Land-use history impacts spatial patterns and composition of woody plant species across a 35-hectare temperate forest plot

David A. Orwig[1], Jason A. Aylward[1], Hannah L. Buckley[2], Bradley S. Case[2] and Aaron M. Ellison[1,3]

[1] Harvard Forest, Harvard University, Petersham, MA, United States of America
[2] School of Science, Auckland University of Technology, Auckland, New Zealand
[3] Sound Solutions for Sustainable Science, Boston, MA, United States of America

ABSTRACT

Land-use history is the template upon which contemporary plant and tree populations establish and interact with one another and exerts a legacy on the structure and dynamics of species assemblages and ecosystems. We use the first census (2010–2014) of a 35-ha forest-dynamics plot at the Harvard Forest in central Massachusetts to describe the composition and structure of the woody plants in this plot, assess their spatial associations within and among the dominant species using univariate and bivariate spatial point-pattern analysis, and examine the interactions between land-use history and ecological processes. The plot includes 108,632 live stems ≥ 1 cm in diameter (2,215 individuals/ha) and 7,595 standing dead stems ≥ 5 cm in diameter. Live tree basal area averaged 42.25 m^2/ha, of which 84% was represented by *Tsuga canadensis* (14.0 m^2/ ha), *Quercus rubra* (northern red oak; 9.6 m2/ ha), *Acer rubrum* (7.2 m^2/ ha) and *Pinus strobus* (eastern white pine; 4.4 m^2/ ha). These same four species also comprised 78% of the live aboveground biomass, which averaged 245.2 Mg/ ha. Across all species and size classes, the forest contains a preponderance (> 80,000) of small stems (<10-cm diameter) that exhibit a reverse-J size distribution. Significant spatial clustering of abundant overstory species was observed at all spatial scales examined. Spatial distributions of *A. rubrum* and *Q. rubra* showed negative intraspecific correlations in diameters up to at least a 150-m spatial lag, likely indicative of crowding effects in dense forest patches following intensive past land use. Bivariate marked point-pattern analysis, showed that *T. canadensis* and *Q. rubra* diameters were negatively associated with one another, indicating resource competition for light. Distribution and abundance of the common overstory species are predicted best by soil type, tree neighborhood effects, and two aspects of land-use history: when fields were abandoned in the late 19th century and the succeeding forest types recorded in 1908. In contrast, a history of intensive logging prior to 1950 and a damaging hurricane in 1938 appear to have had little effect on the distribution and abundance of present-day tree species. Our findings suggest that current day composition and structure are still being influenced by anthropogenic disturbances that occurred over a century ago.

Subjects Ecology, Plant Science, Natural Resource Management, Forestry

Keywords ForestGEO, Harvard Forest, Land-use history, Spatial point-pattern analysis, Temperate forest, *Tsuga canadensis*

Submitted 8 September 2021
Accepted 6 December 2021
Published 3 January 2022

Corresponding author
David A. Orwig,
orwig@fas.harvard.edu

Academic editor
Xugao Wang

Additional Information and Declarations can be found on page 25

DOI **10.7717/peerj.12693**

How to cite this article Orwig DA, Aylward JA, Buckley HL, Case BS, Ellison AM. 2022. Land-use history impacts spatial patterns and composition of woody plant species across a 35-hectare temperate forest plot. *PeerJ* **10**:e12693 http://doi.org/10.7717/peerj.12693

INTRODUCTION

In forested landscapes around the world, legacies of human activities have shaped the composition, size structure, and spatial patterns of trees, understory vegetation, and associated ecosystem processes (*Birks et al., 1988*; *Turner et al., 1990*; *Russell, 1997*; *Foster & Aber, 2004*; *Ellison et al., 2014*). The extent of the interactions between anthropogenic effects and abiotic factors such as climate, soils, and episodic disturbances in shaping vegetation patterns depends on the intensity of the effects and the spatial scale of analysis (*Rackham, 1986*; *Glitzenstein et al., 1990*; *Zimmerman et al., 1995*). A complex interplay of succession, competition, disturbance, environment, and land use shape dynamics and patterns of forests at local-to-regional scales (*Condit et al., 2000*; *Thompson et al., 2002*; *Chazdon, 2003*; *Van Gemerden et al., 2003*).

By further examining the spatial patterns of trees within a forest, ecologists can begin to uncover the underlying processes and mechanisms that led to those patterns (*e.g.*, are species randomly distributed, aggregated, or dispersed in space and why? (*Wiegand, He & Hubbell, 2013*)). A growing number of studies have used point pattern analysis to examine the spatial structure of forests by using fully mapped plots, as each tree, or point, has a mapped location (*Zhang et al., 2010*; *Wang et al., 2011*; *Lutz et al., 2013*; *Fibich et al., 2016*; *Nguyen, Uria-Diez & Wiegand, 2016*). A variety of univariate and bivariate point-pattern analysis methods and summary characteristics have been used to characterize the spatial patterning of trees (*Wiegand & Moloney, 2004*; *Wiegand & Moloney, 2014*). Since each method tells something different about the spatial structure of the data within a forest, it is more desirable to use multiple summary characteristics to better describe the patterns of tree species and among species associations (*Illian et al., 2008*; *Wiegand, He & Hubbell, 2013*).

The forests of New England, USA have been shaped by a variety of natural and anthropogenic factors. As in other forests, the geology and climate of New England define the broad patterns of current forest composition (*Foster et al., 1992*; *Hall et al., 2002*), but the shifts in species abundance and distribution patterns that have occurred since Europeans colonized New England more than 400 years ago have resulted in a relatively homogenous assemblage of young, even-aged stands with fewer late-successional species (*Thompson et al., 2013*). In Massachusetts, modern vegetation exhibits only weak relationships to broad climatic gradients because of the overwhelming influence of past land use (*Foster, Motzkin & Slater, 1998*). An increasing emphasis in ecological studies is evaluating the relative importance of historic land-clearing, agriculture, intensive harvesting (*Foster, 1992*; *Thompson et al., 2002*; *Rhemtulla, Mladenoff & Clayton, 2009*; *Hogan et al., 2016*), and natural episodic storms (*Foster & Boose, 1992*; *Zimmerman et al., 1995*) on current-day structure and species composition of forest stands (*Motzkin et al., 1996*; *Motzkin et al., 1999*; *Bürgi, Östlund & Mladenoff, 2017*). Although evaluating which of these variables are most important in shaping current day structure and composition is challenging, the development and use of statistical approaches like recursive partitioning and conditional inference trees has aided the interpretation and prediction of these types of analyses (*Hothorn, Hornik & Zeileis, 2006*).

Harvard Forest is an ideal location to investigate how spatial patterns and composition of woody plants are influenced by land-use history. For more than a century, Harvard Forest (HF) researchers have investigated and recorded impacts of land use on forests and how New England's forests are continuing to change as the regional climate changes, populations of large herbivores wax and wane, and nonnative insects and pathogens establish, irrupt, and kill tree species (*Foster & Aber, 2004*).

Here, we describe the results of the first census of a fully mapped 35-ha forest-dynamics plot at the Harvard Forest and examine how its structure and composition relates to interactions between land-use history and ecological processes. We first describe the composition and structure of the woody plants in this plot and assess spatial associations within and among the dominant species using univariate and bivariate spatial point-pattern analysis. Second, we uncover the influence of historical land use and natural disturbances on the current-day structure and composition of this forest plot. We pay particular attention to patterns of distribution and abundance of *Tsuga canadensis* (eastern hemlock) and its relationship to other species in the plot because previous work has shown it to be a foundation species in this forest (*e.g.*, a species that defines ecosystems, controlling the biological diversity of associated species and modulating critical ecosystem processes; *sensu* (*Ellison, 2019*)). *Tsuga canadensis* is currently threatened and declining throughout much of the HF plot and its range due to a nonnative insect, *Adelges tsugae* (hemlock woolly adelgid; HWA) and its decline and loss are likely to have profound impacts on forest structure and composition (*Orwig et al., 2013*; *Foster, 2014*).

MATERIALS & METHODS

Site description

The 35-ha (500 × 700 m) HF forest-dynamics plot is part of a global network of Forest Global Earth Observatory (ForestGEO) plots established to monitor, understand, and predict forest dynamics and responses to global change (*Anderson-Teixeira et al., 2015*). The HF ForestGEO plot (southwest corner at 42.5386°N, 72.1798°W) is located within the 380-ha HF Prospect Hill tract in Petersham, Massachusetts, USA within the Worcester/Monadnock Plateau ecoregion (*Griffith, Omernik & Kiilsgaard, 1994*) of Transition Hardwoods-White Pine-Hemlock forests (*Westveld, 1956*) (Fig. 1). Elevations in the plot range from 340.2 to 367.8 m a.s.l. Soils include Gloucester stony loam, Acton stony loam and Whitman very stony silt loams, all of which are gravelly and fine sandy loam soils that developed in glacial tills overlying gneiss and schist bedrock (*Simmons, 1941*). The north-central portion of the plot contains a 3-ha peat swamp with muck soils that has been colonized at intervals by *Castor canadensis* (beaver). Average (1964-2019) annual temperature at the site is 7.9 °C and the annual precipitation of 1090 mm is distributed evenly throughout the year (*Boose & Gould, 2019*).

Land-use history

We examined the influence of past land-use history (derived from forest stand descriptions of dates of field abandonment, areas used as woodlot, pasture, or cultivation; presence of distinct plow horizon; silviculture treatments; and salvage operations); historical events

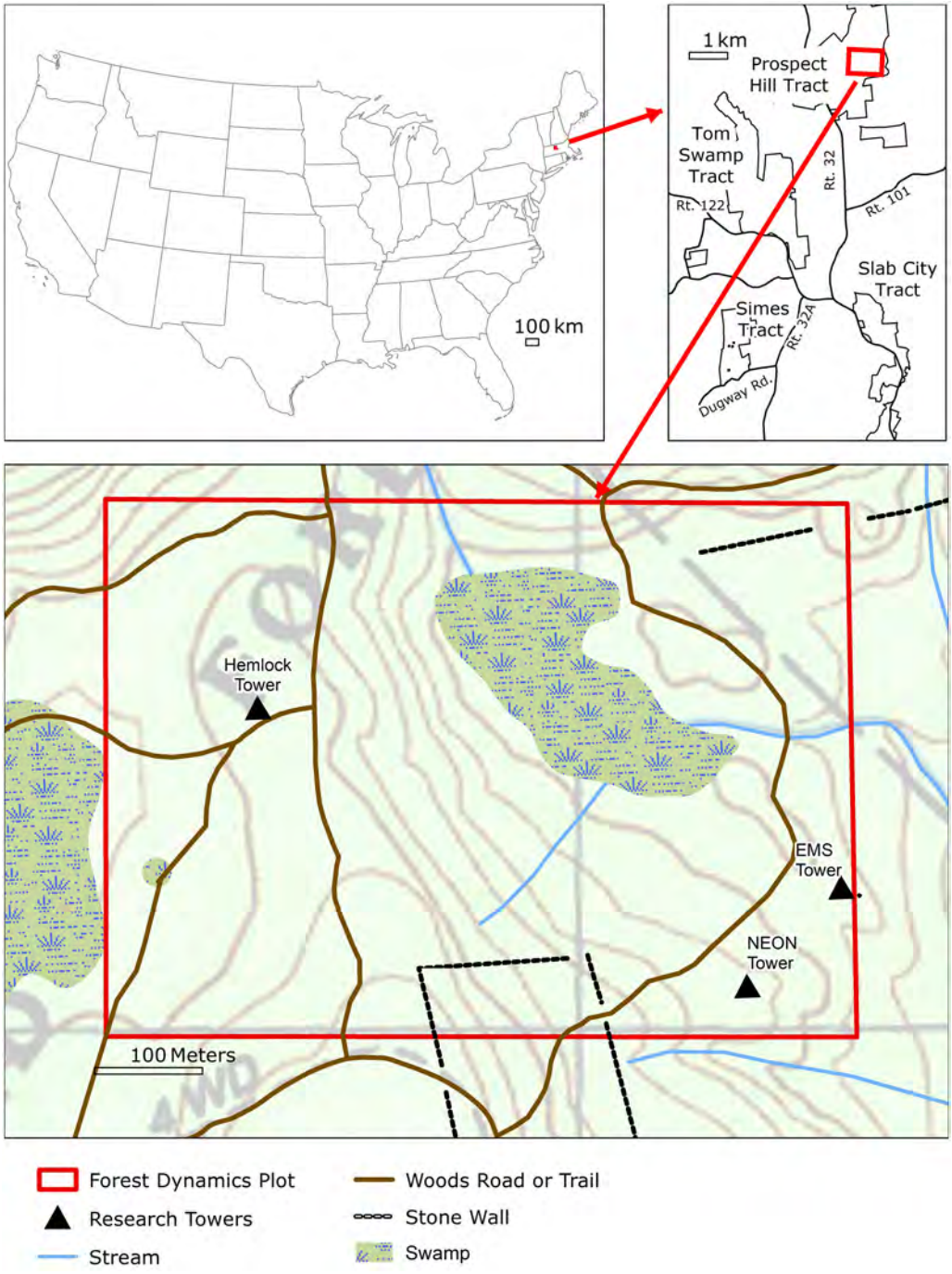

Figure 1 The 500 × 700 m ForestGEO plot located in the town of Petersham, MA on the Prospect Hill tract of HF (upper right panel). Locations of three eddy-flux towers (that measure net ecosystem exchange of carbon and water between the atmosphere and the ecosystem), old forest roads, stone walls (denoted by dotted lines), and the central swamp area are superimposed on topographic contour lines (lower panel).

Full-size DOI: 10.7717/peerj.12693/fig-1

(*e.g.*, insect outbreaks, storms and associated degree of forest damage *Rowlands, 1941*); and biophysical attributes (roads, soil type, slope, aspect, elevation, and distance to streams) on current forest composition and species distribution within the plot by using data from the document archives at HF (http://harvardforest.fas.harvard.edu/document-archive). Original maps of activity were manually transcribed to standardized base maps and then scanned and digitized as shapefiles in ArcView GIS 3.2. The shapefiles were then transformed to Massachusetts State Plane Meters (NAD83 projection) in ArcGIS to align better with aerial photographs and linear features (trails, stonewalls, *etc.*) downloaded from MassGIS (*Hall, 2005*) and used in spatial analyses (see below).

Pollen evidence suggests that prior to European settlement, Prospect Hill was a mixture of old-growth northern hardwoods, *T. canadensis*, and *Pinus strobus* (eastern white pine). Following European arrival, the site then experienced complex ownership and intensive land-use over the next few centuries, both of which are largely representative of the New England region (*Ellison et al., 2014*). Forest clearing began in 1750 and reached a maximum in the 1840s, by which time close to 80% of the original forests had been cleared for agriculture (*Fisher, 1933*; *Raup & Carlson, 1941*). Field abandonment began in 1850 and continued through 1905 in the southern half of the plot (Fig. 2A). Reforestation of those fields continued through the 20th century (*Foster, 1992*). The western, northern, and northeastern areas of the plot remained permanently wooded, but experienced various types of selective cutting in the 1790s and 1870s (*Foster, 1992*). The first maps characterizing forest types of individual stands were completed in 1908 and classified the permanent woodlots in the western third of the plot as being comprised of hardwoods, white pine-hardwoods, hemlock, and red maple (Fig. 2B). Many *Castanea dentata* (American chestnut) died in 1912–1914 from infection by *Endothia parasitica* (chestnut blight) (*McLachlan, Foster & Menalled, 2000*) and forests were damaged by natural disturbances including an ice storm in 1921 and one of the most damaging hurricanes to hit New England in 1938. The hurricane and subsequent salvage logging resulted in the loss of as much as 70% of the standing timber on HF properties (*Foster & Boose, 1992*).

The central sections of the plot, containing mostly stony loam soils and no visible signs of a plow layer, were unimproved pastures abandoned in the mid-19th century (*Motzkin et al., 1999*) (Fig. 2C). These areas reforested and were classified as cordwood (poor hardwood) in 1908 (Fig. 2B), except for an area classified as open, which is the beaver swamp. Much of the cordwood section was subsequently clear-cut in the 1920s and then thinned or salvaged in the late 1940s following the 1938 hurricane. *Pinus resinosa* (red pine) and *Picea abies* (Norway spruce) plantations were established in portions of these abandoned pastures in the mid-1920s and early 1930s. The southcentral area of the plot contained areas of improved pasture and cultivation and was classified as containing white pine in 1908. This area was clear-cut in the 1920s and a portion of it was clear-cut again in 1980, resulting in many small diameter, multi-stemmed trees. Additional biotic changes that impacted the plot included the exotic *Lymantria dispar* (gypsy moth), which lead to widespread defoliation of hardwoods during 1944–45 and 1981; *Cryptococcus fagisuga* (beech scale insect) combined with *Neonectria* fungal spp. (beech-bark disease), which has led to the decline and death of larger *Fagus grandifolia* (American beech); and *Adelges*

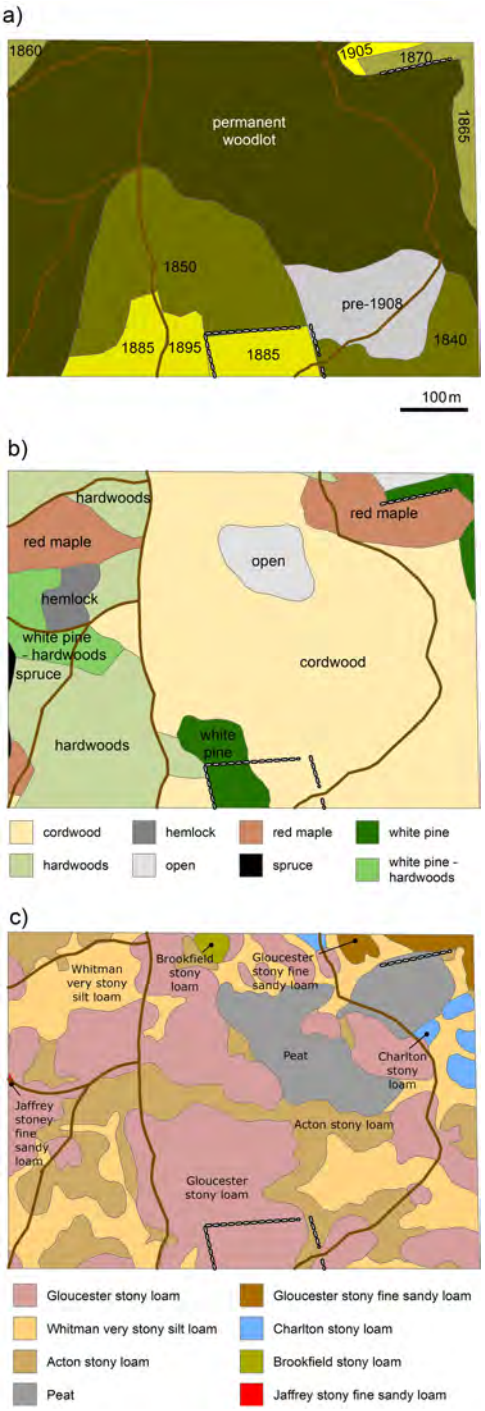

Figure 2 Historical GIS layers of the HF ForestGEO plot. Location of (A) historical fields and their agricultural date of abandonment, (B) forest stands as described in 1908, and (C) soil type. GIS layers obtained from Harvard Forest Data Archive HF 110.

Full-size DOI: 10.7717/peerj.12693/fig-2

tsugae, which was first observed in the plot in 2008 and then rapidly spread throughout the plot, subsequently killing hundreds of *T. canadensis* stems and threatening the rest.

Plot establishment and woody stem census

During March 2010, professional surveyors delineated the plot boundaries, established a continuous grid of 20 × 20-m quadrats, and measured the elevation at each post using a Sokkia SET600 Total Station (Olathe, Kansas, USA). During the summers of 2010 and 2011, all woody stems ≥ 1 cm in diameter at breast height (DBH; 1.3 m above the ground level) were uniquely tagged, identified (nomenclature follows *Haines, 2011*), and measured to the nearest 0.1 cm DBH (*Condit, 1998*). All dead stems ≥ 5 cm diameter that were standing and >45 degrees from horizontal also were tagged, identified, and measured. The swamp located in the center of the plot was sampled when the ground was frozen during the winter months of 2012–2014. Each tagged stem was mapped within one of four 10 × 10 m subquadrats within each quadrat on a scale-drawn map data sheet. Each map was then scanned and individual stems were digitized using the ImageJ processing program (*Rasband 2012*), and converted to local (x, y) coordinates within a quadrat using R (v.3.6.1) (*R Core Team, 2013*) and the CTFS R package (*Condit, 2014*).

Forest species composition and stand structure

Estimates of stem densities were derived from total counts in which multi-stemmed individuals were considered as a single stem, whereas estimates of basal area and biomass were derived from the sum of all stems ≥ 1 cm DBH (*Gilbert et al., 2010*). Biomass of living woody stems was estimated from DBH using allometric equations (Table S1).

Spatial analysis

We assessed the spatial patterns of all stems of the seven most abundant overstory tree species across the entire plot using the pair-correlation function ($g(r)$; *Wiegand & Moloney, 2014*), for which the value of the function represents the degree of clustering ($g(r) > 1$) or overdispersion ($g(r) < 1$) at a given spatial lag (distance between neighboring trees). We compared the observed pair-correlation statistic to that expected if trees were distributed randomly ($g(r) = 1$) within the plot using 199 Monte Carlo CSR (complete spatial randomness) simulations of the tree map for each species.

To test for the effects of intraspecific competition we used the univariate mark-correlation function ($kmm(r)$; *Wiegand & Moloney, 2004*; *Wiegand & Moloney, 2014*) to test whether the size (DBH) of each of the seven most abundant tree species depended on its proximity to neighbors of its own species. The value of $kmm(r)$ represents the relative sizes of trees at a given spatial lag and indicates if trees are larger or smaller than expected at a given spatial lag. We compared the observed univariate mark-correlation function statistic to that expected if the sizes of trees were randomly assigned across individuals using 199 Monte Carlo simulations for each species, *i.e.,* the spatial pattern of the trees remained the same, but their sizes were shuffled (*Jacquemyn et al., 2010*). Spatial analyses were not conducted on shrub species as many only occurred in the central swamp area.

Prior work has shown that the shade-tolerant *T. canadensis* is an important foundation tree species, creating and strongly controlling the microenvironment, understory

vegetation, and ecosystem dynamics (*Ellison et al., 2005*; *Orwig et al., 2013*). Thus, we assessed the potential influence of *T. canadensis* on the sizes of each of the other most common tree species in the plot using a bivariate marked point-pattern analysis (Schlather's version of Moran's I mark-correlation function ($Im1m2(r)$; *Wiegand & Moloney, 2014*). This statistic determines if tree sizes are spatially correlated: individuals are smaller or larger than expected at various distances from a neighbor. We compared the observed $Im1m2(r)$ to that expected if the sizes of trees were randomly assigned across individuals using 199 Monte Carlo simulations for each species (*Jacquemyn et al., 2010*). All spatial pattern analyses were performed using the 2018 version of the software Programita (*Wiegand & Moloney, 2004*; *Wiegand & Moloney, 2014*).

GIS overlays of past land use, historical events, and biophysical attributes were used as covariates in a conditional inference regression-tree model to predict diameter and abundance of the most common overstory species in the plot (Table 1). Species-specific abundances or sizes were predicted for each of the seven most abundant overstory species conditional on their observed locations. A mean relative abundance (stems/ha) associated with each tree location was calculated using raster-based tools within the GIS. First, a 3×3 cell moving focal window analysis was used to generate a surface of mean tree abundances across the plot at a 20-m cell resolution. Subsequently, to associate a mean relative abundance value with each tree, this generated surface was sampled in the GIS at the location of each tree. Using the 'cforest' function in the R package 'party' (Version 1.3-5) (*Hothorn, Hornik & Zeileis, 2006*), the outcomes of 500 conditional inference tree models (*Hothorn, Hornik & Zeileis, 2006*) were compiled and the relative importance of explanatory variables were ranked across all models. The conditional inference algorithm is based on a random forest machine-learning algorithm (*Breiman, 2001*) used in many ecological modeling contexts (*e.g.*, *Fox et al., 2017*; *Mi et al., 2017*; *Mohapatra et al., 2019*; *Shearman et al., 2019*). The conditional inference method improves on the variable ranking methodology by applying a permutation importance algorithm that corrects for variable selection bias resulting from a mix of categorical and continuous explanatory variables that are correlated to varying degrees or that have complex interactions (*Strobl, Boulesteix & Hothorn, 2007*). Variable importance scores are calculated by determining the marginal loss of prediction accuracy from any given model iteration after removing each explanatory variable. Overall variable importance is determined by averaging the variable-wise decrease in accuracy scores over all 500 model iterations to rank the overall importance of each variable across all models.

Data availability

Data associated with this study are publicly available from the Environmental Data Initiative (*Orwig, Foster & Ellison, 2015*: https://doi.org/10.6073/pasta/18c01a2bb5f5bdf98846542ebbdbad65).

RESULTS

Composition and stand structure

Within the 35-ha HF ForestGEO plot, we identified 108,632 live stems ≥ 1 cm DBH, representing 77,536 individuals (2215 ha^{-1}) of 51 woody species in 17 families (Table S2).

Table 1 Description of land-use history, disturbance, stand, and biophysical predictor variables converted to GIS shapefiles and used to predict current tree species abundance and DBH values across the Harvard Forest ForestGEO plot.

Predictor	Description
Land use	
Stand type –1908, 1947, 1986	Early forest stand descriptions in plot recorded by forest type and year
Allen Land Use	Land-use descriptions derived from degree of soil disturbance, including plow (Ap) horizon presence and depth, recorded by previous HF soil scientist, Arthur Allen.
Field Abandonment	Years since the date of field abandonment
20th C. Salvage cutting	Areas that experienced cutting following wind damage or other natural disturbance in the early to mid-1900s
20th C. intensive cutting	Areas that experienced clearcut, shelterwood or reproduction cuts during the early to mid-1900s
Natural disturbance	
Hurricane damage	Data collected between 1939-1941 on degree of overstory trees uprooted, leaning or broken following 1938 hurricane (*Rowlands, 1941*).
Stand features	
Mean DBH of trees within 10 m	Mean DBH of trees within 10 m of individual tree stem
CV DBH of trees within 10 m	Coefficient of variation of DBH of trees within 10 m
Number of trees within 10 m	Number of trees within 10 m of individual tree stem
Mean distance to trees within 10 m	Mean distance to trees within 10 m of individual tree stem
CV distance to trees within 10 m	Coefficient of variation in distance to trees within 10 m of individual tree stem
Biophysical features	
Elevation	Elevation of quadrat as determined from NASA Goddard's Lidar, hyperspectral and thermal (G-LiHT) airborne imager.
Distance to streams (m)	Distance from individual tree stem to streams as identified by the National Hydrography Dataset
Soil drainage class	USDA Natural Resources Conservation Service Soil Survey Geographic (SSURGO) database soil attribute
Simmons soil type	Soil Classification from 1:24000 scale surveys (*Simmons, 1941*)

Common families were Betulaceae, Rosaceae, and Pinaceae (six species each), and Fagaceae and Adoxaceae (five species each). Four tree species (*T. canadensis*, *Acer rubrum* [red maple], *Q. rubra* [northern red oak], and *P. strobus*) and one shrub, *Ilex verticillata* (winterberry), accounted for 63% of all stems (Table 2). Live tree basal area averaged 42.25 m^2 ha^{-1} and average live aboveground biomass was 245.2 Mg ha^{-1}. Eighty-four percent of the basal area and 78% of the biomass was represented by *T. canadensis* (14.0 m^2 ha^{-1}; 61.1 Mg ha^{-1}), *Q. rubra* (9.6 m^2 ha^{-1}; 75.1 Mg ha^{-1}), *A. rubrum* (7.2 m^2 ha^{-1}; 33.8 Mg ha^{-1}) and *P. strobus* (4.4 m^2 ha^{-1}; 20.7 Mg ha^{-1}). The live tree diameter distributions of *T. canadensis* and *F. grandifolia* were strongly right-skewed (reverse-J shaped), whereas

Table 2 List of total live woody plant density, basal area, and biomass within the 35 ha HF ForestGEO plot in 2014.

Scientific name	Total live tree Density (35 ha^{-1})	Total live Basal area (m^2)	Total live Biomass (Mg)
Acer pensylvanicum	339	0.59	1.13
Acer rubrum	9,723	253.54	1182.86
Acer saccharum	1	3.12e−03	0.02
Alnus incana	479	0.68	0.60
Amelanchier laevis	572	0.35	0.61
Aronia melanocarpa	413	0.07	0.10
Betula alleghaniensis	4,059	36.96	207.73
Betula lenta	1,430	21.14	124.04
Betula papyrifera	537	14.80	72.76
Betula populifolia	108	1.49	7.18
Castanea dentata	732	1.12	4.35
Crataegus spp.	180	0.14	0.27
Fagus grandifolia	3,802	20.93	138.58
Frangula alnus	3	7.42e−04	4.90e−04
Fraxinus americana	186	3.84	23.73
Fraxinus nigra	34	0.17	0.82
Hamamelis virginiana	1,931	3.10	5.77
Ilex laevigata	2	1.39e−03	2.76e−03
Ilex mucronata	598	0.64	0.58
Ilex verticillata	9,874	3.62	6.15
Juniperus communis	1	4.52e−04	4.20e−04
Kalmia latifolia	3,914	3.27	7.64
Lindera benzoin	66	0.02	0.04
Lyonia ligustrina	1,178	0.41	2.04
Nyssa sylvatica	180	2.63	11.25
Ostrya virginiana	24	0.06	0.19
Picea abies	900	24.43	93.11
Picea rubens	101	3.61	15.15
Pinus resinosa	790	67.23	330.28
Pinus strobus	2,126	155.68	724.64
Populus grandidentata	2	0.03	0.14
Populus tremuloides	1	0.01	0.05
Prunus pensylvanica	11	0.05	0.98
Prunus serotina	250	5.48	34.85
Quercus alba	38	1.89	14.53
Quercus rubra	3,896	334.99	2,627.07
Quercus velutina	206	19.28	164.46
Rhododendron prinophyllum	127	0.05	0.25
Salix spp.	2	1.59e−04	1.50e−03

(*continued on next page*)

Table 2 (*continued*)

Scientific name	Total live tree Density (35 ha^{-1})	Total live Basal area (m^2)	Total live Biomass (Mg)
Sambucus racemosa	2	5.65e−04	4.03e−03
Sorbus americana	66	0.26	2.78
Toxicodendron radicans	1	1.13e−04	1.05e−04
Toxicodendron vernix	521	0.32	0.38
Tsuga canadensis	22,880	491.07	2138.00
Ulmus americana	1	2.84e−04	3.85e−04
Vaccinium corymbosum	3,531	2.39	9.58
Viburnum acerifolium	39	0.01	0.07
Viburnum dentatum	325	0.08	0.52
Viburnum lantanoides	75	0.01	0.01
Viburnum nudum	1,182	0.44	2.27

those of *A. rubrum*, *Q. rubra*, *P. strobus*, *Betula lenta* (black birch), and *B. alleghaniensis* (yellow birch) were less right-skewed (Fig. 3).

In contrast, 73% of tagged stems and 69% of live individuals within the plot were <10-cm DBH (Fig. 4). These same stems comprised only 5% of the total live plot basal area and 3% of the total live plot biomass (Table 2). Shrub species made up many of these stems with reverse-J size distributions and included *I. verticillata*, *Vaccinium corymbosum* (highbush blueberry), and *Kalmia latifolia* (mountain laurel). Nonnative species in the plot included 1687 stems of *Picea abies* (Norway spruce) and *Pinus resinosa* (red pine) that remained from early 20th-century conifer plantings and three stems of *Frangula alnus* (glossy false buckthorn). Ten species had only one or two stems within the plot (Table 2). Finally, there were 7595 dead stems $\geq$ 5 cm DBH within the plot, >50% of which were *T. canadensis*, *P. strobus*, or *A. rubrum*. Dead tree basal area was 4.18 m^2 ha^{-1} and dead aboveground biomass was 17.53 Mg ha^{-1}.

Spatial structure related to past land-use impacts

The spatial distributions of the seven most common species varied across the plot (Fig. 5). *Pinus strobus* was common throughout the plot. *Tsuga canadensis* was most abundant in the western, northern, and eastern portions of the plot, whereas *Q. rubra* and *A. rubrum* dominated the central and southern areas. Both *Betula* species were most abundant in the central and eastern sections, and *F. grandifolia* was most common in the southeastern section.

Shrubs were also often spatially aggregated with respect to hydrology and topography. *Ilex verticillata V. corymbosum*, *Viburnum nudum* (withe-rod), and *Lyonia ligustrina* (maleberry) dominated the poorly drained beaver swamp (Fig. 6). *Hamamelis virginiana* (witch-hazel) was found in a narrow elevational band (342–346 m) surrounding the swamp and a dense patch of *K. latifolia* was located in the northwest corner of the plot.

The seven most abundant canopy tree species were significantly clustered in the plot at all spatial lags up to 50 m relative to a CSR null expectation (Fig. 7). The effect of intraspecific competition also was apparent for these seven species as smaller than expected

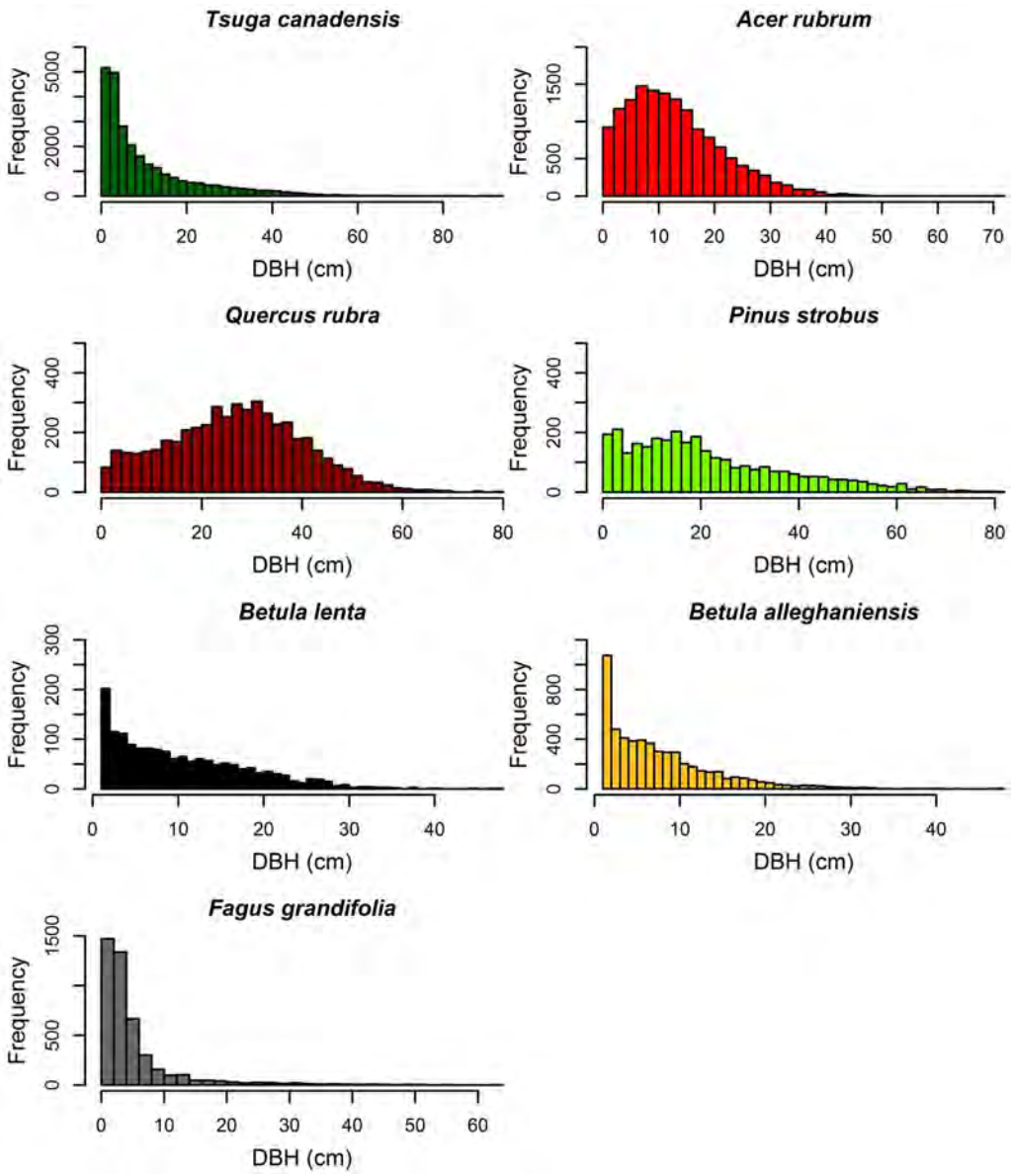

Figure 3 **Diameter distribution of the seven most common overstory species within the HF Forest-GEO plot.**

Full-size DOI: 10.7717/peerj.12693/fig-3

diameters were observed when nearby individuals of the same species. For example, spatial distributions of *A. rubrum*, *Q. rubra*, and *F. grandifolia* showed negative intraspecific correlations in diameters up to at least a 150-m spatial lag, whereas the other species had intraspecific negative correlations at ≤ 50-m spatial lags (Fig. 8). *Tsuga canadensis*, *B. alleghaniensis*, and *P. strobus* had positive spatial correlations (larger diameters than expected among DBHs at spatial lags >150 m. Interspecific correlations in diameters between species suggest that the impact of *T. canadensis* on *Q. rubra* was negative at

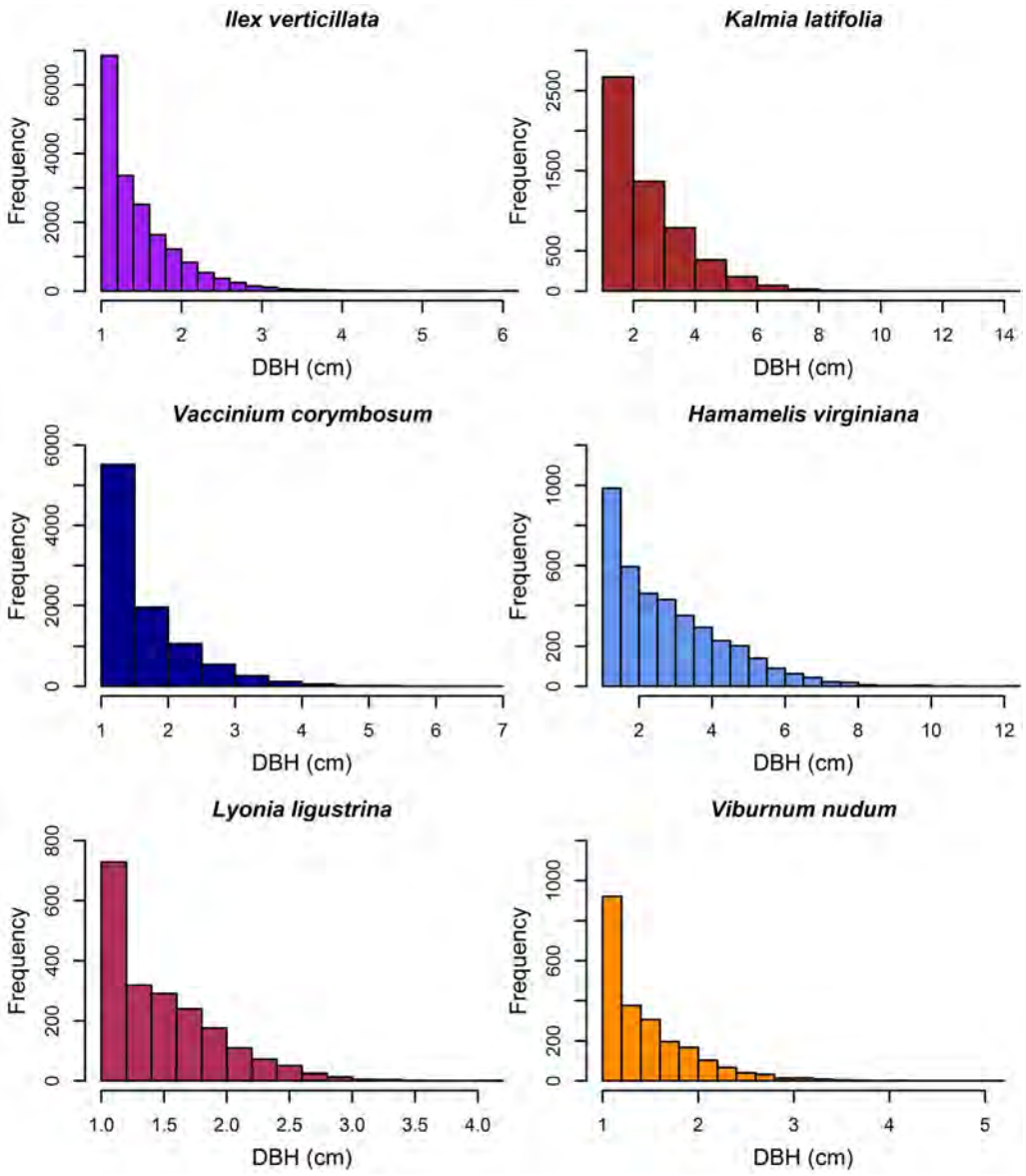

Figure 4 **Diameter distribution of the six most common understory species within the HF ForestGEO plot.**

Full-size DOI: 10.7717/peerj.12693/fig-4

intermediate spatial lags (25–75 m) but positive between *T. canadensis* and the other five species at most spatial scales up to 150 m (Fig. 9).

The abundances and sizes of the most common overstory species were predicted best by a variety of historical factors and competitive interactions. Conditional inference random-forest modeling revealed that the abundances of *T. canadensis*, *P. strobus, Q. rubra, A. rubrum* and *F. grandifolia* were strongly associated with neighborhood effects (size of neighboring trees within 10 m; Fig. 10). The date of field abandonment was a strong predictor of *Q. rubra*, *P. strobus*, and *B. lenta* abundance, whereas the forest type in 1908

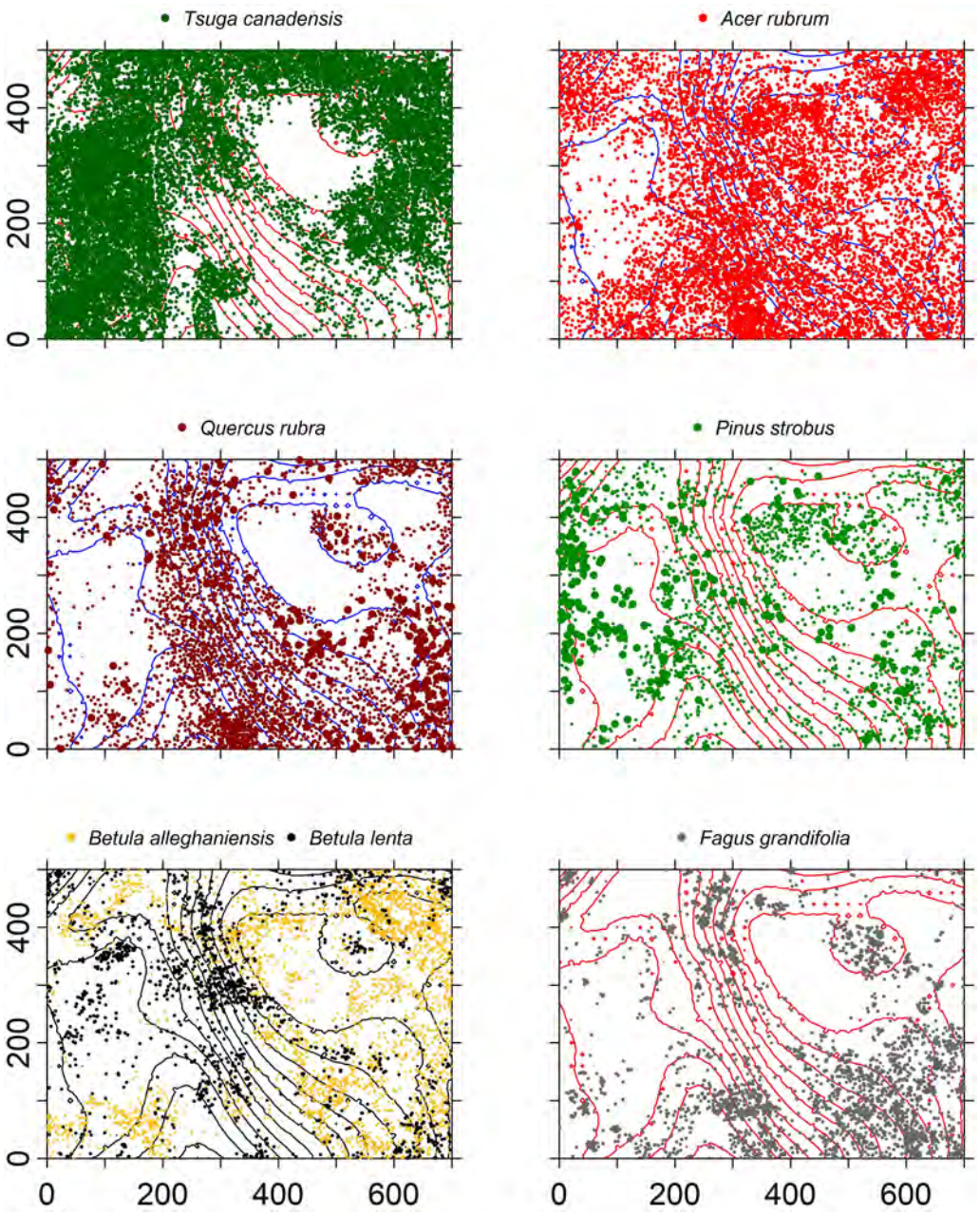

Figure 5 **Spatial distribution of stems ≥1 cm DBH of the seven most common overstory species within the HF ForestGEO plot with 3-m elevation contour lines.**

Full-size DOI: 10.7717/peerj.12693/fig-5

was the best predictor of *B. alleghaniensis* and *A. rubrum* abundance. *Betula* species also were strongly associated with Simmons soil type. Overstory species diameters were best predicted by neighborhood effects for *T. canadensis*, *B. lenta*, and *F. grandifolia*; date of field abandonment for *P. strobus* and *B. alleghaniensis*; and the 1947 stand type for *Q. rubra* and *A. rubrum* (Fig. 11). The predictive power of the conditional inference forest model

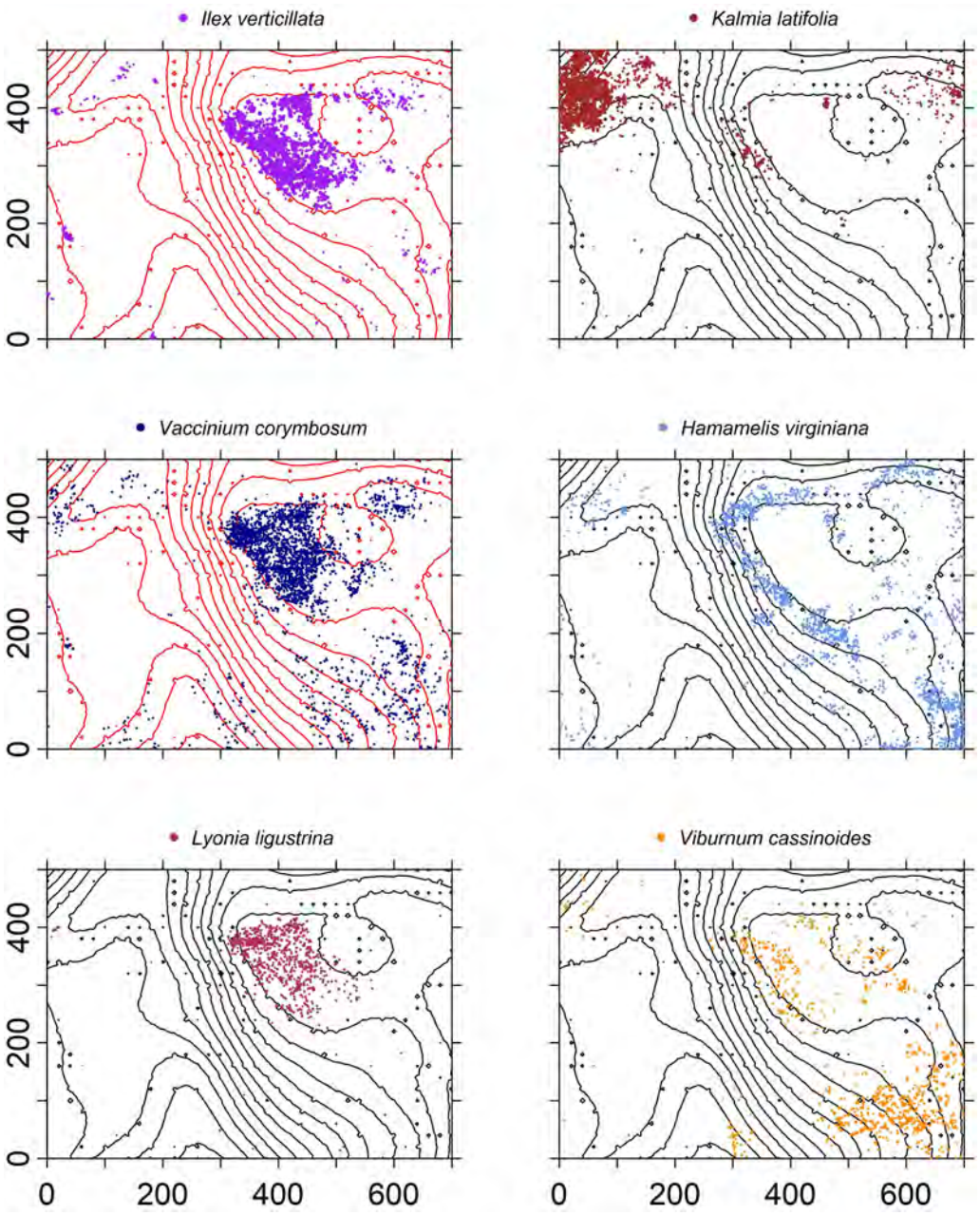

Figure 6 **Spatial distribution of stems ≥1 cm DBH of the six most common understory species within the HF ForestGEO plot with 3 -m elevation contour lines.**

Full-size DOI: 10.7717/peerj.12693/fig-6

regressions was much higher ($R^2 = 0.79$–0.95) for species abundance in the plot compared to species size ($R^2 = 0.11$–0.53).

DISCUSSION

We censused all woody stems ≥1-cm DBH within a 35-ha forest-dynamics plot in north-central Massachusetts to examine the spatial patterns of trees and shrubs at a scale rarely

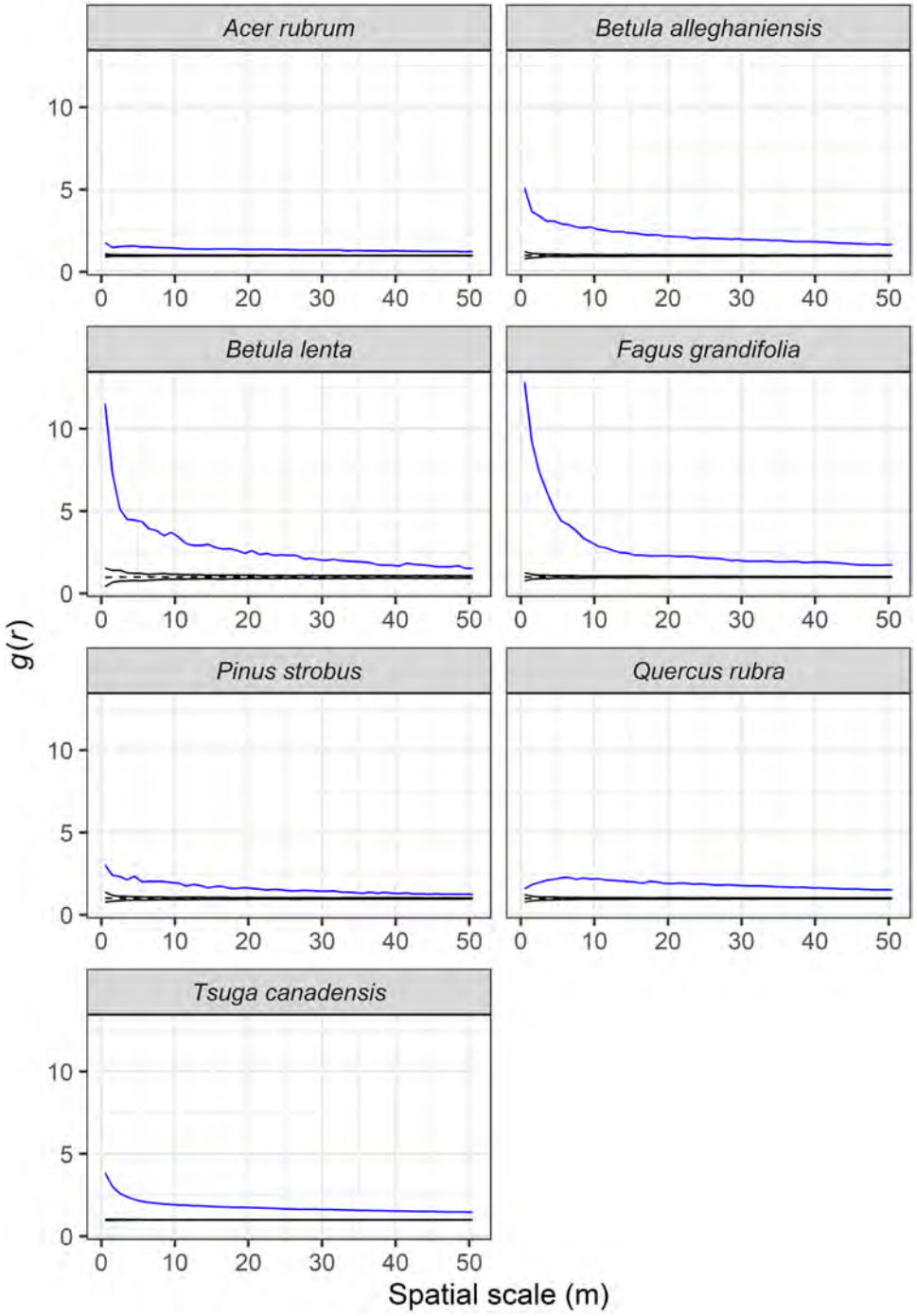

Figure 7 Observed (blue line) and expected (black dashed line) values of the pair correlation function, $g(r)$, showing the degree of spatial clustering (values > 1) of the seven most dominant tree species in the Harvard Forest plot. Expected values were obtained from 199 Monte Carlo simulations to completely randomize the spatial position of trees (complete spatial randomness; CSR).

Full-size DOI: 10.7717/peerj.12693/fig-7

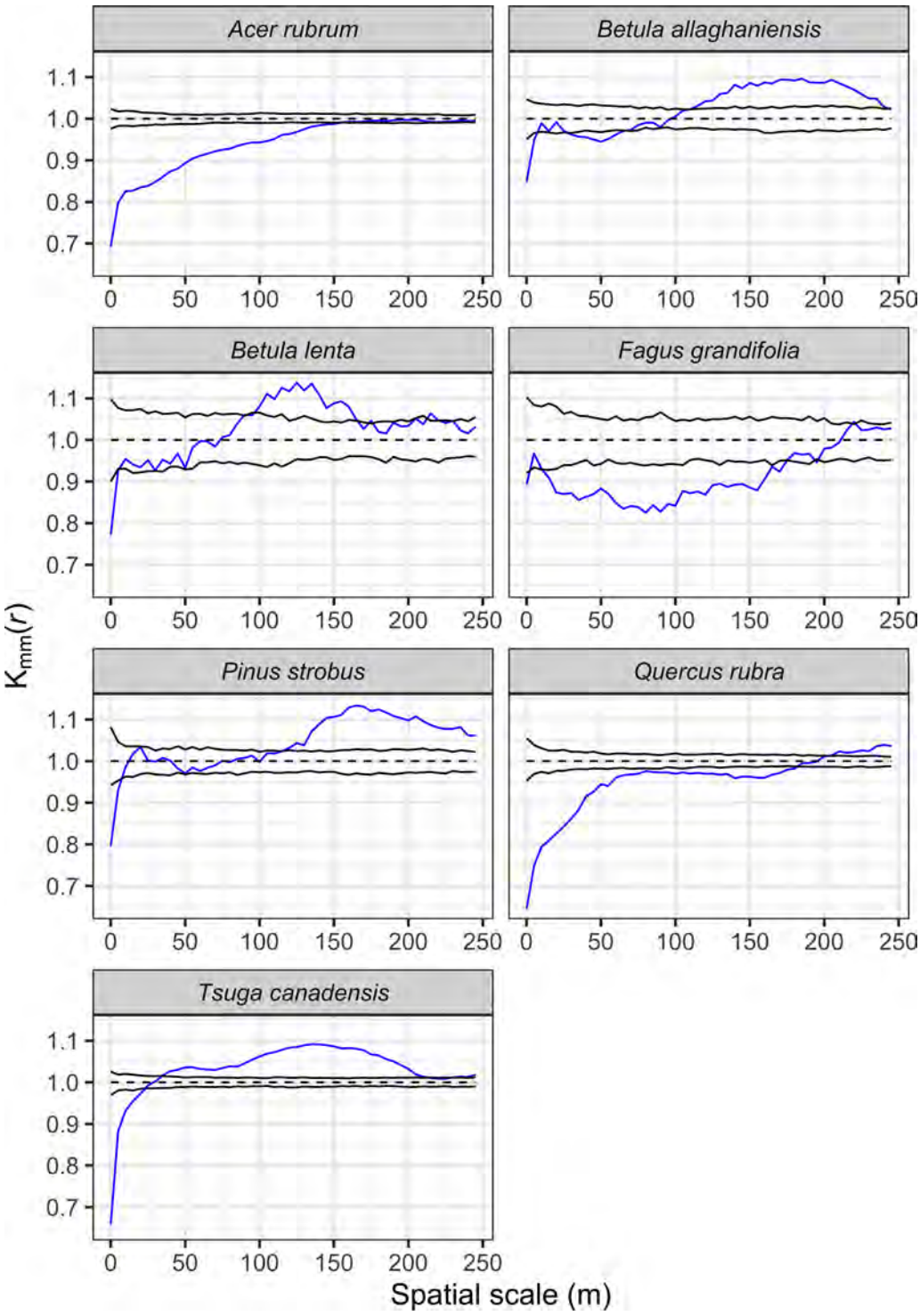

Figure 8 Univariate mark correlation function analysis results showing the effects of the underlying spatial pattern of trees on the size of conspecific individuals for seven dominant species in the Harvard Forest plot across a range of scales. The significance of this effect was (continued on next page…)

Full-size DOI: 10.7717/peerj.12693/fig-8

Figure 8 (…continued)
evaluated by comparing the calculated *kmm* (*r*) against values simulated under a null expectation, where tree sizes were randomly shuffled over all trees for each of the 199 simulations. The blue line indicates calculated *kmm* (*r*) values, while the black lines demark the 95% confidence envelope around simulated *kmm* (*r*) values under the null model. A blue line falling below, within, or above the upper confidence limit, indicates significant negative, independent, or positive correlations among DBH marks for the given species, respectively.

attempted in temperate forests. We have shown that broad patterns in land use and historical disturbance that occurred up to a century ago remain dominant controls on present-day spatial distribution and structure of overstory species. Tree species were significantly clumped within the plot and *T. canadensis* affected the distribution of other dominant canopy species in different ways. Topography and hydrology also affected the distribution and abundance of understory stems. Detailed abundance and species distribution data provided in this study will provide invaluable information on forest dynamics in the future as the currently most abundant species—*Tsuga canadensis*—is declining because of a non-native insect (*Orwig et al., 2018*).

Forest structure is contingent on past land use

The forest canopy within the HF ForestGEO plot, dominated by *T. canadensis*, *Q. rubra*, *A. rubrum*, and *P. strobus*, is representative of many central New England forests. Like other temperate ForestGEO plots, a relatively small number of species dominated the HF plot (13 species were represented by over 1000 stems). However, this number was higher than the 5–10 species that reached this abundance in other temperate ForestGEO plots (*Wang et al., 2010*; *Wang et al., 2011*; *Lutz et al., 2012*; *Bourg et al., 2013*; *Lutz et al., 2013*) and likely reflects the varied habitats, high intensity of prior land use, and early stages of stand development at HF. Although we have much historical knowledge regarding land-use change at HF, the conditional regression random-forest modeling enabled us to explore more quantitatively how patterns of tree size and stem density for the seven most abundant species have been affected by tradeoffs between legacy effects of past land uses, management interventions, disturbances, and local-scale variation in stand structure and environmental conditions. This combination of quantitative modeling with historical knowledge contributes to a deeper understanding of historical human impacts on current forest structure.

For example, our modeling results suggested that *T. canadensis* diameters and stem densities across the full plot are most strongly associated with local stand structural characteristics and neighborhood effects, whereas stem densities are only moderately associated with land-use history. This result is consistent with the relatively undisturbed appearance of the older portion of the HF plot where *T. canadensis* is most common, has persisted through time for thousands of years (*Foster & Zebryk, 1993*), and excluded of other species under its canopy. *Tsuga canadensis* is most abundant on land that was consistently used as a woodlot but never completely cleared for agriculture. The western portion of the plot was one of the few locations at HF that was mapped as *T. canadensis* forest in 1908 (*Spurr, 1956*). The high abundance of *T. canadensis* is the result of its shade tolerance and

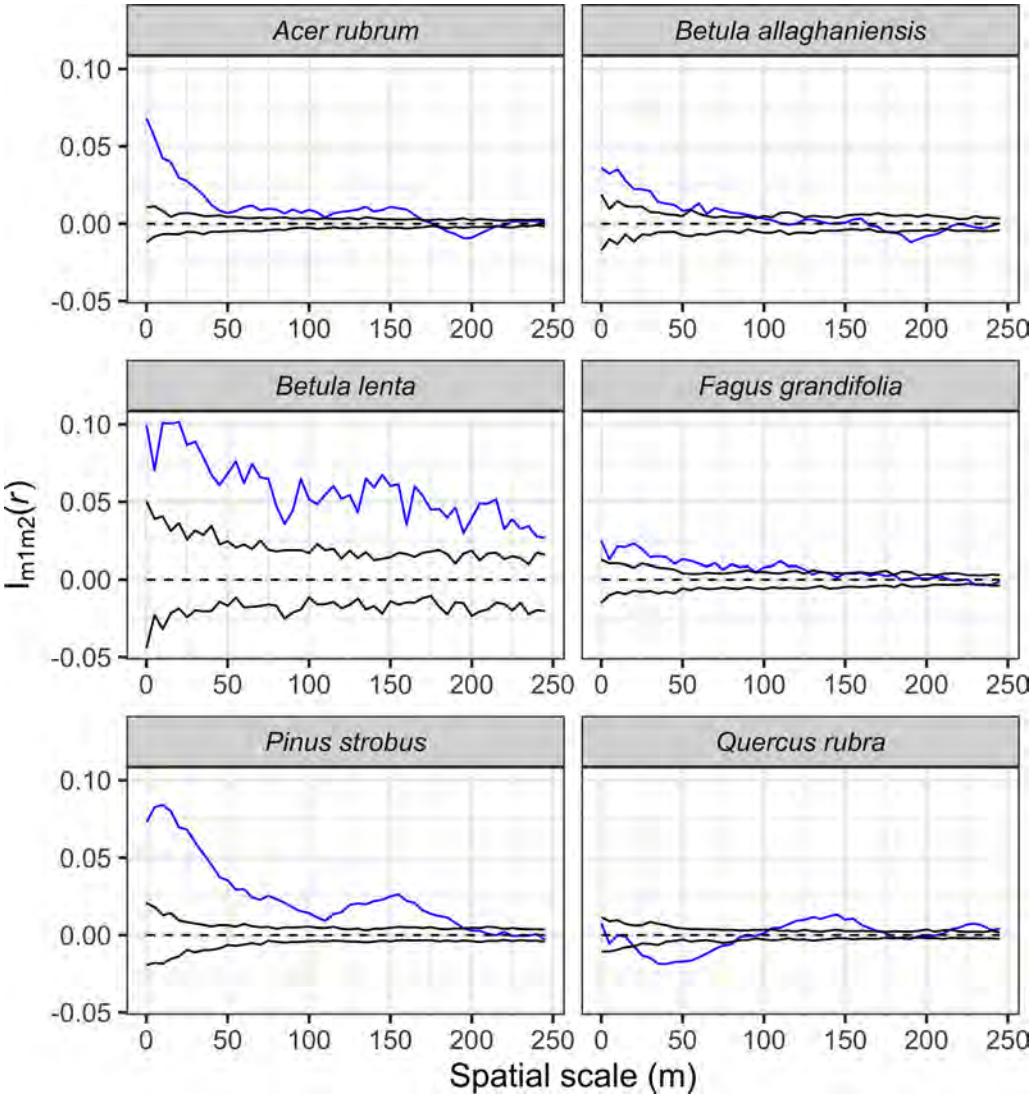

Figure 9 Bivariate marked point pattern analysis results showing the effects of the size of focal *Tsuga canadensis* individuals on the size of six other non-focal species in the HF ForestGEO plot across a range of scales. The significance of this effect was evaluated by comparing the calculated Schlather's I (*Im1m2*(*r*)) bivariate correlation statistic against values simulated under a null expectation, where non-focal species' tree sizes were randomly shuffled over trees for each of 199 simulations. The blue line indicates calculated *Im1m2*(*r*) values, whereas the black lines demark the 95% confidence envelope around simulated *Im1m2*(*r*) values under the null model. A blue line falling below, within, or above the upper confidence limit indicates significant negative, independent, or positive correlations of DBH marks of the given species with the DBH of *T. canadensis* individuals found at a range of distances, respectively.

Full-size DOI: 10.7717/peerj.12693/fig-9

deep crowns, which enable it to persist for decades, modify the understory environment by transmitting very little light, prevent other species from getting established (*Canham et al., 1994*), and gain dominance following partial cuttings, the death and subsequent salvage of *C. dentata* and *F. grandifolia*, and moderate damage from the 1938 hurricane (*Foster et al., 1992*; *Motzkin et al., 1999*; *McLachlan, Foster & Menalled, 2000*). These same disturbances

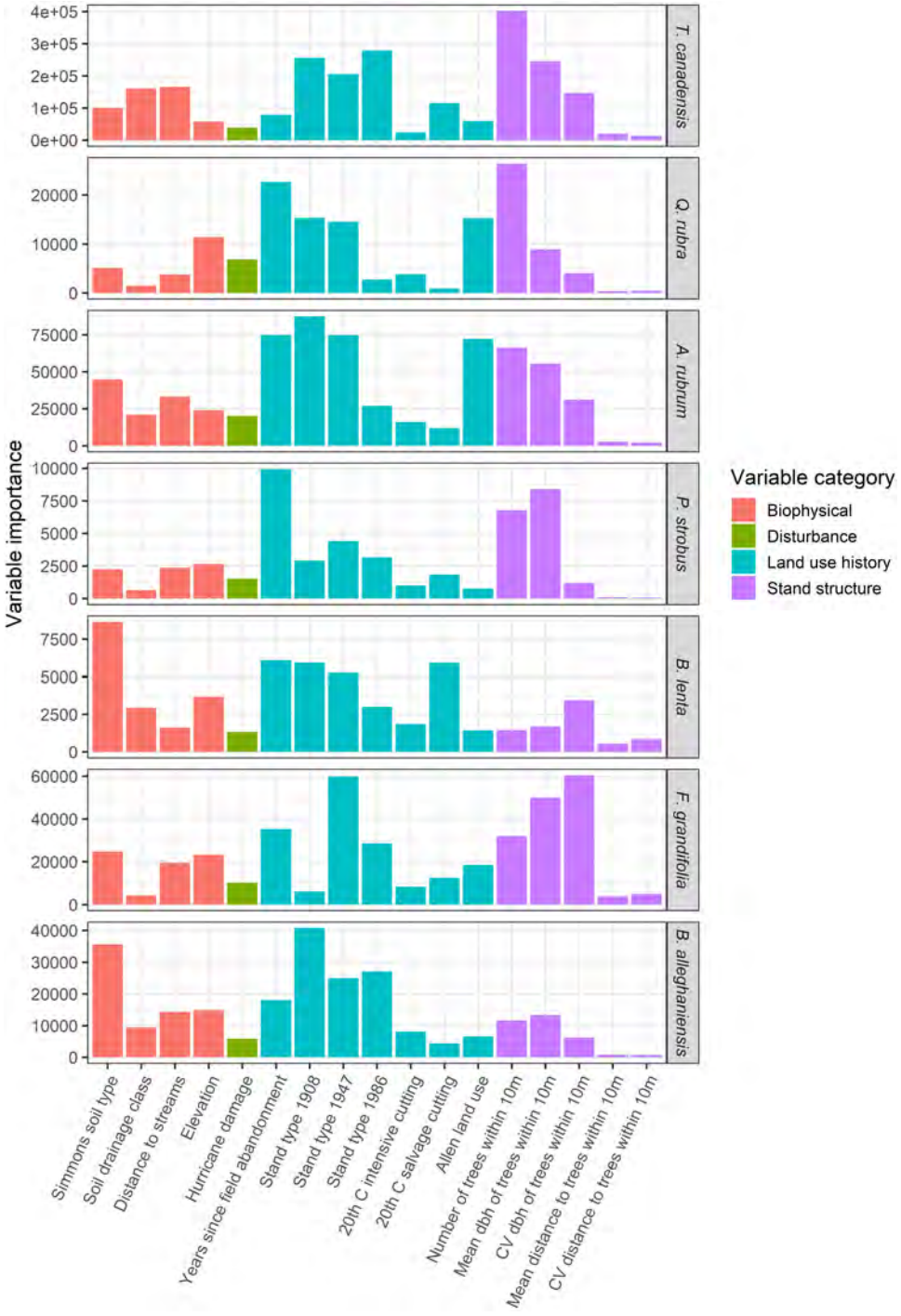

Figure 10 Variable importance scores, based on the mean decrease in prediction accuracy, from a conditional inference random-forest model predicting tree species abundance values (stems/ha) for the seven most common trees as a function of possible predictors. Variable importance scores were calculated across 400 random forest iterations and the range of values is from 0–100,000, reflecting the range of the response variable, abundance.

Full-size DOI: 10.7717/peerj.12693/fig-10

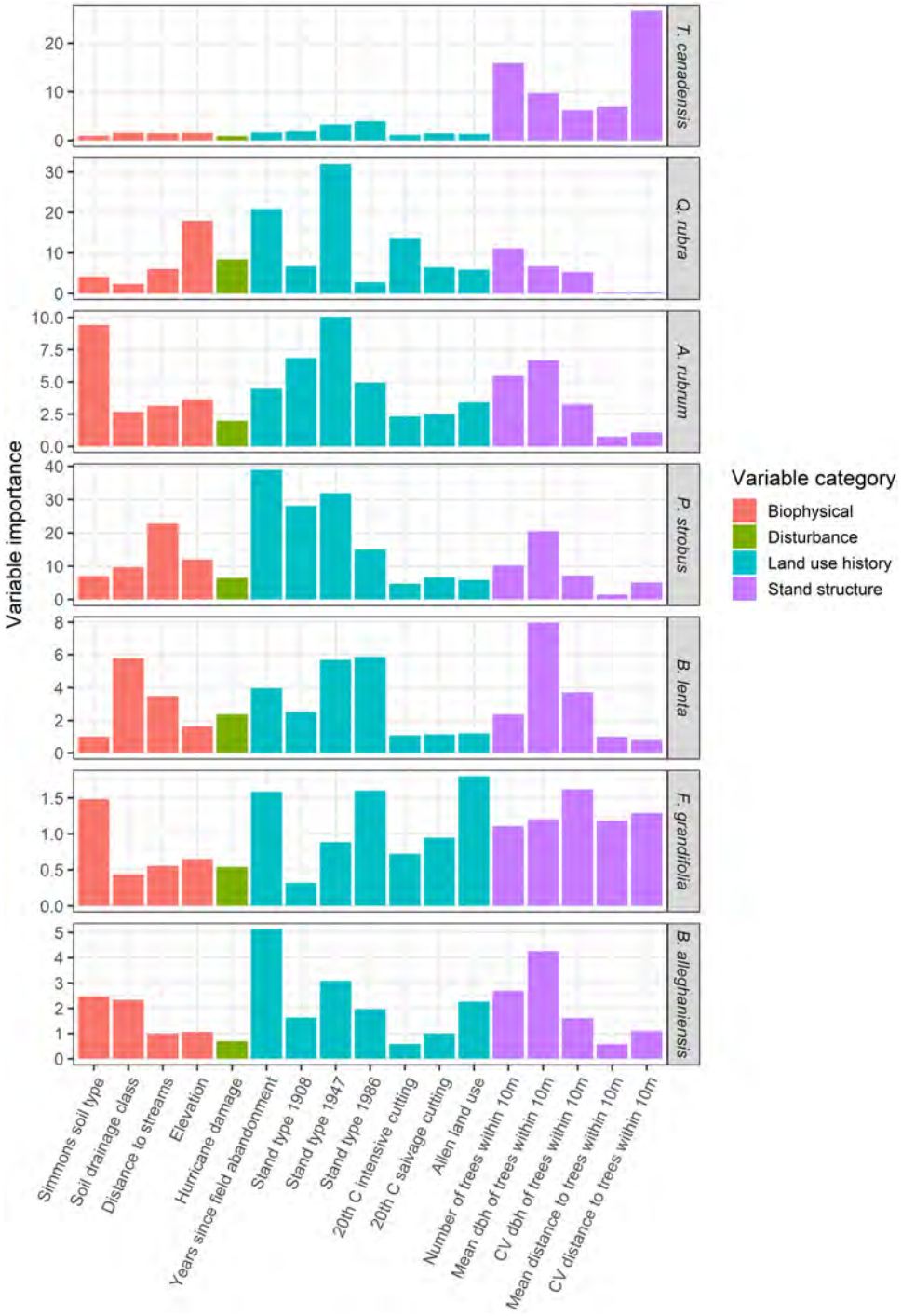

Figure 11 Variable importance scores, based on the mean decrease in prediction accuracy, from a conditional inference random-forest model predicting tree species diameter at breast height (DBH) for the seven most common trees as a function of possible predictors. Variable importance scores were calculated across 400 random forest iterations and the range of values is from 0–40, reflecting the range of the response variable, diameter.

Full-size DOI: 10.7717/peerj.12693/fig-11

also likely led to growth increases and additional establishment of *P. strobus* (*Hibbs, 1982b*), as the largest pine stems also occur on the western edge of the HF plot.

In contrast, modeling revealed stronger effects of both land-use history and stand structural variables on the sizes and stem densities of the other six dominant species. Field abandonment date and stand types present in the early- and mid-20th century are particularly strong predictors of diameters and densities of these species. This is consistent with recorded historical knowledge. For example, *Pinus strobus* and *Q. rubra* are most abundant on areas that were formerly pasture or fields in the mid- to late-1800s that also experienced intensive past silvicultural cuts, thinning, and weeding in the 1920s–1940s, and more severe damage from the 1938 hurricane (*Motzkin et al., 1999*; *Hall, 2005*). *Quercus rubra* trees had larger mean diameters and crown sizes than *F. grandifolia* or *A. rubrum*, consistent with past investigations that highlighted the ability of *Q. rubra* to overtop canopy associates and rapidly expand laterally into gaps (*Oliver, 1978*; *Hibbs, 1982a*). *Acer rubrum* and *B. alleghaniensis* are more closely associated with mesic locations such as swamp borders with silt loam soils and low-lying sites with peaty soils in the northeast corner of the plot; indeed, random-forest models supported the relatively strong importance of soil type for these species and *B. lenta* relative to the other species. The south-central portion of the plot experienced the most intensive land use. It was the only area that experienced historical cultivation and multiple periods of subsequent clear-cutting, including a harvest in 1980. This area is dominated by smaller, multi-stemmed *A. rubrum*, *Q. rubra*, *B. populifolia* and *B. papyrifera* (grey and paper birch), and *Prunus* (cherry) species, which are much more common in forests that have experienced intense human impacts (*Del Tredici, 2001*). The relationship between current stem-density patterns for *A. rubrum* and these two *Betula* species and intensive historical land-use activities can be explained by the ability of these species to sprout following cutting and take advantage of high-light environments (*Burns & Honkala, 1990*).

Understory composition, dominated by woody shrubs, appears to be determined by soil drainage and the ability of individual species to tolerate standing water, poorly drained soils, or subtle topographic variation. Historically, the swamp contained pasture on its western edge and a woodlot in the remaining portion. Today, the wetland shrubs *I. verticillata*, *Va. corymbosum*, *L. ligustrina*, and *Vi. nudum* are found in high abundance in the central beaver swamp, which otherwise is devoid of trees. The northwest section of the plot has the highest elevation and is dominated by *K. latifolia*. *Hamamelis virginiana* appears to be restricted to a narrow elevation west of the swamp and in the southeast corner of the plot. Previous work at HF related *K. latifolia* abundance to nitrogen-poor sites and *H. virginiana* to continuously forested sites (*Motzkin et al., 1999*), which is consistent with our findings.

Across all species and size classes, the forest contains a preponderance (>80,000) of small stems (<10-cm DBH) that exhibit a reverse-J size distribution. The high abundance of stems in this size class (*e.g.*, several shrub species, *T. canadensis,* and *A. rubrum*) is in contrast to several other temperate forest plots (*Lutz et al., 2012*; *Bourg et al., 2013*; *Lutz et al., 2013*), and is more similar to results from tropical evergreen (*Memiaghe et al., 2016*) or Mediterranean forests (*Gilbert et al., 2010*). Most of the abundant overstory and all the abundant shrub species also have reverse-J distributions, indicative of stable populations

and adequate regeneration. For overstory species, this likely is a result of the mix of even-age and varying-aged cohorts and single trees establishing following anthropogenic disturbances and natural gap-phase dynamics that are frequent in this region (*Oliver & Stephens, 1977*; *Hibbs, 1982a*; *Pederson, 2005*). The greater ages of the shade-tolerant *T. canadensis* that occur on primary woodland are approaching a structure and diameter distribution that resembles old-growth forest (*D'Amato, Orwig & Foster, 2008*; *Janowiak, Nagel & Webster, 2008*). In contrast, *A. rubrum* and *Q. rubra* had skewed unimodal size distributions more indicative of managed forests (*Janowiak, Nagel & Webster, 2008*).

Overstory spatial patterns

We observed significant spatial clustering among abundant overstory species at all spatial scales examined. Aggregated species distribution patterns are common in both temperate (*Hou et al., 2004*; *Hao et al., 2007*; *Wang et al., 2011*) and tropical forests (*Condit et al., 2000*; *Plotkin et al., 2000*; *Réjou-Méchain et al. et al. 2011*, *Nguyen, Uria-Diez & Wiegand, 2016*). Both external factors (habitat heterogeneity) and internal factors (dispersal limitation, succession, gap dynamics) can lead to clumped distributions at various spatial scales (*Getzin et al., 2008*; *Réjou-Méchain et al. et al. 2011*). Within the HF ForestGEO plot, high habitat heterogeneity caused by complex past land use (differing field abandonment dates followed by repeated cutting and thinning; *Motzkin et al., 1999*) has likely led to high densities of *A. rubrum* and *Q. rubra* stems in the central portion of the plot. These non-random patches of individuals with lower than average DBH (as seen in the mark correlation analysis) may reflect strong competition for light as seen elsewhere (*Fibich et al., 2016*). Similar patterns seen in *B. alleghaniensis*, *B. lenta*, *P. strobus*, and *T. canadensis* in close proximity to other conspecifics (0–20-m scale) likely reflect crowding effects, and for *T. canadensis*, the ability of thousands of small stems to persist in the understory for decades (*Marshall, 1927*). These effects disappear at intermediate scales and even become positive at distances >100 m, indicating that trees greater than the mean DBH are more broadly distributed. The negative correlation observed for *F. grandifolia* at most spatial lags $\leq$ 150 m may be more reflective of its overall size distribution with most of its stems <10-cm DBH. Beech-bark disease is present at HF, and has likely contributed, along with past cutting, to the absence of large *F. grandifolia* in the plot.

Bivariate mark correlation functions have been underused in large, stem-mapped plots but hold great promise in ecological research (*Velázquez et al., 2016*). We used this method to examine the relationship between the size of individuals of *T. canadensis*, an important foundation species within the plot, with the size of six other important canopy species some distance away. Apart from *Q. rubra*, diameters of the other five species were positively correlated with the diameters of *T. canadensis* at all spatial scales. This pattern is consistent with *T. canadensis* being a foundation species in this forest (*Buckley, Case & Ellison, 2016*; *Ellison et al., 2019*), but it also simply could indicate a "habitat" effect: all these species are growing well everywhere and are found at a wide range of sizes. This effect was particularly strong for *B. lenta* and *P. strobus,* but weaker for *A. rubrum*, *B. alleghaniensis*, and *F. grandifolia* and disappeared after 100–150 m. Diameter of *Q. rubra* was on average smaller than expected by chance when within 20–80 m of *T. canadensis*. Historical factors play a

role here, as the spatial distribution of these species highlight that oak abundance is the lowest within the *T. canadensis*-dominated portions of the plot that were woodlots and suggest that *T. canadensis* and the dense shade cast by their crowns limited establishment of the more intolerant *Q. rubra*.

CONCLUSIONS

The HF ForestGEO plot is the largest mapped temperate-forest plot in North America and joins the growing array of temperate forest-dynamics plots worldwide (*Anderson-Teixeira et al., 2015*). The species composition and aggregated spatial patterns within the plot are still being influenced by a land-use legacy of anthropogenic and natural disturbances that occurred decades to over a century ago. Despite extensive 20th-century harvesting, silvicultural thinning, and salvage operations following the 1938 hurricane, the most common overstory species in the HF ForestGEO plot today can best be predicted by longer-term land-use legacies represented by the 1908 forest type and the date of late 19th-century field abandonment, and tree neighborhood effects. At smaller scales, there is evidence of crowding effects of many common species, likely a result of successional dynamics of these aggrading forests following intensive land use. The increasing importance of *T. canadensis* during the last century across the plot negatively affected the distribution of *Q. rubra*. Its location and five-year schedule of plot sampling highlight the plot as valuable long-term infrastructure that will complement Harvard Forest, LTER, NEON, and ForestGEO research efforts (*Orwig et al., 2018*). Because all woody stems $\geq$ 1-cm DBH are mapped and measured, the data have been used in a variety of complementary ways, including to examine species codispersion patterns and spatial patterns of species co-occurrence (*Buckley, Case & Ellison, 2016*; *Case et al., 2016*), help inform a simulation model of forest dynamics (SORTIE; *Case et al., 2017*), assist with investigating crown allometry (*Sullivan et al., 2017*) and crown mapping (*Hastings et al., 2020*), develop maps of tree-mycorrhizal associations (*Sousa et al., 2021*) and aid in identifying statistical fingerprints of foundation species (*Ellison et al., 2019*). In addition, the data enable us to document changing species distribution patterns at an uncommonly large scale, while focusing on elements of the landscape that are often ignored, like beaver swamps and shrub thickets, and examine their contribution to overall forest structure and composition.

ACKNOWLEDGEMENTS

We thank the many field technicians who helped census the plot including Jerry Breault, Kyle Gay, Mike Babineau, Christian Foster, Jeff Hutchins, Tamara Martz, Elizabeth Barnes, Rachel DeMatte, Bianca Kubierschky, Ellen Kujawa, Audrey Lamb, Ben Misiuk, Finn Olcott, Hallie Schwab, Alida Mau, Joe Horn, Taylor Lucey, Brett Gelinas, Danelle Laflower, Sarah Meyers, and Kyle Krigest. We are grateful to John Wisnewski and the woods crew at HF for providing materials, supplies, and invaluable field assistance with plot logistics. Joel Botti and Frank Schiappa provided survey expertise to establish the 35-ha plot. Special thanks to Stuart Davies and Rick Condit for field training, database assistance, and plot advice. Sean McMahon and Suzanne Lao were extremely helpful with field planning,

data questions, and many plot logistics. Thanks to Jeannette Bowlen for administrative assistance, Emery Boose and Paul Siqueira for help with plot coordinates, Brian Hall for map and GIS layer production, and Matthew Duveneck and Danelle Laflower for assistance with figures. We also thank David Foster for his support and assistance with plot design, location, and integration with other long-term studies at HF. We thank three anonymous reviewers as well as Jonathan Thompson, Neil Pederson, Audrey Barker-Plotkin and lab group members for providing critical comments on earlier versions of the manuscript. There were no conflicts of interest.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

Funding for the Harvard ForestGEO Forest Dynamics plot was provided by the Center for Tropical Forest Science and Smithsonian Institute's Forest Global Earth Observatory (CTFS-ForestGEO), the National Science Foundation's LTER program (DEB 06-20443, DEB 12-37491, DEB 18-32210) and Harvard University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
The Center for Tropical Forest Science and Smithsonian Institute's Forest Global Earth Observatory (CTFS-ForestGEO).
The National Science Foundation's LTER program: DEB 06-20443, DEB 12-37491, DEB 18-32210.
Harvard University.

Competing Interests

Aaron M. Ellison is employed by Sound Solutions for Sustainable Science. All other authors declare that they have no competing interests.

Author Contributions

- David A. Orwig conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Jason A. Aylward performed the experiments, prepared figures and/or tables, and approved the final draft.
- Hannah L. Buckley, Bradley S. Case and Aaron M. Ellison conceived and designed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The data are available from the Environmental Data Inititative: Orwig, D., D. Foster, and A. Ellison. 2021. Harvard Forest CTFS-ForestGEO Mapped Forest

Plot since 2014 ver 3. Environmental Data Initiative. https://doi.org/10.6073/pasta/18c01a2bb5f5bdf98846542ebbdbad65 (accessed 2021-12-07).

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.12693#supplemental-information.

REFERENCES

Anderson-Teixeira KJ, Davies SJ, Bennett AC, Gonzalez-Akre EB, Muller-Landau HC, Joseph Wright S, Abu Salim K, Almeyda Zambrano AM, Alonso A, Baltzer JL. 2015. CTFS-Forest GEO: a worldwide network monitoring forests in an era of global change. *Global Change Biology* **21**:528–549 DOI 10.1111/gcb.12712.

Birks HH, Birks H, Kaland PE, Moe D. 1988. *The cultural landscape: past, present and future*. Cambridge: Cambridge University Press.

Boose E, Gould E. 2019. Harvard Forest Climate Data since 1964. Harvard Forest Data Archive: HF300. *Available at http://harvardforest.fas.harvard.edu*.

Bourg NA, McShea WJ, Thompson JR, McGarvey JC, Shen X. 2013. Initial census, woody seedling, seed rain, and stand structure data for the SCBI SIGEO large forest dynamics plot. *Ecology* **94**:2111–2112 DOI 10.1890/13-0010.1.

Breiman L. 2001. Random forests. *Machine Learning* **45**:5–32 DOI 10.1023/A:1010933404324.

Buckley HL, Case BS, Ellison AM. 2016. Using codispersion analysis to characterize spatial patterns in species co-occurrences. *Ecology* **97**:32–39 DOI 10.1890/15-0578.1.

Bürgi M, Östlund L, Mladenoff DJ. 2017. Legacy effects of human land use: ecosystems as time-lagged systems. *Ecosystems* **20**:94–103 DOI 10.1007/s10021-016-0051-6.

Burns RM, Honkala BH. 1990. *Silvics of North America. Volume 2. Hardwoods*. Washington, D.C.: United States Department of Agriculture.

Canham CD, Finzi AC, Pacala SW, Burbank DH. 1994. Causes and consequences of resource heterogeneity in forests: interspecific variation in light transmission by canopy trees. *Canadian Journal of Forest Research* **24**:337–349 DOI 10.1139/x94-046.

Case BS, Buckley HL, Barker Plotkin A, Ellison A. 2016. Using codispersion analysis to quantify temporal changes in the spatial pattern of forest stand structure. *Journal of Chilean Statistics* **7**:3–15.

Case BS, Buckley HL, Barker-Plotkin AA, Orwig DA, Ellison AM. 2017. When a foundation crumbles: forecasting forest dynamics following the decline of the foundation species Tsuga canadensis. *Ecosphere* **8**:e01893.

Chazdon RL. 2003. Tropical forest recovery: legacies of human impact and natural disturbances. *Perspectives in Plant Ecology, Evolution and Systematics* **6**:51–71 DOI 10.1078/1433-8319-00042.

Condit R. 1998. *Tropical forest census plots: methods and results from Barro Colorado Island, Panama and a comparison with other plots*. Berlin: Springer Science & Business Media.

Condit R. 2014. CTFS R Package. Smithsonian Tropical Research Institute. *Available at https://forestgeo.si.edu/explore-data/r-package-0*.

Condit R, Ashton PS, Baker P, Bunyavejchewin S, Gunatilleke S, Gunatilleke N, Hubbell SP, Foster RB, Itoh A, LaFrankie JV. 2000. Spatial patterns in the distribution of tropical tree species. *Science* **288**:1414–1418 DOI 10.1126/science.288.5470.1414.

D'Amato AW, Orwig DA, Foster DR. 2008. The influence of successional processes and disturbance on the structure of Tsuga canadensis forests. *Ecological Applications* **18**:1182–1199 DOI 10.1890/07-0919.1.

Del Tredici P. 2001. Sprouting in temperate trees: a morphological and ecological review. *The Botanical Review* **67**:121–140 DOI 10.1007/BF02858075.

Ellison AM. 2019. Foundation species, non-trophic interactions, and the value of being common. *Iscience* **13**:254–268 DOI 10.1016/j.isci.2019.02.020.

Ellison AM, Bank MS, Clinton BD, Colburn EA, Elliott K, Ford CR, Foster DR, Kloeppel BD, Knoepp JD, GM Lovett. 2005. Loss of foundation species: consequences for the structure and dynamics of forested ecosystems. *Frontiers in Ecology and the Environment* **3**:479–486 DOI 10.1890/1540-9295(2005)003[0479:LOFSCF]2.0.CO;2.

Ellison AM, Buckley HL, Case BS, Cardenas D, Duque ÁJ, Lutz JA, Myers JA, Orwig DA, Zimmerman JK. 2019. Species diversity associated with foundation species in temperate and tropical forests. *Forests* **10**:128 DOI 10.3390/f10020128.

Ellison AM, Lavine M, Kerson PB, Barker Plotkin AA, Orwig DA. 2014. Building a foundation: Land-use history and dendrochronology reveal temporal dynamics of a Tsuga canadensis (Pinaceae) forest. *Rhodora* **116**:377–427 DOI 10.3119/14-04.

Fibich P, Lepš J, Novotný V, Klimeš P, Těšitel J, Molem K, Damas K, Weiblen GD. 2016. Spatial patterns of tree species distribution in New Guinea primary and secondary lowland rain forest. *Journal of Vegetation Science* **27**:328–339 DOI 10.1111/jvs.12363.

Fisher RT. 1933. New England forests: biological factors. New England's prospect. *American Geographical Society. Special Publication* **21**:3–223.

Foster DR. 1992. Land-use history (1730-1990) and vegetation dynamics in central New England, USA. *Journal of Ecology* **75**:3–771.

Foster D. 2014. *Hemlock: a forest giant on the edge*. New Haven, Connecticut: Yale University Press.

Foster DR, Aber J. 2004. *Forests in time. Ecosystem structure and function as a consequence of 1000 years of change*. New Haven, Connecticut: Synthesis volume of the Harvard Forest LTER Program Yale University Press.

Foster DR, Boose ER. 1992. Patterns of forest damage resulting from catastrophic wind in central New England, USA. *Journal of Ecology* **7**:9–98.

Foster DR, Motzkin G, Slater B. 1998. Land-use history as long-term broad-scale disturbance: regional forest dynamics in central New England. *Ecosystems* **1**:96–119 DOI 10.1007/s100219900008.

Foster DR, Zebryk TM. 1993. Long-term vegetation dynamics and disturbance history of a Tsuga-dominated forest in New England. *Ecology* **74**:982–998 DOI 10.2307/1940468.

Foster DR, Zebryk T, Schoonmaker P, Lezberg A. 1992. Post-settlement history of human land-use and vegetation dynamics of a Tsuga canadensis (hemlock) woodlot in central New England. *Journal of Ecology* **77**:3–786.

Fox EW, Hill RA, Leibowitz SG, Olsen AR, Thornbrugh DJ, Weber MH. 2017. Assessing the accuracy and stability of variable selection methods for random forest modeling in ecology. *Environmental Monitoring and Assessment* **189**:1–20 DOI 10.1007/s10661-016-5706-4.

Getzin S, Wiegand T, Wiegand K, He F. 2008. Heterogeneity influences spatial patterns and demographics in forest stands. *Journal of Ecology* **96**:807–820 DOI 10.1111/j.1365-2745.2008.01377.x.

Gilbert GS, Howard E, Ayala-Orozco B, Bonilla-Moheno M, Cummings J, Langridge S, Parker IM, Pasari J, Schweizer D, Swope S. 2010. Beyond the tropics: forest structure in a temperate forest mapped plot. *Journal of Vegetation Science* **21**:388–405 DOI 10.1111/j.1654-1103.2009.01151.x.

Glitzenstein JS, Canham CD, McDonnell MJ, Streng DR. 1990. Effects of environment and land-use history on upland forests of the Cary Arboretum, Hudson Valley, New York. *Bulletin of the Torrey Botanical Club* **10**:6–122.

Griffith G, Omernik S. Pierson, J, Kiilsgaard C. 1994. The Massachusetts ecological regions project. US Environmental Protection Agency. Corvallis: Environmental Research Laboratory.

Haines A. 2011. *New England Wild Flower Society's Flora Novae Angliae: a manual for the identification of native and naturalized higher vascular plants of New England.* New Haven: Yale University Press.

Hall B. 2005. Historical GIS Data for Harvard Forest Properties from 1908 to Present HF110. *Available at http://harvardforest.fas.harvard.edu*.

Hall B, Motzkin G, Foster DR, Syfert M, Burk J. 2002. Three hundred years of forest and land-use change in Massachusetts, USA. *Journal of Biogeography* **29**:1319–1335 DOI 10.1046/j.1365-2699.2002.00790.x.

Hao Z, Zhang J, Song B, Ye J, Li B. 2007. Vertical structure and spatial associations of dominant tree species in an old-growth temperate forest. *Forest Ecology and Management* **252**:1–11 DOI 10.1016/j.foreco.2007.06.026.

Hastings JH, Ollinger SV, Ouimette AP, Sanders-DeMott R, Palace MW, Ducey MJ, Sullivan FB, Basler D, Orwig DA. 2020. Tree species traits determine the success of lidar-based crown mapping in a mixed temperate forest. *Remote Sensing* **12**:309 DOI 10.3390/rs12020309.

Hibbs DE. 1982a. Gap dynamics in a hemlock–hardwood forest. *Canadian Journal of Forest Research* **12**:522–527 DOI 10.1139/x82-081.

Hibbs DE. 1982b. White pine in the transition hardwood forest. *Canadian Journal of Botany* **60**:2046–2053 DOI 10.1139/b82-252.

Hogan JA, Zimmerman JK, Thompson J, Nytch CJ, Uriarte M. 2016. The interaction of land-use legacies and hurricane disturbance in subtropical wet forest: twenty-one years of change. *Ecosphere* **7**:e01405.

Hothorn T, Hornik K, Zeileis A. 2006. Unbiased recursive partitioning: a conditional inference framework. *Journal of Computational and Graphical Statistics* **15**:651–674 DOI 10.1198/106186006X133933.

Hou J, Mi X, Liu C, Ma K. 2004. Spatial patterns and associations in a Quercus-Betula forest in northern China. *Journal of Vegetation Science* **15**:407–414.

Illian J, Penttinen A, Stoyan H, Stoyan D. 2008. *Statistical analysis and modelling of spatial point patterns (Vol. 70)*. West Sussex: John Wiley & Sons.

Jacquemyn H, Endels P, Honnay O, Wiegand T. 2010. Evaluating management interventions in small populations of a perennial herb Primula vulgaris using spatio-temporal analyses of point patterns. *Journal of Applied Ecology* **47**:431–440 DOI 10.1111/j.1365-2664.2010.01778.x.

Janowiak MK, Nagel LM, Webster CR. 2008. Spatial scale and stand structure in northern hardwood forests: implications for quantifying diameter distributions. *Forest Science* **54**:497–506 DOI 10.17221/49/2008-JFS.

Lutz JA, Larson AJ, Freund JA, Swanson ME, Bible KJ. 2013. The importance of large-diameter trees to forest structural heterogeneity. *PLOS ONE* **8**:e82784 DOI 10.1371/journal.pone.0082784.

Lutz JA, Larson AJ, Swanson ME, Freund JA. 2012. Ecological importance of large-diameter trees in a temperate mixed-conifer forest. *PLOS ONE* **7**:e36131 DOI 10.1371/journal.pone.0036131.

Marshall R. 1927. The growth of hemlock before and after release from suppression. *Harvard Forest Bulletin* **11**:1–42.

McLachlan JS, Foster DR, Menalled F. 2000. Anthropogenic ties to late-successional structure and composition in four New England hemlock stands. *Ecology* **81**:717–733 DOI 10.1890/0012-9658(2000)081[0717:ATTLSS]2.0.CO;2.

Memiaghe HR, Lutz JA, Korte L, Alonso A, Kenfack D. 2016. Ecological importance of small-diameter trees to the structure, diversity and biomass of a tropical evergreen forest at Rabi, Gabon. *PLOS ONE* **11**:e0154988 DOI 10.1371/journal.pone.0154988.

Mi C, Huettmann F, Guo Y, Han X, Wen L. 2017. Why choose Random Forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. *PeerJ* **5**:e2849 DOI 10.7717/peerj.2849.

Mohapatra J, Singh CP, Hamid M, Verma A, Semwal SC, Gajmer B, Khuroo AA, Kumar A, Nautiyal MC, Sharma N. 2019. Modelling *Betula utilis* distribution in response to climate-warming scenarios in Hindu-Kush Himalaya using random forest. *Biodiversity and Conservation* **28**:2295–2317 DOI 10.1007/s10531-019-01731-w.

Motzkin G, Foster D, Allen A, Harrod J, Boone R. 1996. Controlling site to evaluate history: vegetation patterns of a New England sand plain. *Ecological Monographs* **66**:345–365 DOI 10.2307/2963522.

Motzkin G, Wilson P, Foster DR, Allen A. 1999. Vegetation patterns in heterogeneous landscapes: the importance of history and environment. *Journal of Vegetation Science* **10**:903–920 DOI 10.2307/3237315.

Nguyen HH, Uria-Diez J, Wiegand K. 2016. Spatial distribution and association patterns in a tropical evergreen broad-leaved forest of north-central Vietnam. *Journal of Vegetation Science* **27**:318–327 DOI 10.1111/jvs.12361.

Oliver CD. 1978. The development of northern red oak in mixed stands in central New England. *Yale School of Forestry & Environmental Studies Bulletin Series* **91**:1–63.

Oliver CD, Stephens EP. 1977. Reconstruction of a mixed-species forest in central New England. *Ecology* **58**:562–572 DOI 10.2307/1939005.

Orwig DA, Barker Plotkin AA, Davidson EA, Lux H, Savage KE, Ellison AM. 2013. Foundation species loss affects vegetation structure more than ecosystem function in a northeastern USA forest. *PeerJ* **1**:e41 DOI 10.7717/peerj.41.

Orwig DA, Boucher P, Paynter I, Saenz E, Li Z, Schaaf C. 2018. The potential to characterize ecological data with terrestrial laser scanning in Harvard Forest, MA. *Interface Focus* **8(2)**:20170044 DOI 10.1098/rsfs.2017.0044.

Orwig D, Foster D, Ellison A. 2015. Harvard Forest CTFS-ForestGEO mapped forest plot since 2014. Harvard Forest Data Archive: HF253. *Available at http://harvardforest.fas.harvard.edu*.

Pederson N. 2005. *Climatic sensitivity and growth of southern temperate tress in the Eastern US: implicatins for the carbon cycle*. New York: Columbia University New York.

Plotkin JB, Potts MD, Leslie N, Manokaran N, LaFrankie J, Ashton PS. 2000. Species–area curves, spatial aggregation, and habitat specialization in tropical forests. *Journal of Theoretical Biology* **207**:81–99 DOI 10.1006/jtbi.2000.2158.

R Core Team. 2013. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.

Rackham O. 1986. *The history of the countryside*. London: JM Dent.

Rasband W. 2012. ImageJ: image processing and analysis in Java. Bethesda: US National Institutes of Health.

Raup HM, Carlson RE. 1941. The history of land use in the Harvard Forest. *The Harvard Forest Bulletin* **20**:4–62.

Réjou-Méchain M, Flores O, Bourland N, Doucet JL, Fétéké RF, Pasquier A, Hardy OJ. 2011. Spatial aggregation of tropical trees at multiple spatial scales. *Journal of Ecology* **99**:1373–1381 DOI 10.1111/j.1365-2745.2011.01873.x.

Rhemtulla JM, Mladenoff DJ, Clayton MK. 2009. Legacies of historical land use on regional forest composition and structure in Wisconsin, USA (mid-1800s to 1930s to 2000s). *Ecological Applications* **19**:1061–1078 DOI 10.1890/08-1453.1.

Rowlands W. 1941. Damage to even-aged stands in Petersham, Massachusetts by the 1938 hurricane as influenced by stand condition. MF thesis, Harvard University, Cambridge, Massachusetts.

Russell EWB. 1997. *People and the land through time: linking ecology and history*. New Haven: Yale University Press.

Shearman TM, Varner JM, Hood SM, Cansler CA, Hiers JK. 2019. Modelling post-fire tree mortality: can random forest improve discrimination of imbalanced data? *Ecological Modelling* **414**:108855 DOI 10.1016/j.ecolmodel.2019.108855.

Simmons CS. 1941. Prospect hill 1941 soil survey. *Harvard Forest Archives- HF 1941-B (with map)*.

Sousa D, Fisher JB, Galvan FR, Pavlick RP, Cordell S, Giambelluca TW, Giardina CP, Gilbert GS, Imran-Narahari F, Litton CM, Lutz JA. 2021. Tree canopies reflect mycorrhizal composition. *Geophysical Research Letters* **48(10)**:e2021GL092764.

Spurr SH. 1956. Forest associations in the Harvard Forest. *Ecological Monographs* **26**:245–262 DOI 10.2307/1948491.

Strobl C, Boulesteix A. Zeileis, A-L, Hothorn T. 2007. Bias in random forest variable importance measures: illustrations, sources, and a solution. *BMC Bioinformatics* **8**:25 DOI 10.1186/1471-2105-8-25.

Sullivan FB, Ducey MJ, Orwig DA, Cook B, Palace MW. 2017. Comparison of lidar- and allometry-derived canopy height models in an eastern deciduous forest. *Forest Ecology and Management* **406**:83–94 DOI 10.1016/j.foreco.2017.10.005.

Thompson J, Brokaw N, Zimmerman JK, Waide RB, Everham EM, Lodge DJ, Taylor CM, García-Montiel D, Fluet M. 2002. Land use history, environment, and tree composition in a tropical forest. *Ecological Applications* **12**:1344–1363 DOI 10.1890/1051-0761(2002)012[1344:LUHEAT]2.0.CO;2.

Thompson JR, Carpenter DN, Cogbill CV, Foster DR. 2013. Four centuries of change in northeastern United States forests. *PLOS ONE* **8**:e72540 DOI 10.1371/journal.pone.0072540.

Turner B, Clark W, Kates R, Richards J, Mathews J. 1990. The earth as transformed by human action: global and regional changes in the biosphere over the past 300 years. Cambridge: Cambridge University Press.

Van Gemerden BS, Olff H, Parren MP, Bongers F. 2003. The pristine rain forest? Remnants of historical human impacts on current tree species composition and diversity. *Journal of Biogeography* **30**:1381–1390 DOI 10.1046/j.1365-2699.2003.00937.x.

Velázquez E, Martínez I, Getzin S, Moloney KA, Wiegand T. 2016. An evaluation of the state of spatial point pattern analysis in ecology. *Ecography* **39**:1042–1055 DOI 10.1111/ecog.01579.

Wang X, Wiegand T, Hao Z, Li B, Ye J, Lin F. 2010. Species associations in an old-growth temperate forest in north-eastern China. *Journal of Ecology* **98**:674–686 DOI 10.1111/j.1365-2745.2010.01644.x.

Wang X, Wiegand T, Wolf A, Howe R, Davies SJ, Hao Z. 2011. Spatial patterns of tree species richness in two temperate forests. *Journal of Ecology* **99**:1382–1393 DOI 10.1111/j.1365-2745.2011.01857.x.

Westveld M. 1956. Natural forest vegetation zones of New England. *Journal of Forestry* **54**:332–338.

Wiegand T, He F, Hubbell SP. 2013. A systematic comparison of summary characteristics for quantifying point patterns in ecology. *Ecography* **36**:92–103 DOI 10.1111/j.1600-0587.2012.07361.x.

Wiegand T, Moloney KA. 2004. Rings, circles, and null-models for point pattern analysis in ecology. *Oikos* **104**:209–229 DOI 10.1111/j.0030-1299.2004.12497.x.

Wiegand T, Moloney KA. 2014. *Handbook of spatial point-pattern analysis in ecology*. Baca Raton: Chapman and Hall/CRC.

Zhang ZH, Hu G, Zhu D, Luo DH, Ni J. 2010. Spatial patterns and interspecific associations of dominant tree species in two old-growth karst forests, SW China. *Ecological Research* **25(6)**:1151–1160 DOI 10.1007/s11284-010-0740-0.

Zimmerman JK, Aide TM, Rosario M, Serrano M, Herrera L. 1995. Effects of land management and a recent hurricane on forest structure and composition in the Luquillo Experimental Forest, Puerto Rico. *Forest Ecology and Management* **77**:65–76 DOI 10.1016/0378-1127(95)03575-U.