

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

4 Simon J. Goring¹

5 David J. Mladenoff²

6 Charles V. Cogbill³

7 Sydne Record^{3,4}

8 Christopher J. Paciorek⁵

9 Stephen T. Jackson⁶

10 Michael C. Dietze⁷

11 Andria Dawson⁵

12 Jaclyn Hatala Matthes⁸

13 Jason S. McLachlan⁹

14 John W. Williams^{1,10}

15 ¹Department of Geography, University of Wisconsin, Madison, 550 N Park St, Madison WI

16 53706

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

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

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

21 ⁵Department of Statistics, University of California, Berkeley, 367 Evans Hall, Berkeley CA
22 94720

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

26 ⁷Department of Earth and Environment, Boston University, 685 Commonwealth Ave,
27 Boston, MA 02215

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

29 ⁹Department of Biological Sciences, University of Notre Dame, 100 Galvin Life Sciences
30 Center, Notre Dame, IN 46556

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

33

34

35 **Abstract**

36 *EuroAmerican land use and its legacies have transformed forest structure and composition*
37 *across the United States (US). More accurate reconstructions of historical states are critical to*
38 *understanding the processes governing past, current, and future forest dynamics. Gridded*
39 *(8x8km) estimates of pre-settlement (1800s) forests from the upper Midwestern US*
40 *(Minnesota, Wisconsin, and most of Michigan) using 19th Century Public Land Survey (PLS)*
41 *records provide relative composition, biomass, stem density, and basal area for 26 tree*
42 *genera. This mapping is more robust than past efforts, using spatially varying correction*
43 *factors to accommodate sampling design, azimuthal censoring, and biases in tree selection.*

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

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

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

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

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

70 **Introduction:**

71 The composition, demography, and structure of forests in eastern North America have
72 changed continuously over the last millennium, driven by human land use (Foster et al.
73 1998, Ramankutty and Foley 1999, Ellis and Ramankutty 2008, Thompson et al. 2013,
74 Munoz et al. 2014) and climate variability (Umbanhowar et al. 2006, Hotchkiss et al. 2007,
75 Booth et al. 2012, Pederson et al. 2014). While human effects have been a component of
76 these systems for millenia, the EuroAmerican settlement and industrialization period have
77 increased anthropogenic effects by orders of magnitude (Brugam 1978, McAndrews 1988,

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

85 At a regional scale many forests in the upper Midwest (*i.e.*, Minnesota, Wisconsin and
86 Michigan) now have decreased species richness and functional diversity relative to forests
87 of the pre-EuroAmerican settlement (hereafter pre-settlement) period (Schulte et al. 2007,
88 Hanberry et al. 2012a, Li and Waller 2014) due to near complete logging. For example,
89 forests in Wisconsin are in a state of regrowth, with an unfilled carbon sequestration
90 potential of 69 TgC (Rhemtulla et al. 2009a) as a consequence of these extensive land cover
91 conversions and subsequent partial recovery following abandonment of farm lands in the
92 1930s.

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

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

109 While there remains debate over the utility of the concept of novel ecosystems (Hobbs et al.
110 2014a, Murcia et al. 2014), the fact remains that there are now forest and vegetation
111 communities on the landscape without past analogues. The long term management of the
112 systems and their associated services requires a broad understanding of the extent to
113 which landscapes have been modified, and the extent to which land use change has
114 potentially masked underlying processes. It also requires a better understanding of the
115 spatial (and temporal) scales at which novel ecosystems operate. While much restoration
116 effort has focused on ecosystems at local scales there is an increasing need to focus on
117 management and restoration at landscape scales (Menz et al. 2013). Thus a better
118 understanding of the landscape-scale processes driving novelty, the spatial structure of
119 novel ecosystems and their ecological correlates, is increasingly important. It can help
120 prioritize intervention strategies at local scales (Hobbs et al. 2014b), and give us a better
121 understanding of the role of patches in restoring hybrid or novel landscapes. In particular,
122 how important is the species pool to the development of novel landscapes? Are novel

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

127 The upper Midwestern United States represents a unique ecological setting, with multiple
128 major ecotones, including the prairie-forest boundary, historic savanna, and the Tension
129 Zone between southern deciduous forests and northern evergreen forests. The extent to
130 which these ecotones have shifted, and their extent both prior to and following
131 EuroAmerican settlement is of critical importance to biogeochemical and biogeophysical
132 vegetation-atmosphere feedbacks (Matthes et al. in review), carbon sequestration
133 (Rhemtulla et al. 2009a), and regional management and conservation policy (Radeloff et al.
134 2000, Fritschle 2008, Knoot et al. 2010, Gimmi and Radeloff 2013).

135 Land use change at the local and state-level has affected both the structure and
136 composition of forests in the Midwestern United States (*e.g.* Schulte et al. 2007, Hanberry et
137 al. 2012a). Homogenization and shifts in overall forest composition are evident, but the
138 spatial extent and structure of this effect is less well understood. Studies in Wisconsin have
139 shown differential patterns of change in the mixedwood and evergreen dominated north
140 versus the southern driftless and hardwood south. Does this pattern of differential change
141 extend to Minnesota and Michigan? To what extent are land-use effects common across the
142 region, and where are responses ecozone-specific? Has homogenization (*e.g.*, Schulte et al.
143 2007) resulted in novel forest assemblages relative to pre-settlement baselines across the
144 region, and the loss of pre-settlement forest types? Are the spatial distributions of these

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

149 Modern forest structure and composition data (*e.g.*, from the United States Department of
150 Agriculture Forest Service's Forest Inventory and Analysis National Program, FIA; Gray et
151 al. 2012) play a ubiquitous role in forest management, conservation, carbon accounting,
152 and basic research on forest ecosystems and community dynamics. These recent surveys
153 (the earliest FIA surveys began in the 1930s) can be extended with longer-term historical
154 data to understand how forest composition has changed since EuroAmerican settlement.
155 The Public Land Survey was carried out ahead of mass EuroAmerican settlement west and
156 south of Ohio to provide for delineation and sale of the public domain beyond the original
157 East Coast states (Stewart 1935, White 1983). Because surveyors used trees to locate
158 survey points, recording the identity, distance, and directory of two to four trees next to
159 each survey marker, we can make broad-scale inferences about forest composition and
160 structure in the United States prior to large-scale EuroAmerican settlement (Almendinger
161 1996, Liu et al. 2011, Williams and Baker 2011, Tomscha and Gergel 2014). In general, FIA
162 datasets are systematically organized and widely available to the forest ecology and
163 modeling community, whereas most PLS data compilations are of local or, at most, state-
164 level extent. This absence of widely available data on settlement-era forest composition
165 and structure limits our ability to understand and model the current and future processes
166 governing forest dynamics at broader, regional scales. For example, distributional models
167 of tree species often rely upon FIA or other contemporary observational data to build

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

170 Here we use survey data from the original Public Lands Surveys (PLS) in the upper
171 Midwest to derive estimates of pre-settlement (*ca.* mid-late 1800s) forest composition,
172 basal area, stem density, and biomass. This work builds upon prior digitization and
173 classification of PLS data for Wisconsin (Manies and Mladenoff 2000, Schulte et al. 2002)
174 and for parts of Minnesota (Friedman and Reich 2005, Hanberry et al. 2012a) and Michigan
175 Michigan (USFS-NCRS <http://www.ncrs.fs.fed.us/gla/>). Most prior PLS-based
176 reconstructions are for individual states or smaller extents (among others: Duren et al.
177 (2012); Hanberry et al. (2012a); Rhemtulla et al. (2009a); Friedman and Reich (2005))
178 often aggregated at the scale of regional forest zones (Schulte et al. 2007, Hanberry et al.
179 2012a), although aggregation may also occur at the section (Rhemtulla et al. 2009a) or
180 township scale (Kronenfeld et al. 2010). Our work develops new approaches to address
181 major challenges to PLS data, including lack of standardization in tree species names,
182 azimuthal censoring by surveyors, variations in sampling design over time, and differential
183 biases in tree selection among different kinds of survey points within the survey design at
184 any point in time. The correction factors developed here are spatially varying, allowing us
185 to accommodate temporal and spatial variations in surveyor methods.

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

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

197 **Methods:**

198 **Public Lands Survey Data: Assembly, and Standardization**

199 The PLS was designed to facilitate the division and sale of federal lands from Ohio
200 westward and south. The survey created a 1 mile² (2.56 km²) grid (sections) on the
201 landscape. At each section corner, a stake was placed as the official location marker. To
202 mark these survey points, PLS surveyors recorded tree stem diameters, measured
203 distances and azimuths of the two to four trees 'closest' to the survey point and identified
204 tree taxa using common (and often regionally idiosyncratic) names. PLS data thus
205 represent measurements by hundreds of surveyors from 1832 until 1907, with changing
206 sets of instructions over time (Stewart, 1979).

207 The PLS was undertaken to survey land prior to assigning ownership (Stewart 1935, White
208 1983), replacing earlier town proprietor surveys (TPS) used for the northeastern states
209 (Cogbill et al. 2002, Thompson et al. 2013). The TPS provided estimates of relative forest
210 composition at the township level, but no structural attributes. The PLS produced spatially

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

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

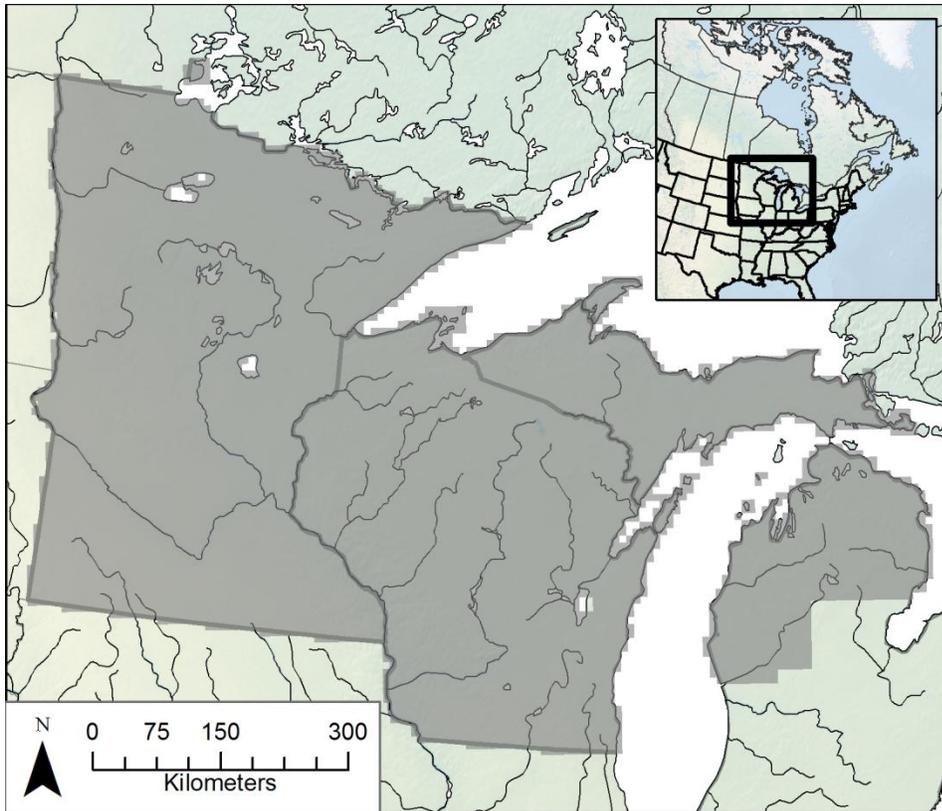
227 Kronenfeld and Wang (2007), working with historical land cover datasets in western New
228 York indicate that direct estimates of density using plotless estimators may be off by nearly
229 37% due to azimuthal censoring (*i.e.*, the tendency of surveyors to avoid trees close to
230 cardinal directions), while species composition estimates may be adjusted by between -4 to
231 +6%, varying by taxon, although Kronenfeld (2014) shows adjustments of less than 1%.

232 These biases can be minimized by appropriate analytical decisions; many efforts over the

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

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

244



245

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

250

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

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

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

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

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

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

303 **Data Aggregation**

304 We binned the point data using an 64km² grid (Albers Gt. Lakes St Lawrence projection;
305 Supplement 1: *base_calculations.R*) to create a dataset that has sufficient numerical power
306 for spatial statistical modeling and sufficient resolution for regional scale analysis (???).
307 This resolution is finer than the 100km² gridded scale used in Freidman and Reich (2005),
308 but coarser than township grids used in other studies (Rhemtulla et al. 2009a, Kronenfeld
309 2014) to provide a scale comparable to aggregated FIA data at a broader scale. Forest
310 compositional data is based on the number of individuals of each genus or plant functional
311 type (PFT) present at all points within a cell. Stem density, basal area and biomass are
312 averaged across all trees at all points within the cell.

313 **Stem Density**

314 Estimating stem density from PLS data is based on a plotless density estimator using the
315 measured distances from each survey point to the nearest trees (Morisita 1957, Persson
316 1971). This Morisita density estimator is then modified to minimize error due to different
317 sampling geometries and several known surveyor biases (Manies et al. 2001, Kronenfeld
318 and Wang 2007, Bouldin 2008, Hanberry et al. 2011, 2012a, 2012b, Liu et al. 2011,
319 Williams and Baker 2011, Cogbill et al. in prep). Survey sampling instructions changed
320 throughout the implementation of the PLS in this region and differed between section and
321 quarter section points and between internal and external points within a township (White
322 1983, Liu et al. 2011). Our approach allows for spatial variation in surveyor methods by
323 applying various spatially different correction factors based not only on the empirical

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

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

332 We estimate the basic stem density (stems m⁻²) using the point-to-tree distances for the
333 closest trees to each point within a defined number of sectors around the point (Morisita
334 (1957) eqn 31.):

$$335 \quad \lambda \hat{m}_2 = \frac{k-1}{\pi \times n} \times \sum_{i=1}^N \frac{k}{\sum_{j=1}^k (r_{ij})^2} \quad (1)$$

336 where λ is density ; k is the number of sectors within which trees are sampled, N is the
337 number of points over which estimates are aggregated, r is the distance of point-to-tree (as
338 m). This estimate can be modified by a refinement of the Cottam quadrant factors (Morisita
339 1954, Cottam and Curtis 1956) which recognizes that different sampling designs, and the
340 order of the distances in different quadrants (or sectors) carry specific weights. This
341 correction, herein called κ , accounts for different sampling designs. When either four
342 quadrants or trees are sampled (point quarter design), or when two trees in opposite
343 semicircles (point halves design) are sampled, the equation is accurate and $\kappa = 1$; when the
344 two trees are in the nearest of two quadrants (two nearest quadrants design), $\kappa = 0.857$;

345 and when two trees are in quadrants on the same side of the direction of travel (one-sided
346 or interior half design), $\kappa = 2$. This parameter, in Cottam's notation (Cottam and Curtis
347 1956), is a divisor of the denominator above, or here, the mathematically equivalent
348 multiplier in the numerator of the reciprocal of the squared distances.

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

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

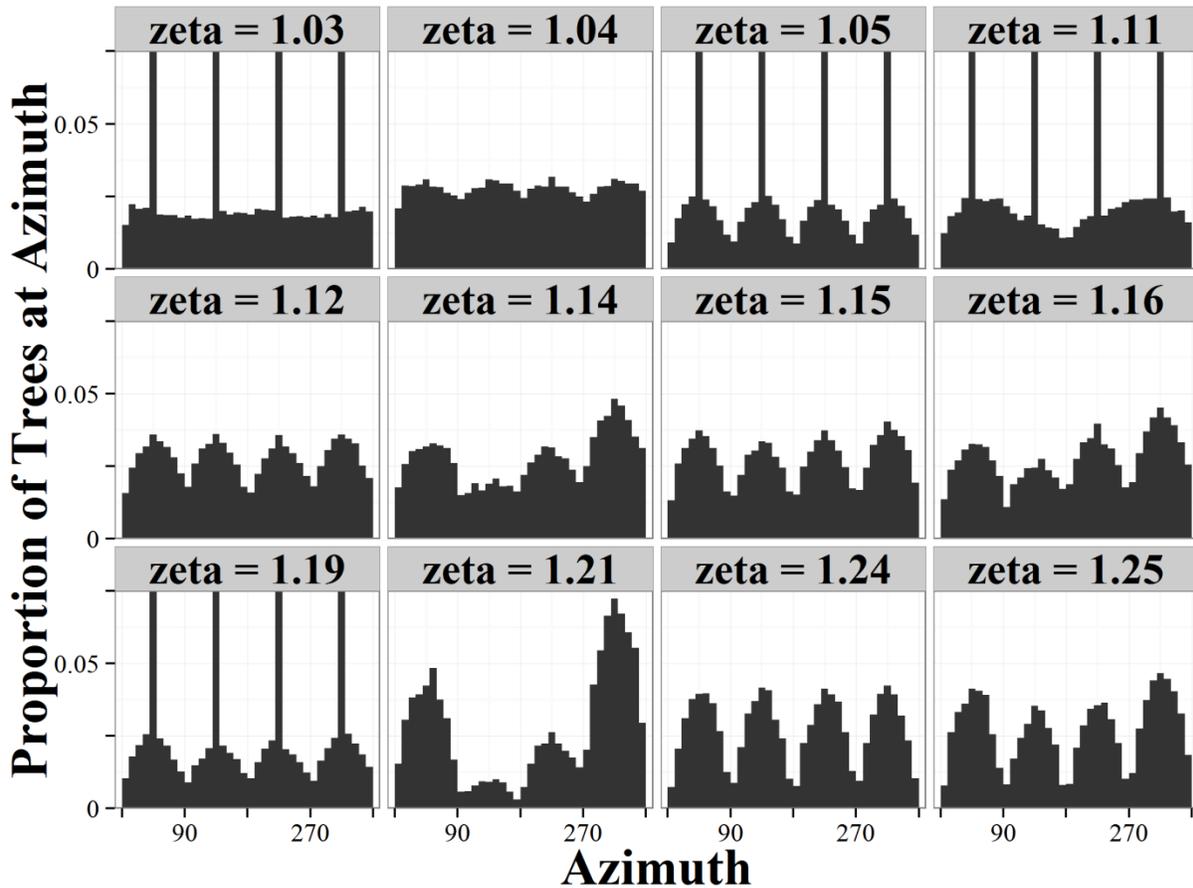
351 Then the point values for any sampling design can be Cottam corrected ($\kappa \times \lambda_M$). For
352 example, the basic Morisita equation for two sectors assumes trees are located in opposite
353 halves, so if the actual design is the nearest tree in the two nearest quadrants, the density
354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

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

361 Sector bias (θ). Although the density model for two tree points assumes that the trees are
362 on opposite sides of a sample line (point halves), often the actual sample is more restricted
363 ($< 180^\circ$) within the sector or is a less restricted ($> 180^\circ$) angle beyond the sector. This
364 deviation from the equation's assumption of equal distribution of angles across the 180°

365 sector is quantified using the empirical angle between the bearings of the two trees (pair
366 angle). In the pair angle frequency plot (Figure 2), the observed proportion of trees (p)
367 within any restricted sector divided by the proportion of that angle within the circle (α is
368 an estimate of the bias imposed by the actual sampling (inspired by Kronenfeld & Wang
369 (2007)). This factor ($\theta = p/\alpha$) indicates bias associated with differences in geometry of two
370 tree samples. This parameter (θ) varies from 0.71 to 1.27, indicating sampling from
371 effectively 253° to 141° sectors.

372



373

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

378 Azimuthal censoring (ζ). In addition to sector bias, surveyors did not always sample trees

379 near the cardinal directions (Kronenfeld and Wang 2007, Bouldin 2008, Hanberry et al.

380 2012b). This azimuthal censoring is commonly found along the line of travel on section

381 lines and sometimes on the perpendicular quarter-section lines. Trees near the cardinal

382 directions were passed over, and a replacement was found within a more restricted

383 angular region. The correction for this bias is calculated following Kronenfeld and Wang

384 (2007) in a manner similar to the sector bias. The factor ζ is the ratio of the proportion of

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

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

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

$$402 \lambda_{\text{Mcorrected}} = \kappa \times \theta \times \zeta \times \phi \times \lambda_{\text{M}} \quad (3)$$

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

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

410 **Table 1.** Correction values based on plot level survey design using state, year, and location
 411 within township as a basis for assignment. Years reported represent the upper bound for each
 412 set of survey years. Internal points are points within the township, external points are on the
 413 township boundary; no sampling occurred outside of a township boundary so plots were
 414 limited to half of the space for internal points. Townships are divided into Section and Quarter
 415 Sections, at most section points and some quarter section points, *r* instructions indicated four
 416 trees were to be sampled, these were '2nQ' plots, whereas others surveyed only two points in
 417 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

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

418

419 Basal Area and Biomass Estimates

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

425 Biomass equations share the basic form:

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

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

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

431 **Table 2.** Biomass parameters used for the calculation of biomass in the pre-settlement
432 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

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

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

440 **FIA Stem Density, Basal Area and Biomass**

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

444 in the United States classified as forested. Each FIA plot consists of four 7.2m fixed-radius
445 subplots in which measurements are made of all trees >12.7cm dbh (Woudenberg et al.
446 2010). We used data from the most recent full plot inventory (2007-2011). The FIA plot
447 inventory provides a median of 3 FIA plots per cell using the 64km² grid.

448 We calculated mean basal area (m² ha⁻¹), stem density (stems ha⁻¹), mean diameter at
449 breast height (cm), and mean biomass (Mg ha⁻¹) for all live trees with dbh greater than
450 20.32cm (8in). Biomass calculations used the same set of allometric regression equations
451 as for the PLS data (Jenkins et al. 2004).

452 **Gridding and Analysing PLS and FIA Data**

453 Spatial maps of stem density, basal area and biomass were generated by averaging all PLS
454 point or FIA plot estimates within a 64km² raster cell. Differences in sampling design
455 between PLS and FIA data combined with spatially structured forest heterogeneity will
456 affect the partitioning of within-cell versus between-cell variance, but not the expected
457 estimates. Most 64km² cells have one or a few intensively sampled FIA plots. Therefore at
458 this scale of aggregation, the low density of FIA plots in heterogeneous forests could result
459 in high within-cell variance and high between-cell variability. For the PLS plotless (point
460 based) estimates, stem density estimates are sensitive to trees close to the plot center.
461 Point-level estimates with very high stem densities can skew the rasterized values, and it is
462 difficult to distinguish artifacts from locations truly characterized by high densities. To
463 accommodate points with exceptionally high densities we carry all values through the
464 analysis, but exclude the top 2.5 percentile when reporting means and standard deviations
465 in our analysis. PLS-based estimates are highly variable from point to point due to the small

466 sample size, but have low variance among 64 km² raster cells due to the uniform sampling
467 pattern of the data. Thus within-cell variance is expected to be high for the PLS point data,
468 but spatial patterns are expected to be robust at the cell level. The base raster and all
469 rasterized data are available as Supplement 3.

470 Standard statistical analysis of the gridded data, including correlations and regression, was
471 carried out in R (Team 2014), and is documented in supplementary material that includes a
472 subset of the raw data to allow reproducibility. Analysis and presentation uses elements
473 from the following R packages: *cluster* (Maechler et al. 2014), *ggplot2* (Wickham 2009a,
474 2009b), *gridExtra* (Augue 2012), *igraph* (Csardi and Nepusz 2006), *mgcv* (Wood 2011),
475 *plyr* (Wickham 2011), *raster* (Hijmans 2014), *reshape2* (Wickham 2007), *rgdal* (Bivand
476 et al. 2014), *rgeos* (Bivand and Rundel 2014), *sp* (Pebesma and Bivand 2005, Bivand et al.
477 2013), and *spdep* (Bivand 2014).

478 We identify analogs and examine differences in composition between and within PLS and
479 FIA datasets using Bray-Curtis dissimilarity (*vegdist* in *vegan*; Oksanen et al. 2014) for
480 proportional composition within raster cells using basal area measurements. For the
481 analog analysis we are interested only in the minimum compositional distance between a
482 focal cell and its nearest compositional (not spatial) neighbor. The distribution of
483 compositional dissimilarities within datasets indicates forest heterogeneity within each
484 time period, while the search for closest analogs between datasets indicates whether
485 contemporary forests lack analogs in pre-settlement forests ('novel forests'), or vice versa
486 ('lost forests'). For the analog analyses, we compute Bray-Curtis distance between each
487 64km² cell in either the FIA or the PLS periods to all other cells within the other dataset

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

492 The differences in sampling design and scale between the PLS and FIA datasets, described
493 above, potentially affect between-era assessments of compositional similarity (*e.g.*,
494 Kronenfeld et al. 2010). The effects of differences in scale should be strongest in regions
495 where there are few FIA plots per 64 km² cell, or where within-cell heterogeneity is high.
496 For the analog analyses, this effect should increase the compositional differences between
497 the FIA and PLS datasets. We test for the importance of this effect on our analog analyses
498 via a sensitivity analysis in which we test whether dissimilarities between FIA and PLS grid
499 cells are affected by the number of PLS plots per cell. We find a small effect, suggesting that
500 our analyses are mainly sensitive to the compositional and structural processes operating
501 on large spatial scales.

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

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

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

521 **Results:**

522 **Data Standardization**

523 The original PLS dataset contains 490,818 corner points (excluding line and meander
524 points), with 166,607 points from Wisconsin, 231,083 points from Minnesota and 93,095
525 points from Michigan. Standardizing data and accounting for potential outliers, described
526 above, removed approximately 1.5% points from the dataset, yielding a final total of
527 366,993 points with estimates used in our analysis.

528 Rasterizing the PLS dataset to the Albers 64km² grid produces 7,939 raster cells with data.
529 Each cell contains between 1 and 94 corner points, with a mean of 61.8 ($\sigma = 15$) and a
530 median of 67 corners (Supplement 3). Cells with a low number of points were mainly near

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

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

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

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

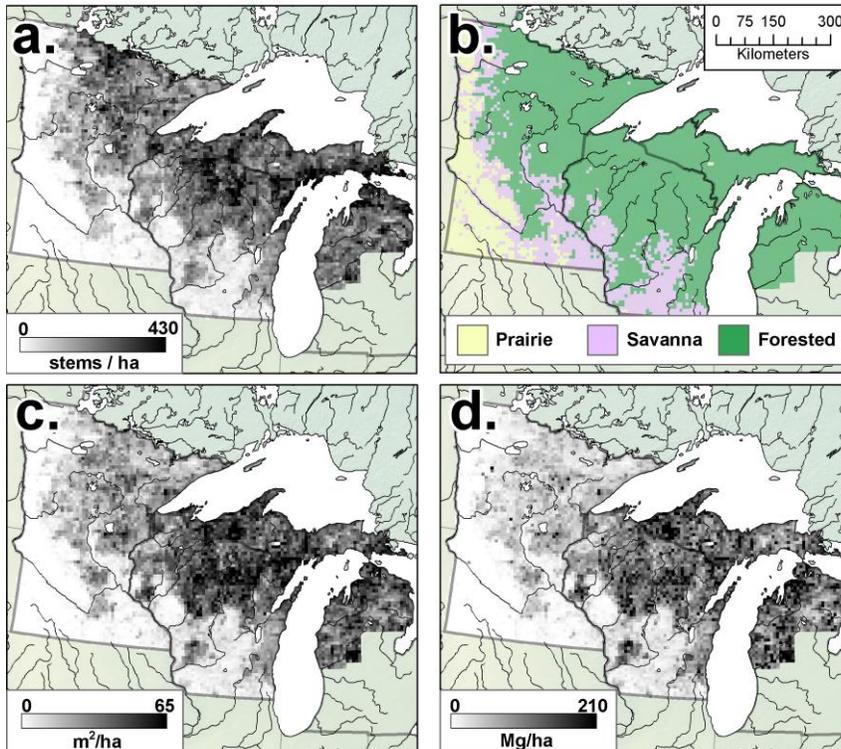
551 These patterns are repeated throughout the data. For spruce this within-genera confusion
552 is even greater, with 3, 50141, 43, 1 assignments to genera-level classes and only 20 to
553 either black or white spruce.

554

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

556 **Stem Density, Basal Area and Biomass**

557 The mean stem density for the region (Figure 3a) is 153 stems ha⁻¹. Stem density exclusive
558 of prairie is 172 stems ha⁻¹ and is 216 stems ha⁻¹ when both prairie and savanna are
559 excluded. The 95th percentile range is 0 - 423 stems ha⁻¹, and within-cell standard
560 deviations between 0 and 441 stems ha⁻¹. Basal area in the domain (Figure 3c) has a 95th
561 percentile range between 0 and 63.5 m² ha⁻¹, a mean of 22.2 m² ha⁻¹, within-cell standard
562 deviations range from 0 to 76.7 m² ha⁻¹. Biomass ranges from 0 to 209 Mg ha⁻¹ (Figure 3d),
563 with cell level standard deviations between 0 and 569 Mg ha⁻¹. High within-cell standard
564 deviations relative to mean values within cells for density, basal area and biomass indicate
565 high levels of heterogeneity within cells, as expected for the PLS data, given its dispersed
566 sampling design.



567

568 **Figure 3.** Total stem density (a) in the Upper Midwest, along with forest type classification
569 (b) based on PLS data and the stem density thresholds defined by Anderson and Anderson
570 (1975); Table 3). Fine lines represent major rivers. To a first order, basal area (c) and biomass
571 (d) show similar patterns to stem density (but see Figure 5).

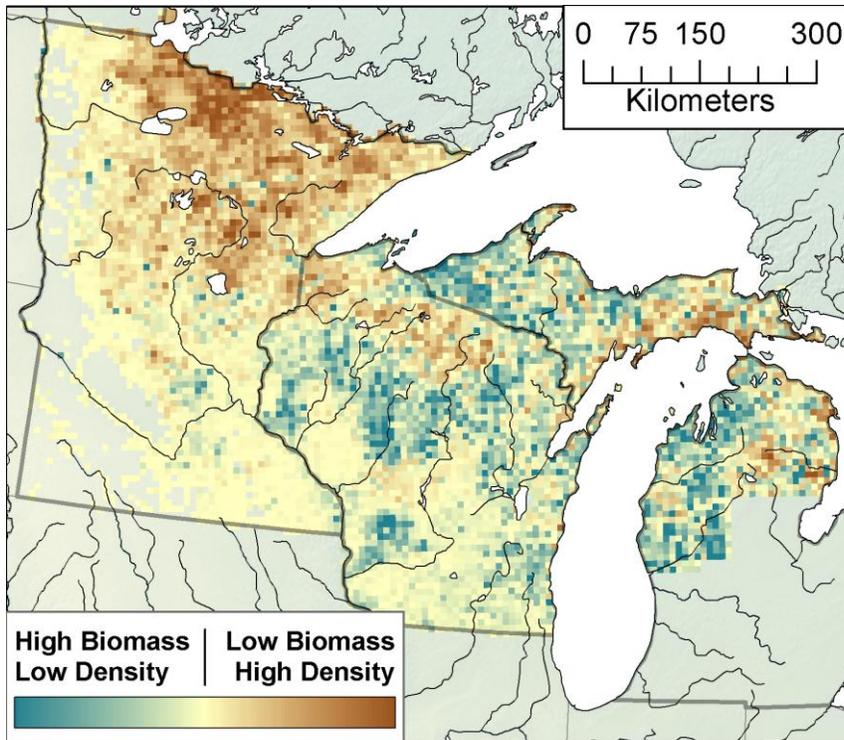
572 In the PLS data, stem density is lowest in the western and southwestern portions of the
573 region, regions defined as prairie and savanna (Figure 3b, Table 3). When the Anderson
574 and Anderson (1975) stem density thresholds (<47 stems ha⁻¹ for Savanna, Table 3) are
575 used, the extent of area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

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

582 *(PFTs) for grasslands (CG, grassland; Non-Tree samples in the PLS), broad leafed deciduous*
 583 *taxa (BDT) and needleleaded evergreen taxa (NET) are used, but leaf area index used in*
 584 *Haxeltine and Prentice (1996) is replaced by stem density classes from Anderson and*
 585 *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

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

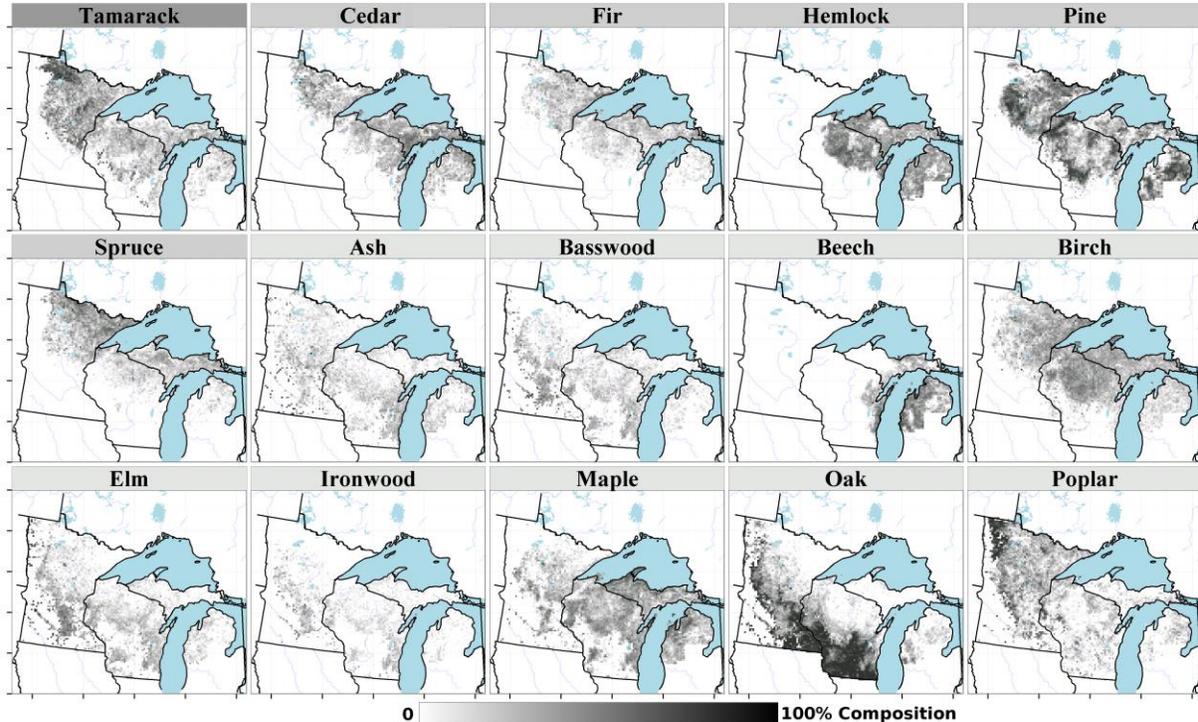


597

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

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

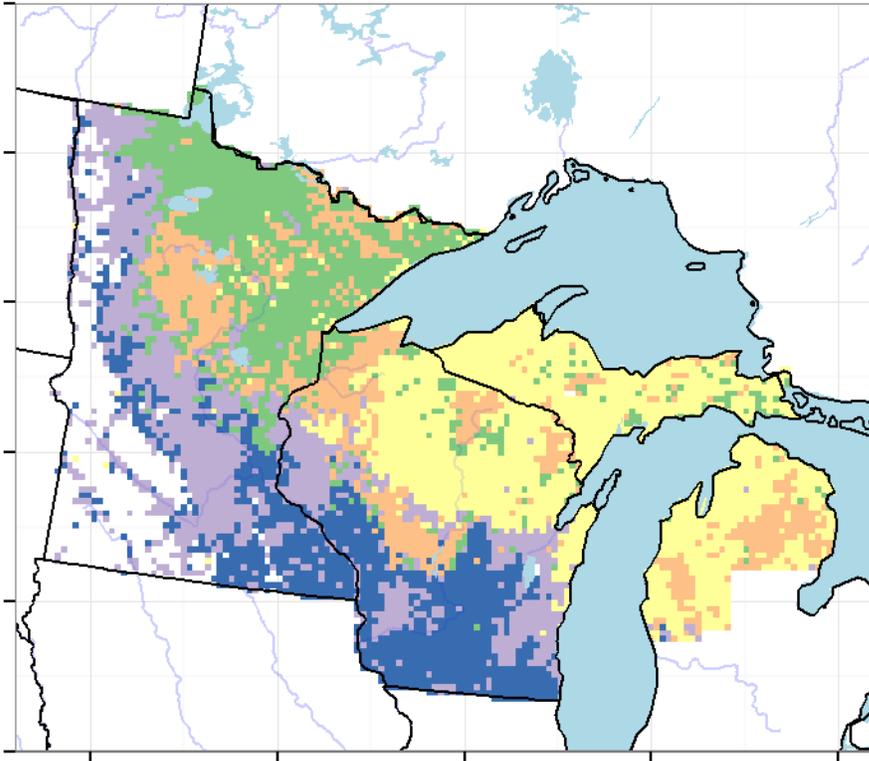
612 composition in forested cells include maple (10%), birch (10%), tamarack (9%) and
613 hemlock (8%).



614 **Figure 5.** Forest composition (%) for the 15 most abundant tree taxa. The scale is drawn
615 using a square-root transform to emphasize low abundances. Shading of the bar above
616 individual taxon maps indicates plant functional type assignments (dark gray: needleleafed
617 deciduous; light gray: needleleafed evergreen; white: broadleafed deciduous).
618

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

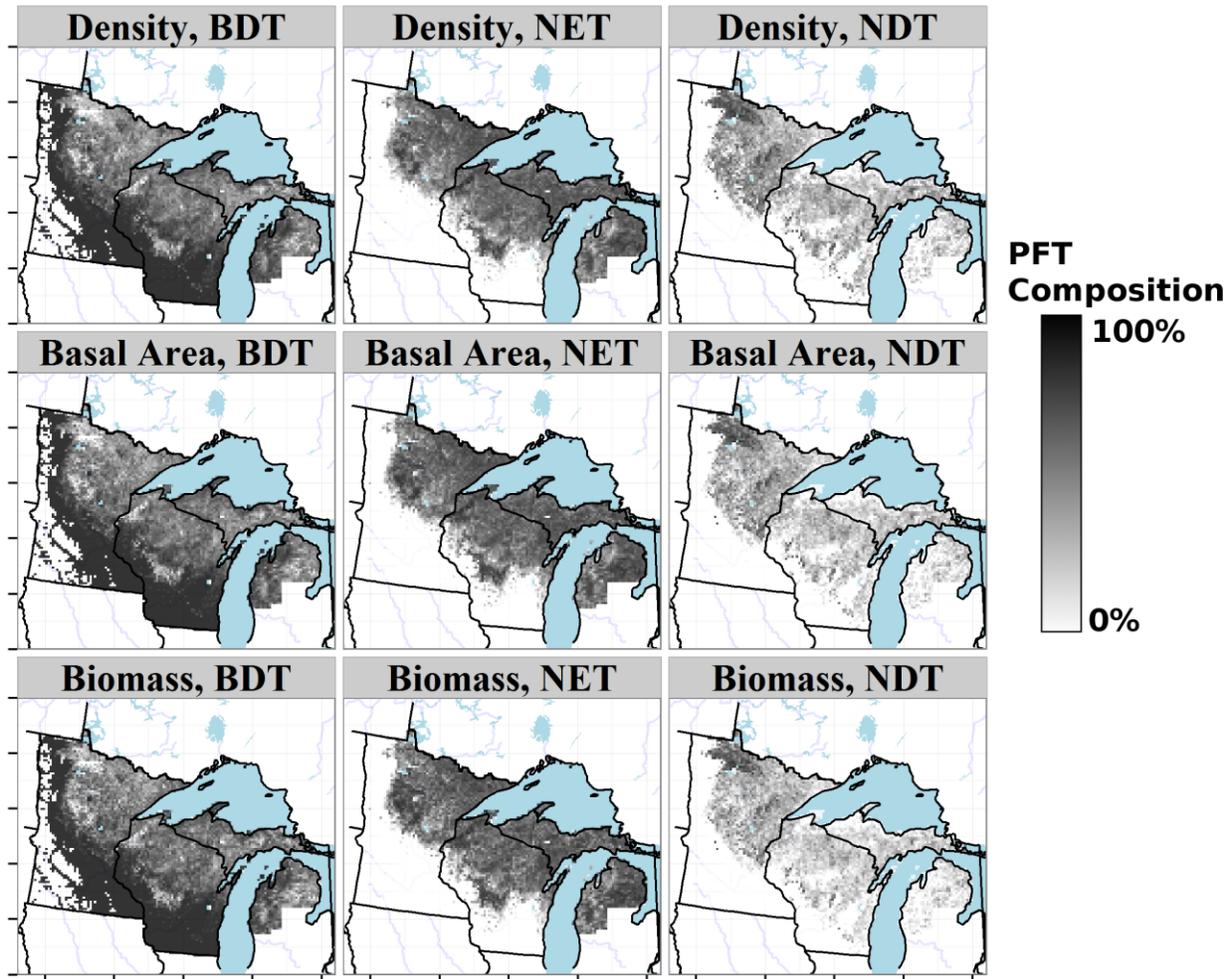
627 Spruce in the PLS represents two species (*Picea glauca*, *Picea mariana*) with overlapping
628 distributions, but complex site preferences that vary in space. *P. glauca* is generally
629 associated with dry upland to wet-mesic sites, while *P. mariana* is associated with hydric
630 sites, but *P. mariana* also frequently occupies upland sites in northern Minnesota. Both
631 cedar (*Thuja occidentalis*) and fir (*Abies balsamea*) are mono-specific genera in this region.
632 Northern hardwoods, such as yellow birch and sugar maple, and beech, are much less
633 common in the lower peninsula of Michigan, and southern Wisconsin, except along Lake
634 Michigan. Birch has extensive cover in the north, likely reflecting high pre-settlement
635 proportions of yellow birch (*Betula alleghaniensis*) on mesic soils, and paper birch on sandy
636 fire-prone soils and in northern Minnesota (birch proportions reach upwards of 34.1% in
637 northeastern Minnesota). Hardwoods in the southwest, such as oak, elm, ironwood and
638 basswood, are most typically mono-specific groupings, with the exception of oak, which
639 comprises 7 species (see Supplement 2). Hardwoods in the southwest are located primarily
640 along the savanna and southern forest margins, or in the southern temperate deciduous
641 forests. Finally, maple and poplar (aspen) have a broad regional distribution, occupying
642 nearly the entire wooded domain. Poplar comprises four species in the region, while maple
643 comprises five species (Supplement 2). Both hardwood classes, those limited to the
644 southern portions of the region, and those with distributions across the domain,
645 correspond to well-defined vegetation patterns for the region (Curtis 1959).



646

647 **Figure 6.** *The five dominant forest types in the Upper Midwest as defined by k-medoid*
648 *clustering. Forest types (from largest to smallest) include Hemlock/Cedar/Birch/Maple*
649 *(yellow), Oak/Poplar/Basswood/Maple (light purple), Tamarack/Pine/Spruce/Poplar (light*
650 *green), Oak Savanna (dark purple) and Pine (orange). These forest types represent meso-*
651 *scale (64km²) forest associations, rather than local-scale associations.*

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

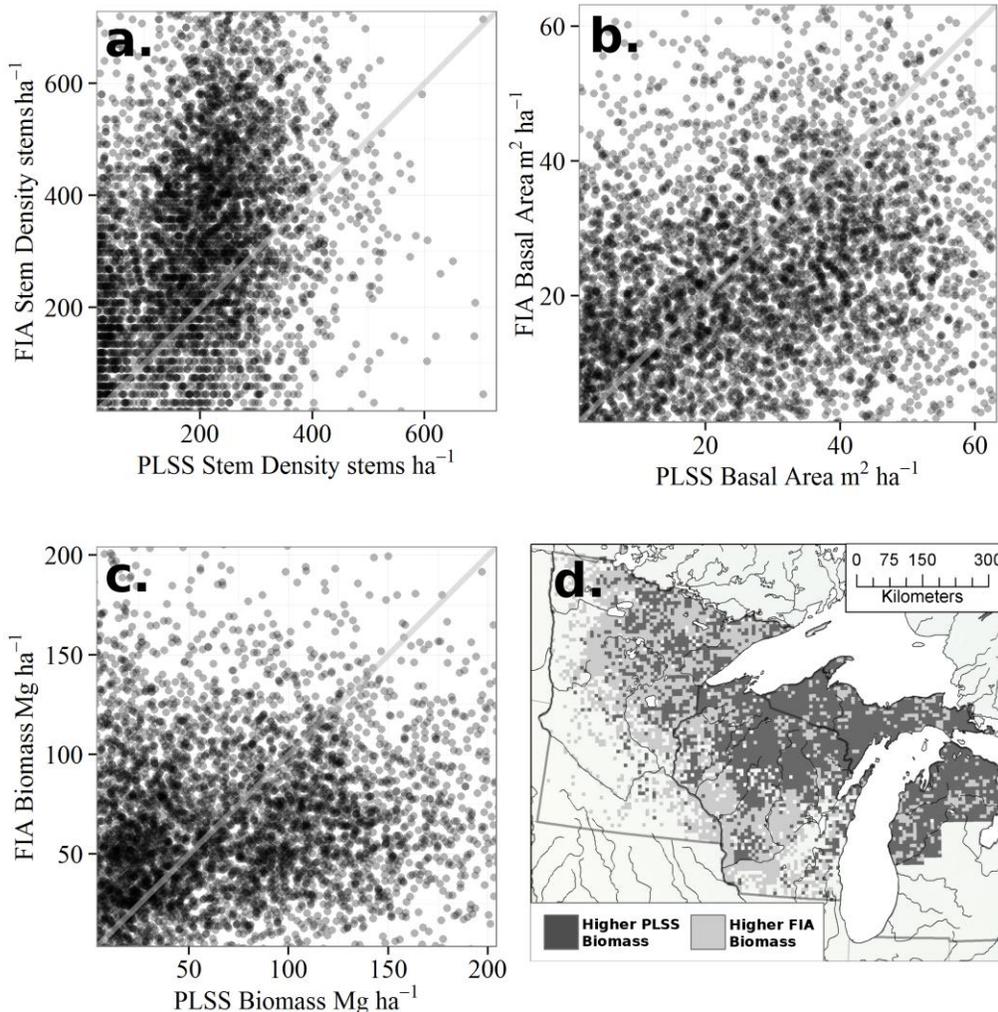


662

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

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

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



675
676 **Figure 8.** *The relationship between average stem density, total basal area and biomass values*
677 *in the PLS and FIA datasets. Stem density and total basal area are higher in the FIA than in*
678 *the PLS, however average cell biomass is higher in the PLS.*

679 Structural Changes Between PLS and FIA Forests

680 Modern forests (FIA) have higher stem densities (146 stems ha⁻¹, $t_{1,5177} = 51.8$, $p < 0.01$)
681 and basal areas (-4.5 m² ha⁻¹, $t_{1,5177} = -16.4$, $p < 0.01$) than PLS forests, but overall, lower
682 biomass (-8.72 Mg ha⁻¹, $t_{1,5177} = -6.55$, $p < 0.01$) than historical forests (Figure 8). We use

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

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

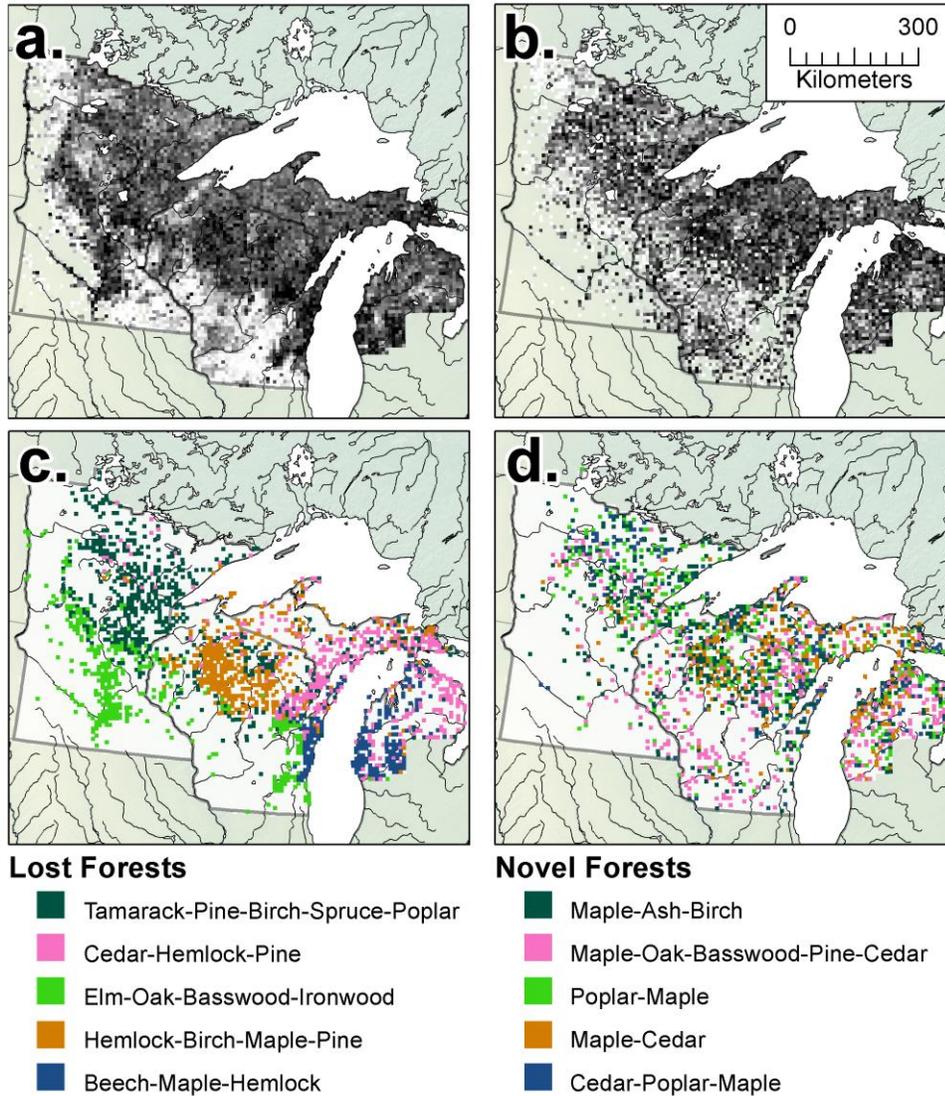
706 diameter, and, subsequently biomass, is occurring in southeastern mixedwood/hardwood
707 forest, while diameter and biomass increases are occurring in the northwest.

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

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

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

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



735

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

741 We define no-analog communities as those whose nearest neighbour is beyond the 95%ile

742 for dissimilarities within a particular dataset. In the PLS dataset, forests that have no

743 modern analogs are defined as "lost forests", while forest types in the FIA with no past

744 analogs are defined as "novel forests". More than 25% of PLS sites have no analog in the

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

749 Lost forests are drawn from across the domain, and show strong ecological and spatial
750 coherence (Figure 9c). Forest classes generally fall into five classes: Tamarack-Pine-Birch-
751 Spruce-Poplar accounts for 28.8% of all lost forests and 7.97% of the total region. This
752 forest type is largely found in north eastern Minnesota, extending southward to central
753 Minnesota, into Wisconsin and along the Upper Peninsula of Michigan, as well as in
754 scattered locations on the Lower Peninsula of Michigan (Figure 9c). This forest likely
755 represents a mesic to hydric forest assemblage, particularly further eastward. Modern
756 forests spatially overlapping this lost type are largely composed of poplar ($\bar{x}_{FIA} = 12\%$) and
757 oak ($\bar{x}_{FIA} = 12\%$). Tamarack in these forests has declined significantly, from 23% to only
758 5% in the FIA, while Poplar has increased from 10% to 22%, resulting in forests that look
759 less mesic and more like early seral forests.

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

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

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

781 Lastly, Beech-Maple-Hemlock accounts for 12.6% of all lost forests and 3.49% of the total
782 region. This forest type is found exclusively on the central, western shore of Lake Michigan
783 and in the Lower Peninsula, in part due to the limited geographic range of Beech in the PLS
784 dataset (Figure 5). Beech is almost entirely excluded from the modern forests in this
785 region, declining from $\bar{x}_{PLS} = 37\%$ to $\bar{x}_{FIA} = 4\%$. Pine in the region increases from 9% to
786 16%, while maple, the dominant taxa in the modern forests, increases from 16 - 25%.

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

790 The distribution of novel ecosystems (Figure 9d) is spatially diffuse relative to the lost
791 forest of the PLS and the forest types tend to have fewer co-dominant taxa. FIA novel forest
792 types also have a more uneven distribution in proportion than the PLS lost forests. Overall,
793 novel forests are associated with higher proportions of maple ($r = 0.02$), ash ($r = 0.03$) and
794 basswood ($r = -0.04$), although basswood is dominant in only one forest type (Poplar-
795 Cedar/juniper-Maple). Novel forests are associated with lower proportions of oak ($r = -$
796 0.28), and pine ($r = -0.11$). This analysis suggests that the loss of particular forest types
797 associated with post-settlement land use was concentrated in mesic deciduous forests and
798 the ecotonal transition between southern and northern hardwood forests, while the gains
799 in novelty were more dispersed, resulting from an overall decline in seral age.

800 By far the largest novel forest type is Maple, which accounts for 37.2% of all novel forests
801 and 2.68% of the total region. As with all novel forest types, this forest type is broadly
802 distributed across the region. This forest type is associated with co-dominant maple ($\bar{x}_{\text{FIA}} =$
803 23%) and ash ($\bar{x}_{\text{FIA}} = 22\%$). Hemlock has declined significantly across this forest type, from
804 $\bar{x}_{\text{PLS}} = 24\%$ to $\bar{x}_{\text{FIA}} = 4\%$.

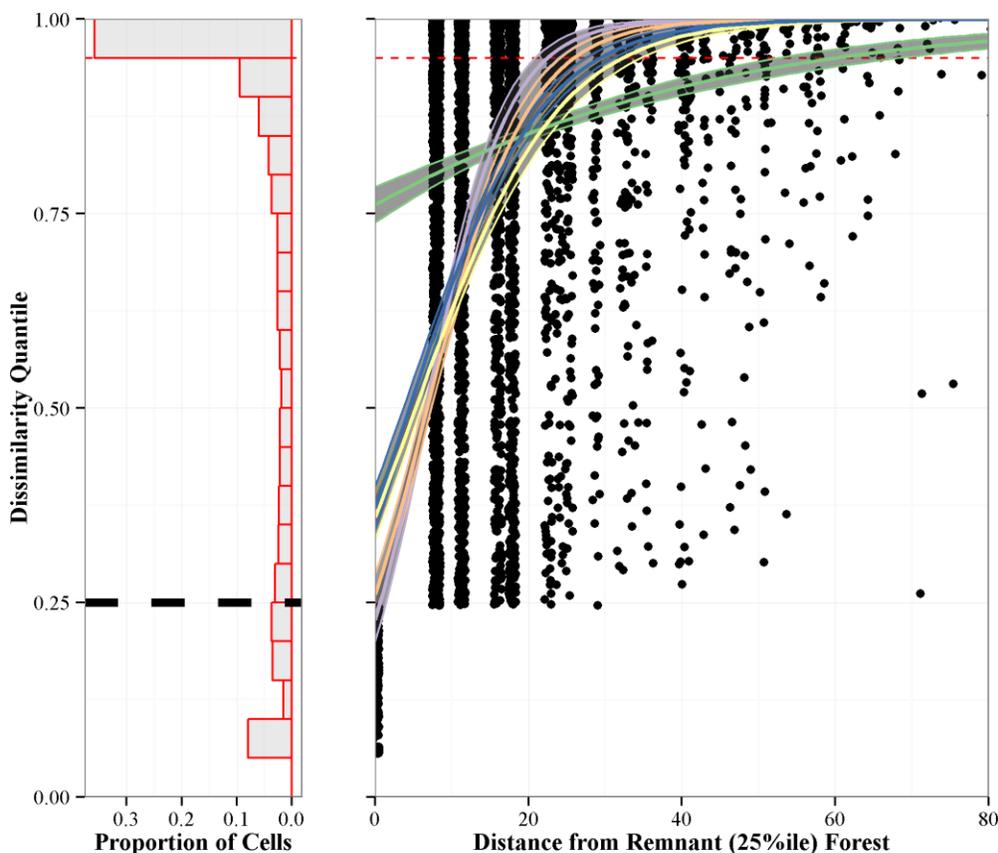
805 Poplar-Cedar/juniper-Maple, accounts for 28.8% of all novel forests and 2.08% of the total
806 region. The broad distributiof these novel forests makes assigning a past forest type more
807 difficult than for the PLS lost forests, the distribution replaces two classes of past forest,

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

810 Pine-Cedar/juniper-Poplar-Maple forest accounts for 16.3% of all novel forests and 1.17%
811 of the total region. This forest type is again broadly distributed, and is widely distributed
812 across the region, representing a homogenous, early seral forest type, likely associated
813 with more mesic sites. Oak forest accounts for 13.3% of all novel forests and 0.96% of the
814 total region. This grouping again shows a pattern of broad distribution across the region,
815 associated with cedar/juniper percentages near 40%, with smaller components of poplar
816 (14%) and maple (13%).

817 **Spatial Correlates of Novelty**

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



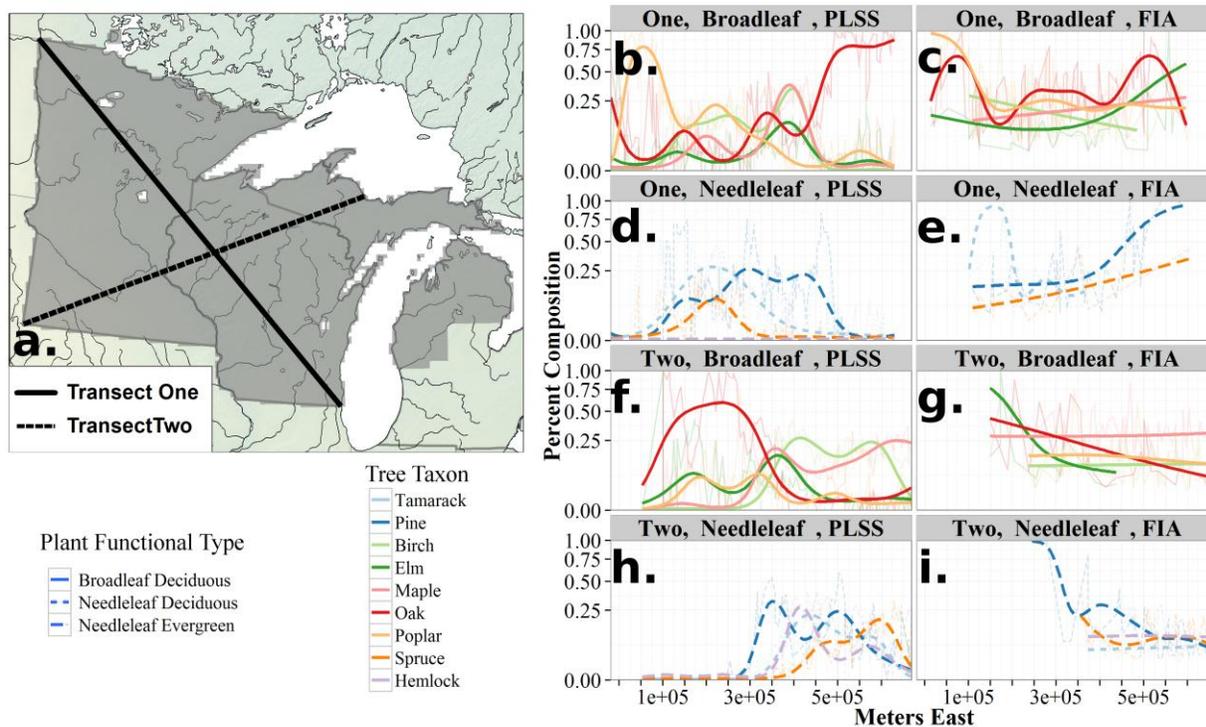
828

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

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

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

849



850
 851 **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.*

858 Compositional Changes Between PLS and FIA Forests: Ecotone Structure

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

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

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

881 The second transect shows a similar pattern, with well defined ecotones in the pre-
882 settlement period (Figure 11f and h), that are largely absent from the FIA data (Figure 11g
883 and i). Oak forest, with a component of elm and poplar in the southwest grades slowly to a
884 rapid transition zone where pine, elm, maple (first), then rapidly birch, hemlock and
885 tamarack, and later, spruce, increase. This region, the Tension Zone, extends from 3×10^5 to
886 4.5×10^5 meters East, eventually becoming a forest that shows co-dominance between birch,
887 pine, maple, spruce and tamarack, likely reflecting some local variability as a result of

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

890 Contemporary forests show broader homogenization and increased heterogeneity
891 (evidenced by the lower within-FIA Moran's I estimates for near-neighbor distances) at a
892 local scale in the region. Homogenization is evident across T1, where Bray-Curtis
893 dissimilarity between adjacent cells declines from the PLSS to the FIA ($\delta_{\text{beta}} = -0.22$, $t_{113} = -$
894 7.93 , $p < 0.001$), mirroring declines in the pine barrens between the 1950s and the present
895 (Li and Waller 2014). The PLS shows strong differentiation in the central region of T2
896 where maple-pine-oak shifts to pine-poplar-birch forest (Figure 11d). This sharp ecotone is
897 not apparent in the FIA data, which shows gradual and blurred changes in species
898 composition across the ecotone (Figure 11i). β -diversity along T2 is lower in the FIA than
899 in the PLSS ($\delta_{\text{beta}} = -0.19$, $t_{65} = -7.34$, $p < 0.01$), indicating higher heterogeneity in the PLS
900 data at the 64 km² meso-scale.

901 Across the entire domain, β diversity is lower in the FIA than in the PLS ($\delta_{\beta} = -0.172$, $t_{1.3e7}$
902 $= 2480$, $p < 0.001$), lending support to the hypothesis of overall homogenization. Differences
903 in sampling design between PLS and FIA data cannot explain this homogenization, since its
904 effect would have been expected to increase β -diversity along linear transects and at larger
905 spatial scales.

906 Discussion

907 Many forests of the PLS, are no longer a part of the modern landscape. Forest types have
908 been lost at the 64 km² mesoscale, and new forest types have been gained. The joint

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

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

928 Decreased β diversity along regional transects indicates homogenization at meso-scales of
929 100s of km², while the overall reduction in Moran's I for dissimilarity in the FIA indicates a
930 regional reduction in heterogeneity on the scale of 1000s of km². The selective loss or

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

938 Anthropogenic shifts in forest composition over decades and centuries seen here and
939 elsewhere (Cogbill et al. 2002, Thompson et al. 2013) are embedded within a set of
940 interacting systems that operate on multiple scales of space and time (macrosystems, *sensu*
941 Heffernan et al. 2014). Combining regional historical baselines, long term ecological studies
942 and high frequency analyses can reveal complex responses to climate change at local and
943 regional scales (Groffman et al. 2012). Estimates of pre-settlement forest composition and
944 structure are critical to understanding the processes that govern forest dynamics because
945 they represent a snapshot of the landscape prior to major EuroAmerican land-use
946 conversion (Schulte and Mladenoff 2001, Liu et al. 2011). Pre-settlement vegetation
947 provides an opportunity to test forest-climate relationships prior to land-use conversion
948 and to test dynamic vegetation models in a data assimilation framework (*e.g.*, Hartig et al.
949 2012). For these reason, the widespread loss of regional forest associations common in the
950 PLS (Figure 9d), and the rapid rise of novel forest assemblages (Figure 9e) have important
951 implications for our ability to understand ecological responses to changing climate. The
952 loss of historical forest types implies that the modern understanding of forest cover,
953 climate relationships, realized and potential niches and species associations may be

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

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

971 This study also points to the need for a deeper understanding of some of the landscape-
972 and regional-scale drivers of novelty, given the likely role for climatic and land use change
973 (including land abandonment) to continue to drive ecological novelty (Martinuzzi et al.
974 2015, Radeloff et al. in press). In particular the role of regional species pools and remnant
975 patches of forest in driving or mitigating compositional novelty. This work shows that the

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

999 The analysis relating to the distance-to-novelty (Figure 10) also points to the possibility
1000 that should the public push for landscape-scale restoration, or management using historical
1001 baselines, focusing on restoration in locations where restoration potential is high, as
1002 suggested for Hemlock/Hardwood forests in northern Wisconsin (Bolliger et al. 2004) has
1003 the possibility of contributing to landscape scale change. If some of the novelty is driven by
1004 depauperate species pools beyond certain threshold distances from remnant forests then it
1005 should also be possible to restore these forest through translocation of key species (Seddon
1006 2010).

1007 Methodological advances of the current work include 1) the systematic standardization of
1008 PLS data to enable mapping at broad spatial extent and high spatial resolution, 2) the use of
1009 spatially varying correction factors to accommodate variations among surveyors in
1010 sampling design, and 3) parallel analysis of FIA datasets to enable comparisons of forest
1011 composition and structure between contemporary and historical time periods. This
1012 approach is currently being extended to TPS and PLS datasets across the north-central and
1013 northeastern US, with the goal of providing consistent reconstructions of forest
1014 composition and structure for northeastern US forests at the time of EuroAmerican forests.

1015 Our results support the consensus that robust estimates of pre-settlement forest
1016 composition and structure can be obtained from PLS data (*e.g.*, Wisconsin: Schulte et al.
1017 2002, Iowa: Rayburn and Schulte 2009, California: Williams and Baker 2011, Oregon:
1018 Duren et al. 2012). Patterns of density, basal area and biomass are roughly equivalent to
1019 previous estimates (Schulte et al. 2007, Rhemtulla et al. 2009a). Our results for stem
1020 density are lower than those estimated by Hanberry *et al.* (Hanberry et al. 2012a) for

1021 eastern Minnesota, but density and basal area are similar to those in the northern Lower
1022 Peninsula of Michigan (Leahy and Pregitzer 2003) and biomass estimates are in line with
1023 estimates of aboveground carbon for Wisconsin (Rhemtulla et al. 2009a).

1024 These maps of settlement-era forest composition and structure can also provide a useful
1025 calibration dataset for pollen-based vegetation reconstructions for time periods prior to
1026 the historic record. Many papers have used calibration datasets comprised of modern
1027 pollen samples to build transfer functions for inferring past climates and vegetation from
1028 fossil pollen records (Jacques et al. 2008, Goring et al. 2009, Paciorek and McLachlan 2009,
1029 Birks et al. 2010). However, modern pollen datasets are potentially confounded by recent
1030 land use, which can alter paleoclimatic reconstructions using pollen data (Jacques et al.
1031 2008). By linking pollen and vegetation at modern and historical periods we develop
1032 capacity to provide compositional datasets at broader spatio-temporal scales, providing
1033 more data for model validation and improvement. Ultimately, it should be possible to
1034 assimilate these empirical reconstructions of past vegetation with dynamic vegetation
1035 models in order to infer forest composition and biomass during past climate changes. Data
1036 assimilation, however, requires assessment of observational and model uncertainty in the
1037 data sources used for data assimilation. Spatiotemporal models of uncertainty are being
1038 developed for the compositional data (Paciorek et al. in review) and biomass data (Feng *et*
1039 *al.* in prep.).

1040 Ultimately the pre-settlement vegetation data present an opportunity to develop and refine
1041 statistical and mechanistic models of terrestrial vegetation that can take multiple structural
1042 and compositional forest attributes into account. The future development of uncertainty

1043 estimates for the data remains an opportunity that can help integrate pre-settlement
1044 estimates of composition and structure into a data assimilation framework to build more
1045 complete and more accurate reconstructions of past vegetation dynamics, and to help
1046 improve predictions of future vegetation under global change scenarios.

1047 **Literature Cited**

- 1048 Almendinger, J. C. 1996. Minnesota's bearing tree database. Minn. Department of Natural
1049 Resources.
- 1050 Anderson, R. C., and M. R. Anderson. 1975. The presettlement vegetation of Williamson
1051 county, Illinois. *Castanea*:345–363.
- 1052 Auguie, B. 2012. GridExtra: Functions in Grid graphics.
- 1053 Bellemare, J., G. Motzkin, and D. R. Foster. 2002. Legacies of the agricultural past in the
1054 forested present: An assessment of historical land-use effects on rich mesic forests. *Journal*
1055 *of Biogeography* 29:1401–1420.
- 1056 Birks, H. B., O. Heiri, H. Seppä, and A. E. Bjune. 2010. Strengths and weaknesses of
1057 quantitative climate reconstructions based on late-Quaternary biological proxies. *Open*
1058 *Ecology Journal* 3:68–110.
- 1059 Bivand, R. 2014. spdep: Spatial dependence: Weighting schemes, statistics and models.
- 1060 Bivand, R., and C. Rundel. 2014. rgeos: Interface to geometry engine - open source (GEOS).
- 1061 Bivand, R., T. Keitt, and B. Rowlingson. 2014. rgdal: Bindings for the Geospatial Data
1062 Abstraction Library.

- 1063 Bivand, R., E. Pebesma, and V. Gomez-Rubio. 2013. Applied spatial data analysis with R.
1064 Second editions. Springer.
- 1065 Bolliger, J., L. A. Schulte, S. N. Burrows, T. A. Sickley, and D. J. Mladenoff. 2004. Assessing
1066 ecological restoration potentials of Wisconsin (USA) using historical landscape
1067 reconstructions. *Restoration Ecology* 12:124–142.
- 1068 Booth, R. K., S. T. Jackson, V. A. Sousa, M. E. Sullivan, T. A. Minckley, and M. J. Clifford. 2012.
1069 Multi-decadal drought and amplified moisture variability drove rapid forest community
1070 change in a humid region. *Ecology* 93:219–226.
- 1071 Bouldin, J. 2008. Some problems and solutions in density estimation from bearing tree
1072 data: A review and synthesis. *Journal of Biogeography* 35:2000–2011.
- 1073 Bourdo, E. A. 1956. A review of the General Land Office survey and of its use in quantitative
1074 studies of former forests. *Ecology*:754–768.
- 1075 Bowman, D. M., G. L. Perry, and J. Marston. 2015. Feedbacks and landscape-level vegetation
1076 dynamics. *Trends in ecology & evolution* 30:255–260.
- 1077 Brugam, R. B. 1978. Pollen indicators of land-use change in southern connecticut.
1078 *Quaternary Research* 9:349–362.
- 1079 Cogbill, C. V., J. Burk, and G. Motzkin. 2002. The forests of presettlement New England, USA:
1080 Spatial and compositional patterns based on town proprietor surveys. *Journal of*
1081 *Biogeography* 29:1279–1304.

- 1082 Cogbill, C. V., S. J. Goring, and A. Thurman. in prep. Estimation of robust correction factors
1083 for Public Land Survey Data.
- 1084 Cottam, G., and J. T. Curtis. 1956. The use of distance measures in phytosociological
1085 sampling. *Ecology*:451–460.
- 1086 Cramer, V. A., R. J. Hobbs, and R. J. Standish. 2008. What's new about old fields? Land
1087 abandonment and ecosystem assembly. *Trends in Ecology & Evolution* 23:104–112.
- 1088 Csardi, G., and T. Nepusz. 2006. The igraph software package for complex network
1089 research. *InterJournal Complex Systems*:1695.
- 1090 Curtis, J. T. 1959. *The vegetation of Wisconsin: An ordination of plant communities*.
1091 University of Wisconsin Press.
- 1092 Dupouey, J.-L., E. Dambrine, J.-D. Laffite, and C. Moares. 2002. Irreversible impact of past
1093 land use on forest soils and biodiversity. *Ecology* 83:2978–2984.
- 1094 Duren, O. C., P. S. Muir, and P. E. Hosten. 2012. Vegetation change from the Euro-American
1095 settlement era to the present in relation to environment and disturbance in southwest
1096 Oregon. *Northwest Science* 86:310–328.
- 1097 Ellis, E. C., and N. Ramankutty. 2008. Putting people in the map: Anthropogenic biomes of
1098 the world. *Frontiers in Ecology and the Environment* 6:439–447.
- 1099 Etienne, D., P. Ruffaldi, J. L. Dupouey, M. Georges-Leroy, F. Ritz, and E. Dambrine. 2013.
1100 Searching for ancient forests: A 2000 year history of land use in northeastern French

- 1101 forests deduced from the pollen compositions of closed depressions. *The Holocene* 23:678–
1102 691.
- 1103 Ewing, H. A. 2002. The influence of substrate on vegetation history and ecosystem
1104 development. *Ecology* 83:2766–2781.
- 1105 Faith, D. P., P. R. Minchin, and L. Belbin. 1987. Compositional dissimilarity as a robust
1106 measure of ecological distance. *Vegetatio* 69:57–68.
- 1107 Foster, D. R., G. Motzkin, and B. Slater. 1998. Land-use history as long-term broad-scale
1108 disturbance: Regional forest dynamics in central New England. *Ecosystems* 1:96–119.
- 1109 Foster, D., F. Swanson, J. Aber, I. Burke, N. Brokaw, D. Tilman, and A. Knapp. 2003. The
1110 importance of land-use legacies to ecology and conservation. *BioScience* 53:77–88.
- 1111 Friedlingstein, P., M. Meinshausen, V. K. Arora, C. D. Jones, A. Anav, S. K. Liddicoat, and R.
1112 Knutti. 2014. Uncertainties in cMIP5 climate projections due to carbon cycle feedbacks.
1113 *Journal of Climate* 27:511–526.
- 1114 Friedman, S. K., and P. B. Reich. 2005. Regional legacies of logging: Departure from
1115 presettlement forest conditions in northern Minnesota. *Ecological applications* 15:726–
1116 744.
- 1117 Fritschle, J. A. 2008. Reconstructing historic ecotones using the Public Land Survey: The
1118 lost prairies of Redwood National Park. *Annals of the Association of American Geographers*
1119 98:24–39.

- 1120 Fuller, J. L., D. R. Foster, J. S. McLachlan, and N. Drake. 1998. Impact of human activity on
1121 regional forest composition and dynamics in central New England. *Ecosystems* 1:76–95.
- 1122 Gimmi, U., and V. C. Radeloff. 2013. Assessing naturalness in northern Great Lakes forests
1123 based on historical land-cover and vegetation changes. *Environmental management*
1124 52:481–492.
- 1125 Goring, S., M. G. Pellatt, T. Lacourse, I. R. Walker, and R. W. Mathewes. 2009. A new
1126 methodology for reconstructing climate and vegetation from modern pollen assemblages:
1127 An example from British Columbia. *Journal of Biogeography* 36:626–638.
- 1128 Gray, A. N., T. J. Brandeis, J. D. Shaw, W. H. McWilliams, and P. D. Miles. 2012. Forest
1129 Inventory and Analysis database of the United States of America (FIA). Pages 255–264 *in* J.
1130 Dengler, J. Oldeland, F. Jansen, M. Chytry, J. Ewald, M. Finckh, F. Glockler, G. Lopez-Gonzalez,
1131 R. K. Peet, and J. H. J. Schaminee, editors. *Vegetation databases for the 21st century.*–
1132 *Biodiversity & Ecology.*
- 1133 Groffman, P. M., L. E. Rustad, P. H. Templer, J. L. Campbell, L. M. Christenson, N. K. Lany, A.
1134 M. Socci, M. A. Vadeboncoeur, P. G. Schaberg, G. F. Wilson, and others. 2012. Long-term
1135 integrated studies show complex and surprising effects of climate change in the northern
1136 hardwood forest. *Bioscience* 62:1056–1066.
- 1137 Grossmann, E. B., and D. J. Mladenoff. 2008. Farms, fires, and forestry: Disturbance legacies
1138 in the soils of the northwest Wisconsin (USA) sand plain. *Forest Ecology and Management*
1139 256:827–836.

- 1140 Hanberry, B. B., S. Fraver, H. S. He, J. Yang, D. C. Dey, and B. J. Palik. 2011. Spatial pattern
1141 corrections and sample sizes for forest density estimates of historical tree surveys.
1142 *Landscape Ecology* 26:59–68.
- 1143 Hanberry, B. B., B. J. Palik, and H. S. He. 2012a. Comparison of historical and current forest
1144 surveys for detection of homogenization and mesophication of Minnesota forests.
1145 *Landscape Ecology* 27:1495–1512.
- 1146 Hanberry, B. B., J. Yang, J. M. Kabrick, and H. S. He. 2012b. Adjusting forest density estimates
1147 for surveyor bias in historical tree surveys. *The American Midland Naturalist* 167:285–306.
- 1148 Hartig, F., J. Dyke, T. Hickler, S. I. Higgins, R. B. OâHara, S. Scheiter, and A. Huth. 2012.
1149 Connecting dynamic vegetation models to data—an inverse perspective. *Journal of*
1150 *Biogeography* 39:2240–2252.
- 1151 Haxeltine, A., and I. C. Prentice. 1996. BIOME3: An equilibrium terrestrial biosphere model
1152 based on ecophysiological constraints, resource availability, and competition among plant
1153 functional types. *Global Biogeochemical Cycles* 10:693–709.
- 1154 Heffernan, J. B., P. A. Soranno, M. J. Angilletta Jr, L. B. Buckley, D. S. Gruner, T. H. Keitt, J. R.
1155 Kellner, J. S. Kominoski, A. V. Rocha, J. Xiao, T. K. Harms, S. J. Goring, L. E. Koenig, W. H.
1156 McDowell, H. Powell, A. D. Richardson, C. A. Stow, R. Vargas, and K. C. Weathers. 2014.
1157 *Macrosystems ecology: Understanding ecological patterns and processes at continental*
1158 *scales. Frontiers in Ecology and the Environment* 12:5–14.
- 1159 Hijmans, R. J. 2014. raster: Geographic data analysis and modeling.

- 1160 Hobbs, R. J., E. S. Higgs, and J. A. Harris. 2014a. Novel ecosystems: Concept or inconvenient
1161 reality? A response to Murcia et al. *Trends in ecology & evolution* 29:645–646.
- 1162 Hobbs, R. J., E. Higgs, C. M. Hall, P. Bridgewater, F. S. Chapin III, E. C. Ellis, J. J. Ewel, L. M.
1163 Hallett, J. Harris, K. B. Hulvey, S. T. Jackson, P. L. Kennedy, C. Kueffer, L. Lach, T. C. Lantz, A.
1164 E. Lugo, J. Mascaro, S. D. Murphy, C. R. Nelson, M. P. Perring, D. M. Richardson, T. R. Seastedt,
1165 R. J. Standish, B. M. Starzomski, K. N. Suding, P. M. Tognetti, L. Yakob, and L. Yung. 2014b.
1166 Managing the whole landscape: Historical, hybrid, and novel ecosystems. *Frontiers in*
1167 *Ecology and the Environment* 12:557–564.
- 1168 Hotchkiss, S. C., R. Calcote, and E. A. Lynch. 2007. Response of vegetation and fire to Little
1169 Ice Age climate change: Regional continuity and landscape heterogeneity. *Landscape*
1170 *Ecology* 22:25–41.
- 1171 Iverson, L. R., and D. McKenzie. 2013. Tree-species range shifts in a changing climate:
1172 Detecting, modeling, assisting. *Landscape Ecology* 28:879–889.
- 1173 Iverson, L. R., and A. M. Prasad. 1998. Predicting abundance of 80 tree species following
1174 climate change in the eastern United States. *Ecological Monographs* 68:465–485.
- 1175 Jacques, J.-M. S., B. F. Cumming, and J. P. Smol. 2008. A pre-European settlement pollen–
1176 climate calibration set for Minnesota, USA: Developing tools for palaeoclimatic
1177 reconstructions. *Journal of Biogeography* 35:306–324.
- 1178 Jenkins, J. C., D. C. Chojnacky, L. S. Heath, R. A. Birdsey, and others. 2004. Comprehensive
1179 database of diameter-based biomass regressions for North American tree species.

- 1180 Knoot, T. G., L. A. Schulte, J. C. Tyndall, and B. J. Palik. 2010. The state of the system and
1181 steps toward resilience of disturbance-dependent oak forests. *Ecology and Society* 15:5.
- 1182 Kronenfeld, B. J. 2014. Validating the historical record: A relative distance test and
1183 correction formula for selection bias in presettlement land surveys. *Ecography*.
- 1184 Kronenfeld, B. J., and Y.-C. Wang. 2007. Accounting for surveyor inconsistency and bias in
1185 estimation of tree density from presettlement land survey records. *Canadian Journal of*
1186 *Forest Research* 37:2365–2379.
- 1187 Kronenfeld, B. J., Y.-C. Wang, and C. P. Larsen. 2010. The influence of the “Mixed Pixel”?
1188 Problem on the detection of analogous forest communities between presettlement and
1189 present in western New York. *The Professional Geographer* 62:182–196.
- 1190 Leahy, M. J., and K. S. Pregitzer. 2003. A comparison of presettlement and present-day
1191 forests in northeastern lower Michigan. *The American Midland Naturalist* 149:71–89.
- 1192 Li, D.-J., and D. M. Waller. 2014. Drivers of observed biotic homogenization in pine barrens
1193 of central Wisconsin. *Ecology*.
- 1194 Liu, F., D. J. Mladenoff, N. S. Keuler, and L. S. Moore. 2011. BROADSCALE variability in tree data
1195 of the historical Public Land Survey and its consequences for ecological studies. *Ecological*
1196 *Monographs* 81:259–275.
- 1197 Lynch, E. A., R. Calcote, S. C. Hotchkiss, and M. Tweiten. 2014. Presence of lakes and
1198 wetlands decreases resilience of jack pine ecosystems to late-holocene climatic changes.
1199 *Canadian Journal of Forest Research* 44:1331–1343.

- 1200 Maechler, M., P. Rousseeuw, A. Struyf, M. Hubert, and K. Hornik. 2014. Cluster: Cluster
1201 analysis basics and extensions.
- 1202 Manies, K. L., and D. J. Mladenoff. 2000. Testing methods to produce landscape-scale
1203 presettlement vegetation maps from the US Public Land Survey records. *Landscape Ecology*
1204 15:741–754.
- 1205 Manies, K. L., D. J. Mladenoff, and E. V. Nordheim. 2001. Assessing large-scale surveyor
1206 variability in the historic forest data of the original US Public Land Survey. *Canadian*
1207 *Journal of Forest Research* 31:1719–1730.
- 1208 Martinuzzi, S., G. I. Gavier-Pizarro, A. E. Lugo, and V. C. Radeloff. 2015. Future land-use
1209 changes and the potential for novelty in ecosystems of the united states. *Ecosystems*:1–11.
- 1210 Matthes, J. H., S. Goring, J. W. Williams, and M. C. Dietze. in revision. Historical vegetation
1211 reconstruction benchmarks CMIP5 pre-colonial land-climate feedbacks across the upper
1212 Midwest and northeastern United States. *Global Change Ecology*.
- 1213 McAndrews, J. H. 1988. Human disturbance of North American forests and grasslands: The
1214 fossil pollen record. Pages 673–697 *in* *Vegetation history*. Springer.
- 1215 Menz, M. H., K. W. Dixon, and R. J. Hobbs. 2013. Hurdles and opportunities for landscape-
1216 scale restoration. *Science* 339:526–527.
- 1217 Mladenoff, D. J., S. E. Dahir, E. V. Nordheim, L. A. Schulte, and G. G. Guntenspergen. 2002.
1218 Narrowing historical uncertainty: Probabilistic classification of ambiguously identified tree
1219 species in historical forest survey data. *Ecosystems* 5:539–553.

- 1220 Morisita, M. 1954. Estimation of population density by spacing method. *Memoirs of the*
1221 *Faculty of Science Kyushu University, Series E* 1:187–197.
- 1222 Morisita, M. 1957. A new method for the estimation of density by the spacing method
1223 applicable to non-randomly distributed populations (in Japanese). *Physiology and Ecology*
1224 (Kyoto) 7:134–144.
- 1225 Munoz, S. E., D. J. Mladenoff, S. Schroeder, and J. W. Williams. 2014. Defining the spatial
1226 patterns of historical land use associated with the indigenous societies of eastern North
1227 America. *Journal of Biogeography*.
- 1228 Murcia, C., J. Aronson, G. H. Kattan, D. Moreno-Mateos, K. Dixon, and D. Simberloff. 2014. A
1229 critique of the 'novel ecosystem' concept. *Trends in ecology & evolution* 29:548–553.
- 1230 Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P.
1231 Solymos, M. H. H. Stevens, and H. Wagner. 2014. *vegan: Community ecology package*.
- 1232 Paciorek, C. J., and J. S. McLachlan. 2009. Mapping ancient forests: Bayesian inference for
1233 spatio-temporal trends in forest composition using the fossil pollen proxy record. *Journal*
1234 *of the American Statistical Association* 104:608–622.
- 1235 Paciorek, C., S. J. Goring, A. Thurman, C. Cogbill, J. Williams, D. Mladenoff, J. Peters, J. Zhu,
1236 and M. J.S. in review. Statistically-estimated tree composition for the northeastern United
1237 States at the time of Euro-American settlement. *PLoS One*.
- 1238 Pebesma, E., and R. Bivand. 2005. Classes and methods for spatial data in R. *R News* 5.

- 1239 Pederson, N., J. M. Dyer, R. W. McEwan, A. E. Hessler, C. J. Mock, D. A. Orwig, H. E. Rieder, and
1240 B. I. Cook. 2014. The legacy of episodic climatic events in shaping temperate, broadleaf
1241 forests. *Ecological Monographs* 84:599–620.
- 1242 Persson, O. 1971. The robustness of estimating density by distance measurements. *in* G. P.
1243 Patil, E. C. Pielou, and W. E. Waters, editors. *Statistical Ecology: Sampling and Modeling*
1244 *Biological Populations and Population Dynamics*. Pennsylvania University Press, University
1245 Park, Pennsylvania.
- 1246 Radeloff, V. C., D. J. Mladenoff, and M. S. Boyce. 2000. A historical perspective and future
1247 outlook on landscape scale restoration in the northwest Wisconsin pine barrens.
1248 *Restoration Ecology* 8:119–126.
- 1249 Radeloff, V. C., J. W. Williams, B. L. Bateman, K. D. Burke, S. K. Carter, E. S. Childress, K.
1250 Cromwell, C. Gratton, A. O. Hasley, B. M. Kraemer, A. W. Latzka, E. Marin-Spiotta, C. D.
1251 Meine, S. E. Munoz, T. M. Neeson, A. M. Pidgeon, A. R. Rissman, R. J. Rivera, L. M. Szymanski,
1252 and J. Usinowicz. *in press*. The rise of novelty in ecosystems. *Ecological Applications*.
- 1253 Ramankutty, N., and J. A. Foley. 1999. Estimating historical changes in global land cover:
1254 Croplands from 1700 to 1992. *Global biogeochemical cycles* 13:997–1027.
- 1255 Ratajczak, Z., J. B. Nippert, and S. L. Collins. 2012. Woody encroachment decreases diversity
1256 across North American grasslands and savannas. *Ecology* 93:697–703.
- 1257 Rayburn, A. P., and L. A. Schulte. 2009. Integrating historic and contemporary data to
1258 delineate potential remnant natural woodlands within Midwestern agricultural landscapes.
1259 *Natural Areas Journal* 29:4–14.

- 1260 Record, S., M. C. Fitzpatrick, A. O. Finley, S. Veloz, and A. M. Ellison. 2013. Should species
1261 distribution models account for spatial autocorrelation? A test of model projections across
1262 eight millennia of climate change. *Global ecology and biogeography* 22:760–771.
- 1263 Rhemtulla, J. M., D. J. Mladenoff, and M. K. Clayton. 2009a. Historical forest baselines reveal
1264 potential for continued carbon sequestration. *Proceedings of the National Academy of*
1265 *Sciences* 106:6082–6087.
- 1266 Rhemtulla, J. M., D. J. Mladenoff, and M. K. Clayton. 2009b. Legacies of historical land use on
1267 regional forest composition and structure in wisconsin, USA (mid-1800s-1930s-2000s).
1268 *Ecological Applications* 19:1061–1078.
- 1269 Schulte, L. A., and D. J. Mladenoff. 2001. The original US Public Land Survey records: Their
1270 use and limitations in reconstructing presettlement vegetation. *Journal of Forestry* 99:5–
1271 10.
- 1272 Schulte, L. A., and D. J. Mladenoff. 2005. Severe wind and fire regimes in northern forests:
1273 Historical variability at the regional scale. *Ecology* 86:431–445.
- 1274 Schulte, L. A., D. J. Mladenoff, and E. V. Nordheim. 2002. Quantitative classification of a
1275 historic northern Wisconsin (USA) landscape: Mapping forests at regional scales. *Canadian*
1276 *Journal of Forest Research* 32:1616–1638.
- 1277 Schulte, L. A., D. J. Mladenoff, S. N. Burrows, T. A. Sickley, and E. V. Nordheim. 2005. Spatial
1278 controls of pre-Euro-American wind and fire disturbance in northern Wisconsin (USA)
1279 forest landscapes. *Ecosystems* 8:73–94.

- 1280 Schulte, L. A., D. J. Mladenoff, T. R. Crow, L. C. Merrick, and D. T. Cleland. 2007.
1281 Homogenization of northern US Great Lakes forests due to land use. *Landscape Ecology*
1282 22:1089–1103.
- 1283 Seddon, P. J. 2010. From reintroduction to assisted colonization: Moving along the
1284 conservation translocation spectrum. *Restoration Ecology* 18:796–802.
- 1285 Staver, A. C., S. Archibald, and S. A. Levin. 2011. The global extent and determinants of
1286 savanna and forest as alternative biome states. *Science* 334:230–232.
- 1287 Stewart, L. O. 1935. *Public Land Surveys: History, instructions, methods*. Page 202.
1288 Collegiate Press, Incorporated, Ames, Iowa, USA.
- 1289 Stoy, P. C., M. Mauder, T. Foken, B. Marcolla, E. Boegh, A. Ibrom, M. A. Arain, A. Arneth, M.
1290 Aurela, C. Bernhofer, and others. 2013. A data-driven analysis of energy balance closure
1291 across FLUXNET research sites: The role of landscape scale heterogeneity. *Agricultural and*
1292 *forest meteorology* 171:137–152.
- 1293 Team, R. C. 2014. *R: A language and environment for statistical computing (version 3.1.0)*.
1294 vienna, Austria: R Foundation for Statistical Computing.
- 1295 Terrail, R., D. Arseneault, M.-J. Fortin, S. Dupuis, and Y. Boucher. 2014. An early forest
1296 inventory indicates high accuracy of forest composition data in pre-settlement land survey
1297 records. *Journal of Vegetation Science* 25:691–702.
- 1298 Thompson, J. R., D. N. Carpenter, C. V. Cogbill, and D. R. Foster. 2013. Four centuries of
1299 change in northeastern United States forests. *PloS one* 8:e72540.

- 1300 Tomscha, S. A., and S. E. Gergel. 2014. Historic land surveys present opportunities for
1301 reconstructing frontier settlement patterns in North America. *Landscape Ecology* 30:203–
1302 213.
- 1303 Umbanhowar, C. E., P. Camill, C. E. Geiss, and R. Teed. 2006. Asymmetric vegetation
1304 responses to mid-Holocene aridity at the prairie–forest ecotone in south-central Minnesota.
1305 *Quaternary Research* 66:53–66.
- 1306 White, C. A. 1983. A history of the rectangular survey system. Page 774. US Department of
1307 the Interior, Bureau of Land Management, Washington D.C., USA.
- 1308 Wickham, H. 2007. Reshaping data with the reshape package. *Journal of Statistical Software*
1309 21:1–20.
- 1310 Wickham, H. 2009a. *ggplot2: Elegant graphics for data analysis*. Springer.
- 1311 Wickham, H. 2009b. *ggplot2: Elegant graphics for data analysis*. Springer New York.
- 1312 Wickham, H. 2011. The split-apply-combine strategy for data analysis. *Journal of Statistical*
1313 *Software* 40:1–29.
- 1314 Williams, M. A., and W. L. Baker. 2011. Testing the accuracy of new methods for
1315 reconstructing historical structure of forest landscapes using GLO survey data. *Ecological*
1316 *Monographs* 81:63–88.
- 1317 Wolkovich, E., B. Cook, K. McLauchlan, and T. Davies. 2014. Temporal ecology in the
1318 anthropocene. *Ecology letters* 17:1365–1379.

- 1319 Wood, S. 2011. Fast stable restricted maximum likelihood and marginal likelihood
1320 estimation of semiparametric generalized linear models. *Journal of the Royal Statistical*
1321 *Society (B)* 73:3–36.
- 1322 Woudenberg, S. W., B. L. Conkling, B. M. O’Connell, E. B. LaPoint, J. A. Turner, and K. L.
1323 Waddell. 2010. The Forest Inventory and Analysis database: Database description and
1324 users manual version 4.0 for phase 2.