Ecosystems
https://doi.org/10.1007/s10021-022-00765-6

Shifting Forests and Carbon: Linking Community Composition and Aboveground Carbon Attributes

Jonathan Knott,[1]* Grant Domke,[1] Christopher Woodall,[2] Brian Walters,[1] Michael Jenkins,[3] and Songlin Fei[3]

[1]*USDA Forest Service Northern Research Station, 1992 Folwell Ave, St. Paul, Minnesota 55108, USA;* [2]*USDA Forest Service Northern Research Station, 271 Mast Rd, Durham, New Hampshire 03824, USA;* [3]*Department of Forestry and Natural Resources, Purdue University, 715 W State St, West Lafayette, Indiana 47907, USA*

Abstract

Forest communities—assemblages of tree species at stand- to landscape-scales—vary in their capacity to capture and store carbon, but previous attempts using discrete forest types to quantify this relationship often fail to consider gradual changes to the seedling, live tree, and standing dead tree demographic components of the community. We analyzed two decades of plot remeasurements (20,680 remeasurements of 5594 unique plots) from the United States Department of Agriculture Forest Service Forest Inventory and Analysis (FIA) Program across the Great Lakes region of the eastern U.S. to assess shifts in forest communities using Latent Dirichlet Allocation, a continuous model of forest community composition. We tested for links between aboveground live and standing dead tree carbon dynamics and shifts in community composition (the relative proportion of different communities within a stand) of three demographic components (seedlings < 2.54 cm, live trees ≥ 2.54 cm, and standing dead trees ≥ 2.54 cm). Live tree carbon varied across six unique communities ranging from 5.2 Mg C ha^{-1} (SE = 0.77 Mg C ha^{-1}) in the spruce-tamarack community to 88.4 Mg C ha^{-1} (SE = 0.65 Mg C ha^{-1}) in the sugar maple-basswood community. On average, forests contained less standing dead tree carbon (2.2 Mg C ha^{-1}) than live tree carbon (40.4 Mg C ha^{-1}), because of lower stem density and fewer species of standing dead trees. Both live and standing dead tree carbon increased over time across the study area (on average 0.274 and 0.045 Mg C ha^{-1} year^{-1}, respectively). Remeasurements of FIA plots revealed that shifts in community composition of the live and standing dead demographic components were associated with decreases in live tree carbon. Conversely, larger degree of divergence between the community composition of the live and standing dead demographic components and shifts in stand structure between remeasurements (mean, variance, and skewness of the live tree diameter distribution) were associated with increases in live tree carbon. Shifts in stand structure were significant yet weaker predictors of standing dead tree carbon dynamics. Instead, shifts in community composition of the standing dead component over time and degree of divergence between the community composition of the live tree and standing dead tree components were more closely linked to changes in standing dead tree carbon stocks. The effects of shifting communities on carbon dynamics

Received 12 November 2021; accepted 3 April 2022

Supplementary Information: The online version contains supplementary material available at https://doi.org/10.1007/s10021-022-00765-6.

Author contributions: JK and GD designed the study. BW and JK compiled data. JK analyzed data and wrote the manuscript. All authors contributed to interpreting results, framing the manuscript, and editing/revising manuscript drafts.

**Corresponding author; e-mail:* jonathan.knott@usda.gov

Published online: 02 May 2022

are important for understanding carbon cycling in forests and management under continued climate change.

Key words: carbon; forest type; forest community; Forest Inventory and Analysis (FIA); Latent Dirichlet Allocation; standing dead trees; live trees; seedlings; structure.

Introduction

Forests provide valuable resources and services, especially in their ability to capture and store carbon. Forests often sequester more carbon than they emit (Luyssaert and others 2008; Pan and others 2011; Wilson and others 2013) but may become carbon sources (Gundersen and others 2021), especially when faced with major disturbance events such as pest and disease outbreaks, stand-replacing fires, and anthropogenic disturbances such as harvesting (Luyssaert and others 2008; Fei and others 2019). To utilize carbon sequestration in forests as a natural climate solution, scientists and stakeholders must understand how the defining characteristics of forests (for example, species composition and structure) affect the ability of forests to capture and store carbon (Bastin and others 2019; Domke and others 2020).

Forest communities—defined here as assemblages of species at stand- to landscape-scales—can serve as a proxy for many above and belowground biotic and abiotic factors that influence carbon dynamics (Orihuela-Belmonte and others 2013; Hoover and Smith 2021). Forest communities represent the coalescence of a variety of tree species and their associated traits reflective of past climate and other legacy effects. These include links to soil productivity and mycorrhizal associations (Wilson and others 2013; Averill and others 2018; Jo and others 2019), phenology and photosynthetic capacity (Richardson and others 2010; Piao and others 2019), and species and structural diversity (Gough and others 2010, 2020; LaRue and others 2019, 2023), among others.

There have been many efforts to classify forests into communities, and how best to do so has remained a classic debate in ecology. In an early effort, Braun (1950) identified forest regions of the eastern United States (U.S.) using a limited sample of forest plots. Contemporary data from the U.S. Department of Agriculture (USDA) Forest Service, Forest Inventory and Analysis (FIA) Program have been used to test and affirm these patterns (Dyer 2006; Costanza and others 2017; Knott and others 2020). The FIA Program itself uses a classification method that assigns each forest inventory plot into discrete forest types and type groups (note that the terms ''forest type'' and ''forest community'' are often used interchangeably throughout the literature, including by FIA, but ''forest type'' can also use to describe more general forest characteristics, for example, mixed-coniferous forest type; Bechtold and Patterson 2005; Costanza and others 2018). Under discrete classifications, the community composition of a stand (that is, the relative proportion of different unique communities within a stand, in contrast to ''species composition'' which represents the identity of species within a community) is limited to a single community. This prevents discrete classifications from capturing the magnitude of forest change over time: a discrete forest type is either the same between remeasurements or has transitioned from one forest type to another (Valle and others 2014; Knott and others 2020). Consequently, both abrupt changes and gradual shifts in the dominance of community-defining species (for example, stand-replacing fires versus gradual shifts in species dominance during succession) could change a discrete community classification. Without auxiliary data to ascertain the drivers/magnitude of change (for example, records of disturbance or land use change), these shifts are indistinguishable from one another. Therefore, shifts in discrete classification offer limited information to predict how shifts in community composition influence patterns of forest dynamics and carbon cycling with relevance to adaptive management and the delivery of forest ecosystem services (for example, intentionally favoring one community over another by tree planting initiatives or maximizing carbon sequestration and storage by forest managers in response to, or in anticipation of, global change).

To reduce the limitations of discrete classification systems on the detection of community change, a method to classify forest communities across a continuous distribution is necessary. The Latent Dirichlet Allocation model (LDA), a model originally developed for identifying populations within genetic data (Pritchard and others 2000) and topics within text data (Blei and others 2003), has been used as a novel tool to identify forest communities across a continuous distribution, partitioning each stand into a mixture of multiple forest communities and each community into a mixture of multiple species (Valle and others 2014; Knott and others 2020; Hudon and others 2021). Knott and others (2020) described spatial and temporal shifts in

eastern US forest communities using LDA and showed how these shifts correlated to landscape-scale drivers of change. This method quantifies the magnitude of change, ranging from small shifts in species dominance to large shifts due to disturbance, land use change, and climate change.

Research in forest community ecology at the macroscale has focused on shifts in spatial distribution or species-level changes within forest communities (Costanza and others 2017; Knott and others 2020), but little attention has been paid to the resulting shifts in carbon stocks and stock changes. Studies at the local scale in tropical ecosystems have shown that carbon stock changes are closely related to changes in forest type (Orihuela-Belmonte and others 2013), but it is unclear whether these associations hold in temperate forests at larger spatial and temporal scales or across live and dead tree carbon pools. Hoover and Smith (2021) provided estimates of live tree carbon stocks and stock changes across FIA-defined discrete forest type groups at the regional level, but few studies have quantified forest carbon stocks across a diversity of pools (for example, live aboveground biomass, dead wood, litter, and soil) and the changes to carbon stocks resulting from shifts in community composition and degree of divergence between different demographic components of the community (seedlings, live trees, and standing dead trees) that can be used to inform local level management (Figure 1). In addition, forest structure (defined here as characteristics of the size distribution of trees within a stand), productivity and carbon dynamics, and the forest community are interdependent and may vary with environmental context (Gough and others 2010; LaRue and others 2019, 2023; Ullah and others 2021). This interdependence makes it unclear which connections are stronger, whether forest carbon stocks and stock changes more closely linked to forest structure or to community composition.

Here, we used 5594 unique forested plots from the FIA Program across the Great Lakes states (Michigan, Minnesota, and Wisconsin) that have been measured three or more times over about the last 20 years (20,680 measurements total) to (1) quantify shifts in community composition across three demographic components (seedlings, live, and standing dead trees) using the LDA method, (2) estimate aboveground carbon stocks and stock changes across carbon pools associated with two of these demographic components (live and standing dead trees), and (3) assess the impact of shifts in the community composition, structure, and degree of divergence of the three demographic components on carbon dynamics.

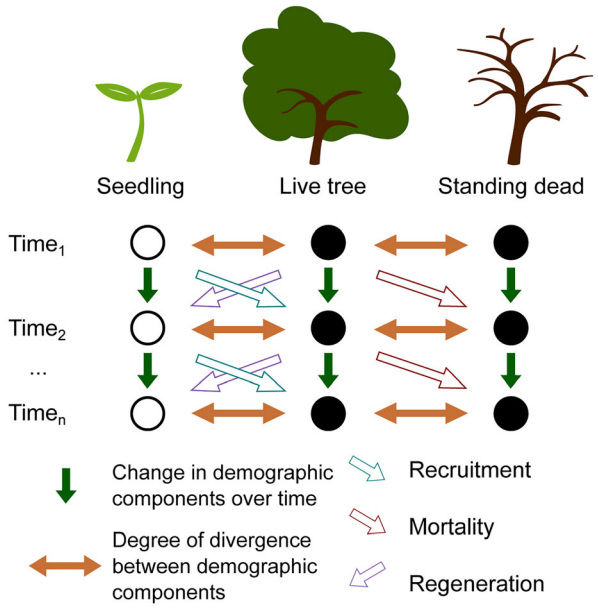

Figure 1. Comparisons across demographic components of forest communities and their associated carbon pools. Forest communities were delineated into three demographic components: seedlings, live trees, and standing dead trees. Circles indicate availability of carbon estimates from FIA for the three components (open = unavailable, filled = available). Vertical arrows represent changes in community composition and carbon stocks of demographic components over time, and horizontal bidirectional arrows represent degree of divergence in community composition between demographic components in each measurement. Diagonal arrows represent comparisons between demographic components over time, recruitment, mortality, and regeneration. Arrow fill represents the ability to directly compare these demographic components within the limits of the FIA data (open = indirect/incomparable, filled = direct comparisons).

Methods

The Forest Inventory and Analysis database

The USDA Forest Service FIA Program conducts the U.S. national forest inventory (NFI). Before 1999, FIA collected data on NFI plots periodically, but since the late 1990s/early 2000s, NFI plot remeasurements have been collected every five to seven years in the eastern U.S at base-level sampling intensity of approximately one inventory plot per 2430 hectares (Bechtold and Patterson 2005). We queried the FIA database (available through the

FIA DataMart, https://apps.fs.usda.gov/fia/datamart/datamart.html) for NFI plots in the Great Lakes region (Michigan, Wisconsin, and Minnesota, Figure 2) with at least three measurements from 1999 to 2019 that were classified as forest at each remeasurement. Plots with only one remeasurement (2 total measurements) were excluded because they were likely newly established, some of their measurements were assigned to a different plot identifier in the FIA database due to missing and/or incorrect data such as plot coordinates, or they were classified as non-forest in at least one remeasurement.

We aimed to assess the community dynamics and carbon stocks of three main demographic components: seedlings [> 30.48 cm tall but < 2.54 cm in diameter at root collar (DRC), or breast height (DBH), depending on the overall height of the seedling], live trees (≥ 2.54 cm DBH), and standing dead trees (≥ 2.54 cm DBH, > 1.37 m length, and < 45° lean). To do so, we compiled two main datasets: (1) plot × species matrices for each demographic component to identify and assess changes to forest communities using LDA, and (2) per-unit-area estimates of live tree and standing dead tree carbon and stand structure for each plot. We were able to compile dataset (1) for all three demographic components, but dataset (2) was only available for live and standing dead trees because seedling carbon is considered *de minimis* and is not estimated by the FIA Program due to differences in sampling protocol between seedlings and live and standing dead trees (tallies of seedlings meeting height and diameter criteria versus DBH measurements). Details for how we constructed each dataset follow, and a description of the FIA database and query criteria can be found in Appendix 1.

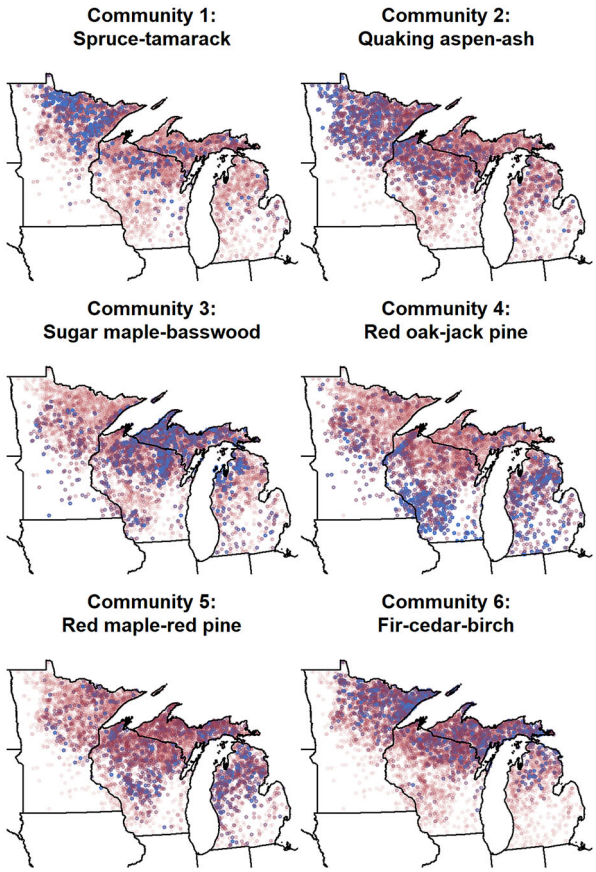

Figure 2. Approximate FIA plot locations and the proportion of each community in the live tree demographic component during the earliest measurement of each plot. FIA plot coordinates are fuzzed (small random noise added to actual plot coordinates) and swapped (some plot locations are switched between similar plots within a county) to anonymize forest landowner data. Color and opacity indicate community proportion: pale red = lower proportion, dark blue = higher proportion.

Dataset (1): plot × species matrices

The LDA model requires a plot × species matrix, where each row is a plot, and each column is a measure of abundance of an individual species. While raw counts can be used (Valle and others 2014), we relativized our data to reduce the effects of structure (to balance out plots with a few large trees with plots with many small trees, Knott and others 2020) and the FIA sampling protocol (trees > 12.7 cm DBH are measured on 168 m^2 subplots, but seedlings and smaller trees 2.54–12.7 cm DBH are tallied and measured on 13.5 m^2 microplots nested within the subplots) on community composition (see Appendix 1 for more details on the NFI plot design). We used a relative measure, importance value (IV), for each species on each plot by averaging relative stem density (stems ha^{-1}) and relative basal area (m^2 ha^{-1}), converting to percentage, and rounding up to the nearest integer percent (to prevent the loss of species with IV > 0% but < 0.5% and accommodate LDA model fitting). This process was completed for live trees and standing dead trees. Because no diameter information is available for seedlings, we computed relative stem density instead of IV. Each plot represented approximately 100 IV or relative stem density units, with some plots containing more units due to rounding (maximum 110 for live trees, 105 for standing dead trees, and 112 for seedlings).

To compare the seedling count data to the live tree abundance and attribute data, we only included species that occurred in both the seedling

and live tree datasets. This excluded 8 species from the live tree and seedling data, but only represented 34 live trees (a very small portion, 0.007%, of the total live tree carbon). We also only included plots that had both seedling and live tree data for all remeasurements. Not all species were present in the standing dead tree data (18 out of 91 species) and some plots contained no standing dead trees (~ 21% of measurements). Species that did not occur in the standing dead tree data were imputed as zeros, and plots that did not contain any standing dead trees were removed to accommodate the LDA model fitting process. Our final plot × species matrices included a total of 91 species across 20,680 total measurements of seedlings (887,878 stems) and live trees (731,888 stems) on 5594 unique plots and 73 species across 16,274 total measurements of standing dead trees (66,137 stems) on 5153 unique plots.

Dataset (2): estimates of aboveground carbon and forest structure

From the individual tree records, we obtained estimates of aboveground carbon from the FIA database. The FIA Program uses a component ratio method to estimate carbon based on the volume of different components of the tree if DBH was above 12.7 cm or based on allometric relationships between DBH and biomass for smaller trees if DBH was below 12.7 cm (Woodall and others 2011; Wilson and others 2013). We used FIA expansion factors to calculate per-unit-area carbon density (megagrams carbon per hectare, Mg C ha^{-1}) represented by each tree. We summed the carbon estimates at the plot-level for live and standing dead trees to estimate total aboveground live and standing dead tree carbon stocks. Note that only plots with standing dead trees had estimates of both the community composition and standing dead tree carbon, that is the standing dead tree carbon estimates were zero-truncated instead of zero-inflated (Woodall and Macfarlane 2010; An and MacFarlane 2013).

Shifts in community composition, structure, and structural diversity are all linked to forest carbon and provide different perspectives on how carbon stocks change over time (LaRue and others 2019, 2023). We used DBH measurements of all live trees to calculate plot-level measures of forest structure to compare to community-level estimates of change. Specifically, we calculated three measures of structure, (1) mean DBH, which estimates the average size of trees on the plot, (2) variance of DBH, which estimates the diversity of size classes (a measure of horizontal structural diversity, LaRue and others 2023), and (3) skewness of the DBH distribution, which estimates the asymmetry of the DBH distribution. We also calculated changes to these three measures over time to compare the effects of shifts in structure versus shifts in the community composition of the seedling, live tree, and standing dead demographic components of the community on forest carbon stocks and stock changes.

Modeling community composition and dynamics with Latent Dirichlet Allocation

The LDA model is most often used as a topic model for text mining that predicts (1) the proportion of each text document (sentence, paragraph, full document, and so on, depending at which level the document × term matrix—equivalent to our plot × species matrix—is constructed) composed of each topic, and (2) the proportion of each topic composed of each term (Blei and others 2003). The text mining algorithm has a clear analogue for forest ecosystems: forests are composed of communities and communities are composed of a variety of species (Valle and others 2014). The mathematical basis for this model can be found in Blei and others (2003) and Pritchard and others (2000), and detailed overviews of how it can be used in biodiversity and conservation science can be found in Valle and others (2014) and Hudon and others (2021). Briefly, LDA (in the context of forest inventory data) is a hierarchical Bayesian model where each plot is modeled as a mixture of underlying forest communities (assemblages of species across plots) and each community is modeled as a mixture of underlying species proportions (Valle and others 2014). The resulting posterior distributions of the LDA model contain the proportion of each community in each plot and the proportion of each species in each community, which can be further analyzed for changes over time or space using more conventional methods such as regression, randomization tests, and multivariate distance metrics (Knott and others 2020).

Like many clustering methods, LDA requires specifying a priori the number of clusters (communities) to be identified by the model. When the number of communities is unknown, previous studies suggest iteratively increasing the number of communities (k) and evaluating model fit for each iteration using AIC (Valle and others 2014) or a suite of metrics (Nikita 2016; Knott and others 2020). We ran the LDA model on the live tree data

from the first measurement of each plot, increasing k by 1 from 2 to 12 (the number of communities found across the entire eastern U.S. in Knott and others 2020). Models were fit in the *topicmodels R* package (Hornik and Grün 2011) via Gibbs sampling with 5 repeated random starts, 5000 total iterations each, with an initial burn-in period of 500 iterations. For each level of k, the model resulting from the random start with the maximum posterior likelihood was retained. We compared models with k communities using four metrics of model fit from the *ldatuning R* package (Nikita 2016). Two metrics are maximized and 2 are minimized at the best fit number of communities. A visual inspection of these metrics showed that 2 metrics were uninformative (never reaching a local maximum or minimum across the entire range of k), but two others had maximum and minimum values at $k = 6$ communities (Figure S1). We explored the posterior distributions of the $k = 6$ model to see how species composition of the communities aligned with our expectations about the species composition of forests in the Great Lakes region, for example, whether any communities were comprised of only one species or whether well-known species associations were detected by the model (for example, sugar maple and basswood, Braun 1950). The second-best model, $k = 10$ communities, was also evaluated based on the resulting species composition, but the species composition of some communities did not align with previously published forest communities of the Great Lakes region (for example, Braun 1950; Bechtold and Patterson 2005; Dyer 2006; Costanza and others 2017; or Knott and others 2020). Therefore, we focused the rest of the results and discussion on the $k = 6$ community model (see Appendix 2 for more information about the model selection process).

The posterior distribution of the final $k = 6$ community model was used to predict the proportion of each community in each of the remeasurements in the live tree data ($n = 15{,}294$ remeasurements). The $k = 6$ model was also used to predict the proportion of each live tree community represented in the seedling and standing dead plot × species datasets. Species composition of each community remained static over time using this approach. We used this approach as opposed to fitting new models for the seedling and standing dead tree data to make direct comparisons across demographic components of the community, but our methods presented a tradeoff. We were able to assess changes in community proportions within each plot separate from changes in species associations across plots, which allowed us to track specific communities through time across multiple plot remeasurements versus a comparison across a single interval (for example, Valle and others 2014; Knott and others 2020). However, this method precludes the formation of novel communities over time or shifts in species composition of the communities.

Neither a discrete model of forest communities nor the LDA model perfectly reflects the ecological transition from one forest community to another. However, LDA allows for a range of magnitude in shifts in community composition versus only abrupt transitions from one community to another that can be identified from other discrete approaches. LDA also partitions carbon pools of each demographic component into a variety of communities, which, hypothetically, links carbon pools to the different communities more accurately than discrete classifications. To test this assumption, we compared models with LDA-identified communities to FIA-defined forest types and broader forest type groups (55 forest types and 15 forest type groups in the Great Lakes region).

We compared community proportions among and within demographic components using Jensen-Shannon Distance (JSD), a continuous distance measure for compositional data (Lin 1991; Hall and others 2008). JSD is symmetric and bounded by [0,1], where 0 represents identical community composition, and 1 represents completely different community composition (Lin 1991; Hall and others 2008; Knott and others 2020). JSD was used to track changes in each demographic component over time: seedlings, representing community shifts within the understory; live trees, representing shifts in the main carbon-capturing species; and standing dead trees, representing the risk of mortality within certain communities (Figure 1). JSD was also used to estimate the degree of divergence in community composition of the different components of the community within a stand. The degree of divergence estimates the diversity of communities across demographic components as well as identifies stands where the current (live) community composition varies from the past (standing dead) and future (seedling) community composition of the stand.

We used generalized linear mixed effects models (GLMMs) and mixture design models to address four main comparisons: (1) How are live and standing dead tree carbon partitioned across the six communities? (2) Which communities have the most changes to live tree carbon (which are cap-

turing or losing carbon most rapidly) and standing dead tree carbon (most at risk for mortality)? (3) Are shifts in community composition or structure of the different demographic components (seedlings, live trees, and standing dead trees) indicative of changes in carbon stocks? (4) Are forests with larger degree of divergence between the seedling, live tree, and standing dead community components more/less carbon dense than those with smaller degree of divergence? Because community composition is a proportion/relative frequency (that is, the sum of all community proportions in a plot equals to one; in a six-community model, if the proportion of communities 1–5 are known, the proportion of community 6 is also known via subtraction from one), regression models containing all community proportions as predictors cannot be fit due to singularity in the matrix calculations. While there are methods for analyzing compositional data (for example, mixture design used in manufacturing, Eriksson and others 1998) and remeasurements (for example, GLMMs with random effects of sampling unit, Zuur and others 2009) there are not, to our present knowledge, methods to account for both compositional predictors and repeated measures in a single model without significant transformations and/or variable reduction which makes interpretation difficult. Instead, we used two approaches which produced similar results: (1) GLMMs for each individual community of each demographic component which account for remeasurements but not compositional data (using *R* package *lme4*, Bates and others 2015) and (2) mixture design models which account for compositional data but not remeasurements (using *R* package *mixexp*, Lawson and Willden 2016).

We fit GLMMs to assess carbon stocks as a function of community composition of the seedling, live tree, and standing dead demographic components and three measures of stand structure: live tree carbon stocks as a function of the proportion of the six communities (individually) across the seedling and live tree demographic components (12 models with one fixed effect each); standing dead tree carbon stocks as a function of the proportion of the six communities across the standing dead and live tree demographic components (12 models with one fixed effect each); and live and standing dead tree carbon stocks as a function of the mean, variance, and skewness of the DBH distribution (two models with three fixed effects each). GLMMs contained a random intercept term for plot ID (to account for remeasurements) and used a log-normal error distribution for carbon stocks (Gaussian distribution with a log link function). We also fit GLMMs using FIA-defined forest types (55 forest types/54 indicator variables) and forest type groups (15 forest type groups/14 indicator variables) to compare to a GLMM containing all but one LDA-identified community and the mixture design model.

We also fit linear (normal error distribution) mixed effects models for change in live tree and standing dead tree carbon as a function of changes in community composition (JSD value) in the seedling (β_{seed}), live tree (β_{live}), and standing dead (β_{dead}) demographic components of the community over time and degree of divergence in community composition between seedling versus live tree (β_{seed_live}), and live tree versus standing dead tree (β_{dead_live}) demographic components of the community within a stand at the time of each remeasurement. These models used the same random intercept term for plot ID to account for remeasurements. The intercept (β_0) values of these models represent the mean change in carbon stocks over time if the seedling, live tree, and standing dead communities are identical ($JSD_{seed\_live} = JSD_{dead\_live} = 0$) and static between remeasurements ($JSD_{seed} = JSD_{live} = JSD_{dead} = 0$). We fit similar models using mean DBH (β_{mean_dbh}), variance of DBH (β_{var_dbh}), and skewness of the DBH distribution (β_{skew_dbh}) as predictors of change in live tree and standing dead tree carbon.

Mixture design models were fit with live or standing dead tree carbon stocks as a function of all six community proportions as predictors. We fit two models for live tree carbon, one with live tree community proportions and one with seedling community proportions as predictors. Similarly, we fit two models for standing dead tree carbon, one with live tree community proportions and one with standing dead community proportions as predictors. Coefficient estimates from mixture design models were used to estimate carbon stocks in stands dominated by each community in each demographic component (Table 1). For both GLMMs and mixture design models we corrected for multiple testing by using a stricter $\alpha = 0.01$ level to assess significance of predictors (that is testing if 99% confidence intervals contain/exclude zero).

Results

The best-fit LDA model identified six unique, distinct forest communities within the FIA live tree data (Table 1, Figure 2). Our initial live tree dataset contained 91 species, but only 38 species were strongly associated with a certain community

Table 1. Carbon Stocks and Carbon Stock Change Across Forest Communities of the Great Lake States

Community	Species composition (proportion)	Mean community proportion	Estimated carbon density (Mg C ha^{-1})	Estimated carbon change (Mg C ha^{-1} y^{-1})
1: Spruce-tamarack	*Picea mariana (0.500)* *Larix laricina (0.261)* *Populus balsamifera (0.110)* *Fraxinus pennsylvanica (0.100)* *Crataegus* spp. *(0.013)*	Live: 0.149 Standing dead: 0.128 Seedlings: 0.118	Live: 5.18 (0.77) Standing dead: 2.18 (0.11)	Live: 0.20 (0.08) Standing dead: 0.17 (0.03)
2: Quaking aspen-black ash	*Populus tremuloides (0.725)* *Fraxinus nigra (0.210)* *Carpinus caroliniana (0.028)* *Prunus pensylvanica (0.012)*	Live: 0.174 Standing dead: 0.218 Seedlings: 0.213	Live: 10.93 (0.75) Standing dead: 1.79 (0.08)	Live: 0.64 (0.08) Standing dead: 0.07 (0.02)
3: Sugar maple-basswood	*Acer saccharum (0.552)* *Tilia americana (0.125)* *Ostrya virginiana (0.086)* *Betula alleghaniensis (0.066)* *Tsuga canadensis (0.061)* *Fagus grandifolia (0.043)* *Fraxinus americana (0.042)* *Acer saccharinum (0.019)*	Live: 0.153 Standing dead: 0.194 Seedlings: 0.140	Live: 88.40 (0.65) Standing dead: 2.39 (0.10)	Live: − 0.02 (0.07) Standing dead: 0.06 (0.02)
4: Red oak-jack pine	*Quercus rubra (0.177)* *Pinus banksiana (0.172)* *Populus grandidentata (0.126)* *Prunus serotina (0.103)* *Ulmus americana (0.088)* *Quercus alba (0.075)* *Quercus macrocarpa (0.067)* *Quercus ellipsoidalis (0.050)* *Quercus velutina (0.045)* *Acer negundo (0.017)* *Carya ovata (0.016)* *Carya cordiformis (0.012)*	Live: 0.186 Standing dead: 0.167 Seedlings: 0.191	Live: 37.85 (0.75) Standing dead: 2.66 (0.08)	Live: 0.40 (0.08) Standing dead: 0.06 (0.02)
5: Red maple-red pine	*Acer rubrum (0.603)* *Pinus resinosa (0.237)* *Pinus strobus (0.146)* *Amelanchier* spp. *(0.013)*	Live: 0.161 Standing dead: 0.156 Seedlings: 0.145	Live: 54.45 (0.87) Standing dead: 1.99 (0.13)	Live: 0.61 (0.09) Standing dead: − 0.04 (0.03)
6: Fir-cedar-birch	*Abies balsamea (0.417)* *Thuja occidentalis (0.252)* *Betula papyrifera (0.220)* *Picea glauca (0.080)* *Acer spicatum (0.029)*	Live: 0.176 Standing dead: 0.136 Seedlings: 0.192	Live: 37.06 (0.73) Standing dead: 2.33 (0.08)	Live: 0.10 (0.07) Standing dead: − 0.02 (0.02)

Species list contains all species with proportion $> 1/m_{spp}$. *Carbon stocks (density) and carbon stock change were estimated for plots dominated by each community using mixture design model coefficients (standard error in parentheses). Note that standing dead estimates are only calculated on plots with standing dead trees (zero-truncated, not zero-inflated).*

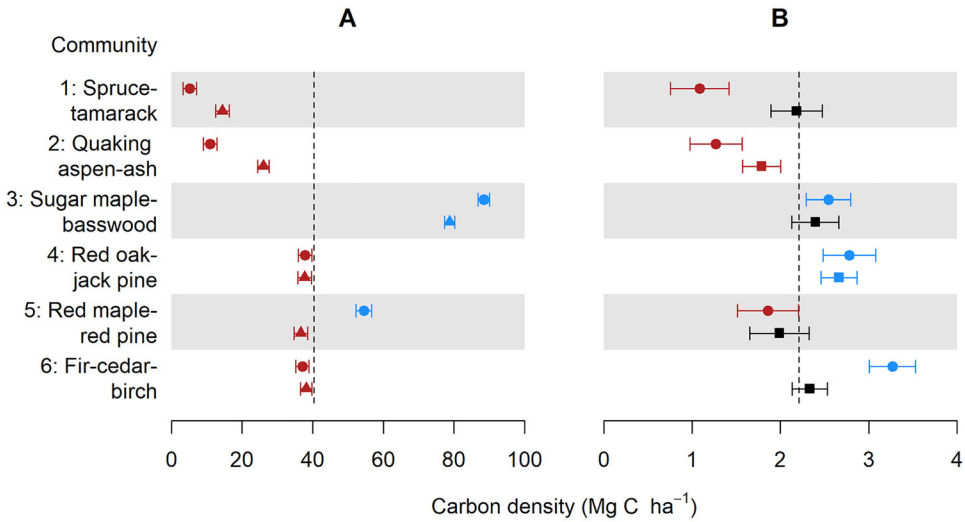

Figure 3. Estimates of live and standing dead tree carbon stocks by community type. Coefficients (Mg C ha^{-1}) by community type were estimated from mixture design models and represent the **A** live tree carbon stocks in forests dominated by each community type in the live tree (circles) and seedling (triangles) components of the community and **B** standing dead tree carbon stocks in forests dominated by each community type in the live tree (circles) and standing dead (squares) components of the community. Error bars indicate 99% confidence intervals (CI). Vertical lines indicate mean carbon density (40.35 and 2.21 Mg C ha^{-1} in **A** and **B** for live and standing dead tree carbon, respectively). Colors indicate significance of relationship (red = CI below the mean, blue = CI above the mean, black = CI contains the mean).

(species proportion $>$ 1/91, $\sim$ 0.011, in any community). The number of species in a community meeting this criterion ranged from 4 to 12. The mean proportion of each community was similar across the live tree communities, ranging from 0.149 (community 1, spruce-tamarack) to 0.186 (community 4, red oak-jack pine). Species composition was similar but not identical to other subcontinental studies of forest communities (for example, Braun 1950; Dyer 2006; Knott and others 2020) due to fitting the model within the Great Lakes region. For example, the conifer-birch community in Knott and others (2020) and fir-cedar-birch (community 6) in this study both contained balsam fir (*Abies balsamea*), northern white cedar (*Thuja occidentalis*), and paper birch (*Betula papyrifera*). The spruce-tamarack community in Knott and others (2020) was similar to the spruce-tamarack community (community 2), both containing black spruce (*Picea mariana*) and tamarack (*Larix laricina*) as dominant species. Similarly, Braun (1950) and Dyer (2006) identified maple-basswood as a defining community of the Great Lakes region, which matched with our sugar maple-basswood community (community 3).

The second-best LDA model contained 10 communities (Figure S1) but did not conform to prior observations in the region. Although some communities identified in the k = 10 community model were expected for the region (such as sugar maple-basswood), others did not align with previously published work (for example, Braun 1950; Bechtold and Patterson 2005; Dyer 2006; Costanza and others 2017; or Knott and others 2020) or poorly reflected the underlying FIA data (such as a quaking aspen-only community, which only occurred as a monoculture on $<$ 2% of FIA plots where quaking aspen was found in the Great Lakes region, Appendix 2).

Aboveground live tree carbon varied across the six communities (Table 1). The mean plot-level estimate of aboveground live tree carbon was 40.4 Mg C ha^{-1} (standard error, SE = 0.18 Mg C ha^{-1}). Coefficient estimates from the mixture design model for plots completely dominated by each community type ($P < 0.001$, corrected R^2 = 0.31, Figure 3a, Table S1) indicated that community 3, the sugar maple-basswood community, had the highest carbon density at 88.4 Mg C ha^{-1} (SE = 0.65 Mg C ha^{-1}) whereas community 1, the spruce-tamarack community, had the lowest carbon density at 5.2 Mg C ha^{-1} (SE = 0.77 Mg C ha^{-1}). Community 5 (red maple-red pine) was predicted to have higher than average carbon density at 54.5 Mg C ha^{-1} (SE = 0.87 Mg C ha^{-1}), and Community 2 (quaking aspen-black ash) was predicted to have the second lowest carbon density at 10.9 Mg C ha^{-1} (SE = 0.75 Mg C ha^{-1}). Communities 4 (red oak-jack pine) and 6 (fir-cedar-birch) were predicted to have

carbon density similar to but slightly below the average (total = 37.9 and 37.1 Mg C ha^{-1}, SE = 0.75 and 0.73 Mg C ha^{-1}, respectively). Generalized linear mixed-effects models with the proportion of each community in the tree data as a predictor of forest carbon density (log-normal error distribution, with a random intercept for plot ID to account for the multiple remeasurements) confirmed these patterns. The proportion of communities 3 (sugar maple-basswood), 5 (red maple-red pine), and 6 (fir-cedar-birch) in the live tree component were positively associated with live tree carbon density and communities 1 (spruce-tamarack), 2 (quaking aspen-black ash), and 4 (red oak-jack pine) were negatively associated with live tree carbon density. The mixture design model with seedling community proportions as predictors of live tree carbon stocks found similar patterns, except community 5 had only small differences from the mean live tree carbon density ($P < 0.001$, corrected $R^2 = 0.22$, Figure 3a, Table S1).

Two GLMMs—one containing FIA-defined forest types (55 forest types represented by 54 indicator variables) and one containing FIA-defined forest type groups (15 forest type groups represented by 14 indicator variables)—performed significantly worse than the mixture design model and a GLMM with the first five LDA-identified communities as predictors of live tree carbon (AICc weight = 1 for the mixture design model compared to all models and AICc weight = 1 for the GLMM with LDA-identified communities compared to GLMMs with FIA-defined forest types and forest type groups). Comparisons of these models also showed that FIA-defined forest type groups were better predictors of live tree carbon stocks than the more specific FIA-defined forest types because of the rarity of some forest types (12 out of 55 forest types were found on < 10 plots).

Standing dead tree carbon stocks were much lower than live tree carbon stocks. The average plot held 2.2 Mg C ha^{-1} (SE = 0.02 Mg C ha^{-1}) standing dead tree carbon (versus 40.4 Mg C ha^{-1} mean live tree carbon, approximately 5.5% of the live tree carbon). The relationship between standing dead tree carbon stocks and standing dead community composition was weaker than the analogous relationship in the live tree community ($P < 0.001$ but corrected $R^2 < 0.01$ for the standing dead mixture design model, Figure 3b). Standing dead tree carbon stocks varied significantly across different community compositions of the live tree and standing dead components of the community (based on mixture design model coefficients like the live tree carbon model, Figure 3b, Table S1). The standing dead tree carbon stocks were the highest for community 4 (red oak-jack pine) (2.6 Mg C ha^{-1}, SE = 0.079 Mg C ha^{-1}) and lowest for community 2 (quaking aspen-ash) (1.8 Mg C ha^{-1}, SE = 0.084 Mg C ha^{-1}). Communities 3 (sugar maple-basswood) and 6 (fir-cedar-birch) also had higher than average standing dead tree carbon density (total = 2.4 and 2.3 Mg C ha^{-1}, SE = 0.103 and 0.077 Mg C ha^{-1}, respectively). Community 5 (red maple-red pine) had lower than average standing dead tree carbon density (2.0 Mg C ha^{-1}, SE = 0.134) while Community 1 (spruce-tamarack) had similar to (but slightly lower than) average standing dead tree carbon density (2.2 Mg C ha^{-1}, SE = 0.129).

Variables related to forest structure were significant predictors of carbon stocks. Mean DBH, variance of DBH, and skewness of the DBH distribution were all positive predictors of live and standing dead tree carbon. Unsurprisingly, this indicated that plots with larger trees, more variation in tree size, and more positive/right skewed distributions (indicative of plots with the presence of large trees) had more carbon across live and standing dead components. However, despite the significance of these model coefficients ($P < 0.001$ for all coefficients in both the live and standing dead models), they generally explained a low amount of variance relative to the random plot effect [GLMM marginal R^2 (fixed effects only) < 0.01 for both models, GLMM conditional R^2 (fixed and random effects) = 0.23 and 0.28 for live and standing dead models, respectively].

Carbon stocks changed over time, but stock changes were only weakly associated with individual community proportions (GLMM marginal $R^2 < 0.07$ for all communities). Measures of change over time across all components of the community (JSD between remeasurements for seedling, live, and standing dead components) and within stand degree of divergence in community composition between components of the community (JSD between seedlings, live trees, and standing dead trees within a measurement period) were generally better predictors of changes in forest carbon stocks than the proportion of any community alone (GLMM marginal $R^2 = 0.11$). The model with shifts in forest structure alone did not perform as well as the model with shifts in community composition alone, but the best model included both structure and community shifts as predictors of live tree carbon change (GLMM marginal $R^2 = 0.16$, Figure 4a). Specifically, the best model indicated that plots exhibited general increases in live tree carbon over time (significant positive

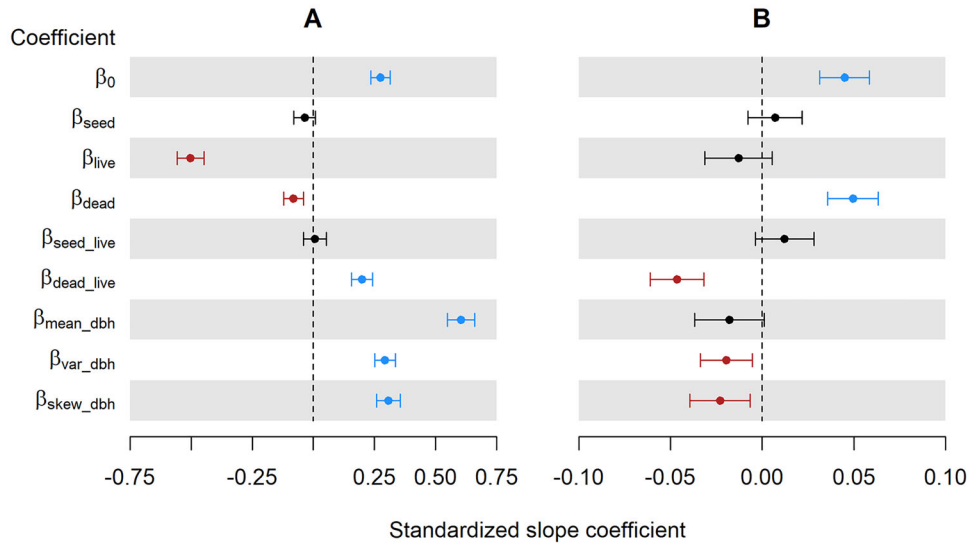

Figure 4. Effects of changes in the components of forest communities and shifts in stand structure on changes in carbon stocks. Coefficients were estimated using a linear mixed effects model with Jensen-Shannon Distance (JSD) in the seedling (β_{seed}), the live tree (β_{live}), and standing dead (β_{dead}) components of the community over time, JSD as an estimate of the degree of divergence in species composition between the seedling and live tree (β_{seed_tree}), and live tree and standing dead (β_{dead_tree}) demographic components of the community within stands at each remeasurement, and shifts in forest structure (mean DBH, β_{mean_dbh}, variance of DBH, β_{var_dbh}, and skewness of diameter distribution, β_{skew_dbh},) as predictors of **A** live and **B** standing dead tree carbon stock change (Mg C ha^{-1} y^{-1}). Intercept (β_0) represents the change in carbon stocks over time if the seedling, live tree, and standing dead communities are identical (β_{seed_live} and β_{dead_live} = 0), and both community composition and structure are static between remeasurements (β_{seed}, β_{live}, β_{dead}, β_{mean_dbh}, β_{var_dbh}, and β_{skew_dbh} = 0). Standardized slope coefficients (β values) are estimates of effect size for change in carbon stocks when each predictor increased by one standard deviation (while holding others fixed). Error bars indicate 99% confidence intervals (CI). Colors indicate significance of relationship (red = CI below 0, blue = CI above 0, black = CI contains 0).

intercept, β_0 = 0.43 Mg C ha^{-1} $year^{-1}$, $P < 0.001$, Figure 4a, Table S2) except when the community composition of the live or standing dead components change over time (β_{tree} = − 0.502 and β_{dead} = − 0.081, $P < 0.001$ for both predictors, Figure 4a, Table S2). A significant positive coefficient for β_{dead_live} showed that plots tended to increase live tree carbon over time when the community composition of the standing dead component of the community diverged from the live tree component (β_{dead_live} = 0.199, $P < 0.001$, Figure 4a). Shifts in forest structure were also positive predictors of live tree carbon shifts: increases in mean, variance, or skewness of the DBH distribution (a shift toward larger and/or more size classes) were associated with increases in live tree carbon (β_{mean_dbh} = 0.604, β_{var_dbh} = 0.293, and β_{skew_dbh} = 0.307, $P < 0.001$ for all three predictors, Figure 4a, Table S2).

Similarly, forests generally increased in standing dead tree carbon over time (β_0 = 0.066 Mg C ha^{-1} $year^{-1}$, $P < 0.001$, Figure 4b, Table S2). Change in community composition of the standing dead component over time was the strongest positive predictor of change in standing dead tree carbon (β_{dead} = 0.049, $P < 0.001$, Figure 4b, Table S2), and degree of divergence between the live tree and standing dead components of the community was the strongest negative predictor of change in standing dead tree carbon (β_{dead_live} = − 0.046, $P < 0.001$, Figure 4b, Table S2). Changes in variance and skewness of the diameter distribution were significant, yet weaker, negative predictors of shifts in standing dead tree carbon (β_{var_dbh} = − 0.020, $P < 0.001$; and β_{skew_dbh} = − 0.023, $P < 0.001$, Figure 4b, Table S2).

Discussion

In contrast to discrete approaches, the continuous distribution of forest communities identified by LDA showed how forest carbon pools are partitioned across unique communities of the northern forest. Our comparison of models using LDA-identified communities versus FIA-defined forest types showed that carbon stocks are better attributed to communities using the continuous LDA model. The variability of carbon stocks per com-

munity was similar between our study and Hoover and Smith (2021) (SE = 0.65–0.87 vs. 0.36–0.85 Mg C ha^{-1} in Hoover and Smith 2021), but community-level mean carbon stocks ranged from 5.2 to 88.4 Mg C ha^{-1} for LDA-identified communities versus 29.3–59.7 Mg C ha^{-1} for FIA-defined forest type group. While discrepancies are partially due to plot selection criteria (5594 plots remaining forest over three or more measurements in our study versus 8094 plots remaining forest over two measurements in Hoover and Smith 2021), the reduced variability of carbon stocks between FIA-defined discrete forest type groups showed how carbon stock estimates trend toward the mean when they are not partitioned into different communities/forest types within a plot. Additionally, the continuous distribution of the LDA-identified communities addressed gradual changes to community composition even when forests did not shift in FIA-defined forest type.

We found that the two maple-dominated communities (3, sugar maple-basswood and 5, red maple-red pine) had the highest live tree carbon density, but were less prevalent than the other communities except spruce-tamarack (ranked 5th and 4th, respectively, for mean community proportion across the region). Conversely, forests with a larger proportion of the spruce-tamarack (community 1) and quaking aspen-black ash (community 2) had the lowest live tree carbon density. This may in part be due to the higher proportion of early successional species in these communities: tamarack is often the first species to establish in hydric areas such as filled-lake bogs, and quaking aspen is a common early successional species on drier sites (Burns and Honkala 1990a, 1990b). All four of the defining species in these communities are classified as shade intolerant (Burns and Honkala 1990b, 1990a), indicating that they may be limited in capturing carbon in denser stands when competing for sunlight.

We did not find strong relationships between the rate of change of either live or standing dead tree carbon stocks and the proportion of any one forest community; instead, changes in carbon stocks were more closely related to overall changes in forest community composition. Not surprisingly, changes to live and standing dead tree carbon pools were most closely linked to changes in community composition of the respective demographic components. However, we found it interesting that the degree of divergence in community composition between the live and standing dead demographic components (that is, JSD between live tree versus standing dead communities) was a significant predictor of change in live tree carbon stocks, indicating that stands where the past (standing dead) community composition differed from the current (live) community composition (more dynamic stands) tended to capture carbon more rapidly than stands where the past and current community composition was static (Figure 4). This may indicate forests transitioning from less carbon dense early successional communities to more carbon dense later successional communities and/or increases in species and structural diversity when more community types are present in the different demographic components (Gough and others 2010; Niklaus and others 2017; LaRue and others 2019). Our analysis of shifts in forest structure generally supported this interpretation: stands where the DBH distribution shifted toward larger trees and/or stands that were more structurally diverse (increase in DBH variance) tended to increase in live tree carbon over time. It is also consistent with broader landscape-level land use change, where more dynamic landscapes sequestered more carbon (Woodall and others 2015).

Our analysis was at the plot/community level, where comparisons could be made within demographic components of the community over time or between demographic components of the community at a given time (Figure 1). However, linking changes between different carbon pools and demographic components over time remains a challenge—for example, estimating regeneration, recruitment, or mortality (empty diagonal arrows in Figure 1). Limitations in the FIA sampling protocol (for example, seedling data collected as stem counts versus DBH measurements; Shriver and others 2021), human intervention (for example, thinning of the understory and removal of trees for timber), and physical/biological processes (for example, probability of a dead tree remaining standing due to topography and physiography, traits including rooting depth, decay rate, and wood density, and other stand-level characteristics; Oberle and others 2018) could bias the estimation of regeneration, recruitment, and mortality rates without more precise individual tree demographic data and modeling (for example, see Shriver and others 2021).

Here, we focused our analyses on the aboveground carbon pools of the live and standing dead demographic components of the community, but these represent only part of the carbon pools within forests. Soils and the belowground interactions between tree roots and microbial communities are important parts of the forest carbon cycle (Silva and Lambers 2021). Averill and others (2018) and Jo

and others (2019) showed that forest carbon stored in soils is closely linked to mycorrhizal associations of the tree community and nitrogen deposition favors arbuscular mycorrhizal (AM) associated tree species. This has the potential to further shift forest communities and belowground carbon stocks, as root-derived carbon is generally greater for AM trees (Keller and others 2021). Although FIA collects soil data on a limited subsample (approximately 1/16) of forest plots in our study area, resampling does not occur at the same frequency as the forest inventory and is not yet available in the FIA database to allow comparisons between changes in forest communities and aboveground carbon stocks to changes in soil attributes.

We used the live tree LDA model to predict the proportion of each community in the seedling and standing dead tree data, which represented a tradeoff in methodology: by keeping the species compositions fixed, we more directly compared the three demographic components both over time and within each remeasurement, but we lost the ability to assess changes in species composition within a community type over time, which also precludes the formation of additional unique communities (Knott and others 2020). This forced understory communities, which may be more dynamic than live tree communities and vary across timescales shorter than the remeasurement period (that is, changes that occur in less than 5–7 years; Small and McCarthy 2003), to align with the mid- and overstory components of the live tree data. Unique mixtures of species may exist within the understory that never become part of the mid-or overstory (and eventually deadwood when mortality occurs) because of the strong environmental filter that prevents members of the seedling layer from recruiting into the upper strata (Shriver and others 2021). Similarly, the presence and abundance of standing dead trees may be species- or cohort-specific. It can be difficult for field crews to identify the species of dead trees, and some species (in this study, 18 out of 91 species) do not produce stems large enough to meet standing dead criteria (Woodall and Nagel 2006). Forest structure and/or species traits may influence the probability of trees meeting standing dead criteria (for example, more dense forests may prevent trees from falling, and trees with deeper roots may be more likely to remain standing after dying; Oberle and others 2018). Certain species or cohorts may experience periodic dieback (for example, due to disease or insect damage) which affects the relative proportion of species within a community (Heitzman 2003; Knott and others 2020). Still, our analysis represents one of the first attempts at linking community and carbon dynamics across three demographic components of the forest community.

Trends in community composition change may be region-specific and confounded by other legacy effects and/or data availability. For example, comparing temporal trends in community composition and carbon dynamics in our study region to other parts of the U.S. may be difficult due to FIA protocol. The remeasurement period is much longer in the western U.S. (usually 10 years), which influences the probability of detecting subtle shifts in composition and/or carbon dynamics because of the frequency of large disturbance events in the western U.S. (for example, extensive wildfire; Noss and others 2006; Schoennagel and Nelson 2011). Legacy effects (biogeographical and anthropogenic) vary across the U.S., differentially affecting the natural progression of forest communities through succession and structural complexity (Cuddington 2011; Gough and others 2020). Because succession represents both shifts in community composition (for example, trends toward the sugar maple-basswood association in much of our study area) and changes to structure (for example, trends toward larger trees), it is difficult (if not impossible) to distinguish purely successional and stand development processes from changes in forest communities in response to recent global change (Cuddington 2011).

The community composition and carbon pools we analyzed are also part of a larger socio-ecological system. The fate of forest carbon depends not only on ecological processes but also human and wildlife use. Standing dead trees may fall, and carbon may end up in downed deadwood and eventually soil, and the probability of standing dead trees entering these pools varies among community types (Woodall and Nagel 2006; Oberle and others 2018). Human activities can affect carbon cycling directly through harvest, where carbon is either quickly released (for example, burned) or stored long-term in wood products (for example, building materials), and indirectly by speeding up carbon cycling through nitrogen deposition (Magnani and others 2007; Jo and others 2019; Averill and others 2018). There are also implications beyond carbon cycling such as the provision of wildlife habitat through standing dead trees and coarse woody debris and production of mast and other plant tissue by the seedling and live tree demographic components as food sources for wildlife (Marcot and others 2010). These interactions with wildlife can feedback with the forest community dynamics through herbivory, seed

dispersal, and management for wildlife habitat (Vellend and others 2006).

Conclusions

We found shifts in communities at a temporal scale spanning a relatively short period of ca. 20 years (with five-year remeasurements). Our study aimed to detect changes in the community composition of forest stands in the Great Lakes region and link these changes to shifts in forest carbon using a continuous model of forest community classification in contrast to traditional discrete approaches. Discrepancies between the continuous LDA model and FIA-defined forest type groups suggest that entities such as the FIA Program aiming to estimate carbon attributes by forest type/community could consider a continuous model of forest type/community classification to improve how they partition and report carbon stocks and stock changes across forest communities. Although our analysis tracked a set of six communities across two decades, it is possible that additional novel communities will form and/or current communities will break apart as a result of climate change, further altering the ability of forests to capture and store carbon. Forests are part of an important feedback with the climate system, sequestering carbon to mitigate climate change, while in turn climate change has caused forest tree species to shift their distributions, leading to altered functions and the disruption of preexisting communities (Woodall and others 2009; Fei and others 2017; Knott and others 2020). Commitments to tree planting projects to increase carbon sequestration on forestlands may need to consider additional adaptive management of forests, such as aligning planting efforts with recent or future projected species distributions and novel communities (Bastin and others 2019; Domke and others 2020). The linkages we found between forest community change and carbon dynamics provide insight into the ability of forests to mitigate climate change, especially when faced with climate-driven range shifts and adaptive management efforts that may induce a shift from one community to another.

Acknowledgements

This research was supported by the U.S. Department of Agriculture Forest Service, Northern Research Station, Forest Inventory and Analysis Program. The findings and conclusions in this publication are those of the authors and should not be construed to represent any official U.S. Department of Agriculture or U.S. Government determination or policy. The authors would like to thank the associate editor and three anonymous reviewers for providing valuable feedback on our manuscript.

Data Availability

Forest Inventory and Analysis database used for this project is available online from the U.S. Forest Service FIA Program's DataMart (https://apps.fs.usda.gov/fia/datamart/datamart.html), with selection criteria in supplementary Appendix 1 and the Methods section.

References

An HS, MacFarlane DW. 2013. Comparing a new model-based method to fixed-area sampling for estimating the abundance of standing dead trees. Forestry 86:231–239.

Averill C, Dietze MC, Bhatnagar JM. 2018. Continental-scale nitrogen pollution is shifting forest mycorrhizal associations and soil carbon stocks. Global Change Biology 24:4544–4553.

Bastin J-F, Finegold Y, Garcia C, Mollicone D, Rezende M, Routh D, Zohner CM, Crowther TW. 2019. The global tree restoration potential. Science 365:76–79.

Bates D, Maechler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4.

Bechtold WA, Patterson PL. 2005. The enhanced forest inventory and analysis program–national sampling design and estimation procedures. Southern Research Station: USDA Forest Service.

Blei DM, Ng AY, Jordan MI. 2003. Latent dirichlet allocation. Journal of Machine Learning Research 3:993–1022.

Braun EL. 1950. Deciduous forests of eastern North America. Caldwell, NJ: The Blackburn Press.

Burns RM, Honkala BH. 1990a. Silvics of North America. Volume 2. Hardwoods. Agriculture Handbook 654.

Burns RM, Honkala BH. 1990b. Silvics of North America. Volume 1. Conifers. Agriculture Handbook 654.

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: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. Forest Ecosystems 5:8.

Cuddington K. 2011. Legacy effects: the persistent impact of ecological interactions. Biological Theory 6:203–210.

Domke GM, Oswalt SN, Walters BF, Morin RS. 2020. Tree planting has the potential to increase carbon sequestration capacity of forests in the United States. Proceedings of the National Academy of Sciences 117:24649–24651.

Dyer JM. 2006. Revisiting the deciduous forests of eastern North America. Bioscience 56:341–352.

Eriksson L, Johansson E, Wikström C. 1998. Mixture design—design generation, PLS analysis, and model usage. Chemometrics and Intelligent Laboratory Systems 43:1–24.

Fei S, Desprez JM, Potter KM, Jo I, Knott JA, Oswalt CM. 2017. Divergence of species responses to climate change. Science Advances 3:e1603055.

Fei S, Morin RS, Oswalt CM, Liebhold AM. 2019. Biomass losses resulting from insect and disease invasions in US forests. Proceedings of the National Academy of Sciences 116:17371–17376.

Gough CM, Atkins JW, Fahey RT, Hardimm BS, LaRue EA. 2020. Community and strucutral constraints on the complexity of North American forests. Global Ecology and Biogeography 29:2107–2118.

Gough CM, Vogel CS, Hardiman B, Curtis PS. 2010. Wood net primary production resilience in an unmanaged forest transitioning from early to middle succession. Forest Ecology and Management 260:36–41.

Gundersen P, Thybring EE, Nord-Larsen T, Vesterdal L, Nadelhoffer KJ, Johannsen VK. 2021. Old-growth forest carbon sinks overestimated. Nature 591:E21-23.

Hall D, Jurafsky D, Manning CD. 2008. Studying the history of ideas using topic models. In: Proceedings of the 2008 conference on empirical methods in natural language processing. pp 363–71.

Heitzman E. 2003. Effects of oak decline on species composition in a northern Arkansas forest. Southern Journal of Applied Forestry 27:264–268.

Hoover CM, Smith JE. 2021. Current aboveground live tree carbon stocks and annual net change in forests of conterminous United States. Carbon Balance and Management 2:1–12.

Hornik K, Grün B. 2011. topicmodels: an R package for fitting topic models. Journal of Statistical Software 40:1–30.

Hudon SF, Zaiats A, Roser A, Roopsind A, Barber C, Robb BC, Pendleton BA, Camp MJ, Clark PE, Davidson MM, Frankel-Bricker J, Fremgen-Tarantino M, Forbey JS, Hayden EJ, Richards LA, Rodriguez OK, Caughlin TT. 2021. Unifying community detection across scales from genomes to landscapes. Oikos 130:831–843.

Jo I, Fei S, Oswalt CM, Domke GM, Phillips RP. 2019. Shifts in dominant tree mycorrhizal associations in response to anthropogenic impacts. Science Advances 5:eaav6358.

Keller AB, Brzostek ER, Craig ME, Fisher JB, Phillips RP. 2021. Root-derived inputs are major contributors to soil carbon in temperate forests, but vary by mycorrhizal type. Ecol Lett 24:626–635. https://doi.org/10.1111/ele.13651.

Knott JA, Jenkins MA, Oswalt CM, Fei S. 2020. Community-level responses to climate change in forests of the eastern United States. Global Ecology and Biogeography 29:1299–1314.

LaRue EA, Hardiman BS, Elliott JM, Fei S. 2019. Structural diversity as a predictor of ecosystem function. Environmental Research Letters 14:114011.

LaRue EA, Knott JA, Domke GM, Chen H, Guo Q, Hisano M, Oswalt CM, Oswalt S, Kong N, Potter K, Fei S. 2023. Structural diversity as a novel predictor for ecosystem productivity. Frontiers in Ecology and the Environment. (in press).

Lawson J, Willden C. 2016. Mixture experiments in R using mixexp. Journal of Statistical Software 72.

Lin J. 1991. Divergence measures based on the Shannon entropy. IEEE Transactions on Inferecnce Theory 37:145–151.

Luyssaert S, Schulze ED, Börner A, Knohl A, Hessenmöller D, Law BE, Ciais P, Grace J. 2008. Old-growth forests as global carbon sinks. Nature 455:213–215.

Magnani F, Mencuccini M, Borghetti M, Berbigier P, Berninger F, Delzon S, Grelle A, Hari P, Jarvis PG, Kolari P, Kowalski AS, Lankreijer H, Law BE, Lindroth A, Loustau D, Manca G, Moncrieff JB, Rayment M, Tedeschi V, Valentini R, Grace J. 2007. The human footprint in the carbon cycle of temperate and boreal forests. Nature 447:848–850.

Marcot BG, Ohmann JL, Mellen-McLean KL, Waddell KL. 2010. Synthesis of regional wildlife and vegetation field studies to guide management of standing and down dead trees. Forest Science 56:391–404.

Nikita M. 2016. ldatuning: tuning of the latent dirichlet allocation models parameters.

Niklaus PA, Baruffol M, He JS, Ma K, Schmid B. 2017. Can niche plasticity promote biodiversity–productivity relationships through increased complementarity? Ecology 98:1104–1116.

Noss RF, Franklin JF, Baker WL, Schoennagel T, Moyle PB. 2006. Managing fire-prone forests in the western United States. Frontiers in Ecology and the Environment 4:481–487.

Oberle B, Ogle K, Zanne AE, Woodall CW. 2018. When a tree falls: controls on wood decay predict standing dead tree fall and new risks in changing forests. PLoS One 13:1–22.

Orihuela-Belmonte DE, De Jong BHJ, Mendoza-Vega J, Van der Wal J, Paz-Pellat F, Soto-Pinto L, Flamenco-Sandoval A. 2013. Carbon stocks and accumulation rates in tropical secondary forests at the scale of community, landscape and forest type. Agriculture Ecosystems and Environment 171:72–84.

Pan YD, Birdsey RA, Fang JY, Houghton R, Kauppi PE, Kurz WA, Phillips OL, Shvidenko A, Lewis SL, Canadell JG, Ciais P, Jackson RB, Pacala SW, McGuire AD, Piao SL, Rautiainen A, Sitch S, Hayes D. 2011. A large and persistent carbon sink in the world's forests. Science 333:988–993.

Piao S, Liu Q, Chen A, Janssens IA, Fu Y, Dai J, Liu L, Lian X, Shen M, Zhu X. 2019. Plant phenology and global climate change: current progresses and challenges. Global Change Biology 25:1922–1940.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Richardson AD, Black TA, Ciais P, Delbart N, Friedl MA, Gobron N, Hollinger DY, Kutsch WL, Longdoz B, Luyssaert S, Migliavacca M, Montagnani L, Munger JW, Moors E, Piao S, Rebmann C, Reichstein M, Saigusa N, Tomelleri E, Vargas R, Varlagin A, Heinrich J. 2010. Influence of spring and autumn phenological transitions on forest ecosystem productivity. Transactions of the Royal Society B 365:3227–3246.

Schoennagel T, Nelson CR. 2011. Restoration relevance of recent national fire plan treatments in forests of the western United States. Frontiers in Ecology and the Environment 9:271–277.

Shriver RK, Yackulic CB, Bell DM, Bradford JB. 2021. Quantifying the demographic vulnerabilities of dry woodlands to climate and competition using range-wide monitoring data. Ecology e03425.

Silva LCR, Lambers H. 2021. Soil-plant-atmosphere interactions: structure, function, and predictive scaling for climate change mitigation. Plant Soil 461:5–27.

Small CJ, McCarthy BC. 2003. Spatial and temporal variability of herbaceous vegetation in an eastern deciduous forest. Plant Ecology 164:37–48.

Ullah F, Gilani H, Sanaei A, Hussain K, Ali A. 2021. Stand structure determines aboveground biomass across temperate forest types and species mixture along a local-scale elevational gradient''. Forest Ecology and Management 486:118984.

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

Vellend M, Knight TM, Drake JM. 2006. Antagonistic effects of seed dispersal and herbivory on plant migration. Ecology Letters 9:319–326.

Wilson BT, Woodall CW, Griffith DM. 2013. Imputing forest carbon stock estimates from inventory plots to a nationally continuous coverage. Carbon Balance and Management 8:1.

Woodall CW, Heath LS, Domke GM, Nichols MC. 2011. Methods and equations for estimating aboveground volume, biomass, and carbon for trees in the U.S. forest inventory, 2010. USDA Forest Service General Technical Report NRS-88.

Woodall CW, Macfarlane DW. 2010. The zero inflation of standing dead tree carbon stocks. USDA Forest Service General Technical Report SRS-157.

Woodall CW, Nagel LM. 2006. Coarse woody type: a new method for analyzing coarse woody debris and forest change. Forest Ecology and Management 227:115–121.

Woodall CW, Oswalt CM, Westfall JA, Perry CH, Nelson MD, Finley AO. 2009. An indicator of tree migration in forests of the eastern United States. Forest Ecology and Management 257:1434–1444.

Woodall CW, Walters BF, Coulston JW, D'Amato AW, Domke GM, Russell MB, Sowers PA. 2015. Monitoring network confirms land use change is a substantial component of the forest carbon sink in the eastern United States. Scientific Reports 5:1–9.

Zuur AF, Ieno EN, Walker NJ, Salveliev AA, Smith GM. 2009. Mixed effects models and extensions in ecology with R. 1st ed. (In: Gail M, Krickeberg K, Samet J, Tsiatis A, Wong W, Eds. New York (NY): Springer