bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Changes in Forest Composition, Stem Density, and Biomass from the Settlement Era (1800s) to Present in the Upper Midwestern United States

Simon J. Goring[1]

David J. Mladenoff[2]

Charles V. Cogbill[3]

Sydne Record[3,4]

Christopher J. Paciorek[5]

Stephen T. Jackson[6]

Michael C. Dietze[7]

Andria Dawson [5]

Jaclyn Hatala Matthes[8]

Jason S. McLachlan[9]

John W. Williams[1,10]

[1]Department of Geography, University of Wisconsin, Madison, 550 N Park St, Madison WI 53706

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[2]Department of Forest and Wildlife Ecology, University of Wisconsin-Madison, 1630 Linden Dr, Madison WI 53706

[3]Harvard Forest, Harvard University, 324 N Main St, Petersham MA 01366

[4]Department of Biology, Bryn Mawr College, 101 North Merion Ave., Bryn Mawr PA 19010

[5]Department of Statistics, University of California, Berkeley, 367 Evans Hall, Berkeley CA 94720

[6]Department of the Interior Southwest Climate Science Center, U.S. Geological Survey, 1955 E. Sixth St. Tucson, AZ 85719; School of Natural Resources and the Environment and Department of Geosciences, University of Arizona, Tucson AZ 85721

[7]Department of Earth and Environment, Boston University, 685 Commonwealth Ave, Boston, MA 02215

[8]Department of Geography, Dartmouth College, 6017 Fairchild, Hanover, NH 03755

[9]Department of Biological Sciences, University of Notre Dame, 100 Galvin Life Sciences Center, Notre Dame, IN 46556

[10]Center for Climatic Research, University of Wisconsin, Madison, 1225 W Dayton St, Madison WI 53706

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

EuroAmerican land use and its legacies have transformed forest structure and composition across the United States (US). More accurate reconstructions of historical states are critical to understanding the processes governing past, current, and future forest dynamics. Gridded (8x8km) estimates of pre-settlement (1800s) forests from the upper Midwestern US (Minnesota, Wisconsin, and most of Michigan) using 19th Century Public Land Survey (PLS) records provide relative composition, biomass, stem density, and basal area for 26 tree genera. This mapping is more robust than past efforts, using spatially varying correction factors to accommodate sampling design, azimuthal censoring, and biases in tree selection.

We compare pre-settlement to modern forests using Forest Inventory and Analysis (FIA) data, with respect to structural changes and the prevalence of lost forests, pre-settlement forests with no current analogue, and novel forests, modern forests with no past analogs. Stem density, basal area and biomass are higher in contemporary forests than in settlement-era forests, but this pattern is spatially structured. Modern biomass is higher than pre-settlement biomass in the northwest (Minnesota and northern Wisconsin), and lower in the east, due to shifts in species composition and, presumably, average stand age. Modern forests are more homogeneous, and ecotonal gradients are more diffuse today than in the past. Novel forest represent 29% of all FIA cells, while 25% of pre-settlement forests no longer exist in a modern context.

Lost forests are centered around the forests of the Tension Zone, particularly in hemlock dominated forests of north-central Wisconsin, and in oak-elm-basswood forests along the forest-prairie boundary in south central Minnesota and eastern Wisconsin. Novel FIA forest

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

assemblages are distributed evenly across the region, but novelty shows a strong relationship to spatial distance from remnant forests in the upper Midwest, with novelty predicted at between 20 to 60km from remnants, depending on historical forest type.

The spatial relationships between remnant and novel forests, shifts in ecotone structure and the loss of historic forest types point to significant challenges to land managers if landscape restoration is a priority in the region. The spatial signals of novelty and ecological change also point to potential challenges in using modern spatial distributions of species and communities and their relationship to underlying geophysical and climatic attributes in understanding potential responses to changing climate. The signal of human settlement on modern forests is broad, spatially varying and acts to homogenize modern forests relative to their historic counterparts, with significant implications for future management.

Key Words: euroamerican settlement, land use change, public land survey, historical ecology, novel ecosystems, biomass, forest inventory and analysis, ecotone, forest ecology

Introduction:

The composition, demography, and structure of forests in eastern North America have changed continuously over the last millennium, driven by human land use (Foster et al. 1998, Ramankutty and Foley 1999, Ellis and Ramankutty 2008, Thompson et al. 2013, Munoz et al. 2014) and climate variability (Umbanhowar et al. 2006, Hotchkiss et al. 2007, Booth et al. 2012, Pederson et al. 2014). While human effects have been a component of these systems for millenia, the EuroAmerican settlement and industrialization period have increased anthropogenic effects by orders of magnitude (Brugam 1978, McAndrews 1988,

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Fuller et al. 1998). Legacies of post-settlement land use in the upper Midwest (Grossmann and Mladenoff 2008) and elsewhere have been shown to persist at local and regional scales (Foster et al. 1998, Dupouey et al. 2002, Etienne et al. 2013), and nearly all North American forests have been affected by the intensification of land use in the past three centuries. Hence, contemporary ecological processes in North American forests integrate the anthropogenic impacts of the post-EuroAmerican period and natural influences at decadal to centennial scales.

At a regional scale many forests in the upper Midwest (*i.e.*, Minnesota, Wisconsin and Michigan) now have decreased species richness and functional diversity relative to forests of the pre-EuroAmerican settlement (hereafter pre-settlement) period (Schulte et al. 2007, Hanberry et al. 2012a, Li and Waller 2014) due to near complete logging. For example, forests in Wisconsin are in a state of regrowth, with an unfilled carbon sequestration potential of 69 TgC (Rhemtulla et al. 2009a) as a consequence of these extensive land cover conversions and subsequent partial recovery following abandonment of farm lands in the 1930s.

Legacies of land use are unavoidable at regional scales (Foster et al. 2003). Under intensive land use change the natural processes of secession, senescense and the replacement of tree species in forests may be masked, or heavily modified by historically recent land use change. These changes can result in non-stationarity within ecosystems that may not be apparent on the relatively narrow time scales at which ecology traditionally operates (Wolkovich et al. 2014). There is a history of recolonization of forested landscapes following agricultural clearance in the upper Midwest (Rhemtulla et al. 2009b), pointing to

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

the importance of understanding ecological trajectories and land use legacies in understanding modern forest dynamics (Foster et al. 2003). Cramer et al. (2008) point to the literature of succession theory to indicate the likelihood that many old fields will return to a 'natural' state, but point out that recovery is not universal. In particular, intense fragmentation of the landscape can deplete the regional species pool, leading to failures of recruitment that would favor species with longer distance seed dispersersal (Bellemare et al. 2002). In the upper Midwest long seed dispersal would favor species such as poplar (*Populus* sp.), white birch (*Betula papyrifera*) and some maple species (*Acer* sp.), at the expense of large-seeded species such as walnut (*Juglans* sp.), oak (*Quercus* sp.) and others.

While there remains debate over the utility of the concept of novel ecosystems (Hobbs et al. 2014a, Murcia et al. 2014), the fact remains that there are now forest and vegetation communities on the landscape without past analogues. The long term management of the systems and their associated services requires a broad understanding of the extent to which landscapes have been modified, and the extent to which land use change has potenitally masked underlying processes. It also requires a better understanding of the spatial (and temporal) scales at which novel ecosystems operate. While much restoration effort has focused on ecosystems at local scales there is an increasing need to focus on management and restoration at landscape scales (Menz et al. 2013). Thus a better understanding of the landscape-scale processes driving novelty, the spatial structure of novel ecosystems and their ecological correlates, is increasingly important. It can help prioritize intervention strategies at local scales (Hobbs et al. 2014b), and give us a better understanding of the role of patches in restoring hybrid or novel landscapes. In particular, how important is the species pool to the development of novel landscapes? Are novel

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

forests further from remnant forests than might otherwise be expected? Is novelty operating at landscape scales in the upper Midwest, and is the spatial distribution of new forests tied to historical patterns vegetation or losses of forest types from the historical landscape?

The upper Midwestern United States represents a unique ecological setting, with multiple major ecotones, including the prairie-forest boundary, historic savanna, and the Tension Zone between southern deciduous forests and northern evergreen forests. The extent to which these ecotones have shifted, and their extent both prior to and following EuroAmerican settlement is of critical importance to biogeochemical and biogeophysical vegetation-atmosphere feedbacks (Matthes et al. in review), carbon sequestration (Rhemtulla et al. 2009a), and regional management and conservation policy (Radeloff et al. 2000, Fritschle 2008, Knoot et al. 2010, Gimmi and Radeloff 2013).

Land use change at the local and state-level has affected both the structure and composition of forests in the Midwestern United States (*e.g.* Schulte et al. 2007, Hanberry et al. 2012a). Homogenization and shifts in overall forest composition are evident, but the spatial extent and structure of this effect is less well understood. Studies in Wisconsin have shown differential patterns of change in the mixedwood and evergreen dominated north versus the southern driftless and hardwood south. Does this pattern of differential change extend to Minnesota and Michigan? To what extent are land-use effects common across the region, and where are responses ecozone-specific? Has homogenization (*e.g.*, Schulte et al. 2007) resulted in novel forest assemblages relative to pre-settlement baselines across the region, and the loss of pre-settlement forest types? Are the spatial distributions of these

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

novel and lost forest types overlapping, or do they have non-overlapping extents? If broad-scale reorganization is the norm following EuroAmerican settlement, then the ecosystems that we have been studying for the past century may indeed be novel relative to the reference conditions of the pre-settlement era.

Modern forest structure and composition data (*e.g.*, from the United States Department of Agriculture Forest Service's Forest Inventory and Analysis National Program, FIA; Gray et al. 2012) play a ubiquitous role in forest management, conservation, carbon accounting, and basic research on forest ecosystems and community dynamics. These recent surveys (the earliest FIA surveys began in the 1930s) can be extended with longer-term historical data to understand how forest composition has changed since EuroAmerican settlement. The Public Land Survey was carried out ahead of mass EuroAmerican settlement west and south of Ohio to provide for delineation and sale of the public domain beyond the original East Coast states (Stewart 1935, White 1983). Because surveyors used trees to locate survey points, recording the identity, distance, and directory of two to four trees next to each survey marker, we can make broad-scale inferences about forest composition and structure in the United States prior to large-scale EuroAmerican settlement (Almendinger 1996, Liu et al. 2011, Williams and Baker 2011, Tomscha and Gergel 2014). In general, FIA datasets are systematically organized and widely available to the forest ecology and modeling community, whereas most PLS data compilations are of local or, at most, state-level extent. This absence of widely available data on settlement-era forest composition and structure limits our ability to understand and model the current and future processes governing forest dynamics at broader, regional scales. For example, distributional models of tree species often rely upon FIA or other contemporary observational data to build

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

species-climate relationships that can be used to predict potential range shifts (Iverson and Prasad 1998, Iverson and McKenzie 2013).

Here we use survey data from the original Public Lands Surveys (PLS) in the upper Midwest to derive estimates of pre-settlement (*ca*. mid-late 1800s) forest composition, basal area, stem density, and biomass. This work builds upon prior digitization and classification of PLS data for Wisconsin (Manies and Mladenoff 2000, Schulte et al. 2002) and for parts of Minnesota (Friedman and Reich 2005, Hanberry et al. 2012a) and Michigan Michigan (USFS-NCRS http://www.ncrs.fs.fed.us/gla/). Most prior PLS-based reconstructions are for individual states or smaller extents (among others: Duren et al. (2012); Hanberry et al. (2012a); Rhemtulla et al. (2009a); Friedman and Reich (2005)] often aggregated at the scale of regional forest zones (Schulte et al. 2007, Hanberry et al. 2012a), although aggregation may also occur at the section (Rhemtulla et al. 2009a) or township scale (Kronenfeld et al. 2010). Our work develops new approaches to address major challenges to PLS data, including lack of standardization in tree species names, azimuthal censoring by surveyors, variations in sampling design over time, and differential biases in tree selection among different kinds of survey points within the survey design at any point in time. The correction factors developed here are spatially varying, allowing us to accommodate temporal and spatial variations in surveyor methods.

We aggregate point based estimates of stem density, basal area and biomass to an 8 x 8km grid, and classify forest types in the upper Midwest to facilitate comparisons between FIA and PLS data. We compare the PLS data to late-20th-century estimates of forest composition, tree stem density, basal area and biomass. We explore how forest

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

homogenization has changed the structure of ecotones along two major ecotones from southern deciduous to northern evergreen forests and to the forest-prairie boundary. Using analog analyses, we identify lost forests that have no close compositional counterpart today and novel forests with no close historical analogs. This work provides insight into the compositional and structural changes between historic and contemporary forests, while setting the methodological foundation for a new generation of maps and analyses of settlement-era forests in the Eastern US.

Methods:

Public Lands Survey Data: Assembly, and Standardization

The PLS was designed to facilitate the division and sale of federal lands from Ohio westward and south. The survey created a 1 mile^2 (2.56 km^2) grid (sections) on the landscape. At each section corner, a stake was placed as the official location marker. To mark these survey points, PLS surveyors recorded tree stem diameters, measured distances and azimuths of the two to four trees 'closest'to the survey point and identified tree taxa using common (and often regionally idiosyncratic) names. PLS data thus represent measurements by hundreds of surveyors from 1832 until 1907, with changing sets of instructions over time (Stewart, 1979).

The PLS was undertaken to survey land prior to assigning ownership (Stewart 1935, White 1983), replacing earlier town proprietor surveys (TPS) used for the northeastern states (Cogbill et al. 2002, Thompson et al. 2013). The TPS provided estimates of relative forest composition at the township level, but no structural attributes. The PLS produced spatially

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

explicit point level data, with information about tree spacing and diameter, that can be used to estimate absolute tree density and biomass. PLS notes include tree identification at the plot level, disturbance (Schulte and Mladenoff 2005) and other features of the pre-settlement landscape. However, uncertainties exist within the PLS and township level dataset (Bourdo 1956).

Ecological uncertainty in the PLS arises from the dispersed spatial sampling design (fixed sampling every 1 mile), precision and accuracy in converting surveyor's use of common names for tree species to scientific nomenclature (Mladenoff et al. 2002), digitization of the original survey notes, and surveyor bias during sampling (Bourdo 1956, Manies et al. 2001, Schulte and Mladenoff 2001, Liu et al. 2011). Estimates vary regarding the ecological significance of surveyor bias. Terrail *et al*. (2014) show strong fidelity between taxon abundance in early land surveys versus old growth plot surveys. Liu *et al* (2011) estimate the ecological significance of some of the underlying sources of bias in the PLS and show ecologically significant (>10% difference between classes) bias in species and size selection for corner trees. However Liu *et al*. (2011) also indicate that the true sampling error cannot be determined, particularly because most of these historic ecosystems are largely lost to us.

Kronenfeld and Wang (2007), working with historical land cover datasets in western New York indicate that direct estimates of density using plotless estimators may be off by nearly 37% due to azimuthal censoring (*i.e.*, the tendency of surveyors to avoid trees close to cardinal directions), while species composition estimates may be adjusted by between -4 to +6%, varying by taxon, although Kronenfeld (2014) shows adjustments of less than 1%. These biases can be minimized by appropriate analytical decisions; many efforts over the

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

years have assessed and corrected for the biases and idiosyncrasies in the original surveyor data (Manies et al. 2001, Kronenfeld and Wang 2007, Bouldin 2008, Hanberry et al. 2011, 2012a, 2012b, Liu et al. 2011, Williams and Baker 2011, Cogbill et al. in prep). And, even given these caveats, PLS records remain the best source of data about both forest composition and structure in the United States prior to EuroAmerican settlement.

This analysis builds upon and merges previous state-level efforts to digitize and database the point-level PLS data for Wisconsin, Minnesota and the Upper Peninsula and upper third of the Lower Peninsula of Michigan. These datasets were combined using spatial tools in R (package *rgdal*: Bivand et al. 2014, Team 2014) to form a common dataset for the upper Midwest (Figure 1) using the Albers Great Lakes and St Lawrence projection (see code in Supplement 1, file: *step_one_clean_bind.R*; proj4: *+init:EPSG:3175*).

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

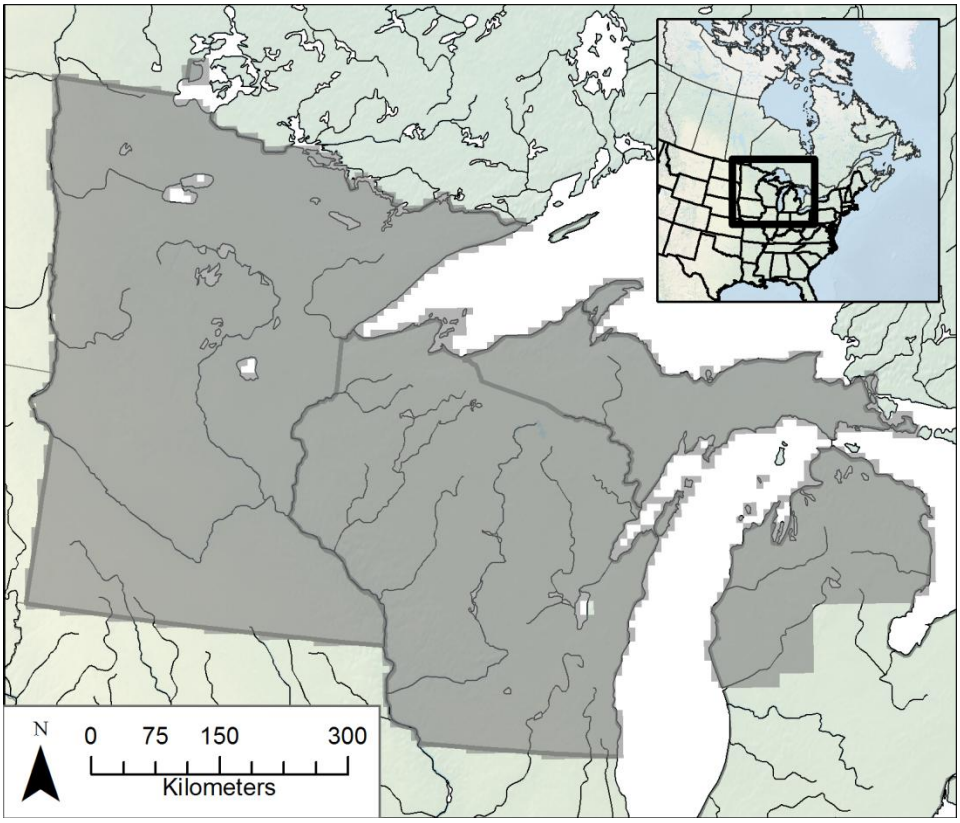

Figure 1. *The domain of the Public Land Survey investigated in this study. The broad domain includes Minnesota, Wisconsin and the upper two thirds of Michigan state. A 8x8km grid is superimposed over the region to aggregate data, resulting in a total of 7940 cells containing data.*

We took several steps to standardize the dataset and minimize the potential effects of surveyor bias upon estimates of forest composition, density, and biomass. All steps are preserved in the supplementary R code (Supplement 1: *step_one_clean_bind.R*). First, we excluded line and meander trees (i.e. trees encountered along survey lines, versus trees located at section or quarter corners) because surveyor selection biases appear to have been more strongly expressed for line trees, meander trees have non-random habitat preferences (Liu et al. 2011), and the inherent differences in sampling design between line, meander and corner points. We used only the closest two trees at each corner point

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

because the third and fourth furthest trees have stronger biases with respect to species composition and diameter (Liu et al. 2011). Corner points were used only if 1) there were at least two trees at a survey point, 2) the two trees were from different quadrants (defined by the cardinal directions), and 3) there were valid azimuths to the trees (a defined quadrant with an angle between 0 and 90) and valid diameters (numeric, non-zero).

Many species-level identifications used by PLS surveyors are ambiguous. Statistical models can predict the identity of ambiguous species (Mladenoff et al. 2002), but these models introduce a second layer of uncertainty into the compositional data, both from the initial surveyors' identification, and from the statistical disambiguation. Given the regional scale of the analysis, and the inherent uncertainty in the survey data itself, we chose to avoid this layer of taxonomic uncertainty, and retained only genus-level identification (Supplement 2, *Standardized Taxonomy*). The ecological implications for the use of genera-level taxonomies are important for this region. While fire tolerance is fairly well conserved within genera, shade tolerance can vary. *Betula* contains shade intolerant *B. paperyfera* and the intermediate *B. alleghaniensis*, while *Pinus* contains the very shade intolerant *P. banksana*, the intolerant *P. resinosa* and the shade tolerant *P. strobus*. For cases where shade tolerance (of fire tolerance) varies strongly within a genera we examine the data to determine the suitability of the assignment, or extent of confusion within the assigned genera.

In areas of open prairie or other treeless areas, *e.g.* southwestern Minnesota, surveyors recorded distances and bearings to 'Non Tree' objects. When points were to be located in water bodies the point data indicates 'Water'. Points recorded "No Tree" are considered to

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

have been from extremely open vegetation, with an estimated point-level stem density of 0 stems ha^{-1}. We based our estimates on terrestrial coverage, so water cells are excluded completely. Hence, absence of trees at "No Tree"€™ locations does reduce the gridded estimates of terrestrial stem density, but absence of trees at 'Water' locations does not.

Digitization of the original surveyor notebooks introduces the possibility of transcription errors. The Wisconsin dataset was compiled by the Mladenoff lab group, and has undergone several revisions over the last two decades in an effort to provide accurate data (Manies and Mladenoff 2000, Radeloff et al. 2000, Mladenoff et al. 2002, Schulte et al. 2002, Liu et al. 2011). The Minnesota transcription error rate is likely between 1 and 5%, and the treatment of azimuths to trees varies across the dataset (Almendinger 1996). Michigan surveyor observations were transcribed to mylar sheets overlaid on State Quadrangle maps, so that the points were displayed geographically, and then digititized to a point based shapefile (Ed Schools, pers. comm.; Great Lakes Ecological Assessment. USDA Forest Service Northern Research Station. Rhinelander, WI. http://www.ncrs.fs.fed.us/gla/), carrying two potential sources of transciption error. Preliminary assessment of Southern Michigan data indicates a transcription error rate of 3 - 6%. To reduce errors associated with transcription across all datasets, we exclude sites for which multiple large trees have a distance of 1 link (20.12 cm) to plot center, trees with very large diameters (diameter at breast height - dbh > 100 in; 254 cm), plots where the azimuth to the tree is unclear, and plots where the tree is at plot center but has a recorded azimuth. All removed plots are documented in the code used for analysis (Supplement 1: *step_one_clean_bind.R*) and are commented for review.

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Data Aggregation

We binned the point data using an 64km^2 grid (Albers Gt. Lakes St Lawrence projection; Supplement 1: *base_calculations.R*) to create a dataset that has sufficient numerical power for spatial statistical modeling and sufficient resolution for regional scale analysis (**???**). This resolution is finer than the 100km^2 gridded scale used in Freidman and Reich (2005), but coarser than township grids used in other studies (Rhemtulla et al. 2009a, Kronenfeld 2014) to provide a scale comparable to aggregated FIA data at a broader scale. Forest compositional data is based on the number of individuals of each genus or plant functional type (PFT) present at all points within a cell. Stem density, basal area and biomass are averaged across all trees at all points within the cell.

Stem Density

Estimating stem density from PLS data is based on a plotless density estimator using the measured distances from each survey point to the nearest trees (Morisita 1957, Persson 1971). This Morisita density estimator is then modified to minimize error due to different sampling geometries and several known surveyor biases (Manies et al. 2001, Kronenfeld and Wang 2007, Bouldin 2008, Hanberry et al. 2011, 2012a, 2012b, Liu et al. 2011, Williams and Baker 2011, Cogbill et al. in prep). Survey sampling instructions changed throughout the implementation of the PLS in this region and differed between section and quarter section points and between internal and external points within a township (White 1983, Liu et al. 2011). Our approach allows for spatial variation in surveyor methods by applying various spatially different correction factors based not only on the empirical

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

sample geometry, but also on known surveyor biases deviating from this design (Cogbill et al. in prep).

We estimate stem density (stems m^{-2}) based on a on a modified form of the Morisita two-tree density estimator, which uses the distance-to-tree measurements for the two closest trees at each point (Morisita 1954). Our modified form uses explicit and spatially varying correction factors, modeled after the Cottam correction factor (Cottam and Curtis 1956), that account for variations in sampling designs over time and among surveyors. All code to perform the analysis is included in Supplement 1.

We estimate the basic stem density (stems m^{-2}) using the point-to-tree distances for the closest trees to each point within a defined number of sectors around the point (Morisita (1957) eqn 31.):

$$\lambda\hat{m}_2 = \frac{k-1}{\pi \times n} \times \sum_{i=1}^{N} \frac{k}{\sum_{j=1}^{k} (r_{ij})^2} \quad (1)$$

where λ is density ; k is the number of sectors within which trees are sampled, N is the number of points over which estimates are aggregated, r is the distance of point-to-tree (as m). This estimate can be modified by a refinement of the Cottam quadrant factors (Morisita 1954, Cottam and Curtis 1956) which recognizes that different sampling designs, and the order of the distances in different quadrants (or sectors) carry specific weights. This correction, herein called κ, accounts for different sampling designs. When either four quadrants or trees are sampled (point quarter design), or when two trees in opposite semicircles (point halves design) are sampled, the equation is accurate and κ = 1; when the two trees are in the nearest of two quadrants (two nearest quadrants design), κ = 0.857;

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and when two trees are in quadrants on the same side of the direction of travel (one-sided or interior half design), $\kappa = 2$. This parameter, in Cottam's notation (Cottam and Curtis 1956), is a divisor of the denominator above, or here, the mathematically equivalent multiplier in the numerator of the reciprocal of the squared distances.

We further simplify the density estimate in equation (1) so that it is calculated at each point (N=1) and for two sample trees only (k=2):

$$\lambda_M = \frac{2}{\pi \times \sum_{j=1}^{2} r_j{}^2}$$

Then the point values for any sampling design can be Cottam corrected ($\kappa \times \lambda_M$). For example, the basic Morisita equation for two sectors assumes trees are located in opposite halves, so if the actual design is the nearest tree in the two nearest quadrants, the density from equation 2 will be overestimated and must be correspondingly corrected by multiplying by $\kappa = 0.857$.

Further corrections account for the restriction of trees to less than the full sector (θ), censoring of trees near the cardinal azimuths (ζ), and undersampling of trees smaller than a certain diameter limit (ϕ). These parameters are derived from analyses of measurements of bearing angles and diameters actually observed in surveys of witness trees within a subset of townships across the upper Midwest.

Sector bias (θ). Although the density model for two tree points assumes that the trees are on opposite sides of a sample line (point halves), often the actual sample is more restricted ($< 180^o$) within the sector or is a less restricted ($> 180^o$) angle beyond the sector. This deviation from the equation's assumption of equal distribution of angles across the 180°

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

sector is quantified using the empirical angle between the bearings of the two trees (pair angle). In the pair angle frequency plot (Figure 2), the observed proportion of trees (p) within any restricted sector divided by the proportion of that angle within the circle (α is an estimate of the bias imposed by the actual sampling (inspired by Kronenfeld & Wang (2007)). This factor ($\theta = p/\alpha$) indicates bias associated with differences in geometry of two tree samples. This parameter (θ) varies from 0.71 to 1.27, indicating sampling from effectively 253° to 141° sectors.

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

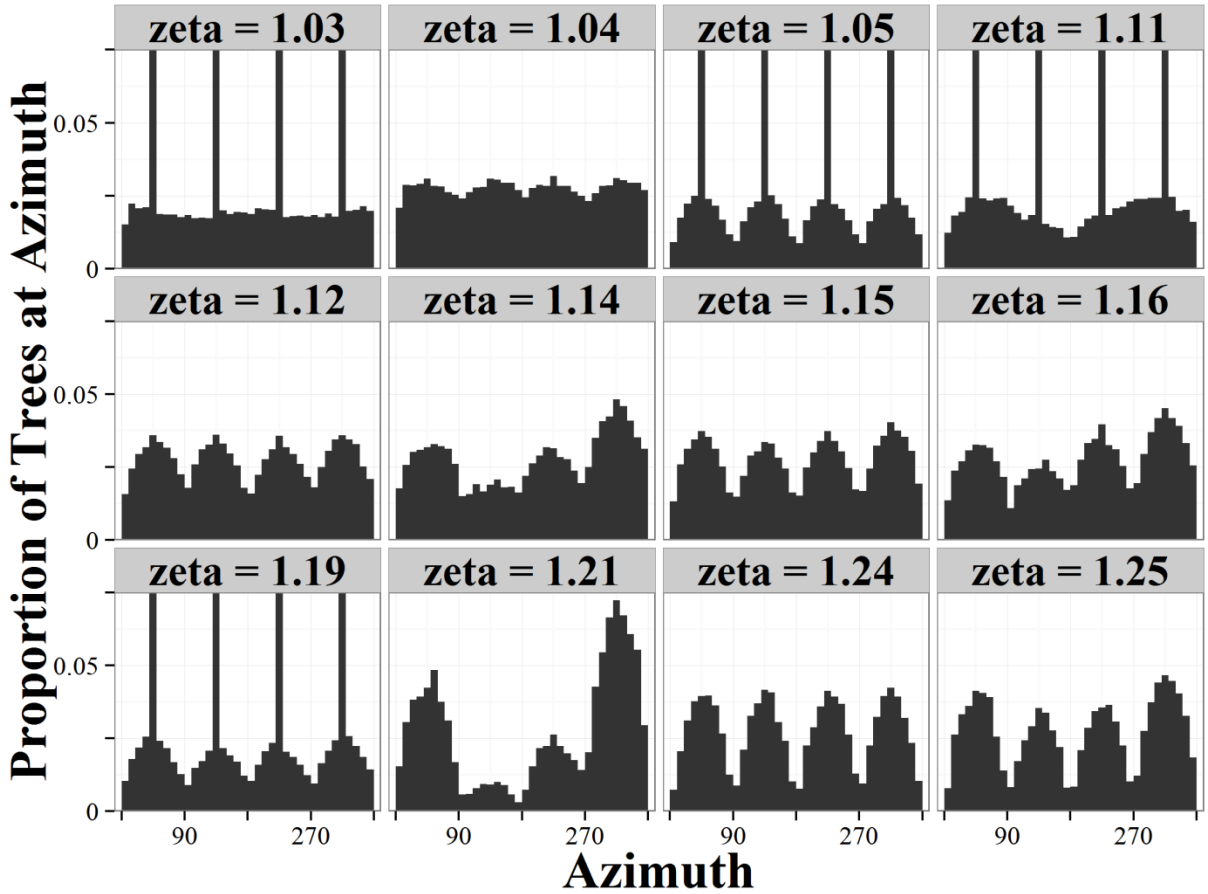

Figure 2. *Correction factors for ζ in the PLS data, and the associated distribution of azimuths for each ζ value, by panel. High peaks represent midpoints for quadrants where azimuth is defined as* e.g.*, NE or SW. Greater differences between cardinal directions and other azimuths result in higher ζ values, excluding the peaked values.*

Azimuthal censoring (ζ). In addition to sector bias, surveyors did not always sample trees near the cardinal directions (Kronenfeld and Wang 2007, Bouldin 2008, Hanberry et al. 2012b). This azimuthal censoring is commonly found along the line of travel on section lines and sometimes on the perpendicular quarter-section lines. Trees near the cardinal directions were passed over, and a replacement was found within a more restricted angular region. The correction for this bias is calculated following Kronenfeld and Wang (2007) in a manner similar to the sector bias. The factor ζ is the ratio of the proportion of

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

trees in the restricted area (p) divided by the proportion of the complete circle (α) that is used. The azimuthal censoring parameter (ζ) ranges from 1.03 to 1.25 indicating an equivalent to complete elimination of trees from 10° to 72° azimuths adjacent to the cardinal directions.

Diameter limit (ϕ). Examination of the diameter distributions from settlement era surveys across the upper Midwest clearly demonstrate witness trees less than 8 inches in diameter were undersampled (Bouldin 2008, Liu et al. 2011, Cogbill et al. in prep). We have confirmed this bias in our own inspection of plots of diameter frequency in the PLS data, which show a strong mode at 8"€. This bias can be accommodated by setting a diameter limit, and only calculating the density for trees with diameters above this limit. Total density calculated from all trees is reduced to this reference limit by simply multiplying the total by the percentage of trees above this limit. This effectively eliminates the smaller trees from the total and normalizes the value of trees above this standard. The parameter (ϕ) represents diameter size bias is simply the percentage of trees ≥ 8" and, in practice, ranges from 0.6 - 0.9.

Because all surveyor bias corrections are simple multipliers of the model density and should be independent, the bias-minimized estimate of the point density of trees ≥ 8" is:

$$\lambda_{\mathrm{Mcorrected}} = \kappa \times \theta \times \zeta \times \phi \times \lambda_{\mathrm{M}} \quad (3)$$

Estimates for each point *i* can be averaged for all *N* points in any region. Correction factors are calculated separately for different regions, years, internal versus external lines, section versus quarter-section points, and surveyor sampling designs (Table 1). All code to perform the analyses is included in Supplement 1 and the full rationale for and calculation

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

of these measures is described further in Cogbill et al. (in prep). Further, simulation used stem mapped stands from the region presented in Cogbill et al. (in prep) supports the robustness of this method, as opposed to other methods presented in the literature.

Table 1. *Correction values based on plot level survey design using state, year, and location within township as a basis for assignment. Years reported represent the upper bound for each set of survey years. Internal points are points within the township, external points are on the township boundary; no sampling occurred outside of a township boundary so plots were limited to half of the space for internal points. Townships are divided into Section and Quarter Sections, at most section points and some quarter section points, r instructions indicated four trees were to be sampled, these were '2nQ' plots, whereas others surveyed only two points in adjacent plot halves ('P' plots).*

State	Survey Year	Internal	Section	Trees	kappa	theta	zeta	phi
Wisc	1845	ext	Sec	P	2	0.82	1.14	0.89
Wisc	1845	ext	QSec	P	1	1.29	1.11	0.89
Wisc	1845	int	Sec	P	1	1.14	1.17	0.89
Wisc	1845	int	QSec	P	1	1.08	1.06	0.85
Wisc	1845	ext	Sec	2nQ	0.86	1	1.21	0.86
Wisc	1845	ext	QSec	2nQ	0.8563	1	1.11	0.91
Wisc	1845	int	Sec	2nQ	0.86	1	1.24	0.92
Wisc	1845	int	QSec	2nQ	0.86	1	0.75	0
Wisc	1907	ext	Sec	P	2	0.89	1.16	0.9
Wisc	1907	ext	QSec	P	2	0.9	1.14	0.84
Wisc	1907	int	Sec	P	1	1.07	1.12	0.9
Wisc	1907	int	QSec	P	1	1.04	1.04	0.8
Wisc	1907	ext	Sec	2nQ	0.86	1	1.13	0.99
Wisc	1907	ext	QSec	2nQ	0.86	1	1.12	0
Wisc	1907	int	Sec	2nQ	0.8563	1	1.24	0.83
Wisc	1907	int	QSec	2nQ	0.8563	1	1	0
Mich	all	ext	Sec	P	2	0.87	1.25	0.85
Mich	all	ext	QSec	P	1	0.94	1.21	0.76
Mich	all	int	Sec	P	1	1.27	1.24	0.85
Mich	all	int	QSec	P	1	1.26	1.15	0.77
Mich	all	ext	Sec	2nQ	0.86	1	1.24	0.84
Mich	all	ext	QSec	2nQ	0.86	1	1.35	0.85
Mich	all	int	Sec	2nQ	0.8563	1	1.26	0.84
Mich	all	int	QSec	2nQ	0.8563	1	1.28	0.68
Minn	1855	ext	Sec	P	2	0.71	1.19	0.67
Minn	1855	ext	QSec	P	1	1.05	1.11	0.68
Minn	1855	int	Sec	P	1	0.71	1.05	0.76

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

State	Survey Year	Internal	Section	Trees	kappa	theta	zeta	phi
Minn	1855	int	QSec	P	1	1.09	1.03	0.6
Minn	1855	ext	Sec	2nQ	0.86	1	1.17	0.66
Minn	1855	ext	QSec	2nQ	0.86	1	1	0.68
Minn	1855	int	Sec	2nQ	0.8563	1	1.5	0.59
Minn	1855	int	QSec	2nQ	0.8563	1	1	0.25
Minn	1907	ext	Sec	P	2	0.71	1.19	0.67
Minn	1907	ext	QSec	P	1	1.05	1.11	0.68
Minn	1907	int	Sec	P	1	0.71	1.05	0.76
Minn	1907	int	QSec	P	1	1.09	1.03	0.6
Minn	1907	ext	Sec	2nQ	0.86	1	1.17	0.66
Minn	1907	ext	QSec	2nQ	0.86	1	1	0.68
Minn	1907	int	Sec	2nQ	0.8563	1	1.5	0.59
Minn	1907	int	QSec	2nQ	0.8563	1	1	0.25

Basal Area and Biomass Estimates

Forest basal area is calculated by multiplying the point-based stem density estimate by the average stem basal area from the reported diameters at breast height for the closest two trees at the point (n=2). Aboveground dry biomass (Mg ha^{-1}) is calculated using the USFS FIA tree volume and dry aboveground biomass equations for the United States (Jenkins et al. 2004).

Biomass equations share the basic form:

$$m = \mathrm{Exp}(\beta_0 + \beta_1 * \mathrm{lndbh})$$

where m represents stem biomass for an individual tree in kg. β_0 and β_1 are the parameters described in Table 2 and dbh is the stem diameter at breast height (converted to cm) recorded in the survey notes. The biomass estimates are summed across both trees at a

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

survey point and multiplied by the stem density calculated at that point to produce an estimate of aboveground biomass reported in Mg ha^{-1} (Jenkins et al. 2004).

Table 2. *Biomass parameters used for the calculation of biomass in the pre-settlement dataset(rounded for clarity).*

Jenkins Species Group	β_0	β_1	PalEON Taxa Included (Supp. 2)
Aspen, Alder, Poplar, Willow	-2.20	2.38	Poplar, Willow, Alder
Soft Maple, Birch	-1.91	2.36	Birch
Mixed Hardwood	-2.48	2.48	Ash, Elm, Maple, Basswood, Ironwood, Walnut, Hackberry, Cherries, Dogwood, Buckeye
Hard Maple, Oak, Hickory, Beech	-2.01	2.43	Oak, Hickory, Beech, Other Hardwood
Cedar and Larch	-2.03	2.26	Tamarack, Cedar
Fir and Hemlock	-2.54	2.43	Fir, Hemlock
Pine	-2.54	2.43	Pine
Spruce	-2.08	2.33	Spruce

Matching PLSS tree genera to the species groups defined by Jenkins *et al*. (2004) is straightforward, placing the 22 genera used in this study into 9 allometric groups (Table 2). However, all maples are assigned to the generic "Hardwood" group since separate allometric relationships exist for soft and hard maple (Table 2). Biomass estimates for "Non tree" survey points are assigned 0 Mg ha^{-1}.

We use the stem density thresholds of Anderson and Anderson (1975) to discriminate prairie, savanna, and forest.

FIA Stem Density, Basal Area and Biomass

The United States Forest Service has monitored the nation's forests through the FIA Program since 1929, with an annualized state inventory system implemented in 1998 (Woudenberg et al. 2010). On average there is one permanent FIA plot per 2,428 ha of land

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

in the United States classified as forested. Each FIA plot consists of four 7.2m fixed-radius subplots in which measurements are made of all trees >12.7cm dbh (Woudenberg et al. 2010). We used data from the most recent full plot inventory (2007-2011). The FIA plot inventory provides a median of 3 FIA plots per cell using the 64km^2 grid.

We calculated mean basal area (m^2 ha^{-1}), stem density (stems ha^{-1}), mean diameter at breast height (cm), and mean biomass (Mg ha^{-1}) for all live trees with dbh greater than 20.32cm (8in). Biomass calculations used the same set of allometric regression equations as for the PLS data (Jenkins et al. 2004).

Gridding and Analysing PLS and FIA Data

Spatial maps of stem density, basal area and biomass were generated by averaging all PLS point or FIA plot estimates within a 64km^2 raster cell. Differences in sampling design between PLS and FIA data combined with spatially structured forest heterogeneity will affect the partitioning of within-cell versus between-cell variance, but not the expected estimates. Most 64km^2 cells have one or a few intensively sampled FIA plots. Therefore at this scale of aggregation, the low density of FIA plots in heterogeneous forests could result in high within-cell variance and high between-cell variability. For the PLS plotless (point based) estimates, stem density estimates are sensitive to trees close to the plot center. Point-level estimates with very high stem densities can skew the rasterized values, and it is difficult to distinguish artifacts from locations truly characterized by high densities. To accommodate points with exceptionally high densities we carry all values through the analysis, but exclude the top 2.5 percentile when reporting means and standard deviations in our analysis. PLS-based estimates are highly variable from point to point due to the small

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

sample size, but have low variance among 64 km^2 raster cells due to the uniform sampling pattern of the data. Thus within-cell variance is expected to be high for the PLS point data, but spatial patterns are expected to be robust at the cell level. The base raster and all rasterized data are available as Supplement 3.

Standard statistical analysis of the gridded data, including correlations and regression, was carried out in R (Team 2014), and is documented in supplementary material that includes a subset of the raw data to allow reproducibility. Analysis and presentation uses elements from the following R packages: `cluster` (Maechler et al. 2014), `ggplot2` (Wickham 2009a, 2009b), `gridExtra` (Auguie 2012), `igraph` (Csardi and Nepusz 2006), `mgcv` (Wood 2011), `plyr` (Wickham 2011), `raster` (Hijmans 2014), `reshape2` (Wickham 2007), `rgdal` (Bivand et al. 2014), `rgeos` (Bivand and Rundel 2014), `sp` (Pebesma and Bivand 2005, Bivand et al. 2013), and `spdep` (Bivand 2014).

We identify analogs and examine differences in composition between and within PLS and FIA datasets using Bray-Curtis dissimilarity (*vegdist* in `vegan`; Oksanen et al. 2014) for proportional composition within raster cells using basal area measurements. For the analog analysis we are interested only in the minimum compositional distance between a focal cell and its nearest compositional (not spatial) neighbor. The distribution of compositional dissimilarities within datasets indicates forest heterogeneity within each time period, while the search for closest analogs between datasets indicates whether contemporary forests lack analogs in pre-settlement forests ('novel forests'), or vice versa ('lost forests'). For the analog analyses, we compute Bray-Curtis distance between each 64km^2 cell in either the FIA or the PLS periods to all other cells within the other dataset

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(FIA to FIA, PLS to PLS), and between datasets (PLS to FIA and FIA to PLS), retaining only the minimum. For within era analyses (FIA - FIA and PLS - PLS), cells were not allowed to match to themselves. We define vegetation classes for lost and novel forests using k-medoid clustering (*pam* in `cluster`; Maechler et al. 2014)).

The differences in sampling design and scale between the PLS and FIA datasets, described above, potentially affect between-era assessments of compositional similarity (*e.g.*, Kronenfeld et al. 2010). The effects of differences in scale should be strongest in regions where there are few FIA plots per 64 km^2 cell, or where within-cell heterogeneity is high. For the analog analyses, this effect should increase the compositional differences between the FIA and PLS datasets. We test for the importance of this effect on our analog analyses via a sensitivity analysis in which we test whether dissimilarities between FIA and PLS grid cells are affected by the number of PLS plots per cell. We find a small effect, suggesting that our analyses are mainly sensitive to the compositional and structural processes operating on large spatial scales.

To understand the extent to which novelty operates at landscape scales we relate novelty to the distance to the nearest 'remnant' forest cell. Here we use a threshold of the 25%ile of compositional dissimilarity within the PLSS data, meaning the dissimilarity of any one 'remnant' cell is well within the historical bounds of dissimilarity. We use a binomial-logit regression to relate the degree of novelty to the spatial distance to the nearest 'remnant' cell, and examine this relationship for four major forest types within the PLSS data (Oak savanna, Oak/Poplar savanna, Pine/Spruce forest and Maple/Cedar/Hemlock/Birch mixedwood forests). We expect that a weak relationship will indicate that forest recovery

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

following landscape-scale land use change is moderated by a species pool culled from from small remnant patches, individual specimens, or local scale restoration efforts (for example during the 1930s). A significant relationship between distance to remant forest and novelty indicates that small patches have been insufficient to restore natural forest cover within the region, and would indicate that greater efforts are needed to restore landscapes at regional scales.

All datasets and analytic codes presented here are publicly available and open source at (http://github.som/SimonGoring/WitnessTrees), with the goal of enabling further analyses of ecological patterns across the region and the effects of post-settlement land use on forest composition and structure. Data are also archived at the Long Term Ecological Research Network Data Portal (https://portal.lternet.edu/nis/home.jsp).

Results:

Data Standardization

The original PLS dataset contains 490,818 corner points (excluding line and meander points), with 166,607 points from Wisconsin, 231,083 points from Minnesota and 93,095 points from Michigan. Standardizing data and accounting for potential outliers, described above, removed approximately 1.5% points from the dataset, yielding a final total of 366,993 points with estimates used in our analysis.

Rasterizing the PLS dataset to the Albers 64km^2 grid produces 7,939 raster cells with data. Each cell contains between 1 and 94 corner points, with a mean of 61.8 (σ = 15) and a median of 67 corners (Supplement 3). Cells with a low number of points were mainly near

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

water bodies or along political boundaries such as the Canadian/Minnesota border, or southern Minnesota and Wisconsin borders. Only 2.44% of cells have fewer than 10 points per cell.

Species assignments to genera were rarely problematic. Only 18 PLS trees were assigned to the Unknown Tree category, representing less than 0.01% of all points. These unknown trees largely consisted of corner trees for which taxon could not be interpreted, but for which diameter and azimuth data was recorded. A further 0.011% of trees were assigned to the "Other hardwood" taxon (*e.g.*, hawthorn, "may cherry", and "white thorn").

For maple the class has very high within-genera specificity for a number of assignments. A total of 78478 trees are assigned to "Maple". Of these, surveyors do use common names that can be ascribed to the species level (e.g., *A. saccharum*, n = 56331), but a large number of the remaining assignments are above the species level (n = 21356). This lack of specificity for a large number of records causes challenges in using the species level data. A similar pattern is found for pine, where many individual trees (125639) can be identified to the level of species (*P. strobus*, n = 41673; *P. bansiana*, n = 28784; *P. resinosa*, n = 28766), but there remains a large class of pine identified only at the genus level, or with unclear assignment (n = 17606).

For ash the data includes both attributions to black or brown ash (n=9312) and white ash (n = 2350), but again, also includes a large class of ash for which no distinction is made within the genera (n = 2, 7393, 16, 12).

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

These patterns are repeated throughout the data. For spruce this within-genera confusion is even greater, with 3, 50141, 43, 1 assignments to genera-level classes and only 20 to either black or white spruce.

Spatial Patterns of Settlement-Era Forest Composition: Taxa and PFTs

Stem Density, Basal Area and Biomass

The mean stem density for the region (Figure 3a) is 153 stems ha^{-1}. Stem density exclusive of prairie is 172 stems ha^{-1} and is 216 stems ha^{-1} when both prairie and savanna are excluded. The 95th percentile range is 0 - 423 stems ha^{-1}, and within-cell standard deviations between 0 and 441 stems ha^{-1}. Basal area in the domain (Figure 3c) has a 95th percentile range between 0 and 63.5 m^2 ha^{-1}, a mean of 22.2 m^2 ha^{-1}, within-cell standard deviations range from 0 to 76.7 m^2 ha^{-1}. Biomass ranges from 0 to 209 Mg ha^{-1} (Figure 3d), with cell level standard deviations between 0 and 569 Mg ha^{-1}. High within-cell standard deviations relative to mean values within cells for density, basal area and biomass indicate high levels of heterogeneity within cells, as expected for the PLS data, given its dispersed sampling design.

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

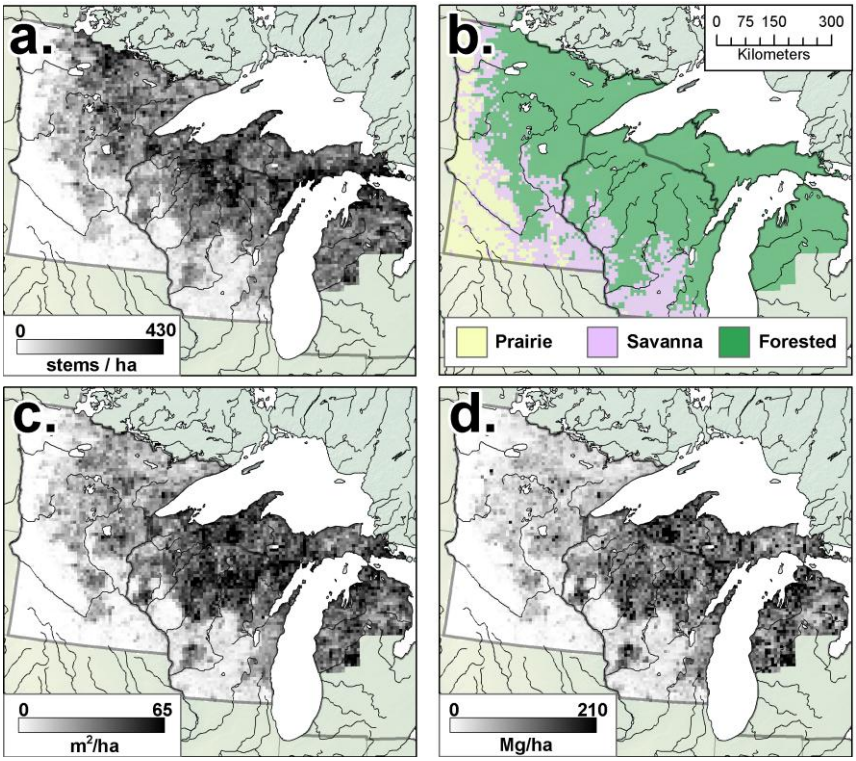

Figure 3. *Total stem density (a) in the Upper Midwest, along with forest type classification (b) based on PLS data and the stem density thresholds defined by Anderson and Anderson (1975); Table 3). Fine lines represent major rivers. To a first order, basal area (c) and biomass (d) show similar patterns to stem density (but see* Figure 5*).*

In the PLS data, stem density is lowest in the western and southwestern portions of the region, regions defined as prairie and savanna (Figure 3b, Table 3). When the Anderson and Anderson (1975) stem density thresholds (<47 stems ha^{-1} for Savanna, Table 3) are used, the extent of area classified as savanna is roughly equivalent to prior reconstructions (Curtis 1959, Bolliger et al. 2004, Rhemtulla et al. 2009b) (Figure 3b). The highest stem densities occur in north-central Minnesota and in north-eastern Wisconsin (Figure 3a), indicating younger forests and/or regions of lower forest productivity.

Table 3. *Forest classification scheme used in this paper for comparison between pre-settlement forests and the Haxeltine and Prentice (1996) potential vegetation classes represented in Ramankutty and Foley (Ramankutty and Foley 1999). Plant functional types*

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(PFTs) for grasslands (CG, grassland; Non-Tree samples in the PLS), broad leafed deciduous taxa (BDT) and needleleaded evergreen taxa (NET) are used, but leaf area index used in Haxeltine and Prentice (1996) is replaced by stem density classes from Anderson and Anderson (Anderson and Anderson 1975).

Forest Class	**Haxeltine & Prentice Rules**	**Current Study**
Prairie	Dominant PFT CG, LAI > 0.4	Stem dens. < 0.5 stem/ha
Savanna	Dominant PFT CG, LAI > 0.6	1 < Stem dens. < 47 stems ha^{-1}
Temperate Deciduous	Dominant PFT BDT, LAI > 2.5	Stem dens. > 48 stems ha^{-1}, BDT > 70% composition
Temperate Conifer	Dominant PFT (NET + NDT), LAI > 2.5	Stem dens. > 47 stems ha^{-1}, NET + NDT > 70% composition
Mixedwood	Both BDT (LAI > 1.5) & NET (LAI > 2.5) present	Stem dens. > 47 stems ha^{-1}, BDT & NET both < 70% composition

Forest structure during the settlement era can be understood in part by examining the ratio of stem density to biomass, a measure that incorporates both tree size and stocking. Regions in northern Minnesota and northwestern Wisconsin have low biomass and high stem densities (Figure 4, brown). This indicates the presence of young, small-diameter, even-aged stands, possibly due to frequent stand-replacing fire disturbance in the pre-EuroAmerican period or to poor edaphic conditions. Fire-originated vegetation is supported by co-location with fire-prone landscapes in Wisconsin (Schulte et al. 2005). High-density, low-biomass regions also have shallower soils, colder climate, and resulting lower productivity. High-biomass values relative to stem density (Figure 4, green) are found in Michigan and southern Wisconsin. These regions have higher proportions of deciduous species, with higher tree diameters than in northern Minnesota.

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

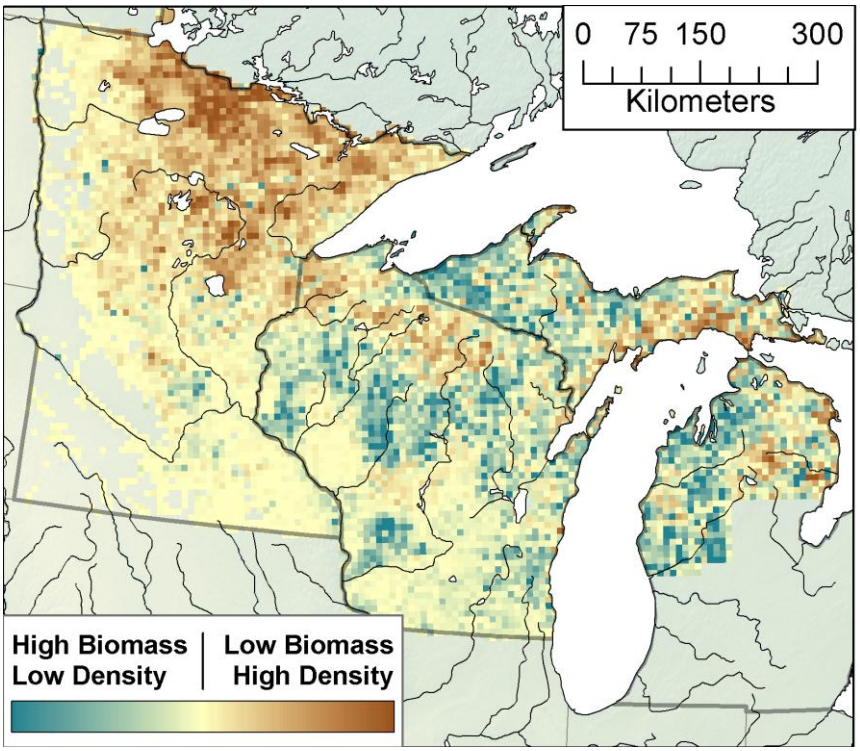

Figure 4. *Regions with high stem density to biomass ratios (blue) indicate dense stands of smaller trees, while regions with low stem density to biomass ratios (red) indicate larger trees with wider spacings.*

Taxon composition within settlement-era forests is spatially structured along dominant gradients from south to north (deciduous dominated to conifer dominated forests) and from east to west (mixed wood forests to open prairie) (Figure 5). Oak is dominant in the south of the region, with an average composition of 21%, however, that proportion drops to 8% when only forested cells are considered, due to its prevalence as a monotypic dominant in the savanna and prairie. Pine shows the opposite trend, with average composition of 14% and 17% in unforested and forested cells respectively. Pine distributions represent three dominant taxa, *Pinus strobus*, *Pinus resinosa* and *Pinus banksiana*. These three species have overlapping but ecologically dissimilar distributions, occuring in close proximity in some regions, such as central Wisconsin, and are typically associated with sandy soils with low water availability. Other taxa with high average

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

composition in forested cells include maple (10%), birch (10%), tamarack (9%) and hemlock (8%).

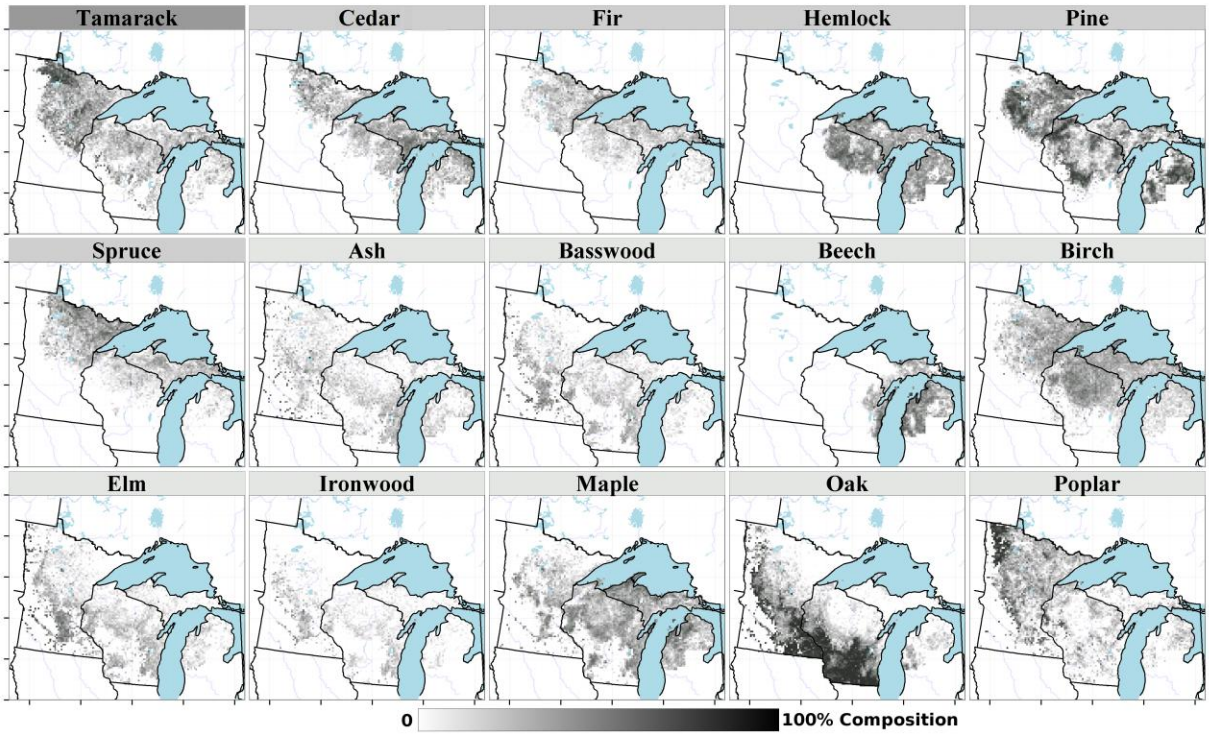

Figure 5. *Forest composition (%) for the 15 most abundant tree taxa. The scale is drawn using a square-root transform to emphasize low abundances. Shading of the bar above individual taxon maps indicates plant functional type assignments (dark gray: needleleafed deciduous; light gray: needleleafed evergreen; white: broadleafed deciduous).*

For a number of taxa, proportions are linked to the total basal area within the cell. For 4 taxa - hemlock, birch, maple and cedar - taxon proportions are positively related to total basal area. For 17 taxa including oak, ironwood, poplar, tamarack and elm, high proportions are strongly associated with lower basal areas (Figures 3 and 5). This suggests that hemlock, birch, maple and cedar occured in well-stocked forests, with higher average dbh. These taxa are most common in Michigan and in upper Wisconsin. Taxa with negative relationships to total basal area (*e.g.*, spruce and tamarack) are more common in the northwestern part of the domain.

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Spruce in the PLS represents two species (*Picea glauca*, *Picea mariana*) with overlapping distributions, but complex site preferences that vary in space. *P. glauca* is generally associated with dry upland to wet-mesic sites, while *P. mariana* is associated with hydric sites, but *P. mariana* also frequently occupies upland sites in northern Minnesota. Both cedar (*Thuja occidentalis*) and fir (*Abies balsamea*) are mono-specific genera in this region.

Northern hardwoods, such as yellow birch and sugar maple, and beech, are much less common in the lower peninsula of Michigan, and southern Wisconsin, except along Lake Michigan. Birch has extensive cover in the north, likely reflecting high pre-settlement proportions of yellow birch (*Betula alleghaniensis*) on mesic soils, and paper birch on sandy fire-prone soils and in northern Minnesota (birch proportions reach upwards of 34.1% in northeastern Minnesota). Hardwoods in the southwest, such as oak, elm, ironwood and basswood, are most typically mono-specific groupings, with the exception of oak, which comprises 7 species (see Supplement 2). Hardwoods in the southwest are located primarily along the savanna and southern forest margins, or in the southern temperate deciduous forests. Finally, maple and poplar (aspen) have a broad regional distribution, occupying nearly the entire wooded domain. Poplar comprises four species in the region, while maple comprises five species (Supplement 2). Both hardwood classes, those limited to the southern portions of the region, and those with distributions across the domain, correspond to well-defined vegetation patterns for the region (Curtis 1959).

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

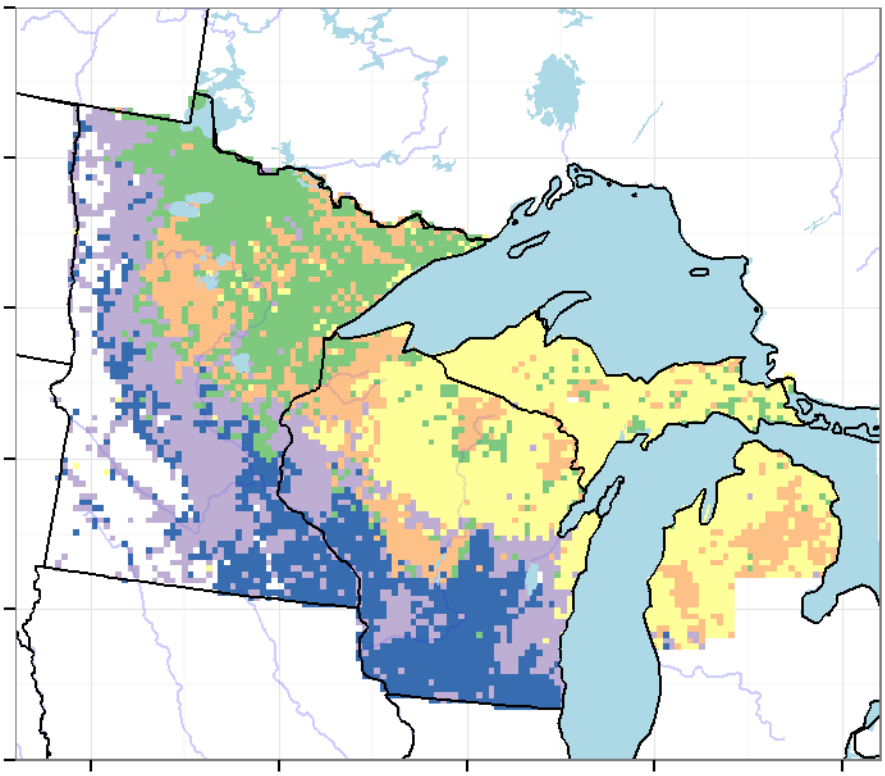

Figure 6. *The five dominant forest types in the Upper Midwest as defined by k-medoid clustering. Forest types (from largest to smallest) include Hemlock/Cedar/Birch/Maple (yellow), Oak/Poplar/Basswood/Maple (light purple), Tamarack/Pine/Spruce/Poplar (light green), Oak Savanna (dark purple) and Pine (orange). These forest types represent meso-scale (64km^2) forest associations, rather than local-scale associations.*

These individual species distributions result in a mosaic of forest classes accross the region (Figure 6). The dominant class is the Hemlock/Cedar/Birch/Maple assemblage in northern Wisconsin, and upper Michigan (Figure 6, yellow). This mixedwood assemblage is interspursed by both Pine dominated stands (Figure 6, orange) and, to a lesser degree, the softwood assemblage Tamarack/Pine/Spruce/Poplar (Figure 6, green), which dominates in north-eastern Minnesota. The softwood assemblage is itself interspursed with Pine dominated stands, and grades into a mixed-hardwood assemblage of Oak/Poplar/Basswood/Maple (Figure 6, light purple) to the west. Thismixed- softwood forest assemblage grades south into mono-specific Oak savanna (Figure 6, dark blue).

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

661

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

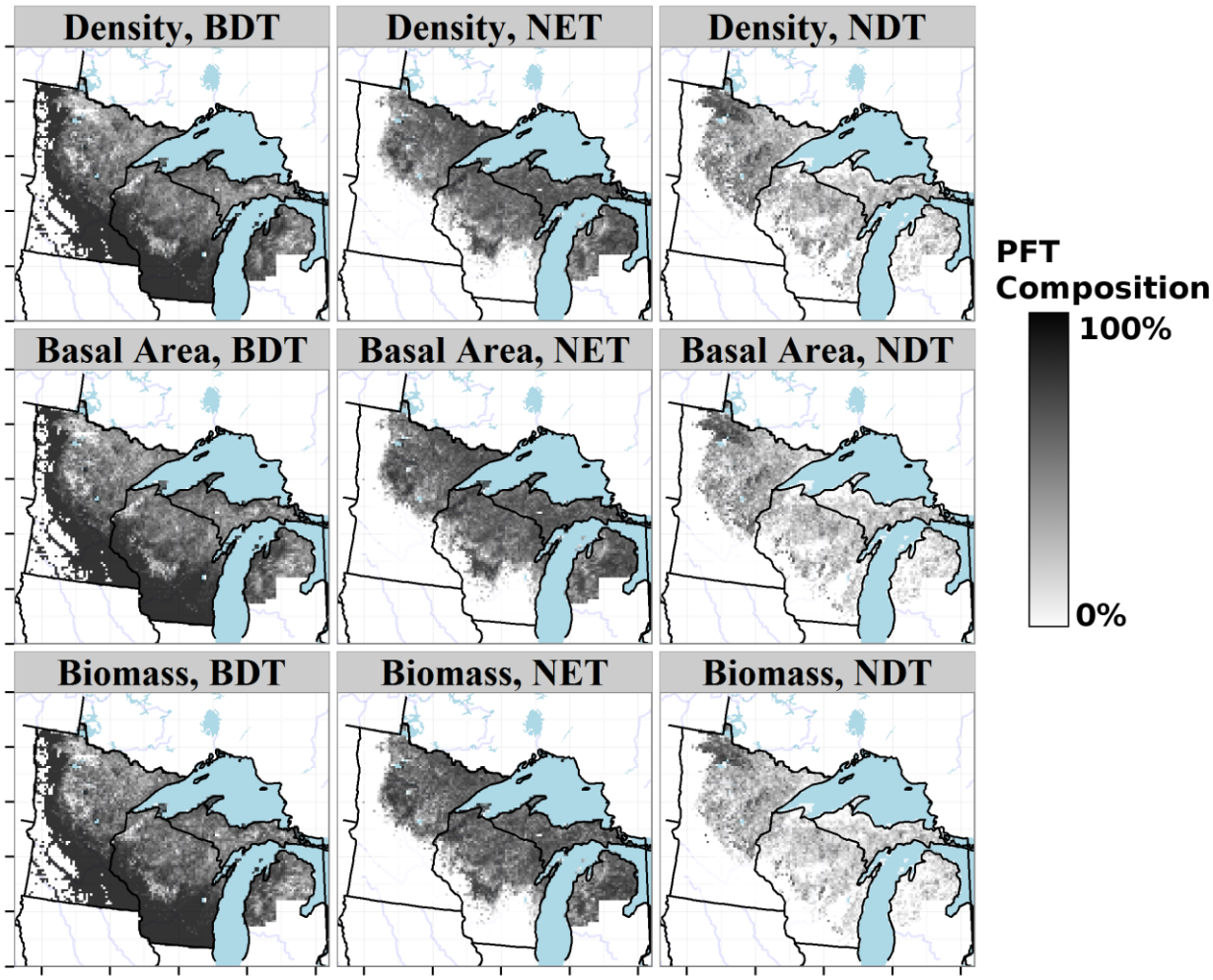

Figure 7. *Proportional distribution of Plant Functional Types (PFTs) in the upper Midwest from PLS data, for broadleaved deciduous trees (BDT), needleleaved deciduous trees (NDT), and needleleaved evergreen trees (NET). Distributions are shown as proportions relative to total basal area, total biomass, and composition (*Figure 3*). The grassland PFT is mapped onto non-tree cells with the assumption that if trees were available surveyors would have sampled them.*

The broad distributions of most plant functional types results in patterns within individual PFTs that are dissimilar to the forest cover classes (Figure 6). Thus overlap among PFT distributions (Figure 7) emerges from the changing composition within the plant functional type from deciduous broadleaved species associated with the southern, deciduous

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

dominated region, to broadleafed deciduous species associated with more northern regions in the upper Midwest.

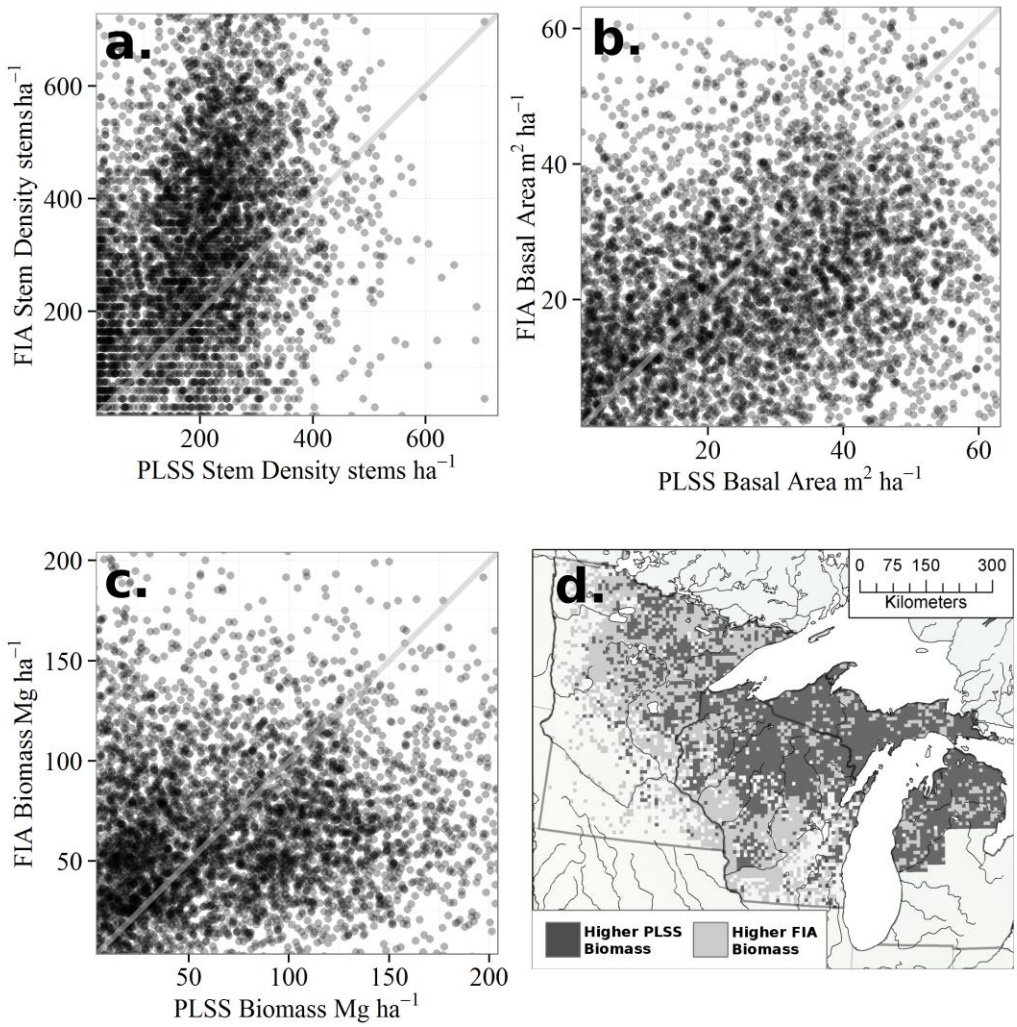

Figure 8. *The relationship between average stem density, total basal area and biomass values in the PLS and FIA datasets. Stem density and total basal area are higher in the FIA than in the PLS, however average cell biomass is higher in the PLS.*

Structural Changes Between PLS and FIA Forests

Modern forests (FIA) have higher stem densities (146 stems ha^{-1}, $t_{1,5177} = 51.8$, $p < 0.01$) and basal areas (-4.5 m^2 ha^{-1}, $t_{1,5177} = -16.4$, $p < 0.01$) than PLS forests, but overall, lower biomass (-8.72 Mg ha^{-1}, $t_{1,5177} = -6.55$, $p < 0.01$) than historical forests (Figure 8). We use

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

only point pairs where both FIA and PLS data occur since non-forested regions are excluded from the FIA and as such . The similarity in biomass despite lower stem density and total basal area in the PLS data is surprising. Two likely factors are shifts in allometric scaling associated with changes in species composition, or a higher mean diameter of PLS trees (Figure 8d).

The PLS has a lower overall mean diameter than the FIA (δ_{diam} = -2.9 cm, 95%CI from -17.3 to 8.18cm). FIA diameters are higher than PLS diameters in the northwestern parts of the domain (on average 6.47 cm higher), overlapping almost exactly with regions where we have shown low biomass-high density stands (Figure 4). At the same time, regions with high biomass and low density stands, in northeastern Wisconsin, and the Upper and Lower Peninsulas of Michigan, had higher average diameters during the PLS than in the FIA, on average 3.65 cm higher. Thus we are seeing an overal increase in tree size in the sub-boreal region and a decrease in temperate mixedwood forests, where we find tree species with much higher dbh:biomass ratios (Jenkins et al. 2004). This is coupled with declining variance in dbh across the domain (from within cell variance of 37.9 cm the PLS to 30.7 cm in the FIA). Thus, the mechanism by which low density and basal area produce roughly equivalent biomass estimates between the FIA and PLS is likely due to the differential impacts of land clearence and subesequent forest management in the south east vs the northwest. The loss of high biomass southern hardwood forests is balanced by higher biomass in the northeast due to fire supression and regeneration of hardwoods in the northwest. Declining diameters from the PLS to FIA are most strongly associated with higher abundances of poplar, ironwood and oak, while declining diameters are associated with maple and hemlock, further supporting the assertion that much of the loss in

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

diameter, and, subsequently biomass, is occuring in southeastern mixedwood/hardwood forest, while diameter and biomass increases are occuring in the northwest.

Differences between FIA and PLS data in sampling design are unlikely to be a factor; these differences are expected to affect how these datasets sample local- to landscape-scale heterogeneity, but should not affect the overall trends between datasets. Differences in variability introduce noise into the relationship, but given the large number of samples used here, the trends should be robust.

Compositional Changes Between PLS and FIA Forests: Novel and Lost Forests

Both the PLS- and FIA-era compositional data show similar patterns of within-dataset dissimilarity, with the highest dissimilarities found in central Minnesota and northwestern Wisconsin. High within-PLS dissimilarities are associated with high proportions of maple, birch and fir while high within-FIA dissimilarities are associated with high proportions of hemlock, cedar and fir. Dissimilarity values in the FIA dataset are less spatially structured than in the PLSS. Moran's I for dissimilarities within the FIA ($I_{FIA} = 0.198$, $p < 0.001$) are lower than the dissimilarities within the PLSS ($I_{PLSS} = 0.496$, $p < 0.001$), suggesting lower spatial autocorrelation in the FIA dataset. Cells with identical pairs represent 5.6% of the PLS cells and 7.44% of FIA cells. Identical cells in the PLS are largely located along the southern margin and most (69.5%) are composed entirely of oak. Cells in the FIA with identical neighbors are composed of either pure oak (19.4%), pure poplar (26%) or pure ash (14%).

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

There is a small but significant positive relationship ($F_{1,5964} = 920$, $p < 0.001$) between the number of FIA plots and within-FIA dissimilarity. The relationship accounts for 13% of total variance and estimates an increase of $\delta_d = 0.0134$ for every FIA plot within a cell. This increase represents only 3.08% of the total range of dissimilarity values for the FIA data. There is a gradient of species richness that is co-linear with the number of FIA plots within a cell, where plot number increases from open forest in the south-west to closed canopy, mixed forest in the Upper Peninsula of Michigan. Hence, differences in within- and between-cell variability between the PLS and FIA datasets seem to be having only a minor effect on these regional-scale dissimilarity analyses.

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

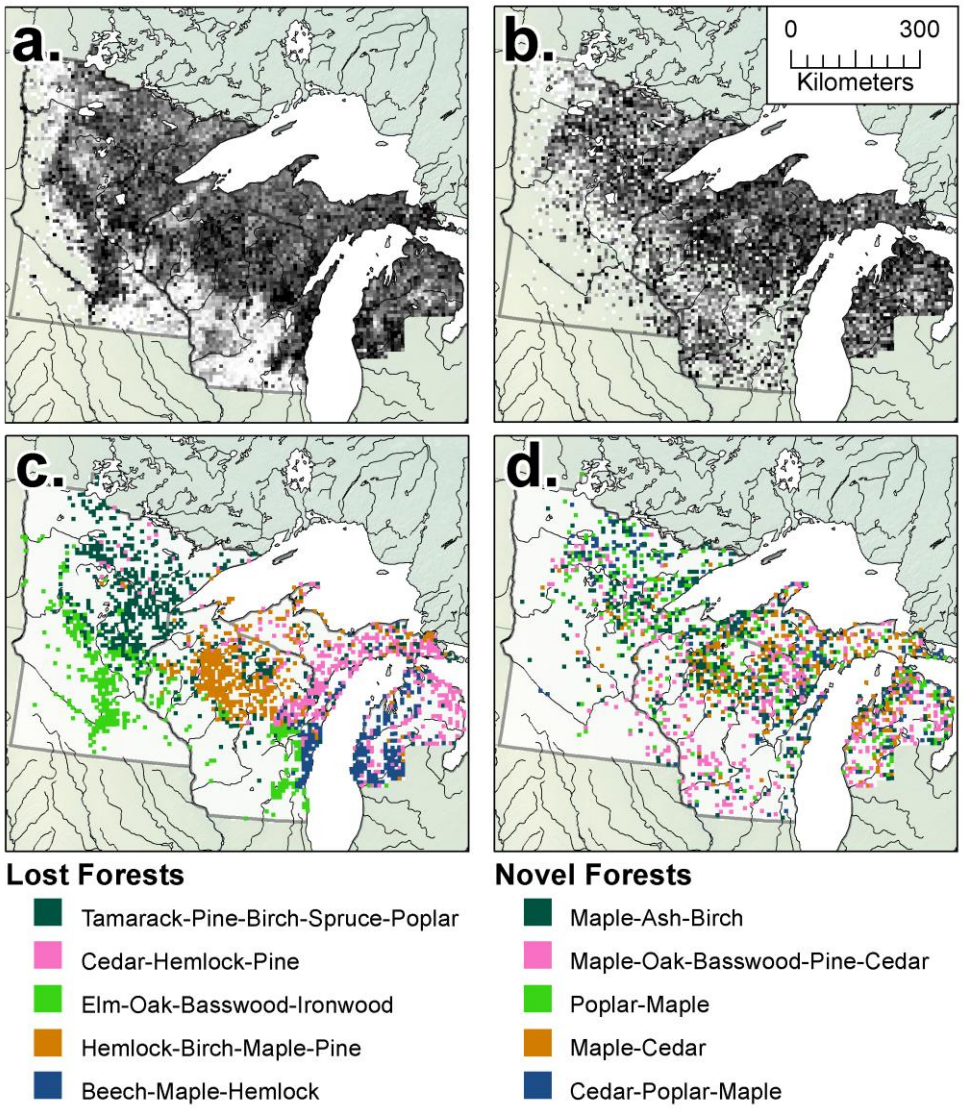

Figure 9. *Minimum dissimilarity maps. Distributions of minimum (within dataset) dissimilarities during the PLS (a) and FIA (b) show spatially structured patterns of dissimilarity, with stronger spatial coherence for the PLS. Lost forests (c) show strong compositional and spatial coherence, and have more taxa with percent composition > 10% than within Novel forests during the FIA era (d).*

We define no-analog communities as those whose nearest neighbour is beyond the 95%ile for dissimilarities within a particular dataset. In the PLS dataset, forests that have no modern analogs are defined as "lost forests", while forest types in the FIA with no past analogs are defined as "novel forests". More than 25% of PLS sites have no analog in the

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

FIA dataset ('lost forests'; PLS-FIA dissmimilarity, Figure 9c), while 29% of FIA sites have no analog in the PLS data ('novel forests'; FIA-PLS dissimilarity, Figure 9d). Lost forests show strong spatial coherence, centered on the "Tension Zone" (Curtis 1959), the ecotone between deciduous forests and hemlock-dominated mixed forest (Figure 5).

Lost forests are drawn from across the domain, and show strong ecological and spatial coherence (Figure 9c). Forest classes generally fall into five classes: Tamarack-Pine-Birch-Spruce-Poplar accounts for 28.8% of all lost forests and 7.97% of the total region. This forest type is largely found in north eastern Minnesota, extending southward to central Minnesota, into Wisconsin and along the Upper Peninsula of Michigan, as well as in scattered locations on the Lower Peninsula of Michigan (Figure 9c). This forest likely represents a mesic to hydric forest assemblage, particularly further eastward. Modern forests spatially overlapping this lost type are largely composed of poplar ($\overline{x}_{FIA}$ = 12%) and oak ($\overline{x}_{FIA}$ = 12%). Tamarack in these forests has declined significantly, from 23% to only 5% in the FIA, while Poplar has increased from 10% to 22%, resulting in forests that look less mesic and more like early seral forests.

Cedar/juniper-Hemlock-Pine accounts for 19.8% of all lost forests and 5.49% of the total region. This forest type is found largely in northeastern Wisconsin, and the Upper and Lower Peninsulas of Michigan. This lost forest type has been predominantly replaced by maple, poplar, and pine, retaining relatively high levels of cedar ($\overline{x}_{PLS}$ = 19%; $\overline{x}_{FIA}$ = 14%). The loss of hemlock is widespread across the region, but particularly within this forest type, declining to only 3% from a pre-settlement average of 18%.

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Elm-Oak-Basswood-Ironwood accounts for 19.6% of all lost forests and 5.42% of the total region. The region is centered largely within savanna and prairie-forest margins, both in south-central Minnesota and in eastern Wisconsin, but, is largely absent from savanna in the Driftless area of southwestern Wisconsin. These forests were historically elm dominated ($\bar{x}_{PLS}$ = 25%), not oak dominated savanna, as elsewhere (particularly in the Driftless). Modern forests replacing these stands are dominated by oak and ash, with strong components of maple, and basswood. Elm has declined strongly in modern forests ($\bar{x}_{FIA}$ = 1%), possibly in part due to Dutch Elm Disease and land use. The increase in ash in these forests is substantial, from $\bar{x}_{PLS}$ = 5% to $\bar{x}_{FIA}$ = 15%.

Hemlock-Birch-Maple-Pine accounts for 19.2% of all lost forests and 5.33% of the total region. This forest type, dominant in north central Wisconsin, was dominated by hemlock ($\bar{x}_{PLS}$ = 26%) and what was likely late seral yellow birch ($\bar{x}_{PLS}$ = 24%), replaced largely by maple (from $\bar{x}_{PLS}$ = 12% to $\bar{x}_{FIA}$ = 27%). Poplar increases from 1% to 13% in the FIA, again indicating a shift to earlier seral forests in the FIA. Hemlock is almost entirely lost from the forests, declining from 26% to 4% in the FIA.

Lastly, Beech-Maple-Hemlock accounts for 12.6% of all lost forests and 3.49% of the total region. This forest type is found exclusively on the central, western shore of Lake Michigan and in the Lower Peninsula, in part due to the limited geographic range of Beech in the PLS dataset (Figure 5). Beech is almost entirely excluded from the modern forests in this region, declining from $\bar{x}_{PLS}$ = 37% to $\bar{x}_{FIA}$ = 4%. Pine in the region increases from 9% to 16%, while maple, the dominant taxa in the modern forests, increases from 16 - 25%.

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

On average lost forests contain higher proportions of ironwood (r = 0.203), beech (r = 0.2), birch (r = 0.189) and hemlock (r = 0.188) than the average PLS forest, and lower proportions of oak (r = -0.28), poplar (r = -0.145), and pine (r = -0.107).

The distribution of novel ecosystems (Figure 9d) is spatially diffuse relative to the lost forest of the PLS and the forest types tend to have fewer co-dominant taxa. FIA novel forest types also have a more uneven distribution in proportion than the PLS lost forests. Overall, novel forests are associated with higher proportions of maple (r = 0.02), ash (r = 0.03) and basswood (r = -0.04), although basswood is dominant in only one forest type (Poplar-Cedar/juniper-Maple). Novel forests are associated with lower proportions of oak (r = -0.28), and pine (r = -0.11). This analysis suggests that the loss of particular forest types associated with post-settlement land use was concentrated in mesic deciduous forests and the ecotonal transition between southern and northern hardwood forests, while the gains in novelty were more dispersed, resulting from an overall decline in seral age.

By far the largest novel forest type is Maple, which accounts for 37.2% of all novel forests and 2.68% of the total region. As with all novel forest types, this forest type is broadly distributed across the region. This forest type is associated with co-dominant maple ($\overline{x}_{FIA}$ = 23%) and ash ($\overline{x}_{FIA}$ = 22%). Hemlock has declined significantly across this forest type, from $\overline{x}_{PLS}$ = 24% to $\overline{x}_{FIA}$ = 4%.

Poplar-Cedar/juniper-Maple, accounts for 28.8% of all novel forests and 2.08% of the total region. The broad distributiof these novel forests makes assigning a past forest type more difficult than for the PLS lost forests, the distribution replaces two classes of past forest,

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

one dominated by oak, in southern Wisconsin and Minnesota, the other by mixed hemlock, beech, birch and cedar forests.

Pine-Cedar/juniper-Poplar-Maple forest accounts for 16.3% of all novel forests and 1.17% of the total region. This forest type is again broadly distributed, and is widely distributed across the region, representing a homogenous, early seral forest type, likely associated with more mesic sites. Oak forest accounts for 13.3% of all novel forests and 0.96% of the total region. This grouping again shows a pattern of broad distribution across the region, associated with cedar/juniper percentages near 40%, with smaller components of poplar (14%) and maple (13%).

Spatial Correlates of Novelty

Modern compositional dissimilarity from the PLSS data is related to distance from 'remnant' forest. The dissimilarity quantile of FIA-PLSS distances increases with increasing distance to remnant cells. While it is difficult to quantify exactly what is meant by remnant, given the strong compositional variability within the PLS dataset, we use the 25%ile quantile of within dataset nearest-neighbor disimilarities for the PLS as a useful indicator. Results are robust to higher levels of dissimilarity, up to the 90%ile. Using the 25%ile for within PLS dissimilarity, approximately 17% of FIA cells can be classed as 'remnant' forest. The mean distance to remnant forests for cells with dissimilarities above the 25%ile is 17.5 km, higher than the mean of ~9.6km expected if each 8x8km cell had at least one adjacent 'remnant' cell.

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

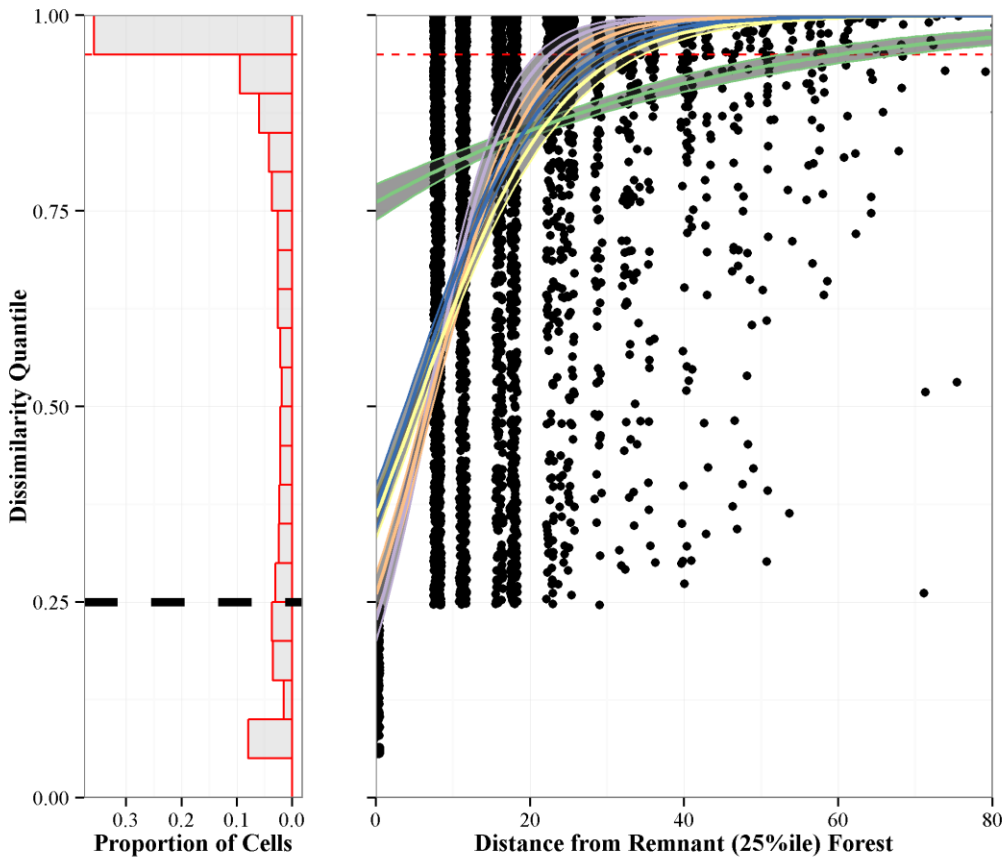

Figure 10. *(a) Distribution of dissimilarity quantiles for FIA forests from historical PLS forests. The quantile scale is defined using nearest neighbor distances within the PLS data, provding a measure of internal variability. (b) The model relating novelty to spatial distance from remnant forest. Here the 25%ile is used to indicate remnant forest. The red dashed line indicates forested cells above the 95%ile of dissimilarity, or novel forests. The curves represent the relationship between spatial distance and compositional dissimilarity for each of the five major historic forest types defined here as Hemlock/Cedar/Birch/Maple (yellow), Oak/Poplar/Basswood/Maple (light purple), Tamarack/Pine/Spruce/Poplar (light green), Oak Savanna (dark purple) and Pine (orange).*

The GLM shows that distance from remnant forests in the FIA is significantly related to the dissimilarity quantile for FIA data ($\chi_{1,4}$=3271, $p < 0.001$). The mean distance to novelty varies by PLS forest type, but is between approximately 20 and 60km for the four forest types examined here (Figure 11b). The least sensitive forest type appears to be the northern softwood forests (Figure 6, green), which reach 'novelty' only when separated from remnant forests by 60km. Hemlock/Cedar/Birch/Maple forest type (Figure 6, yellow).

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The most sensitive forest type, the hardwood Oak/Poplar/Basswood/Maple forest type (Figure 6, light purple) reaches novelty at only 23km from a focal remnant cell, while Pine dominated forests (Figure 6, orange) reach novelty at 26km. Oak savanna and Hemlock/Cedar/Birch/Maple forest are less sensitive, reaching novelty at 30 and 33km respectively.

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

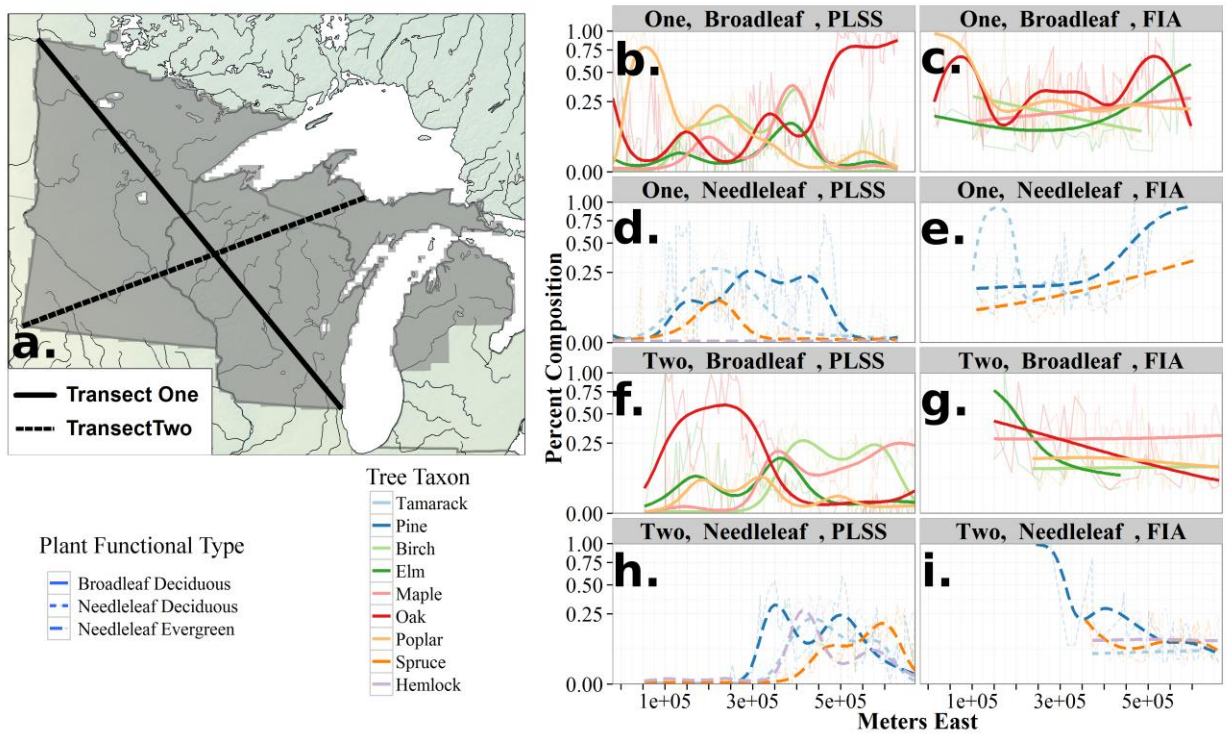

Figure 11. *Transects (a) across the region show clear changes in the ecotonal strength. Transect One shows shifts in broad-leafed taxon distributions from the PLS to FIA (b and c) and in needle-leafed distributions (d and e). Transect Two broadleaf (f and g) and needleleaf (h and i) taxa show shifts that again appear to represent regional scale homogenization. Ecotones in the pre-settlement era were stronger in the past than they are in the present. Fitted curves represent smoothed estimates across the transects using Generalized Additive Models using a beta family.*

Compositional Changes Between PLS and FIA Forests: Ecotone Structure

To understand how the ecotonal structure has been transformed by post-settlement land use, we constructed two transects of the FIA and PLS data (Figure 11a), and fitted GAM models to genus abundances along these transects. Transect One (T1) runs from northern prairie (in northern Minnesota) to southern deciduous savanna in southeastern Wisconsin (left to right in Figures 11c-f), while Transect Two (T2) runs from southern prairie in southwestern Minnesota to northern mixedwood forest in the Upper Peninsula of Michigan (left to right in Figures 11g-j). In general, these transect analyses show: 1) significant

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

differences in ecotonal structure between the present and pre-settlement, and 2) steeper ecotones in the past and more diffuse ecotones today.

For T1, GAM models show significant differences (using AIC) between time periods in curves for all broadleafed taxa (Figure 11b & c) and for al needleleafed taxa (Figures 10d and e). The PLS curves show a rapid transition in the northwest from oak to poplar dominated open forest that then transitions to a needleleafed forest composed of pine, spruce and tamarack, with high proportions of tamarack grading to pine further to the south east. Tamarack and poplar proportions decline gradually from the east, being replaced first by pine, then briefly by maple and birch, and then. ultimately by oak as the transect grades into oak savanna. In the FIA dataset oak and poplar in the northwest appears to decline simultaneously, grading into needleleafed forests that are absent from the FIA dataset in the first 100km along the transect. While the PLS transect shows distinct vegetation types in the central partof the transect, the FIA shows relatively constant proportions of oak, pine, spruce, poplar and maple before pine, oak and elm increase in the southeastern portions of the transect.

The second transect shows a similar pattern, with well defined ecotones in the pre-settlement period(Figure 11f and h), that are largely absent from the FIA data (Figure 11g and i). Oak forest, with a component of elm and poplar in the southwest grades slowly to a rapid transition zone where pine, elm, maple (first), then rapidly birch, hemlock and tamarack, and later, spruce, increase. This region, the Tension Zone, extends from 3×10^5 to 4.5×10^5 meters East, eventually becoming a forest that shows co-dominance between birch, pine, maple, spruce and tamarack, likely reflecting some local variability as a result of

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

topographic and hydrological factors. Missing data at the beginning of the FIA transect reflects a lack of FIA plots in unforested regions in the west

Contemporary forests show broader homogenization and increased heterogeneity (evidenced by the lower within-FIA Moran's I estimates for near-neighbor distances) at a local scale in the region. Homogenization is evident across T1, where Bray-Curtis dissimilarity between adjacent cells declines from the PLSS to the FIA (δ_{beta} = -0.22, t_{113} = -7.93, p<0.001), mirroring declines in the pine barrens between the 1950s and the present (Li and Waller 2014). The PLS shows strong differentiation in the central region of T2 where maple-pine-oak shifts to pine-poplar-birch forest (Figure 11d). This sharp ecotone is not apparent in the FIA data, which shows gradual and blurred changes in species composition across the ecotone (Figure 11i). β-diversity along T2 is lower in the FIA than in the PLSS (δ_{beta} = -0.19, t_{65}=-7.34, p < 0.01), indicating higher heterogeneity in the PLS data at the 64 km^2 meso-scale.

Across the entire domain, β diversity is lower in the FIA than in the PLS (δ_{β} = -0.172, $t_{1.3e7}$ = 2480, p <0.001), lending support to the hypothesis of overall homogenization. Differences in sampling design between PLS and FIA data cannot explain this homogenzation, since its effect would have been expected to increase β-diversity along linear transects and at larger spatial scales.

Discussion

Many forests of the PLS, are no longer a part of the modern landscape. Forest types have been lost at the 64 km^2 mesoscale, and new forest types have been gained. The joint

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

controls of broad-scale climatic structuring and local hydrology on forest composition and density can be seen in the pre-settlement forests, particularly along the Minnesota River in south-western Minnesota, where a corridor of savanna was sustained in a region mostly occupied by prairie (Figure 3b), but ecotones in the modern forest composition data are weaker now than in the past (Fig. 10), with clear signs of increased homogenization at local and regional scales and decreased spatial structure in vegetation assemblages (Figure 9).

The loss of ecotones in the upper Midwestern United States suggests that our ability to predict the abiotic controls on species distributions at the landscape scale may be weaker than in the past, reducing the influence of variables such as climate or edaphic factors, and increasing the relative influence of recent land use history. Work in eastern North America suggests the utility of including spatial structure in species distribution models to improve predictive ability (Record et al. 2013). The spatial random effects may improve models by capturing missing covariates within SDMs (Record et al. 2013), but if recent land use history has strongly shaped species distributions, or co-occurence, then the spatial effect is likely to be non-stationary at longer temporal scales. Given the implicit assumption of stationarity in many ecological models (Wolkovich et al. 2014), the need for longer time-scale observations, or multiple baselines from which to build our distributional models becomes critical if we are to avoid conflating recent land use effects with the long term ecological processes structuring the landscape.

Decreased β diversity along regional transects indicates homogenization at meso-scales of 100s of km^2, while the overall reduction in Moran's I for dissimilarity in the FIA indicates a regional reduction in heterogeneity on the scale of 1000s of km^2. The selective loss or

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

weakening of major vegetation ecotones, particularly in central Wisconsin, and the development of novel species assemblages across the region. These changes are the result of land use, both agricultural and logging, but affect forests in contrasting ways across the domain. Maple has become one of the most dominant taxa across the region, while in northern Minnesota, forest biomass has increased and species shifts have reflected increases in poplar and pine, while in southern Wisconsin, biomass has declined, and hemlock has been lost almost completely.

Anthropogenic shifts in forest composition over decades and centuries seen here and elsewhere (Cogbill et al. 2002, Thompson et al. 2013) are embedded within a set of interacting systems that operate on multiple scales of space and time (macrosystems, *sensu* Heffernan et al. 2014). Combining regional historical baselines, long term ecological studies and high frequency analyses can reveal complex responses to climate change at local and regional scales (Groffman et al. 2012). Estimates of pre-settlement forest composition and structure are critical to understanding the processes that govern forest dynamics because they represent a snapshot of the landscape prior to major EuroAmerican land-use conversion (Schulte and Mladenoff 2001, Liu et al. 2011). Pre-settlement vegetation provides an opportunity to test forest-climate relationships prior to land-use conversion and to test dynamic vegetation models in a data assimilation framework (*e.g.*, Hartig et al. 2012). For these reason, the widespread loss of regional forest associations common in the PLS (Figure 9d), and the rapid rise of novel forest assemblages (Figure 9e) have important implications for our ability to understand ecological responses to changing climate. The loss of historical forest types implies that the modern understanding of forest cover, climate relationships, realized and potential niches and species associations may be

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

strongly biased in this region, even though 29% of the total regional cover is novel relative to forests only two centuries ago.

Beyond shifts in composition at a meso-scale, the broader shifts in ecotones can strongly impact models of species responses and co-occurence on the landscape. For example, the heterogeneity, distribution, and control of savanna-forest boundaries (Staver et al. 2011) is of particular interest to ecologists and modelers given the ecological implications of current woody encroachment on savanna ecosystems (Ratajczak et al. 2012). Declines in landscape heterogeneity may also strongly affect ecosystem models, and predictions of future change. Recent work using the FLUXNET tower network has shown that energy budgets are strongly related to landscape measures of heterogeneity in both vegetation and topography (Stoy et al. 2013). Our data show higher levels of vegetation heterogeneity at mesoscales during the pre-settlement era, and greater fine scaled turnover along transects. Lower β diversity shown here and elsewhere (Li and Waller 2014) indicate increasing homogeneity at a very large spatial scale, and the loss of resolution along major historical ecotones. Increasing heterogeneity in the pre-settlement time would introduce non-stationarity into energy budgets, and would likely increase the uncertainty in vegetation-atmosphere processes, a key uncertainty in CMIP5 models (Friedlingstein et al. 2014).

This study also points to the need for a deeper understanding of some of the landscape- and regional-scale drivers of novelty, given the likely role for climatic and land use change (including land abandonment) to continue to drive ecological novelty (Martinuzzi et al. 2015, Radeloff et al. in press). In particular the role of regional species pools and remnant patches of forest in driving or mitigating compositional novelty. This work shows that the

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

baseline forest type, and its structure on the landscape moderates the degree to which landscape scale patterns can drive compositional novelty. To some degree relationships between compositional novelty and distance from remnant patches may be dependent on the simplicity or complexity of the species pool and the sensitivity of dissimilarity metrics to β diversity (Faith et al. 1987). Our results indicate that this cannot be the driving factor, both the simplest forest class (Pine) and one of the most complex (Oak/Poplar/Basswood/Maple) show strong spatial effects. These forest types are also one of the most fragmented across the region, indicating that fragmentation, both in the modern sense, driven by land use change and subsequent reforestation, and in the historic sense, driven by biotic and abiotic factors at local and landscape scale, resulting in the patchy distributions of one forest type within the matrix of one or more other forest types (as is the case with Pine forests). This may well point to the role of landscape-level controls in moderating alternate stable states (Bowman et al. 2015) within these forest types. There is strong evidence that in some locations pine forests have persisted over long timescales in the region (Ewing 2002), although there is also evidence, in other regions, that these states may shift strongly in response to interactions between landscape level processes such as fire and geophysical features (Lynch et al. 2014). Thus complex interactions between landscape scale processes, whether they be fire, land use change, or geophysical features, and the species assemblages themselves, point to the difficulty in making simplifying assumptions about species assemblages, whether they be plant functional types, species richness, or phylogenetic metrics, since we know that this region is dominated by forests that respond very differently to the settlement-era disturbance, but that are composed of different species of the same genera.

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The analysis relating to the distance-to-novelty (Figure 10) also points to the possibility that should the public push for landscape-scale restoration, or management using historical baselines, focusing on restoration in locations where restoration potential is high, as suggested for Hemlock/Hardwood forests in northern Wisconsin (Bolliger et al. 2004) has the possibility of contributing to landscape scale change. If some of the novelty is driven by depauparate species pools beyond certain threshold distances from remnant forests then it should also be possible to restore these forest through translocation of key species (Seddon 2010).

Methodological advances of the current work include 1) the systematic standardization of PLS data to enable mapping at broad spatial extent and high spatial resolution, 2) the use of spatially varying correction factors to accommodate variations among surveyors in sampling design, and 3) parallel analysis of FIA datasets to enable comparisons of forest composition and structure between contemporary and historical time periods. This approach is currently being extended to TPS and PLS datasets across the north-central and northeastern US, with the goal of providing consistent reconstructions of forest composition and structure for northeastern US forests at the time of EuroAmerican forests.

Our results support the consensus that robust estimates of pre-settlement forest composition and structure can be obtained from PLS data (*e.g.*, Wisconsin: Schulte et al. 2002, Iowa: Rayburn and Schulte 2009, California: Williams and Baker 2011, Oregon: Duren et al. 2012). Patterns of density, basal area and biomass are roughly equivalent to previous estimates (Schulte et al. 2007, Rhemtulla et al. 2009a). Our results for stem density are lower than those estimated by Hanberrry *et al*. (Hanberry et al. 2012a) for

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

eastern Minnesota, but density and basal area are similar to those in the northern Lower Peninsula of Michigan (Leahy and Pregitzer 2003) and biomass estimates are in line with estimates of aboveground carbon for Wisconsin (Rhemtulla et al. 2009a).

These maps of settlement-era forest composition and structure can also provide a useful calibration dataset for pollen-based vegetation reconstructions for time periods prior to the historic record. Many papers have used calibration datasets comprised of modern pollen samples to build transfer functions for inferring past climates and vegetation from fossil pollen records (Jacques et al. 2008, Goring et al. 2009, Paciorek and McLachlan 2009, Birks et al. 2010). However, modern pollen datasets are potentially confounded by recent land use, which can alter paleoclimatic reconstructions using pollen data (Jacques et al. 2008). By linking pollen and vegetation at modern and historical periods we develop capacity to provide compositional datasets at broader spatio-temporal scales, providing more data for model validation and improvement. Ultimately, it should be possible to assimilate these empirical reconstructions of past vegetation with dynamic vegetation models in order to infer forest composition and biomass during past climate changes. Data assimilation, however, requires assessment of observational and model uncertainty in the data sources used for data assimilation. Spatiotemporal models of uncertainty are being developed for the compositional data (Paciorek et al. in review) and biomass data (Feng *et al*. in prep.).

Ultimately the pre-settlement vegetation data present an opportunity to develop and refine statistical and mechanistic models of terrestrial vegetation that can take multiple structural and compositional forest attributes into account. The future development of uncertainty

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

estimates for the data remains an opportunity that can help integrate pre-settlement estimates of composition and structure into a data assimilation framework to build more complete and more accurate reconstructions of past vegetation dynamics, and to help improve predictions of future vegetation under global change scenarios.

Literature Cited

Almendinger, J. C. 1996. Minnesota's bearing tree database. Minn. Department of Natural Resources.

Anderson, R. C., and M. R. Anderson. 1975. The presettlement vegetation of Williamson county, Illinois. Castanea:345–363.

Auguie, B. 2012. GridExtra: Functions in Grid graphics.

Bellemare, J., G. Motzkin, and D. R. Foster. 2002. Legacies of the agricultural past in the forested present: An assessment of historical land-use effects on rich mesic forests. Journal of Biogeography 29:1401–1420.

Birks, H. B., O. Heiri, H. Seppä, and A. E. Bjune. 2010. Strengths and weaknesses of quantitative climate reconstructions based on late-Quaternary biological proxies. Open Ecology Journal 3:68–110.

Bivand, R. 2014. spdep: Spatial dependence: Weighting schemes, statistics and models.

Bivand, R., and C. Rundel. 2014. rgeos: Interface to geometry engine - open source (GEOS).

Bivand, R., T. Keitt, and B. Rowlingson. 2014. rgdal: Bindings for the Geospatial Data Abstraction Library.

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Bivand, R., E. Pebesma, and V. Gomez-Rubio. 2013. Applied spatial data analysis with R. Second editions. Springer.

Bolliger, J., L. A. Schulte, S. N. Burrows, T. A. Sickley, and D. J. Mladenoff. 2004. Assessing ecological restoration potentials of Wisconsin (USA) using historical landscape reconstructions. Restoration Ecology 12:124–142.

Booth, R. K., S. T. Jackson, V. A. Sousa, M. E. Sullivan, T. A. Minckley, and M. J. Clifford. 2012. Multi-decadal drought and amplified moisture variability drove rapid forest community change in a humid region. Ecology 93:219–226.

Bouldin, J. 2008. Some problems and solutions in density estimation from bearing tree data: A review and synthesis. Journal of Biogeography 35:2000–2011.

Bourdo, E. A. 1956. A review of the General Land Office survey and of its use in quantitative studies of former forests. Ecology:754–768.

Bowman, D. M., G. L. Perry, and J. Marston. 2015. Feedbacks and landscape-level vegetation dynamics. Trends in ecology & evolution 30:255–260.

Brugam, R. B. 1978. Pollen indicators of land-use change in southern connecticut. Quaternary Research 9:349–362.

Cogbill, C. V., J. Burk, and G. Motzkin. 2002. The forests of presettlement New England, USA: Spatial and compositional patterns based on town proprietor surveys. Journal of Biogeography 29:1279–1304.

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Cogbill, C. V., S. J. Goring, and A. Thurman. in prep. Estimation of robust correction factors for Public Land Survey Data.

Cottam, G., and J. T. Curtis. 1956. The use of distance measures in phytosociological sampling. Ecology:451–460.

Cramer, V. A., R. J. Hobbs, and R. J. Standish. 2008. What's new about old fields? Land abandonment and ecosystem assembly. Trends in Ecology & Evolution 23:104–112.

Csardi, G., and T. Nepusz. 2006. The igraph software package for complex network research. InterJournal Complex Systems:1695.

Curtis, J. T. 1959. The vegetation of Wisconsin: An ordination of plant communities. University of Wisconsin Press.

Dupouey, J.-L., E. Dambrine, J.-D. Laffite, and C. Moares. 2002. Irreversible impact of past land use on forest soils and biodiversity. Ecology 83:2978–2984.

Duren, O. C., P. S. Muir, and P. E. Hosten. 2012. Vegetation change from the Euro-American settlement era to the present in relation to environment and disturbance in southwest Oregon. Northwest Science 86:310–328.

Ellis, E. C., and N. Ramankutty. 2008. Putting people in the map: Anthropogenic biomes of the world. Frontiers in Ecology and the Environment 6:439–447.

Etienne, D., P. Ruffaldi, J. L. Dupouey, M. Georges-Leroy, F. Ritz, and E. Dambrine. 2013. Searching for ancient forests: A 2000 year history of land use in northeastern French

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

forests deduced from the pollen compositions of closed depressions. The Holocene 23:678–691.

Ewing, H. A. 2002. The influence of substrate on vegetation history and ecosystem development. Ecology 83:2766–2781.

Faith, D. P., P. R. Minchin, and L. Belbin. 1987. Compositional dissimilarity as a robust measure of ecological distance. Vegetatio 69:57–68.

Foster, D. R., G. Motzkin, and B. Slater. 1998. Land-use history as long-term broad-scale disturbance: Regional forest dynamics in central New England. Ecosystems 1:96–119.

Foster, D., F. Swanson, J. Aber, I. Burke, N. Brokaw, D. Tilman, and A. Knapp. 2003. The importance of land-use legacies to ecology and conservation. BioScience 53:77–88.

Friedlingstein, P., M. Meinshausen, V. K. Arora, C. D. Jones, A. Anav, S. K. Liddicoat, and R. Knutti. 2014. Uncertainties in cMIP5 climate projections due to carbon cycle feedbacks. Journal of Climate 27:511–526.

Friedman, S. K., and P. B. Reich. 2005. Regional legacies of logging: Departure from presettlement forest conditions in northern Minnesota. Ecological applications 15:726–744.

Fritschle, J. A. 2008. Reconstructing historic ecotones using the Public Land Survey: The lost prairies of Redwood National Park. Annals of the Association of American Geographers 98:24–39.

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Fuller, J. L., D. R. Foster, J. S. McLachlan, and N. Drake. 1998. Impact of human activity on regional forest composition and dynamics in central New England. Ecosystems 1:76–95.

Gimmi, U., and V. C. Radeloff. 2013. Assessing naturalness in northern Great Lakes forests based on historical land-cover and vegetation changes. Environmental management 52:481–492.

Goring, S., M. G. Pellatt, T. Lacourse, I. R. Walker, and R. W. Mathewes. 2009. A new methodology for reconstructing climate and vegetation from modern pollen assemblages: An example from British Columbia. Journal of Biogeography 36:626–638.

Gray, A. N., T. J. Brandeis, J. D. Shaw, W. H. McWilliams, and P. D. Miles. 2012. Forest Inventory and Analysis database of the United States of America (FIA). Pages 255–264 *in* J. Dengler, J. Oldeland, F. Jansen, M. Chytry, J. Ewald, M. Finckh, F. Glockler, G. Lopez-Gonzalez, R. K. Peet, and J. H. J. Schaminee, editors. Vegetation databases for the 21st century.– Biodiversity & Ecology.

Groffman, P. M., L. E. Rustad, P. H. Templer, J. L. Campbell, L. M. Christenson, N. K. Lany, A. M. Socci, M. A. Vadeboncoeur, P. G. Schaberg, G. F. Wilson, and others. 2012. Long-term integrated studies show complex and surprising effects of climate change in the northern hardwood forest. Bioscience 62:1056–1066.

Grossmann, E. B., and D. J. Mladenoff. 2008. Farms, fires, and forestry: Disturbance legacies in the soils of the northwest Wisconsin (USA) sand plain. Forest Ecology and Management 256:827–836.

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Hanberry, B. B., S. Fraver, H. S. He, J. Yang, D. C. Dey, and B. J. Palik. 2011. Spatial pattern corrections and sample sizes for forest density estimates of historical tree surveys. Landscape Ecology 26:59–68.

Hanberry, B. B., B. J. Palik, and H. S. He. 2012a. Comparison of historical and current forest surveys for detection of homogenization and mesophication of Minnesota forests. Landscape Ecology 27:1495–1512.

Hanberry, B. B., J. Yang, J. M. Kabrick, and H. S. He. 2012b. Adjusting forest density estimates for surveyor bias in historical tree surveys. The American Midland Naturalist 167:285–306.

Hartig, F., J. Dyke, T. Hickler, S. I. Higgins, R. B. OâHara, S. Scheiter, and A. Huth. 2012. Connecting dynamic vegetation models to data–an inverse perspective. Journal of Biogeography 39:2240–2252.

Haxeltine, A., and I. C. Prentice. 1996. BIOME3: An equilibrium terrestrial biosphere model based on ecophysiological constraints, resource availability, and competition among plant functional types. Global Biogeochemical Cycles 10:693–709.

Heffernan, J. B., P. A. Soranno, M. J. Angilletta Jr, L. B. Buckley, D. S. Gruner, T. H. Keitt, J. R. Kellner, J. S. Kominoski, A. V. Rocha, J. Xiao, T. K. Harms, S. J. Goring, L. E. Koenig, W. H. McDowell, H. Powell, A. D. Richardson, C. A. Stow, R. Vargas, and K. C. Weathers. 2014. Macrosystems ecology: Understanding ecological patterns and processes at continental scales. Frontiers in Ecology and the Environment 12:5–14.

Hijmans, R. J. 2014. raster: Geographic data analysis and modeling.

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Hobbs, R. J., E. S. Higgs, and J. A. Harris. 2014a. Novel ecosystems: Concept or inconvenient reality? A response to murcia et al. Trends in ecology & evolution 29:645–646.

Hobbs, R. J., E. Higgs, C. M. Hall, P. Bridgewater, F. S. Chapin III, E. C. Ellis, J. J. Ewel, L. M. Hallett, J. Harris, K. B. Hulvey, S. T. Jackson, P. L. Kennedy, C. Kueffer, L. Lach, T. C. Lantz, A. E. Lugo, J. Mascaro, S. D. Murphy, C. R. Nelson, M. P. Perring, D. M. Richardson, T. R. Seastedt, R. J. Standish, B. M. Starzomski, K. N. Suding, P. M. Tognetti, L. Yakob, and L. Yung. 2014b. Managing the whole landscape: Historical, hybrid, and novel ecosystems. Frontiers in Ecology and the Environment 12:557–564.

Hotchkiss, S. C., R. Calcote, and E. A. Lynch. 2007. Response of vegetation and fire to Little Ice Age climate change: Regional continuity and landscape heterogeneity. Landscape Ecology 22:25–41.

Iverson, L. R., and D. McKenzie. 2013. Tree-species range shifts in a changing climate: Detecting, modeling, assisting. Landscape Ecology 28:879–889.

Iverson, L. R., and A. M. Prasad. 1998. Predicting abundance of 80 tree species following climate change in the eastern United States. Ecological Monographs 68:465–485.

Jacques, J.-M. S., B. F. Cumming, and J. P. Smol. 2008. A pre-European settlement pollen–climate calibration set for Minnesota, USA: Developing tools for palaeoclimatic reconstructions. Journal of Biogeography 35:306–324.

Jenkins, J. C., D. C. Chojnacky, L. S. Heath, R. A. Birdsey, and others. 2004. Comprehensive database of diameter-based biomass regressions for North American tree species.

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Knoot, T. G., L. A. Schulte, J. C. Tyndall, and B. J. Palik. 2010. The state of the system and steps toward resilience of disturbance-dependent oak forests. Ecology and Society 15:5.

Kronenfeld, B. J. 2014. Validating the historical record: A relative distance test and correction formula for selection bias in presettlement land surveys. Ecography.

Kronenfeld, B. J., and Y.-C. Wang. 2007. Accounting for surveyor inconsistency and bias in estimation of tree density from presettlement land survey records. Canadian Journal of Forest Research 37:2365–2379.

Kronenfeld, B. J., Y.-C. Wang, and C. P. Larsen. 2010. The influence of the "Mixed Pixel"? Problem on the detection of analogous forest communities between presettlement and present in western New York. The Professional Geographer 62:182–196.

Leahy, M. J., and K. S. Pregitzer. 2003. A comparison of presettlement and present-day forests in northeastern lower Michigan. The American Midland Naturalist 149:71–89.

Li, D.-J., and D. M. Waller. 2014. Drivers of observed biotic homogenization in pine barrens of central Wisconsin. Ecology.

Liu, F., D. J. Mladenoff, N. S. Keuler, and L. S. Moore. 2011. Broadscale variability in tree data of the historical Public Land Survey and its consequences for ecological studies. Ecological Monographs 81:259–275.

Lynch, E. A., R. Calcote, S. C. Hotchkiss, and M. Tweiten. 2014. Presence of lakes and wetlands decreases resilience of jack pine ecosystems to late-holocene climatic changes. Canadian Journal of Forest Research 44:1331–1343.

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Maechler, M., P. Rousseeuw, A. Struyf, M. Hubert, and K. Hornik. 2014. Cluster: Cluster analysis basics and extensions.

Manies, K. L., and D. J. Mladenoff. 2000. Testing methods to produce landscape-scale presettlement vegetation maps from the US Public Land Survey records. Landscape Ecology 15:741–754.

Manies, K. L., D. J. Mladenoff, and E. V. Nordheim. 2001. Assessing large-scale surveyor variability in the historic forest data of the original US Public Land Survey. Canadian Journal of Forest Research 31:1719–1730.

Martinuzzi, S., G. I. Gavier-Pizarro, A. E. Lugo, and V. C. Radeloff. 2015. Future land-use changes and the potential for novelty in ecosystems of the united states. Ecosystems:1–11.

Matthes, J. H., S. Goring, J. W. Williams, and M. C. Dietze. in revision. Historical vegetation reconstruction benchmarks CMIP5 pre-colonial land-climate feedbacks across the upper Midwest and northeastern United States. Global Change Ecology.

McAndrews, J. H. 1988. Human disturbance of North American forests and grasslands: The fossil pollen record. Pages 673–697 *in* Vegetation history. Springer.

Menz, M. H., K. W. Dixon, and R. J. Hobbs. 2013. Hurdles and opportunities for landscape-scale restoration. Science 339:526–527.

Mladenoff, D. J., S. E. Dahir, E. V. Nordheim, L. A. Schulte, and G. G. Guntenspergen. 2002. Narrowing historical uncertainty: Probabilistic classification of ambiguously identified tree species in historical forest survey data. Ecosystems 5:539–553.

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Morisita, M. 1954. Estimation of population density by spacing method. Memoirs of the Faculty of Science Kyushu University, Series E 1:187–197.

Morisita, M. 1957. A new method for the estimation of denisty by the spacing method applicable to non-randomly distributed populations (in Japanese). Physiology and Ecology (Kyoto) 7:134–144.

Munoz, S. E., D. J. Mladenoff, S. Schroeder, and J. W. Williams. 2014. Defining the spatial patterns of historical land use associated with the indigenous societies of eastern North America. Journal of Biogeography.

Murcia, C., J. Aronson, G. H. Kattan, D. Moreno-Mateos, K. Dixon, and D. Simberloff. 2014. A critique of the 'novel ecosystem'concept. Trends in ecology & evolution 29:548–553.

Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and H. Wagner. 2014. vegan: Community ecology package.

Paciorek, C. J., and J. S. McLachlan. 2009. Mapping ancient forests: Bayesian inference for spatio-temporal trends in forest composition using the fossil pollen proxy record. Journal of the American Statistical Association 104:608–622.

Paciorek, C., S. J. Goring, A. Thurman, C. Cogbill, J. Williams, D. Mladenoff, J. Peters, J. Zhu, and M. J.S. in review. Statistically-estimated tree composition for the northeastern United States at the time of Euro-American settlement. PLoS One.

Pebesma, E., and R. Bivand. 2005. Classes and methods for spatial data in R. R News 5.

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Pederson, N., J. M. Dyer, R. W. McEwan, A. E. Hessl, C. J. Mock, D. A. Orwig, H. E. Rieder, and B. I. Cook. 2014. The legacy of episodic climatic events in shaping temperate, broadleaf forests. Ecological Monographs 84:599–620.

Persson, O. 1971. The robustness of estimating density by distance measurements. *in* G. P. Patil, E. C. Pielou, and W. E. Waters, editors. Statistical Ecology: Sampling and Modeling Biological Populations and Population Dynamics. Pennsylvania University Press, University Park, Pennsylvania.

Radeloff, V. C., D. J. Mladenoff, and M. S. Boyce. 2000. A historical perspective and future outlook on landscape scale restoration in the northwest Wisconsin pine barrens. Restoration Ecology 8:119–126.

Radeloff, V. C., J. W. Williams, B. L. Bateman, K. D. Burke, S. K. Carter, E. S. Childress, K. Cromwell, C. Gratton, A. O. Hasley, B. M. Kraemer, A. W. Latzka, E. Marin-Spiotta, C. D. Meine, S. E. Munoz, T. M. Neeson, A. M. Pidgeon, A. R. Rissman, R. J. Rivera, L. M. Szymanski, and J. Usinowicz. in press. The rise of novelty in ecosystems. Ecological Applications.

Ramankutty, N., and J. A. Foley. 1999. Estimating historical changes in global land cover: Croplands from 1700 to 1992. Global biogeochemical cycles 13:997–1027.

Ratajczak, Z., J. B. Nippert, and S. L. Collins. 2012. Woody encroachment decreases diversity across North American grasslands and savannas. Ecology 93:697–703.

Rayburn, A. P., and L. A. Schulte. 2009. Integrating historic and contemporary data to delineate potential remnant natural woodlands within Midwestern agricultural landscapes. Natural Areas Journal 29:4–14.

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Record, S., M. C. Fitzpatrick, A. O. Finley, S. Veloz, and A. M. Ellison. 2013. Should species distribution models account for spatial autocorrelation? A test of model projections across eight millennia of climate change. Global ecology and biogeography 22:760–771.

Rhemtulla, J. M., D. J. Mladenoff, and M. K. Clayton. 2009a. Historical forest baselines reveal potential for continued carbon sequestration. Proceedings of the National Academy of Sciences 106:6082–6087.

Rhemtulla, J. M., D. J. Mladenoff, and M. K. Clayton. 2009b. Legacies of historical land use on regional forest composition and structure in wisconsin, USA (mid-1800s-1930s-2000s). Ecological Applications 19:1061–1078.

Schulte, L. A., and D. J. Mladenoff. 2001. The original US Public Land Survey records: Their use and limitations in reconstructing presettlement vegetation. Journal of Forestry 99:5–10.

Schulte, L. A., and D. J. Mladenoff. 2005. Severe wind and fire regimes in northern forests: Historical variability at the regional scale. Ecology 86:431–445.

Schulte, L. A., D. J. Mladenoff, and E. V. Nordheim. 2002. Quantitative classification of a historic northern Wisconsin (USA) landscape: Mapping forests at regional scales. Canadian Journal of Forest Research 32:1616–1638.

Schulte, L. A., D. J. Mladenoff, S. N. Burrows, T. A. Sickley, and E. V. Nordheim. 2005. Spatial controls of pre–Euro-American wind and fire disturbance in northern Wisconsin (USA) forest landscapes. Ecosystems 8:73–94.

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Schulte, L. A., D. J. Mladenoff, T. R. Crow, L. C. Merrick, and D. T. Cleland. 2007. Homogenization of northern US Great Lakes forests due to land use. Landscape Ecology 22:1089–1103.

Seddon, P. J. 2010. From reintroduction to assisted colonization: Moving along the conservation translocation spectrum. Restoration Ecology 18:796–802.

Staver, A. C., S. Archibald, and S. A. Levin. 2011. The global extent and determinants of savanna and forest as alternative biome states. Science 334:230–232.

Stewart, L. O. 1935. Public Land Surveys: History, instructions, methods. Page 202. Collegiate Press, Incorporated, Ames, Iowa, USA.

Stoy, P. C., M. Mauder, T. Foken, B. Marcolla, E. Boegh, A. Ibrom, M. A. Arain, A. Arneth, M. Aurela, C. Bernhofer, and others. 2013. A data-driven analysis of energy balance closure across FLUXNET research sites: The role of landscape scale heterogeneity. Agricultural and forest meteorology 171:137–152.

Team, R. C. 2014. R: A language and environment for statistical computing (version 3.1. 0). vienna, Austria: R Foundation for Statistical Computing.

Terrail, R., D. Arseneault, M.-J. Fortin, S. Dupuis, and Y. Boucher. 2014. An early forest inventory indicates high accuracy of forest composition data in pre-settlement land survey records. Journal of Vegetation Science 25:691–702.

Thompson, J. R., D. N. Carpenter, C. V. Cogbill, and D. R. Foster. 2013. Four centuries of change in northeastern United States forests. PloS one 8:e72540.

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Tomscha, S. A., and S. E. Gergel. 2014. Historic land surveys present opportunities for reconstructing frontier settlement patterns in North America. Landscape Ecology 30:203–213.

Umbanhowar, C. E., P. Camill, C. E. Geiss, and R. Teed. 2006. Asymmetric vegetation responses to mid-Holocene aridity at the prairie–forest ecotone in south-central Minnesota. Quaternary Research 66:53–66.

White, C. A. 1983. A history of the rectangular survey system. Page 774. US Department of the Interior, Bureau of Land Management, Washington D.C., USA.

Wickham, H. 2007. Reshaping data with the reshape package. Journal of Statistical Software 21:1–20.

Wickham, H. 2009a. ggplot2: Elegant graphics for data analysis. Springer.

Wickham, H. 2009b. ggplot2: Elegant graphics for data analysis. Springer New York.

Wickham, H. 2011. The split-apply-combine strategy for data analysis. Journal of Statistical Software 40:1–29.

Williams, M. A., and W. L. Baker. 2011. Testing the accuracy of new methods for reconstructing historical structure of forest landscapes using GLO survey data. Ecological Monographs 81:63–88.

Wolkovich, E., B. Cook, K. McLauchlan, and T. Davies. 2014. Temporal ecology in the anthropocene. Ecology letters 17:1365–1379.

bioRxiv preprint doi: https://doi.org/10.1101/026575; this version posted September 11, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Wood, S. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society (B) 73:3–36.

Woudenberg, S. W., B. L. Conkling, B. M. O'Connell, E. B. LaPoint, J. A. Turner, and K. L. Waddell. 2010. The Forest Inventory and Analysis database: Database description and users manual version 4.0 for phase 2.