Witness Tree paper

Simon Goring et al.

28 January, 2015

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

Simon J. Goring[1]

John W. Williams[1,2]

David J. Mladenoff[3]

Charles V. Cogbill[4]

Sydne Record[4]

Christopher J. Paciorek[6]

Stephen T. Jackson[5]

Michael C. Dietze[7]

Jaclyn Hatala Matthes[7]

Jason S. McLachlan[8]

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

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

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

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

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

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

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

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

Abstract

EuroAmerican land use and its legacies have transformed forest structure and composition in many regions of the United States. More accurate reconstructions of historical states are critical to understanding the processes governing past, current, and future forest dynamics.

We develop and present a new dataset of gridded (8x8km) estimates of pre-settlement vegetation from the upper Midwestern United States (Minnesota, Wisconsin, and most of Michigan) using the Public Land Survey

(PLS), with estimates of relative composition for XX tree genera and total biomass, stem density, and basal area. These estimates of stem density correct for variations in surveyor sampling by applying spatially varying correction factors that accommodate alternate sampling designs, azimuthal censoring, and biases in tree selection. We applied these estimates to map forest, prairie and savanna distributions, and to reconstruct spatial patterns in forest composition and structure, which we compare to modeled potential vegetation maps that are widely used by terrestrial ecosystem modelers as historic baselines. We compare pre-settlement to modern forests using Forest Inventory and Analysis (FIA) data, both with respect to structural changes and the prevalence of forests with no current or past analogs.

Stem density and total basal area are higher in contemporary forests than in settlement-era forests, but aboveground biomass is higher in settlement-era forests, in part because individual settlement-era trees are larger on average. Modern forests are more homogenous than settlement-era forests, and ecotonal gradients are less sharp today than in the past. Almost 24% of FIA cells represent novel forests, with no close analog to settlement-era forest composition, while 20% of pre-settlement forests no longer exist in a modern context. The loss of PLS forest types is spatially structured, concentrated particularly in hemlock- and beech-dominated forests at the tension zone between deciduous and evergreen forests. This loss changes the structure of major ecotones across the region, including forest-prairie, sub-boreal-deciduous and mixed-wood to prairie ecotones. Novel FIA forest assemblages are distributed evenly across the region, representing a broad-scale homogenization of forest composition and structure that is strongly influenced by current and historic land use. All datasets and open-source analytical scripts are publicly available, and work is underway to extend these settlement-era reconstructions across the northeastern US forests.

Key Words: forest composition, land use change, paleoecology, public land survey

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 (Pederson et al. 2014). Legacies of past 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 impacted by changing land use in the past three centuries. Hence, observed ecological processes in North American forests reflect both anthropogenic and natural influences at decadal to centennial scales. These dual influences pose the challenge that natural processes may be masked, or heavily modified by anthropogenic effects.

At a regional scale many forests in the upper Midwest (Minnesota, Wisconsin and Michigan) now have decreased species richness and functional diversity relative to forests of the pre-Euro-American settlement period [hereafter pre-settlement) (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 recovery. 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 settlement at a regional scale is of critical importance to biogeochemical and biogeophysical vegetation-atmosphere feedbacks (Matthes et al. in revision), 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).

The extent to which changes in forest composition since Euro-American settlement [e.g. Schulte et al. (2007);hanberry2012homogenization] has altered species co-occurance and affects our ability to understand individual or community responses to climate change is unknown: Are we measuring contemporary forests that have no past analogs to predict a future with no contemporary analogs? At what scales is homogenization ocuring? How has forest change affected ecotones across the region? Pockets of primary forest originating before the settlement period exist in the Midwestern United States and have been heavily studied by ecologists

REFS, but because land use has been both extensive and strongly selective, these pockets are an incomplete and unrepresentative sample of past vegetation.

Modern forest structure and composition data (e.g., from the USDA 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 Euro-American settlement. The Public Land Survey provides broad coverage of the United States prior to Euro-American settlement (Almendinger 1996, Liu et al. 2011). In general, FIA datasets are well organized and widely available to the forest ecology and modeling community, whereas most PLS studies have been limited to local to state-level extents. This absence of widely available data on settlement-era forest composition and structure severely limits our ability to understand and model the current and future processes governing forest dynamics. For example, distributional models of tree species often rely upon FIA or other contemporary observational data to build species-climate relationships that can be used to predict potential range shifts (Iverson and Prasad 1998, Iverson and McKenzie 2013). If land use change is impacting the shape of the realized niche then species distribution modeling, based on correlative models relating current extent to local climatic and edaphic factors, may have much weaker ability to interpret species responses to climate change than model diagnostics might indicate. Indeed, re-organization following land use change may result in shifts in species co-occurance, and the development of novel ecosystems. Thus, while Hobbs *et al.* (2006) argued that novel ecosystems were understudied, it may well be the case that in some regions they are all that's been studied.

Here we use survey data from the original Public Lands Surveys (PLS) in the upper Midwest (here defined as Minnesota, Wisconsin, and Michigan) 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. 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 with coarser spatial aggregations (Schulte et al. 2007, Hanberry et al. 2012a), although aggregation may also occur at the township scale (Rhemtulla et al. 2009a, Kronenfeld et al. 2010). Our work addresses several major challenges to using 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.

We aggregate point based estimates of stem density, basal area and biomass to an 8 x 8km grid to help reduce local scale uncertainty, 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 from FIA data. Using the gridded data we examine species co-occurence and two spatial transects across the domain to understand the extent of change along forest ecotones at a broad spatial scale, and along well-understood ecotones from southern deciduous to northern evergreen forests and to the forest-prairie boundary.

Methods:

Public Lands Survey Data: Assembly, and Standardization

The U.S. Public Land Survey (PLS) was designed to facilitate the division and sale of land 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 two to four trees near sample points 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, 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 explicit point level data, providing information about tree spacing and diameter across an extensive region. 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 dispersed spatial sampling within the dataset (fixed sampling every 1 mile), uncertainty 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 survey bias. Terrail *et al.* (2014) show strong fidelity between taxon abundance in early land surveys and from 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 the authors also indicate that the true sampling error cannot be determined, particularly since 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 angular bias bias (*i.e.*, 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 years have assessed and corrected for the idiosyncrasies of the original surveyor data to minimize sampling bias (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). Use of the PLS data requires caution, and careful consideration of scale, bias and variability. 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 Euro-American settlement.

This analysis builds upon and merges prior 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*).

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 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, the non-random habitat preferences of meander trees (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 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 the 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 added layer of taxonomic uncertainty, and retained only genus level identification (Supplement 2, *Standardized taxonomy*). 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 have been extremely open, with an estimated point level stem density of 0 stems/ha. We based our estimates on terrestrial coverage, so water cells are excluded completely so that absence does not reduce overal terrestrial stem density.

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 transcrition 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 (dbh > 100“; 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.

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 (Thurman et al. in prep). This scale is smaller than the 100km^2 gridded scale used in Freidman and Reich (2005), but larger than township grids used other studies(Rhemtulla et al. 2009a, Kronenfeld 2014) to provide a scale comparable to aggregated FIA data at a broader scale. Forest composition data is based on the number of individuals of each taxon (genera or plant functional types, PFTs) 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, Basal Area and Biomass Estimates

Estimating stem density, basal area and biomass from PLS data requires various corrections to minimize sampling bias in the original surveyor measurements (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 townships (White 1983, Liu et al. 2011). The changing plot geometry across the region thus precludes uniform methods for correcting for surveyor bias when aggregating from points to grids because there is significant variability between sampling geometries among plots (Cogbill et al. in prep). Our approach allows for spatial variation in surveyor methods by applying spatially varying correction factors that are based on plot geometry, angle distributions and diameter censoring within the data set.

We estimate stem density (stems m^{-2}) based on a modification of the distance-to-tree measurements for the two closest trees at each point (Morisita 1954) using explicit and spatially varying correction factors, modeled after the Cottam correction factor (Cottam and Curtis 1956), that accounts for variations in sampling designs over time and among surveyors. All code to perform the analysis is included in Supplement 3.

Stem density is calculated using a modified form of the Morisita equation. The original Morisita equation is expressed as:

$d_i = \frac{q-1}{\pi \times n} \times \frac{q}{\sum_{j=1}^{q} (r_j)^2}$

Here, d_i is the density estimate at point i, q is the number of sectors sampled (all $q = 2$), r_j is the distance from the plot center to tree j (in meters).

The modified form of the Morisita equations is expressed as:

$d_i = \kappa \times \frac{q-1}{\pi \times n} \times \frac{q}{\sum_{j=1}^{q}(r_j)^2} \times \theta \times \zeta \times \phi$

The modified form of the Morisita equation adds several parameters to adjust for 1) departures from the standard point-centered quarter method assumed by Morisita and 2) several potential sources of surveyor bias when selecting trees. We account for variations in sample design, in particular the lack of a true point-centered quarter method using κ, the modified Cottam correction factor. The parameter θ accounts for deviation from the expected pair angle distribution associated with differences in plot geometry. As plots change from two trees in a 180^o semi-circle (exterior points) to true point-centered quarters, the distribution of pair angles changes along with the sampled area. Our θ estimates are empirically derived with $\theta = 1$ representing a true point-quarter sampling design.

The parameter ζ is used to correct for azimuthal censoring, since surveyors appear to censor certain azimuths (Kronenfeld and Wang 2007), particularly along cardinal directions of travel, a pattern that varies across plot types. In this model a $\zeta = 1$ would be applied to sets of plots for which no azimuthal censoring is visible. Plots with a greater degree of censoring, particularly along the cardinal directions obtain higher correction factors, up to 25% (Figure 2). The last correction factor, ϕ is used to correct for the inclusion of stems under 8" in diameter within a region, which would otherwise replace trees of larger diameters at greater distances. The full application rationale for and calculation of these measures is described further in Cogbill *et al.* (in prep). Correction factors were calculated separately for different regions, years, internal versus external section points, and surveyor sampling designs (Table 1).

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) 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:

$m = Exp(\beta_0 + \beta_1 * \ln dbh)$

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 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).

Matching PLSS tree taxa to the species groups defined by Jenkins *et al.* (2004) is straightforward, placing the 22 taxa 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.

Forest Classification

We convert the PLS data to global vegetation categories as defined by Haxeltine and Prentice (1996) (Table 3) in order to create vegetation maps that can be used as initial states for ecosystem models. We use the stem density thresholds of Anderson and Anderson (1975) to discriminate prairie, savanna, and forest. We set >70% relative abundance of NE and ND, or BD as a threshold for PFT dominance and to discriminate among the temperate deciduous, temperate evergreen, and mixedwood forest types. (Table 3). We compare these PLS-based maps to prior maps of potential vegetation by Ramankutty and Foley (1999), which are widely used by ecosystem modelers to set initial states for 20th- and 21st-century simulations.

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 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), stem density (stems/ha), mean diameter at breast height (cm), and mean biomass (Mg/ha) 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 are generated by averaging all PLS point or FIA plot estimates within a 64km^2 raster cell. Most 64km^2 cells have one or few FIA plots that are sampled intensively. Therefore at this scale of aggregation, the relatively low density of FIA plots in heterogeneous forests could result in high within-cell variance and high between-cell variability. Stem density estimates from the PLS data are highly 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 sample size, but have low variance across the landscape due to the uniform sampling pattern of the data. Thus while within-cell variance is expected to be high for the PLS point data and 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: `ggplot2` (Wickham 2009a,Wickham (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).

Differences in composition between and within PLS and FIA datasets are examined using Bray-Curtis dissimilarity (vegdist in vegan; (Oksanen et al. 2014)) for proportional composition within raster cells using basal area measurements. For the purposes of this analysis we are interested only in the minimum compositional distance between a focal cell and its nearest compositional (not spatial) neighbor. The distribution of closest analogs within datasets provides information about forest heterogeneity within each time period, while the search for closest analogs between datasets provides information about whether contemporary forests lack analogs in pre-settlement forests, and vice versa. 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 (FIA to FIA, PLS to PLS), and between datasets (PLS to FIA and FIA to PLS), retaining only the minimum value. For the FIA to FIA and PLS to PLS analyses, cells were not allowed to match to themselves.

Comparisons between the gridded PLS and FIA potentially are affected by differences in sampling design. The PLS usually samples across different kinds of forest types within a given 64km^2 cell (a “mixed pixel” problem, (Kronenfeld et al. 2010)), whereas each FIA plot samples from a single forest type. This difference should mainly manifest as differences in within-cell and between-cell heterogeneity and should not affect expected values. The effects of differences in scale should be strongest in regions where there are few FIA plots per 64 km2 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 gridcells 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.

Results:

Data Standardization

The original PLS dataset contains 519235 corner points (excluding line and meander points), with 171433 points from Wisconsin, 251660 points from Minnesota and 96142 points from Michigan. Standardizing data and accounting for potential outliers, described above, removed approximately 1.5% points from the dataset.

Rasterizing the PLS dataset to the Albers 64km^2 grid produces 7970 raster cells with data. Each cell contains between 1 and 94 corner points, with a mean of 65 ($\sigma = 15$) and a median of 69 corners (Supplement 3). Cells with a low number of points were mainly near water bodies or along political boundaries such as the Canadian/Minnesota border, or southern Minnesota and Wisconsin borders. Only 2% of cells have less than 10 points per cell.

Species assignments to genera were rarely problematic. Only 0.076% of PLS trees were assigned to the Unknown Tree category, largely consisting of corner trees for which taxon could not be interpreted, but for which diameter and azimuth data was recorded. A further 0.0105% of trees were assigned to the "Other hardwood" taxon.

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 140 stems/ha. Stem density exclusive of prairie is 160 stems/ha and is 210 stems/ha when both prairie and savanna are excluded. Stem density has a 95th percentile range from 0 to 400 stems/ha with cell-level standard deviations between 0 and 430 stems/ha. Basal area in the domain (Figure 3c) has a 95th percentile range between 0 and 59 m^2/ha, a mean of 20.7 m^2/ha and cell level standard deviations between 0 and 72 m^2/ha. Biomass ranges from 0 to 200 Mg/ha (Figure 3d), with cell level standard deviations between 0 and 590 Mg/ha. High cell level standard deviations relative to mean values within cells for density, basal area and biomass indicate high levels of heterogeneity within cells, as might be expected for the PLS data, which samples only two trees every mile.

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 for Savanna, Table 3) are used, the extent of area classified as savanna is more limited in southern Wisconsin (Figure 3b) than in prior reconstructions (Curtis 1959, Bolliger et al. 2004, Rhemtulla et al. 2009b), despite generally low stem densities (Figure 3a). 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. The interplay between broad-scale climatic structuring and local hydrological controls on forest composition and density can be seen, particularly along the Minnesota River in south-western Minnesota, where a corridor of savanna is sustained in a region mostly occupied by prairie (Figure 3b).

Forest structure during the settlement era can be understood in part by 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, blue). This indicates the presence of young, small-diameter, even-aged stands, possibly due to frequent stand-replacing fire disturbance in the pre-EuroAmerican period. 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, red) are found in Michigan and southern Wisconsin, indicating regions with greater proportions of deciduous species and higher diameters than northern Minnesota, due to higher productivity soil and climate and lower incidences of stand-replacing disturbance.

Taxon composition within settlement-era forests is heterogenous (Figure 5). Oak is dominant across the region, with an average composition of 21%, however, that proportion drops to 8% when only forested cells are considered, due to its prevelance in the savanna and prairie. pine shows the opposite trend, with average composition of 14% and 18% in unforested and forested cells. 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 composition include maple (8%), birch (8%), tamarack (8%) and hemlock (6%).

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 18 taxa including oak, poplar, ironwood, 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 are more common in the northwestern part of the domain.

Spruce in the PLS represents two species (*Picea glauca*, *Picea mariana*) with overlapping distributions, but strongly different site preferences (dry upland and moist sites respectively). 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% in the northeast). Hardwoods that occupy the southern and western portions of the region, such as oak, elm, basswood and beech, are most typically mono-specific groupings, with the exception of oak, which comprises seven species (see Supplementary Table 1). These taxa 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 (see Supplemental material). These hardwood classes correspond to well-defined vegetation patterns for the region (Curtis 1959). Thus overlap among PFT distributions (Figure 6) emerges from the changing composition within the plant functional type from deciduous broadleaved species associated with the southern, deciduous dominated region, to broadleafed deciduous species associated with more northern regions in the upper Midwest.

Settlement-Era Vegetation Types

Comparison to FIA Composition

Comparing PLS and FIA estimates for stem density, basal area and biomass indicates that the modern forests (FIA) have higher stem densities and total basal areas, but overall, comparable biomass between the PLS and FIA data (Figure 8). 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 FIA appears to have higher average diameters in northern regions, where the relationship between increases in diameter and increases in biomass is expected to be lower because of allometric scaling. The higher average diameters in the Mixedwood and Temperate Deciduous forests of the east (Figure 8d) may be the mechanism by which low density and basal area produce roughly equivalent biomass estimates between the FIA and PLS. 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

Both the PLS- and FIA-era compositional data show similar patterns of within-dataset dissimilarity, wth the highest dissimilarities found in central Minnesota and northwestern Wisconsin (Figure 10b,c). In the

PLS-PLS comparisons (10b) high dissimilarities are associated with high proportions of maple, birch and fir (Figure 5). High FIA-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} = 0.1979074, p < 0.001) are lower than the dissimilarities within the PLSS (I_{PLSS} = 0.4923701, p < 0.001), suggesting lower spatial autocorrelation in the FIA dataset. Cells with identical pairs represent 5.6% of the PLS cells and 7.4 of FIA cells. Identical cells in the PLS are largely located along the southern margin and most (69%) are composed entirely of oak. Cells in the FIA with identical neighbors are composed of either pure oak (19%), pure poplar (26%) or pure ash (14%).

There is a small but significant relationship ($F_{1,5964}$= 920, NA, p < 0.001) between the number of FIA plots and FIA-FIA dissimilarity. The relationship accounts for 10% of total variance and estimates an increase of δ_d = 0.0134277 for every FIA plot within a cell. This increase represents only 3.1% 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.

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 here as lost forests, while forest types in the FIA data with no past analogs are defined as novel. More than 25% of PLS sites have no analogue in the FIA dataset (PLS-FIA dissmimilarity, Figure 10d), while 29% of FIA sites have no analogue in the PLS data (FIA-PLS dissimilarity, Figure 10e). PLS-FIA no-analogues show strong spatial coherence, centered on the "Tension Zone" (Curtis 1959) ecotone between deciduous forests and hemlock-dominated mixed forest (Figure 6). The distribution of FIA-PLS no-analogs (Figure 10e) is spatially diffuse and appears to be shifted north of the PLS-FIA no-analog distribution. PLS-FIA no-analogs (lost forests) are related to higher proportions of beech (r = 0.17), ironwood (r = 0.23), and hemlock (r = 0.06). This loss is reflected in shifts in the transects shown above (Figure 8). Pine and oak show no significant relationship to novelty since they are present throughout the region. This analysis suggests that post-settlement land use had the greatest effect on mesic deciduous forests and the ecotonal transition between southern and northern hardwood forests.

To understand how the ecotone has been transformed by post-settlement land use, we constructed two transects of the FIA and PLS data, and fitted GAM models to genus abundances along these transects. Transect One runs from northern prairie (in northern Minnesota) to southern deciduous savanna in southeastern Wisconsin, while Transect Two runs from southern prairie in southwestern Minnesota to northern mixedwood forest in the Upper Peninsula of Michigan.

For Transect One, GAM models shows significant differences (using AIC) in the curves fit to tamarack, pine, Birch, Oak, Poplar and Spruce between PLS and FIA data, but not for elm, maple or hemlock (which is almost entirely absent along the transect). Pine used to be a major component of the tension zone, reaching proportions of up to 80% in north central Wisconsin. Declines in pine proportions are likely due to logging and subsequent regeneration of other taxa. Increases in pine in southeastern Wisconsin are possibly as a result of the cultivation of pine plantations in the region. Transect 2 also shows significant changes in taxon composition between PLS and FIA data, particularly for tamarack, pine, birch, maple, oak, and hemlock. Oak is almost completely eliminated from the central to western portions of the transect in the FIA data, largely due to land conversion to agriculture in the modern era. The eastern portion of Transect Two shows a shift from a landscape dominated by multiple taxa, pine, birch, maple and hemlock, to one dominated by Maple, along with lower proportions of Poplar (Transect Two, FIA: Figure 9).

Contemporary forests show broader homogenization and increased heterogeneity (evidenced by the lower FIA-FIA Moran's I estimates) at a local scale in the region. Homogenization is evident across Transect One, 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 (2014). The PLS shows strong differentiation in the central region of Transect Two where Maple-Pine-Oak shifts to Pine-Poplar-Birch forest. This sharp ecotone is not apparent in the FIA data, which shows gradual and blurred changes in species composition across the ecotone. β-diversity along Transect Two 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_{beta} = -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

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 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.

Results show clear signs of increased homogenization at local and regional scales and decreased spatial structure in vegetation assemblages. 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. This homogenization also causes the selective loss or weakening of major vegetation ecotones, particularly in central Wisconsin, and the development of novel species assemblages across the region.

All datasets and analytic codes presented here are publicly available and open source (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. 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 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).

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 Euro-American land-use conversion. 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 9), and the rapid rise of novel forest assemblages (Figure 9) have important implications for our ability to understand ecological responses to changing climate. The loss of these forests implies that the modern understanding of forest cover, climate relationships, realized and potential niches and species associations may be strongly biased toward a single state, even though 38% 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-occurrence 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 (Stoy et al. 2013). Our data show higher levels of 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 heterogeneity at a very large spatial scale, and the loss of resolution along major historical ecotones. Thus analysis of the processes governing vegetation heterogeneity and ecotones mayinadvertently and substantially incorporate anthropogenic processes.

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 (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 (Thurman et al. in prep) 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 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.

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 occured 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 andsome quarter section points, suvey instructions indicated four trees were to be sampled, these were '2nQ' plots, wheras 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

State	Survey Year	Internal	Section	Trees	kappa	theta	zeta	phi
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
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

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

Jenkins Species Group	β_0	β_1	PalEON Taxa Included (Supp. 2)
Mixed Hardwood	-2.48	2.48	Ash, Elm, Maple, Basswood, Ironwood, Walnut, Hackberry, Cherries, Do
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

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 (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
Temperate Deciduous	Dominant PFT BDT, LAI > 2.5	Stem dens. > 48 stems/ha, BDT > 70% con
Temperate Conifer	Dominant PFT (NET + NDT), LAI > 2.5	Stem dens. > 47 stems/ha, NET + NDT >
Mixedwood	Both BDT (LAI > 1.5) & NET (LAI > 2.5) present	Stem dens. > 47 stems/ha, BDT & NET bo

Table 4. *Classification proportions, patch size and edge cell proportion for various classification schemes used with the Public Land Survey data. All patch size estimates are in 1000s of km^2.*

Classification Metric	Prairie	Savanna	Temperate Deciduous	Temperate Evergreen	Mixed Wood	Mean
Ramankutty and Foley (1999)	86	63	9	50	241	184
PLS Data	53	93	79	151	98	36

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.*

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.*

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 4*).*

Figure 4. *Variations in the relationship between biomass and stem density can be used to understand forest structure. 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. Only cells greater than 1 standard deviation from the ratio mean are classified as low or high.*

Figure 5. *Forest composition as a percent 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 (white: needleleafed deciduous; light gray: needleleafed evergreen; dark gray: broadleafed deciduous).*

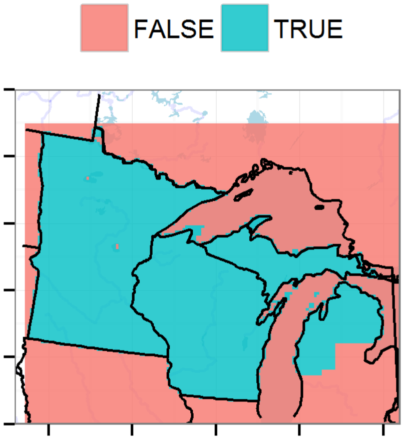
Figure 6. *Proportional distribution of Plant Functional Types (PFTs) in the upper Midwest from PLS data, for broadleaved deciduous trees (BDT), needleleaved deciduous trees (NDT), and needleleaved evergreen trees (NET). Distributions are shown as proportions relative to total basal area, total biomass, 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.*

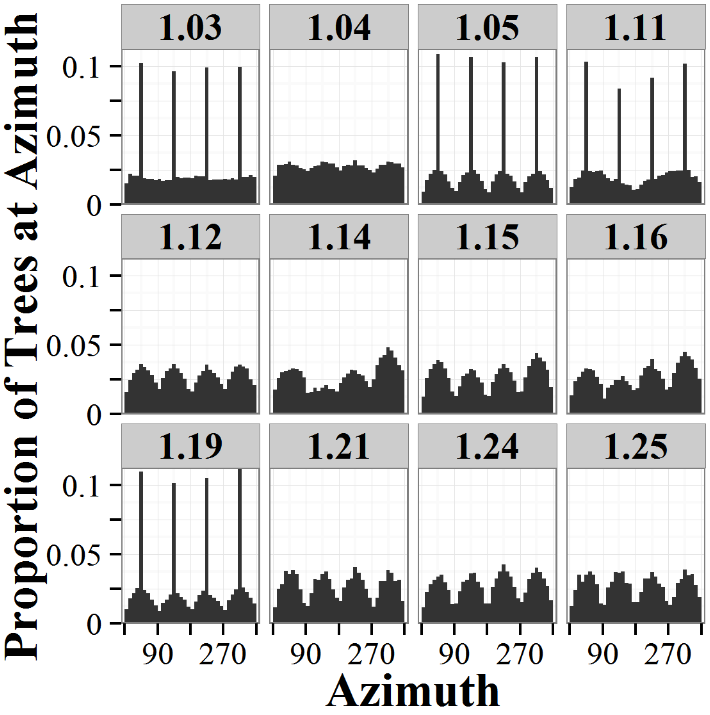
Figure 7. *Forest classification using Ramankutty and Foley (1999) estimates and those using baseline PLS data (aggregated to PFTs for equivalence) with the Haxeltine and Prentice (1996) classification scheme (Table 2). Colors represent forest classes for the region as defined by Ramankutty and Foley (1999): (Pr) Prairie; (Sa) Savanna; (TD) Temperate deciduous; (TE) Temperate evergreen; (MW) Mixedwood forest.*

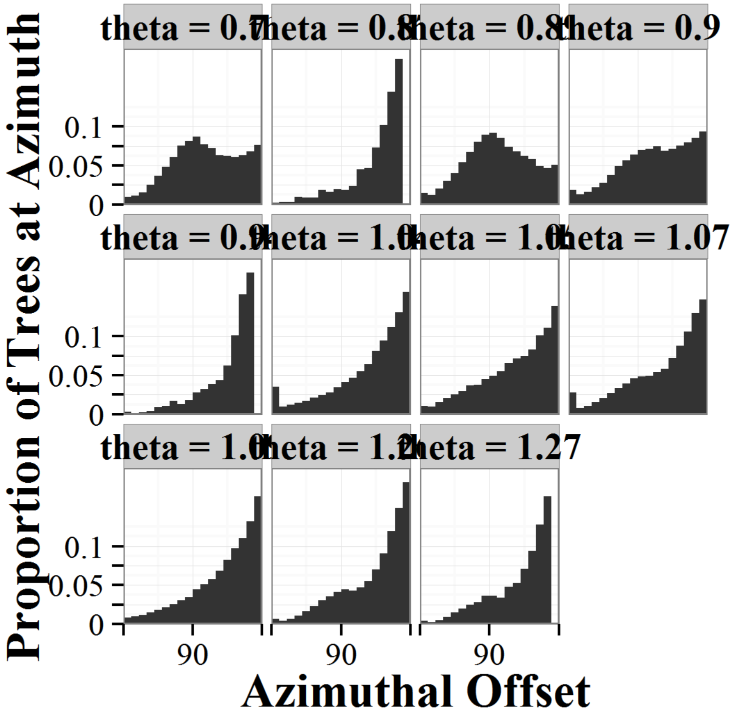
Figure 8. *The relationship between average stem density, total basal area and biomass values in the PLS and FIA datasets. Stem density and total basal area are higher in the FIA than in the PLS, however average cell biomass is higher in the PLS.*

Figure 9. *Minimum dissimilarity maps. Cells with a high minimum dissimilarity lack close compositional analogs to other cells in the reference dataset. The top two panels show compositional heterogeneity within the PLS and FIA data, while the bottom two panels identify PLS locations with no close analogs in the FIA data (PLS to FIA) and FIA locations with no close PLS analogs (FIA to PLS). More than half (55%) of points in the PLS have minimum dissimilarities to the FIA data that are greater than the 95 percentile for minimum dissimilarities within either the FIA or PLS datasets.*

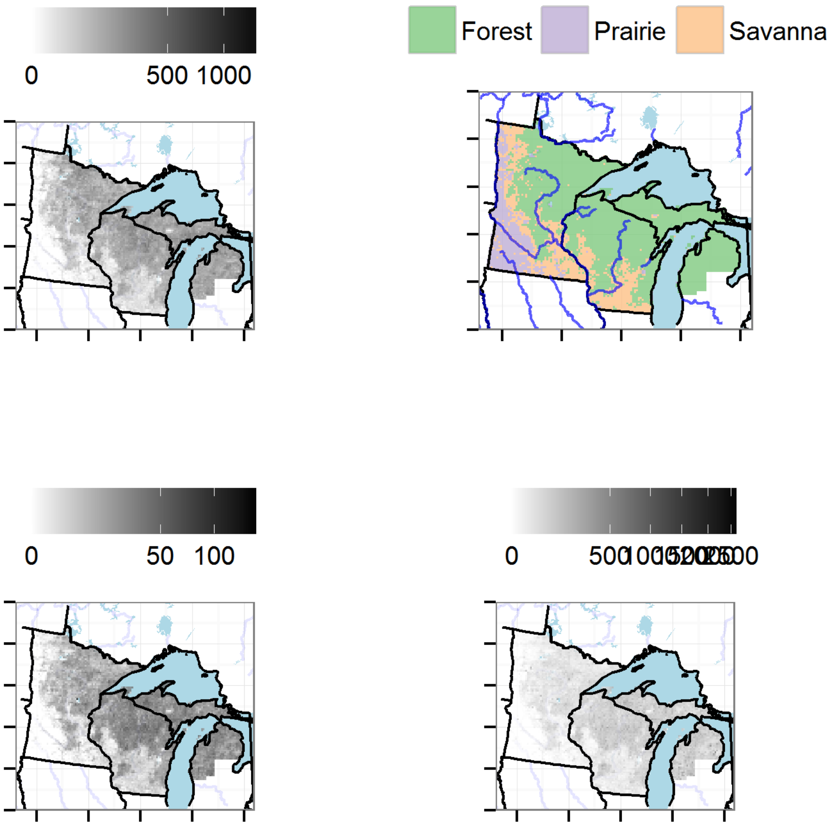
Figure 10. *Transects across the region show clear changes in the forest ecotones from one region to another. Fitted curves represent smoothed estimates across the transects using Generalized Additive Models using beta fits.*

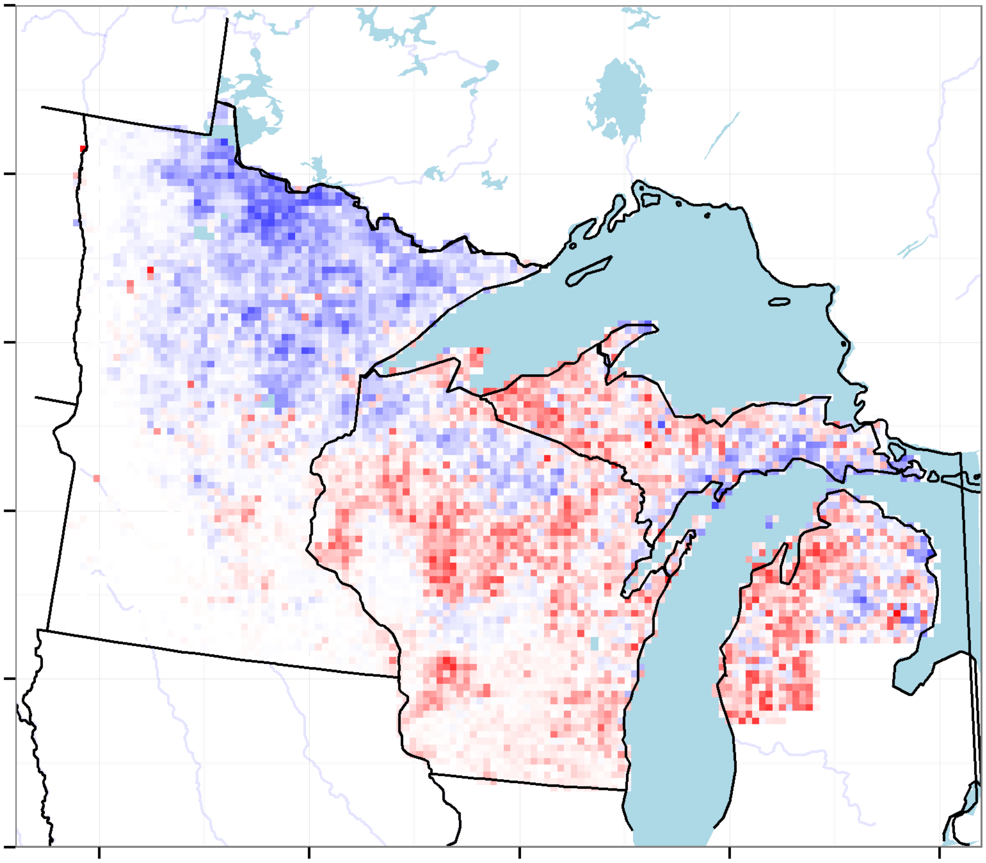

FALSE
TRUE

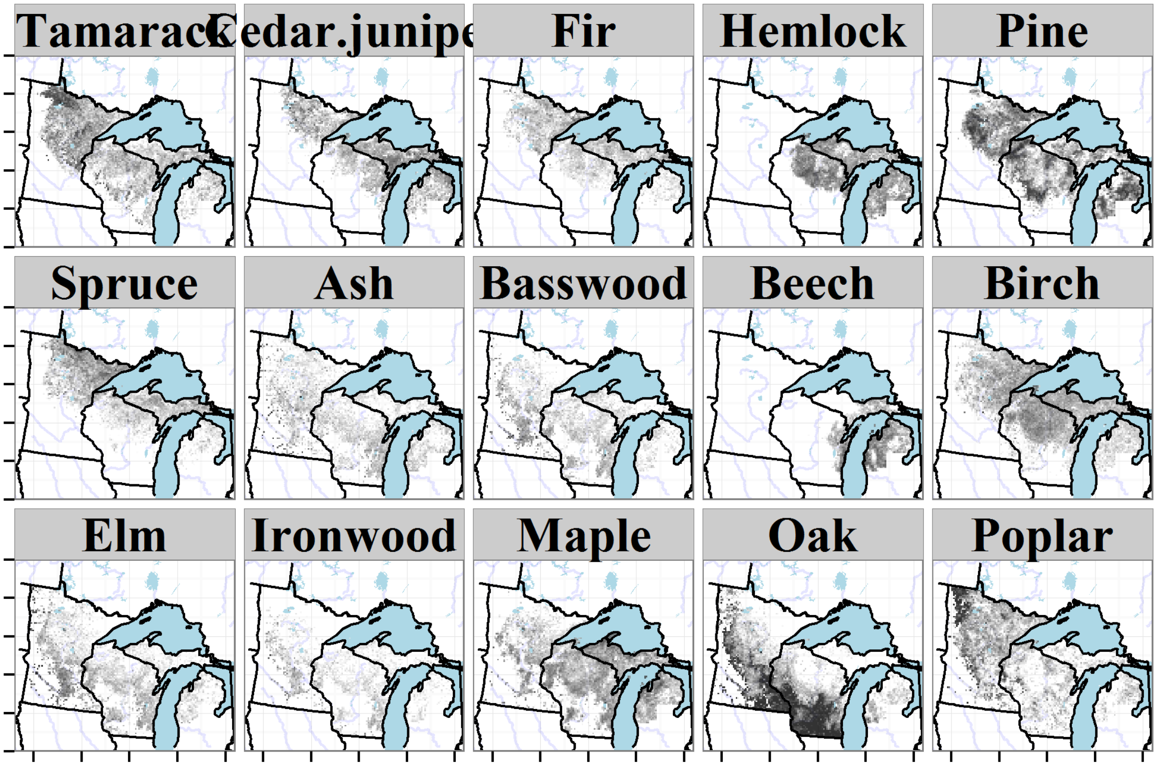

1.03
1.04
1.05
1.11
1.12
1.14
1.15
1.16
1.19
1.21
1.24
1.25
0.1
0.05
0
90
270
Proportion of Trees at Azimuth
Azimuth

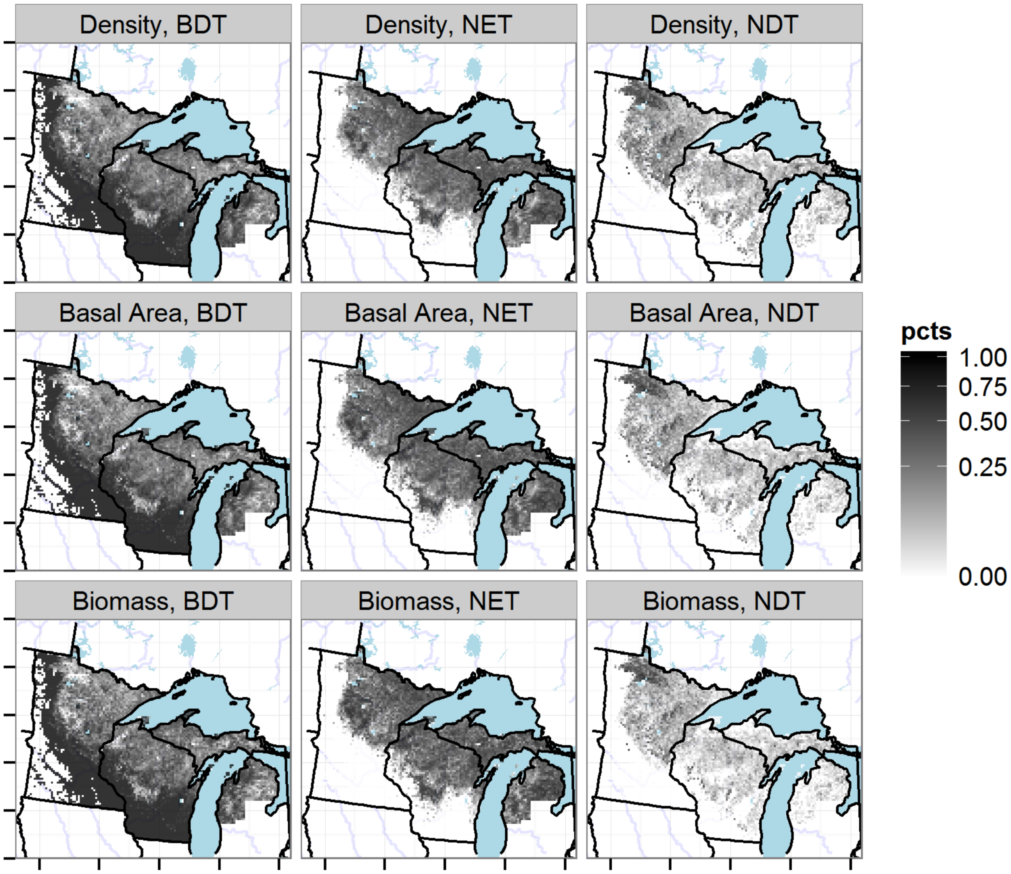


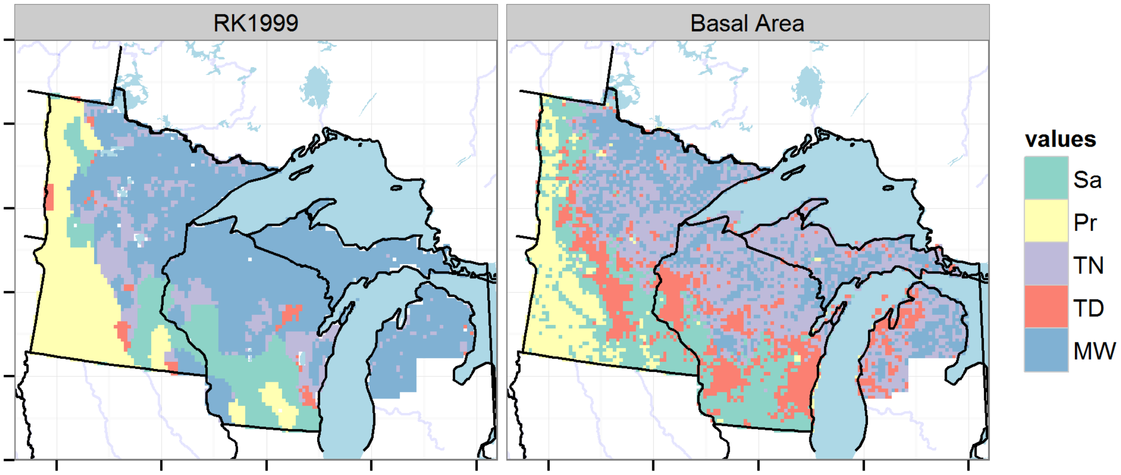
Currently Unscripted Plot Angle pairs for theta.

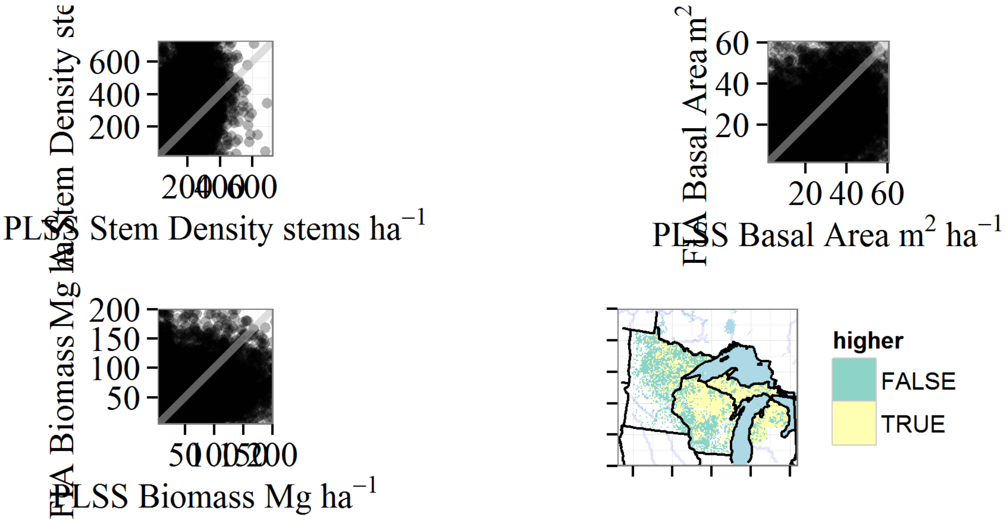

0
500 1000
Forest
Prairie
Savanna
0
50 100
0
500 1000 1500 2000 2500

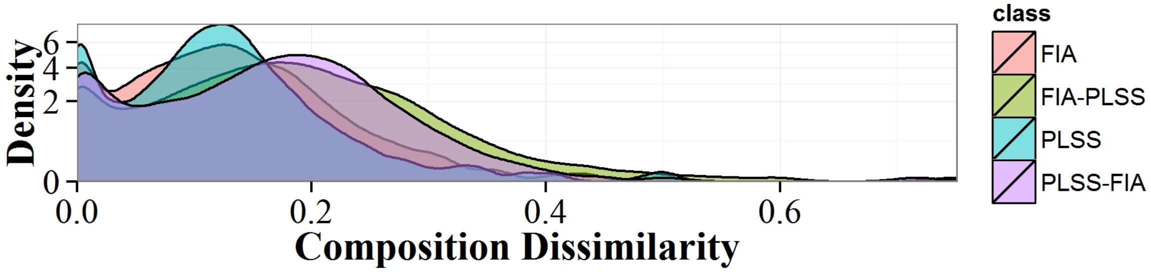


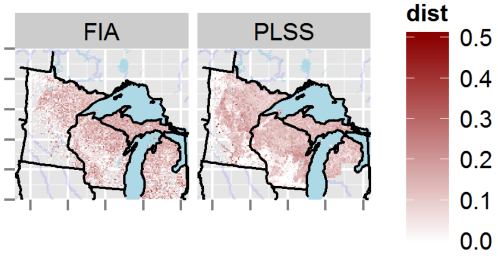

Tamarack
Cedar.junipe
Fir
Hemlock
Pine
Spruce
Ash
Basswood
Beech
Birch
Elm
Ironwood
Maple
Oak
Poplar

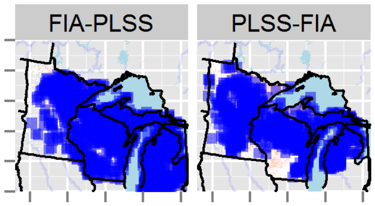

Density, BDT
Density, NET
Density, NDT
Basal Area, BDT
Basal Area, NET
Basal Area, NDT
Biomass, BDT
Biomass, NET
Biomass, NDT
pcts
1.00
0.75
0.50
0.25
0.00

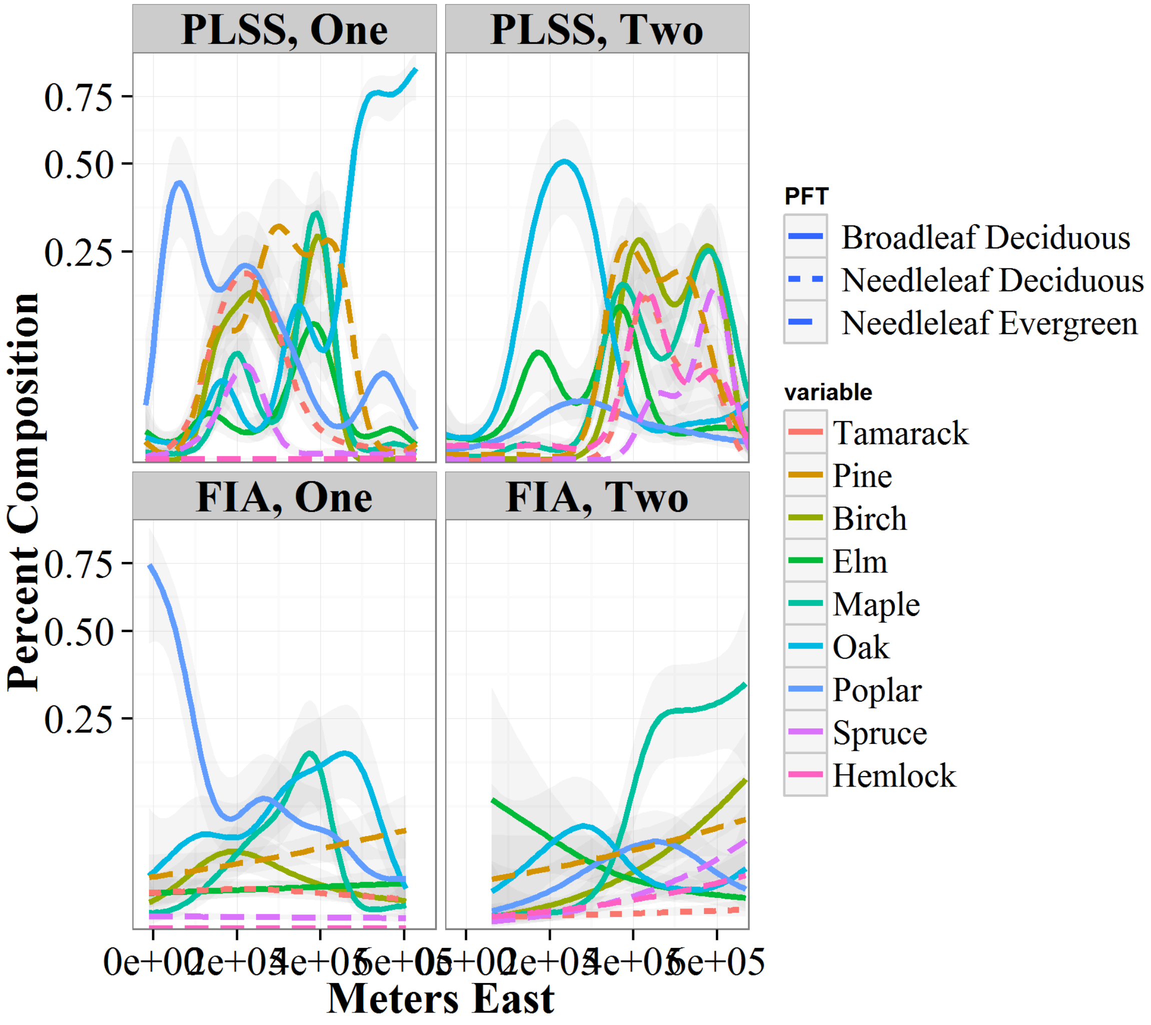

RK1999
Basal Area
values
Sa
Pr
TN
TD
MW


FIA Stem Density ste
600
400
200
200 400 600
PLSS Stem Density stems ha^{-1}
FIA Basal Area m^2
60
40
20
20 40 60
PLSS Basal Area m^2 ha^{-1}
FIA Biomass Mg ha^{-1}
200
150
100
50
50 100 150 200
PLSS Biomass Mg ha^{-1}
higher
FALSE
TRUE


Density
6
4
2
0
0.0
0.2
0.4
0.6
Composition Dissimilarity
class
FIA
FIA-PLSS
PLSS
PLSS-FIA


FIA
PLSS
dist
0.5
0.4
0.3
0.2
0.1
0.0


FIA-PLSS
PLSS-FIA



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.

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.

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.

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.

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.

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 Pres.

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

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.

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.

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, P. D. Miles, and others. 2012. Forest Inventory and Analysis database of the United States of America (FIA). Vegetation databases for the 21st century.–Biodiversity & Ecology 4:255–264.

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.

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. Oâ€™Hara, 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.

Hobbs, R. J., S. Arico, J. Aronson, J. S. Baron, P. Bridgewater, V. A. Cramer, P. R. Epstein, J. J. Ewel, C. A. Klink, A. E. Lugo, and others. 2006. Novel ecosystems: theoretical and management aspects of the new ecological world order. Global ecology and biogeography 15:1–7.

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.

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 âœMixed Pixelâ? 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.

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.

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.

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.

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

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.

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.

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

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.

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.

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.

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.

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.

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.

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.

Thurman, A., C. J. Paciorek, S. J. Goring, and J. W. Williams. in prep. Estimating uncertainty from pre-settlement forest survey data.

Umbanhowar Jr, 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. US Department of the Interior, Bureau of Land Management.

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.

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. Oâ€™Connell, E. B. LaPoint, J. A. Turner, K. L. Waddell, and others. 2010. The Forest Inventory and Analysis database: Database description and users manual version 4.0 for phase 2.