

# Community-level responses to climate change in forests of the eastern United States

Jonathan A. Knott<sup>1</sup>  | Michael A. Jenkins<sup>1</sup>  | Christopher M. Oswalt<sup>2</sup>  | Songlin Fei<sup>1</sup> 

<sup>1</sup>Department of Forestry and Natural Resources, Purdue University, West Lafayette, Indiana

<sup>2</sup>USDA Forest Service Southern Research Station, Knoxville, Tennessee

## Correspondence

Songlin Fei, Department of Forestry and Natural Resources, Purdue University, 715 W State St, West Lafayette, IN, 47907, USA.  
Email: sfei@purdue.edu

## Funding information

Purdue Climate Change Research Center; National Institute of Food and Agriculture, Grant/Award Number: 11207356; U.S. Forest Service, Grant/Award Number: 19-JV-11242305-102; National Institute of Food and Agriculture; Forest Service

Editor: Naia Morueta-Holme

## Abstract

**Aim:** Climate change has impacted forest ecosystems, leading to species-level tree migration. However, the impacts of climate change on forest communities are mostly unknown. Here, we assess changes to forest communities at three scales: within-community changes in species composition, individual community spatial shifts and changes across all communities.

**Location:** Eastern USA.

**Major taxa studied:** Forest tree species.

**Methods:** Using a region-wide forest inventory dataset from the United States Department of Agriculture Forest Service's Forest Inventory and Analysis Program with > 70,000 plots, we identified forest communities using the latent Dirichlet allocation method. We analysed changes in species composition within communities and assessed community-level spatial shifts over the last three decades to quantify the responses of individual communities to climate change. We used the distribution of forest communities across climate conditions to predict where communities could migrate to during the study period and compared climate-predicted shifts with observed community shifts. Changes across all communities were modelled as a function of climatic and non-climatic variables using generalized linear mixed-effects models.

**Results:** We identified 12 regional forest communities of the eastern USA, which varied in their stability of species composition over the study period. All communities experienced relatively short yet significant shifts in their spatial distribution (median = 8.0 km/decade). Historical climate and changes in seasonal temperature variability were the best predictors of change across all communities. However, the distance and direction of individual community migration were poorly predicted by climate change, and the observed direction was often the opposite of the predicted direction.

**Main conclusions:** Forest communities shifted their distributions over the last three decades, but climate change outpaced the rate of community migration. Continued lags between climate change and forest community responses and the lack of migration in the direction predicted by climate change might lead to the inability of forests to keep up with changing climate.

## KEYWORDS

climate change, forest communities, FIA, latent Dirichlet allocation, spatial shifts, tree migration

## 1 | INTRODUCTION

Forest ecosystems across the USA are threatened by human disturbance (Riitters, Coulston, & Wickham, 2012; Vanderwel & Purves, 2014), invasion of non-native plants and pests (Fei, Morin, Oswalt, & Liebhold, 2019; Oswalt et al., 2015), and climate change (Fei et al., 2017; Iverson & Prasad, 1998). Among these major threats, climate change has been shown to cause tree species to shift their distributions. Often, species have moved to higher latitudes or elevations in response to temperature change (Lenoir, Gégout, Marquet, De Ruffray, & Brisse, 2008; Woodall et al., 2009; Zhu, Woodall, & Clark, 2012), but westward movement has also been observed in response to change in precipitation (Fei et al., 2017). Changes in the distribution of forest species can have severe consequences for ecosystem functioning; however, changes at the community level, impacting not only the species composition but also the interactions among species, might be even more important to ecosystem functioning (Loreau et al., 2001; Symstad, Tilman, Willson, & Knops, 1998). Understanding the effects of climate change and other large-scale threats to forest communities is important for predicting the future sustainability of forests and the services they provide.

Ecologists have attempted to define communities for more than a century, but communities are dynamic, and their stability is increasingly questioned owing to climate change. Early attempts at defining communities, such as those of Clements (1916) and Gleason (1926), revealed multiple conflicting views of ecological communities. Clements (1916) argued that communities acted as superorganisms that matured into a climax community determined by regional macroclimate, whereas Gleason (1926) argued that communities were artificial groupings of overlapping species distributions created by individualistic responses to environmental gradients. In addition to the difficulty of defining communities, Braun (1950) identified large-scale, subcontinental patterns of forest species distributions, which Davis (1983) showed were the result of post-glaciation migrations of individual species over the last five millennia. As a result of these migration patterns, distinct forest regions emerged that were considered mostly stable in the recent past (Braun, 1950), but their current and future stability is questionable given decade-scale tree species migrations in response to recent climate change (e.g., Fei et al., 2017; Woodall et al., 2009; Zhu et al., 2012) that rival the distance of these historical millennium-scale post-glaciation migrations.

In addition to difficulty in defining communities, data availability and computational complexity often limited earlier attempts at description and analysis of changes to ecological communities. Many early studies were based on limited, local-scale samples and used simplified statistical tests or anecdotal evidence. The availability of large-scale datasets (e.g., Forest Inventory and Analysis from the United States Department of Agriculture [USDA] Forest Service; FIA) containing tens of thousands to millions of observations and the advancements in analytical techniques for big data (e.g., Bayesian statistics and machine learning) have allowed researchers to confirm the patterns in species co-existence observed by earlier

studies and provide greater detail about these communities (Dyer, 2006; Peters et al., 2014). However, these large-scale community analyses still have underlying problems related to the abundance and complexity of data. First, many multivariate methods rely on the creation of a distance matrix in multidimensional species space to determine whether they exhibit different species compositions (Costanza, Coulston, & Wear, 2017; Legendre & Legendre, 2012). These approaches can be influenced by the abundance of non-overlapping species (i.e., the “double-zero” problem), where patterns are driven by the lack of species co-occurrence (Legendre & Legendre, 2012). As such, distance metrics used in multivariate analyses need to be considered carefully in order to minimize these effects (Dyer, 2006). Multivariate methods are good at detecting abrupt changes between the communities, but may fail to detect differences when transitions between communities occur slowly (Valle, Baiser, Woodall, & Chazdon, 2014). In addition, the creation and analysis of a distance matrix for tens of thousands of samples can become computationally unfeasible (e.g., a distance matrix between 80,000 samples, the approximate size of the FIA database for the eastern USA, contains 6.4 billion entries). Second, irregular sampling density can cause community delineations to be biased by areas where forests are more abundant. Regions such as the southern Appalachian Mountains or the upper Midwest that have large regions of continuous forest with many sampling sites in the FIA database can drive the patterns found in traditional clustering and multivariate methods. Thus, the results of these methods may provide great detail about one region but little detail about a more sparsely forested or sampled region. There is a need to address these problems by using different models of species co-existence.

Here, we take a multiple-lens view of forest community changes over time by assessing changes at multiple scales: (a) within-community changes in species composition; (b) spatial shifts of individual communities; and (c) changes in composition and dominance across all communities. We first needed to identify forest communities, and we used the latent Dirichlet allocation (LDA) topic model (Blei, Ng, & Jordan, 2003; Valle et al., 2014) to identify forest communities within two datasets (T1, 1980s and T2, 2010s), each containing > 70,000 FIA plots. As such, communities in this study are similar to those considered in traditional local-scale ecological studies but are more generalized given the large amount of data used to identify consistent patterns in species associations. The resulting communities, therefore, represent regional species assemblages or regional forest types similar to those defined by Braun (1950), Dyer (2006) or Costanza et al. (2017). Our first assessment of forest community changes was to analyse: (a) changes in species rank within a community (i.e., changes in dominant species of the community); (b) changes in the contribution of species to a community (i.e., shifts in the proportion/abundance of each species in each community); and (c) the gain or loss of species over time. These analyses would reveal whether new assemblages of species are formed or if the species composition of existing communities is shifting. Our second assessment of forest community changes was to test whether communities were shifting their geographical distributions and if the community migrations were tracking with climate change. Given recent

knowledge of spatial shifts at the species level (Fei et al., 2017; Zhu et al., 2012), we aimed to test whether forest communities were also shifting their distributions in response to climate change. Finally, our third assessment of forest community changes was to identify areas with high levels of community change and test whether these areas were also experiencing greater levels of climate change. By taking this multiple-lens view of forest community responses to climate change, we can gain a better understanding of the ways in which forest communities change over time and can use this information to address the future sustainability of forest communities in a continuously changing global climate.

## 2 | METHODS

### 2.1 | Forest inventory data

The USDA Forest Service's FIA program is a nationwide sampling effort of approximately one plot per 2,428 ha (6,000 acres) of forested land, containing c. 130,000 plots nationwide and c. 80,000 plots in the eastern USA. We selected FIA data from the eastern USA, encompassing 37 states and the following four ecoregions (Cleland et al., 2007): the northern hardwood region (NHR), the central hardwood region (CHR), the forest-prairie transition region (FPTR) and the southern pine-hardwood region (SPHR) (Supporting Information Table S1; Figure S1). We compiled two datasets (T1 and T2) to assess changes over the last three decades. Before the early 2000s, the FIA program used a periodic sampling protocol, in which states completed sampling every 10–15 years. More recently, a panel system for FIA sampling has been used, where each year partial sampling is completed and states report sampling every 5–7 years when all plots in the state have been sampled. Therefore, the year of completion varied across the eastern USA for both time periods. For the first period (T1), we used the first measurement available in 1980 or later (a full periodic sample from each state completed in a single year from 1980 to 1995). For the second period (T2), we used measurements from sampling that concluded in 2015–2017 (the compiled sampling that was carried out over the previous 5–7 years). The median interval between T1 and T2 sampling periods in each state was 31 years.

Ideally, we would have used data for all species available in the FIA database. However, some species were aggregated to the genus level in surveys from certain states at T1. For example, coastal states from Virginia to Florida have *Celtis* genus-level records at T1 but have *Celtis occidentalis* and *Celtis laevigata* recorded separately at T2. Likewise, some genera (such as the *Carya* genus) were identified to the genus level at T1 across most of the range. Therefore, we assigned species with only genus-level samples at T1 to their genus-level species code even when species-level data were collected at T2. Likewise, ash species (*Fraxinus* species) were aggregated to the genus level to reduce the effects of uncertainty from known identification issues (e.g., overlapping ranges of species with similar morphological characteristics). Additional species that were not aggregated to the genus level were

considered rare and removed from the dataset if they occurred in <300 plots (following Fei et al., 2017). The final species list included 138 species of interest (out of c. 263 species in the FIA database for the eastern USA) that were aggregated into 85 new species or genus labels (hereafter, "species"; Supporting Information Table S2).

We used three metrics to measure species abundance in each plot: relative basal area, relative stem density and importance value (average of relative basal area and relative density). These metrics were derived for all stems and two separate size classes: < 5.1 cm diameter at breast height (d.b.h.) (hereafter, saplings) and > 5.1 cm d.b.h. (hereafter, trees). The use of relative metrics allowed each plot to be weighted equally (i.e., the sum of the importance values for all species in each plot was equal). All values were rounded up to the nearest integer percentage to accommodate for the LDA model (which was built using integer word counts) and to avoid removing species from samples where they had low abundance (with a metric value between 0 and .5%). The plot-level data were used to fit the LDA model. However, to account for the spatial mismatch between T1 and T2 (i.e., not all plots inventoried at T1 were re-inventoried at T2 or vice versa), we aggregated plot-level results to a 1,452 km<sup>2</sup> hexagon tessellation following Fei et al. (2017). To reduce the potential biases in our analysis of spatial shifts caused by hexagons with few FIA plots (i.e., dramatic community turnover attributable to mismatch in sampling), we selected hexagons that contained > 10 FIA plots at both T1 and T2. A final total of 89,231 plots at T1 and 75,715 plots T2 fell into 1,813 hexagons with > 10 FIA plots.

### 2.2 | Identifying forest communities with the latent Dirichlet allocation model

The latent Dirichlet allocation topic model is a Bayesian hierarchical model that was developed to identify topics in text data based on the frequency and co-occurrence of words across hundreds to thousands of documents (Blei et al., 2003). It has been used widely for text mining (e.g., Tirunillai & Tellis, 2014), but has recently been expanded to other uses in the natural sciences, such as remote sensing, vegetation classification and community detection (Damgaard, 2015; Tang et al., 2013; Valle et al., 2014). For community detection, the analogy to traditional text mining is clear; communities (or topics) are based on the frequency and co-occurrence of species (or words) across hundreds to thousands of plots (or documents). The LDA model is a Bayesian hierarchical model that assumes a generative process to forest (or document) sampling (or creation) (Blei et al., 2003; Valle et al., 2014). This assumption means that first a forest (or document) selects a number of communities (or topics) to include in the sample, following a Dirichlet distribution. Then, for each community (or topic), a number of species (or words) related to the community (or topic) are chosen. The posterior distributions of the LDA model contain two main components: (a) the proportion of each community in each sample (or each topic in each document); and (b) the proportion of each species in each community (or each word in each topic).

We ran the LDA model on the T1 data using the built-in Gibbs sampler in the “topicmodels” package in R (Grün & Hornik, 2011), with 5,000 iterations and a 500 iteration burn-in period. The LDA model requires specification of the number of groups a priori. Given that we did not have a priori knowledge of the exact number of communities in the FIA data, we used an iterative approach, increasing the number of communities,  $k$ , by one, from two to 50. We used the plot-level data with importance value for all stems to fit the LDA model. The models were evaluated for the goodness-of-fit and complexity using a suite of four metrics (Arun, Suresh, Veni Madhavan, & Narasimha Murthy, 2010; Cao, Xia, Li, Zhang, & Tang, 2009; Deveaud, SanJuan, & Bellot, 2014; Griffiths & Steyvers, 2004) from the “ldatuning” package in R (Nikita, 2016). The value of  $k$  is chosen by maximizing two of the metrics (Deveaud et al., 2014; Griffiths & Steyvers, 2004) and minimizing the other two (Arun et al., 2010; Cao et al., 2009); therefore, we included possible values of  $k$  that were at these maximum and minimum values of the four metrics. Owing to the asymptotic nature of the metrics, we also included breakpoints indicated by multivariate adaptive regression splines (MARS regression; Friedman, 1991) using the “earth” package in R (Milborrow, 2017). Finally, we used the posterior distributions from the candidate models to evaluate an ecologically based metric, the number of species with proportion  $> 1/n_{\text{species}}$  in each community, to remove models that contained communities that were composed of only one species. To test whether there were differences between input data type (i.e., relative basal area, relative density, or importance value for saplings, adults or all stems), we used the best-fitting model to initialize new models with different input data and tested for differences in species composition using Mantel tests using the “ade4” package in R (Dray & Dufour, 2007). After identifying the best-fitting model from the T1 data, we used the posterior distributions to initialize the T2 model. Doing so allowed the species composition of the communities to begin aligning with the T1 communities but then change progressively to fit the T2 data, reflecting the way that communities change naturally over time. We compared the species composition between T1 and T2 to identify communities that changed the dominant species or lost/gained species.

To visualize the distribution of identified communities and calculate measures of community change over time, we aggregated our results to the 1,452 km<sup>2</sup> hexagon tessellation (Supporting Information Figure S1) by taking the average community proportion across all plots within a hexagon. We then created regional forest community maps for T1 and T2 by mapping the community with the highest proportion in each hexagon. We used these maps to identify hexagons where the dominant community type changed between T1 and T2 and create a transition matrix containing the proportion of hexagons changing from one dominant community type to another throughout the study period. Non-forested areas were masked out using the 2011 National Land Cover Database for visualization purposes only (<https://www.mrlc.gov/data>). A GIS shapefile with the posterior distribution of communities in each hexagon and a spreadsheet of the posterior distribution of species in each community are available at <https://www.doi.org/10.4231/GCE5-ZY59>.

### 2.3 | Climatic and non-climatic forest condition data

The PRISM climate dataset (4 km resolution) was used to create normal values of climate for 30 years leading up to the T1 FIA sampling (1951–1980; hereafter, “historical climate conditions”) and 30 years during the study period (1986–2015; hereafter, “study period conditions”; PRISM Climate Group, Oregon State University, <http://prism.oregonstate.edu/>). The PRISM dataset contains the monthly mean, maximum and minimum temperature and monthly accumulated precipitation. We calculated the mean of monthly mean temperature (MAT), the minimum of monthly minimum temperature (TMIN), the maximum of monthly maximum temperature (TMAX) and the sum of monthly precipitation (TAP) across the 12 months in each year. We also calculated climatic variability by taking the standard deviation (SD) of monthly precipitation (PSD), SD of monthly minimum temperature (TMINS), SD of monthly maximum temperature (TMAXSD) and the temperature range (annual maximum temperature minus annual minimum temperature; TRANGE). Therefore, TRANGE, TMINS and TMAXSD are measures of seasonal temperature variability, and PSD is a measure of seasonal precipitation variability. Temperature and precipitation conditions are closely linked to drought severity (Hu & Willson, 2000; McEwan, Dyer, & Pederson, 2011); therefore, we used Palmer drought severity index (PDSI) data (from the same two periods, 1951–1980 and 1986–2015) to test whether drought severity was a better predictor of community change than the PRISM-derived temperature and precipitation variables (available at <https://wrcc.dri.edu/wwdt/>). Then, we averaged each variable across 30 years before and during the study period (1951–1980 and 1986–2015, respectively). We calculated the difference between historical climate conditions and study period conditions for each variable to use as indicators of climate change (hereafter, “climate-change variables”).

A variety of non-climatic factors can influence forest dynamics, and we aimed to test forest developmental stage, nitrogen deposition and fire frequency as potential drivers of community change (hereafter, “non-climatic variables”). Given that we do not have reliable information on the tree or stand age, we used hexagon-level total basal area (BA) at T1 (across all species in all plots in each hexagon) as a proxy for forest developmental stage, following Fei et al. (2017). We also calculated hexagon-level total basal area at T2 and took the difference between T1 and T2 as a measure of forest development over the study period. Nitrogen deposition data (NDEP) during a portion of the study period (2000–2015) were aggregated at the hexagon level (available at <http://nadp.slh.wisc.edu/>), following Jo, Fei, Oswald, Domke, and Phillips (2019). We used the kernel density tool in ArcGIS (v.10.4; Esri Inc., USA) on fire-occurrence point data during a majority of the study period (1992–2015) as a proxy for fire frequency within each hexagon (Jo et al., 2019; Short, 2017). We ln-transformed fire frequency [LOG(FIRE)] to normalize the predictor. Maps of predictor variables are available in the Supporting Information (Figure S2).

## 2.4 | Analysis of individual community changes over time

To assess changes in the distribution of communities over time, we computed a proportion-weighted community centroid and community area for T1 and T2. We used the proportion of the community in each hexagon to weight the  $x$ - $y$  coordinates of the hexagon centroid and took the difference in community centroid between the two time periods to assess the distance and direction of movement of the centroid (following Fei et al., 2017). This approach allowed shifts in abundance within the community to influence the centroid of the community, even if the spatial extent did not change. We assessed the significance of the change in latitude, longitude and overall distance of the shift using randomization tests by randomly disassociating the  $x$ - $y$  coordinates with the community proportion and recalculating the centroid shifts. We then repeated the randomization process 1,000 times and calculated a  $p$ -value for each community (i.e., the proportion of random shifts larger than the observed shift). All  $p$ -values were corrected for multiple testing using the Benjamini-Hochberg method ( $P_{BH}$ ; Benjamini & Hochberg, 1995), and shifts were considered significant when  $P_{BH} \leq .05$ . We also assessed changes in community spatial coverage by weighting the hexagon area by the community proportion at T1 and T2. We used randomization tests to assess the significance of changes in the community area by randomly reassigning community proportions to T1 or T2 and recalculating the difference in the proportion-weighted area between the randomly assigned T1 and T2 groups. This process was repeated 1,000 times to test how extreme the observed expansions or contractions were relative to random increases or decreases in area.

For each community, we tested whether climate change predicted shifts in its distribution. We first calculated the relative frequency of the community under each of the nine historical climate conditions (eight PRISM-derived variables and PDSI). Then, a kernel density tool (density.default in the "stats" package in R; R Core Team, 2019) was used to create a smoothed probability of occurrence distribution across each of the historical climate conditions. Locations with study period conditions within the range of observed historical climate conditions were assigned the corresponding probability of occurrence for each of the eight PRISM climate variables and PDSI. The probability of occurrence was averaged across the nine variables to generate a probability of occurrence in each hexagon for each community based solely on climate. Next, we calculated the climate-predicted centroids for the communities by weighting the  $x$ - $y$  coordinates of each hexagon by the probability of occurrence and compared these centroids with the observed T2 centroids.

## 2.5 | Analysis of change across communities

The communities in each hexagon are compositional data; that is, the proportion of all communities in a hexagon sums to one, which leads to a lack of independence between the communities (i.e., if the proportion of  $k - 1$  out of the  $k$  communities is known, the proportion

of the  $k$ th community can be derived as one minus the sum of the  $k - 1$  community proportions). To reduce the issue of violating independence and identify which forest regions are changing the most, we calculated the Jensen-Shannon divergence (JSD; Hall, Jurafsky, & Manning, 2008; Lin, 1991) between the  $k$ -dimensional community composition at T1 and T2 using the "phlentropy" package in R (Drost, 2018). The JSD is a measure of dissimilarity between compositional data or probability distributions based on the Kullback-Liebler divergence (Lin, 1991). It is calculated by taking the Shannon entropy of the average of the T1 and T2 community proportions and subtracting the average of the Shannon entropy of T1 and T2 community proportions:  $JSD = H\left(\frac{X_{T1} + X_{T2}}{2}\right) - \frac{H(X_{T1}) + H(X_{T2})}{2}$ , where  $H$  is the Shannon entropy,  $H = -\sum_i p_i \log_2 p_i$ , and  $X_{T1}$  and  $X_{T2}$  are vectors of community proportions in a hexagon at T1 and T2, respectively (Lin, 1991). We then square-root transformed JSD to calculate the Jensen-Shannon distance ( $JSD_{T1,T2}$ ), which is a distance metric that is symmetric and approximately normal, but bounded by [0, 1] (Lin, 1991).

To test the importance of drivers of community composition, we modelled  $JSD_{T1,T2}$  as a function of initial climate conditions, climate-change variables and non-climatic variables. Theoretically,  $JSD_{T1,T2}$  is approximately normal but bounded by [0, 1]; however, we found that our measures of  $JSD_{T1,T2}$  were slightly right skewed; therefore, we fitted generalized linear mixed-effects models (GLMMs) with a beta distribution and logit link (which fitted our  $JSD_{T1,T2}$  values better) using the "glmmTMB" package in R (Brooks et al., 2017). Owing to collinearity between predictor variables (Supporting Information Figure S2), we fitted each model with two predictor variables: the historical climate conditions and the change between historical and study period conditions. Fire frequency and nitrogen deposition contained data during the study period only; therefore, we fitted these GLMMs as univariate models. ANOVA revealed significant differences in  $JSD_{T1,T2}$  across the 12 communities ( $F_{11,1801} = 15.93$ ,  $p < .001$ ); therefore, we included random slopes and intercepts for each dominant community at T1 in our GLMMs. However, models that included random slopes failed to converge (owing to some communities being dominant in only a few hexagons); therefore, we used a random intercept only. Given that the predictor variables were standardized, coefficient values are a measure of effect size: a predictor with a larger coefficient value indicates that it influences the response more strongly. A total of 1,757 hexagons were used in the GLMMs, owing to hexagons with missing climatic and non-climate data. All analyses were conducted in R v.3.6.1 (R Core Team, 2019), and all maps were projected to the Albers equal-area conic projection. A full list of R packages used in these analyses is available in the Supporting Information (Table S3).

## 3 | RESULTS

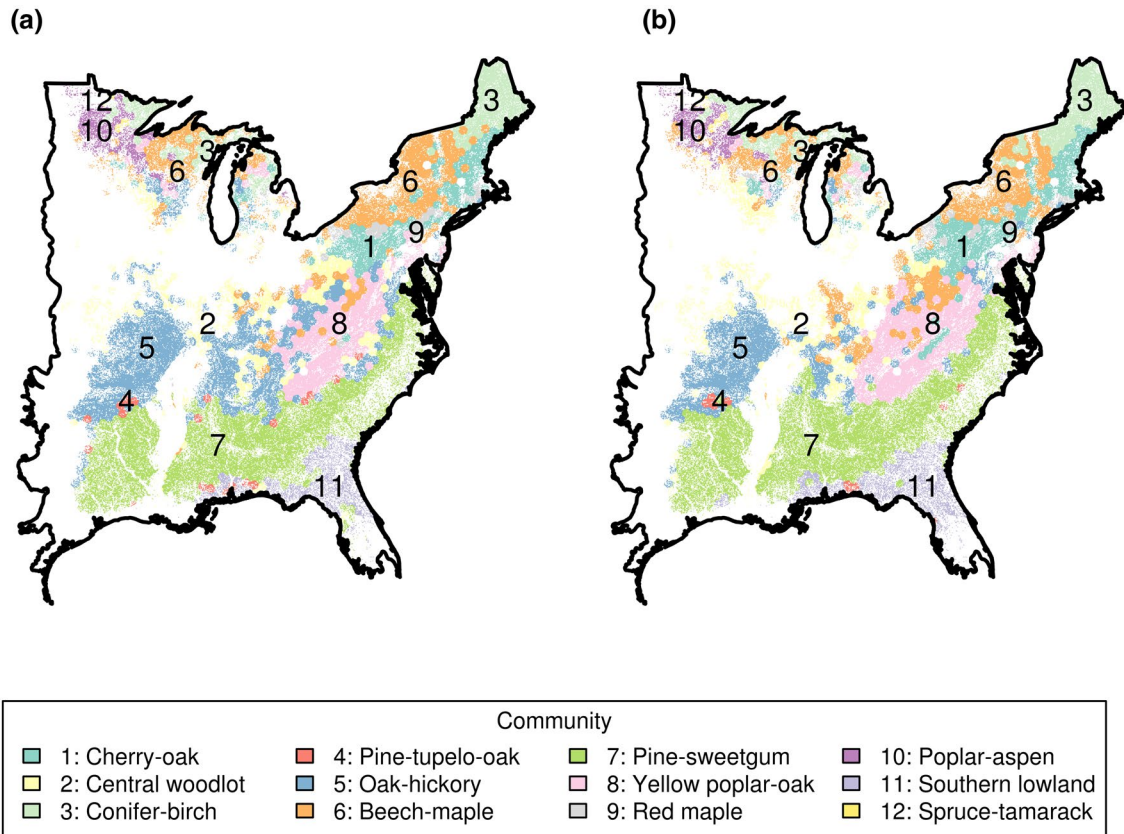
### 3.1 | Assessment of communities identified by LDA

Using the LDA model on FIA data, we identified 12 dominant forest communities in the eastern USA (Table 1; Figure 1). Model selection

TABLE 1 Top species in each community at the two sampling periods

Community	Time	Species
1. Cherry-oak	T1	<i>Quercus rubra</i> (.360), <i>Prunus</i> spp. (.286), <i>Pinus strobus</i> (.177), <i>Tsuga canadensis</i> (.126), <i>Betula lenta</i> (.051)
	T2	<i>Prunus</i> spp. (.318), <i>Quercus rubra</i> (.238), <i>Pinus strobus</i> (.216), <i>Tsuga canadensis</i> (.148), <i>Betula lenta</i> (.081)
2. Central woodlot	T1	<i>Ulmus</i> spp. (.346), <i>Pinus virginiana</i> (.117)*, <i>Sassafras albidum</i> (.095), <i>Quercus laevis</i> (.069)*, <i>Celtis</i> spp. (.061), <i>Robinia pseudoacacia</i> (.054), <i>Juglans nigra</i> (.046), <i>Acer negundo</i> (.043), <i>Platanus occidentalis</i> (.034), <i>Cercis canadensis</i> (.032), <i>Acer saccharinum</i> (.029), <i>Quercus muehlenbergii</i> (.021), <i>Gleditsia triacanthos</i> (.014), <i>Populus deltoides</i> (.013), <i>Maclura pomifera</i> (.012)
	T2	<i>Ulmus</i> spp. (.382), <i>Celtis</i> spp. (.108), <i>Acer negundo</i> (.088), <i>Juglans nigra</i> (.072), <i>Sassafras albidum</i> (.067), <i>Robinia pseudoacacia</i> (.052), <i>Platanus occidentalis</i> (.047), <i>Acer saccharinum</i> (.040), <i>Cercis canadensis</i> (.036), <i>Quercus muehlenbergii</i> (.025), <i>Gleditsia triacanthos</i> (.023), <i>Maclura pomifera</i> (.023), <i>Populus deltoides</i> (.020), <i>Quercus imbricaria</i> (.012)*
3. Conifer-birch	T1	<i>Abies balsamea</i> (.343), <i>Betula papyrifera</i> (.260), <i>Thuja occidentalis</i> (.183), <i>Populus grandidentata</i> (.096), <i>Picea glauca</i> (.052), <i>Picea rubens</i> (.046), <i>Betula alleghaniensis</i> (.021)
	T2	<i>Abies balsamea</i> (.371), <i>Thuja occidentalis</i> (.131), <i>Betula alleghaniensis</i> (.113), <i>Betula papyrifera</i> (.106), <i>Picea rubens</i> (.098), <i>Populus grandidentata</i> (.075), <i>Picea glauca</i> (.059), <i>Acer pensylvanicum</i> (.048)*
4. Pine-tupelo-oak	T1	<i>Pinus echinata</i> (.316), <i>Nyssa sylvatica</i> (.235), <i>Pinus palustris</i> (.195), <i>Quercus falcata</i> (.165), <i>Diospyros virginiana</i> (.066), <i>Quercus lyrata</i> (.022)
	T2	<i>Nyssa sylvatica</i> (.318), <i>Pinus echinata</i> (.199), <i>Pinus palustris</i> (.179), <i>Quercus falcata</i> (.163), <i>Diospyros virginiana</i> (.074), <i>Quercus lyrata</i> (.031), <i>Quercus laevis</i> (.030)*
5. Oak-hickory	T1	<i>Carya</i> spp. (.302), <i>Quercus alba</i> (.294), <i>Quercus velutina</i> (.160), <i>Quercus stellata</i> (.134), <i>Juniperus virginiana</i> (.076), <i>Quercus marilandica</i> (.035)*
	T2	<i>Carya</i> spp. (.322), <i>Quercus alba</i> (.287), <i>Juniperus virginiana</i> (.144), <i>Quercus velutina</i> (.123), <i>Quercus stellata</i> (.114)
6. Beech-maple	T1	<i>Acer saccharum</i> (.372), <i>Fraxinus</i> spp. (.302), <i>Fagus grandifolia</i> (.120), <i>Tilia</i> spp. (.078), <i>Ostrya virginiana</i> (.066), <i>Betula alleghaniensis</i> (.045)* <i>Acer pensylvanicum</i> (.014)*
	T2	<i>Acer saccharum</i> (.351), <i>Fraxinus</i> spp. (.310), <i>Fagus grandifolia</i> (.182), <i>Ostrya virginiana</i> (.093), <i>Tilia</i> spp. (.065)
7. Pine-sweetgum	T1	<i>Pinus taeda</i> (.472), <i>Liquidambar styraciflua</i> (.298), <i>Quercus nigra</i> (.108)*, <i>Quercus laurifolia</i> (.054)*, <i>Quercus virginiana</i> (.027)*, <i>Quercus phellos</i> (.027), <i>Quercus pagoda</i> (.015)*
	T2	<i>Pinus taeda</i> (.674), <i>Liquidambar styraciflua</i> (.301), <i>Quercus phellos</i> (.026)
8. Yellow poplar-oak	T1	<i>Liriodendron tulipifera</i> (.279), <i>Quercus prinus</i> (.186), <i>Pinus banksiana</i> (.162), <i>Pinus resinosa</i> (.118), <i>Oxydendrum arboreum</i> (.112), <i>Quercus coccinea</i> (.109), <i>Pinus rigida</i> (.023)
	T2	<i>Liriodendron tulipifera</i> (.330), <i>Quercus prinus</i> (.178), <i>Pinus resinosa</i> (.131), <i>Oxydendrum arboreum</i> (.092), <i>Pinus virginiana</i> (.084)*, <i>Quercus coccinea</i> (.080), <i>Pinus banksiana</i> (.066), <i>Pinus rigida</i> (.024)
9. Red maple	T1	<i>Acer rubrum</i> (.964), <i>Salix</i> spp. (.036)
	T2	<i>Acer rubrum</i> (.959), <i>Salix</i> spp. (.041)
10. Poplar-aspen	T1	<i>Populus tremuloides</i> (.804), <i>Populus balsamifera</i> (.105), <i>Quercus macrocarpa</i> (.091),
	T2	<i>Populus tremuloides</i> (.715), <i>Quercus macrocarpa</i> (.121), <i>Populus balsamifera</i> (.069), <i>Betula papyrifera</i> (.056)*, <i>Fraxinus</i> spp. (.038)*
11. Southern lowland	T1	<i>Pinus elliotii</i> (.539), <i>Magnolia virginiana</i> (.114), <i>Taxodium ascendens</i> (.101), <i>Pinus serotina</i> (.078), <i>Taxodium distichum</i> (.051), <i>Persea borbonia</i> (.049), <i>Nyssa aquatica</i> (.041), <i>Gordonia lasianthus</i> (.023)
	T2	<i>Quercus nigra</i> (.302)*, <i>Pinus elliotii</i> (.279), <i>Quercus laurifolia</i> (.108)*, <i>Magnolia virginiana</i> (.075), <i>Taxodium ascendens</i> (.049), <i>Taxodium distichum</i> (.047), <i>Quercus virginiana</i> (.042)*, <i>Persea borbonia</i> (.030), <i>Nyssa aquatica</i> (.028), <i>Gordonia lasianthus</i> (.020), <i>Pinus serotina</i> (.019)
12. Spruce-tamarack	T1	<i>Picea mariana</i> (.383), <i>Larix laricina</i> (.213), <i>Carpinus caroliniana</i> (.212), <i>Ilex opaca</i> (.102), <i>Quercus michauxii</i> (.025), <i>Betula populifolia</i> (.024), <i>Quercus palustris</i> (.020)
	T2	<i>Picea mariana</i> (.274), <i>Carpinus caroliniana</i> (.220), <i>Larix laricina</i> (.195), <i>Ilex opaca</i> (.137), <i>Quercus pagoda</i> (.072)*, <i>Betula populifolia</i> (.026), <i>Quercus palustris</i> (.023), <i>Quercus shumardii</i> (.021)*, <i>Quercus michauxii</i> (.020)

Note: Species are included only if their proportion is  $> 1/n_{\text{species}}$  ( $> 1/85$ ). The species proportion in each community is given in parentheses. Asterisks indicate species that were present in the community in one sampling period but not the other.



**FIGURE 1** Maps of 12 regional forest communities across the eastern USA. Hexagon sampling units are mapped according to the community with the highest proportion in each hexagon at (a) T1 and (b) T2. Non-forested areas are masked out by the 2011 National Land Cover Database for visualization purposes only (<https://www.mrlc.gov/data>). Community 12 (spruce–tamarack) was the dominant community in only eight hexagons at T1 and 11 hexagons at T2 and therefore does not show clearly on the maps. Maps are projected to the Albers equal-area conic projection. The species composition of the 12 communities can be found in Table 1, and a map showing the difference between T1 and T2 can be found in Figure 5. GIS shapefiles of these maps are available for download from <https://www.doi.org/10.4231/GCE5-ZY59>

via an iterative approach identified a candidate set of models with values of  $k$ , the number of communities in the model, ranging from 12 to 43 (out of possible  $k$  values ranging from two to 50). However, when we applied an ecologically based metric, the number of species with a proportion in the community  $> 1/n_{\text{species}}$  ( $> 1/85$ ), the  $k = 12$  community model was chosen because it did not include any communities composed of only one species. Models with varying inputs (relative density, relative basal area, and importance value for saplings, trees and all stems) generally agreed in species composition and were not significantly different from one another (Mantel test correlation  $r > .99$ ,  $p < .01$  for all tests). Therefore, results shown are based on the model including the importance value of all stems; it was the most generalized community, incorporating density, basal area and all size classes.

Given that the LDA model does not incorporate spatial constraints on the communities, the clustering of communities in a specific region would indicate that the model is generating meaningful communities. Indeed, the observed communities were centred on a specific geographical region (Figure 1; Supporting Information Figure S3). Likewise, we used the species composition of each community to assess whether it aligned with previous knowledge of the forest

communities in these different regions. We found similarities between our communities and those from the studies by Braun (1950), Dyer (2006) and Costanza et al. (2017) (see Discussion), indicating that the LDA model was identifying meaningful ecological communities. The spatial and ecological consistency of the communities indicated that further analysis using communities as a unit was justified.

### 3.2 | Community compositional shift between T1 and T2

Generally, communities were identifiably similar between T1 and T2, and correlation tests revealed a significant correlation between T1 and T2 species composition for all communities ( $r = .67-.99$ ,  $p < .001$  for all tests). Three main community groups were observed in terms of their compositional shifts, although the groupings were not mutually exclusive. The first group of communities (Communities 5, oak-hickory; 6, beech-maple; 8, yellow poplar-oak; 9, red maple; 10, poplar-aspen; and 12, spruce-tamarack) remained stable throughout the study period, with only minor shifts in species composition or gain/loss of minor component species (Table 1). For example,

Community 6 (beech–maple) lost only the bottom two species in the community and had no changes in dominance, and Community 9 (red maple) had very minor changes in the relative proportion of red maple (*Acer rubrum*) versus willow species (*Salix* species). The second group of communities (Communities 2, central woodlot; 7, pine–sweetgum; and 11, southern lowland) gained or lost major species over time. Community 11 (southern lowland) gained oak species (*Quercus nigra*, *Quercus laurifolia* and *Quercus virginiana*) from Community 7 (pine–sweetgum), and Community 2 (central woodlot) lost Virginia pine (*Pinus virginiana*), the species with the second highest proportion in Community 2 at T1. All other communities except Communities 1 (cherry–oak) and 9 (red maple) gained or lost species; however, these gains and losses were generally a minor component of the community. The third community group (Communities 1, cherry–oak; 4, pine–sweetgum; and 11, southern lowland) shifted the dominant species in the community. For example, Community 1 (cherry–oak) switched from being dominated by red oak (*Quercus rubra*) to being dominated by cherry species (*Prunus* species), and Community 4 (pine–tupelo–oak) switched from being dominated by shortleaf pine (*Pinus echinata*) to being dominated by black gum (*Nyssa sylvatica*). The most dramatic shift in species composition occurred in Community 11 (southern lowland), which switched from being dominated by slash pine (*Pinus elliotii*) to a new species in the community, water oak (*Quercus nigra*).

### 3.3 | Shifts in spatial distributions of communities

We found that there was large variability in the distance that communities shifted over the last three decades. Compared with the

species level (i.e., observed shifts from Fei et al., 2017), community-level shifts in spatial distribution were relatively short, ranging from 3.6 to 24.6 km/decade, with a median of 8.0 km/decade (Table 2). All communities had significant shift distances when tested with randomization tests (Figure 2a). The forest community with the largest spatial shift, Community 11 (southern lowland), was centred in the SPHR. This community shifted its centroid 24.6 km/decade to the southwest (Benjamini–Hochberg-adjusted  $p$ -value,  $P_{BH} < .001$ ). Communities 2 (central woodlot), 3 (conifer–birch) and 4 (pine–tupelo–oak) also had relatively large shift distances (24.2, 21.2 and 21.5 km/decade, respectively,  $P_{BH} < .001$  for all three communities; Table 2). The community with the smallest shift, Community 1 (cherry–oak) had a marginally significant south-eastward shift (3.6 km/decade,  $P_{BH} = .045$ ).

The direction of shift also varied across the 12 communities (Table 2; Figure 2b). Half of the communities shifted eastward (five significant: Communities 1, cherry–oak, 2.8 km/decade,  $P_{BH} = .035$ ; 3, conifer–birch, 21.2 km/decade,  $P_{BH} < .001$ ; 4, pine–tupelo–oak, 11.3 km/decade,  $P_{BH} < .001$ ; 5, oak–hickory, 2.5 km/decade,  $P_{BH} = .035$ ; and 8, yellow poplar–oak, 3.7 km/decade,  $P_{BH} = .015$ ) and the other half shifted westward (five significant: Communities 2, central woodlot, 23.2 km/decade,  $P_{BH} < .001$ ; 6, beech–maple, 11.4 km/decade,  $P_{BH} < .001$ ; 9, red maple, 6.7 km/decade,  $P_{BH} < .001$ ; 11, southern lowland, 12.2 km/decade,  $P_{BH} < .001$ ; and 12, spruce–tamarack, 6.4 km/decade,  $P_{BH} < .001$ ). A total of five communities shifted northward (four significant: Communities 2, central woodlot, 6.9 km/decade,  $P_{BH} = .002$ ; 4, pine–tupelo–oak, 18.3 km/decade,  $P_{BH} < .001$ ; 5, oak–hickory, 5.8 km/decade,  $P_{BH} < .001$ ; and 7, pine–sweetgum, 8.6 km/decade,  $P_{BH} < .001$ ), and seven communities shifted southward (four significant: Communities 6, beech–maple,

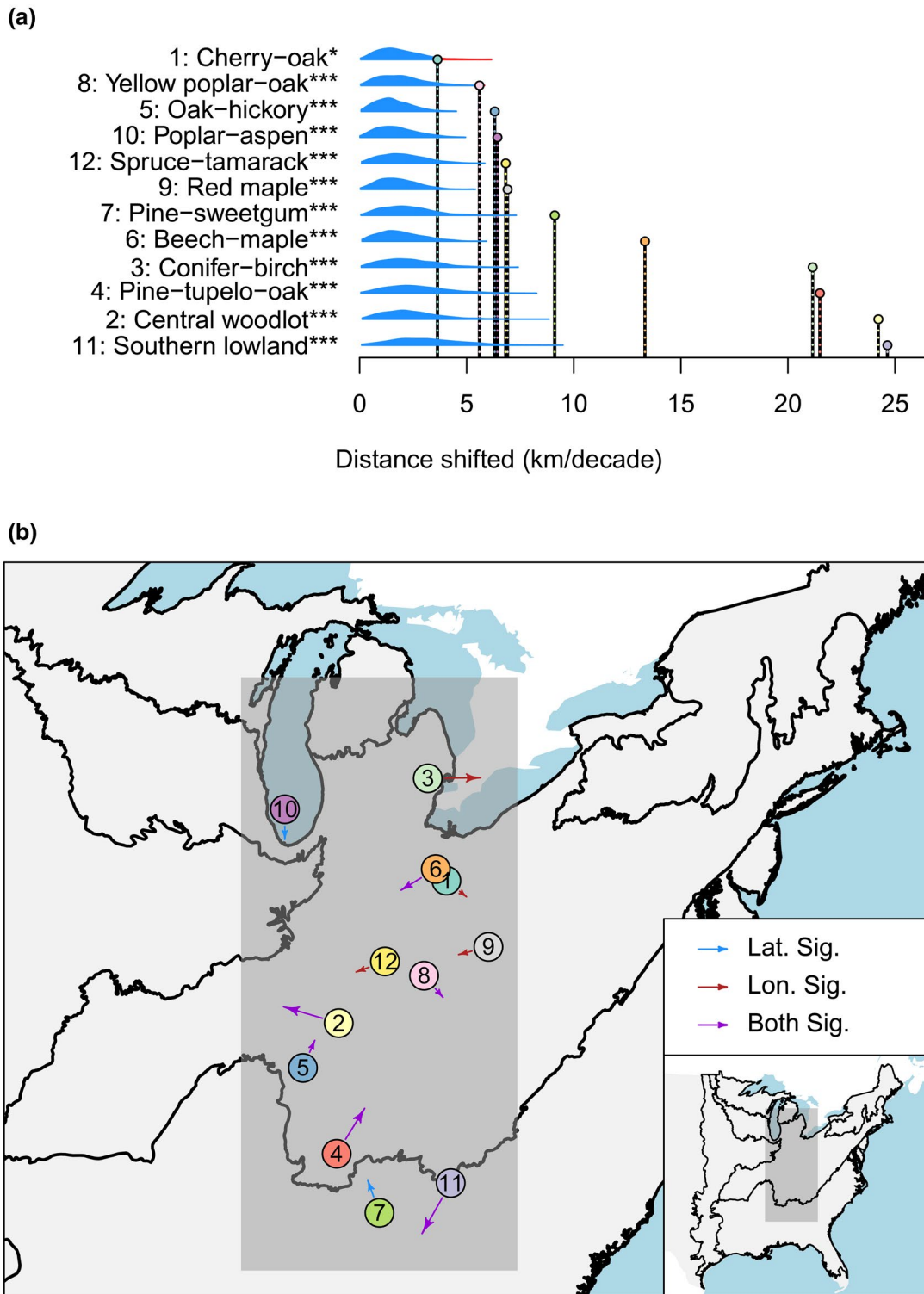
**TABLE 2** Shifts in forest communities over the last three decades

Community	Latitude shift (km/decade)	Longitude shift (km/decade)	Observed direction	Predicted direction	Distance shift (km/decade)	Area change (km <sup>2</sup> /decade)
1. Cherry–oak	–2.3	2.8*	South-east	North-west	3.6*	4,578* (+7.1%)
2. Central woodlot	6.9**	–23.2***	North-west	North-east	24.2***	175 (+.2%)
3. Conifer–birch	.2	21.2***	North-east	North-west	21.2***	1,259 (+2.5%)
4. Pine–tupelo–oak	18.3***	11.3***	<b>North-east</b>	<b>North-east</b>	21.5***	–11506*** (–18.5%)
5. Oak–hickory	5.8***	2.5*	<b>North-east</b>	<b>North-east</b>	6.3***	–17089*** (–15.6%)
6. Beech–maple	–6.9***	–11.4***	South-west	North-east	13.3***	–205 (–.2%)
7. Pine–sweetgum	8.6***	–3.1	North-west	North-east	9.1***	1,830 (+1.8%)
8. Yellow poplar–oak	–4.2**	3.7*	South-east	South-west	5.6***	2,878 (+5.0%)
9. Red maple	–1.7	–6.7***	South-west	North-west	6.9***	3,423* (+5.1%)
10. Poplar–aspen	–6.4***	+0.0	South-east	North-east	6.4***	–2005 (–4.7%)
11. Southern lowland	–21.4***	–12.2***	<b>South-west</b>	<b>South-west</b>	24.6***	13,759*** (+26.7%)
12. Spruce–tamarack	–2.3	–6.4***	South-west	North-west	6.8***	2,903** (+7.7%)

Note: Shifts were measured by the movement of the community centroid and changes in community spatial coverage. Positive values of latitude and longitude shift represent northward and eastward movement, respectively; conversely, negative values represent southward and westward movement. The predicted direction was based on expected shifts owing to climate change (communities shown in bold had the same observed and predicted direction of shift). The percentage change in the area is in parentheses. Asterisks indicate significant Benjamini–Hochberg-adjusted  $p$ -values ( $P_{BH}$ ) when tested by randomization tests.

\* $P_{BH} \leq .05$ ; \*\* $P_{BH} \leq .01$ ; \*\*\* $P_{BH} \leq .001$ .





**FIGURE 2** Centroid shift of forest communities. (a) Frequency distribution of randomized centroid shifts (blue area) and observed shift distance (dashed lines). The proportion of the frequency distribution greater than the observed shift distance is in red. Asterisks indicate significant Benjamini–Hochberg adjusted  $p$ -values ( $P_{BH}$ ): \* $P_{BH} \leq .05$  and \*\*\* $P_{BH} \leq .001$ . (b) Direction and distance of forest community shifts. Arrow colours represent the significance of shifts in distribution. The map is projected to the Albers equal-area conic projection, and the inset map shows the location displayed relative to the study area (white areas included for display purposes only)

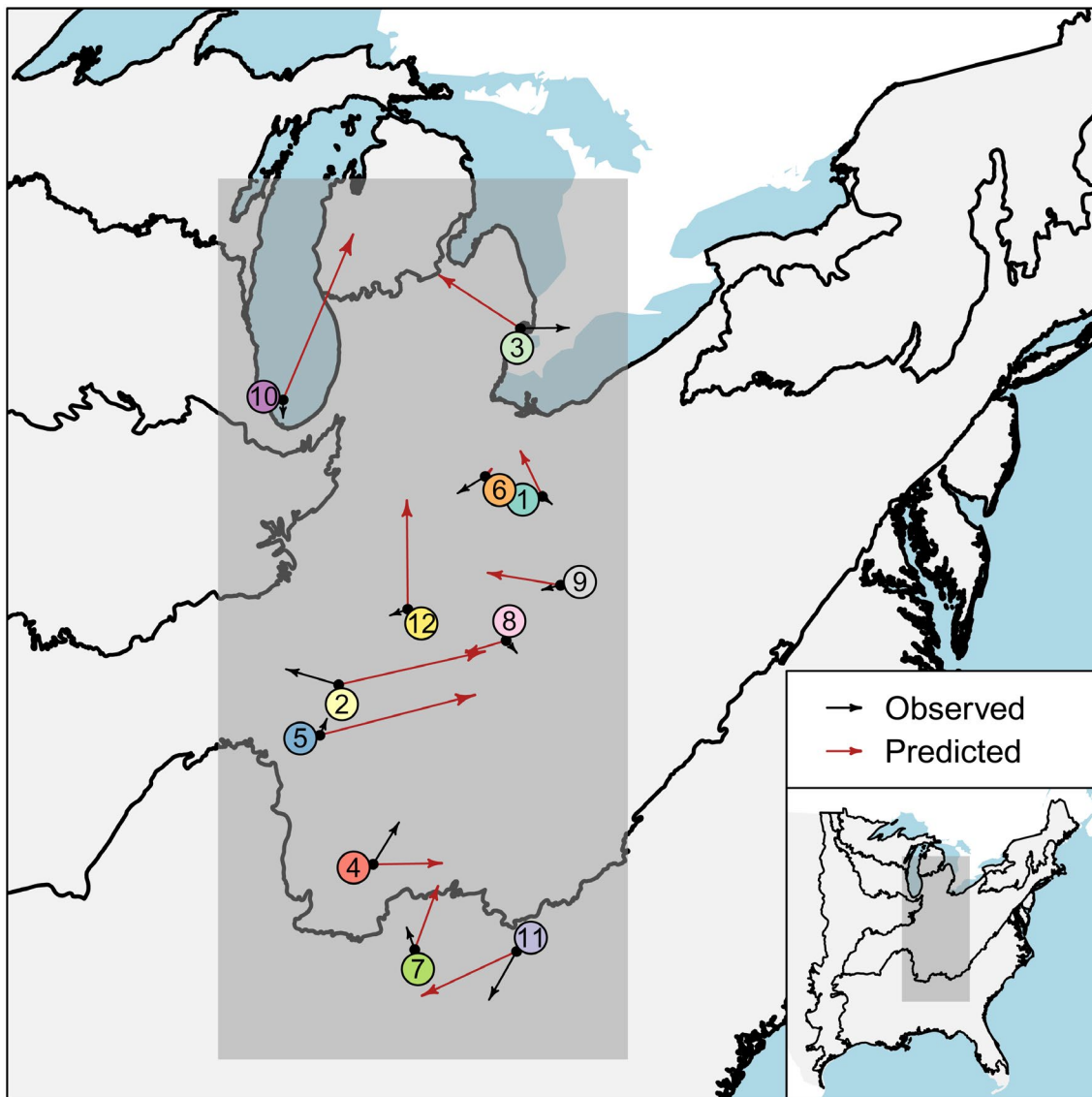
6.9 km/decade,  $P_{BH} < .001$ ; 8, yellow poplar-oak, 4.2 km/decade,  $P_{BH} = .009$ ; 10, poplar-aspen, 6.4 km/decade,  $P_{BH} < .001$ ; and 11, southern lowland, 21.4 km/decade,  $P_{BH} < .001$ ).

Additionally, eight communities expanded their area of coverage, and four communities contracted their area of coverage (Table 2). Of the eight expansions, four communities gained spatial

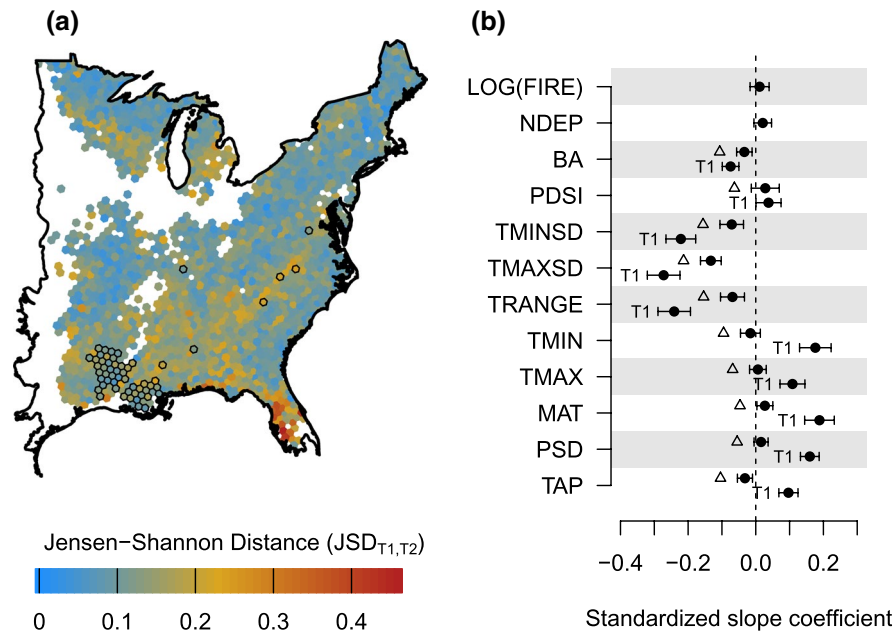
coverage significantly (Communities 1, cherry-oak, 4,578 km<sup>2</sup>/decade,  $P_{BH} = .03$ ; 9, red maple, 3,423 km<sup>2</sup>/decade,  $P_{BH} = .026$ ; 11, southern lowland, 13,759 km<sup>2</sup>/decade,  $P_{BH} < .001$ ; and 12, spruce-tamarack, 2,903 km<sup>2</sup>/decade,  $P_{BH} = .003$ ). Of the four contractions, two communities lost spatial coverage significantly (Communities 4, pine-tupelo-oak, -11,506 km<sup>2</sup>/decade,  $P_{BH} < .001$ ; and 5, oak-hickory, -17,089 km<sup>2</sup>/decade,  $P_{BH} < .001$ ; Table 2). The largest increase in community area was in Community 11 (southern lowland), which increased by 13,759 km<sup>2</sup>/decade, a 26.7% expansion in area ( $P_{BH} < .001$ ). The largest decrease in area was in Community 5 (oak-hickory), which decreased by 17,089 km<sup>2</sup>/decade, a 15.6% contraction in area ( $P_{BH} < .001$ ). Community 2 (central woodlot) had the smallest change in area, increasing by 175 km<sup>2</sup>/decade (0.2%

expansion), and Community 6 (beech-maple) had a similar decrease in area (-205 km<sup>2</sup>/decade, 0.2% contraction).

Observed T2 locations of communities were predicted well by climate conditions during the study period (Supporting Information Figure S4a,b), but shifts in the communities were smaller than predicted by climate change except for Community 6 (beech-maple; Figure 3). Climate-predicted longitude was moderately associated with observed longitude of the 12 communities ( $R^2 = 0.278$ ), and the slope of the association was not different from the 1:1 line (slope = 1.14,  $p = .826$ ; Supporting Information Figure S4a). However, the relationship between observed and predicted latitude was very strong ( $R^2 = 0.96$ ) and significantly smaller than the 1:1 line (slope = 0.84,  $p < .001$ ; Figure S4b), indicating that the predicted



**FIGURE 3** Predicted versus observed shifts in forest communities. Black dots represent community centroids at T1, and arrows represent climate-predicted (red) and observed (black) shifts in the community centroid. All predicted shifts are larger than the observed shifts except for Community 6 (beech-maple). The inset map indicates the location of community centroids within the study area. White areas outside of the study area are included for display purposes only. The map is projected to the Albers equal-area conic projection. See Figure 2 for a significance of observed community shifts



**FIGURE 4** Changes in forest communities and associated factors. (a) Response variable from generalized linear mixed-effects models (GLMMs): change in community composition measured by Jensen–Shannon distance in  $k = 12$ -dimensional community space between T1 and T2 ( $JSD_{T1,T2}$ ). Areas in red have greater dissimilarity in community composition between T1 and T2. Hexagons with black borders were removed from the analysis because of missing data. The map is projected to the Albers equal-area conic projection. (b) Predictors of change in community composition over the last three decades. Dots represent standardized slope coefficient estimates from GLMMs with a beta distribution and logit link function. Shading indicates individual models. All models except fire frequency and nitrogen deposition [LOG(FIRE) and NDEP, respectively] contained the historical climate conditions (indicated by “T1”) and change between the historical and study period conditions (indicated by “Δ”). Bars represent 95% confidence intervals, and those that cross the vertical line at zero are considered non-significant. Abbreviations: BA = basal area; LOG(FIRE) = fire frequency; MAT = mean annual temperature; NDEP = nitrogen deposition; PDSI = Palmer drought severity index; PSD = SD of monthly precipitation; TAP = total annual precipitation; TMAX = annual maximum temperature; TMAXSD = SD of maximum monthly temperature; TMIN = annual minimum temperature; TMINS D = SD of minimum monthly temperature; TRANGE = annual temperature range

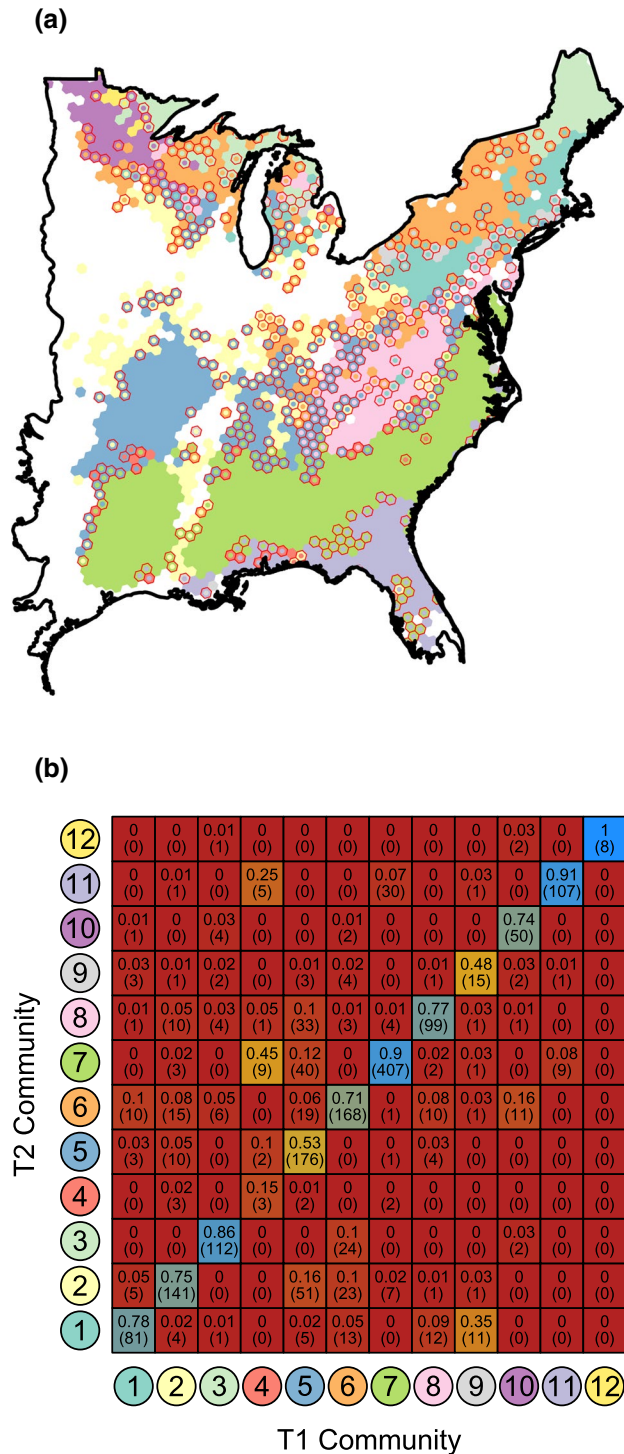
locations are generally farther from the centre of the study region than observed. In contrast, there were very low  $R^2$  values between predicted and observed shifts in the distance in latitude and longitude, with slopes close to zero and far from the 1:1 line ( $R^2 = 0.02$  and  $0.05$ , and slope =  $0.047$  and  $-0.06$  for latitude and longitude, respectively, Figure 3; Supporting Information Figure S4c,d). The map depicting the observed and predicted shifts (Figure 3) showed many communities shifting in directions different from those predicted by climate change, and only three communities had the same observed and predicted direction of shift (Table 2).

### 3.4 | Across-community responses to climatic and non-climatic factors

We assessed changes in all communities collectively across the study region to identify where forest communities are changing most rapidly and compare those areas with a variety of climatic and non-climatic factors. Areas in the southern portion of the study region tended to have the largest changes in community composition over time (i.e., the largest  $JSD_{T1,T2}$ ; Figure 4a). Our GLMMs with climatic and non-climatic predictors of  $JSD_{T1,T2}$  showed that eight PRISM-derived historical climatic conditions were significant

predictors of community change over time (TAP, PSD, MAT, TMAX, TMIN, TRANGE, TMAXSD and TMINS D;  $P_{BH} < .001$  for all eight), but PDSI exhibited only marginal significance ( $P_{BH} = .047$ ; Figure 4b). Of the nine climate variables, only changes in temperature variability (TRANGE, TMINS D and TMAXSD;  $P_{BH} < .001$  for all three) and MAT ( $P_{BH} = .025$ ) were significant (Figure 4b). Of the non-climatic variables we tested, only basal area (BA; a proxy for forest developmental stage) was significant ( $P_{BH} < .001$  for initial BA and  $P_{BH} = .004$  for change in BA). Areas that had forests that were initially older or that had increases in basal area (forests that got older progressively throughout the study period) tended to be more stable over time.

About one-quarter of hexagons showed turnover in the dominant community type (24.6%, 446 out of 1813 hexagons; Figure 5), although correlation tests revealed that the overall community composition was consistent between T1 and T2 across all 12 communities ( $r = .80$ – $.97$ ,  $p < .001$  for all tests). The map of change in the dominant community between T1 and T2 (Figure 5a) showed that areas changing from one community to another tended to occur along the boundaries of the T1 communities (e.g., along the northern and southern edges of Community 7, pine–sweetgum, in the SPHR, or the boundary between Communities 8, yellow poplar–oak, and 5, oak–hickory, in the CHR). Hexagons dominated by Communities 3



**FIGURE 5** Change in dominant forest communities between T1 and T2. (a) Hexagons with red borders indicate where the dominant community at T2 differs from the dominant community at T1. Hexagon outer colour represents the dominant community at T1 and the inner point colour represents the dominant community at T2. The map is projected to the Albers equal-area conic projection. (b) The proportion matrix between T1 and T2. Each box contains the proportion of hexagons with the dominant community at T1 (x axis) transitioning to the dominant community at T2 (y axis). The number of hexagons in each transition is in parentheses. Colours represent higher proportions (red = low, blue = high). A total of 446 out of 1813 hexagons (24.6%) changed dominance between T1 and T2

(conifer–birch), 7 (pine–sweetgum), 11 (southern lowland) and 12 (spruce–tamarack) at T1 were most frequently classified under the same dominant community at T2 (86.2, 90.0, 91.5 and 100.0%, respectively; Figure 5b).

## 4 | DISCUSSION

We showed that the LDA model works for identifying meaningful communities of co-occurring species at a regional scale. In general, the 12 communities aligned with previous studies of forest communities. The boundaries of our regional forest community maps (Figure 1) were similar to those found by Dyer (2006) and Braun (1950). Likewise, the species composition and distribution of some of our communities aligned closely with those identified by Costanza et al. (2017). For example, the balsam fir–quaking aspen community identified by Costanza et al. (2017) is similar in species composition and spatial distribution to Community 3 (conifer–birch). A benefit of the LDA model is that it provides a relative frequency of each forest community within each sample versus other methods that assign a single dominant community type to each sample (Valle et al., 2014). This allows for the analysis of community dynamics across the entire study area, including areas where a community is a minor component (e.g., community centroids were weighted by the community proportion in all hexagons, not only in the hexagons where it was dominant). However, our maps in Figure 1 and analysis of change in the dominant community (Figure 5) indicate only the community with the highest proportion in each aggregated hexagon unit; that is, many hexagons contained multiple community types (median number of communities with proportion > 1/12 in each hexagon = 4; minimum = 1; maximum = 8).

Given the wide range of responses to climate change at the species level (e.g., Fei et al., 2017; Woodall et al., 2009; Zhu et al., 2012), we expected to find communities responding to climate change with different spatial dynamics. For example, we found a wide range of spatial shifts, such as the southern lowland community (Community 11), which had a large southwest shift and large increase in area, versus Communities 1 (cherry–oak) and 8 (yellow poplar–oak), which had highly stable spatial distributions (both showing marginally significant south-eastern shifts and small increases in the area). The dramatic shifts in Community 11 are likely to be attributable to increases in the slash pine (*Pinus elliottii*) component of this community in this region (Fox, Jokela, & Allen, 2007; Knott, Desprez, Oswald, & Fei, 2019) and to the addition of *Quercus* species over the study period. In contrast, Communities 1 (cherry–oak) and 8 (yellow poplar–oak) have contrasting mechanisms for stability: the cherry–oak community is most commonly found as a mid- to late successional component of forests in the CHR and NHR versus the yellow poplar–oak community, which is dominated by yellow poplar (*Liriodendron tulipifera*), a disturbance-dependent species that can grow in a wide range of conditions and create a stable community, where it dominates multi-tree canopy openings (Burns & Honkala, 1990).

In addition to variability in the spatial dynamics of the communities, we also expected to find variability in the number of changes to the species composition of the communities. Again, Community 11 (southern lowland) was most dynamic, gaining oak species (*Quercus nigra*, *Quercus laurifolia* and *Quercus virginiana*) at T2 from Community 7 (pine-sweetgum). In contrast, Communities 6 (beech-maple) and 9 (red maple) were highly stable in species composition. Although they both had significant spatial shifts (indeed, Community 6 was the only community that shifted farther than climate predicted), the stability of their species composition is not surprising given that beech-maple forests are considered to be a climax community (Braun, 1950), and red maple (*Acer rubrum*) has increased consistently over recent decades (Fei & Steiner, 2007).

The strongest predictors of community change over time were related to historical climatic conditions and changes in seasonal temperature variability. In general, forests that had the greatest changes in community composition tended to have a wetter and warmer historical climate, higher precipitation variability or lower seasonal temperature variability (Figure 4b). In addition, forests that decreased in seasonal temperature variability over the study period tended to have larger changes over time. Measures of change in precipitation were not significant in relationship to community spatial shifts despite species-level migration in response to change in precipitation (Fei et al., 2017), and measures of temperature change (MAT) were less significant than measures of change in temperature variability. The nature of the relationship between seasonal temperature variability and  $JSD_{T1,T2}$  (areas with larger climate variability and areas that increased in climate variability led to more stable communities) generally supports the hypothesis that fluctuations in temperature can act as stabilizing processes in vegetation dynamics (Lloret, Escudero, Iriondo, Martínez-Vilalta, & Valladares, 2012).

Although the non-climatic variables we tested play an important role in local forest dynamics, they generally lack strong relationships with community change at the regional scale (the largest of the three scales studied in this research), which aligns with the view of Ricklefs (1987) that regional-scale processes are more important than local-scale processes. Fire frequency is often considered to be an important disturbance metric in forest ecosystems, especially at the stand level (Briggs, Knapp, & Brock, 2002; Hutchinson, Sutherland, & Yaussy, 2005; Nowacki & Abrams, 2008), but was found to have little effect on the overall changes to forest communities at the regional scale (Figure 4b). In addition, nitrogen deposition was found to have a non-significant effect on forest community change despite the regulatory influence of the nitrogen and carbon cycles (and the mycorrhizal communities that influence these cycles) in forest ecosystems (Jo et al., 2019; Lovett, Weathers, & Arthur, 2002; Pellegrini et al., 2017). Areas with an initially higher total basal area or that gained basal area, generally associated with older forests, were found to be more stable over time, consistent with comparisons of old-growth forests versus younger forests (Fralish, Crooks, Chambers, & Harty, 1991); however, the relationship was weaker than the effect of climate.

Previous studies at the local scale have shown dramatic within-site turnover in forest communities as a response to climate change (Feeley, Davies, Perez, Hubbell, & Foster, 2011; Lebrija-Trejos, Pérez-García, Meave, Bongers, & Poorter, 2010; Ozier, Groninger, & Ruffner, 2006). At the regional scale, c. 25% of the study region experienced turnover in the dominant community (Figure 5), with some communities losing dominance at higher rates than other communities. This trend was also indicated by the significant community shift distances but lack of directionality; as one community moves out of an area, another fills in behind, leading to a lack of consistent directional shifts.

The rate of change in communities is likely to be lagging behind climate change. Although we showed many dramatic changes in species composition, spatial distribution and turnover in the dominant community type, climate change outpaced the rate of community migration (Figure 3). However, this is not surprising because there are expected to be significant time lags between climate change and community-level responses (Bertrand et al., 2011). It is possible that communities have not yet accumulated enough species-level responses (e.g., increased mortality in unsuitable areas or differential recruitment rates across the region) to produce a community-wide response to climate change, especially throughout the relatively short study period. Our study encompassed three decades of forest inventory data, but a more direct response to climate change might emerge when considering a longer time interval.

A second reason for climate lags is the resilience of complex systems. Generally, with increased complexity, there is a greater resilience to stressors (Loreau et al., 2001; Symstad et al., 1998). With these complex communities composed of multiple species, it is likely that species with large responses are balanced out by others with small responses, and the diversity and plasticity of traits within a community can allow it to compensate for climate change. Additionally, climate effects may be masked out by other unaccounted factors, such as trait variability (e.g., variation in the dispersal ability of the component species), invasive species and disturbances (such as deer browse and change in land use) that affect communities differentially (Côté, Rooney, Tremblay, Dussault, & Waller, 2004; Jo, Potter, Domke, & Fei, 2018; Knott et al., 2019; Oswalt et al., 2015).

There are a few caveats of our study that need to be considered, especially when using the results of our study for management and extrapolation to other regions. First, our results are aggregated to a larger scale (i.e., the 1,452 km<sup>2</sup> hexagon tessellation) than the scale at which most management occurs. This is important for managers who intend to use our results as an indicator of changes beyond their local plot-level observations. Managers should consider the dynamics of the various regional communities represented in their forests when making decisions, but they also should recognize that the trends in regional forest communities might not always reflect the variability in local-scale forest dynamics. Second, it is also important to consider the location of our study. Forest communities near the coast (hard boundaries) might

tend to shift inland, and forest communities along the soft boundaries of our study area (i.e., Canada and the central USA) might have shifted into these regions but are not observable in the FIA data used in this study. Likewise, our analysis of climate-predicted shifts was limited to climate data within the study region boundaries, and it is likely that some communities (especially those near the Canadian border) will find a suitable climate outside the study area. However, many of the northern communities did not move as far north as predicted by climate change even when limited to the study area; therefore, extending our modelling approach to include potentially suitable climate in Canada might further increase the divergence between observed and predicted shifts. Finally, although we found many significant shifts in eastern U.S. communities, communities in other locations (e.g., in other parts of North America or on other continents) might be more or less stable than those found in our study area; however, to our knowledge, there have not been studies similar to ours in other areas of North America or globally.

The consistently significant shifts in forest communities can serve as a warning sign of the continued impact of anthropogenic activities. Although we detected changes in forest communities within three decades that surpass rates of change observed during times of historical climate change (e.g., century to millennium time-scales of post-glaciation migration; Davis, 1983), climate change is currently outpacing the rate of community migration. Forest communities not only migrated shorter distances than climate change predicted, but also shifted in the direction opposite to climate change. This is alarming in that forest communities are unable to keep up with either the pace or the direction of climate change. Future analyses incorporating individual species traits within a community can help to elucidate the susceptibility of certain ecosystem functions to climate change. In addition, understanding other potential threats to forest communities, such as invasive species, specific management practices and other climate-related factors, can help to quantify further the sustainability of forest ecosystems and the services they provide. Nevertheless, our analysis presents one of the first attempts at quantification of the redistribution of regional forest communities, and our results can aid in the monitoring and management of forest ecosystems in a rapidly changing global environment.

#### ACKNOWLEDGMENTS

We would like to thank our funding support of this project through the United States Department of Agriculture National Institute of Food and Agriculture (USDA NIFA 11207356), USDA Forest Service (19-JV-11242305-102) and the Purdue Climate Change Research Center to S.F. We would also like to thank R. Swihart, J. Dukes, N. Morueta-Holme (associate editor for *Global Ecology and Biogeography*) and two anonymous reviewers for helpful comments on the manuscript.

#### DATA ACCESSIBILITY

The original FIA data used in this project are available through the USDA Forest Service (<https://www.fia.fs.fed.us/>). The climate data

are available through the PRISM Climate Group, Oregon State University (for temperature and precipitation; <http://prism.oregonstate.edu/>) and the West Wide Drought Tracker (for PDSI; <https://wrcc.dri.edu/wwdt/>). Nitrogen deposition data are available from the National Atmospheric Deposition Program <http://nadp.slh.wisc.edu/>, and fire frequency data are available from Short (2017). More details about the publicly available data are in the Methods Sections 2.1 and 2.3. GIS shapefiles of the hexagon-level output for mapping individual communities or creating maps of the top communities and a full list of community species compositions are available through the Purdue University Research Repository (PURR; <https://www.doi.org/10.4231/GCE5-ZY59>).

#### ORCID

Jonathan A. Knott  <https://orcid.org/0000-0003-3856-8454>

Michael A. Jenkins  <https://orcid.org/0000-0002-1612-6383>

Christopher M. Oswalt  <https://orcid.org/0000-0001-8183-7651>

Songlin Fei  <https://orcid.org/0000-0003-2772-0166>

#### REFERENCES

- Arun, R., Suresh, V., Veni Madhavan, C. E., & Narasimha Murthy, M. N. (2010). On finding the natural number of topics with latent Dirichlet allocation: Some observations. In: M. J. Zaki, J. X. Yu, B. Ravindran, & V. Pudi (eds) *Advances in Knowledge Discovery and Data Mining*. PAKDD 2010. Lecture Notes in Computer Science, vol 6118. Berlin, Heidelberg.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57, 289–300.
- Bertrand, R., Lenoir, J., Piedallu, C., Riofrío-Dillon, G., de Ruffray, P., Vidal, C., ... Gégout, J.-C. (2011). Changes in plant community composition lag behind climate warming in lowland forests. *Nature*, 479(7374), 517–520.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Braun, E. L. (1950). *Deciduous forests of eastern North America*. Caldwell, NJ: The Blackburn Press.
- Briggs, J. M., Knapp, A. K., & Brock, B. L. (2002). Expansion of woody plants in tallgrass prairie: A fifteen-year study of fire and fire-grazing interactions. *The American Midland Naturalist*, 147, 287–294.
- Brooks, M. E., Kristensen, K., Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., ... Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, 9, 378–400.
- Burns, R. M., & Honkala, B. H. (1990). *Silvics of North America: Volume 2, hardwoods*. United States Department of Agriculture (USDA), Forest Service, *Agriculture Handbook*, 654. <https://www.srs.fs.usda.gov/pubs/1548>
- Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing*, 72, 1775–1781.
- Cleland, D., Freeouf, J., Keys, J., Nowacki, G., Carpenter, C., & McNab, W. (2007). Ecological subregions: Sections and subsections for the conterminous United States. *General Technical Report WO-76D*, 76. Washington, DC: U.S. Department of Agriculture, Forest Service.
- Clements, F. E. (1916). *Plant succession: An analysis of the development of vegetation*. Washington, DC: Carnegie Institute of Washington Publication.
- Costanza, J. K., Coulston, J. W., & Wear, D. N. (2017). An empirical, hierarchical typology of tree species assemblages for assessing forest dynamics under global change scenarios. *PLoS ONE*, 12, e0184062.

- Côté, S. D., Rooney, T. P., Tremblay, J.-P., Dussault, C., & Waller, D. M. (2004). Ecological impacts of deer overabundance. *Annual Review of Ecology, Evolution, and Systematics*, 35, 113–147.
- Damgaard, C. (2015). Modelling pin-point cover data of complementary vegetation classes. *Ecological Informatics*, 30, 179–184.
- Davis, M. B. (1983). Quaternary history of deciduous forests of eastern North America and Europe. *Annals of the Missouri Botanical Garden*, 550–563.
- Deveaud, R., SanJuan, E., & Bellot, P. (2014). Accurate and effective latent concept modeling for ad hoc information retrieval. *Document Numerique*, 17, 61–84.
- Dray, S., & Dufour, A. (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22(4), 1–20.
- Drost, H. (2018). philentropy: Information theory and distance quantification with R. *Journal of Open Source Software*, 3, 765.
- Dyer, J. M. (2006). Revisiting the deciduous forests of eastern North America. *BioScience*, 56, 341–352.
- Feeley, K. J., Davies, S. J., Perez, R., Hubbell, S. P., & Foster, R. B. (2011). Directional changes in the species composition of a tropical forest. *Ecology*, 92, 871–882.
- Fei, S., Desprez, J. M., Potter, K. M., Jo, I., Knott, J. A., & Oswalt, C. M. (2017). Divergence of species responses to climate change. *Science Advances*, 3, e1603055.
- Fei, S., Morin, R. S., Oswalt, C. M., & Liebhold, A. M. (2019). Biomass losses resulting from insect and disease invasions in US forests. *Proceedings of the National Academy of Sciences USA*, 116, 17371–17376.
- Fei, S., & Steiner, K. C. (2007). Evidence for increasing red maple abundance in the eastern United States. *Forest Science*, 53, 473–477.
- Fox, T. R., Jokela, E. J., & Allen, H. L. (2007). The development of pine plantation silviculture in the southern United States. *Journal of Forestry*, 105, 337–347.
- Fralish, J. S., Crooks, F. B., Chambers, J. L., & Harty, F. M. (1991). Comparison of presettlement, second-growth and old-growth forest on six site types in the Illinois Shawnee Hills. *The American Midland Naturalist*, 125, 294–309.
- Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, 19, 1–67.
- Gleason, H. A. (1926). The individualistic concept of the plant association. *Bulletin of the Torrey Botanical Club*, 53, 7–26.
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences USA*, 101(suppl 1), 5228–5235.
- Grün, B., & Hornik, K. (2011). topicmodels: An R package for fitting topic models. *Journal of Statistical Software*, 40(13), 1–30.
- Hall, D., Jurafsky, D., & Manning, C. D. (2008). Studying the history of ideas using topic models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP '08)*, Honolulu, Hawaii. Association for Computational Linguistics, USA, 363–371.
- Hu, Q., & Willson, G. D. (2000). Effects of temperature anomalies on the Palmer Drought Severity Index in the central United States. *International Journal of Climatology*, 20, 1899–1911.
- Hutchinson, T. F., Sutherland, E. K., & Yaussy, D. A. (2005). Effects of repeated prescribed fires on the structure, composition, and regeneration of mixed-oak forests in Ohio. *Forest Ecology and Management*, 218, 210–228.
- Iverson, L. R., & Prasad, A. M. (1998). Predicting abundance of 80 tree species following climate change in the eastern United States. *Ecological Monographs*, 68, 465–485.
- Jo, I., Fei, S., Oswalt, C. M., Domke, G. M., & Phillips, R. P. (2019). Shifts in dominant tree mycorrhizal associations in response to anthropogenic impacts. *Science Advances*, 5, eaav6358.
- Jo, I., Potter, K. M., Domke, G. M., & Fei, S. (2018). Dominant forest tree mycorrhizal type mediates understory plant invasions. *Ecology Letters*, 21, 217–224.
- Knott, J. A., Desprez, J. M., Oswalt, C. M., & Fei, S. (2019). Shifts in forest composition in the eastern United States. *Forest Ecology and Management*, 433, 176–183.
- Lebrija-Trejos, E., Pérez-García, E. A., Meave, J. A., Bongers, F., & Poorter, L. (2010). Functional traits and environmental filtering drive community assembly in a species-rich tropical system. *Ecology*, 91, 386–398.
- Legendre, P., & Legendre, L. F. (2012). *Numerical ecology* (Vol. 24). Oxford, UK: Elsevier.
- Lenoir, J., Gégout, J.-C., Marquet, P., De Ruffray, P., & Brisse, H. (2008). A significant upward shift in plant species optimum elevation during the 20th century. *Science*, 320(5884), 1768–1771.
- Lin, J. (1991). Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*, 37, 145–151.
- Lloret, F., Escudero, A., Iriondo, J. M., Martínez-Vilalta, J., & Valladares, F. (2012). Extreme climatic events and vegetation: The role of stabilizing processes. *Global Change Biology*, 18, 797–805.
- Loreau, M., Naeem, S., Inchausti, P., Bengtsson, J., Grime, J., Hector, A., ... Schmid, B. (2001). Biodiversity and ecosystem functioning: Current knowledge and future challenges. *Science*, 294(5543), 804–808.
- Lovett, G. M., Weathers, K. C., & Arthur, M. A. (2002). Control of nitrogen loss from forested watersheds by soil carbon: Nitrogen ratio and tree species composition. *Ecosystems*, 5, 0712–0718.
- McEwan, R. W., Dyer, J. M., & Pederson, N. (2011). Multiple interacting ecosystem drivers: Toward an encompassing hypothesis of oak forest dynamics across eastern North America. *Ecography*, 34, 244–256.
- Milborrow, S. (2017). earth: Multivariate adaptive regression splines (Version 4.5.0). Retrieved from <https://CRAN.R-project.org/package=earth>
- Nikita, M. (2016). ldatuning: Tuning of the latent Dirichlet allocation models parameters (Version 0.2.0). Retrieved from <https://CRAN.R-project.org/package=ldatuning>
- Nowacki, G. J., & Abrams, M. D. (2008). The demise of fire and “mesophication” of forests in the eastern United States. *BioScience*, 58, 123–138.
- Oswalt, C. M., Fei, S., Guo, Q., Iannone, B. V. III, Oswalt, S. N., Pijanowski, B. C., & Potter, K. M. (2015). A subcontinental view of forest plant invasions. *NeoBiota*, 24, 49–54.
- Ozier, T. B., Groninger, J. W., & Ruffner, C. M. (2006). Community composition and structural changes in a managed Illinois Ozark Hills forest. *The American Midland Naturalist*, 155, 253–269.
- Pellegrini, A. F. A., Ahlström, A., Hobbie, S. E., Reich, P. B., Nieradzik, L. P., Staver, A. C., ... Jackson, R. B. (2017). Fire frequency drives decadal changes in soil carbon and nitrogen and ecosystem productivity. *Nature*, 553(7687), 194–198.
- Peters, D. P. C., Havstad, K. M., Cushing, J., Tweedie, C., Fuentes, O., & Villanueva-Rosales, N. (2014). Harnessing the power of big data: Infusing the scientific method with machine learning to transform ecology. *Ecosphere*, 5, art67.
- R Core Team (2019). *R: A language and environment for statistical computing* (Version 3.6.1). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Ricklefs, R. E. (1987). Community diversity: Relative roles of local and regional processes. *Science*, 235(4785), 167–171.
- Riitters, K. H., Coulston, J. W., & Wickham, J. D. (2012). Fragmentation of forest communities in the eastern United States. *Forest Ecology and Management*, 263, 85–93.
- Short, K. C. (2017). *Spatial wildfire occurrence data for the United States, 1992–2015 [FPA\_FOD\_20170508]*. (4th edition). Fort Collins, CO: Forest Service Research Data Archive.
- Symstad, A. J., Tilman, D., Willson, J., & Knops, J. M. (1998). Species loss and ecosystem functioning: Effects of species identity and community composition. *Oikos*, 389–397.

- Tang, H., Shen, L., Qi, Y., Chen, Y., Shu, Y., Li, J., & Clausi, D. A. (2013). A multiscale latent Dirichlet allocation model for object-oriented clustering of VHR panchromatic satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, *51*, 1680–1692.
- Tirunillai, S., & Tellis, G. J. (2014). Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent Dirichlet allocation. *Journal of Marketing Research*, *51*, 463–479.
- Valle, D., Baiser, B., Woodall, C. W., & Chazdon, R. (2014). Decomposing biodiversity data using the latent Dirichlet allocation model, a probabilistic multivariate statistical method. *Ecology Letters*, *17*, 1591–1601.
- Vanderwel, M. C., & Purves, D. W. (2014). How do disturbances and environmental heterogeneity affect the pace of forest distribution shifts under climate change? *Ecography*, *37*, 10–20.
- Woodall, C., Oswald, C., Westfall, J., Perry, C., Nelson, M., & Finley, A. (2009). An indicator of tree migration in forests of the eastern United States. *Forest Ecology and Management*, *257*, 1434–1444.
- Zhu, K., Woodall, C. W., & Clark, J. S. (2012). Failure to migrate: Lack of tree range expansion in response to climate change. *Global Change Biology*, *18*, 1042–1052.

## BIOSKETCH

**Jonathan Knott** received his PhD from the Department of Forestry and Natural Resources at Purdue University, working in the Natural Resources Spatial Analysis Laboratory of Songlin Fei (<https://web.ics.purdue.edu/~sfei/index.php>). His research focuses on the impacts of climate change in eastern U.S. forests.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the Supporting Information section.

**How to cite this article:** Knott JA, Jenkins MA, Oswald CM, Fei S. Community-level responses to climate change in forests of the eastern United States. *Global Ecol Biogeogr.* 2020;00:1–16. <https://doi.org/10.1111/geb.13102>