Landsc Ecol (2022) 37:1819–1837
https://doi.org/10.1007/s10980-022-01436-6

RESEARCH ARTICLE

Defining landscape-level forest types: application of latent Dirichlet allocation to species distribution models

Matthew P. Peters · Steve N. Matthews · Anantha M. Prasad · Louis R. Iverson

Received: 12 October 2021 / Accepted: 14 March 2022 / Published online: 12 April 2022

Abstract

Context Forest type (FT) classification provides useful information to ecologists and forest managers by representing similar sites based on species dominance. Various methods have been developed using stand-level or plot-level information, however, these classifications are not always effective at representing broader landscape patterns of species diversity.

Objectives We classified landscape-level FTs from species habitat models and compared against classifications intended for stand-level information. We used a departure score to assess potential changes to current FT from projected changes in climate and habitat suitability (HS).

Methods We applied a text mining algorithm, latent Dirichlet allocation (LDA), to 125 species HS models within the eastern United States to define 11 FTs under current conditions. We compared the LDA model against two summations of relative abundance. We then developed a departure score to characterize potential changes to current FTs under projected climate change.

Results The LDA model showed broad spatial agreement with summations of species relative abundance. However, LDA's landscape-level dominance of species differed from stand-level classifications of species summations. Varying degrees of pressure from climate change and HS indicated that future FTs could face conditions that result in departures. However, the overall departure scores tended to be lower due to reduced pressure from modeled changes in HS for much of the eastern US.

Conclusions LDA results are promising for classifying landscape-level FTs. Portraying potential changes in future FTs with departure scores may facilitate better management by aligning the spatial scales of information and not attributing changes to specific species or conditions.

Disclaimer The use of trade, firm, or corporation names in this publication is for the information and convenience of the reader. Such use does not constitute an official endorsement or approval by the U.S. Department of Agriculture or the Forest Service of any product or service to the exclusion of others that may be suitable.

Supplementary Information The online version of this article contains supplementary material available (https://doi.org/10.1007/s10980-022-01436-6).

M. P. Peters (✉) · S. N. Matthews · A. M. Prasad
USDA Forest Service, Northern Research Station, Northern Institute of Applied Climate Science, Delaware, OH, USA
e-mail: matthew.p.peters@usda.gov

S. N. Matthews
School of Environment and Natural Resources, The Ohio State University, Columbus, OH, USA

L. R. Iverson
USDA Forest Service, Northern Research Station, Northern Institute of Applied Climate Science, Retired, Delaware, OH, USA

Keywords Climate change · Eastern United States · Habitat suitability · Importance values · Text mining algorithm

Introduction

The practice of classifying forest vegetation into communities (e.g., Clements 1916; Gleason 1926) or species assemblages (e.g., Fauth et al. 1996) provides ecologists and foresters a baseline expectation of what taxa and site conditions exist within a landscape. Classification often occurs at the stand- or site-level, where tree species diversity is almost always lower than at a landscape-level (e.g., $\geq 10\,\text{km}^2$), in which broader patterns of forest composition are influenced by land use patterns, other disturbances, and geomorphology. Forest managers can use information from forest types (FTs) and communities (e.g., Eyre 1980; Arner et al. 2003; Ruefenacht et al. 2008) when making decisions about which species to promote in response to disturbances. These stand-level classifications act as a type of ecological division at local scales by providing context to existing vegetation and possible associated species which may not already be present. Yet, this fine-scaled information is often unavailable, and managers, researchers, and decision makers tend to rely on coarse scaled, landscape-level, information to supplement unsampled locations. The concept of hierarchical scales is foundational in landscape ecology (see Wu and Qi 2000; Turner et al. 2003; Wu and Hobbs 2007) and can be useful when aggregating information to appropriate spatial scales to meaningfully represent patterns and processes.

In addition, when planning improvements for adaptive capacities to mitigate potential impacts from climate change (e.g., Nagel et al. 2017), information about potential future FTs can be desirable (Nevins et al. 2021), but not widely available. As species' distributions have changed independently from other species (Webb and Bartlein 1992) and novel communities will likely emerge under future climate change, it can be important to consider how range shifts affect current communities into the future. Therefore, species distribution models are most often developed to examine the impacts of climate change on tree species' habitat suitability (HS) using coarse resolutions to capture broad-scale patterns and align various data sets (e.g., $\geq 1\,\text{km}^2$) than are intended for FT classifications developed for stand-level information (Fig. 1). While examples of combining individual species models (Iverson and Prasad 2001; Iverson et al. 2008) or modeling FTs (Joshi et al. 2012; Tchebakova et al. 2016) under climate change scenarios are available, we submit that the use of *a priori* groups of species, intended to represent stand-level dominance, can be problematic when applied to landscape-level information (Fig. 1); especially when projecting into the future as outlined below. With updated models of individual species' relative abundance represented at 10 and $20\,\text{km}^2$ grid cells (Iverson et al. 2019b; Peters et al. 2019), we sought to assign the dominant FT in a way that addressed the following issues of a previous aggregation approach that applied a stand-level classification to $20\,\text{km}^2$ grid cells (described in Iverson and Prasad 2001; Iverson et al. 2008): (1) FTs with many species had the potential to have greater summed importance values (e.g., relative abundance) than those with fewer species, (2) a small percentage of cells had equal dominance among two or more FTs, and the classification logic defaulted to the first class encountered, (3) FTs defined by complex stand-level criteria may not be reflective of broader landscape diversity, (4) because modeled HS does not translate to actual abundance/presence everywhere (e.g., commission/omission errors), and modeling resolutions could contain multiple FTs, especially near transition zones, a single class based on *a priori* definitions is potentially misrepresentative, and (5) *a priori* groupings of traditional FTs do not allow for the possibility of novel combination of species.

To overcome the limitations of using stand-level definitions and simple aggregation methods to define FTs from modeled HS at landscape-level scales, we explored the possibility of using latent Dirichlet allocation (LDA), a machine learning algorithm for text mining to determine the likelihood of species associated with various FTs. We compare the results of the LDA model with those of simple summation of relative abundance, and a more complex summation used by the USDA Forest Service Forest Inventory and Analysis program to define FTs from modeled HS. We then calculated a departure score to represent the potential change to future forest communities based on projected climate change and predicted HS. The purpose of our analysis is to show that broader landscape-level assessments of HS via machine learning algorithms provides useful insights over methods

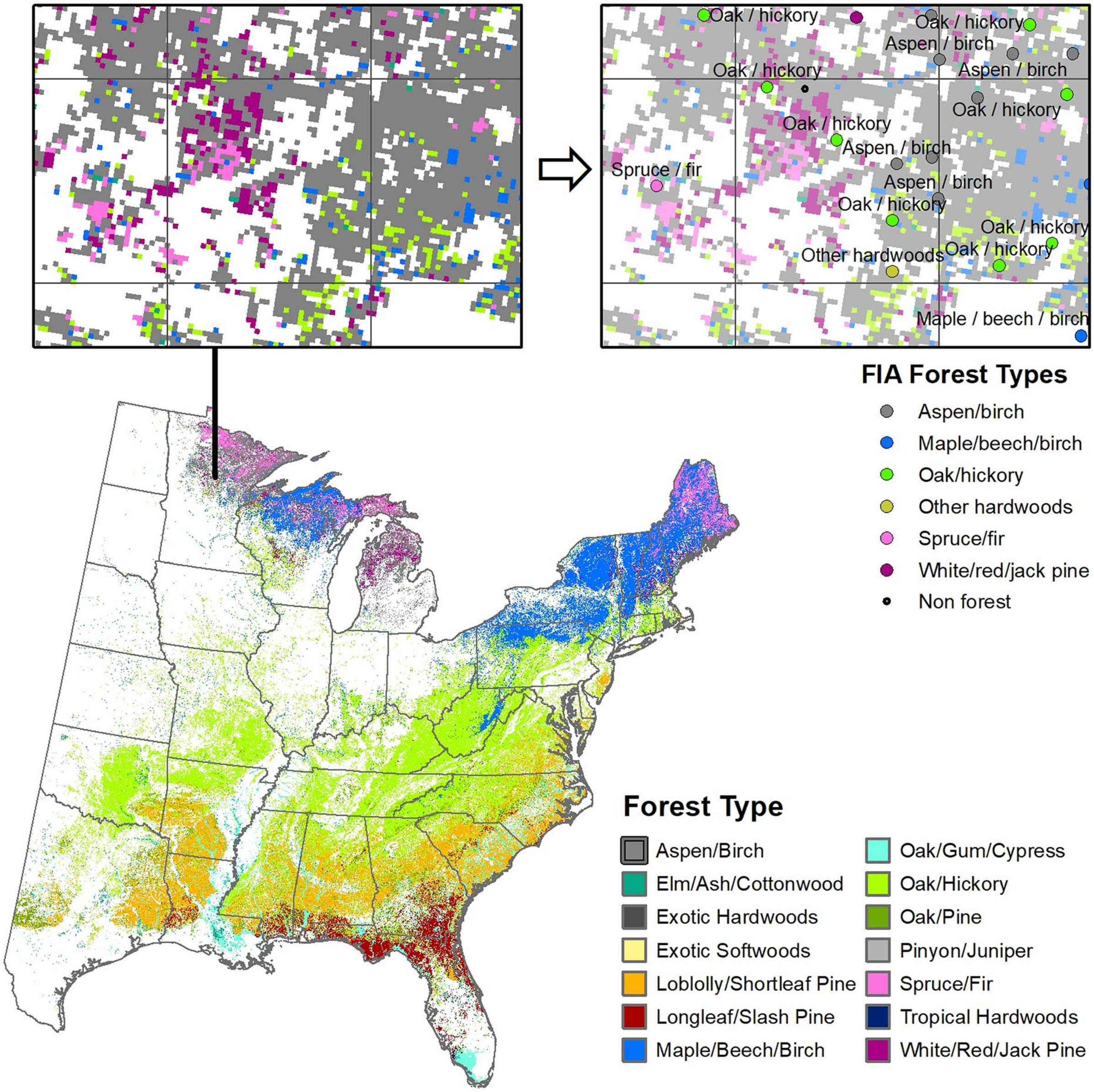


Fig. 1 Classification of MODIS imagery for the eastern U.S. depicting 14 forest communities (Ruefenacht et al. 2008). Left inset shows varying proportions of different FTs among the DISTRIB-II modeling grid cells (thin black squares), while the right inset indicates the field assigned FT for the Forest Inventory and Analysis (FIA) plot. Locations of FIA plots are approximate, but indicate which communities were likely surveyed when compared to the left inset. Note that the FTs in the right inset were lightened to show FIA plots and labels

developed to summarize local conditions, and hence should be considered for future assessments of forested landscapes. Developing metrics that combine the influence of projected changes in climate with HS to indicate departure from current conditions provides better understanding of complex interactions.

Methods

Modeled habitat suitability

We modeled HS under current and potential future conditions for 125 eastern U.S. tree species (Iverson et al. 2019b; Peters et al. 2019) using the randomForest package (Liaw and Wiener 2002) for R (R Core Team 2018). These models, DISTRIB-II, were derived from USDA Forest Service Forest Inventory and Analysis (FIA) data (www.fia.fs.fed.us) that were used to calculate individual species importance values (IVs) from relative species density and dominance (e.g., basal area) at the plot level (hereafter also referred as relative abundance). These plot data were aggregated (hereafter FIA actual IV) to a hybrid lattice of 10×10 and 20×20 km grid cells to provide a coarse landscape-level representation of species' relative abundance. The hybrid lattice is used to increase the spatial resolution and average among fewer FIA inventory plots where higher densities of plots exist (Peters et al. 2019). IVs represent a species' relative abundance as a continuous value ranging from 0 to 100; to limit uncertainty of very low abundance, IVs between 0 and 1 were assigned to 0 for purposes of this study as these values are likely not dominant among other species within the grid cell. The DISTRIB-II models were built from 45 environmental predictors (climate, elevation, and soil) (Supplemental Material Table 1) with more than 84,000 FIA plots aggregated among the grid cells under current climate conditions (1981–2010). Seven climate variables (annual precipitation, growing season precipitation, annual mean temperature, growing season mean temperature, mean temperature of the coldest month, mean temperature of the warmest month, and annual aridity) were included and swapped for future projected conditions while other environmental predictors were held constant. Not all species models were created equal, and a series of metrics were applied to score and categorize, low to high, model reliability of each species (Peters et al. 2019). The models were then applied to all 29,357 grid cells in the eastern U.S. and six scenarios of climate change from three general circulation models (GCM) and two representative concentration pathways (RCP, Moss et al. 2008), or scenarios of emissions: 4.5 (lower emissions) and 8.5 (higher emissions) (Iverson et al. 2019b). However, to simplify outputs in this analysis, the predicted IVs from the three GCMs were averaged to produce only two RCP scenarios: low vs. high emissions. The FIA actual IVs were used to train the FT models, which were then applied to a modeled estimate of current IVs. Average species IVs for each RCP were used to examine potential future departures from the current FT assignments.

Latent Dirichlet allocation model

Latent Dirichlet allocation (LDA) was first developed for population genetics by Pritchard et al. (2000) and independently developed for machine learning as a text mining approach by Blei et al. (2003) to describe 16,000 digital documents using the frequency of individual words within a document to define topics based on many documents. LDA has also been used for image classification (Ye et al. 2009) and has only recently been applied to define FTs (Valle et al. 2014; Knott et al. 2020) from observed vegetation. Using FIA data, Knott et al. (2020) calculated IVs for species and defined forest communities with LDA. We took a similar approach because our DISTRIB-II models are derived from IVs representing the relative abundance of a species and can be a surrogate for frequency. Thus, in LDA terms, we used IVs (*frequency*) of individual tree species (*words*) to define FTs (*topics*) for each modeling grid cell (*documents*). Output consists of the likelihood for each FT occurring within the modeling grid cells and the likelihood for each species being associated with each FT. Data preparation included (1) removing grid cells where no species were recorded by FIA or modeled under current or future climate change scenarios, and (2) modifying IVs by rounding to integers, except for values < 1.0 which were recorded as zeros. Using integers, we hope to create more competition among FT assignment in the LDA algorithm between species with very close IVs.

The LDA model was performed in Python (Python Software Foundation) using the guidedlda library (Singh 2017) which uses a collapsed Gibbs sampling approach (Griffiths and Steyvers 2004). The guidedlda library allowed us to compare how an unseeded and a seeded LDA model defines individual FTs. Seeding LDA topics helps to ensure certain groupings of words emerge and are not included in more generalizable classes (Jagarlamudi et al. 2012). Specifically, the guidedlda implementation uses a parameter seed confidence, which ranges from 0 to 1, that biases seeded species to corresponding FTs. Preliminary parameterizations suggested that some of the more regional FTs with few associated species may benefit from some initial definition to support optimization across the full data space. The seeded model used the species in Table 1 to assist in defining five FTs. We set the parameters to define 11 FTs, used 2000 iterations, applied a seed confidence of 0.15, and retained log-likelihoods for every tenth iteration. We used default values of 0.01 for the two Dirichlet parameters controlling the distribution of FTs (*topics*) and species (*words*). We also evaluated classification of nine and 13 FTs but chose 11 based on evaluation metrics (Supplemental Material Fig. 1) from the tmtoolkit library (Konrad 2020) and because some of resulting FTs shared species, suggesting that combining these FTs would reduce the number of classes within the eastern United States. It can be desirable to have uniquely defined FT classes that reflect patterns across broad landscapes rather than many classes with shared characteristics that could easily transition to another class with small changes. The LDA models were developed using FIA actual IVs initially among 125 species but were iteratively reduced to 104 non-rare species to predict potential current FTs. The 21 species were excluded from the LDA model following an approach used by Knott et al. (2020) where the species' likelihood across all FTs was compared to a threshold of $1/n_{species}$.

Table 1 Forest types defined by U.S. Department of Agriculture Forest Service and 108 associated species among the 125 modeled by DISTRIB-II. Species are listed in alphabetical order, those in bold were included as seeded topics for the latent Dirichlet allocation (LDA) model

Forest type	Species
White/Red/Jack Pine	*Betula populifolia*†, *Picea rubens*†, *Pinus banksiana, P. resinosa, P. strobus, Tsuga canadensis*
Spruce/Fir	*Abies balsamea, Larix laricina, Picea glauca, P. mariana, P. rubens, Thuja occidentalis*
Longleaf/Slash Pine	***Pinus elliottii***, ***P. palustris***
Loblolly/Shortleaf Pine	*Liquidambar styraciflua*†, *Pinus clausa*, ***P. echinata***, *P. glabra, P. pungens, P. rigida, P. serotina*, ***P. taeda***, *P. virginiana*
Oak/Pine	*Fraxinus americana*, Juniperus virginiana, Pinus echinata, P. elliottii*, P. palustris, P. rigida*†, *P. serotina*†, *P. strobus, P. taeda, P. virginiana, Quercus coccinea*†, *Q. falcata, Q. incana*, Q. laevis*, Q. marilandica*, Q. nigra*†, *Q. rubra, Q. stellata*†, *Q. velutina*†
Oak/Hickory	*Acer rubrum*, Aesculus flava*, A. glabra*, Carpinus caroliniana*, Carya alba*, C. cordiformis, C. glabra*, C. laciniosa*†, *C. ovata, C. texana, Cercis canadensis, Cornus florida*, Diospyros virginiana, Fraxinus americana*, F. nigra*, F. pennsylvanica*, F. quadrangulata*, Gleditsia triacanthos*, Juglans nigra, Liquidambar styraciflua, Liriodendron tulipifera, Morus rubra*, Nyssa sylvatica*, Oxydendrum arboreum*, Prunus pensylvanica*, P. serotina*, Quercus alba, Q. bicolor*†, *Q. coccinea, Q. ellipsoidalis*†, *Q. falcata, Q. imbricaria, Q. incana*, Q. laevis*, Q. laurifolia*, Q. lyrata*†, *Q. macrocarpa, Q. marilandica, Q. michauxii*†, *Q. muehlenbergii*†, *Q. nigra*, Q. pagoda*†, *Q. phellos, Q. prinus, Q. rubra, Q. shumardii*†, *Q. stellata, Q. texana*†, *Q. velutina, Q. virginiana*, Robinia pseudoacacia, Sassafras albidum, Ulmus alata*, U. americana*, U. crassifolia*, U. rubra**
Oak/Gum/Cypress	*Acer rubrum*, ***Carya aquatica***, *C. laciniosa*, Chamaecyparis thyoides*†, *Liquidambar styraciflua, Magnolia virginiana**, ***Nyssa aquatica***, ***N. biflora***, *Quercus lyrata, Q. michauxii*, ***Q. pagoda****, *Q. phellos, Q. texana*, ***Taxodium ascendens***, ***T. distichum***, *Ulmus americana*†
Elm/Ash/Cottonwood	***Acer negundo***, *A. rubrum*, ***A. saccharinum****, ***Betula nigra***, ***Carya illinoinensis***, ***Celtis laevigata***, ***C. occidentalis***, *Fraxinus americana*†, ***F. nigra***, *F. pennsylvanica*, ***Platanus occidentalis***, ***Populus deltoides***, ***Salix nigra***, *Ulmus alata, U. americana, U. crassifolia, U. rubra*
Maple/Beech/Birch	*Acer barbatum**, *A. nigrum, A. pensylvanicum*†, *A. rubrum*†, *A. saccharum, (A) spicatum*†, *Betula alleghaniensis, (B) lenta*†, *B. populifolia*†, *Fagus grandifolia, Fraxinus quadrangulata*†, *Gleditsia triacanthos*†, *Juglans nigra*†, *Prunus serotina, Robinia pseudoacacia*†, *Tilia americana*
Aspen/Birch	*Betula lenta*†, *B. papyrifera, B. populifolia, Populus balsamifera, P. grandidentata, P. tremuloides, Prunus pensylvanica*
Pinyon/Juniper	***Juniperus ashei***

Species are listed in alphabetical order, those in bold were included as seeded topics for the latent Dirichlet allocation (LDA) model. Species or groups not used to seed LDA model were used in simple summation and FIA summation.

Species with * were used in FIA summation but not simple summation. Species with † were used in simple summation but not FIA summation

Simple summation

The FIA program uses criteria defined by the Society of American Foresters (SAF, Eyre 1980) to classify eastern forests into 11 distinct groups (Table 1) which have been mapped from 250 m satellite imagery (Ruefenacht et al. 2008). Previous DISTRIB models predicted relative abundance (i.e., IV) for 80 tree species among counties (Iverson and Prasad 2001) and 134 species among a uniform grid of 20 km^2 cells (Iverson et al. 2008) and applied SAF classifications. Using IVs from 92 species (Table 1), SAF classifications were assigned according to simple, IV-based assignment rules (Iverson and Prasad 2001; Iverson et al. 2008) and the dominant group was assigned to the DISTRIB-II modeling grid cells. Assignment of loblolly/shortleaf pine (*Pinus taeda/P. echinata*), oak/pine (*Quercus/Pinus*), and oak/hickory (*Quercus/Carya*) FTs used special rules to determine proportional dominance described in Iverson and Prasad (2001). If (1) loblolly/shortleaf pine was less than oak/hickory or both of these two groups were absent and oak/pine was dominant, then the grid cell was assigned as oak/pine; (2) loblolly/shortleaf pine was greater than oak/hickory and oak/pine was the dominant class, then the grid cell was reclassified as loblolly/shortleaf pine since pines exceeded 50%; (3) both loblolly/shortleaf pine and oak/hickory were greater than zero having the same summed IV and oak/pine was dominant, then the grid cell was assigned oak/pine. The earlier approaches (Iverson and Prasad 2001; Iverson et al. 2008) did not account for equal dominance among multiple groups, therefore, we used a “multiple” class when a grid cell resulted in two or more FTs having equal dominance. Additionally, a similar approach using a more rigorous algorithm (Arner et al. 2003) developed for stand-level data was applied (Supplemental Material FIA Summation). However, we present the simple summation and two LDA approaches to focus on improvements between the two versions of DISTRIB.

Evaluation

The LDA model calculates a probability for each FT, whereby the log-likelihoods can be used to calculate the Akaike information criterion (AIC) statistic and a harmonic mean (Griffiths and Steyvers 2004) to compare the rate of performance among iterations of various parameterizations. However, results from the two summations do not produce such statistics to compare among parameterizations or between methods. Additionally, to evaluate LDA performance and determine an ideal number of FTs, we examined three metrics (Supplemental Material Fig. 1) from the tmtoolkit library that evaluate: (1) the distribution of variance between FTs and species and the marginal FT (Arun et al. 2010), (2) the posterior distributions between FTs and species (Cao et al. 2009), and (3) coherence of species within each FT (Mimno et al. 2011). Because the resulting FT classes themselves (e.g., class names) differed between the LDA and the summation methods, we were unable to quantify differences among all the classifications. However, statistics such as a fuzzy kappa (Hagen-Zanker 2009) can be used to compare agreement among categorical values or information generated from difference metrics [e.g., diffeR package for R (Pontius Jr. and Santacruz 2019)] can calculate agreement/disagreement between reference and comparison maps, which we used to evaluate differences among the two LDA models and the two summation outputs. The difference between FT assignments resulting from the FIA actual IV and the modeled current IV (i.e., training against prediction) used in the two summation classification methods, as well as the unseeded and seeded LDA models were compared. Additionally, differences between the unseeded and seeded LDA models for FTs resulting from the FIA actual IV and modeled current IV were compared (i.e., unseeded against seeded).

Departure analysis

The three methods used to define FTs can be applied to potential future HS to map direct changes. We however chose to develop a departure score to indicate grid cells that show departure from current conditions by the end of the century, resulting in a different FT. Using departure scores to represent potential future changes in forest communities is likely to reduce some of the perceived certainty reflected by mapped projections and provides a way to account for complex changes in *both* climate and HS. This is because the departure scores identify where changes might occur but do not directly attribute the changes to certain climate conditions or specific species. The DISTRIB-II models included seven climate variables

(Supplemental Material Table 1) which were swapped for future projected conditions by holding the other environmental predictors constant. We calculated a climatic vulnerability (CV) score following the approach described by Triepke et al. (2019) for each climate variable. The CV score is derived by dividing the projected change in climate by two standard deviations of the interannual variation of observed climate values and subtracting one so that future changes within 95% (assuming a normal distribution) of the current variability are represented within a range of -1–0.

$$CV\ score = \frac{|projected\ change|}{2s} - 1$$

where the projected change is the difference between the baseline condition and the future projected value for each climate variable, and s is the standard deviation of the interannual variability during the baseline period. For each climate variable, CV scores were calculated at each modeling grid and assumes climate values are normally distributed. The seven CV scores were reclassed following Triepke et al. (2019) to low, moderate, high, and very high using the ranges ≤ 0, >0–0.49, 0.5–0.99, and ≥ 1 to generalize vulnerability across proportionally increasing classes of departure. Each vulnerability class was given numerical values of 0.25, 0.5, 0.75, and 1, respectively, and were then weighted equally by 1/7, as seven climate variables were included in the DISTRIB-II models; and combined into a single value to represent the potential climate pressure that could influence future departures. Because the randomForest models may correlate each of the seven climate variables differently among the 125 species and each variable could have both positive (i.e., increased suitability) or negative (i.e., decreased suitability) affects, aggregating the seven CV scores consolidates these variations into a single value.

Changes in climate may partly explain why a FT might depart from the current classification, as changes in species' HS can vary as well. Therefore, we considered how climate related changes may influence departure of each grid cell to another FT via three species groupings: (a) species associated with the current dominant FT, (b) other species with current suitable habitat, and (c) species having newly suitable habitat in the future. Using the species IVs, each were summed among the three groups and assigned a value based on one of the five possible conditions defined in Table 2 (see detailed definitions in Supplemental Material) to represent potential changes in HS for the group on a 0–1 gradient. The three groups were summed and then divided by 2.5, the maximum combined possible value, to represent the potential gradient of changes in HS that could influence future departures. The HS score is designed so lower values (0–0.5) indicate the current FT species could remain dominant under future climate scenarios, while higher values (0.5–1) suggest that the FT species could lose dominance to other species or those with newly suitable habitats. Overlap between lower and higher HS scores reflects the uncertainty of aggregating HS of many species to characterize the dominant FT and

Table 2 Three groups of species and five potential future habitat responses used to represent potential changes among habitat suitability (HS) that could result in future departures from current FTs

	Increase/No change and dominant	Increase/No Change and not dominant	Increase and equal to current species	Decrease and dominant	Decrease and not dominant
Dominant forest type species	0	0.75	–	0.5	1
Other species	1	0.25	–	0.5	0
Future species	1	0.5	0.25	–	0

Only one of the three species groups can be dominant, thus not all possible combination can occur. The corresponding score for each of the three species groups are summed and divided by 2.5, the maximum possible score to produce a 0–1 scaled metric. For example, if the summed importance value (IV) of dominant FT species (e.g., $N=10$) is modeled to increase/no change in HS and remain dominant, the group score is 0. If the summed IV for other species (e.g., $N=6$) is modeled to increase/no change in HS but the group is not dominant, the group score is 0.25. If the HS models indicate newly suitable habitat for species (e.g., $N=4$) and the summed IV is not dominant, the group score is 0.5. The resulting HS score is thus $(0+0.5+0.5)$ / $2.5=0.4$. See definitions in Supplemental Material the competition other species might exhibit in the future.

The resulting climate pressure and HS scores were then averaged to indicate a range of potential departure under future projected conditions on a 0–1 scale for each grid cell. Higher values indicate a greater likelihood of departure from the current FT classification resulting from the combined influence of changes in climate and HS. While lower values at the landscape-level suggest little to no departure, individual species associated with the dominant FT may still show a decrease (or increase) in HS in the future and thus could impact fire regimes, forest structure, and wildlife habitat and food sources.

Case studies

We chose three sites to illustrate the information that can be gained from the departure analysis. Site 1 is the Manistee National Forest in Michigan, buffered to a minimum area of 8000 sq. km, and sites 2 and 3 correspond to a one-degree grid of latitude/longitude in Tennessee (TN) and Maine (ME) for which tree species summaries are available from the Climate Change Tree Atlas (www.fs.fed.us/nrs/atlas/). For each site, descriptions of current mean climate values and species suitable habitat are provided as a baseline.

Site 1 (Manistee NF)

The Manistee National Forest (Fig. 2a) received 874 mm of annual precipitation and had a mean annual temperature of 7.4 °C based on the 1981–2010 30-year mean (Supplemental Material Fig. 3). Growing season (May–Sep) precipitation and mean temperature were 425 mm and 16.8 °C, the mean temperatures of the coldest and warmest months were −6.7 and 20.1 °C, respectively, and an annual aridity index of −0.2 provide a baseline of the seven climate variables included in the DISTRIB-II models. The Climate Change Tree Atlas summary table (https://www.fs.fed.us/nrs/atlas/combined/resources/summaries/NatFor/) reports HS of 47 species and projects newly suitable habitat for 19 species in the future.

Site 2 (TN – S35_E86)

The landscape bounded by −86° W, 35° N and −87° W, 36° N (Fig. 2b) received 1128 mm of annual precipitation and had a mean annual temperature of 12.0 °C based on the 1981–2010 30-year mean (Supplemental Material Fig. 3). Growing season (May–September) precipitation and mean temperature were 449 mm and 18.6 °C, the mean temperatures of the coldest and warmest months were 1.9 and 20.9 °C, respectively, and an annual aridity index of −0.1 provide a baseline of the seven climate variables included in the DISTRIB-II models. The Climate Change Tree Atlas summary table (https://www.fs.fed.us/nrs/atlas/combined/resources/summaries/grid/) reports HS of 71 species and projects newly suitable habitat for 18 species in the future.

Site 3 (ME – S46_E68)

The landscape bounded by −68° W, 46° N and −69° W, 47° N (Fig. 2c) received 836 mm of annual precipitation and had a mean annual temperature of 3.3 °C based on the 1981–2010 30-year mean (Supplemental Material Fig. 3). Growing season (May–September) precipitation and mean temperature were 382 mm and 12.0 °C, the mean temperatures of the coldest and warmest months were −10.3 and 14.9 °C, respectively, and an annual aridity index of −0.01 provide a baseline of the seven climate variables included in the DISTRIB-II models. The Climate Change Tree Atlas summary table (https://www.fs.fed.us/nrs/atlas/combined/resources/summaries/grid/) reports HS for 34 species and projects newly suitable habitat for 16 species in the future.

Results

Visually for some FTs each of the four classifications generally had broad spatial agreement (Fig. 3). The two summations (Fig. 3a and b) varied slightly in the species used to define each of the 11 FTs, and while some classes were similar in area, others like Elm/Ash/Cottonwood (*Ulmus/Fraxinus/Populus*) or Oak/Pine (*Quercus/Pinus*), were not (Table 3). Likewise, the two LDA models (Fig. 3c and d) which used the same species and only differed by using 21 species to

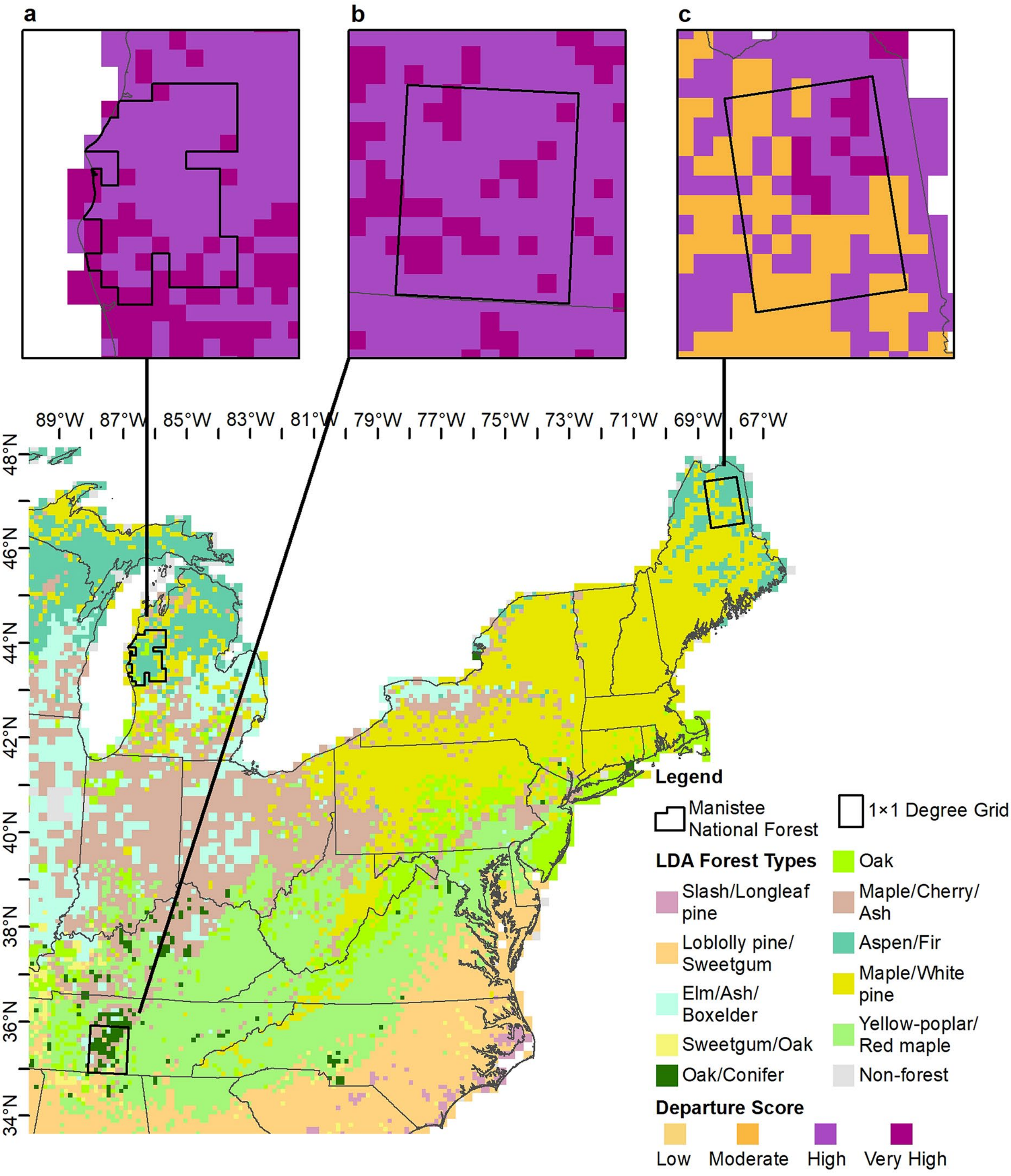


Fig. 2 Forest types classified by the seeded latent Dirichlet allocation (LDA) model under current (1981–2010) climate conditions. Three sites used to illustrate information provided by the departure analysis under climate scenario 8.5 for **a** Manistee National Forest in Michigan, **b** one-degree grid in Tennessee, and **c** one-degree grid in Maine

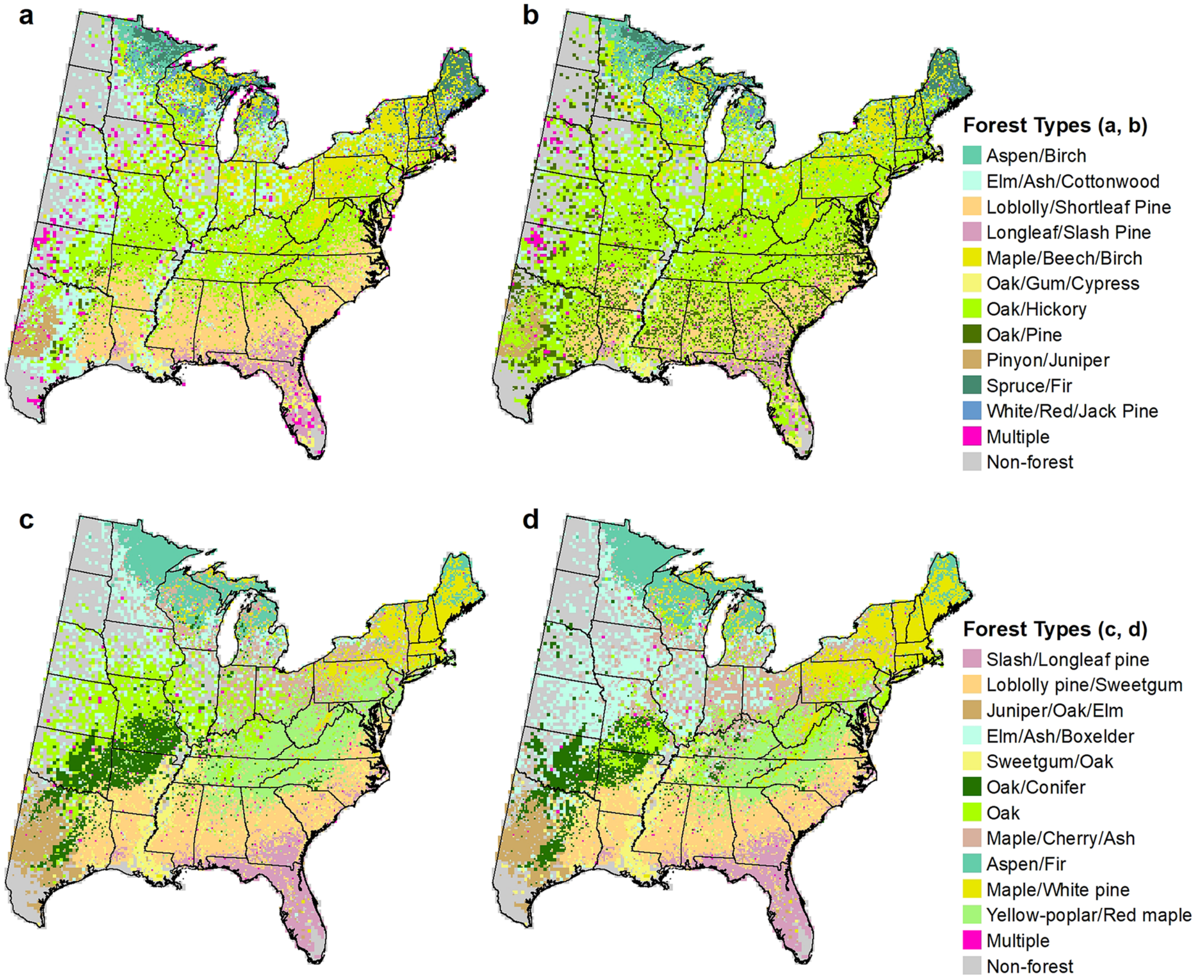


Fig. 3 Forest types defined by **a** a simple summation of species' importance values (IVs), **b** a similar algorithm used by the Forest Inventory and Analysis (FIA) program applied to IVs, **c** using a latent Dirichlet allocation (LDA) model on IVs, and **d** using a seeded LDA model with five FTs seeded by 21 species. Each method defined 11 distinct FTs using combinations of IVs from 92, 100, 104, and 104 species for a, b, c, and d, respectively. Methods **a** and **b** are based on traditional stand-level classifications of dominant species, while methods **c** and **d** use a machine learning algorithm to classify each group

seed five of the 11 FTs resulted in greater similarities among classes (Table 3).

Among the unseeded and seeded LDA models, AIC values were very similar (17,065,763 vs. 17,136,652, respectively), but indicate that the unseeded model has a slightly lower rate of information loss. However, the harmonic mean of log-likelihoods, which is better suited for comparing rates, were somewhat more distinct (−8,753,407 vs. −8,674,283, respectively) and between the harmonic means, the seeded LDA model resulted in a tighter distribution of log-likelihoods among the iterations (Supplemental Material Fig. 4). Among the 11 FTs, the seeded LDA model classified more area of the eastern U.S. as Elm/Ash/Boxelder (7.7%), Maple/White pine (11.2%), and Maple/Cherry/Ash (8.3%) and less area as Yellow-poplar/Red maple (9.6%) than the unseeded LDA model with 4.6, 8.9, 7.3, and 13.7%, respectively (Table 3; Fig. 3). Both LDA models resulted in fewer grid cells having multiple, or co-dominant FTs compared to the two summations. Using a threshold of 20%, the mapped probabilities for each seeded LDA FT shows unique spatial patterns (Fig. 4).

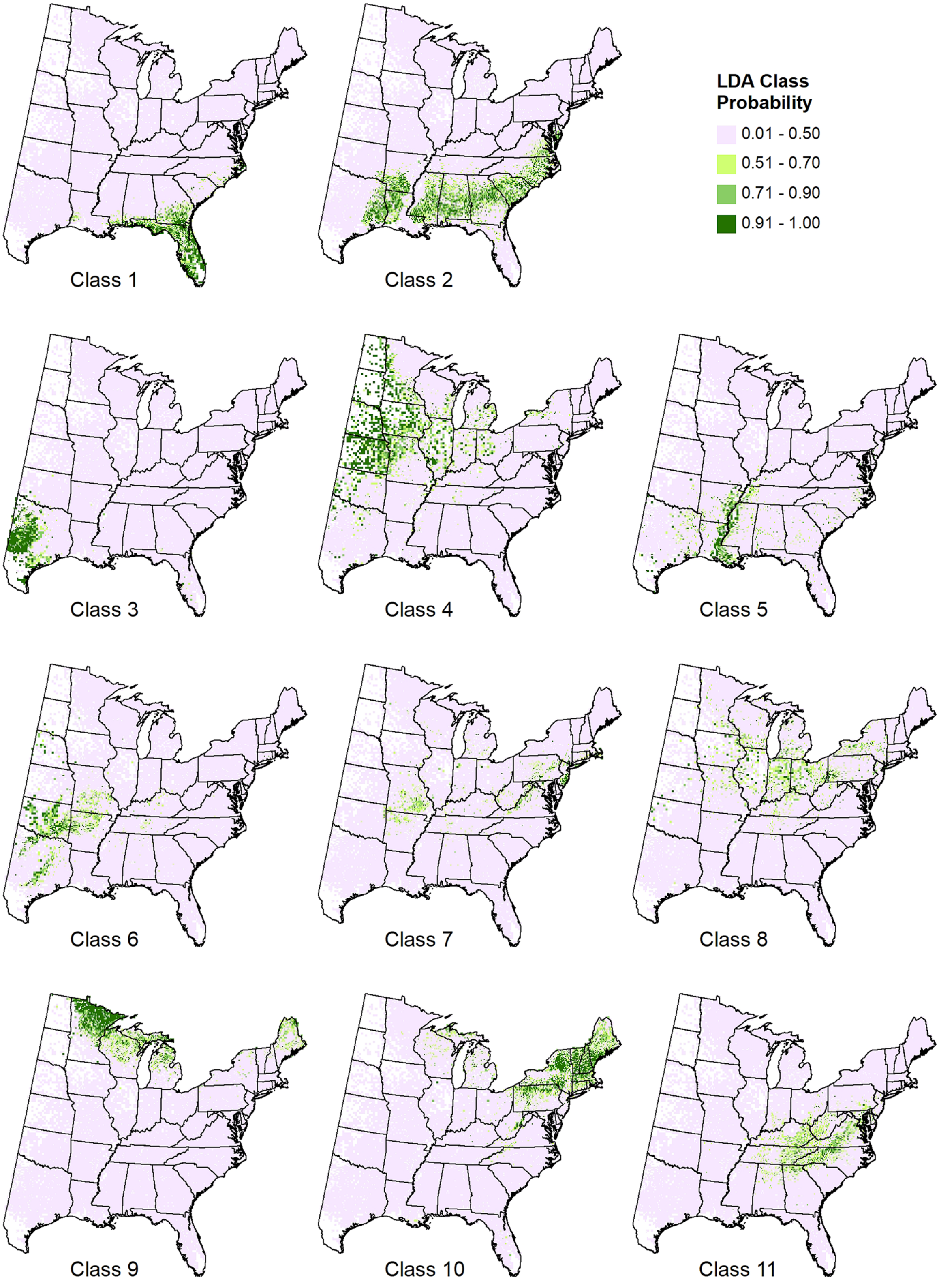
Table 3 Percent area of each FT within the eastern U.S. resulting from the four methods used to define FTs from modeled importance values. The simple and Forest Inventory and Analysis (FIA) summations use the same 92 and 100 species, respectively, but apply a different logic to assign each category. The latent Dirichlet allocation (LDA) model and the seeded LDA model used 104 species to define 11 distinct FTs, with the seeded LDA model using 21 species to aid in defining five of the 11 FTs (see Table 1). Forest type names with * indicate traditional stand-level classifications, while other names describe broad communities resulting from the LDA models (e.g., Aspen/Fir and Yellow-poplar/Red maple). These FTs are depicted in Fig. 3

Forest Type	Simple Summation	FIA Summation	Unseeded LDA Model	Seeded LDA Model
Aspen/Birch*	2.9	4.9	–	–
Elm/Ash/Cottonwood* (Elm/Ash/Boxelder)	10.6	4.0	4.6	7.7
Loblolly/Shortleaf pine* (Loblolly pine/Sweetgum)	19.0	10.9	17.1	17.3
Longleaf/Slash pine* (Slash/Longleaf pine)	3.6	2.9	5.6	5.7
Maple/Beech/Birch* (Maple/White pine)	13.7	7.9	8.9	11.2
Oak/Gum/Cypress* (Sweetgum/Oak)	3.2	1.8	5.5	5.5
Oak/Hickory* (Oak)	24.0	39.6	6.8	5.7
Oak/Pine* (Oak/Conifer)	1.5	10.1	7.1	5.8
Pinyon/Juniper* (Juniper/Oak/Elm)	1.2	0.9	2.7	2.5
Spruce/Fir*	4.0	3.3	–	–
White/Red/Jack pine*	2.8	1.8	–	–
Aspen/Fir	–	–	9.0	9.3
Maple/Cherry/Ash	–	–	7.3	8.3
Yellow-poplar/Red maple	–	–	13.7	9.6
Multiple	3.2	1.2	0.7	0.7
nonforest	10.3	10.8	10.8	10.8

The dominant probabilities ranged from 0.2 to 1, with 20% of the values below 50%. Additionally, the LDA model was more in line with the concept of defining a landscape-level FT and not imposing stand-level definitions to broader areas by allowing the machine learning algorithm to determine species communities.

The diffeR comparison of actual IV-derived dominant FTs (training data set) against the predicted IVs under current climate conditions for the simple and FIA summations resulted in 73.7 and 71.3% agreement, respectively (Supplemental Material Fig. 6a and b). The unseeded and seeded LDA models resulted in 77.6 and 84.6% agreement, respectively (Supplemental Material Fig. 2a and b). The diffeR comparison between the two LDA models for actual IV-derived dominant FTs and the predicted IVs resulted in agreements of 73.3 and 75.7%, respectively (Supplemental Material Fig. 5a and b). This suggests that the seeded LDA model produced similar results between the FIA training data (Fig. 3d) and under current conditions (Fig. 5a) as well as predicted by the unseeded model.

Except for growing season precipitation, mean temperature of the coldest month, and mean temperature of the warmest month under GCM45, CV scores were generally very high across the eastern U.S. (Supplemental Material Fig. 7) resulting from future values that are outside the recent interannual variation which produces the highest class of CV scores. However, when combined into a single climate pressure score, values range from moderate to very high (Fig. 5b, c). Projected changes to individual species' HS vary greatly within the eastern U.S. and combining only the species associated (according to SAF) with each FT provides just partial information. Thus, we also considered changes among other species and those with newly suitable habitats to indicate overall changes that could impact future FTs. The HS scores ranged from very low to very high under both climate scenarios with the GCM85 having more high and very high scores (Fig. 5d, e). Combining the climate


LDA Class Probability
0.01 - 0.50
0.51 - 0.70
0.71 - 0.90
0.91 - 1.00
Class 1
Class 2
Class 3
Class 4
Class 5
Class 6
Class 7
Class 8
Class 9
Class 10
Class 11

◀ **Fig. 4** Mapped probabilities from the seeded latent Dirichlet allocation (LDA) model of FIA actual IV for 11 FTs among the DISTRIB-II hybrid lattice of 10 and 20 km^2 grid cells

pressure and HS scores to reflect potential departure under future climate conditions suggests that 33 and 67% of the area of eastern U.S. FTs could experience high to very high departures from current conditions under the GCM45 and GCM85 scenarios respectively (Fig. 5f, g).

Case Studies

Site 1 (Manistee NF)

The seeded LDA model defined five FTs for this site, Aspen/Fir, Maple/White pine, Oak, Maple/Cherry/Ash, and Elm/Ash/Boxelder, which represent 57.2, 28.8, 10.5, 2.3, and 1.2% of the site, respectively (Fig. 2). Mean CV scores for all FTs within and surrounding the national forest indicate a high degree of pressure from changing conditions under both climate scenarios (Fig. 6). Mean HS scores varied from moderate to very high degree of change in HS, with Oak, Aspen/Fir, and Maple/White pine having an increased degree of change under RCP 8.5, while Elm/Ash/Boxelder and Maple/Cherry/Ash were on average the same under both climate scenarios. The resulting departure scores are moderate to high (Supplemental Material), due to decreased HS among associated species, increased suitability of other species, and those with newly suitable habitats in the future (Supplemental Material).

Site 2 (TN—S35_E86)

The seeded LDA model defined seven FTs for this site, Oak/Conifer, Maple/Cherry/Ash, Yellow-poplar/Red maple, Elm/Ash/Boxelder, Loblolly pine/Sweetgum, Sweetgum/Oak, and Oak, which representing 38.7, 25.9, 18.7, 6.5, 4.4, 3.1, and 2.7% of the site, respectively (Fig. 2). Mean CV scores for all FTs at this site indicate a high degree of pressure from changing conditions under both climate scenarios (Fig. 6). Among the seven FTs, mean HS scores varied from moderate to very high degree of change in HS, with increased potential changes under RCP 8.5. The resulting departure scores are high to very high across the site (Supplemental Material), mostly driven by increases in other species not associated with the current FTs and species with newly suitable habitats (Supplemental Material). Species defining the Oak and Maple/Cherry/Ash types are projected to decline while other species and newly suitable habitats increase in the future, resulting in high and very high potential departure scores.

Site 3 (ME—S46_E68)

The seeded LDA model defined two FTs for this site, Aspen/Fir and Maple/White pine, which occupy 52.7 and 47.3% of the site, respectively (Fig. 2). Mean CV scores for the two FTs indicate a high or very high degree of pressure from changing climatic conditions under each scenario respectively (Fig. 6), however, the Maple/White pine mean HS score suggests a low degree of pressure from changes in HS under both climate scenarios and results in a moderate degree of potential departure; while Aspen/Fir mean HS score suggests a moderate to high degree of HS change from the two climate scenarios resulting in an increased potential of departure, moderate and high, respectively (Supplemental Material). The potential departure of Aspen/Fir results from modeled decreased HS of associated species, increased suitability of other species, and newly suitable habitat from species currently not present (Supplemental Material). Conversely, the relatively lower departure potential for Maple/White pine is due to modest modeled increases in HS for associated species, some decreases in suitability for other species, and lower potential for newly suitable species from becoming dominant, assuming species migrate and establish within these newly suitable habitats as modeled (Prasad et al. 2016).

Discussion

Forest type classification and mapping offer a conceptual composition of forest species, and regionally some indication of site conditions (Iverson et al. 1996; Iverson et al. 2019a). Classification of dominant species developed from stand-level information are critical for local scales to inform management decisions, but when applied to large landscapes (e.g., square kilometers), these classifications can misrepresent compositional mixtures

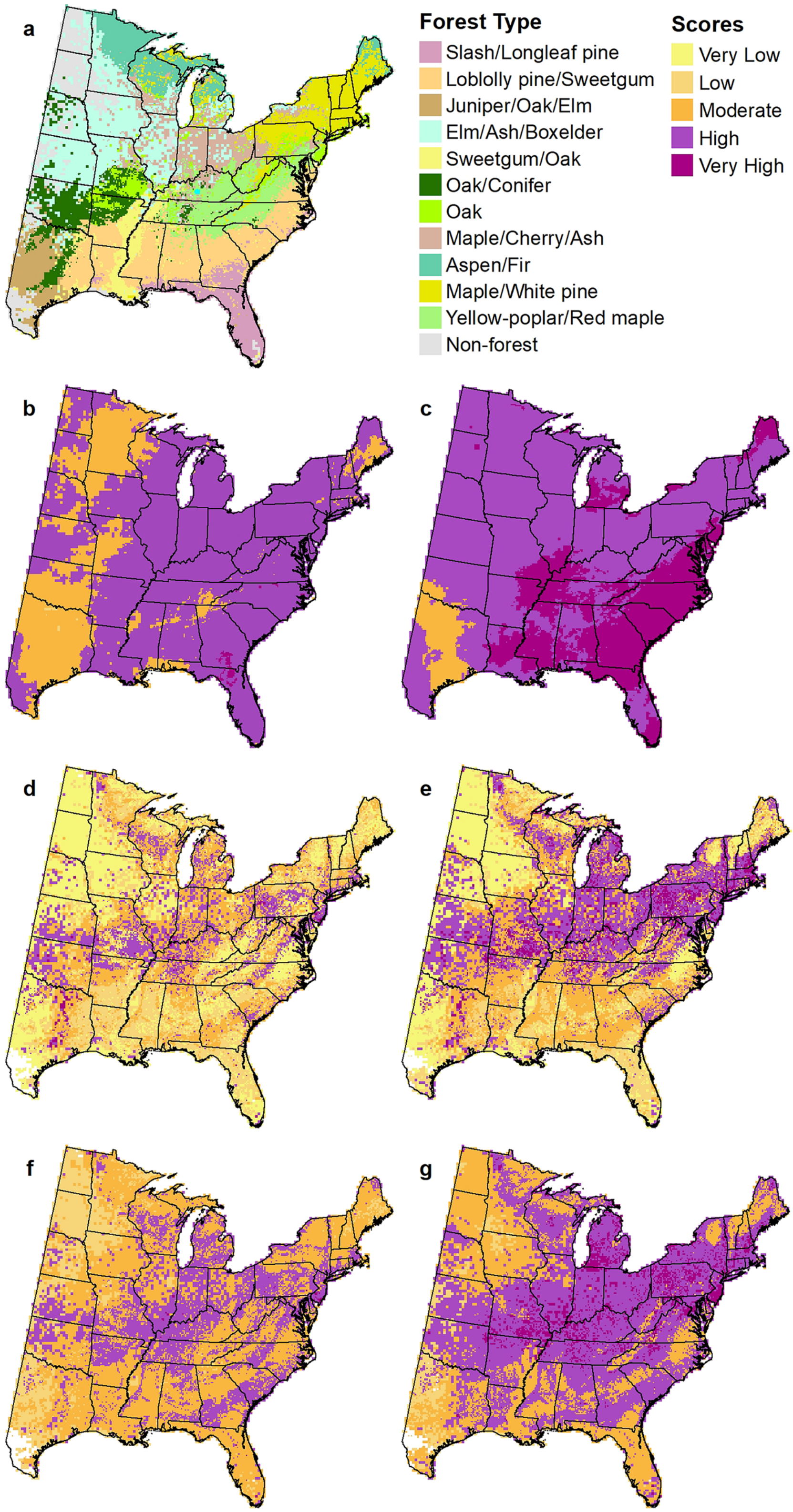


Fig. 5 **a** Forest types indicated by the seeded LDA model under current (1981–2010) climate conditions, climate pressure scores under **b** GCM45 and **c** GCM85 scenarios, habitat suitability scores under **d** GCM45 and **e** GCM85 scenarios, and departure scores under **f** GCM45 and **g** GCM85 scenarios. Departure scores are the average of climate pressure and habitat suitability scores

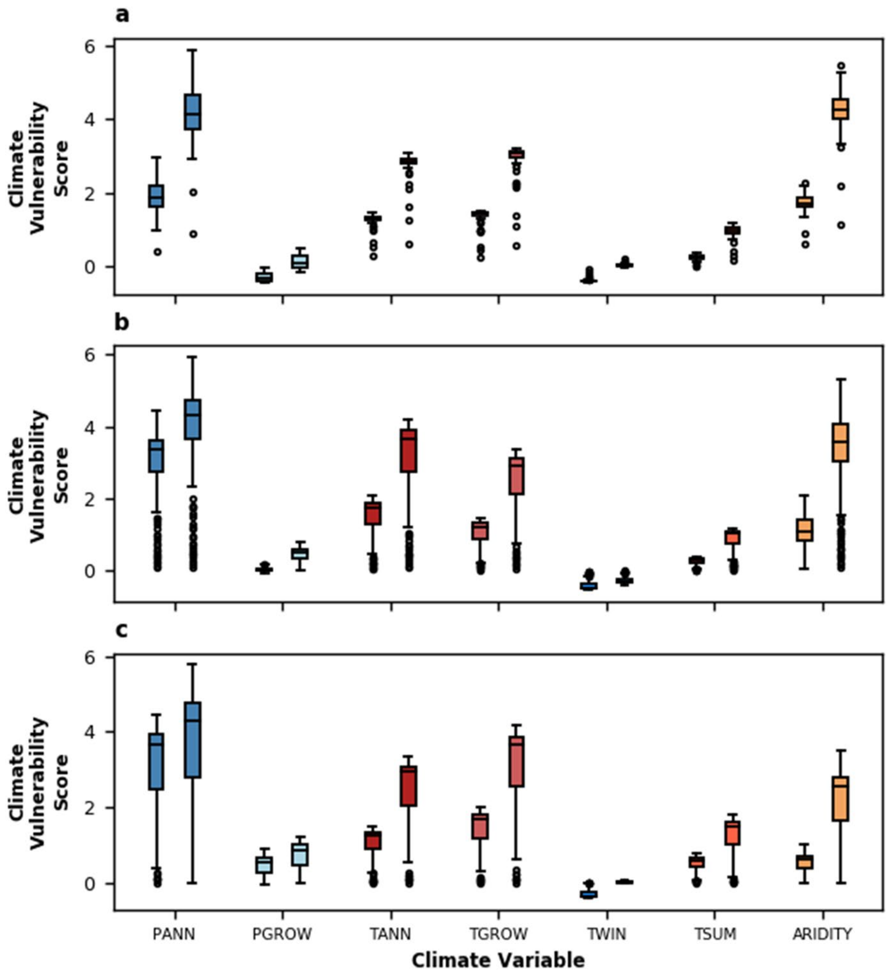


Fig. 6 For case study sites: **a** Manistee National Forest in Michigan, **b** one-degree grid in Tennessee, and **c** one-degree grid in Maine, climate vulnerability scores under climate scenarios 4.5 (left) and 8.5 (right) among the seven climate variables: annual precipitation (PANN), growing season (May-Sep) precipitation (PGROW), mean annual temperature (TANN), growing season mean temperature (TGROW), mean temperature of coldest month (TWIN), mean temperature of warmest month (TSUM), and aridity index (ARIDITY)

across fragmented landscapes. Methods such as hierarchical clustering (e.g., Costanza et al. 2017, 2018), ordination (e.g., Feilhauer et al. 2011), or machine learning algorithms can provide useful insights into unique combinations of species groupings (Valle et al. 2014; Knott et al. 2020) and represents patterns of species diversity at macroscales better than stand-levels. The two FT summations rely on *a priori* groupings of species to determine dominance of relatively large heterogeneous areas, but at the stand-level for which these groups have been developed, species diversity is usually lower than across large landscapes. This results in a scale mismatch where the larger area of the habitat model is not representative of the conditions (i.e., species diversity) defining stand-level dominance. Unsupervised machine learning algorithms such as LDA do not require such *a priori* groupings and can reveal hidden associations from the data that would otherwise be overlooked by other techniques.

Habitat suitability models, such as those used in this analysis, provide potential information (e.g., modeled niche) about where species may exist regionally, but do not replace field surveys or knowledge of where species are across the landscape. Therefore, aggregating modeled HS with spatial resolutions $>1\,km^2$ to classify FTs likely provides limited information at the forest management scale and should be used with caution. Among the resulting 11 FTs, species composition varied from stand-level definitions of species dominance and reflected the broader heterogeneous landscape (e.g., Maple/Cherry/Ash) which the habitat model grid cells represent. However, within these landscapes, managers should find patches that resemble stand-level classes, but as climate change and other disturbances occur, shifts in species composition may facilitate departures to other FTs. The LDA-derived FTs can be further impacted by changes in HS and climate pressures in the future. Therefore, combining the climate pressure and HS scores to examine departures from the former offers greater insight than just comparing which new FT is expected to become dominant.

Like all models, the LDA approach provides an approximate representation of reality, and data quality too can influence predicted results. Combining modeled HS for 104 species with varying degrees of model reliability, low to high, likely influences the LDA performance. However, limiting the analysis to only highly reliable species models may result in an incomplete picture of species composition. Removing species, especially ones that had resulting likelihoods below an acceptable threshold ($1/n_{species}$) of being associated with any of the 11 FTs improved initial performance statistics, as did converting species' frequencies from decimal values to integers. Seeding the LDA model to initiate some FTs is a possible area for greater improvements. We tried to avoid using many species traditionally associated with certain FTs as the scale at which our models were applied do not support these groupings and we wanted the LDA model to define species associations. Though, as shown if Fig. 3c and d, the unseeded model defined some of the seeded FTs. However, the species associated with these FTs and the predicted likelihoods differed among the two LDA models (Supplemental Material). It has also been suggested that spatially constraining the LDA training data may improve results (Wang and Grimson 2007; Philbin et al. 2011). However, for our analysis, spatially constraining the training data did not seem appropriate given our objective to define landscape-level FTs whereby limiting possible species combinations across large ranges of HS could result in missing information. This is potentially another area of improvement and further research.

The CV scores developed by Triepke et al. (2019) used the standard deviation of interannual variability to define vulnerability of future conditions. In developing weights for the CV scores, we applied an equal weighting, one-seventh since each climate variable could have the same chance of being important among the randomForest-derived species HS models. It is possible that deeper examination of the randomForest output could reveal model preference for specific climate variables within other weighting schemes (e.g., precipitation, temperature, aridity equally weighted as one-third, or annual and seasonal variables as one-half). However, this is beyond the scope of the current analysis and would reflect species-specific importance. Determining the influence of climate pressure and attributing changes in HS to evaluate potential departures identifies where changes in FTs could occur but doesn't provide information as to what is responsible for any changes. The emergent patterns resulting from the CV scores (Fig. 5b, c) are reflective of the climate scenarios and influenced by the amount of warming and changes to precipitation. Likewise, the patterns of HS scores (Fig. 5d, e) are to a large degree influenced by species diversity, which tends to be higher in the East and transitions from forests to grassland prairies. As depicted in Fig. 5a, several FTs intermingle within the central portion of the eastern U.S. and these transitions zones appear to have higher scores of potential change.

The three sites are climatically distinct (Supplemental Material Fig. 3) and are projected to have varying amounts of change among the seven climate variables included in the DISTRIB-II models under the two scenarios of climate change, with annual precipitation and aridity having the largest range (Fig. 6). Combining information from potential changes in climate conditions and HS for many species helps to indicate where future conditions may diverge and favor different FTs. The average summed IVs for current FTs (Supplemental Material) illustrate that attributing future conditions to specific species communities contains considerable uncertainties. Climate change alone may not result in expected changes to FTs, as some species groupings may benefit from warmer conditions. For example, HS of oaks (*Quercus*) have been modeled to increase under warmer conditions, especially southern species with newly suitable habitats north of current ranges (Iverson et al. 2019b); however, the absence of fire has been shown to limit oak competitiveness (Nowacki and Abrams 2015; Iverson et al. 2017), and there is uncertainty as to how dry future conditions will be in the East, which would affect the rate and amount of compositional changes (Pederson et al. 2015). Additionally, wetter conditions, if paired with only moderately warming conditions, may favor mesic species like maples (*Acer*) and birch (*Betula*), which would tend to outcompete the more shade intolerant species like the oaks (Nowacki and Abrams 2008, 2015).

Conclusions

Forest type classification developed from stand-level information doesn't always reflect the landscape-level patterns of broader species diversity and fragmentation. Advances in machine learning algorithms and novel techniques provide insights to better use information gleaned from HS modeling efforts. Our application of a text mining algorithm to define FTs from modeled suitable habitats and developing departure scores to assess potential future changes is a novel addition to the important quest for understanding future forest community structures. We intend this information to assist forest managers planning for the next half of the century by identifying areas with greater departures resulting from not only climatic changes, but also species' suitability.

We realize that the FTs defined by the LDA model differ from more traditional stand-level classifications. Although some similarities were apparent, classes may not reflect site-specific expectations, but rather appear more representative of the broader landscapes depicted by the FT models. This context provides a unique perspective that forest managers may find useful when interpreting the DISTRIB-II models.

Acknowledgements We are grateful to the field crews that collect the Forest Inventory and Analysis data used in our models, in which this work would not be possible as well as sources of digital environmental datasets. We thank Jonathan Knott, Andy Gougherty, and two anonymous reviewers who provided comments on earlier versions of this manuscript.

Author contributions All authors contributed to the conceptualization and design of the analysis. Data preparation and modeling was performed by MPP. All authors reviewed model outputs and results. The manuscript was written by MPP with contributions from all co-authors.

Funding This research was funded by the U.S. Department of Agriculture Forest Service.

Availability of data and material DISTRIB-II model output used to classify forest types is available from https://www.fs.fed.us/nrs/atlas, forest type and departure maps will also be included. All other material will be made available upon request.

Declarations

Conflict of interest The authors have no conflicts of interest or competing interests to declare that are relevant to the content of this article.

References

Arner SL, Woudenberg S, Waters S et al (2003) National algorithms for determining stocking class, stand size class, and forest type for forest inventory and analysis plots. U.S. Department of Agriculture, Forest Service, Northeastern Research Station, Newtown Square, p 65

Arun R, Suresh V, Veni Madhavan CE, Narasimha Murthy MN (2010) On Finding the natural number of topics with latent Dirichlet allocation: some observations. In: Zaki MJ, Yu JX, Ravindran B, Pudi V (eds) Advances in knowledge discovery and data mining. PAKDD 2010. Springer, Berlin Heidelberg, pp 391–402

Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. J Mach Learn Res 3(4–5):993–1022

Cao J, Xia T, Li J, Zhang Y, Tang S (2009) A density-based method for adaptive LDA model selection. Neurocomputing 72(7–9):1775–1781

Clements FE (1916) Plant succession: an analysis of the development of vegetation. Carnegie Institution of Washington, Washington

Costanza JK, Coulston JW, Wear DN (2017) An empirical, hierarchical typology of tree species assemblages for assessing forest dynamics under global change scenarios. PLoS ONE 12(9):e0184062

Costanza JK, Faber-Langendoen D, Coulston JW, Wear DN (2018) Classifying forest inventory data into species-based forest community types at broad extents: exploring tradeoffs among supervised and unsupervised approaches. For Ecosyst 5(1):8

Eyre FH (ed) (1980) Forest Cover Types of the United States and Canada. Society of American Foresters, Washington, D.C.

Fauth JE, Bernardo J, Camara M, Resetarits WJ, Van Buskirk J, McCollum SA (1996) Simplifying the Jargon of Community Ecology: a conceptual approach. Am Nat 147(2):282–286

Feilhauer H, Faude U, Schmidtlein S (2011) Combining Isomap ordination and imaging spectroscopy to map continuous floristic gradients in a heterogeneous landscape. Remote Sens Environ 115(10):2513–2524

Gleason HA (1926) The individualistic concept of the plant association. Bull Torrey Bot Club 53(1):7–26

Griffiths TL, Steyvers M (2004) Finding scientific topics. Proc Natl Acad Sci USA 101(suppl 1):5228

Hagen-Zanker A (2009) An improved Fuzzy Kappa statistic that accounts for spatial autocorrelation. Int J Geogr Inf Sci 23(1):61–73

Iverson LR, Scott CT, Dale ME, Prasad A (1996) Development of an Integrated Moisture Index for Predicting Species Composition. In: Kohl M and Gertner GZ (eds) Caring for the forest: research in a changing world, statistics, mathematics and computers, proceedings of the meeting of IUFRO S4.11-00, Tampere, Finland 1995. Swiss Federal Institute for Forest, Snow and Landscape Research, Birmensdorf, Switzerland, pp 101–116

Iverson LR, Prasad AM (2001) Potential changes in tree species richness and forest community types following climate change. Ecosystems 4(3):186–199

Iverson LR, Prasad AM, Matthews SN, Peters M (2008) Estimating potential habitat for 134 eastern US tree species under six climate scenarios. For Ecol Manag 254(3):390–406

Iverson LR, Hutchinson T, Peters M, Yaussy D (2017) Long-term response of oak-hickory regeneration to partial harvest and repeated fires: Influence of light and moisture. Ecosphere 8(1):e01642

Iverson LR, Bartig JL, Nowacki GJ et al (2019a) USDA Forest Service section, subsection, and landtype descriptions for southeastern Ohio. U.S. Department of Agriculture, Forest Service, Northern Research Station. Newtown Square, PA

Iverson LR, Peters MP, Prasad AM, Matthews SN (2019b) Analysis of Climate Change Impacts on Tree Species of the Eastern US: Results of DISTRIB-II Modeling. Forests 10(4):302

Jagarlamudi J, Daumé H, Udupa R (2012) Incorporating lexical priors into topic models. In: Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics. Avignon, France, Association for Computational Linguistics, pp 204–213

Joshi PK, Rawat A, Narula S, Sinha V (2012) Assessing impact of climate change on forest cover type shifts in Western Himalayan Eco-region. J Forestry Res 23(1):75–80

Konrad M (2020) tmtoolkit: Text mining and topic modeling toolkit. Version 0.10.0

Knott JA, Jenkins MA, Oswalt CM, Fei S (2020) Community-level responses to climate change in forests of the eastern United States. Glob Ecol Biogeogr 29(8):1299–1314

Liaw A, Wiener M (2002) Classification and regression by randomForest. R news 2(3):18–22

Mimno DM, Wallach HM, Talley E, Leenders M, McCallum A (2011) Optimizing semantic coherence in topic models. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Edinburgh United Kingdom, pp 262–272

Moss R, Babiker W, Brinkman S et al (2008) Towards new scenarios for the analysis of emissions, climate change, impacts, and response strategies. technical summary. Intergovernmental Panel on Climate Change, Geneva, p 25

Nagel LM, Palik BJ, Battaglia MA et al (2017) Adaptive silviculture for climate change: a national experiment in manager-scientist partnerships to apply an adaptation framework. J For 115(3):167–178

Nevins MT, D'Amato AW, Foster JR (2021) Future forest composition under a changing climate and adaptive forest management in southeastern Vermont, USA. For Ecol Manag 479:118527

Nowacki GJ, Abrams MD (2008) The demise of fire and "Mesophication" of forests in the Eastern United States. Bioscience 58(2):123–138

Nowacki GJ, Abrams MD (2015) Is climate an important driver of post-European vegetation change in the Eastern United States? Glob Chang Biol 21(1):314–334

Pederson N, D'Amato A, Dyer J et al (2015) Climate remains an important driver of post-European vegetation change in the eastern United States. Glob Chang Biol 21(6):2105–2110

Pontius RG Jr, Santacruz A (2019) diffeR: Metrics of Difference for Comparing Pairs of Maps or Pairs of Variables R package version 0.0–6

Peters MP, Iverson LR, Prasad AM, Matthews SN (2019) Utilizing the density of inventory samples to define a hybrid lattice for species distribution models: DISTRIB-II for 135 eastern U.S. trees. Ecol Evol 9(15):8876–8899

Philbin J, Sivic J, Zisserman A (2011) Geometric latent dirichlet allocation on a matching graph for large-scale image datasets. Int J Comput Vis 95(2):138–153

Prasad AM, Iverson LR, Matthews SN, Peters MP (2016) A multistage decision support framework to guide tree species management under climate change via habitat suitability and colonization models, and a knowledge-based scoring system. Landsc Ecol 31(9):2187–2204. https://doi.org/10.1007/s10980-016-0369-7

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155(2):945–959. https://doi.org/10.1093/genetics/155.2.945

Python Software Foundation Python Language Reference. version 2.7.16 edn. Python Software Foundation

R Core Team (2018) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Ruefenacht B, Finco MV, Nelson MD et al (2008) Conterminous US and Alaska forest type mapping using forest inventory and analysis data. Photogramm Eng Remote Sens 74(11):1379–1388

Singh V (2017) Topic Modeling with Guided latent Dirichllocation, GuidedLDA: Guided Topic modeling with latent Dirichlet allocation. Version 2.0.0

Tchebakova NM, Parfenova EI, Korets MA, Conard SG (2016) Potential change in forest types and stand heights in central Siberia in a warming climate. Environ Res Lett 11(3):035016

Triepke FJ, Muldavin EH, Wahlberg MM (2019) Using climate projections to assess ecosystem vulnerability at scales relevant to managers. Ecosphere 10(9):e02854

Turner MG, Gardner RH, O'Neill RV (2003) Landscape ecology in theory and practice pattern and process. Springer, New York

Valle D, Baiser B, Woodall CW, Chazdon R (2014) Decomposing biodiversity data using the Latent Dirichlet Allocation model, a probabilistic multivariate statistical method. Ecol Lett 17(12):1591–1601

Webb T, Bartlein PJ (1992) Global changes during the last 3 million years—climatic controls and biotic responses. Annu Rev Ecol Syst 23:141–173

Wu J, Hobbs RJ (2007) Key topics in landscape ecology. Cambridge University Press, Cambridge

Wu J, Qi Y (2000) Dealing with Scale in Landscape Analysis: An Overview. Geographic Inform Sci 6(1):1–5. https://doi.org/10.1080/10824000009480528

Wang X, Grimson E (2007) Spatial Latent Dirichlet Allocation. In: Platt J, Koller D, Singer Y, Roweis S (eds) 20th International Conference on Neural Information Processing Systems, Vancouver, B.C., Canada 2007. vol 20, pp 1577–1584

Ye F, Shi Z, Shi Z (2009) A comparative study of PCA, LDA and Kernel LDA for image classification. In: 2009 International Symposium on Ubiquitous Virtual Reality, Guangju, Korea (South) 2009. IEEE, pp 51–54. https://doi.org/10.1109/ISUVR.2009.26