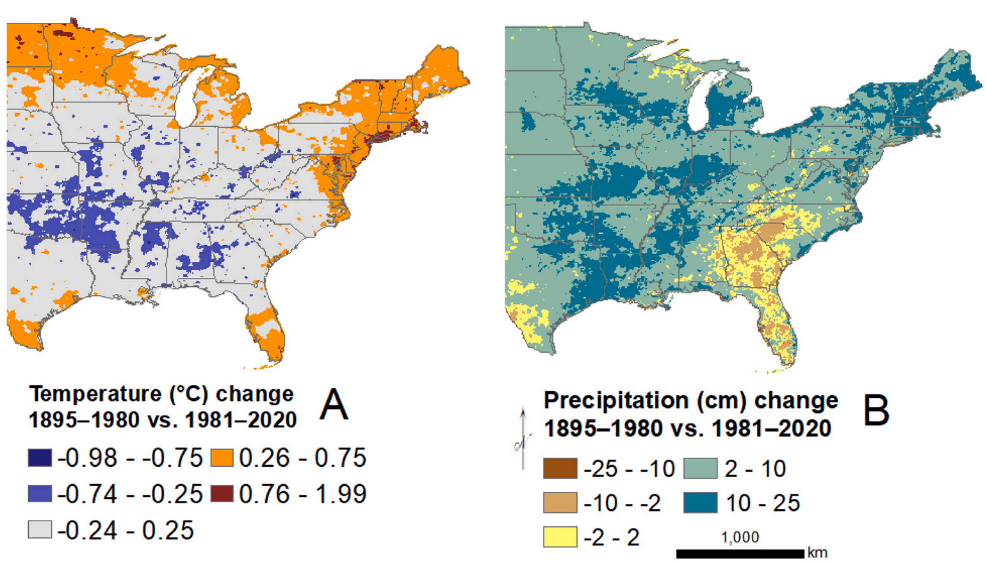

Fig. 1. Study extent of the eastern United States, including changes in mean annual temperature (A) and mean annual total precipitation (cm) during 1981–2020 as compared to values during 1895–1980 (data modified from PRISM Climate Group, 2021).

studies about future biodiversity scenarios during 1990–2014 incorporated only climate change projections (Titeux et al., 2016).

Conversely, land use and disturbance comprise a wide range of disparate factors that are not compiled into a few accessible databases, resulting in lack of information about which measures are available or effective for modeling (Mod et al., 2016). Over time and space, land use and disturbance may be complex to measure, inaccurate, and patchy in distribution. Land use and disturbance are multifaceted, with changes over time and space, resulting in sites and landscapes with unique sequences of land use and disturbance histories (Crimmins et al., 2014; Motzkin et al., 1999). For example, percentage area in agriculture by county is available for the U.S., but this simple metric is not likely to capture historical cycles of agricultural conversion and abandonment in marginal lands that occur even now as crop and ethanol prices fluctuate (Yang et al., 2020). In any event, land use and disturbance variables are less common than climate variables in building plant models (Mod et al., 2016; Sirami et al., 2017; Titeux et al., 2016). Out of 200 plant species distribution studies during years 2010–2015, 17 studies included variables representing such disturbances as fire and grazing, 11 studies included anthropogenic disturbance variables related to population or housing density, roads, agriculture, forest management, ownership status, and non-native species, and 32 studies included current land use and land cover classes based on remote sensing (Mod et al., 2016). Equally, land cover was used in 11% of 190 species distribution studies during 2012–2016 (Fourcade et al., 2018).

Information technology enables modeling of novel, complex data to understand associations between environmental factors and plant distributions and abundances (Recknagel and Michener, 2018). Likely due to data limitations, species distribution models and scenarios seldom include land use and disturbance variables, which typically are limited to current land use and cover classes, relative to climate variables (Crimmins et al., 2014; Fourcade et al., 2018; Mod et al., 2016; Sirami et al., 2017). Indeed, modeling and scenario studies have integrated fewer variables from different categories over time, becoming increasingly imbalanced to focus on climate change alone since the early 2000s (Mod et al., 2016; Sirami et al., 2017; Titeux et al., 2016). To isolate which general land use and disturbance measures may be effective at modeling tree species, I examined the associations of land use or disturbance, climate, and modulating factors of soil, topography, and wind speed with current relative abundance (% of all trees) of 65 common tree species. Because relative abundance requires samples for accuracy, I used relative abundance estimates for plots by ecological subsections in the eastern half of U.S., including part of the central U.S. for this multi-regional, macroecological study (Fig. 2). For modeling, I identified the five most important variables in three separate categories of 1) land use and disturbance, 2) climate, and 3) soil, topography, and wind speed, and then determined variable importance based on modeling five combinations of variables from each category, ranging from one variable per category to five variables per category. In particular, my objective was to isolate important land use and disturbance variables for modeling tree species by assessing how effective land use and disturbance variables were compared to climate variables at modeling the challenging problem of relative abundance of trees (i.e., percent of all trees), not at equilibrium and under nonstationary conditions. Additionally, I contrasted important variables for modeling relative abundance with important variables for modeling distributions,

Fig. 2. Study extent of the eastern United States with 460 ecological subsections.

which have highly accurate associations with climate data, while retaining the same modeling framework for best comparison.

2. Methods

2.1. Data

The USDA Forest Inventory and Analysis, 2021 surveys thousands of forestland long-term plots, which have an area of 0.07 ha, located about every 2500 ha across the U.S. (Bechtold and Patterson, 2005). Plots are surveyed every five to ten years, varying by state, and I used the latest complete cycle of inventories, ranging from years 2004 to 2019 (mean inventory year of 2013). I determined the relative abundance (% of all trees) for trees ≥12.7 cm in diameter (at 1.4 m above ground height), which is an estimate that requires more than one plot sample, as opposed to point locations for species distribution modeling. Therefore, the analysis unit was ecological units of subsections, which are defined by homogenous climate, vegetation, surficial geology, lithology, geomorphic processes, soils, and hydrology (mean area = 632,000 ha, ranging from 22,300 ha to 7,700,000 ha; Fig. 2; Keys et al., 2007). Similar to many multiregional landscape studies, subsamples, or number of plots in ecological subsections, and sampling density, or density of plots relative to forestland area (km^2) in ecological subsections were not even (Supplementary files 1 and 2). Sampling density was least in southeastern forests (0.004 to 0.022 plots per km^2 forestlands), whereas subsections with fewer plots had limited area of forestlands, resulting in greater sampling density (0.027 to 0.81 plots per km^2 forestlands). Sixty-five tree species, or genera, were present in at least 100 subsections out of 460 subsections in the eastern U.S. (Supplementary file 3).

For predictor variables, I determined mean values by ecological subsection for 10 variables from both the land use and disturbance category and climate category and nine variables for a modulating category of soil, topography, and wind speed (Table 1). For spatiotemporal means, I calculated simple means for any repeated measures over time and zonal means by ecological subsection (ESRI, ArcGIS, Redlands, CA). Land use and disturbance changes that initiated during Euro-American settlement influenced tree composition and structure, as described above, and for this initial exploration, I selected general indices, with some covering wide time intervals. Land use and disturbance variables were population density during years 1790 to 2010 (Fang and Jawitz, 2018); agricultural percentage area during years 1850 to 1997 (Maizel et al., 1998); current (year 2016) percentage area in croplands, pasturelands, forestlands, and wildlands (all vegetation types; Homer et al., 2020); current loblolly pine plantations, which are the major plantations in the U.S. with loblolly pine the most dominant species now in the southeastern U.S., represented by loblolly pine basal area percentage area (USDA Forest Inventory and Analysis Evalidator, 2022); historical (pre-Euro-American settlement) fire return intervals for all fires and surface fires (LANDFIRE, 2022a); and vegetation departure since Euro-American settlement (LANDFIRE, 2022b). Climate variables were mean annual temperature, maximum temperature of the hottest month, minimum temperature of the coldest month, mean temperature of the coldest and hottest three months (i.e., quarters), annual precipitation, precipitation during the driest month and driest three months, and precipitation during the wettest month and wettest three months during years 1970 to 2000 (Worldclim 2.1; Fick and Hijmans, 2017). Soil variables were percentage clay, percentage sand, organic matter, pH, water storage, and soil depth (Walkinshaw et al., 2021). Topographic variables were topographic roughness and compound topographic index, which also is known as the topographic wetness index, for quantifying topographic control on water flow (Amatulli et al., 2020). The last variable was wind speed (Global Wind Atlas, 2021) in the soil, topography, and wind speed category.

Table 1
Model results from separate categories of 1) land use and disturbance, 2) climate, and 3) soil, topography, and wind speed; count and sum of importance values for variables in models of 65 tree species in the eastern United States; count of variables for models with importance values >90 and the only variable of models; importance rank and reason.

Variable	Count	Sum	Most important count	One variable count	Rank	Reason
Land use and disturbance						
all fire return	33	2059	15	5	1	high values, many models
agricultural area	29	2049	17	4	2	high values, many models
% forest	23	1079	9	1	3	moderately high values, many models
loblolly plantation	8	600	6	5	4	only variable for five species
% pasture	8	594	4	1	5	equal treatment of all groups
population density	11	587	4	0	6	equal treatment of all groups
surface fire return	15	570	4	0	0	similar to all fire return interval
veg departure	8	469	4	1	0	low values
% crop	17	459	3	0	0	similar to agricultural area, low values
% herbaceous	7	351	2	1	0	low values
Climate						
temp warmest quarter	34	2196	17	1	1	highest values, many models
max temp warmest month	29	1622	12	1	3	high values, many models
min temp coldest month	26	1312	6	0	5	similar cold temperature
temperature	29	1305	9	1	4	unique annual temperature
precip driest month	18	1010	9	1	2	keep precipitation
precip driest quarter	14	968	7	2	0	very similar precipitation
temp coldest quarter	24	943	6	1	6	similar cold temperature
precipitation	10	309	3	2	0	low values
precip wettest month	3	106	1	0	0	low values
precip wettest quarter	1	88	0	0	0	low values
Soil, topography, wind						
organic matter	25	1653	14	4	1	high values, many models
wind speed	23	1554	15	3	2	high values, many models
% clay	21	1173	10	3	3	moderately high values, many models
% sand	20	1051	9	2	4	moderately high values, many models
compound topographic	17	1008	8	1	5	moderately high values, many models
pH	17	694	6	2	6	equal treatment of all groups
soil depth	13	595	5	0	0	low values
topographic roughness	11	481	3	0	0	low values
water storage	8	139	0	0	0	low values

2.2. Correlation issues

Climate variables tend to be correlated because they are ecologically unique variations of either temperature or precipitation. Correlated variables can interfere with measuring the influence of variables. With correlated predictor variables, strong predictors can end up with lower importance values than if only one temperature or precipitation variable was included in modeling. Ensemble, nonlinear models that can determine nonlinear interactions among variables typically are robust in isolating influence of redundant, highly correlated predictors, compared to linear models (Kuhn and Johnson, 2019). That is, ensemble models distinguish correlated strong predictors rather than splitting the explanatory value into intermediate importance. To account for correlation, I applied a nonlinear ensemble method, the random forests regressor. Additionally, with recursive feature elimination, each variable is left out of the model to allow independent selection of variables through backwards stepwise selection. Moreover, at maximum, only five input variables remained for the category comparisons.

2.3. Modeling framework

The basic modeling process reduced the number of model variables to the most important variables and identified R^2 values for models of all retained important variables and the foremost important variable (Figs. 3 and 4). The first step was recursive feature elimination, also known as backwards selection in which variables are ranked and sequentially eliminated, using the random forests nonlinear regressor and 10-fold cross-validation, to reduce number of model variables (Kuhn, 2008; R Core Team, 2021). The remaining variables continued onto another step that applied the random forests nonlinear regressor, which was trained with 10-fold cross-validation, to determine the most important variables with values >40 based on scaled values up to 100. The third step for these variables was determining final importance values again with the random forests nonlinear regressor and 10-fold cross-validation. Then, the same random forests nonlinear regressor and 10-fold cross-validation modeling was applied to determine R^2 values for models of the most important variables and models of the foremost important variable, based on withheld samples that were 25% of the total samples.

To identify the five most important variables from each of the three separate categories (land use and disturbance; climate; and soil, topography and wind speed), I first modeled each of the tree species by category to reduce the initial predictor variables to the five most important variables (Table 1; Fig. 3). With this output, I determined the frequency and mean importance value of the variables and identified the

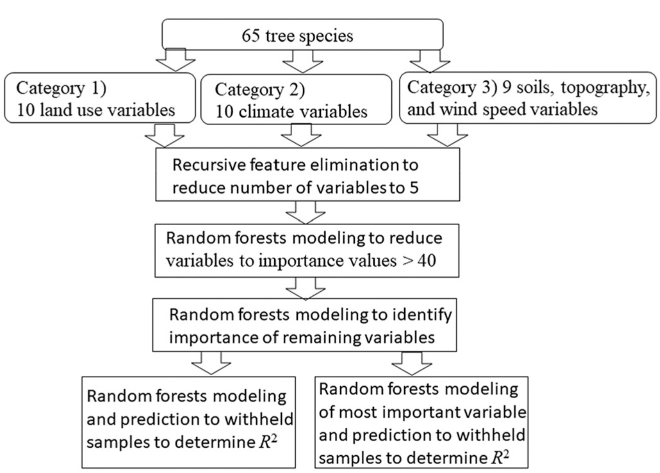

Fig. 3. Modeling steps for separate categories of variables: 1) land use and disturbance, 2) climate, and 3) soil, topography, and wind speed.

foremost important variables (values ≥90) for each tree species and category. As a check to confirm inclusion of the fifth-most important variable and exclusion of sixth-most important variable, I re-ran the random forests models with each of these variables as the fifth variable of five variable models, to isolate the influence of the fifth and sixth importance variables on model accuracy.

For combined important variables from each of the three categories, I identified the most important variables based on five combinations of input variables from each category. That is, each tree species had five models, which consisted of the foremost important variables from each category (i.e., 3 variables total), the most important two variables (i.e., 6 variables total), the most important three variables, the most important four variables, and the most important five variables (Table 2; Fig. 4). I applied recursive feature elimination to first limit models with 12 or 15 variables to half the number of variables. With this output, I determined the frequency and mean importance value of the variables and identified the foremost important variables (values ≥90) for the five models for each tree species. For all the different number of input variables, I determined R^2 values for both the most important variables and the foremost important variable. From the five models for each tree species, I identified the model with the greatest R^2 value.

2.4. Comparison of results to presence and absence models

Regarding contrast of important variables and accuracy of species abundance models to species distribution models, I modeled presence and absence of each tree species while keeping the modeling framework, as described above, fixed for best comparison. I converted any relative abundances of ≥0.5% of all trees in an ecological subsection to the presence class and < 0.5% of all trees as the (pseudo-)absence class, because absences can never be known at landscape scales and < 0.5% allows for some transient outliers that happened to be measured. Classification accuracy measures for withheld samples were generated rather than R^2 values. I used the full 15 variable model of five important variables from each of three categories as variable inputs and compared to the results of the full model of 15 variables from the regression models.

3. Results

3.1. Models from separate categories

Overall, for five variable models of 65 tree species relative abundances from the three separate categories (land use and disturbance; climate; and soil, topography, and wind speed), land use and disturbance variables had similar accuracy to climate variables in predicting to withheld samples (Table 2). Land use and disturbance variables had a mean R^2 value of 0.31 for a mean of 2.45 variables compared to climate variables, which had mean R^2 value of 0.32 for a mean of 2.89 variables. The soil, topography, and wind speed category had weaker associations, resulting in a mean R^2 value of 0.24 for 2.39 variables, with tree species relative abundance.

The five most important variables for 65 tree species from the land use and disturbance category were historical fire return intervals for all fires, agricultural percentage area over time (years 1850 to 1997), current percentage area in forestlands, and pine plantations that was represented by loblolly pine basal area, and percentage area in pasturelands (Table 1). The first two variables were highly important predictors, as measured by use in about 30 species models and summed importance values of >2000 in species models, whereas the third variable was moderately important (i.e., summed importance value of about 1000), and the last two variables had low importance (i.e., summed importance value of ≤600). The five most important variables from the climate category were mean temperature of the hottest quarter, maximum temperature of the hottest month, minimum temperature of the coldest month, annual mean annual temperature, and precipitation during the

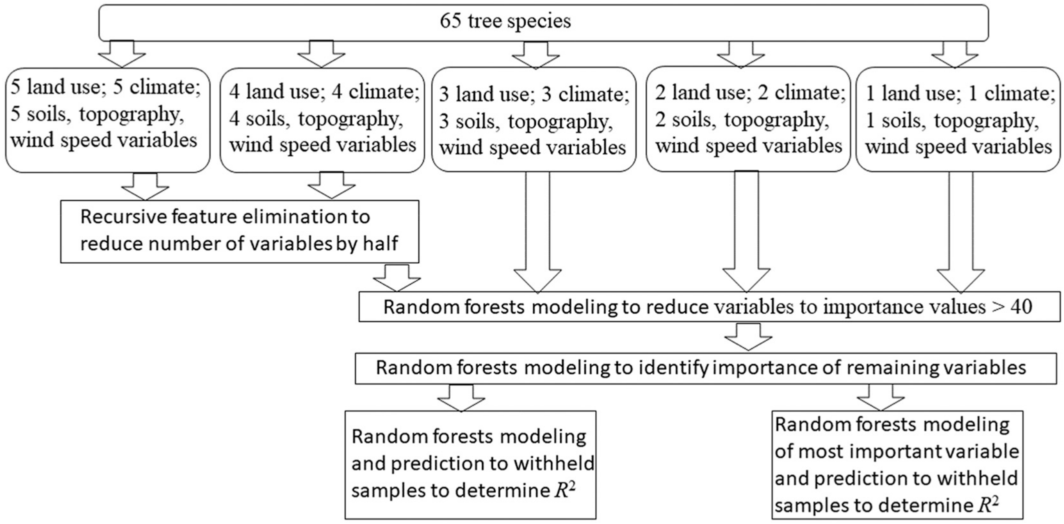

Fig. 4. Modeling steps for combined categories of variables with varying number of input variables.

driest month. All variables were moderately to highly important (i.e., summed importance value ranging from 1000 to 2200). The five most important variables from the soil, topography, and wind speed category were organic matter, wind speed, percentage sand, percentage clay, and compound topographic index. All variables had moderate to moderately high importance (i.e., summed importance value ranging from 1000 to 1700). For each of the three categories, when checking the fifth-most compared to the sixth-most important variables in five variable models, rather than all variable models, the fifth-most important variables continued to provide greater accuracy than the sixth-most important variables.

3.2. Models from combined categories

For all models when the variable categories were combined, the climate variables overall were most important for tree species models with one and two variables from each category, and then became increasingly less important as number of input variables increased to five from each category (Table 2). Land use and disturbance variables were the most important for models with three to five variables from each category. Percentage sand was the third-most important variable for models with five variables from each category, albeit the modulating soil, topography, and wind speed category cumulatively was least important. The values for R^2 increased with number of variables from 0.23 for the three input variable models, for a mean of 1.54 retained variables, to 0.40 with 15 input variable models, for a mean of 2.95 retained variables. Models had a range of R^2 values depending on species and type and number of input predictor variables.

From the five models for each of 65 tree species, I identified the model with the greatest R^2 value (Fig. 5). Out of the best models, 32 species had models with greatest R^2 values of 0.5 or greater, ranging up to 0.81 (Fig. 6; Supplementary file 3). The land use and disturbance category and climate category each were the foremost important variables in models for 25 species, whereas soil variables were the foremost important variable for 11 species and wind speed was the foremost important variable for four species. Cumulatively for all the best species models, the most important variables were temperature of the warmest quarter (climate), historical fire return interval of all fires (land use and disturbance), agricultural area during 1850 to 1997 (land use and disturbance), and precipitation of the driest month (climate). Organic matter and wind speed variables of the soil, topography, and wind speed category were the next two most important variables, but at about half the sum of the importance values and number of species models with the variables. For best models, percentage agricultural area over time (1850 to 1997) generally had greater values where species were present than absent and mean historical fire return interval of all fires generally had greater values where species were (pseudo-)absent than present (Fig. 7; Supplementary file 4). However, when mean fire return interval was the most important variable for species models, mean fire return interval had both negative and positive relationships with species.

Precipitation of the driest month was important for the combined category models but may be less important than it appears for the eastern U.S., which is humid with a moderate range of precipitation. For the initial modeling of the climate set of variables, temperature of the warmest quarter, maximum temperature of hottest month, minimum temperature of the coldest month, and mean temperature were more important than precipitation (Table 1). I ranked precipitation of the driest month as the second-most important variable to allow the models to extract maximum information from climate because precipitation has an east-west gradient whereas temperature has a north-south gradient. But higher ranked variables had more opportunities to exert more influence, particularly the three foremost variables, one from each category, that were inputs in all five models.

3.3. Best species models that were single category models

Most species had best models with variables from two or three categories, but single category models did occur. Climate-only variables were most important for greatest R^2 value for models of seven species: American basswood (*Tilia americana*), blackgum (*Nyssa sylvatica*), eastern redbud (*Cercis canadensis*), eastern white pine (*Pinus strobus*), northern red oak (*Quercus rubra*), quaking aspen (*Populus tremuloides*), and water oak (*Quercus nigra*). By species, land use and disturbance only were most important for models of six species or genera: black locust (*Robinia pseudoacacia*), black walnut (*Juglans nigra*), blackjack oak (*Quercus marilandica*), hawthorn (*Crataegus*), shortleaf pine (*Pinus echinata*), and swamp white oak (*Quercus bicolor*). However, the hawthorn model (R^2 value = 0.11) and shortleaf pine (R^2 value = 0.17) were very weak. No species had greatest R^2 values for one variable models from the soil, topography, and wind speed category.

3.4. Comparison of results to presence and absence models

For the presence and (pseudo-)absence species distribution models, the 65 tree species models had a mean accuracy of 84% (sensitivity = 0.80 and specificity = 0.82) for five variables of each category (Table 2), with a mean number of retained variables of about 3.5. In comparison, the regression models for the 65 tree species models had a mean R^2 value

Table 2
Count and sum of importance values for models with varying number of input variables from categories of 1) land use and disturbance, 2) climate, and 3) soil, topography, and wind speed for 65 tree species in the eastern United States and mean number of variables and R^2 for models.

Number	Variables	Count	Sum	Mean	R^2
1	temp warmest quarter	52	3800	1.54	0.23
1	all fire return	32	1900		
1	organic matter	16	800		
2	temp warmest quarter	44	2181		
2	agriculture	32	1961	2.48	0.36
2	precip driest month	27	1447		
2	all fire return	23	1155		
2	organic matter	15	827		
2	wind	20	693		
3	agriculture	23	1653		
3	temp warmest quarter	25	1360	2.03	0.31
3	all fire return	15	1253		
3	precip driest month	20	1001		
3	organic matter	10	710		
3	max temp warmest month	20	629		
3	clay	9	500		
3	forest	5	427		
3	wind	5	142		
4	agriculture	21	1532	2.34	0.37
4	all fire return	16	1084		
4	min temp coldest month	18	887		
4	forest	11	713		
4	temp warmest quarter	18	674		
4	clay	8	587		
4	precip driest month	13	575		
4	organic matter	7	567		
4	max temp warmest month	19	532		
4	sand	10	528		
4	loblolly	5	478		
4	wind	6	271		
5	agriculture	24	1491	2.95	0.40
5	all fire return	19	1087		
5	sand	16	875		
5	precip driest month	15	729		
5	max temp warmest month	18	728		
5	min temp coldest month	16	664		
5	temperature	19	651		
5	temp warmest quarter	17	569		
5	loblolly	5	472		
5	clay	8	449		
5	organic matter	7	445		
5	forest	9	374		
5	compound topographic	8	350		
5	wind	7	303		
5	pasture	4	200		

of 0.40 for five variables of each category, with a mean number of retained variables of about 3. For the presence and absence species distribution models, mean temperature was the most important variable (sum of importance values = 2116), followed by the other three temperature variables, agricultural area (sum of importance values = 941), and then the precipitation variable. From the soil, topography, and wind speed category, sand and organic matter were the seventh- and eighth-most important variables for predicting species distributions.

4. Discussion

4.1. Key findings

While statistical associations are simply correlations, commonly between species distributions and climate, and not causation, species ranges and species abundances within ranges are the realization of many factors, not only physiological response to climate (Aitken et al., 2008; Early and Sax, 2014; Perret et al., 2019; Van der Veken et al., 2008). However, it is unclear how effective land use and disturbance variables are at modeling species relative abundances compared to other types of variables. For this multi-regional study in the eastern United States, land use and disturbance variables were effective for modeling the challenging problem of relative abundances of tree species, but only historical fire return intervals for all fires and agricultural area during 1850 to 1997 were competitive with climate variables. Climate variables, specifically temperature of the warmest quarter and precipitation of the driest month, were most important for models with one and two input variables from each category, but once the information from climate was gained, the two important land use and disturbance variables (historical fire return intervals for all fires and agricultural area during 1850 to 1997) became the most important for models with three to five input variables from each category. Although the threshold for satisfactory model values depends on the situation, 32 tree species had models with greatest R^2 values of 0.5 or greater, ranging to 0.81. For the best species models, the most important variables were temperature of the warmest quarter, historical fire return interval, agricultural area during years 1850 to 1997, and precipitation of the driest month. Indeed, historical fire return intervals for all fires and agricultural area during 1850 to 1997 were consistently the two most important variables from land use and disturbance category, whereas importance of variables from the climate category and modulating soil, topography, and wind speed category changed depending on number of input variables.

Standard land use and land cover classes are easily available but rare compared to climate as model predictors, similar to soil and topography data, but land use and land cover classes were relatively unimportant for modeling tree species abundances. Notably, current land use and cover classes, which are the most widely applied land use variables for plant species distribution studies (Mod et al., 2016), were not effective variables for modeling tree species relative abundances. Similarly, soil, topography, and wind speed had weaker associations with tree species abundances than the models from either the land use and disturbance category or climate category. For the one category models and best species models from combined categories, soil organic matter and wind speed had relatively low importance compared to the best land use and disturbance variables and climate variables, at least at these coarse resolutions rather than site scales.

The two important land use and disturbance variables of fire return intervals for all fires and agricultural area over time are known mechanisms for affecting species abundances. Both metrics provided historical information, compared to the less effective but most commonly used current land use and land cover classes. Land use change encompasses multiple components, including the actual loss of a vegetation type with replacement by a land use and also disruption in functions and disturbances, such as frequent surface fires (Hanberry, 2021). Frequent (< 30 years) surface fire regimes historically were common in the U.S. Surface fire-sensitive species survived in protected firebreaks, which typically were the moisture extremes of wetlands and rocky outcrops with discontinuous vegetation and also colder northern or mountainous climates where frequent ignitions were unlikely. Fire exclusion occurred concurrently with land use changes, for example, as agricultural fields prevented fire spread through a network of firebreaks (Hanberry, 2021) and created changed competitive dynamics among tree species. Fire-sensitive tree species increased and expanded, replacing fire-tolerant oak and pine species. Despite the importance of fire on species distributions and abundances, Crimmins et al. (2014) found that inclusion of fire occurrence and fire return interval departure variables did not improve species distribution models of 144 vascular plant species in California relative to climate variables, but importantly, climate variables have an ideal spatial structure to map species distributions.

Even though readily available climate data provide an excellent match to presence and absence models of species distributions, predictions of species relative abundances within ranges generally are inaccurate compared to modeling purely ranges. For the full 15 variable models of five input variables from each category, the regression models of relative abundance had a mean R^2 value of 0.40 for the 65 tree species. As a comparison to simple classification models of presence and (pseudo-)absence, the 65 tree species models had a mean accuracy of

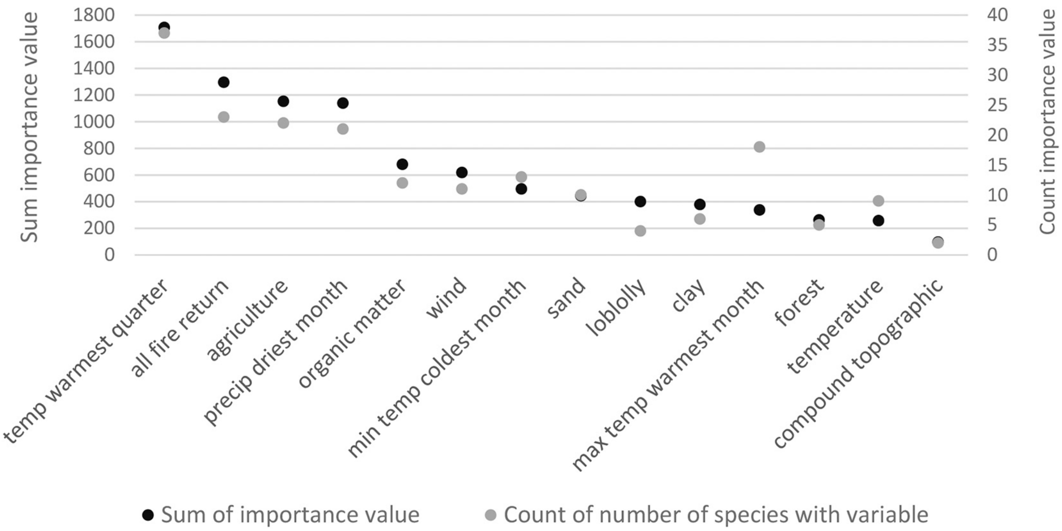

Fig. 5. Sum of importance values for variables and number of models that contain the variable, for 65 species models with greatest R^2 value from the combined categories with varying number of input variables.

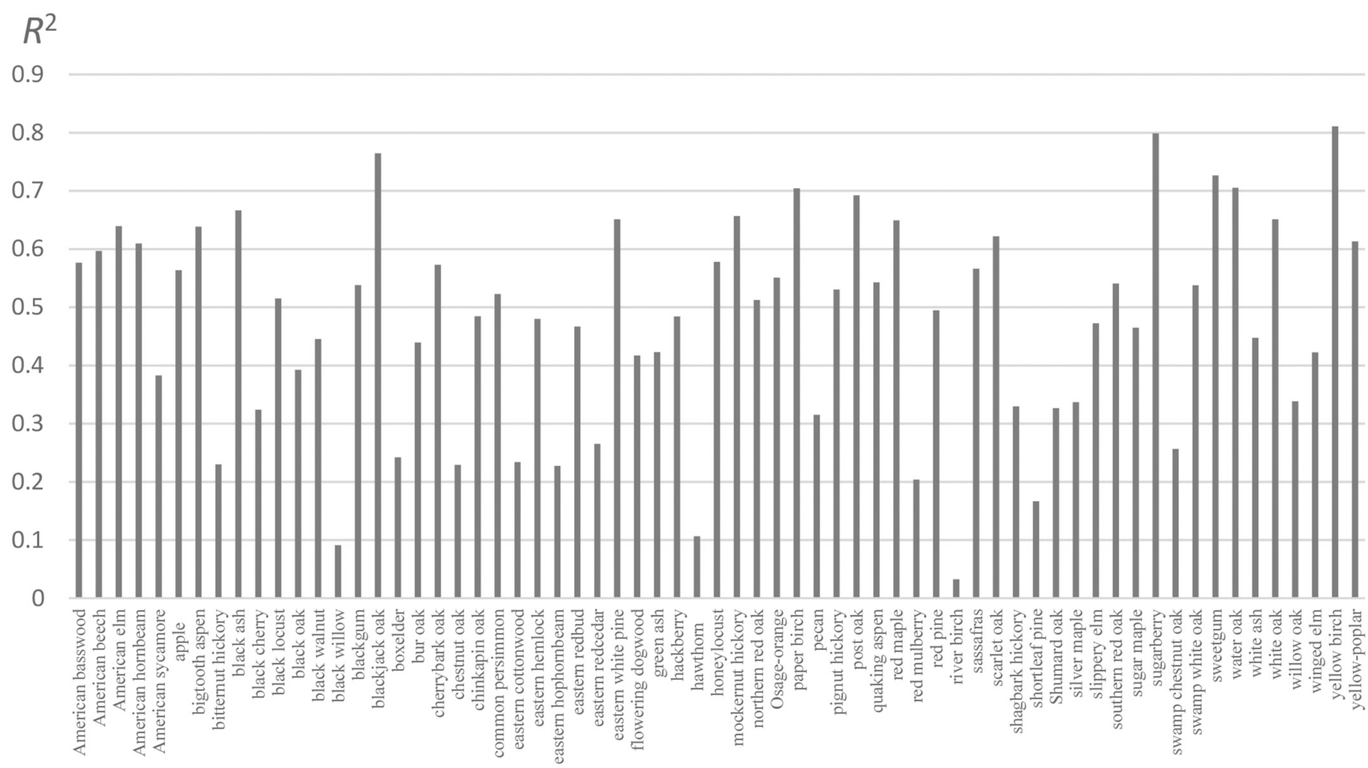

Fig. 6. Greatest R^2 value by tree species for models from the combined categories with varying number of input variables (see Supplementary File 3).

84% for the full 15 variable input models. For predicting species distributions, as opposed to relative abundances, climate variables were more important than land use and disturbance variables, similar to the findings of Crimmins et al. (2014). In terms of variable importance for species distributions, following four temperature variables, agricultural area over time was the fifth-most important variable, and historical fire return interval was the 11th most important variable. In contrast, agricultural area and fire return interval were the two most important variables in the full model of fifteen input variables for regression of relative abundances. Similar to abundance models, land use classes again were not important variables for species distributions, ranking as the last two variables out of fifteen, whereas soil variables were of low to moderate importance. These results support the concept that climate sets general boundaries on distributions, which can expand or contract when other factors change, whereas relative abundance within distributions is influenced by land use and disturbance (Canham and Thomas, 2010; Chambers et al., 2013).

4.2. Variables and species

Agricultural area during years 1850 to 1997 was the most important predictor for black locust and black walnut in land use-only models, whereas agricultural area was the most important variable in mixed category models for apple (*Malus*), black cherry (*Prunus serotina*),

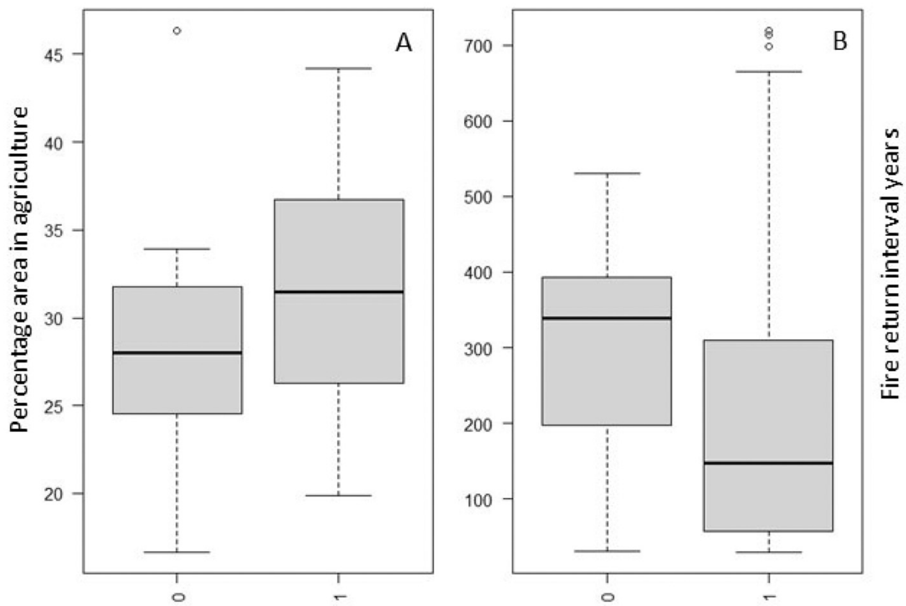

Fig. 7. Distribution of the important land use variables of percentage area in agriculture during 1850 to 1997 (A) and mean historical fire return interval for all fires (B) by whether all species were present (1; ≥ 0.5% of all trees) or (pseudo-) absent (0; < 0.5% of all trees) in ecological subsections.

hackberry (*Celtis occidentalis*), honeylocust (*Gleditsia triacanthos*), shagbark hickory (*Carya ovata),* and white ash (*Fraxinus americana*) for combined category models with R^2 values ranging from 0.32 to 0.58. These species all had positive relationships with agricultural area, in that agricultural area was greater were they were present (mean 37% of area) than where they were not present (mean 23% of area). Of course, apples and black walnut are agroforestry species, and hickories and honeylocust, and perhaps other tree species, were cultivated in the past by native humans (Warren, 2016). In previous research, black walnut, black cherry, hackberry, and white ash were associated with agricultural area (Hanberry, 2022), and these species have traits of rapid growth and some facultative association with wetlands, which are traits shared by black locust, honeylocust, and shagbark hickory. All of these species have increased since Euro-American settlement because in the past they were relegated by disturbance regimes to the borders between upland fires and wetland flooding. Rapid establishment after historically rare severe overstory disturbances, such as flooding in the past, make these species suited to recovery from current disturbance regimes.

Historical fire return interval of all fires was an important predictor for American sycamore (*Platanus occidentalis*), black oak (*Quercus velutina*), blackjack oak, eastern redcedar (*Juniperus virginiana*), post oak (*Quercus stellata*), red maple, Shumard oak (*Quercus shumardii*), sugar maple, swamp white oak, and white oak (*Quercus alba*). Black oak, blackjack oak, post oak, and white oak were favored by historical frequent surface regimes, and mean fire return interval was four-fold longer where they were absent than where they were present. In contrast, sugar maple was favored by very long (centuries and millennia) disturbance-free intervals, and where sugar maple was present had a mean fire return interval that was eight-fold longer than where the species was absent. Eastern redcedar and red maple are the major beneficiaries of land use change. Eastern redcedar primarily is present in the southwestern part of eastern forests, which historically had more frequent fire, and eastern redcedar has been increasing and expanding into grasslands (i.e., the study extent includes the entire eastern half of the U.S., including the tallgrass prairies of the central U.S.), where fire return intervals were short (Hanberry, 2021). Conversely, red maple is most abundant in northeastern forests where fire return intervals were longer and has expanded within eastern forests, including into the northernmost region where fire return intervals were long (Hanberry, 2020). American sycamore, Shumard oak, and swamp white oak are similar to species associated with agricultural area, with rapid growth (or moderate growth rates for Shumard oak) and facultative in wetlands, and generally may be responding to changed disturbance regimes.

4.3. Balancing climate change and land use change

Incorporating both climate change and land use and disturbance change into research is challenging, resulting in about 10% of studies that consider both climate change and land use and disturbance (Sirami et al., 2017; Titeux et al., 2016). Indeed, the imbalance between use of climate data relative to land use and disturbance variables for species distribution models and biodiversity scenarios has increased over time (Mod et al., 2016; Sirami et al., 2017; Titeux et al., 2016). Nevertheless, land use change and direct effects by human activities have been more influential on biodiversity than direct effects of climate change, which is being forced through fossil fuel use by humans (IPBES, 2019). Hominids have affected ecosystems through hunting and land management in the form of fire for hundreds of thousands of years (Roos et al., 2014). Agricultural intensification, land abandonment, fire regimes, and forest management are some of the land use and disturbance practices that affects species composition and abundance, resulting in changing equilibria and biotic interactions, regardless of climate influences (Araújo and Luoto, 2007; García-Valdés et al., 2013; Urbieta et al., 2008). Current species abundances may result due to release from historical disturbance factors rather than from current climate and edaphic conditions (Donohue et al., 2000). Based on a body of research, historical species abundances and distributions generally are well-known for the U.S., with major transitions due to disturbance regime changes in land use (i.e., exclusion of surface disturbance and acceleration of canopy tree disturbance) in all but some of the hemiboreal forests, which historically experienced relatively frequent severe fires, similar to current frequent severe overstory tree disturbance (Hanberry, 2020).

Additionally, although changing patterns can be recognized, attributions of cause require lines of evidence from both the land use and climate perspectives. Even apparent climate limitations are not always consistent with species physiological tolerances when disturbance and competition filters are removed (Aitken et al., 2008; Early and Sax, 2014; Perret et al., 2019; Van der Veken et al., 2008). Climate is attributed to outcomes without support by evidence (Taheri et al., 2021). For example, across most of the eastern half of the U.S., warming overall has not occurred during the past 30 years compared to the previous 85 years, while precipitation has increased (Fig. 1; Partridge et al., 2018; Rojas et al., 2019; Hanberry et al., 2020; PRISM Climate Group, 2021). However, the recent trend in precipitation likewise is not cause for tree species redistributions, because trees started redistributing before 1900, following agricultural and other land use and disturbance changes of Euro-American settlement (Hanberry, 2021). Furthermore, relatively stable site conditions interacted with historical disturbances and land use to produce an outcome different than manifested today. Historical species associations with soils, water availability, topography, and elevation generally are a product of a past land use that is no longer operational, as substantiated by a new suite of species under the same edaphic and topographic conditions (Hanberry et al., 2013). Apparent tolerances and optimal conditions of species will change based on different filters.

Because climate alone does not quantify species relative abundances, identification and development of effective land use variables for modeling is useful. While projection of land use and disturbance to the future is another challenge (Albert et al., 2020) that causes even greater imbalance between the number of studies that incorporate land use relative to climate change (Sirami et al., 2017), current land use and land cover variables were not important for current tree relative abundances. The two most important land use variables for these models were historical land use and disturbance. Therefore, these same variables will remain stable in forecasting species abundances. For future research, different time intervals may be explored as well as different spatial scales, extents, and analysis units. Additionally, different aspects of variables that were not effective may be tested, such as number of windstorms rather than wind speed. Agricultural area was an important variable, but various components may be elucidated, including crop types, agricultural practices, farm sizes, and production intensities that may modulate the effect of this land use type as a disturbance in the landscape, filtering, favoring or hindering the occurrence and abundance of species.

4.4. Considerations

Even though model variables were selected on the basis of a mechanism for influence, important model variables may by chance match well with relative species abundances without being influential due to spurious and confounding correlations. For instance, loblolly pine basal area as a variable, intended to represent the influence of tree removals in widespread loblolly pine plantations, simply provided a match for species of the southeastern U.S. In another example, species associated with agricultural area may have succeeded in the past under conditions ideal for agriculture, such as flat topography and fertile and wet soils; although these were alternative variable options, agricultural area may compress the topographic and soil information more compactly. Additionally, although not effective in these models, wind speed and topographic roughness (i.e., flat or steep topography) can act as proxies for both climate and land use (e.g., warmer temperatures, agriculture, and fires may occur on flat lands). An important variable for bur oak (*Quercus macrocarpa*) was precipitation of the driest month, but bur oak is relatively drought-tolerant (3.85 out of a maximum of 5; Niinemets and Valladares, 2006). Perhaps drought tolerance is the explanation for bur oak's distribution into the drier (north)western part of the eastern half of the U.S., but bur oak also has traits for surface fire tolerance, allowing this species to be successful in protected sites within (historical) grasslands.

In contrast, influential variables may not match well with current, unstable species abundances because of spatiotemporal mismatches or inaccuracies. For example, here, the modeling resolution probably was too coarse for site-specific soil and topographic variables to be important (Sormunen et al., 2011). A different extent likely would highlight more localized relevant variables. A temporal mismatch may arise due to changes in species distributions and abundances during the past century or more; therefore, historical fire return intervals have not been relevant since fire exclusion. Historical fire return intervals also are not based on accurate records, resulting in inaccuracies. Nonetheless, the only important land use and disturbance variables covered a time interval extending into the 1800s. In the future, the influence of current land use may start to be manifested in surveys of tree abundances.

5. Conclusions

Along with climate variables, land use and disturbance variables of agricultural area during years 1850 to 1997 and historical fire return interval of all fires were effective for modeling species abundances. The important two land use and disturbance variables occurred over long intervals while current land class variables, including current crop and pasture use, were not important to models. Climate variables were most important for presence and absence species distributions, which showed that climate is the main control of distribution boundaries, but land use and disturbance influences relative abundance within distributions. Only two of ten examined land use variables were competitive with climate variables for modeling tree species relative abundances, leaving identification or development of additional land use variables as a research need. Condensing multi-dimensional land use histories into metrics is a challenge, but additional land use variables likely are effective for modeling tree species relative abundances, which will enhance the balance between land use change and climate change in modeling studies.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All data are publicly available.

Acknowledgements

I thank the reviewers for their time and comments. 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 U.S. Government determination or policy. The author has no conflicts of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2023.102110.

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. Evol. Appl. 1, 95–111.
Albert, C.H., Hervé, M., Fader, M., Bondeau, A., Leriche, A., Monnet, A.C., Cramer, W., 2020. What ecologists should know before using land use/cover change projections for biodiversity and ecosystem service assessments. Reg. Environ. Chang. 20, 1–12.
Amatulli, G., McInerney, D., Sethi, T., Strobl, P., Domisch, S., 2020. Geomorpho90m, empirical evaluation and accuracy assessment of global high-resolution geomorphometric layers. Sci. Data 7, 1–18.
Araújo, M.B., Luoto, M., 2007. The importance of biotic interactions for modelling species distributions under climate change. Glob. Ecol. Biogeogr. 16 (6), 743–753.
Bechtold, W.A., Patterson, P.L., 2005. The enhanced forest inventory and analysis program–national sampling design and estimation procedures. USDA Forest Service, Southern Research Station, Asheville, NC.

Canham, C.D., Thomas, R.Q., 2010. Frequency, not relative abundance, of temperate tree species varies along climate gradients in eastern North America. Ecology 91, 3433–3440.

Chambers, D., Périé, C., Casajus, N., de Blois, S., 2013. Challenges in modelling the abundance of 105 tree species in eastern North America using climate, edaphic, and topographic variables. For. Ecol. Manag. 291, 20–29.

Crimmins, S.M., Dobrowski, S.Z., Mynsberge, A.R., Safford, H.D., 2014. Can fire atlas data improve species distribution model projections? Ecol. Appl. 24 (5), 1057–1069.

Davidson, R., 1840. An Excursion to the Mammoth Cave and the Barrens of Kentucky. A. T. Silliman and Son. Available at. https://www.google.com/books/edition/An_Excursion_to_the_Mammoth_Cave_and_the/71EVAAAAYAAJ?hl=en&gbpv=1&printsec=frontcover. Available at.

Day, G.M., 1953. The Indian as an ecological factor in the northeastern forest. Ecology 34, 329–346.

Deutsch, C.A., Tewksbury, J.J., Huey, R.B., Sheldon, K.S., Ghalambor, C.K., Haak, D.C., Martin, P.R., 2008. Impacts of climate warming on terrestrial ectotherms across latitude. Proc. Natl. Acad. Sci. 105 (18), 6668–6672.

Donohue, K., Foster, D.R., Motzkin, G., 2000. Effects of the past and the present on species distribution: land-use history and demography of wintergreen. J. Ecol. 88 (2), 303–316.

Early, R., Sax, D.F., 2014. Climatic niche shifts between species' native and naturalized ranges raise concern for eco-logical forecasts during invasions and climate change. Glob. Ecol. Biogeogr. 23, 1356–1365.

Fang, Y., Jawitz, J., 2018. High-resolution reconstruction of the United States human population distribution, 1790-2010. Sci. Data 5, 180067.

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37, 4302–4315.

Fourcade, Y., Besnard, A.G., Secondi, J., 2018. Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. Glob. Ecol. Biogeogr. 27, 245–256.

Gagnon, P.R., Battaglia, L.L., Hanberry, B.B., Conner, W.H., King, S.L., 2021. Fire in floodplain forests of the southeastern USA. In: Greenberg, C.H., Collins, B. (Eds.), Fire Ecology and Management: Past, Present, and Future of US Forested Ecosystems. Springer International Publishing, New York, pp. 201–242.

García-Valdés, R., Zavala, M.A., Araujo, M.B., Purves, D.W., 2013. Chasing a moving target: projecting climate change-induced shifts in non-equilibrial tree species distributions. J. Ecol. 101 (2), 441–453.

Global Wind Atlas, 2021. Global Wind Atlas 3.0. Technical University of Denmark (DTU). Available at. https://globalwindatlas.info/downloads/gis-files.

Grime, J.P., 1977. Evidence for the existence of three primary strategies in plants and its relevance to ecological and evolutionary theory. Am. Nat. 111 (982), 1169–1194.

Hanberry, B.B., 2019. Recent shifts in shade tolerance and disturbance traits in forests of the eastern United States. Ecol. Process. 8 (1), 1–11.

Hanberry, B.B., 2020. Baseline to novel ecosystems in Michigan, USA, with a quantitative and qualitative assessment. Écoscience 27, 255–268.

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. Confronting the issue of invasive native tree species due to land use change in the eastern United States. Land 11, 161.

Hanberry, B.B., Noss, R.F., 2022. Locating potential historical fire-maintained grasslands of the eastern United States based on topography and wind speed. Ecosphere 13 (6), e4098.

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.

Hanberry, B.B., Kabrick, J.M., He, H.S., 2014. Changing tree composition by life history strategy in a grassland-forest landscape. Ecosphere 5 (3), 1–16.

Hanberry, B.B., Abrams, M.D., Arthur, M.A., Varner, J.M., 2020. Reviewing fire, climate, deer, and foundation species as drivers of historically open oak and pine forests and transition to closed forests. Front. For. Glob. Chang. 3, 56.

Homer, C., Dewitz, J., Jin, S., Xian, G., Costello, C., Danielson, P., Gass, L., Funk, M., Wickham, J., Stehman, S., Auch, R., 2020. Conterminous United States land cover change patterns 2001–2016 from the 2016 national land cover database. ISPRS J. Photogramm. Remote Sens. 162, 184–199.

Huang, H., Winter, J.M., Osterberg, E.C., Horton, R.M., Beckage, B., 2017. Total and extreme precipitation changes over the northeastern United States. J. Hydrometeorol. 18, 1783–1798.

Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services [IPBES], 2019. Global Assessment Report on Biodiversity and Ecosystem Services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. IPBES Secretariat. Available at. https://doi.org/10.5281/zenodo.3831673.

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

Kuhn, M., 2008. Building predictive models in R using the caret package. J. Stat. Softw. 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.

LANDFIRE, 2022a. Biophysical Settings 2016 Remap. Available at. https://LANDFIRE.gov/.

LANDFIRE, 2022b. Vegetation Departure Index. Available at. https://LANDFIRE.gov/.

Leiberg, J.B., 1902. Forest conditions in the northern Sierra Nevada. In: California, (Vol. 8).. US Government Printing Office.

Loehle, C., 1998. Height growth rate tradeoffs northern determine and southern range limits for trees. J. Biogeogr. 25, 735–742.

Maizel, M., White, R.D., Root, R., Gage, S., Stitt, S., Osborne, L., Muehlbach, G., 1998. Historical interrelationships between population settlement and farmland in the conterminous United States, 1790 to 1992. Available at. https://archive.usgs.gov/archive/sites/landcover.usgs.gov/luhna/chap2.html.

Mod, H.K., Scherrer, D., Luoto, M., Guisan, A., 2016. What we use is not what we know: environmental predictors in plant distribution models. J. Veg. Sci. 27, 1308–1322.

Motzkin, G., Wilson, P., Foster, D.R., Allen, A., 1999. Vegetation patterns in heterogeneous landscapes: the importance of history and environment. J. Veg. Sci. 10 (6), 903–920.

Niinemets, Ü., Valladares, F., 2006. Tolerance to shade, drought, and waterlogging of temperate northern hemisphere trees and shrubs. Ecol. Monogr. 76, 521–547.

Pan, Y., Chen, J.M., Birdsey, R., McCullough, K., He, L., Deng, F., 2011. Age structure and disturbance legacy of North American forests. Biogeosciences 8, 715–732.

Partridge, T.F., Winter, J.M., Osterberg, E.C., Hyndman, D.W., Kendall, A.D., Magilligan, F.J., 2018. Spatially distinct seasonal patterns and forcings of the US warming hole. Geophys. Res. Lett. 45, 2055–2063.

Perret, D.L., Leslie, A.B., Sax, D.F., 2019. Naturalized distributions show that climatic disequilibrium is structured by niche size in pines (*Pinus* L.). Glob. Ecol. Biogeogr. 28, 429–441.

Prism Climate Group, 2021. PRISM Climate Data. Available online: https://prism.oregonstate.edu/ (accessed on 30 October 2021).

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

Recknagel, F., Michener, W.K., 2018. Ecological informatics: An introduction. In: Recknagel, F., Michener, W.K. (Eds.), Ecological Informatics. Springer International Publishing, New York, pp. 3–12.

Rojas, M., Lambert, F., Ramirez-Villegas, J., Challinor, A.J., 2019. Emergence of robust precipitation changes across crop production areas in the 21st century. Proc. Natl. Acad. Sci. 116 (14), 6673–6678.

Roos, C.I., Bowman, D.M., Balch, J.K., Artaxo, P., Bond, W.J., Cochrane, M., D'Antonio, C.M., DeFries, R., Mack, M., Johnston, F.H., Krawchuk, M.A., 2014. Pyrogeography, historical ecology, and the human dimensions of fire regimes. J. Biogeogr. 41, 833–836.

Sargent, C.S., 1884. Report on the Forests of North America (Exclusive of Mexico). Department of the Interior Census Office. Government Printing Office, Washington, D.C. Available at. https://www.google.com/books/edition/Report_on_the_Forests_of_North_America_e/pfFq1ttP4h8C?hl=en&gbpv=1&printsec=frontcover. Available at.

Sirami, C., Caplat, P., Popy, S., Clamens, A., Arlettaz, R., Jiguet, F., Brotons, L., Martin, J. L., 2017. Impacts of global change on species distributions: obstacles and solutions to integrate climate and land use. Glob. Ecol. Biogeogr. 26, 385–394.

Sormunen, H., Virtanen, R., Luoto, M., 2011. Inclusion of local environmental conditions alters high-latitude vegetation change predictions based on bioclimatic models. Polar Biol. 34, 883–897.

Taheri, S., Naimi, B., Rahbek, C., Araújo, M.B., 2021. Improvements in reports of species redistribution under climate change are required. Sci. Adv. 7, eabe1110.

Titeux, N., Henle, K., Mihoub, J.B., Regos, A., Geijzendorffer, I.R., Cramer, W., Verburg, P.H., Brotons, L., 2016. Biodiversity scenarios neglect future land-use changes. Glob. Chang. Biol. 22, 2505–2515.

Urbieta, I.R., Zavala, M.A., Maranón, T., 2008. Human and non-human determinants of forest composition in southern Spain: evidence of shifts towards cork oak dominance as a result of management over the past century. J. Biogeogr. 35 (9), 1688–1700.

USDA Forest Inventory and Analysis, 2021. Available at. https://apps.fs.usda.gov/fia/datamart/CSV/datamart_csv.html.

USDA Forest Inventory and Analysis, 2022. Forest Inventory EVALIDator Web-Application Version 1.8.0.01. U.S. Department of Agriculture, Forest Service, Northern Research Station. Available at. http://apps.fs.usda.gov/Evalidator/evalidator.jsp.

Van der Veken, S., Hermy, M., Vellend, M., Knapen, A., Verheyen, K., 2008. Garden plants get a head start on climate change. Front. Ecol. Environ. 6, 212–216.

Walkinshaw, M., O'Geen, A.T., Beaudette, D.E., 2021. Soil Properties. California Soil Resource Lab. casoilresource.lawr.ucdavis.edu/soil-properties.

Warren, R.J., 2016. Ghosts of cultivation past-Native American dispersal legacy persists in tree distribution. PLoS One 11, e0150707.

Yang, J., Tao, B., Shi, H., Ouyang, Y., Pan, S., Ren, W., Lu, C., 2020. Integration of remote sensing, county-level census, and machine learning for century-long regional cropland distribution data reconstruction. Int. J. Appl. Earth Obs. Geoinf. 91, 102151.