



An imputed forest composition map for New England screened by species range boundaries



Matthew J. Duveneck^{a,*}, Jonathan R. Thompson^a, B. Tyler Wilson^b

^a Harvard Forest, Harvard University, Petersham, MA, USA

^b Forest Inventory and Analysis, Northern Research Station, USDA Forest Service, Saint Paul, MN, USA

ARTICLE INFO

Article history:

Received 19 December 2014

Received in revised form 3 March 2015

Accepted 7 March 2015

Keywords:

Nearest-neighbor imputation

New England

Forest inventory

Tree species distribution

ABSTRACT

Initializing forest landscape models (FLMs) to simulate changes in tree species composition requires accurate fine-scale forest attribute information mapped continuously over large areas. Nearest-neighbor imputation maps, maps developed from multivariate imputation of field plots, have high potential for use as the initial condition within FLMs, but the tendency for field plots to be imputed over large geographical distances can result in species being mapped outside of their home ranges, which is problematic. We developed an approach for evaluating and imputing field plots based on their similarity across multiple spatial environmental variates, their species composition, and their geographical distance between source and imputation to produce a map that is appropriate for initializing an FLM. We used this approach to map 13 million ha of forest throughout the six New England states (Rhode Island, Connecticut, Massachusetts, New Hampshire, Vermont, and Maine). Using both independent state forest and, more extensive, ecoregion validation data sets, we compared the imputation map to field inventory data, based on the dissimilarity of tree community composition and the rank order correlation of tree species abundance. Average Bray–Curtis dissimilarity between the imputation map and field plots was 0.32 and 0.12, for the state forest and ecoregion validation data sets, respectively. Average Spearman rank order correlation was 0.81 and 0.93 for the state forest and ecoregion validation data sets, respectively. Our analyses suggest that this approach to imputation can realistically capture regional variation in forest composition. We expect the imputation map will be valuable for several regional forest studies and that the approach could be successfully used in other regions.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Forest Landscape Models (FLMs) simulate succession and disturbance over meso-scales (generally 100–10,000 km²) and incorporate spatially interactive processes represented using interacting raster map cells (pixels) (Scheller and Mladenoff, 2007). FLMs typically operate at 30–250 m cell resolution. Increasing use of FLMs to simulate forest change over large spatial and temporal scales is driving demand for fine grain forest attribute data that is mapped continuously over large areas for use as initial (or starting) conditions (He, 2008; Keane et al., 2004; Scheller and Mladenoff, 2007). Because these models simulate processes at the scale of individual trees or cohorts of trees, they require a level of detail that is typically only obtainable via field inventories. For example, the LANDIS-II FLM requires a spatial representation of tree species-age cohorts (Scheller et al., 2007).

Given a detailed representation of initial forest conditions, FLMs are frequently used to simulate the effects of natural and human processes on forests. They offer realistic spatial depictions of future forest conditions at a similarly high level of detail. The FLM approach is valuable for understanding how species distributions and ecosystem conditions may change over large areas and of long time frames in response to climate change, land-use and other environmental stressors (Bettinger et al., 2005; Duveneck et al., 2014a; Gustafson, 2013; Liang et al., 2014; Thompson et al., 2011).

The challenge of model initialization is pronounced for FLMs relative to other spatially explicit forest models. Unlike models used to examine aspects of ecological theory (e.g., gap models (Keane et al., 1996; Shugart et al., 2010)), FLMs are used to examine landscape changes that are specific to a certain place and time making a “spin-up to equilibrium” approach potentially unreliable. Also, because the focus of a FLM is typically on individual species—whether to infer differences in species distributions or to look at species-specific disturbances (e.g., pests, pathogens, timber harvest)—users cannot rely on simplified representations of plant

* Corresponding author. Tel.: +1 978 756 6167.

E-mail address: mduveneck@gmail.com (M.J. Duveneck).

functional types, which is common practice for initializing dynamic vegetation models (Bonan et al., 2011; Haxeltine and Prentice, 1996; Moorcroft et al., 2001). Finally, due to the longevity of trees and the tendency for forests to have strong compositional inertia, the representation or choice of initial conditions can have significant consequences for FLM simulation outcomes.

While remote sensing platforms are able to survey large areas; they are generally unable to discern individual species or stand structure information (e.g., species age). Forest inventory data, such as data from the U.S. Department of Agriculture Forest Service Forest Inventory and Analysis (FIA) program (Bechtold

and Patterson, 2005), are widely available to populate initial conditions within model simulations. However, inventory plots are sparsely distributed relative to the typical grain size used within a FLM. As a result, researchers often impute attributes from field plots (source plots) to each of the raster pixels within a landscape (Hudak et al., 2008; Ohmann and Gregory, 2002). Several methods for spatial imputation exist.

Imputation methods have been used to develop initial conditions for FLMs. “Landscape Builder” (Dijak, 2013) is software designed to stochastically impute source plots within landcover and landform spatial layers. This has the strength of using land



Fig. 1. Study landscape in New England (dark gray) within area where kNN data are available in the eastern United States (light gray).

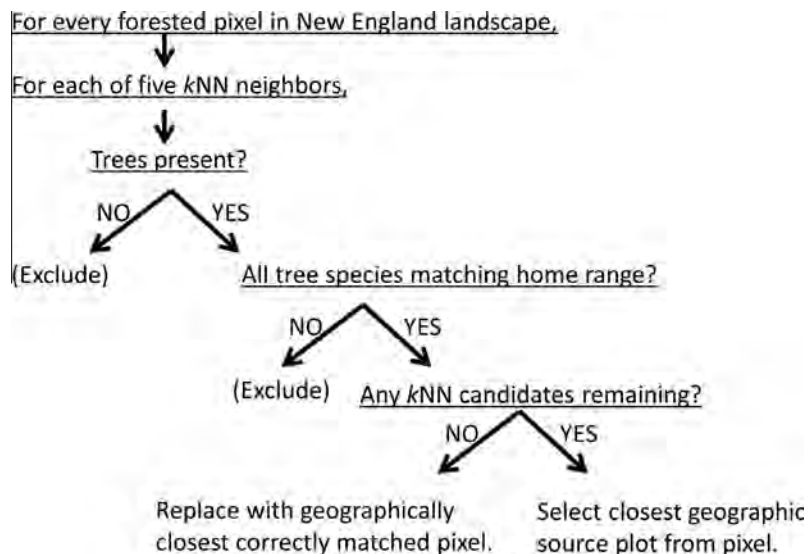


Fig. 2. Decision tree representing our source plot selection criteria for each pixel.

form and land cover in the imputation but the weakness of relying on a stochastic process for the final imputation. Other forest simulation modelers have developed ad hoc methods for initializing forest conditions for FLMs based on stochastically imputing source plots based on previously mapped forest types (Duveneck et al., 2014b; Scheller et al., 2008). These imputation methods include a stochastic component whereby source plots are randomly assigned to a given pixel within some constraint (e.g., a group of source plots matching pixel forest type are eligible for one to be stochastically assigned to that pixel).

The nearest neighbor (kNN) imputation method integrates spatial environmental data with satellite imagery to determine the nearest neighboring source plots across multiple spatial environmental variates for each map pixel. Generally, kNN involves multiple individual source plots (k) assessed for each pixel (Ohmann and Gregory, 2002). Nearest neighboring plots are determined from the lowest multi-dimensional distance of remote sensed and/or other spatial environmental variates (e.g., phenology, topography, and climate), defined as feature-space. Plots measured at each pixel can be equally weighted or weighted in proportion to their distance from source location to potential imputation location in feature-space. In the case of distance-weighted kNN, k_1 for each pixel equals the inventory plot whose distance from source location to potential imputation location in feature-space is smallest (i.e. the nearest neighbor). k_2 equals the source plot whose distance is the next smallest, and so on. Some imputation methods predict a single imputation source plot for each pixel (k_1), while other methods predict values for each pixel by aggregating information from multiple source plots ($k_{n>1}$) (Wilson et al., 2012). Averaged plot conditions are not useful for representing forest conditions used by a FLM as they require an explicit maintenance of the within stand species covariance structure, which is lost when source plots are averaged together.

Imputation methods that provide a single nearest neighbor source plot (i.e. smallest feature-space) include: most-similar-neighbor (MSN) which use a canonical correlation analysis distance metric (Moeur and Stage, 1995) and gradient-nearest-neighbor (GNN) which use a canonical correspondence analysis distance metric (Ohmann and Gregory, 2002). Single k_1 GNN (Spies et al., 2007; Thompson et al., 2011, 2006) and MSM (Hassani et al., 2004; Nothdurft et al., 2009) neighbor products have been used in forest landscape analyses. The single nearest neighbor approaches can be problematic because these imputation methods are typically applied over large areas, which presents a risk of source plots being imputed over long geographic distances, even though the distance from source plot to imputation pixel in feature-space is small (Grossmann et al., 2009; Hudak et al., 2008; Ohmann et al., 2014). Such long distance imputations can result in species being mapped outside their home range (Grossmann et al., 2009; Wilson et al., 2012). Given anticipated species range shifts due to climate change (Duveneck et al., 2014b; Iverson et al., 2008), reliable representation of species ranges at the start of simulations is necessary to understand species migration and other initial condition departures.

We developed and made available an imputation map for New England (Harvard Forest Data Archive #234). We constructed the map using a previously published nearest neighbor analysis using $k_{n>1}$ (Wilson et al., 2012). Our methodology is based on source plot similarity in the predictor variables of feature-space, species composition, and geographical distance between imputed source plots and raster cells. We validated map attributes at an ecoregional scale (Wang et al., 2014). We expect our regional map of imputed forest inventory plots to be useful for a myriad of New England forest landscape ecology projects that depend on reliable knowledge of current conditions and we believe the method could be applied to other regions.

2. Methods

2.1. Study area

Our study area includes the forested areas within the six New England states in the northeastern U.S. (Rhode Island, Connecticut, Massachusetts, New Hampshire, Vermont, and Maine) (Fig. 1). Total forest cover within this 18 million hectare region exceeds 80% but ranges from 50% (Rhode Island) to 90% (Maine). The forests span a diverse ecological gradient that includes spruce-fir forests in northern Maine to oak-maple and pitch pine forests in southern New England (Foster and Aber, 2004; Tang and Beckage, 2010). Elevations range from sea level to nearly 2000 m above sea level in the White Mountain Range in New Hampshire. The region experiences annual mean temperatures ranging from 3 to 10 °C (mean Jan temp = −6 °C; mean July

Table 1

Tree species used in analyses of final imputation map (k_n). SPP_CODE represents first two letters of genus followed by the first two letters of species and are used in Fig. 3.

COMMON_NAME	GENUS	SPECIES	SPP_CODE
balsam fir	<i>Abies</i>	<i>balsamea</i>	abba
red maple	<i>Acer</i>	<i>rubrum</i>	acru
sugar maple	<i>Acer</i>	<i>saccharum</i>	acsa
yellow birch	<i>Betula</i>	<i>alleghaniensis</i>	beal
sweet birch	<i>Betula</i>	<i>lenta</i>	bele
paper birch	<i>Betula</i>	<i>papyrifera</i>	bepa
gray birch	<i>Betula</i>	<i>populifolia</i>	bepo
pignut hickory	<i>Carya</i>	<i>glabra</i>	cagl
American beech	<i>Fagus</i>	<i>grandifolia</i>	fagr
white ash	<i>Fraxinus</i>	<i>americana</i>	fram
black ash	<i>Fraxinus</i>	<i>nigra</i>	frni
tamarack (native)	<i>Larix</i>	<i>laricina</i>	lala
eastern hophornbeam	<i>Ostrya</i>	<i>virginiana</i>	osvi
white spruce	<i>Picea</i>	<i>glauca</i>	pihl
black spruce	<i>Picea</i>	<i>mariana</i>	piha
red spruce	<i>Picea</i>	<i>rubens</i>	piru
red pine	<i>Pinus</i>	<i>resinosa</i>	pire
pitch pine	<i>Pinus</i>	<i>rigida</i>	piri
eastern white pine	<i>Pinus</i>	<i>strobus</i>	pist
balsam poplar	<i>Populus</i>	<i>balsamifera</i>	poba
bigtooth aspen	<i>Populus</i>	<i>grandidentata</i>	pogr
quaking aspen	<i>Populus</i>	<i>tremuloides</i>	potr
black cherry	<i>Prunus</i>	<i>serotina</i>	prse
white oak	<i>Quercus</i>	<i>alba</i>	qual
scarlet oak	<i>Quercus</i>	<i>coccinea</i>	quco
chestnut oak	<i>Quercus</i>	<i>prinus</i>	qupr
northern red oak	<i>Quercus</i>	<i>rubra</i>	quru
black oak	<i>Quercus</i>	<i>velutina</i>	quve
northern white-cedar	<i>Thuja</i>	<i>occidentalis</i>	thoc
American basswood	<i>Tilia</i>	<i>americana</i>	tiam
eastern hemlock	<i>Tsuga</i>	<i>canadensis</i>	tsca
American elm	<i>Ulmus</i>	<i>americana</i>	ulam

Table 2

State forests used in the independent validation. n = number of inventory plots in each forest site. Dissimilarity refers to the compositional dissimilarity (Bray–Curtis) between the inventory plots collected by the state forests and the imputed FIA plots within the boundary of that state forest. Similarly, the correlation refers to the Spearman correlation coefficient between species' basal area in the state forest plots and the imputed FIA plots within the boundary of that state forest.

STATE FOREST	n	DISSIMILARITY	CORRELATION
Warwick state forest	17	0.318	0.858
Arthur Wharton Swann state forest	17	0.313	0.773
Cookson state forest	17	0.431	0.748
Savoy Mountain state forest	20	0.201	0.845
Wendell state forest	22	0.319	0.798
Freetown-Fall River state forest	23	0.410	0.531
Kenneth Dubuque Memorial state forest	40	0.231	0.921
October Mountain state forest	52	0.340	0.844
Victory forest	130	0.305	0.934

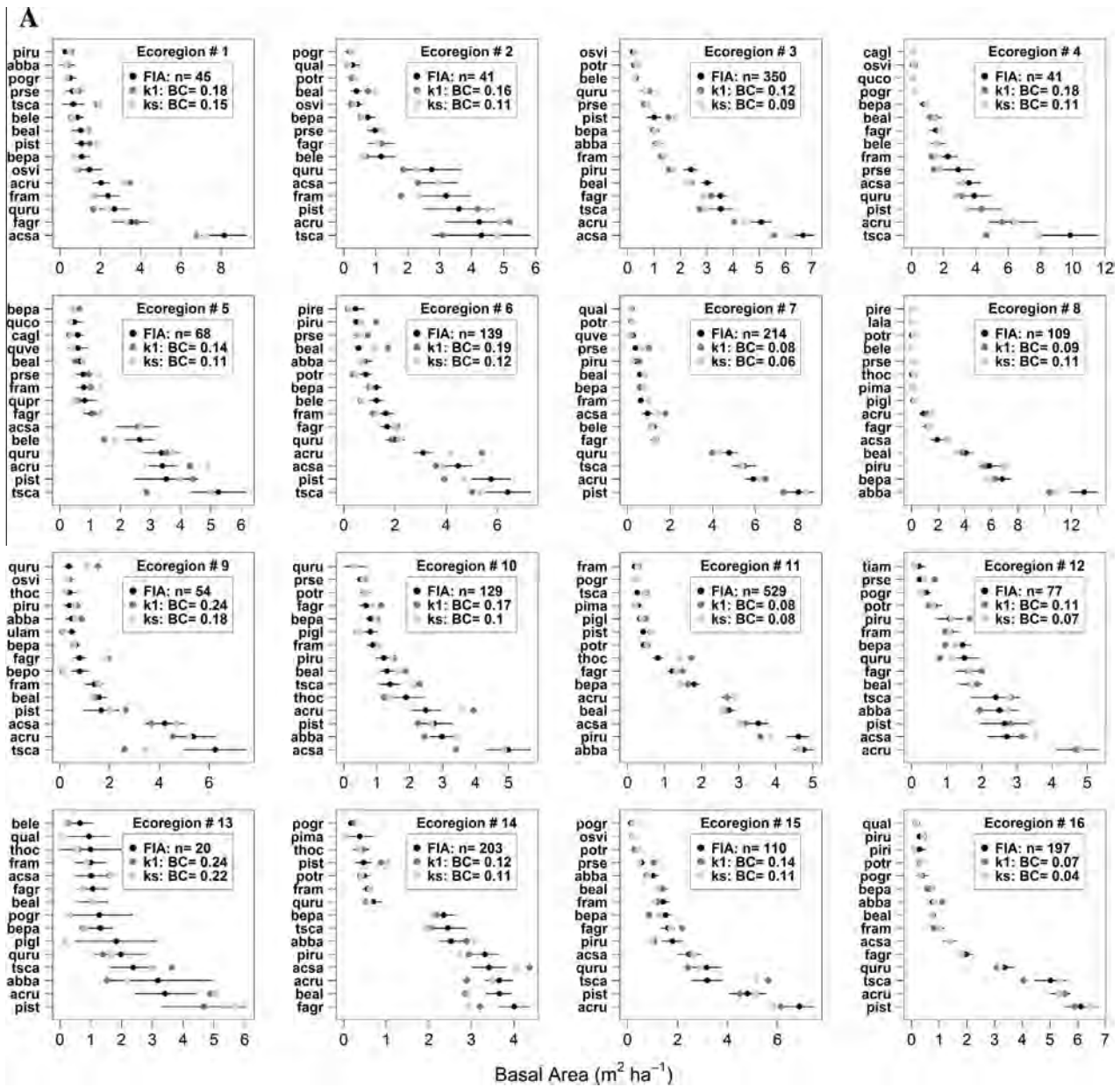


Fig. 3. Observed mean tree species basal area and standard error ($m^2 ha^{-1}$) of FIA plots (black dots and lines) in ecoregions: (A) 1 through 16, (B) 17 through 32, and (C) 33 through 40, compared to imputed pixels of the first neighbor map (k_1) (light gray), and the multiple k selection map (k_s) (dark gray). The number of FIA plots varied in each ecoregion (FIA n). Although all species were used in the dissimilarity calculation, only the 15 most abundant species ranked by basal area were plotted. Species codes correspond to species listed in Table 1. Ecoregion codes correspond to the map in Fig. 5.

temp = 19 °C), and average annual precipitation from 79 to 255 cm. The region was almost completely forested until the 17th C, when two centuries of logging and agricultural clearing removed more than half of the forest cover and cut-over most of the rest (Foster, 1992; Thompson et al., 2013). Forest cover reached its nadir in the mid nineteenth century, after which widespread farm abandonment and population concentration initiated a century of natural reforestation and forest growth. The modern landscape reached its apex of reforestation recently and is again experiencing a slow loss of forest cover in all six states (Foster et al., 2010).

2.2. Nearest-neighbor spatial data selection

We used an existing kNN imputation analysis recently conducted for the eastern United States using equal weighting and a canonical correspondence analysis distance metric (Wilson et al.,

2012). This process used vegetation phenology derived from MODIS imagery and spatial environmental variables to impute source plots from the FIA database to each 250-m pixel across the eastern United States (Fig. 1). Because the original kNN imputation products were derived for all of the eastern United States, all FIA plots from the eastern United States were considered as source plots in the Wilson et al. (2012) imputation. Source plots used in this imputation were measured between 2001 and 2006. For each 250-m pixel in our study area, we assembled the first five nearest-neighbor plots ($k_{1:5}$).

We built a screened-selection map (hereafter, k_s) by selecting the best source plot from the five nearest neighbors in feature-space for each 250-m pixel in our landscape as follows (Fig. 2). Because our final map product was focused on forests rather than other land uses, we included only the map pixels classified as forest within the National Land Cover Database (NLCD 2011) (Homer

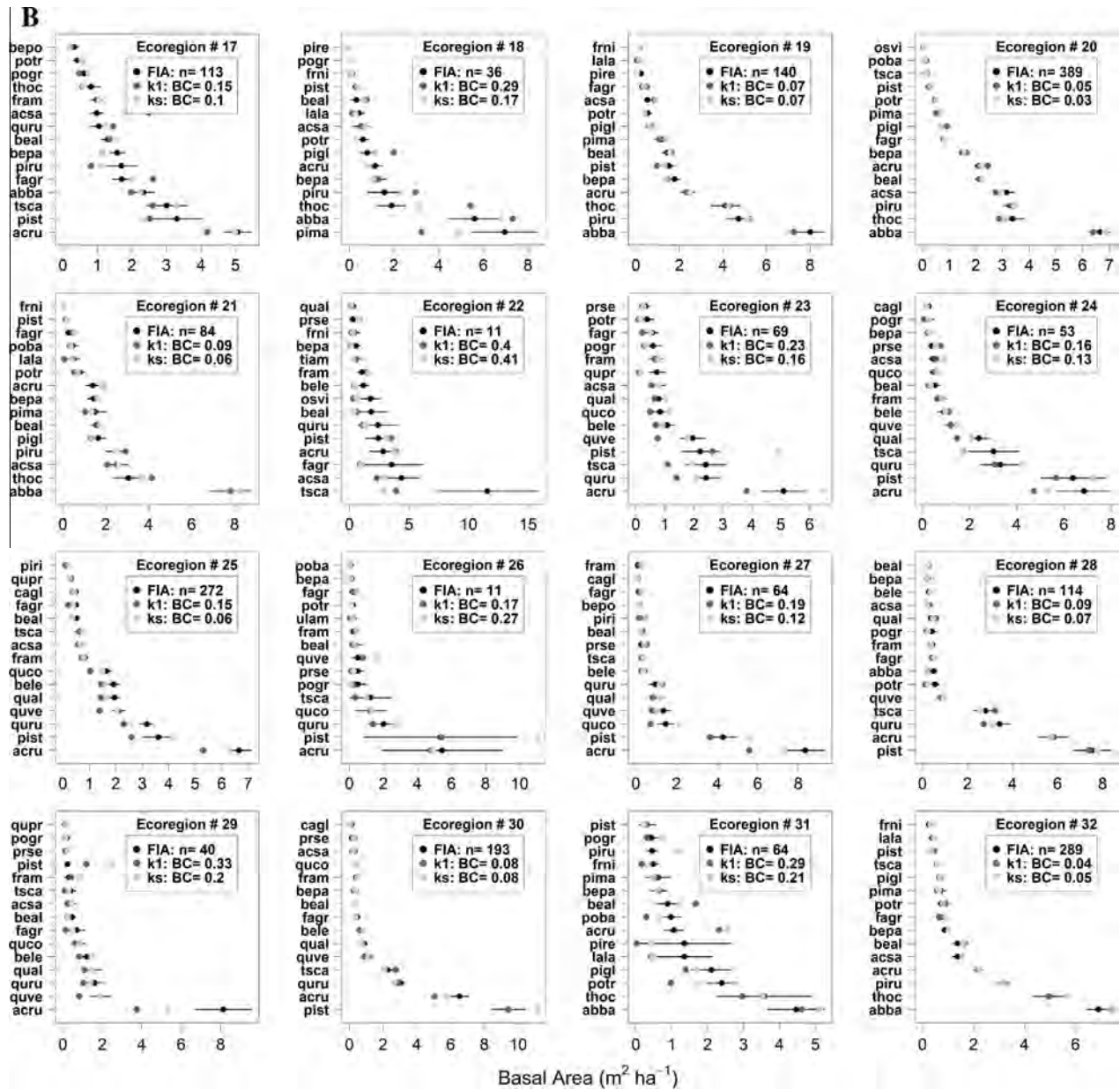


Fig. 3 (continued)

et al., 2012). For each pixel, we screened out source plots that did not include trees and source plots that included tree species that did not match historical species range distribution (Little, 1971). We screened species range distribution in order to reduce erroneous spatial representation of tree species. Of the remaining neighbors for each pixel, we choose the closest source plot based on geographic (Euclidian) distance from source plot to imputation pixel. If none of the five neighbors remained, we selected the geographically closest pixel following the screening process described. For evaluation purposes, we compared k_s to the first-neighbor map (k_1) where, each pixel was based on the nearest neighbor plot without being screened for species distribution, nor geographic distance.

2.3. Spatial imputation evaluation

We evaluated the selection protocol at a sub-regional scale. For validation, we utilized both a study-area-wide network of

inventory plots (not independent of source plots) within ecoregions, and a smaller, independent inventory plot network within state forests. For our ecoregion evaluation, we identified FIA field plots and associated tree data within 40 EPA level IV ecoregions in New England (Environmental Protection Agency, 2012). We compared average tree species (tree species listed in Table 1) basal area ($m^2 ha^{-1}$) of FIA field plots to average tree species basal area of the imputed source plots within each ecoregion. We obtained exact coordinates of FIA plots through a research agreement between Harvard University and the U.S. Forest Service, Northern Research Station. From the imputation map, we removed pixels where actual FIA plots were located in order to reduce circularity in the evaluation.

As an independent validation, we compared the species basal area distribution of empirically measured plot data from several state forests to our final imputation map. We sought sites in New England where inventory data, collected in a method comparable to the FIA, were publically available. Despite contacting state

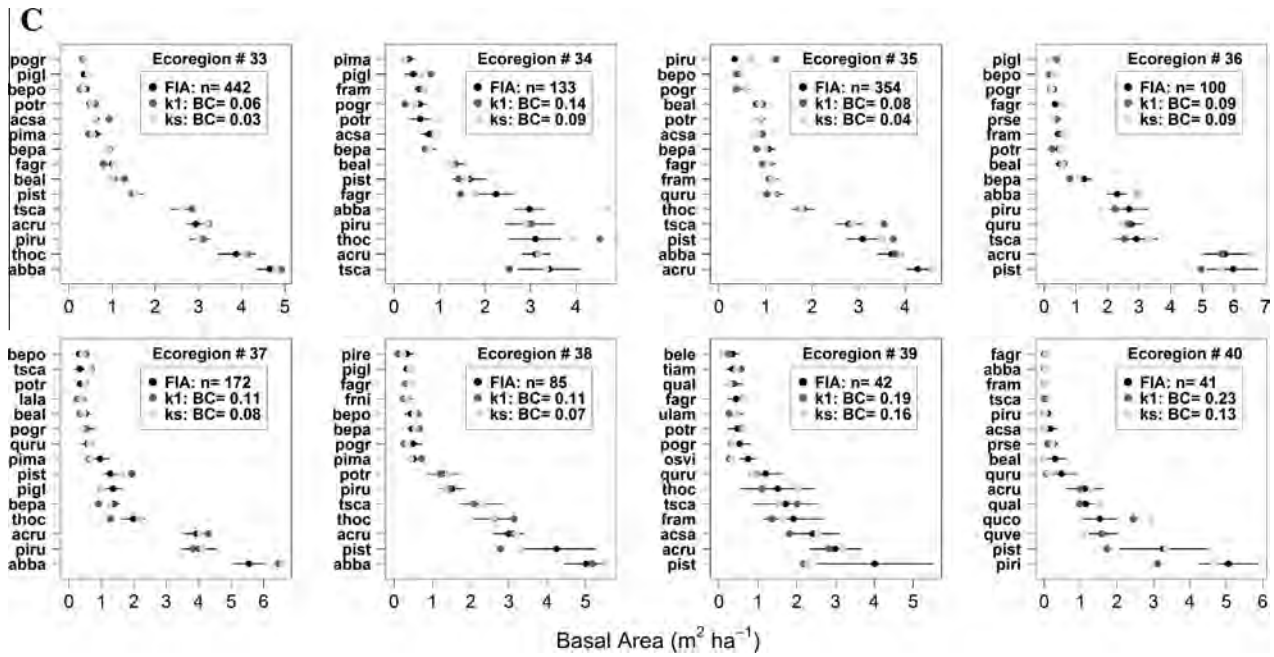


Fig. 3 (continued)

forests in all six states, we were only able to obtain data from Massachusetts and Vermont. State forests in Massachusetts were measured according to state continuous forest inventory protocols (Massachusetts Bureau of Forestry, 2014). The Victory State Forest in Vermont was measured using similar protocols. From these state forests, we compared the species distribution from the state forest inventory data to the aggregated imputed plots representing these forests.

Within ecoregions and state forests, we calculated Bray–Curtis community dissimilarity (Bray and Curtis, 1957) between the field inventory plots and the plot imputed pixels as a measure of imputation quality. Bray–Curtis dissimilarity is most strongly affected by species with high abundance, is not affected by joint absences, and is not overly affected by outliers (Quinn and Keough, 2002). Bray–Curtis dissimilarity values range between 0 and 1 with 0 being the most similar and 1 being the most dissimilar. In addition, we calculated the rank correlation (Spearman) of species basal area between inventory plots and imputed source plots as a measure of how well the order of species abundance was represented by the imputed map in both the ecoregion and state forest data sets.

We summarized the frequency of kNN neighbors ($k_{1:5}$) used in the final map in order to evaluate distribution of the source neighbors. In addition, we summarized the frequency distribution of geographic distance from source plots to imputation pixels in order to evaluate how distant the imputed plots were sourced. As a visual demonstration of the map, we built and plotted a forest type map of the landscape. To build the forest type map, we joined our final imputation map of FIA plot numbers with plot-assigned forest types (Bechtold and Patterson, 2005) and aggregated them according to Canham et al. (2013). All analyses were done using the raster (Hijmans, 2014), Vegan (Oksanen et al., 2013), and base packages in R (R Core Team, 2013).

3. Results

The imputation map included a decreasing selection frequency of the k_1 through k_5 neighbors. Specifically, the nearest neighbor in feature-space (k_1) was used most frequently in the final map

(21.2% of pixels) followed by k_2 (20.6%), k_3 (19.5%), k_4 (19.3%), and k_5 (19.3%) neighbors. In our species selection screen, only 0.6% of pixels were not assigned one of the $k_{1:5}$ neighbors (Fig. 2); these pixels were assigned the geographically nearest source plot. In our final map, 81% of pixels came from source plots within 50 km. 91% of pixels came from source plots within 100 km.

Overall, our screening algorithm produced a 250 m map (k_5) of imputed FIA source plots that captures regional gradients in tree species composition. Our state forest validation resulted in dissimilarity between inventory plots and imputation plots from 0.201 to 0.431 (Table 2). Spearman rank order correlation ranged from 0.531 to 0.934. State forest sites with larger sample sizes generally resulted in lower dissimilarity and higher rank order correlation (Table 2). In our ecoregion evaluation, k_5 largely captured the ecoregion variation in species composition as measured by the actual FIA plots across the region (Fig. 3). Average dissimilarity between FIA plots and the imputation map at the ecoregion scale was 0.12 (sd = 0.07). Average Spearman rank order correlation between FIA plots and imputed map was 0.93 (sd = 0.06) indicating that the order of species dominance was well represented by the imputation map in each ecoregion.

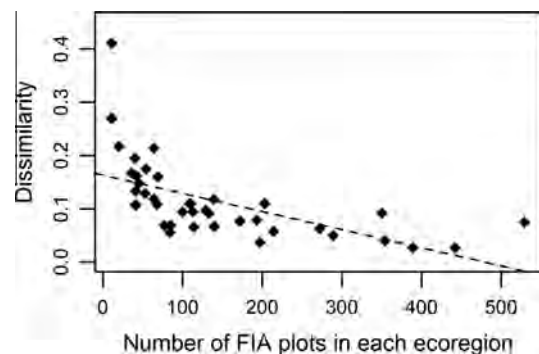


Fig. 4. Number of FIA plots compared to average Bray–Curtis dissimilarity in each ecoregion ($r^2 = 0.35$, $p < 0.05$).

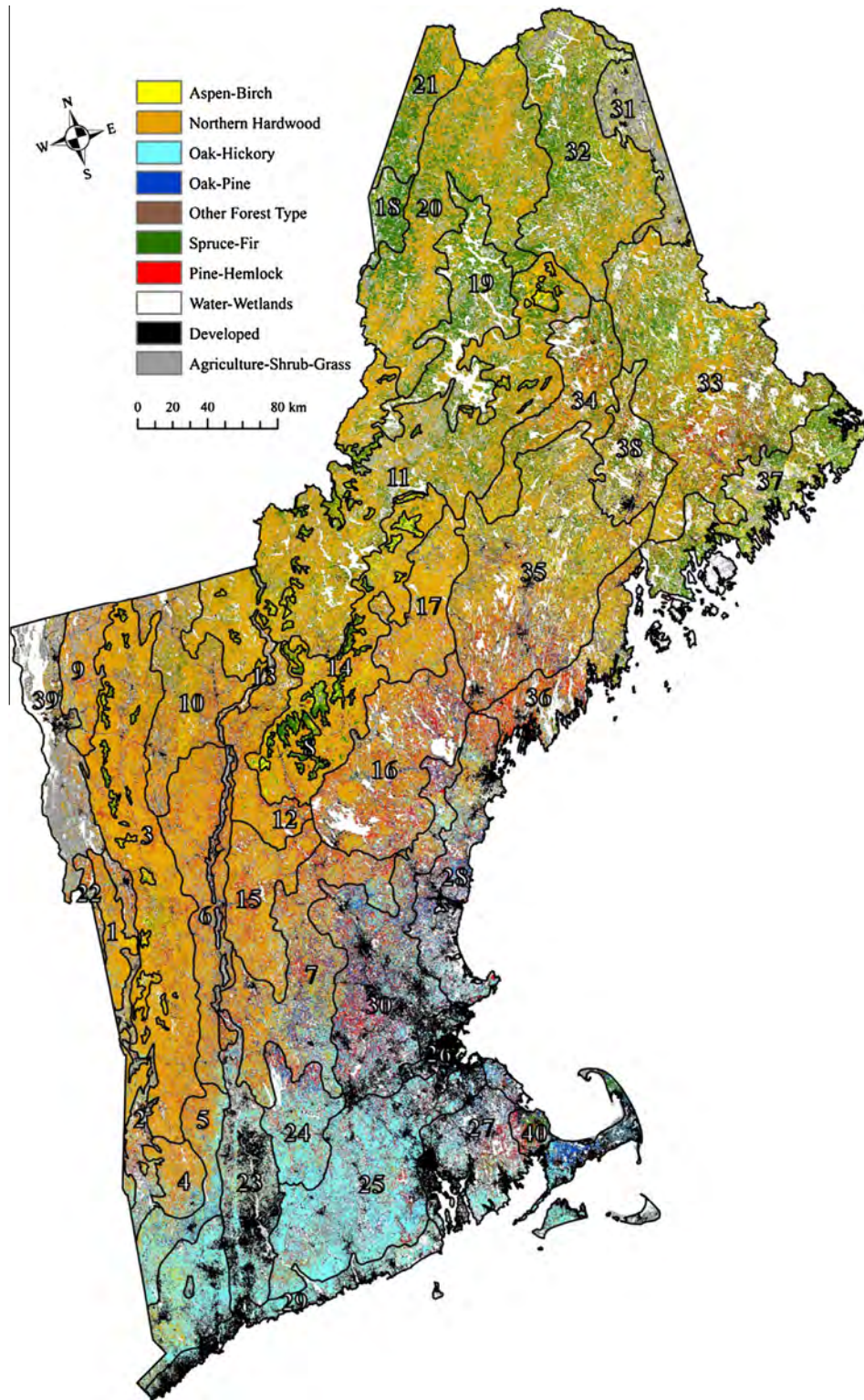


Fig. 5. Forest type assigned to k_s imputed FIA forest plots in New England. Numbers and black polygon boundaries correspond to ecoregions in Fig. 3. Non-forest classifications come from National Land Cover Database (2011).

The quality of the imputation was strongly related to the number of FIA plots measured in the field within each ecoregion. Dissimilarity and number of field plots within each ecoregion were significantly correlated (Fig. 4) ($p < 0.05$). For example, in the southern New Hampshire/North Central Massachusetts ecoregion (ecoregion 7), 214 FIA plots indicated low variance among FIA field

plots (as measured by low standard error) and low dissimilarity between FIA plots and our map (Dissimilarity = 0.06). Alternatively, in northwestern Vermont (ecoregion 39), 42 FIA field plots indicated higher variance among plots and higher dissimilarity between plots and our map (Dissimilarity = 0.16). Finally, Our forest type map, having extracted FIA ‘forest type’ for each pixel,

well-represents the broad ecological types and transitions throughout New England (Fig. 5).

4. Discussion

This imputation map (k_s) captures regional variation in species composition, as evidenced by the low dissimilarity between inventory plots and our imputation map and the high, rank order correlation of tree species. This was evident in both our smaller, state forest data and across ecoregions. Higher dissimilarity measurements in our smaller state forest data compared to the ecoregion evaluation may be the result of specific management regimes practiced at state forests, which is a factor not considered in the Wilson et al. (2012) algorithm. Also, the state forest data sets used different sample plot designs and field sampling protocols than those of the FIA program. Nevertheless, at state forests and ecoregions with larger sample sizes, our imputation map resulted in greater representation of inventory data.

Each pixel in k_s is associated with a single FIA source plot. In Fig. 5 we show the forest type associated with each pixel imputed FIA source plot. Users of our map can link to any FIA attribute measured at those plots (e.g., tree list, stem density, basal area, biomass, etc.).

The most significant contribution of our algorithm was the removal of source plots from pixels based on whether all species were within their historical range. This step reconciles the two major limitations for using KNN maps in FLMS—i.e. the simulation require a single forest condition and thus cannot use averaged conditions across $k_{n>1}$ plots and using k_1 can result in species being imputed outside their home ranges, which is unacceptable for any analysis of long term compositional change. We recognize that variation in precision and accuracy exists in the modern species distribution maps that we used (Little, 1971), and that in some cases species may be migrating from their historical ranges (e.g., Leithead et al., 2012). Nonetheless, we assume that climate change induced migration of tree species in recent decades is negligible. Finally, as compared to the k_1 map, k_s suggests a better representation of regional species composition as measured by reduced ecoregion dissimilarity to FIA plots.

There is considerable spatial variation in the quality of the imputation as measured by community dissimilarity. The variation in dissimilarity was also related to the number of field inventory plots within each ecoregion. Although our validation used field inventory as the known truth, we found greater variance in the field inventory estimate when fewer plots existed in a given ecoregion (Fig. 4), as would be expected from a sample of the population. In some cases, where the field plot density is very low, it is possible that the actual species distribution is better represented by our map than the FIA plots that we used for validation.

Although our neighbor screening process did not require that source plots come from a limited distance to pixels, we selected the closest neighbor in geographic space from the eligible neighbors. As the $k_{1:5}$ eligible neighbors were approximately equidistant in feature-space (Wilson et al., 2012), within the five nearest neighbors, we used Euclidean distance in geographic space over feature-space distance. This resulted in the majority of pixels represented by local/regional source plots. This further provides spatial relevance to k_s where otherwise pixels could have been imputed from long distances. Additionally, the frequency distribution of the $k_{1:5}$ neighbors in k_s puts confidence in the original KNN methods (Wilson et al., 2012).

It would be possible to create a more elaborate screening process, which may result in an improved imputation map. Indeed, we examined several alternative algorithms. For example, we screened neighbors matching 2011 NLCD forest types (Homer

et al., 2012) in addition to the species distribution screen. Compared to FIA forest types (e.g., sugar maple/beechn/yellow birch), NLCD forest types are more general (e.g., deciduous, coniferous, and mixed) and the native pixel resolution was 30 m compared to 250 m of the kNN maps. However, the use of NLCD screen resulted in greater ecoregion dissimilarity to FIA plots and more than 30% of pixels with no neighbor match resulting in target cell replacement. Ultimately, we rejected the NLCD screen and believe that the level of accuracy obtained with the existing algorithm achieves the right balance between accuracy and parsimony.

Our imputation map builds on and adds value to the kNN product previously developed for the region (Wilson et al., 2012). In addition, we introduce methods to further screen and evaluate an existing kNN imputation product for simulation and other applications. Future research may adapt our methods, or a derivation of these, to further provide and evaluate plot level data at regional scales. This imputation map will be valuable to many forest ecology applications in New England. In addition to forest composition change modeling, highly detailed forest maps with species specific attributes from individual plots such as this can be used to monitor habitat structure (e.g., Chandler et al., 2012; Djak et al., 2007; Nonaka et al., 2007) and spatially analyze biodiversity (e.g., Anderson and Ferree, 2010). Although one can create a plot level map of any attribute measured in the field and included in the FIA database, this comes with the caveat that only tree species composition has been validated in this paper. We expect this map to be used as a foundation for many applications that rely on spatial representation of plot-level data in New England (HF LTER, 2012; NECS, 2014; NH EPSCoR, 2011).

Acknowledgements

This research was supported in part by the National Science Foundation Harvard Forest Long Term Ecological Research Program (Grant No. NSF-DEB 12-37491) and the Scenarios Society and Solutions Research Coordination Network (Grant No. NSF-DEB-13-38809). This research was also made possible by MOU (14-MU-11242305-025) between Harvard University and USFS Northern Research Station, to use actual FIA plot locations in our analysis. Our sincere thanks to Brett Butler for his help developing this MOU. We also thank Sofie McComb and Luca Morreale for research assistance, and two anonymous reviewers for valuable feedback on an earlier version of this manuscript.

References

- Anderson, M.G., Ferree, C.E., 2010. Conserving the stage: climate change and the geophysical underpinnings of species diversity. *PLoS One* 5, 1–10.
- Bechtold, W.A., Patterson, P.L., 2005. The enhanced forest inventory and analysis program – national sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. Ashville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station.
- Bettinger, P., Lennette, M., Johnson, K.N., Spies, T.A., 2005. A hierarchical spatial framework for forest landscape planning. *Ecol. Modell.* 182, 25–48. <http://dx.doi.org/10.1016/j.ecolmodel.2004.07.009>.
- Bonan, G.B., Lawrence, P.J., Oleson, K.W., Levis, S., Jung, M., Reichstein, M., Lawrence, D.M., Swenson, S.C., 2011. Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data. *J. Geophys. Res.* 116, G02014. <http://dx.doi.org/10.1029/2010JG001593>.
- Bray, J.R., Curtis, J.T., 1957. An ordination of the upland forest communities of Southern Wisconsin. *Ecol. Monogr.* 27, 325–349. <http://dx.doi.org/10.2307/1942268>.
- Canham, C., Rogers, N., Buchholz, T., 2013. Regional variation in forest harvest regimes in the northeastern United States. *Ecol. Appl.* 23, 515–522.
- Chandler, C.C., King, D.L., Chandler, R.B., 2012. Do mature forest birds prefer early-successional habitat during the post-fledging period? *Forest Ecol. Manage.* 264, 1–9. <http://dx.doi.org/10.1016/j.foreco.2011.09.018>.
- Djak, W., 2013. Landscape Builder: software for the creation of initial landscapes for LANDIS from FIA data. *Comput. Ecol. Softw.* 3, 17–25.

- Dijak, W.D., Rittenhouse, C.D., Larson, M.A., Thompson, F.R., Millspaugh, J.J., 2007. Landscape habitat suitability index software. *J. Wildl. Manage.* 71, 668–670. <http://dx.doi.org/10.2193/2006-341>.
- Duveneck, M.J., Scheller, R.M., White, M.A., 2014a. Effects of alternative forest management on biomass and species diversity in the face of climate change in the northern Great Lakes region (USA). *Can. J. Forest Res.* 44, 700–710.
- Duveneck, M.J., Scheller, R.M., White, M.A., Handler, S.D., Ravenscroft, C., 2014b. Climate change effects on northern Great Lake (USA) forests: a case for preserving diversity. *Ecosphere* 5, 1–26.
- Environmental Protection Agency, 2012. Level IV ecoregions of EPA region 1. U.S. EPA Office of Research and Development (ORD) – National Health and Environmental Effects Research Laboratory (NHEERL), Corvallis, OR.
- Foster, D.R., 1992. Land-use history and vegetation dynamics in central New England, USA. *J. Ecol.* 80, 753–771.
- Foster, D.R., Aber, J.D. (Eds.), 2004. The environmental consequences of 1000 years of change in New England. *Forests in Time*. Yale University Press.
- Foster, D.R., Donahue, B., Kittredge, D.B., Fallon-Lambert, K., Hunter, M., Hall, B., Irland, L.C., Lillieholm, R.J., Orwig, D., D'Amato, A., Colburn, E., Thompson, J.R., Levitt, J.N., Ellison, A., Keeton, W.S., Aber, J., Cogbill, C., Driscoll, C.T., Fahey, T.J., Hart, C.M., 2010. *Wildland and Woodlands: A Forest Vision for New England*. Harvard University Press, Cambridge, MA.
- Grossmann, E., Ohmann, J., Gregory, M., May, H., 2009. Nationwide Forest Imputation Study (NaFIS)–western team final report. In: Final Rep. Nationwide Forest Imput. Study. <<http://www.fsl.orst.edu/lemma/main.php>>.
- Gustafson, E.J., 2013. When relationships estimated in the past cannot be used to predict the future: using mechanistic models to predict landscape ecological dynamics in a changing world. *Landsc. Ecol.* 28, 1429–1437. <http://dx.doi.org/10.1007/s10980-013-9927-4>.
- Hassani, B.T., LeMay, V., Marshall, P.L., Temesgen, H., Zumrawi, A.-A., 2004. Regeneration imputation models for complex stands of southeastern British Columbia. *Forest Chron.* 80, 271–278. <http://dx.doi.org/10.5558/ffc80271-2>.
- Haxeltine, A., Prentice, I.C.L., 1996. BIOME3: an equilibrium terrestrial biosphere model based on ecophysiological constraints, resource availability, and competition among plant functional types. *Global Biogeochem. Cycles* 10, 693–709. <http://dx.doi.org/10.1029/96GB02344>.
- He, H.S., 2008. Forest landscape models: definitions, characterization, and classification. *Forest Ecol. Manage.* 254, 484–498. <http://dx.doi.org/10.1016/j.foreco.2007.08.022>.
- HF LTER, 2012. Harvard Forest-Long Term Ecological Research Proposal V [WWW Document]. <<http://harvardforest.fas.harvard.edu/sites/harvardforest.fas.harvard.edu/files/publications/pdfs/LTERV-2012-proposal.pdf>> (accessed 06.11.14).
- Hijmans, R.J., 2014. Raster geographic data analysis and modeling. R Package Version, 2.2–12.
- Homer, C., Fry, J., Barnes, C., 2012. The national land cover database. US Geol. Surv. Fact Sheet.
- Hudak, A.T., Crookston, N.L., Evans, J.S., Hall, D.E., Falkowski, M.J., 2008. Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. *Remote Sens. Environ.* 112, 2232–2245. <http://dx.doi.org/10.1016/j.rse.2007.10.009>.
- Iverson, L.R., Prasad, A.M., Matthews, S.N., Peters, M., 2008. Estimating potential habitat for 134 eastern US tree species under six climate scenarios. *Forest Ecol. Manage.* 254, 390–406. <http://dx.doi.org/10.1016/j.foreco.2007.07.023>.
- Keane, R.E., Ryan, K.C., Running, S.W., 1996. Simulating effects of fire on northern rocky mountain landscapes with the ecological process model FIRE-BGC. *Tree Physiol.* 16, 319–331. <http://dx.doi.org/10.1093/treephys/16.3.319>.
- Keane, R.E., Cary, G.J., Davies, I.D., Flannigan, M.D., Gardner, R.H., Lavorel, S., Lenihan, J.M., Li, C., Rupp, T.S., 2004. A classification of landscape fire succession models: spatial simulations of fire and vegetation dynamics. *Ecol. Modell.* 179, 3–27. <http://dx.doi.org/10.1016/j.ecolmodel.2004.03.015>.
- Leithead, M., Silva, L.C.R., Anand, M., 2012. Recruitment patterns and northward tree migration through gap dynamics in an old-growth white pine forest in northern Ontario. *Plant Ecol.* 213, 1699–1714. <http://dx.doi.org/10.1007/s11258-012-0116-3>.
- Liang, Y., He, H.S., Wu, Z., Yang, J., 2014. Effects of environmental heterogeneity on predictions of tree species' abundance in response to climate warming. *Environ. Model. Softw.* 59, 222–231. <http://dx.doi.org/10.1016/j.envsoft.2014.05.025>.
- Little, E.L., 1971. Atlas of United States trees. In: Conifers and important hardwoods, vol. 1. U.S. Department of Agriculture Miscellaneous Publication 1146.
- Massachusetts Bureau of Forestry, 2014. Manual for continuous forest inventory field procedures. Massachusetts Dep. Conserv. Recreat., 117.
- Moeur, M., Stage, A.R., 1995. Most similar neighbor: an improved sampling inference procedure for natural resource planning. *Forest Sci.* 41, 337–359.
- Moorcroft, P., Hurtt, G., Pacala, S., 2001. A method for scaling vegetation dynamics: the ecosystem demography model (ED). *Ecol. Monogr.* 71, 557–585.
- NECSC, 2014. North East Climate Science Center [WWW Document]. <<http://necsc.umass.edu/>> (accessed 06.11.14).
- NH EPSCoR, 2011. NH EPSCoR [WWW Document]. <<http://nhepscor.org/changes-land-use-and-climate-affect-our-environment>> (accessed 06.11.14).
- Nonaka, E., Spies, T.A., Wimberly, M.C., Ohmann, J.L., Onaka, E.T.N., 2007. Historical range of variability in live and dead wood biomass: a regional-scale simulation study. *Can. J. Forest Res.* 37, 2349–2364. <http://dx.doi.org/10.1139/x07-064>.
- Nothdurft, A., Saborowski, J., Breidenbach, J., 2009. Spatial prediction of forest stand variables. *Eur. J. Forest Res.* 128, 241–251. <http://dx.doi.org/10.1007/s10342-009-0260-z>.
- Ohmann, J.L., Gregory, M.J., 2002. Predictive mapping of forest composition and structure with direct gradient analysis and nearest-neighbor imputation in coastal Oregon, USA. *Can. J. Forest Res.* 32, 725–741.
- Ohmann, J.L., Gregory, M.J., Roberts, H.M., 2014. Scale considerations for integrating forest inventory plot data and satellite image data for regional forest mapping. *Remote Sens. Environ.* 151, 3–15. <http://dx.doi.org/10.1016/j.rse.2013.08.048>.
- Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H., 2013. Vegan: community ecology package. R Package Version 2, 2.0–10 <<http://cran.r-project.org/package=vegan>>. R Packag. ver. 2.0–8>.
- Quinn, G.P., Keough, M.J., 2002. *Experimental Design and Data Analysis for Biologists*. Cambridge University Press, Cambridge, UK, New York.
- R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, <<http://www.R-project.org/>>.
- Scheller, R.M., Mladenoff, D.J., 2007. An ecological classification of forest landscape simulation models: tools and strategies for understanding broad-scale forested ecosystems. *Landsc. Ecol.* 22, 491–505. <http://dx.doi.org/10.1007/s10980-006-9048-4>.
- Scheller, R.M., Domingo, J.B., Sturtevant, B.R., Williams, J.S., Rudy, A., Gustafson, E.J., Mladenoff, D.J., 2007. Design, development, and application of LANDIS-II, a spatial landscape simulation model with flexible temporal and spatial resolution. *Ecol. Modell.* 201, 409–419. <http://dx.doi.org/10.1016/j.ecolmodel.2006.10.009>.
- Scheller, R.M., Van Tuyl, S., Clark, K., Hayden, N.G., Hom, J., Mladenoff, D.J., 2008. Simulation of forest change in the New Jersey Pine Barrens under current and pre-colonial conditions. *Forest Ecol. Manage.* 255, 1489–1500.
- Shugart, H.H., Saatchi, S., Hall, F.G., 2010. Importance of structure and its measurement in quantifying function of forest ecosystems. *J. Geophys. Res.* 115 <http://dx.doi.org/10.1029/2009JG000993>, G00E13.
- Spies, T.A., McComb, B.C., Kennedy, R.S.H., McGrath, M.T., Olsen, K., Pabst, R.J., 2007. Potential effects of forest policies on terrestrial biodiversity in a multi-ownership province. *Ecol. Appl.* 17, 48–65.
- Tang, G., Beckage, B., 2010. Projecting the distribution of forests in New England in response to climate change. *Divers. Distrib.* 16, 144–158. <http://dx.doi.org/10.1111/j.1472-4642.2009.00628.x>.
- Thompson, J.R., Johnson, K.N., Lennette, M., Spies, T.A., Bettinger, P., 2006. Historical disturbance regimes as a reference for forest policy in a multiowner province: a simulation experiment. *Can. J. Forest Res.* 36, 401–417. <http://dx.doi.org/10.1139/x05-247>.
- Thompson, J.R., Foster, D.R., Scheller, R., Kittredge, D., 2011. The influence of land use and climate change on forest biomass and composition in Massachusetts, USA. *Ecol. Appl.* 21, 2425–2444.
- Thompson, J.R., Carpenter, D.N., Cogbill, C.V., Foster, D.R., 2013. Four centuries of change in northeastern United States forests. *PLoS One* 8, e72540. <http://dx.doi.org/10.1371/journal.pone.0072540>.
- Wang, W.J., He, H.S., Spetich, M.A., Shifley, S.R., Thompson, F.R., Dijak, W.D., Wang, Q., 2014. A framework for evaluating forest landscape model predictions using empirical data and knowledge. *Environ. Model. Softw.* 62, 230–239. <http://dx.doi.org/10.1016/j.envsoft.2014.09.003>.
- Wilson, B.T., Lister, A.J., Riemann, R.I., 2012. A nearest-neighbor imputation approach to mapping tree species over large areas using forest inventory plots and moderate resolution raster data. *Forest Ecol. Manage.* 271, 182–198.