1   # Witness Tree paper

2   Simon Goring *et al.*

3   22 October, 2015

4   **Changes in Forest Composition, Stem Density, and Biomass from the**

5   **Settlement Era (1800s) to Present in the Upper Midwestern United**

6   **States**

7   **Author List**

8   Simon J. Goring[1]

9   David J. Mladenoff[2]

10   Charles V. Cogbill[3]

11   Sydne Record[3,4]

12   Christopher J. Paciorek[5]

13   Stephen T. Jackson[6]

14   Michael C. Dietze[7]

15   Andria Dawson [5]

16   Jaclyn Hatala Matthes[8]

17   Jason S. McLachlan[9]

18   John W. Williams[1,10]

19   [1]Department of Geography, University of Wisconsin, Madison, 550 N Park St, Madison WI

20   53706

21   [2]Department of Forest and Wildlife Ecology, University of Wisconsin-Madison, 1630 Linden

22   Dr, Madison WI 53706

23   [3]Harvard Forest, Harvard University, 324 N Main St, Petersham MA 01366

24   [4]Department of Biology, Bryn Mawr College, 101 North Merion Ave., Bryn Mawr PA 19010

25   [5]Department of Statistics, University of California, Berkeley, 367 Evans Hall, Berkeley CA

26   94720

27   [6]Department of the Interior Southwest Climate Science Center, U.S. Geological Survey, 1955

28   E. Sixth St. Tucson, AZ 85719; School of Natural Resources and the Environment and

29   Department of Geosciences, University of Arizona, Tucson AZ 85721

30   [7]Department of Earth and Environment, Boston University, 685 Commonwealth Ave,

31   Boston, MA 02215

32   [8]Department of Geography, Dartmouth College, 6017 Fairchild, Hanover, NH 03755

33   [9]Department of Biological Sciences, University of Notre Dame, 100 Galvin Life Sciences

34   Center, Notre Dame, IN 46556

35   [10]Center for Climatic Research, University of Wisconsin, Madison, 1225 W Dayton St,

36   Madison WI 53706

37   **Corresponding Author**: Simon Goring - goring@wisc.edu

38   ───────────────────────────────────────

Abstract

40   *EuroAmerican land use and its legacies have transformed forest structure and composition*

41   *across the United States (US). More accurate reconstructions of historical states are critical to*

42   *understanding the processes governing past, current, and future forest dynamics. Gridded*

43   *(8x8km) estimates of pre-settlement (1800s) forests from the upper Midwestern US*

44   *(Minnesota, Wisconsin, and most of Michigan) using 19th Century Public Land Survey (PLS)*

45   *records provide relative composition, biomass, stem density, and basal area for 26 tree*

46   *genera. This mapping is more robust than past efforts, using spatially varying correction*

47   *factors to accommodate sampling design, azimuthal censoring, and biases in tree selection.*

48   *We compare pre-settlement to modern forests using Forest Inventory and Analysis (FIA) data,*

49   *with respect to structural changes and the prevalence of lost forests, pre-settlement forests*

50   *with no current analogue, and novel forests, modern forests with no past analogs. Differences*

51   *between PLSS and FIA forests are spatially structured as a result of differences in the*

52   *underlying ecology and land use impacts in the Upper Midwestern United States. Modern*

53   *biomass is higher than pre-settlement biomass in the northwest (Minnesota and north-*

54   *eastern Wisconsin, including regions that were historically open savanna), and lower in the*

55   *east (eastern Wisconsin and Michigan), due to shifts in species composition and, presumably,*

56   *average stand age. Modern forests are more homogeneous, and ecotonal gradients are more*

57   *diffuse today than in the past. Novel forest assemblages represent 29% of all FIA cells, while*

58   *25% of pre-settlement forests no longer exist in a modern context.*

59   *Lost forests are centered around the forests of the Tension Zone, particularly in hemlock*

60   *dominated forests of north-central Wisconsin, and in oak-elm-basswood forests along the*

61   *forest-prairie boundary in south central Minnesota and eastern Wisconsin. Novel FIA forest*

62   *assemblages are distributed evenly across the region, but novelty shows a strong relationship*

63   *to spatial distance from remnant forests in the upper Midwest, with novelty predicted at*

64   *between 20 to 60km from remnants, depending on historical forest type.*

65   *The spatial relationships between remnant and novel forests, shifts in ecotone structure and*

66   *the loss of historic forest types point to significant challenges to land managers if landscape*

67   *restoration is a priority in the region. The spatial signals of novelty and ecological change also*

68   *point to potential challenges in using modern spatial distributions of species and communities*

69   *and their relationship to underlying geophysical and climatic attributes in understanding*

70   *potential responses to changing climate. The signal of human settlement on modern forests is*

71   *broad, spatially varying and acts to homogenize modern forests relative to their historic*

72   *counterparts, with significant implications for future management.*

Introduction:

76 The composition, demography, and structure of forests in eastern North America have

77 changed continuously over the last millennium, driven by human land use [1–5] and

78 climate variability [6–9]. While human effects have been a component of these systems for

79 millenia, the EuroAmerican settlement and industrialization period have increased

80 anthropogenic effects by orders of magnitude [10–12]. Legacies of post-settlement land use

81 in the upper Midwest [13] and elsewhere have been shown to persist at local and regional

82 scales [5,14,15], and nearly all North American forests have been affected by the

83 intensification of land use in the past three centuries. Hence, contemporary ecological

84 processes in North American forests integrate the anthropogenic impacts of the post-

85 EuroAmerican period and natural influences at decadal to centennial scales.

86 At a regional scale many forests in the upper Midwest (*i.e.*, Minnesota, Wisconsin and

87 Michigan) now have decreased species richness and functional diversity relative to forests

88 of the pre-EuroAmerican settlement (hereafter pre-settlement) period [16–18] due to near

89 complete logging. For example, forests in Wisconsin are in a state of regrowth, with an

90 unfilled carbon sequestration potential of 69 TgC [19] as a consequence of these extensive

91 land cover conversions and subsequent partial recovery following abandonment of farm

92 lands in the 1930s. But while regional patterns may establish themselves across the

93 midwest, the range of ecozones and patterns of land use in space and time result in both

94 broad spatial patterns, but significant local to regional variation. For example, while fire

95 suppression occured throughout the region effects of suppression have and will continue to

96   manifest themselves differently depending on the historical vegetation and biophyssical

97   characteristics of the site or region.

98   Legacies of land use are unavoidable at regional scales [20]. Under intensive land use

99   change the natural processes of secession, senescense and the replacement of tree species

100   in forests may be masked, or heavily modified by historically recent land use change.

101   Broad-scale land use change can result in changes to forest structure and species pools that

102   may result in non-stationarity within ecosystems that may not be apparent on the

103   relatively narrow time scales at which ecology traditionally operates [21], meaning

104   chronosequences may not be sufficeint to understand shifts in structure and composition.

105   There is a history of recolonization of forested landscapes following agricultural clearance

106   in the upper Midwest [22], pointing to the importance of understanding ecological

107   trajectories and land use legacies in understanding modern forest dynamics [20]. Cramer*el*

108   *al.*. [23] point to the literature of succession theory to indicate the likelihood that many old

109   fields will return to a 'natural' state, but point out that recovery is not universal. In

110   particular, intense fragmentation of the landscape can deplete the regional species pool,

111   leading to failures of recruitment that would favor species with longer distance seed

112   dispersersal [24]. In the upper Midwest long seed dispersal would favor species such as

113   poplar (*Populus* sp.), white birch (*Betula papyrifera*) and some maple species (*Acer* sp.), at

114   the expense of large-seeded species such as walnut (*Juglans* sp.), oak (*Quercus* sp.) and

115   others.

116   While there remains debate over the utility of the concept of novel ecosystems [25,26], the

117   fact remains that there are now forest and vegetation communities on the landscape

118   without past analogues. The long term management of the systems and their associated

119   services requires a broad understanding of the extent to which landscapes have been

120   modified, and the extent to which land use change has potenitally masked underlying

121   processes. It also requires a better understanding of the spatial (and temporal) scales at

122   which novel ecosystems operate. While restoration efforts have generally focused on

123   ecosystems at local scales, there is an increasing need to focus on management and

124   restoration at landscape scales [27]. Thus a better understanding of the landscape-scale

125   processes driving novelty, the spatial structure of novel ecosystems and their ecological

126   correlates, is increasingly important. An understanding of landscape level processes

127   driving ecological novelty can help prioritize intervention strategies at local scales [28],

128   and give us a better understanding of the role of patches in restoring hybrid or novel

129   landscapes. In particular, how important is the species pool to the development of novel

130   landscapes? Are novel forests further from remnant forests than might otherwise be

131   expected? Is novelty operating at landscape scales in the upper Midwest, and is the spatial

132   distribution of new forests tied to historical patterns vegetation or losses of forest types

133   from the historical landscape?

134   The upper Midwestern United States represents a unique ecological setting, with multiple

135   major ecotones, including the prairie-forest boundary, historic savanna, and the Tension

136   Zone between southern deciduous forests and northern evergreen forests. The extent to

137   which these ecotones have shifted, and their extent both prior to and following

138   EuroAmerican settlement is of critical importance to biogeochemical and biogeophysical

139   vegetation-atmosphere feedbacks [29], carbon sequestration [19], and regional

140   management and conservation policy [30–33].

141 Land use change at the local and state-level has affected both the structure and

142 composition of forests in the Midwestern United States [16,17]. Homogenization and shifts

143 in overall forest composition are evident, but the spatial extent and structure of this effect

144 is less well understood. Studies in Wisconsin have shown differential patterns of change in

145 the mixedwood and evergreen dominated north, the southern driftless and hardwood

146 dominated forests in south-central Wisconsin, and the prairie and savanna ecosystems that

147 bound the region to the south and west. Does this pattern of differential change extend to

148 Minnesota and Michigan? To what extent are land-use effects common across the region,

149 and where are responses ecozone-specific? Has homogenization [16] resulted in novel

150 forest assemblages relative to pre-settlement baselines across the region, and the loss of

151 pre-settlement forest types? Are the spatial distributions of these novel and lost forest

152 types overlapping, or do they have non-overlapping extents? If broad-scale reorganization

153 is the norm following EuroAmerican settlement, then the ecosystems that we have been

154 studying for the past century may indeed be novel relative to the reference conditions of

155 the pre-settlement era.

156 Modern forest structure and composition data [34] play a ubiquitous role in forest

157 management, conservation, carbon accounting, and basic research on forest ecosystems

158 and community dynamics. These recent surveys (the earliest FIA surveys began in the

159 1930s) can be extended with longer-term historical data to understand how forest

160 composition has changed since EuroAmerican settlement. The Public Land Survey was

161 carried out ahead of mass EuroAmerican settlement west and south of Ohio to provide for

162 delineation and sale of the public domain beyond the original East Coast states [35,36].

163 Because surveyors used trees to locate survey points, recording the identity, distance, and

164   directory of two to four trees next to each survey marker, we can make broad-scale

165   inferences about forest composition and structure in the United States prior to large-scale

166   EuroAmerican settlement [37–40]. In general, FIA datasets are systematically organized

167   and widely available to the forest ecology and modeling community, whereas most PLSS

168   data compilations are of local or, at most, state-level extent. This absence of widely

169   available data on settlement-era forest composition and structure limits our ability to

170   understand and model the current and future processes governing forest dynamics at

171   broader, regional scales. For example, distributional models of tree species often rely upon

172   FIA or other contemporary observational data to build species-climate relationships that

173   can be used to predict potential range shifts [41,42].

174   Here we use survey data from the original Public Lands Surveys (PLS) in the upper

175   Midwest to derive estimates of pre-settlement (*ca*. mid-late 1800s) forest composition,

176   basal area, stem density, and biomass. This work builds upon prior digitization and

177   classification of PLSS data for Wisconsin [43,44] and for parts of Minnesota [17,45] and

178   Michigan Michigan (USFS-NCRS http://www.ncrs.fs.fed.us/gla/). Most prior PLS-based

179   reconstructions are for individual states or smaller extents [17,19,45,46] often aggregated

180   at the scale of regional forest zones [16,17], although aggregation may also occur at the

181   section [19] or township scale [47]. Our work develops new approaches to address major

182   challenges to PLSS data, including lack of standardization in tree species names, azimuthal

183   censoring by surveyors, variations in sampling design over time, and differential biases in

184   tree selection among different kinds of survey points within the survey design at any point

185   in time. The correction factors developed here are spatially varying, allowing us to

186   accommodate temporal and spatial variations in surveyor methods.

187   We aggregate point based estimates of stem density, basal area and biomass to an 8 x 8km

188   grid, and classify forest types in the upper Midwest to facilitate comparisons between FIA

189   and PLSS data. We compare the PLSS data to late-20th-century estimates of forest

190   composition, tree stem density, basal area and biomass. We explore how forest

191   homogenization has changed the structure of ecotones along two major ecotones from

192   southern deciduous to northern evergreen forests and to the forest-prairie boundary.

193   Using analog analyses, we identify lost forests that have no close compositional counterpart

194   today and novel forests with no close historical analogs. This work provides insight into the

195   compositional and structural changes between historic and contemporary forests, while

196   setting the methodological foundation for a new generation of maps and analyses of

197   settlement-era forests in the Eastern US.

198   ## Methods:

199   ### Public Lands Survey Data: Assembly, and Standardization

200   The PLSS was designed to facilitate the division and sale of federal lands from Ohio

201   westward and south. The survey created a 1 mile^2 (2.56 km^2) grid (sections) on the

202   landscape. At each section corner, a stake was placed as the official location marker. To

203   mark these survey points, PLSS surveyors recorded tree stem diameters, measured

204   distances and azimuths of the two to four trees 'closest'to the survey point and identified

205   tree taxa using common (and often regionally idiosyncratic) names. PLSS data thus

206   represent measurements by hundreds of surveyors from 1832 until 1907, with changing

207   sets of instructions over time (Stewart, 1979).

208   The PLSS was undertaken to survey land prior to assigning ownership (Stewart 1935,

209   White 1983), replacing earlier town proprietor surveys (TPS) used for the northeastern

210   states [2,48]. The TPS provided estimates of relative forest composition at the township

211   level, but no structural attributes. The PLSS produced spatially explicit point level data,

212   with information about tree spacing and diameter, that can be used to estimate absolute

213   tree density and biomass. PLSS notes include tree identification at the plot level,

214   disturbance [49] and other features of the pre-settlement landscape. However,

215   uncertainties exist within the PLSS and township level dataset [50].

216   Ecological uncertainty in the PLSS 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 [51], digitization of the original survey

219   notes, and surveyor bias during sampling [38,50,52,53]. Estimates vary regarding the

220   ecological significance of surveyor bias. Terrail *et al.* [54] show strong fidelity between

221   taxon abundance in early land surveys versus old growth plot surveys. Liu *et al* [38]

222   estimate the ecological significance of some of the underlying sources of bias in the PLSS

223   and show ecologically significant (>10% difference between classes) bias in species and

224   size selection for corner trees. However Liu *et al.* [38] also indicate that the true sampling

225   error cannot be determined, particularly because most of these historic ecosystems are

226   largely lost to us.

227   Kronenfeld and Wang [55], 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 [56] shows adjustments of less than 1%. These

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

233   have assessed and corrected for the biases and idiosyncrasies in the original surveyor data

234   [17,38,39,53,55,57–60]. And, even given these caveats, PLSS records remain the best

235   source of data about both forest composition and structure in the United States prior to

236   EuroAmerican settlement.

237   This analysis builds upon and merges previous state-level efforts to digitize and database

238   the point-level PLSS data for Wisconsin, Minnesota and the Upper Peninsula and upper

239   third of the Lower Peninsula of Michigan. These datasets were combined using spatial tools

240   in R [61,62] to form a common dataset for the upper Midwest (Fig 1) using the Albers Great

241   Lakes and St Lawrence projection (see code in Supplement 1, file: *step_one_clean_bind.R*;
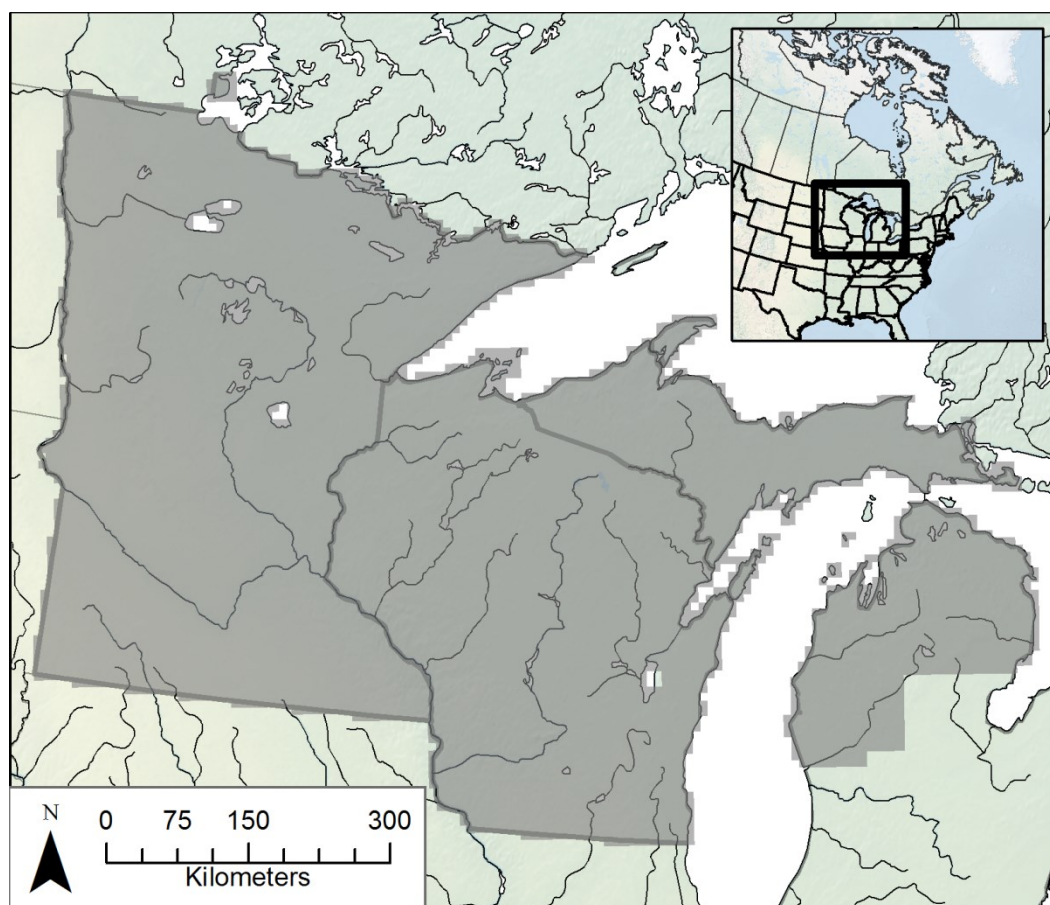
242   proj4: *+init:EPSG:3175*).

243

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

248   We took several steps to standardize the dataset and minimize the potential effects of

249   surveyor bias upon estimates of forest composition, density, and biomass. All steps are

250   preserved in the supplementary R code (Supplement 1: *step_one_clean_bind.R*). First, we

251   excluded line and meander trees (i.e. trees encountered along survey lines, versus trees

252   located at section or quarter corners) because surveyor selection biases appear to have

253   been more strongly expressed for line trees, meander trees have non-random habitat

254   preferences [38], and the inherent differences in sampling design between line, meander

255   and corner points. We used only the closest two trees at each corner point because the

256   third and fourth furthest trees have stronger biases with respect to species composition

257   and diameter [38]. Corner points were used only if 1) there were at least two trees at a

258   survey point, 2) the two trees were from different quadrants (defined by the cardinal

259   directions), and 3) there were valid azimuths to the trees (a defined quadrant with an angle

260   between 0 and 90) and valid diameters (numeric, non-zero).

261   Many species-level identifications used by PLSS surveyors are ambiguous. Statistical

262   models can predict the identity of ambiguous species [51], but these models introduce a

263   second layer of uncertainty into the compositional data, both from the initial surveyors'

264   identification, and from the statistical disambiguation. Given the regional scale of the

265   analysis, and the inherent uncertainty in the survey data itself, we chose to avoid this layer

266   of taxonomic uncertainty, and retained only genus-level identification (Supplement 2,

267   *Standardized Taxonomy*). The ecological implications for the use of genera-level

268   taxonomies are important for this region. While fire tolerance is fairly well conserved

269   within genera, shade tolerance can vary. *Betula* contains shade intolerant *B. paperyfera* and

270   the intermediate *B. alleghaniensis*, while *Pinus* contains the very shade intolerant *P.*

271   *banksiana*, the intolerant *P. resinosa* and the shade tolerant *P. strobus*. For cases where

272   shade tolerance (or fire tolerance) varies strongly within a genera we examine the data to

273   determine the suitability of the assignment, or extent of confusion within the assigned

274   genera.

275   In areas of open prairie or other treeless areas, *e.g.* southwestern Minnesota, surveyors

276   recorded distances and bearings to 'Non Tree' objects. When points were to be located in

277   water bodies the point data indicates 'Water'. Points recorded "No Tree" are considered to

278   have been from extremely open vegetation, with an estimated point-level stem density of 0

279   stems ha^{-1}. We based our estimates on terrestrial coverage, so water cells are excluded

280   completely. Hence, absence of trees at "No Tree" locations does reduce the gridded

281   estimates of terrestrial stem density, but absence of trees at 'Water' locations does not.

282   Digitization of the original surveyor notebooks introduces the possibility of transcription

283   errors. The Wisconsin dataset was compiled by the Mladenoff lab group, and has

284   undergone several revisions over the last two decades in an effort to provide accurate data

285   [30,38,43,44,51]. The Minnesota transcription error rate is likely between 1 and 5%, and

286   the treatment of azimuths to trees varies across the dataset [37]. Michigan surveyor

287   observations were transcribed to mylar sheets overlaid on State Quadrangle maps, so that

288   the points were displayed geographically, and then digititized to a point based shapefile

289   (Ed Schools, pers. comm.; Great Lakes Ecological Assessment. USDA Forest Service

290   Northern Research Station. Rhinelander, WI. http://www.ncrs.fs.fed.us/gla/), carrying two

291   potential sources of transcription error. Preliminary assessment of Southern Michigan data

292   indicates a transcription error rate of 3 - 6%. To reduce errors associated with

293   transcription across all datasets, we exclude sites for which multiple large trees have a

294   distance of 1 link (20.12 cm) to plot center, trees with very large diameters (diameter at

295   breast height - dbh > 100 in; 254 cm), plots where the azimuth to the tree is unclear, and

296   plots where the tree is at plot center but has a recorded azimuth. All removed plots are

297   documented in the code used for analysis (Supplement 1: *step_one_clean_bind.R*) and are

298   commented for review.

Data Aggregation

300   We binned the point data using an 64km^2 grid (Albers Gt. Lakes St Lawrence projection;

301   Supplement 1: *base_calculations.R*) to create a dataset that has sufficient numerical power

302   for spatial statistical modeling and sufficient resolution for regional scale analysis [63].

303   This resolution is finer than the 100km^2 gridded scale used in Freidman and Reich [45], but

304   coarser than township grids used in other studies [19,56] to provide a scale comparable to

305   aggregated FIA data at a broader scale. Forest compositional data is based on the number

306   of individuals of each genus or plant functional type (PFT) present at all points within a

307   cell. Stem density, basal area and biomass are averaged across all trees at all points within

308   the cell.

Stem Density

310   Estimating stem density from PLSS data is based on a plotless density estimator that uses

311   the measured distances from each survey point to the nearest trees at the plot location

312   [64,65]. The Morisita density estimator is then modified to minimize error due to different

313   sampling geometries and several known surveyor biases [17,38,39,53,55,57–60]. The

314   standardized approach for this method is well validated, however surveyors did not use a

315   consistent approach to plot level sampling. Survey sampling instructions changed

316   throughout the implementation of the PLSS in this region and differed between section and

317   quarter section points and between internal and external points within a township [36,38].

318   Our approach allows for spatial variation in surveyor methods by applying various

319   spatially different correction factors based not only on the empirical sample geometry, but

320   also on known surveyor biases deviating from this design [57]. These estimates are based

321   on empirical examination of the underlying data, and have been validated using

322   simulations on stem mapped stands [57].

323   We estimate stem density (stems m^{-2}) based on a on a modified form of the Morisita two-

324   tree density estimator, which uses the distance-to-tree measurements for the two closest

325   trees at each point [66]. Our modified form uses explicit and spatially varying correction

326   factors, modeled after the Cottam correction factor [67], that account for variations in

327   sampling designs over time and among surveyors. All code to perform the analysis is

328   included in Supplement 1.

329   We estimate the basic stem density (stems m^{-2}) using the point-to-tree distances for the

330   closest trees to each point within a defined number of sectors around the point (Reference

331   64 eqn 31.):

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

333   where λ is density ; k is the number of sectors within which trees are sampled, N is the

334   number of points over which estimates are aggregated, r is the distance of point-to-tree (as

335   m). This estimate can be modified by a refinement of the Cottam quadrant factors [66,67]

336   which recognizes that different sampling designs, and the order of the distances in different

337   quadrants (or sectors) carry specific weights. This correction, herein called κ, accounts for

338   different sampling designs. When either four quadrants or trees are sampled (point quarter

339   design), or when two trees in opposite semicircles (point halves design) are sampled, the

340   equation is accurate and $\kappa = 1$; when the two trees are in the nearest of two quadrants (two

341   nearest quadrants design), $\kappa = 0.857$; and when two trees are in quadrants on the same

342   side of the direction of travel (one-sided or interior half design), $\kappa = 2$. This parameter, in

343   Cottam's notation [67], is a divisor of the denominator above, or here, the mathematically

344   equivalent multiplier in the numerator of the reciprocal of the squared distances.

345   We further simplify the density estimate in equation (1) so that it is calculated at each point

346   (N=1) and for two sample trees only (k=2):

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

348   Then the point values for any sampling design can be Cottam corrected ($\kappa \times \lambda_M$). For

349   example, the basic Morisita equation for two sectors assumes trees are located in opposite

350   halves, so if the actual design is the nearest tree in the two nearest quadrants, the density

351   from equation 2 will be overestimated and must be correspondingly corrected by

352   multiplying by $\kappa = 0.857$.

353   Further corrections account for the restriction of trees to less than the full sector (θ),

354   censoring of trees near the cardinal azimuths (ζ), and undersampling of trees smaller than

355   a certain diameter limit (ϕ). These parameters are derived from analyses of measurements

356   of bearing angles and diameters actually observed in surveys of witness trees within a

357   subset of townships across the upper Midwest.

358   Sector bias (θ). Although the density model for two tree points assumes that the trees are

359   on opposite sides of a sample line (point halves), the actual sample is often more restricted

360   (< 180°) within the sector, or is a less restricted (> 180°) angle beyond the sector (see

361   Supplement 3). This deviation from the equation's assumption of equal distribution of

362   angles across the 180° sector is quantified using the empirical angle between the bearings

363   of the two trees (pair angle). The pair angle frequencies (Supplement 3) that the observed

364   proportion of trees (p) within any restricted sector divided by the proportion of that angle

365   within the circle (α) are an estimate of the bias imposed by the actual sampling [55]. The

366   factor ($\theta = p/\alpha$) indicates bias associated with differences in geometry of two tree samples.

367   This parameter (θ) varies from 0.71 to 1.27, indicating sampling from effectively 253° to

368   141° sectors.

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

370   near the cardinal directions [55,58,59]. This azimuthal censoring is commonly found along

371   the line of travel on section lines and sometimes on the perpendicular quarter-section lines.

372   Trees near the cardinal directions were passed over, and a replacement was found within a

373   more restricted angular region. The correction for this bias is calculated following

374   Kronenfeld and Wang [55] in a manner similar to the sector bias. The factor ζ is the ratio of

375   the proportion of trees in the restricted area (p) divided by the proportion of the complete

376   circle (α) that is used. The azimuthal censoring parameter (ζ) ranges from 1.03 to 1.25

377   indicating an equivalent to complete elimination of trees from 10° to 72° azimuths adjacent

378   to the cardinal directions.

379   Diameter limit (ϕ). Examination of the diameter distributions from settlement era surveys

380   across the upper Midwest clearly demonstrate witness trees less than 8 inches in diameter

381   were undersampled [38,57,59]. We have confirmed this bias in our own inspection of plots

382   of diameter frequency in the PLSS data, which show a strong mode at 8". This bias can be

383   accommodated by setting a diameter limit, and only calculating the density for trees with

384   diameters above this limit. Total density calculated from all trees is reduced to this

385   reference limit by simply multiplying the total by the percentage of trees above this limit.

386   This effectively eliminates the smaller trees from the total and normalizes the value of trees

387   above this standard. The parameter (ϕ) represents diameter size bias is simply the

388   percentage of trees $\geq$ 8" and, in practice, ranges from 0.6 - 0.9.

389   Because all surveyor bias corrections are simple multipliers of the model density and

390   should be independent, the bias-minimized estimate of the point density of trees $\geq$ 8" is:

391    $\lambda_{Mcorrected} = \kappa \times \theta \times \zeta \times \phi \times \lambda_M$ (3)

392   Estimates for each point *i* can be averaged for all *N* points in any region. Correction factors

393   are calculated separately for different regions, years, internal versus external lines, section

394   versus quarter-section points, and surveyor sampling designs (Supplement 4). All code to

395   perform the analyses is included in Supplement 1 and the full rationale for and calculation

396   of these measures is described further in Cogbill*el al.* [57]. Further, simulation used stem

397   mapped stands from the region presented in Cogbill*el al.* [57] supports the robustness of

398   this method, as opposed to other methods presented in the literature.

399   **Basal Area and Biomass Estimates**

400   Forest basal area is calculated by multiplying the point-based stem density estimate by the

401   average stem basal area from the reported diameters at breast height for the closest two

402   trees at the point (n=2). Aboveground dry biomass (Mg ha^{-1}) is calculated using the USFS

403   FIA tree volume and dry aboveground biomass equations for the United States [68].

404   Biomass equations share the basic form:

405    $m = Exp(\beta_0 + \beta_1 * lndbh)$

406   where m represents stem biomass for an individual tree in kg. β_0 and β_1 are parameters

407   derived from [68] and described in Table 1. dbh is the stem diameter at breast height

408   (converted to cm) recorded in the survey notes. The biomass estimates are summed across

409   both trees at a survey point and multiplied by the stem density calculated at that point to

410   produce an estimate of aboveground biomass reported in Mg ha^{-1} [68].

411   **Table 1.** *Biomass parameters used for the calculation of biomass in the pre-settlement*

412   *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	-	2.26	Tamarack, Cedar
	2.03		
Fir and Hemlock	-	2.43	Fir, Hemlock
	2.54		
Pine	-	2.43	Pine
	2.54		
Spruce	-	2.33	Spruce
	2.08		

413   Matching PLSS tree genera to the species groups defined by Jenkins *et al.* [68] is

414   straightforward, placing the 22 genera used in this study into 9 allometric groups (Table 1).

415   However, all maples are assigned to the generic "Hardwood" group since separate

416   allometric relationships exist for soft and hard maple (Table 1). Biomass estimates for "Non

417   tree" survey points are assigned 0 Mg ha^{-1}.

418   We use the stem density thresholds of Anderson and Anderson [69] to discriminate prairie,

419   savanna, and forest.

FIA Stem Density, Basal Area and Biomass

421   The United States Forest Service has monitored the nation's forests through the FIA

422   Program since 1929, with an annualized state inventory system implemented in 1998 [70].

423   On average there is one permanent FIA plot per 2,428 ha of land in the United States

424   classified as forested. Each FIA plot consists of four 7.2m fixed-radius subplots in which

425   measurements are made of all trees >12.7cm dbh [70]. We used data from the most recent

426   full plot inventory (2007-2011). The FIA plot inventory provides a median of 3 FIA plots

427   per cell using the 64km² grid.

428   We calculated mean basal area (m² ha^{-1}), stem density (stems ha^{-1}), mean diameter at

429   breast height (cm) for all live trees with dbh greater than 20.32cm (8in). Biomass

430   calculations (mean biomass: Mg ha^{-1}) used the same set of allometric regression equations

431   as for the PLSS data [68]. All calculations followed instructions in Woudenberg *et al* [70]

432   using forested plots only (COND_STATUS_CD 1).

433   One critical issue is the reliance on forested condition for the FIA sampling. This reduces

434   our capacity to compare forest state between PLS and FIA cover in regions with historical

435   prairie and savanna coverage. In addition, it may result in the overestimation of modern

436   density, basal area and biomass at the mesoscale in these same regions by drawing from a

437   sample biased specifically towards regions with > 10% forest cover [70], however, the 10%

438   cover theshold is fairly low, but more likely in line with "open forest" [69] than savanna.

439   **Gridding and Analysing PLSS and FIA Data**

440   Maps of stem density, basal area and biomass were generated by averaging all PLSS point

441   or FIA plot estimates within a 64km² raster cell. Differences in sampling design between

442   PLSS and FIA data combined with spatially structured forest heterogeneity will affect the

443   partitioning of within-cell versus between-cell variance, but not the expected estimates.

444   Most 64km² cells have one or a few intensively sampled FIA plots. Therefore at this scale of

445   aggregation, the low density of FIA plots in heterogeneous forests could result in high

446   within-cell variance and high between-cell variability. For the PLSS plotless (point based)

447   estimates, stem density estimates are sensitive to trees close to the plot center. Point-level

448   estimates with very high stem densities can skew the rasterized values, and it is difficult to

449   distinguish artifacts from locations truly characterized by high densities. To accommodate

450   points with exceptionally high densities we carry all values through the analysis, but

451   exclude the top 2.5 percentile when reporting means and standard deviations in our

452   analysis. PLS-based estimates are highly variable from point to point due to the small

453   sample size, but have low variance among 64 km^2 raster cells due to the uniform sampling

454   pattern of the data. Thus within-cell variance is expected to be high for the PLSS point data,

455   but spatial patterns are expected to be robust at the cell level. The base raster and all

456   rasterized data are available as Supplement 3.

457   Standard statistical analysis of the gridded data, including correlations, paired t-tests and

458   regression, was carried out in R [61], and is documented in supplementary material that

459   includes a subset of the raw data to allow reproducibility. Analysis and presentation uses

460   elements from the following R packages: `cluster` [71], `ggplot2` [72,73], `gridExtra` [74],

461   `igraph` [75], `mgcv` [76], `plyr` [77], `raster` [78], `reshape2` [79], `rgdal` [62], `rgeos` [80], `sp`

462   [81,82], and `spdep` [83].

463   We identify analogs and examine differences in composition between and within PLSS and

464   FIA datasets using Bray-Curtis dissimilarity [84] for proportional composition within

465   raster cells using basal area measurements. For the analog analysis we are interested only

466   in the minimum compositional distance between a focal cell and its nearest compositional

467   (not spatial) neighbor. The distribution of compositional dissimilarities within datasets

468   indicates forest heterogeneity within each time period, while the search for closest analogs

469   between datasets indicates whether contemporary forests lack analogs in pre-settlement

470   forests ('novel forests'), or vice versa ('lost forests'). For the analog analyses, we compute

471   Bray-Curtis distance between each 64km^2 cell in either the FIA or the PLSS periods to all

472   other cells within the other dataset (FIA to FIA, PLSS to PLSS), and between datasets (PLSS

473   to FIA and FIA to PLS), retaining only the minimum. For within era analyses (FIA - FIA and

474   PLSS - PLSS), cells were not allowed to match to themselves. We define vegetation classes

475   for lost and novel forests using k-medoid clustering [71].

476   The differences in sampling design and scale between the PLSS and FIA datasets, described

477   above, potentially affect between-era assessments of compositional similarity [47]. The

478   effects of differences in scale should be strongest in regions where there are few FIA plots

479   per 64 km^2 cell, or where within-cell heterogeneity is high. For the analog analyses, this

480   effect should increase the compositional differences between the FIA and PLSS datasets.

481   We test for the importance of this effect on our analog analyses via a sensitivity analysis in

482   which we test whether dissimilarities between FIA and PLSS grid cells are affected by the

483   number of PLSS plots per cell. We find a small effect (see below), suggesting that our

484   analyses are mainly sensitive to the compositional and structural processes operating on

485   large spatial scales.

486   To understand the extent to which the processes governing novelty operate at landscape

487   scales, we relate the novelty of a cell to the spatial distance between individual novel cells

488   and the nearest 'remnant' forest cell, *i.e.*, how far away can you go from a remnant forest

489   cell before all cells are predicted to be novel. We examine whether this relationship varies

490   between forest types, and whether it is different than the relationship we might see if the

491   dissiminlarity values were distributed randomly on the landscape. The definition of

492   "remnant" forest is likely to be arbitrary and, possibly, contentious. We use a threshold, the

493   lowest 25%ile of compositional dissimilarity within the PLSS data, as our cutoff. This

494   means that all FIA cells with nearest neighbor dissimilarities to the PLSS era forests below

495   this cutoff are considered to be representative of the PLSS era forests. The analysis

496   presented below is robust to higher cutoffs for the remnant forest threshold.

497   We use a generalized linear model with a binomial family to relate novelty (as a binomial,

498   either novel or not) to the spatial distance from the nearest 'remnant' cell for each of the

499   five major forest types within the PLSS data (Oak savanna, Oak/Poplar/Basswood/Maple,

500   Pine, Hemlock/Cedar/Birch/Maple and Tamarack/Pine/Spruce/Poplar forests). Because

501   the geographic extent of this region is complex, with islands, peninsulas and political

502   boundaries, we use permutation, resampling the FIA to PLSS nearest neighbor distances

503   without replacement, to estimate the expected distance to novelty if FIA to PLSS nearest

504   neighbor dissimilarities were distributed randomly on the landscape.

505   We expect that a weak relationship will indicate that novelty, following landscape-scale

506   land use change, is moderated by a species pool culled from small remnant patches,

507   individual specimens, or local scale restoration efforts. A significant relationship between

508   distance from remant forest and novelty indicates that small patches have been insufficient

509   to restore natural forest cover within the region, and would indicate that greater efforts are

510   needed to restore landscapes at regional scales.

511   All datasets and analytic codes presented here are publicly available and open source at

512   (http://github.com/PalEON-Project/WitnessTrees), with the goal of enabling further

513   analyses of ecological patterns across the region and the effects of post-settlement land use

514   on forest composition and structure. Data are also archived at the Long Term Ecological

515   Research Network Data Portal (https://portal.lternet.edu/nis/home.jsp).

Results:

Data Standardization

518   The original PLSS dataset contains 490,818 corner points (excluding line and meander

519   points), with 166,607 points from Wisconsin, 231,083 points from Minnesota and 93,095

520   points from Michigan. Standardizing data and accounting for potential outliers, described

521   above, removed approximately 1.5% points from the dataset, yielding a final total of

522   366,993 points with estimates used in our analysis.

523   Rasterizing the PLSS dataset to the Albers 64km^2 grid produces 7,939 raster cells with data.

524   Each cell contains between 1 and 94 corner points, with a mean of 61.8 (σ = 15) and a

525   median of 67 corners (Supplement 3). Cells with a low number of points were mainly near

526   water bodies or along political boundaries such as the Canadian/Minnesota border, or

527   southern Minnesota and Wisconsin borders. Only 2.44% of cells have fewer than 10 points

528   per cell.

529   Species assignments to genera were rarely problematic. Only 18 PLSS trees were assigned

530   to the Unknown Tree category, representing less than 0.01% of all points. These unknown

531   trees largely consisted of corner trees for which taxon could not be interpreted, but for

532   which diameter and azimuth data was recorded. A further 0.011% of trees were assigned

533   to the "Other hardwood" taxon (*e.g.*, hawthorn, "may cherry", and "white thorn").

534   For maple the class has very high within-genera specificity for a number of assignments. A

535   total of 78478 trees are assigned to "Maple". Of these, surveyors do use common names

536   that can be ascribed to the species level (e.g., *A. saccharum*, n = 56331), but a large number

537   of the remaining assignments are above the species level (n = 21356). This lack of

538   specificity for a large number of records causes challenges in using the species level data. A

539   similar pattern is found for pine, where many individual trees (125639) can be identified to

540   the level of species (*P. strobus*, n = 41673; *P. banksiana*, n = 28784; *P. resinosa*, n = 28766),

541   but there remains a large class of pine identified only at the genus level, or with unclear

542   assignment (n = 17606).

543   For ash the data includes both surveyor attributions to "brown ash" (presumably a

544   colloquial term used by surveyors as this is not currently an accepted common name in the

545   region) and black ash (n=9312), and white ash (n = 2350), but again, also includes a large

546   class of ash for which no distinction is made within the genera (n = 7423).

547   These patterns are repeated throughout the data. For spruce this within-genera confusion

548   is even greater, with 50188 assignments to genera-level classes and only 20 to either black

549   or white spruce.

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

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

552   The mean stem density for the region (Fig 2a) is 153 stems ha^{-1}. Stem density exclusive of

553   prairie is 172 stems ha^{-1} and is 216 stems ha^{-1} when both prairie and savanna are excluded.

554   The 95th percentile range is 0 - 423 stems ha^{-1}, and within-cell standard deviations

555   between 0 and 441 stems ha^{-1}. Basal area in the domain (Fig 2c) has a 95th percentile

556   range between 0 and 63.5 m^2 ha^{-1}, a mean of 22.2 m^2 ha^{-1}, within-cell standard deviations

557   range from 0 to 76.7 m^2 ha^{-1}. Biomass ranges from 0 to 209 Mg ha^{-1} (Fig 2d), with cell level

558   standard deviations between 0 and 569 Mg ha^{-1}. High within-cell standard deviations

559   relative to mean values within cells for density, basal area and biomass indicate high levels

560   of heterogeneity within cells, as expected for the PLSS data, given its dispersed sampling

561   design.

562
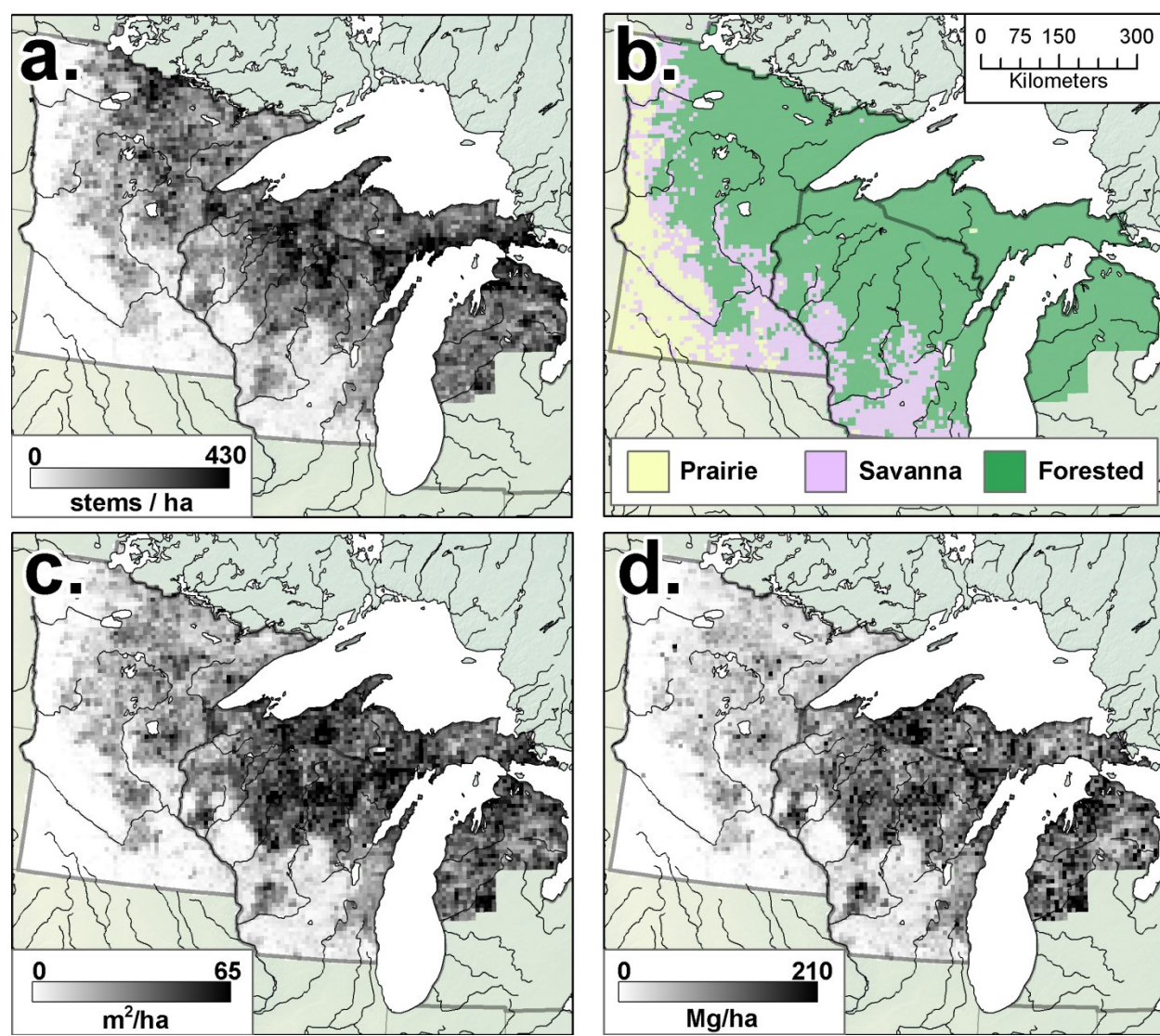
563 **Fig 2.** *Total stem density (a) in the Upper Midwest, along with forest type classification (b)*

564 *based on PLSS data and the stem density thresholds defined by Anderson and Anderson [69];*

565 *Table 2). Fine lines represent major rivers. To a first order, basal area (c) and biomass (d)*

566 *show similar patterns to stem density (but see* Fig 3*).*

567 In the PLSS data, stem density is lowest in the western and southwestern portions of the

568 region, regions defined as prairie and savanna (Fig 2b, Table 2). When the Anderson and

569 Anderson [69] stem density thresholds (<47 stems ha^{-1} for Savanna, Table 2) are used, the

570 extent of area classified as savanna is roughly equivalent to prior reconstructions

571 [22,85,86] (Fig 2b). The highest stem densities occur in north-central Minnesota and in

572 north-eastern Wisconsin (Fig 2a), indicating younger forests and/or regions of lower forest

573 productivity.

574 **Table 2**. *Forest classification scheme used in this paper for comparison between pre-*

575 *settlement forests and the Haxeltine and Prentice [87] potential vegetation classes*

576 *represented in Ramankutty and Foley [1]. Plant functional types (PFTs) for grasslands (CG,*

577 *grassland; Non-Tree samples in the PLS), broad leafed deciduous taxa (BDT) and*

578 *needleleaded evergreen taxa (NET) are used, but leaf area index used in Haxeltine and*

579 *Prentice [87] is replaced by stem density classes from Anderson and Anderson [69].*

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	Dominant PFT BDT, LAI > 2.5	Stem dens. > 48 stems ha^{-1}, BDT > 70%

Deciduous		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

580   Forest structure during the settlement era can be understood in part by examining the

581   ratio of stem density to biomass, a measure that incorporates both tree size and stocking.

582   Regions in northern Minnesota and northwestern Wisconsin have low biomass and high

583   stem densities (Fig 3, blue). This indicates the presence of young, small-diameter, even-

584   aged stands, possibly due to frequent stand-replacing fire disturbance in the pre-

585   EuroAmerican period or to poor edaphic conditions. Fire-originated vegetation is

586   supported by co-location with fire-prone landscapes in Wisconsin [88]. High-density, low-

587   biomass regions also have shallower soils, colder climate, and resulting lower productivity.

588   High-biomass values relative to stem density (Fig 3, red) are found in Michigan and

589   southern Wisconsin. These regions have higher proportions of deciduous species, with

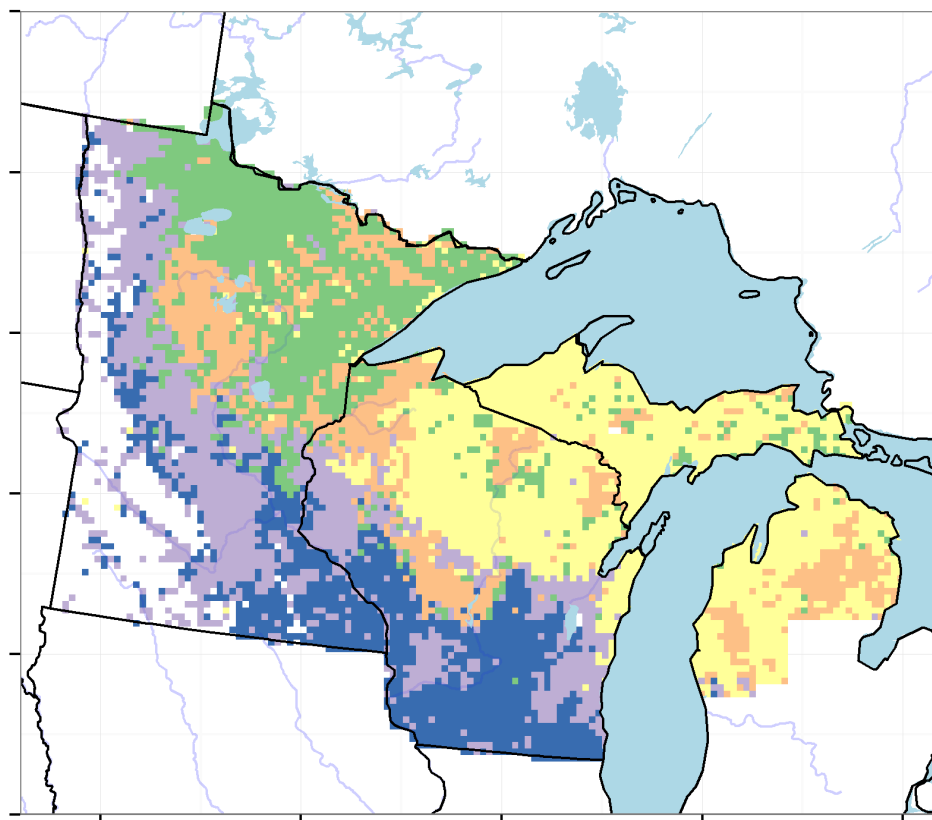590   higher tree diameters than in northern Minnesota.

591

Fig 3. *The major forest types in the pre-settlement Upper Midwest. Five clusters are shown*

using k-medoid clustering. These clusters represent (b) the ratio between biomass and stem

density as an indicator of forest structure. Regions with high stem density to biomass ratios

(blue) indicate dense stands of smaller trees, while regions with low stem density to biomass

ratios (red) indicate larger trees with wider spacings.

Taxon composition within settlement-era forests is spatially structured along dominant

gradients from south to north (deciduous dominated to conifer dominated forests) and

from east to west (mixed wood forests to open prairie) (Fig 4). Oak is dominant in the

south of the region, with an average composition of 21%, however, that proportion drops

to 8% when only forested cells (cells with stem density > 48 stems/ha) are considered, due

602   to its prevalence as a monotypic dominant in the savanna and prairie. Pine shows the

603   opposite trend, with average composition of 14% and 17% in unforested and forested cells

604   respectively. Pine distributions represent three dominant taxa, *Pinus strobus*, *Pinus resinosa*

605   and *Pinus banksiana*. These three species have overlapping but ecologically dissimilar

606   distributions, occuring in close proximity in some regions, such as central Wisconsin, and

607   are typically associated with sandy soils with low water availability. Other taxa with high

608   average composition in forested cells include maple (10%), birch (10%), tamarack (9%)

609   and hemlock (8%).

610
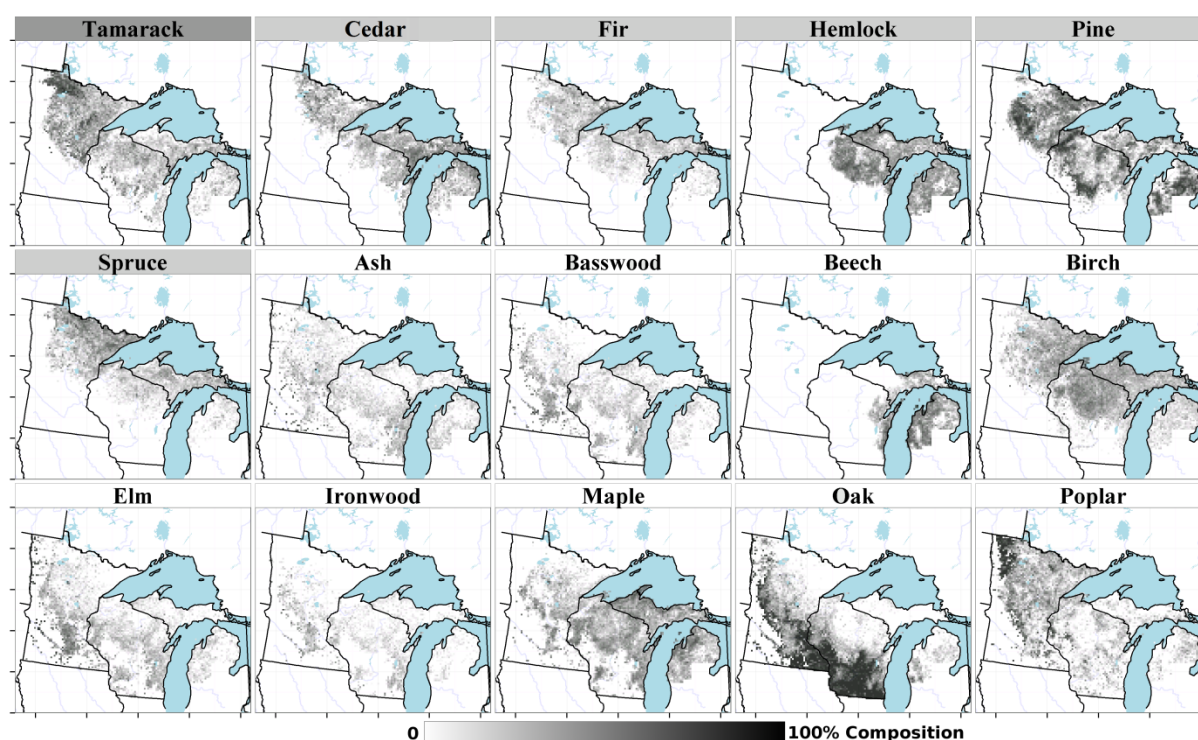
611   **Fig 4**. *Forest composition (%) for the 15 most abundant tree taxa. The scale is drawn using a*

612   *square-root transform to emphasize low abundances. Shading of the bar above individual*

613   *taxon maps indicates plant functional type assignments (dark gray: needleleafed deciduous;*

614   *light gray: needleleafed evergreen; white: broadleafed deciduous).*

615   For a number of taxa, proportions are linked to the total basal area within the cell. For 4

616   taxa - hemlock, birch, maple and cedar - taxon proportions are positively related to total

617   basal area. For 17 taxa including oak, ironwood, poplar, tamarack and elm, high

618   proportions are strongly associated with lower basal areas (Figures 3 and 5). This suggests

619   that hemlock, birch, maple and cedar occurred in well-stocked forests, with higher average

620   dbh. These taxa are most common in Michigan and in upper Wisconsin. Taxa with negative

621   relationships to total basal area (*e.g.*, spruce and tamarack) are more common in the

622   northwestern part of the domain.

623   Spruce in the PLSS represents two species (*Picea glauca*, *Picea mariana*) with overlapping

624   distributions, but complex site preferences that vary in space. *P. glauca* is generally

625   associated with dry upland to wet-mesic sites, while *P. mariana* is associated with hydric

626   sites, but *P. mariana* also frequently occupies upland sites in northern Minnesota. Both

627   cedar (*Thuja occidentalis*) and fir (*Abies balsamea*) are mono-specific genera in this region.

628   Northern hardwoods, such as yellow birch and sugar maple, and beech, are much less

629   common in the lower peninsula of Michigan, and southern Wisconsin, except along Lake

630   Michigan. Birch has extensive cover in the north, likely reflecting high pre-settlement

631   proportions of yellow birch (*Betula alleghaniensis*) on mesic soils, and paper birch on sandy

632   fire-prone soils and in northern Minnesota (birch proportions reach upwards of 34.1% in

633   northeastern Minnesota). Hardwoods in the southwest, such as oak, elm, ironwood and

634   basswood, are most typically mono-specific groupings, with the exception of oak, which

635   comprises 7 species (see Supplement 2). Hardwoods in the southwest are located primarily

636   along the savanna and southern forest margins, or in the southern temperate deciduous

637   forests. Finally, maple and poplar (aspen) have a broad regional distribution, occupying

638   nearly the entire wooded domain. Poplar comprises four species in the region, while maple

639   comprises five species (Supplement 2). Both hardwood classes, those limited to the

640   southern portions of the region, and those with distributions across the domain,

641   correspond to well-defined vegetation patterns for the region [85].

642   These individual species distributions result in a mosaic of forest classes accross the region

643   (Fig 5). The dominant class is the Hemlock/Cedar/Birch/Maple assemblage in northern

644   Wisconsin, and upper Michigan (Fig 5, yellow). This mixedwood assemblage is interspersed

645   by both Pine dominated landscapes (Fig 5, orange) and, to a lesser degree, the softwood

646   assemblage Tamarack/Pine/Spruce/Poplar (Fig 5, green), which dominates in north-

647   eastern Minnesota. The softwood assemblage is itself interspersed with Pine dominated

648   landscapes, and grades into a mixed-hardwood assemblage of

649   Oak/Poplar/Basswood/Maple (Fig 5, light purple) to the west. Thismixed- softwood forest

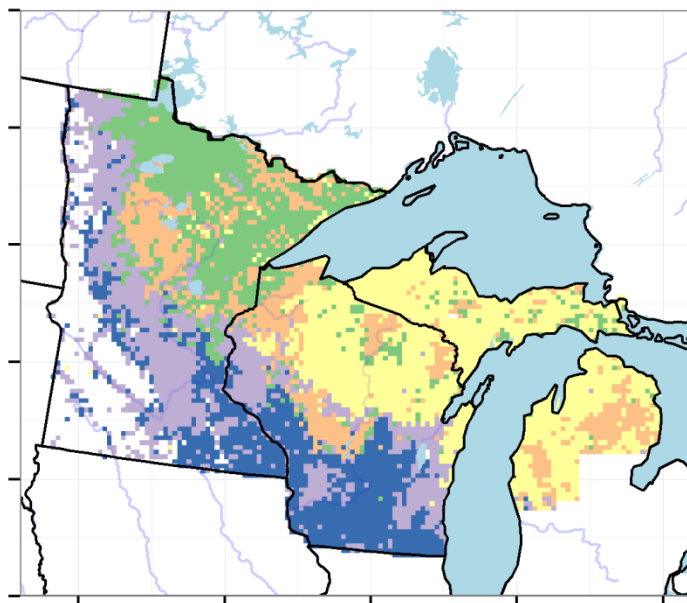650   assemblage grades south into mono-specific Oak savanna (Fig 5, dark blue).

651

Fig 5. *The five dominant forest types in the Upper Midwest as defined by k-medoid clustering.*

Forest types (from largest to smallest) include Hemlock/Cedar/Birch/Maple (yellow),

Oak/Poplar/Basswood/Maple (light purple), Tamarack/Pine/Spruce/Poplar (light green),

Oak Savanna (dark purple) and Pine (orange). These forest types represent meso-scale

(64km²) forest associations, rather than local-scale associations.

The broad distributions of most plant functional types results in patterns within individual

PFTs that are dissimilar to the forest cover classes (Fig 5). Thus overlap among PFT

distributions (Fig 6) emerges from the changing composition within the plant functional

type from deciduous broadleaved species associated with the southern, deciduous

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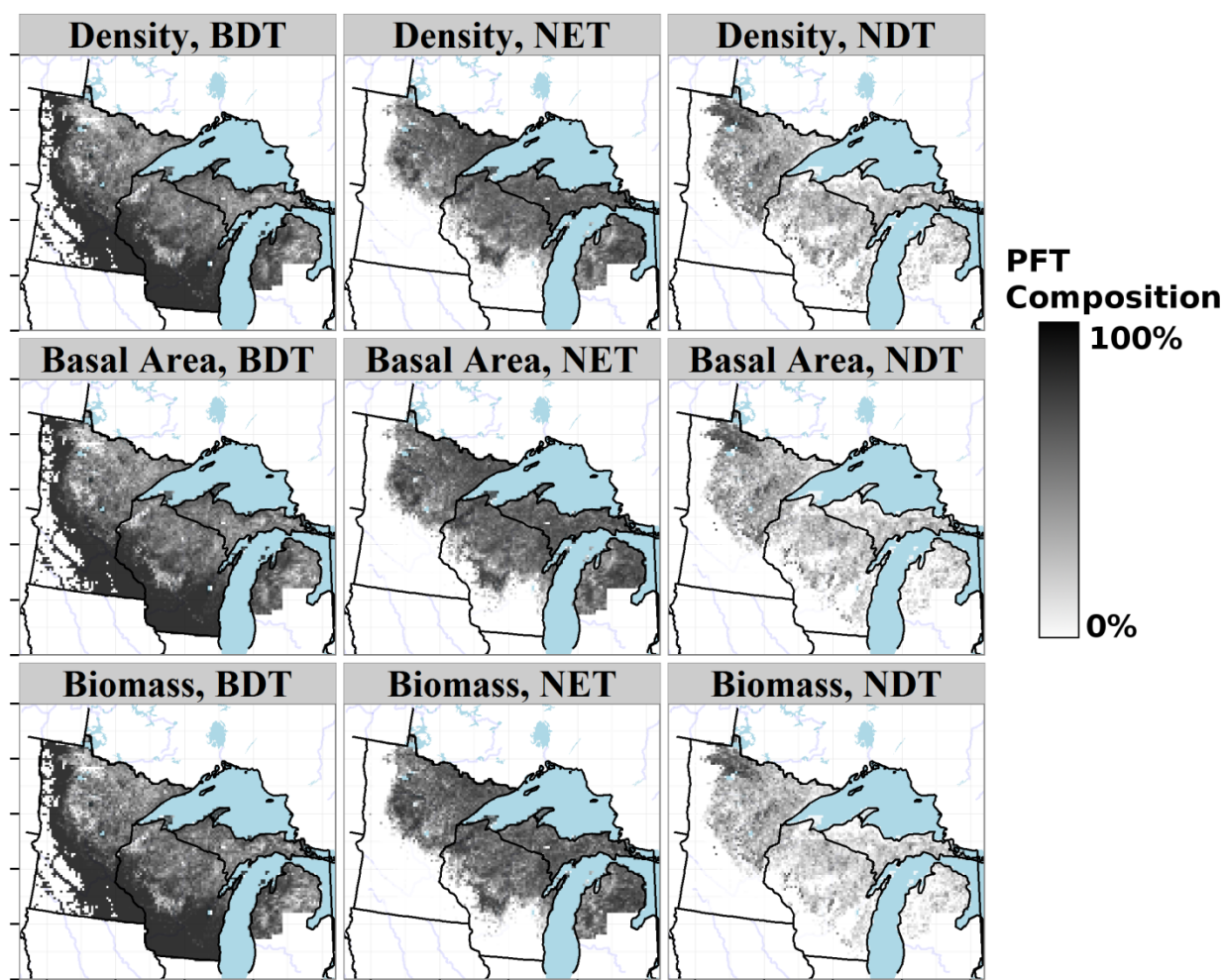
in the upper Midwest.

Fig 6. *Proportional distribution of Plant Functional Types (PFTs) in the upper Midwest from PLSS data, for broadleaved deciduous trees (BDT), needleleaved deciduous trees (NDT), and needleleaved evergreen trees (NET). Distributions are shown as proportions relative to total basal area, total biomass, and composition (Fig 2). The grassland PFT is mapped onto non-tree cells with the assumption that if trees were available surveyors would have sampled them.*

Structural Changes Between PLSS and FIA Forests

671   Differences between PLSS and FIA data shows strong spatial patterns, but overall estimates

672   can be examined. By cell, modern forests (FIA) have higher stem densities (146 stems ha^{-1},

673      $t_{1,5177}$ = 51.8, p < 0.01) than PLSS forests, but slightly lower basal areas (-4.5 m^2 ha^{-1}, $t_{1,5177}$

674   = -16.4, p < 0.01) and lower biomass (-8.72 Mg ha^{-1}, $t_{1,5177}$ = -6.55, p < 0.01) (Fig 7). We use

675   only point pairs where both FIA and PLSS data occur since non-forested regions are

676   excluded from the FIA and as such cannot be directly compared with PLS estimates. The

677   similarity in biomass despite lower stem density and total basal area in the PLSS data is

678   surprising. Two likely factors are shifts in allometric scaling associated with changes in

679   species composition, or a higher mean diameter of PLSS trees (Fig 7d). Total biomass was

680   45,080 Mg higher in the PLSS when summed across all cells coincident between the FIA

681   and PLSS.

682

683 **Table 3.** *Mean cell-wise change in forest zone density, basal area and biomass since PLSS for*

684 *cells with coverage in both PLSS and FIA eras by forest class. All forest zones show increases in*

685 *stem density since the PLSS era (positive values). All zones but Oak Savanna most show*

686 *declines in mean basal area since the PLSS era, while modern biomass is lower in the FIA-era*

687 *for the both Hemlock/Cedar/Birch/Maple and Pine forest zones, but higher in the remaining*

688 *three zones.*

Forest Type	Number of Cells	Stem Density (stems/ha)	Basal Area (m^2/ha)	Biomass (Mg/ha)
Hemlock/Cedar/Birch/Maple	1780	170.5	-13.8	-56.7
Tamarack/Pine/Spruce/Poplar	1105	76.1	-4.2	4.7
Pine	966	191.4	-1.8	-5.2
Oak/Poplar/Basswood/Maple	708	108.4	-0.2	24.8
Oak Savanna	577	182.6	13.1	62.2

689

690 Every one of the five historical PLSS zones shows an increase in stem density (Table 3). The

691 two forest types bordering the prairie, Oak Savanna and Oak/Poplar/Basswood/Maple

692 both show increases in density that likely reflect, in part, the issues addressed earlier with

693 regards to the sampling of forested plots in the FIA (over 10% cover). Density in the Oak

694 Savanna increases from a mean 27 stems/ha to 217 stems/ha, with a mean biomass

695 increase of 62 Mg ha^{-1} per cell (Table 3), the highest of any of the zones. The

696 Oak/Poplar/Basswood/Maple forest had higher PLSS-era densities (90 stems/ha)

697 reflecting open forest status rather than savanna (Table 2), but also shows a large increase

698   in estimated FIA-era stem density (to 218 stems/ha) but with a much lower increase in

699   biomass than the Oak Savanna, and a negligable increase in basal area (Table 3). The

700   largest forest zone, Hemlock/Cedar/Birch/Maple shows the largest decline in biomass (a

701   net loss of 56.7 MG ha[01] since the PLSS-era) and basal area (net loss of 13 m^2 ha^{-1} since the

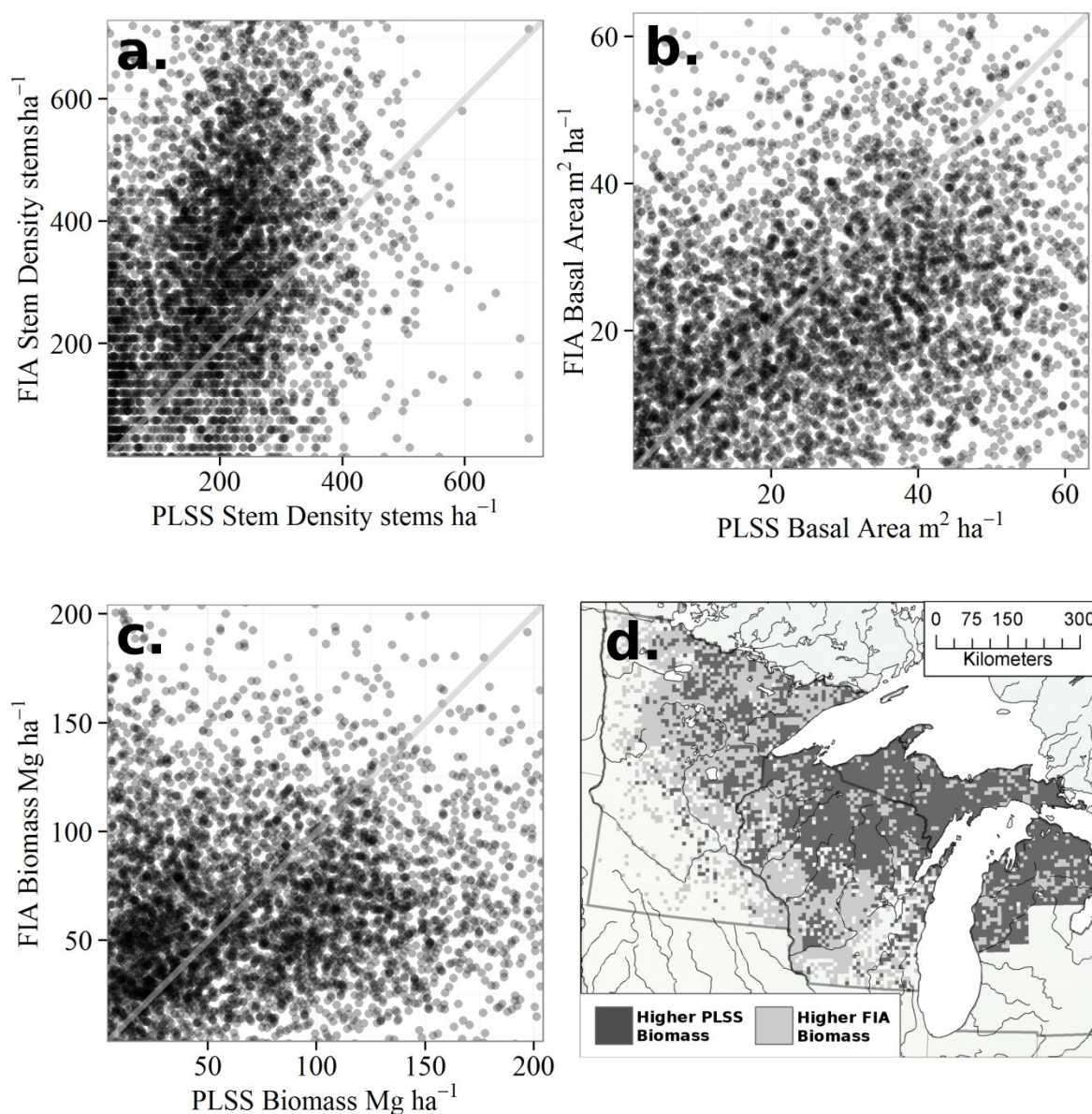702   PLSS-era), but with an average increase in FIA era stem density.

703

704   **Fig 7**. *The relationship between (a) average stem density, (b) total basal area and (c) biomass*

705   *values in the PLSS and FIA datasets. Stem density and total basal area are higher in the FIA*

706   *than in the PLS, however average cell biomass is higher in the PLSS. A 1:1 line has been added*

707   *to panels a-c to indicate equality.*

708   The PLSS has a lower overall mean diameter than the FIA (δ_{diam} = -2.9 cm, 95%CI from -

709   17.3 to 8.18cm). FIA diameters are higher than PLSS diameters in the northwestern parts

710   of the domain (on average 6.47 cm higher), overlapping almost exactly with regions where

711   we have shown low biomass-high density stands (Fig 3). At the same time, regions with

712   high biomass and low density stands, in northeastern Wisconsin, and the Upper and Lower

713   Peninsulas of Michigan, had higher average diameters during the PLSS than in the FIA, on

714   average 3.65 cm higher. Thus we are seeing an overal increase in tree size in the sub-boreal

715   region and a decrease in temperate mixedwood forests, where we find tree species with

716   much higher dbh:biomass ratios [68]. This is coupled with declining variance in dbh across

717   the domain (from within cell variance of 37.9 cm the PLSS to 30.7 cm in the FIA). Thus, the

718   mechanism by which low density and basal area produce roughly equivalent biomass

719   estimates between the FIA and PLSS is likely due to the differential impacts of land

720   clearence and subesequent forest management in the south east vs the northwest. The loss

721   of high biomass southern hardwood forests is balanced by higher biomass in the northeast

722   due to fire suppression and regeneration of hardwoods in the northwest. Declining

723   diameters from the PLSS to FIA are most strongly associated with higher abundances of

724   poplar, ironwood and oak, while declining diameters are associated with maple and

725   hemlock, further supporting the assertion that much of the loss in diameter, and,

726   subsequently biomass, is occuring in southeastern mixedwood/hardwood forest, while

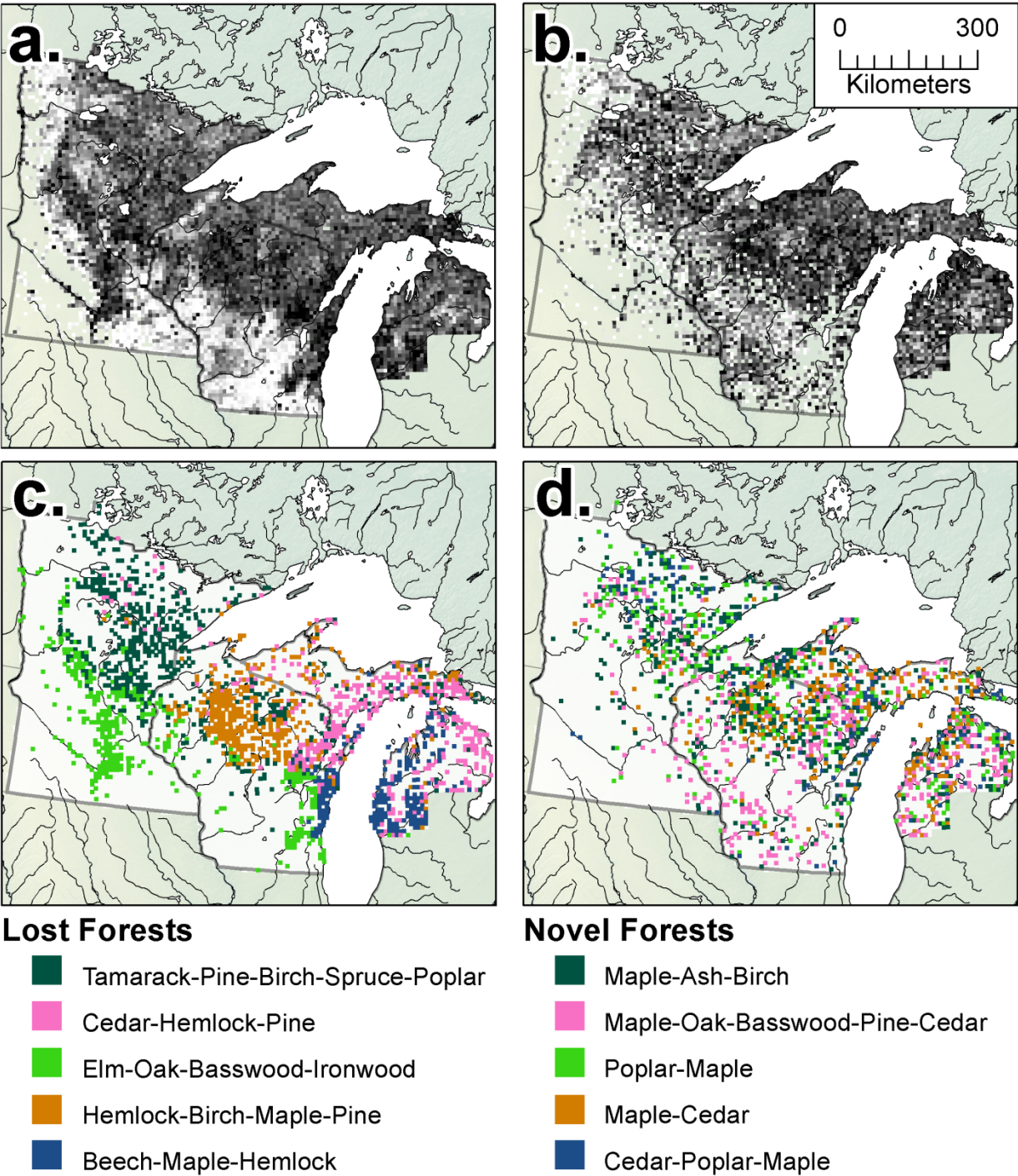727   diameter and biomass increases are occuring in the northwest.

728   Differences between FIA and PLSS data in sampling design are unlikely to be a factor for

729   most measures (see below); these differences are expected to affect how these datasets

730   sample local- to landscape-scale heterogeneity, but should not affect the overall trends

731   between datasets. Differences in variability introduce noise into the relationship, but given

732   the large number of samples used here, the trends should be robust.

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

734   Both the PLS- and FIA-era compositional data show similar patterns of within-dataset

735   dissimilarity, with the highest dissimilarities found in central Minnesota and northwestern

736   Wisconsin. High within-PLSS dissimilarities are associated with high proportions of maple,

737   birch and fir while high within-FIA dissimilarities are associated with high proportions of

738   hemlock, cedar and fir. Dissimilarity values in the FIA dataset are less spatially structured

739   than in the PLSS. Moran's I for dissimilarities within the FIA (I_{FIA} = 0.198, p < 0.001) are

740   lower than the dissimilarities within the PLSS (I_{PLSS} = 0.496, p < 0.001), suggesting lower

741   spatial autocorrelation in the FIA dataset. Cells with identical pairs represent 5.6% of the

742   PLSS cells and 7.44% of FIA cells. Identical cells in the PLSS are largely located along the

743   southern margin and most (69.5%) are composed entirely of oak. Cells in the FIA with

744   identical neighbors are composed of either pure oak (19.4%), pure poplar (26%) or pure

745   ash (14%).

746   There is a small but significant positive relationship ($F_{1,5964}$= 920, p < 0.001) between the

747   number of FIA plots and within-FIA dissimilarity. The relationship accounts for 13% of

748   total variance and estimates an increase of δ_d = 0.0134 for every FIA plot within a cell. This

749   increase represents only 3.08% of the total range of dissimilarity values for the FIA data.

750   There is a gradient of species richness that is co-linear with the number of FIA plots within

751   a cell, where plot number increases from open forest in the south-west to closed canopy,

752   mixed forest in the Upper Peninsula of Michigan. Hence, differences in within- and

753   between-cell variability between the PLSS and FIA datasets seem to have only a minor

754   effect on these regional-scale dissimilarity analyses.

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

756   for dissimilarities within a particular dataset. In the PLSS dataset, forests that have no

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

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

759   FIA dataset ('lost forests'; PLS-FIA dissimilarity, Fig 8c), while 29% of FIA sites have no

760   analog in the PLSS data ('novel forests'; FIA-PLSS dissimilarity, Fig 8d). Lost forests show

761   strong spatial coherence, centered on the "Tension Zone" [85], the ecotone between

762   deciduous forests and hemlock-dominated mixed forest (Fig 4).

Fig 8. *Minimum dissimilarity maps. Distributions of minimum (within dataset) dissimilarities*

during the PLSS (a) and FIA (b) show spatially structured patterns of dissimilarity, with

stronger spatial coherence for the PLS. Lost forests (c) show strong compositional and spatial

767   *coherence, and have more taxa with percent composition > 10% than within Novel forests*

768   *during the FIA era (d).*

769   Lost forests are drawn from across the domain, and show strong ecological and spatial

770   coherence (Fig 8c). Forest classes generally fall into five classes: Tamarack-Pine-Birch-

771   Spruce-Poplar accounts for 28.8% of all lost forests and 7.97% of the total region. This

772   forest type is largely found in north eastern Minnesota, extending southward to central

773   Minnesota, into Wisconsin and along the Upper Peninsula of Michigan, as well as in

774   scattered locations on the Lower Peninsula of Michigan (Fig 8c). This forest likely

775   represents a mesic to hydric forest assemblage, particularly further eastward. Modern

776   forests spatially overlapping this lost type are largely composed of poplar ($\overline{x}_{FIA}$ = 12%) and

777   oak ($\overline{x}_{FIA}$ = 12%). Tamarack in these forests has declined significantly, from 23% to only

778   5% in the FIA, while Poplar has increased from 10% to 22%, resulting in forests that look

779   less mesic and more like early seral forests.

780   Cedar/juniper-Hemlock-Pine accounts for 19.8% of all lost forests and 5.49% of the total

781   region. This forest type is found largely in northeastern Wisconsin, and the Upper and

782   Lower Peninsulas of Michigan. This lost forest type has been predominantly replaced by

783   maple, poplar, and pine, retaining relatively high levels of cedar ($\overline{x}_{PLS}$ = 19%; $\overline{x}_{FIA}$ = 14%).

784   The loss of hemlock is widespread across the region, but particularly within this forest

785   type, declining to only 3% from a pre-settlement average of 18%.

786   Elm-Oak-Basswood-Ironwood accounts for 19.6% of all lost forests and 5.42% of the total

787   region. The region is centered largely within savanna and prairie-forest margins, both in

788   south-central Minnesota and in eastern Wisconsin, but, is largely absent from savanna in

789   the Driftless area of southwestern Wisconsin. These forests were historically elm

790   dominated ($\overline{x}_{PLS}$ = 25%), not oak dominated savanna, as elsewhere (particularly in the

791   Driftless). Modern forests replacing these stands are dominated by oak and ash, with

792   strong components of maple, and basswood. Elm has declined strongly in modern forests

793   ($\overline{x}_{FIA}$ = 1%), possibly in part due to Dutch Elm Disease and land use. The increase in ash in

794   these forests is substantial, from $\overline{x}_{PLS}$ = 5% to $\overline{x}_{FIA}$ = 15%.

795   Hemlock-Birch-Maple-Pine accounts for 19.2% of all lost forests and 5.33% of the total

796   region. This forest type, dominant in north central Wisconsin, was dominated by hemlock

797   ($\overline{x}_{PLS}$ = 26%) and what was likely late seral yellow birch ($\overline{x}_{PLS}$ = 24%), replaced largely by

798   maple (from $\overline{x}_{PLS}$ = 12% to $\overline{x}_{FIA}$ = 27%). Poplar increases from 1% to 13% in the FIA, again

799   indicating a shift to earlier seral forests in the FIA. Hemlock is almost entirely lost from the

800   forests, declining from 26% to 4% in the FIA.

801   Lastly, Beech-Maple-Hemlock accounts for 12.6% of all lost forests and 3.49% of the total

802   region. This forest type is found exclusively on the central, western shore of Lake Michigan

803   and in the Lower Peninsula, in part due to the limited geographic range of Beech in the

804   PLSS dataset (Fig 4). Beech is almost entirely excluded from the modern forests in this

805   region, declining from $\overline{x}_{PLS}$ = 37% to $\overline{x}_{FIA}$ = 4%. Pine in the region increases from 9% to

806   16%, while maple, the dominant taxa in the modern forests, increases from 16 - 25%.

807   On average lost forests contain higher proportions of ironwood (r = 0.203), beech (r = 0.2),

808   birch (r = 0.189) and hemlock (r = 0.188) than the average PLSS forest, and lower

809   proportions of oak (r = -0.28), poplar (r = -0.145), and pine (r = -0.107).

810   The distribution of novel ecosystems (Fig 8d) is spatially diffuse relative to the lost forest of

811   the PLSS and the forest types tend to have fewer co-dominant taxa. FIA novel forest types

812   also have a more uneven distribution in proportion than the PLSS lost forests. Overall,

813   novel forests are associated with higher proportions of maple (r = 0.02), ash (r = 0.03) and

814   basswood (r = -0.04), although basswood is dominant in only one forest type (Poplar-

815   Cedar/juniper-Maple). Novel forests are associated with lower proportions of oak (r = -

816   0.28), and pine (r = -0.11). This analysis suggests that the loss of particular forest types

817   associated with post-settlement land use was concentrated in mesic deciduous forests and

818   the ecotonal transition between southern and northern hardwood forests, while the gains

819   in novelty were more dispersed, resulting from an overall decline in seral age.

820   By far the largest novel forest type is Maple, which accounts for 37.2% of all novel forests

821   and 2.68% of the total region. As with all novel forest types, this forest type is broadly

822   distributed across the region. This forest type is associated with co-dominant maple ($\overline{x}_{FIA}$ =

823   23%) and ash ($\overline{x}_{FIA}$ = 22%). Hemlock has declined significantly across this forest type, from

824     $\overline{x}_{PLS}$ = 24% to $\overline{x}_{FIA}$ = 4%.

825   Poplar-Cedar/juniper-Maple, accounts for 28.8% of all novel forests and 2.08% of the total

826   region. The broad distributiof these novel forests makes assigning a past forest type more

827   difficult than for the PLSS lost forests, the distribution replaces two classes of past forest,

828   one dominated by oak, in southern Wisconsin and Minnesota, the other by mixed hemlock,

829   beech, birch and cedar forests.

830   Pine-Cedar/juniper-Poplar-Maple forest accounts for 16.3% of all novel forests and 1.17%

831   of the total region. This forest type is again broadly distributed, and is widely distributed

832   across the region, representing a homogenous, early seral forest type, likely associated

833   with more mesic sites. Oak forest accounts for 13.3% of all novel forests and 0.96% of the

834   total region. This grouping again shows a pattern of broad distribution across the region,

835   associated with cedar/juniper percentages near 40%, with smaller components of poplar

836   (14%) and maple (13%).

837   ## Spatial Correlates of Novelty

838   Modern compositional dissimilarity from the PLSS data is related to distance from

839   'remnant' forest. The dissimilarity quantile of FIA-PLSS distances increases with increasing

840   distance to remnant cells, and this relationship is robust to higher thresholds for remnant

841   forest classification, up to the 90%ile of within-PLSS near neighbor dissimilarities. Using

842   the 25%ile for within PLSS dissimilarity, approximately 67% of FIA cells can be classed as

843   'remnant' forest. The mean distance to remnant forests for cells with dissimilarities above

844   the 25%ile is 17.7 km, higher than the mean of ~9.6km expected if each 8x8km cell had at

845   least one adjacent 'remnant' cell.

846   **Table 4**. *Spatial distance to novelty - modeled as a binomial - from remnant forests (forests*

847   *within the first 25th percentile of nearest neighbor distances). The null model uses*

848   *permutation (n=100) where quantiles are resampled without replacement.*

Zone	Min	Max	Min (Null)	Max (Null)
Tamarack/Pine/Spruce/Poplar	29	43	11	14
Oak/Poplar/Basswood/Maple	23	33	14	20

Pine	32	56	10	12
Hemlock/Cedar/Birch/Maple	0	undef.	11	13
Oak Savanna	17	23	13	18

849   The GLM shows that distance from remnant forests in the FIA is significantly related to the

850   probability of a cell being novel ($\chi_{1,4}$=623, p < 0.001). The mean distance to novelty varies

851   by PLSS forest type, but is between approximately 20 and 60km for the four forest types

852   examined here (Fig 9), while the null model would predict a distance of 10 - 20km to

853   novelty from remnant cells if dissimilarities were distributed randomly on the landscape

854   (Table 4). Novel forests are generally further from remnant patches than expected in the

855   null model, regardless of forest type, but the distance to novelty is greater for modern

856   forests that are, generally, more similar to their PLSS state (Pine and Tamarack dominated

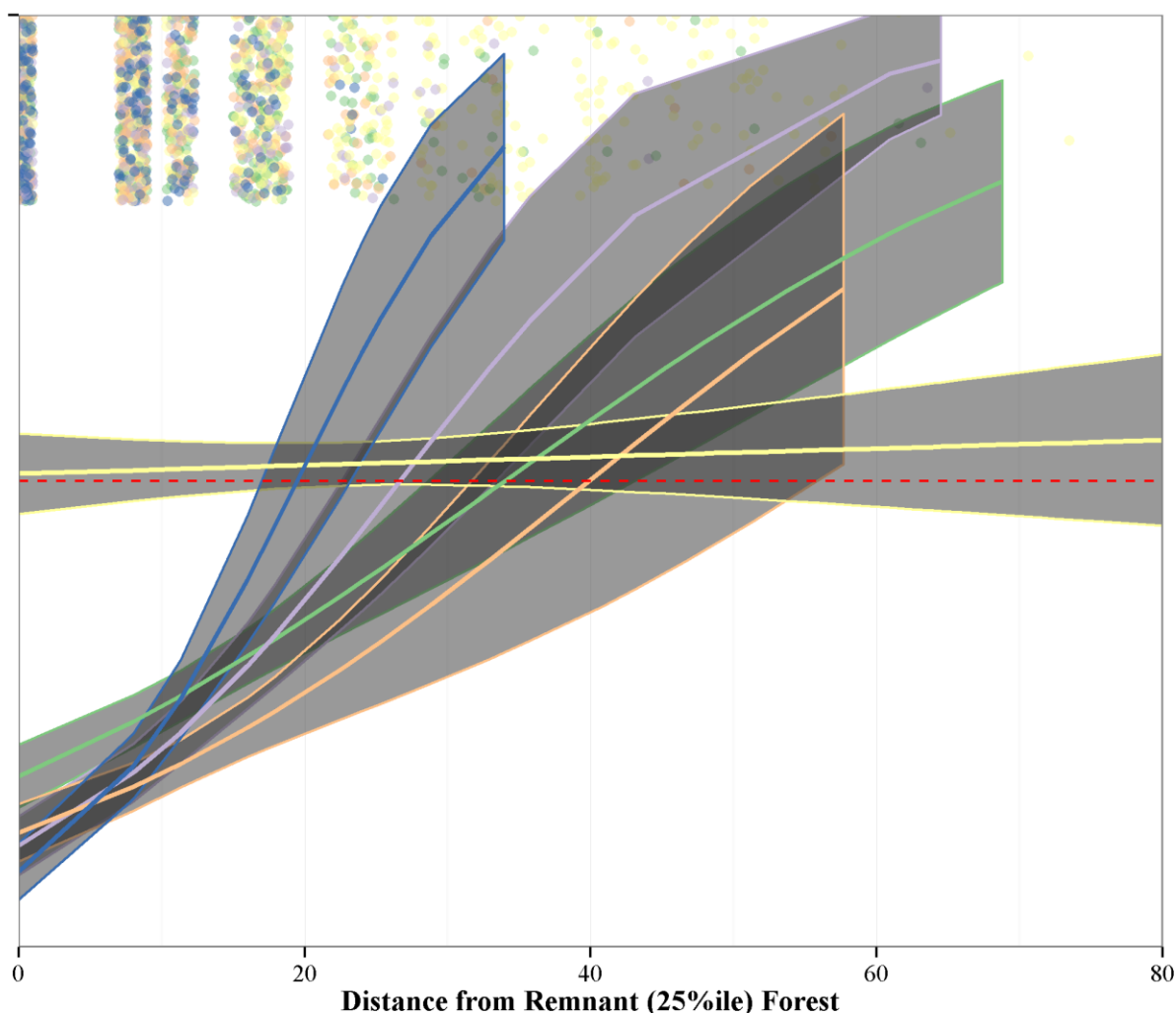857   forests), and closer for forests that are more dissimilar.

858
859 **Fig 9**. *The model relating novelty to spatial distance from remnant forest. Here the 25%ile is*

860 *used to indicate remnant forest, and the 95%ile is defined as novelty. We use a binomial*

861 *regression to predict novelty, the red dashed line indicates a response greater than 0.5. The*

862 *curves represent the relationship between spatial distance and compositional dissimilarity for*

863 *each of the five major historic forest types (Fig 5) defined here as Oak Savanna (blue),*

864 *Oak/Poplar/Basswood/Maple (light purple), Tamarack/Pine/Spruce/Poplar (green),*

865 *Hemlock/Cedar/Birch/Maple (yellow) and Pine (orange).*

866   Critically, we see that the Hemlock/Cedar/Birch/Maple forest class (Fig 5 & 10b, yellow),

867   appearing as a flat line, predicts novelty continuously, from distance 0. This is due, in part,

868   to the very small proporion of Hemlock/Cedar/Birch/Maple cells that are considered

869   residual (only 63 of 1780 cells in the Hemlock zone are considered remnant) and the very

870   high proportion of novel cells in the zone (923 of 1780 cells, or 52% of all cells).

871   Oak Savanna is the most similar to its null model, with a confidence interval that overlaps

872   slightly with the null expectation (Table 4). Northern softwood forests

873   (Tamarack/Pine/Spruce/Poplar, Fig 5, light green) reach novelty at between 29 and 43km,

874   northern Oak forests (Oak/Poplar/Basswood/Maple; Fig 5, light purple) reach novelty at

875   23 - 33 km, slightly higher than the 14 - 19km predicted by the null model. Pine forests (Fig

876   5, orange) are three times further than expected by the null, at 32 - 56km (Table 4).

877   **Compositional Changes Between PLSS and FIA Forests: Ecotone Structure**

878   To understand how the ecotonal structure has been transformed by post-settlement land

879   use, we constructed two transects of the FIA and PLSS data (Fig 10a), and fitted GAM

880   models to genus abundances along these transects. Transect One (T1) runs from northern

881   prairie (in northern Minnesota) to southern deciduous savanna in southeastern Wisconsin

882   (left to right in Figures 11c-f), while Transect Two (T2) runs from southern prairie in

883   southwestern Minnesota to northern mixedwood forest in the Upper Peninsula of Michigan

884   (left to right in Figures 11g-j). In general, these transect analyses show: 1) significant

885   differences in ecotonal structure between the present and pre-settlement, and 2) steeper

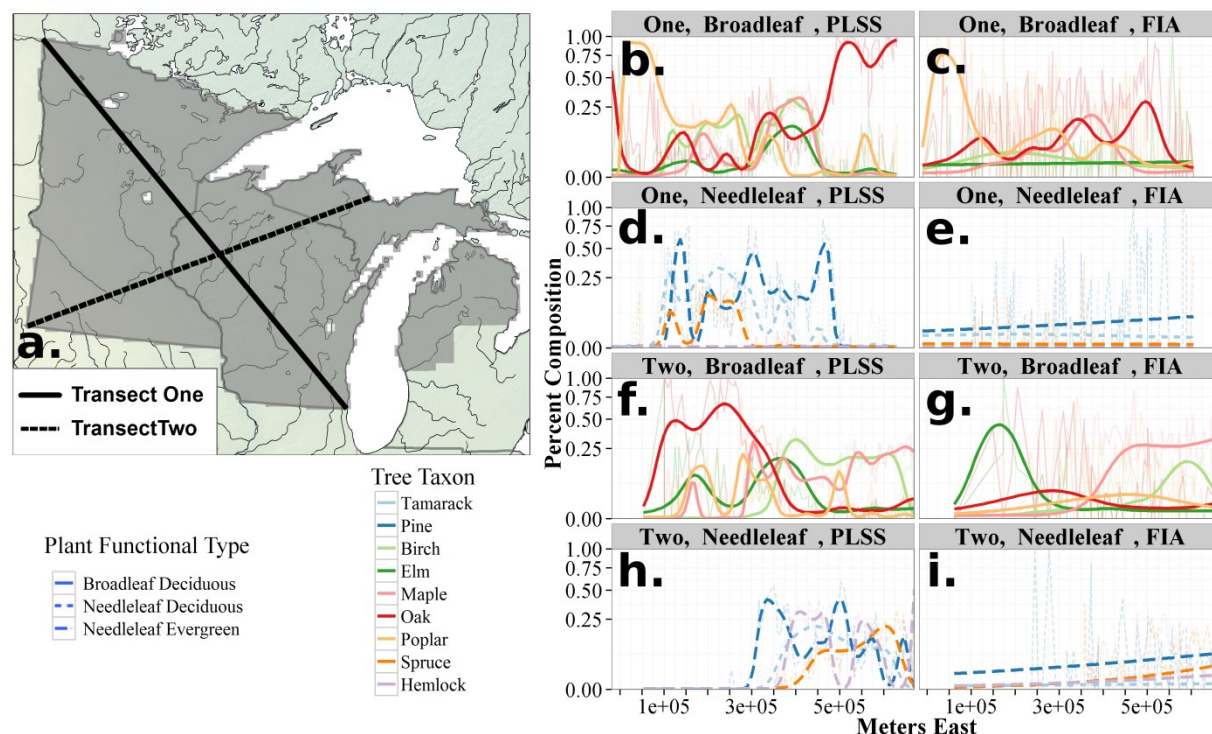886   ecotones in the past and more diffuse ecotones today.

887

Fig 10. *Transects (a) across the region show clear changes in the ecotonal strength. Transect One shows shifts in broad-leafed taxon distributions from the PLSS 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.*

For T1, GAM models show significant differences (using AIC) between time periods in curves for all broadleafed taxa (Fig 10b & c) and for all needleleafed taxa (Figures 10d and e). The PLSS curves show a rapid transition in the northwest from oak to poplar dominated open forest that then transitions to a needleleafed forest composed of pine, spruce and tamarack, with high proportions of tamarack grading to pine further to the south east.

900 Tamarack and poplar proportions decline gradually from the east, being replaced first by

901 pine, then briefly by maple and birch, and then. ultimately by oak as the transect grades

902 into oak savanna. In the FIA dataset oak and poplar in the northwest appears to decline

903 simultaneously, grading into needleleafed forests that are absent from the FIA dataset in

904 the first 100km along the transect. While the PLSS transect shows distinct vegetation types

905 in the central partof the transect, the FIA shows relatively constant proportions of oak,

906 pine, spruce, poplar and maple before pine, oak and elm increase in the southeastern

907 portions of the transect.

908 The second transect shows a similar pattern, with well defined ecotones in the pre-

909 settlement period(Fig 10f and h), that are largely absent from the FIA data (Fig 10g and i).

910 Oak forest, with a component of elm and poplar in the southwest grades slowly to a rapid

911 transition zone where pine, elm, maple (first), then rapidly birch, hemlock and tamarack,

912 and later, spruce, increase. This region, the Tension Zone, extends from 3×10^5 to 4.5×10^5

913 meters East, eventually becoming a forest that shows co-dominance between birch, pine,

914 maple, spruce and tamarack, likely reflecting some local variability as a result of

915 topographic and hydrological factors. Missing data at the beginning of the FIA transect

916 reflects a lack of FIA plots in unforested regions in the west

917 Contemporary forests show broader homogenization and increased heterogeneity

918 (evidenced by the lower within-FIA Moran's I estimates for near-neighbor distances) at a

919 local scale in the region. Homogenization is evident across T1, where Bray-Curtis

920 dissimilarity between adjacent cells declines from the PLSS to the FIA (δ_{beta} = -0.22, t_{113} = -

921 7.93, p<0.001), mirroring declines in the pine barrens between the 1950s and the present

922 [18]. The PLSS shows strong differentiation in the central region of T2 where maple-pine-

923 oak shifts to pine-poplar-birch forest (Fig 10d). This sharp ecotone is not apparent in the

924 FIA data, which shows gradual and blurred changes in species composition across the

925 ecotone (Fig 10i). β-diversity along T2 is lower in the FIA than in the PLSS (δ_{beta} = -0.19,

926 t_{65}=-7.34, p < 0.01), indicating higher heterogeneity in the PLSS data at the 64 km^2 meso-

927 scale.

928 Across the entire domain, β diversity is lower in the FIA than in the PLSS (δ_β = -0.172, $t_{1.3e7}$

929 = 2480, p <0.001), lending support to the hypothesis of overall homogenization. Differences

930 in sampling design between PLSS and FIA data cannot explain this homogenzation, since its

931 effect would have been expected to increase β-diversity along linear transects and at larger

932 spatial scales.

Discussion

934 Many forests of the PLS, are no longer a part of the modern landscape. Forest types have

935 been lost at the 64 km^2 mesoscale, and new forest types have been gained. The joint

936 controls of broad-scale climatic structuring and local hydrology on forest composition and

937 density can be seen in the pre-settlement forests, particularly along the Minnesota River in

938 south-western Minnesota, where a corridor of savanna was sustained in a region mostly

939 occupied by prairie (Fig 2b), but ecotones in the modern forest composition data are

940 weaker now than in the past (Fig 10), with clear signs of increased homogenization at local

941 and regional scales and decreased spatial structure in vegetation assemblages (Fig 8).

942   The loss of ecotones in the upper Midwestern United States suggests that our ability to

943   predict the abiotic controls on species distributions at the landscape scale may be weaker

944   than in the past, reducing the influence of variables such as climate or edaphic factors, and

945   increasing the relative influence of recent land use history. Our results suggest that both

946   recent land use history and historical vegetation cover play a large role in recovery from

947   the large scale disturbance seen following EuroAmerican settlement.

948   Work in eastern North America suggests the utility of including spatial structure in species

949   distribution models to improve predictive ability [89]. The spatial random effects may

950   improve models by capturing missing covariates within SDMs [89], but if recent land use

951   history has strongly shaped species distributions, or co-occurence, then the spatial effect is

952   likely to be non-stationary at longer temporal scales. Given the implicit assumption of

953   stationarity in many ecological models [21], the need for longer time-scale observations, or

954   multiple baselines from which to build our distributional models becomes critical if we are

955   to avoid conflating recent land use effects with the long term ecological processes

956   structuring the landscape.

957   Decreased β diversity along regional transects indicates homogenization at meso-scales of

958   100s of km^2, while the overall reduction in Moran's I for dissimilarity in the FIA indicates a

959   regional reduction in heterogeneity on the scale of 1000s of km^2. The selective loss or

960   weakening of major vegetation ecotones, particularly in central Wisconsin, and the

961   development of novel species assemblages across the region further suggests that modern

962   correlational studies, examining regional relationships between species and climate (for

963   example) may fail to capture the full range of edaphic controls on spcies distributions.

964   These changes are the result of land use, both agricultural and logging, but affect forests in

965   contrasting ways across the domain. Maple has become one of the most dominant taxa

966   across the region, while in northern Minnesota, forest biomass has increased and species

967   shifts have reflected increases in poplar and pine, while in central and eastern Wisconsin,

968   biomass has declined, and hemlock has been lost almost completely.

969   Anthropogenic shifts in forest composition over decades and centuries seen here and

970   elsewhere [2,48] are embedded within a set of interacting systems that operate on multiple

971   scales of space and time [90]. Combining regional historical baselines, long term ecological

972   studies and high frequency analyses can reveal complex responses to climate change at

973   local and regional scales [91]. Estimates of pre-settlement forest composition and structure

974   are critical to understanding the processes that govern forest dynamics because they

975   represent a snapshot of the landscape prior to major EuroAmerican land-use conversion

976   [38,52]. Pre-settlement vegetation provides an opportunity to test forest-climate

977   relationships prior to land-use conversion and to test dynamic vegetation models in a data

978   assimilation framework [92]. For these reason, the widespread loss of regional forest

979   associations common in the PLSS (Fig 8d), and the rapid rise of novel forest assemblages

980   (Fig 8e) have important implications for our ability to understand ecological responses to

981   changing climate. The loss of historical forest types implies that the modern understanding

982   of forest cover, climate relationships, realized and potential niches and species associations

983   may be strongly biased in this region, even though 29% of the total regional cover is novel

984   relative to forests only two centuries ago.

985   Beyond shifts in composition at a meso-scale, the broader shifts in ecotones can strongly

986   impact models of species responses and co-occurence on the landscape. For example, the

987   heterogeneity, distribution, and control of savanna-forest boundaries [93] is of particular

988   interest to ecologists and modelers given the ecological implications of current woody

989   encroachment on savanna ecosystems [94]. Declines in landscape heterogeneity may also

990   strongly affect ecosystem models, and predictions of future change. Our data show higher

991   levels of vegetation heterogeneity at mesoscales during the pre-settlement era, and greater

992   fine scaled turnover along transects. Lower β diversity shown here and elsewhere [18]

993   indicate increasing homogeneity at a very large spatial scale, and the loss of resolution

994   along major historical ecotones.

995   This study also points to the need for a deeper understanding of some of the landscape-

996   and regional-scale drivers of novelty, given the likely role for climatic and land use change

997   (including land abandonment) to continue to drive ecological novelty [95,96]. In particular

998   the role of regional species pools and remnant patches of forest in driving or mitigating

999   compositional novelty. This work shows that the baseline forest type, and its structure on

1000   the landscape moderates the degree to which landscape scale patterns can drive

1001   compositional novelty. To some degree relationships between compositional novelty and

1002   distance from remnant patches may be dependent on the simplicity or complexity of the

1003   species pool and the sensitivity of dissimilarity metrics to β diversity [97]. Our results

1004   indicate that diversity alone cannot be the driving factor in determining post-settlement

1005   dissimilarity (and novelty), since all forest classes show this pattern of change.

1006   Both Pine and the Oak/Poplar/Basswood/Maple forest types are the most fragmented

1007   across the region. There is strong evidence that in some locations pine forests have

1008   persisted over very long timescales in the region [98], although there is also evidence, in

1009   other regions, that these states may shift strongly in response to interactions between

1010   landscape level processes such as fire and geophysical features [99]. Thus complex

1011   interactions between landscape scale processes, whether they be fire, land use change, or

1012   geophysical features, and the species assemblages themselves, point to the difficulty in

1013   making simplifying assumptions about species assemblages. Caution in simplifying species

1014   assignments, whether they be plant functional types, species richness, or phylogenetic

1015   metrics, is neccessary since this region is dominated by forests that respond very

1016   differently to the settlement-era (and pre-settlement) disturbance, but that are composed

1017   of different species of the same genera and plant functional type. This caution is clearly

1018   warranted since recent ecosystem model benchmarking using pre-settlement vegetation

1019   has shown significant mismatch between climate representations of plant functional types

1020   across a range of ecosystem models, with no model accurately representing the true

1021   climate space of plant functional types in the northeastern upper Midwestern United States

1022   [100].

1023   The analysis relating to the distance-to-novelty (Fig 9) points to the possibility that

1024   landscape-scale restoriation has high likelihood of success if local-scale restoration focuses

1025   on sites where restoration potential is high, as suggested for Hemlock/Cedar/Birch/Maple

1026   forests in northern Wisconsin [86]. If some of the novelty is driven by depauparate species

1027   pools beyond certain threshold distances from remnant forests then it should also be

1028   possible to restore these forest at a regional scale through the translocation of key species

1029   [101]. This work is supported by a number of other studies at smaller scales [102–104], for

1030   example, the presence of white pine in mesic sites during the PLS era has been attributed to

1031   its presence as a seed source on marginal sites at scales of of hundreds of meters [105].

1032   Computer simulations [106] show that seed source distribution can affect community

1033   composition over hundreds of years at large spatial scales in a region spatially coincident

1034   with this current study. Thus land use change has significantly altered the landscape, both

1035   by "resetting" the sucessional clock, but also, because of the extent of change, by impacting

1036   the regional species pool and seed source for re-establishing forests that are

1037   compositionally similar to pre-settlement forests.

1038   Methodological advances of the current work include 1) the systematic standardization of

1039   PLSS data to enable mapping at broad spatial extent and high spatial resolution, 2) the use

1040   of spatially varying correction factors to accommodate variations among surveyors in

1041   sampling design, and 3) parallel analysis of FIA datasets to enable comparisons of forest

1042   composition and structure between contemporary and historical time periods. This

1043   approach is currently being extended to TPS and PLSS datasets across the north-central

1044   and northeastern US, with the goal of providing consistent reconstructions of forest

1045   composition and structure for northeastern US forests at the time of EuroAmerican forests.

1046   Our results support the consensus that robust estimates of pre-settlement forest

1047   composition and structure can be obtained from PLSS data [39,44,46,107,108]. Patterns of

1048   density, basal area and biomass are roughly equivalent to previous estimates [16,19], but

1049   show variability across the region, largely structured by historical vegetation type (Table

1050   3). Our results for stem density are lower than those estimated by Hanberrry *et al*. [17] for

1051   eastern Minnesota, but density and basal area are similar to those in the northern Lower

1052   Peninsula of Michigan [109] and biomass estimates are in line with estimates of

1053   aboveground carbon for Wisconsin [19].

1054   These maps of settlement-era forest composition and structure can provide a useful

1055   calibration dataset for pollen-based vegetation reconstructions for time periods prior to

1056   the historic record [110]. Many papers have used calibration datasets comprised of modern

1057   pollen samples to build transfer functions for inferring past climates and vegetation from

1058   fossil pollen records [111–114]. However, modern pollen datasets are potentially

1059   confounded by recent land use, which can alter paleoclimatic reconstructions using pollen

1060   data [113]. By linking pollen and vegetation at modern and historical periods we develop

1061   capacity to provide compositional datasets at broader spatio-temporal scales, providing

1062   more data for model validation and improvement. Ultimately, it should be possible to

1063   assimilate these empirical reconstructions of past vegetation with dynamic vegetation

1064   models in order to infer forest composition and biomass during past climate changes. Data

1065   assimilation, however, requires assessment of observational and model uncertainty in the

1066   data sources used for data assimilation. Spatiotemporal models of uncertainty are being

1067   developed for the compositional data [63].

1068   Ultimately the pre-settlement vegetation data present an opportunity to develop and refine

1069   statistical and mechanistic models of terrestrial vegetation that can take multiple structural

1070   and compositional forest attributes into account. The future development of uncertainty

1071   estimates for the data remains an opportunity that can help integrate pre-settlement

1072   estimates of composition and structure into a data assimilation framework to build more

1073   complete and more accurate reconstructions of past vegetation dynamics, and to help

1074   improve predictions of future vegetation under global change scenarios.

Acknowledgements

Literature Cited

1082   1. Ramankutty N, Foley JA. Estimating historical changes in global land cover: Croplands

1083   from 1700 to 1992. Global biogeochemical cycles. Wiley Online Library; 1999;13: 997–

1084   1027.

1085   2. Thompson JR, Carpenter DN, Cogbill CV, Foster DR. Four centuries of change in

1086   northeastern United States forests. PloS one. Public Library of Science; 2013;8: e72540.

1087   3. Munoz SE, Mladenoff DJ, Schroeder S, Williams JW. Defining the spatial patterns of

1088   historical land use associated with the indigenous societies of eastern North America.

1089   Journal of Biogeography. Wiley Online Library; 2014;

1090   4. Ellis EC, Ramankutty N. Putting people in the map: Anthropogenic biomes of the world.

1091   Frontiers in Ecology and the Environment. Eco Soc America; 2008;6: 439–447.

1092   5. Foster DR, Motzkin G, Slater B. Land-use history as long-term broad-scale disturbance:

1093   Regional forest dynamics in central New England. Ecosystems. Springer; 1998;1: 96–119.

1094   6. Pederson N, Dyer JM, McEwan RW, Hessl AE, Mock CJ, Orwig DA, et al. The legacy of

1095   episodic climatic events in shaping temperate, broadleaf forests. Ecological Monographs.

1096   Eco Soc America; 2014;84: 599–620.

1097   7. Booth RK, Jackson ST, Sousa VA, Sullivan ME, Minckley TA, Clifford MJ. Multi-decadal

1098   drought and amplified moisture variability drove rapid forest community change in a

1099   humid region. Ecology. Eco Soc America; 2012;93: 219–226.

1100   8. Hotchkiss SC, Calcote R, Lynch EA. Response of vegetation and fire to Little Ice Age

1101   climate change: Regional continuity and landscape heterogeneity. Landscape Ecology.

1102   Springer; 2007;22: 25–41.

1103   9. Umbanhowar CE, Camill P, Geiss CE, Teed R. Asymmetric vegetation responses to mid-

1104   Holocene aridity at the prairie–forest ecotone in south-central Minnesota. Quaternary

1105   Research. Elsevier; 2006;66: 53–66.

1106   10. Brugam RB. Pollen indicators of land-use change in southern connecticut. Quaternary

1107   Research. Elsevier; 1978;9: 349–362.

1108   11. Fuller JL, Foster DR, McLachlan JS, Drake N. Impact of human activity on regional forest

1109   composition and dynamics in central New England. Ecosystems. Springer; 1998;1: 76–95.

1110   12. McAndrews JH. Human disturbance of North American forests and grasslands: The

1111   fossil pollen record. Vegetation history. Springer; 1988. pp. 673–697.

1112   13. Grossmann EB, Mladenoff DJ. Farms, fires, and forestry: Disturbance legacies in the soils

1113   of the northwest Wisconsin (USA) sand plain. Forest Ecology and Management. Elsevier;

1114   2008;256: 827–836.

1115   14. Dupouey J-L, Dambrine E, Laffite J-D, Moares C. Irreversible impact of past land use on

1116   forest soils and biodiversity. Ecology. Eco Soc America; 2002;83: 2978–2984.

1117   15. Etienne D, Ruffaldi P, Dupouey JL, Georges-Leroy M, Ritz F, Dambrine E. Searching for

1118   ancient forests: A 2000 year history of land use in northeastern French forests deduced

1119   from the pollen compositions of closed depressions. The Holocene. Sage Publications;

1120   2013;23: 678–691.

1121   16. Schulte LA, Mladenoff DJ, Crow TR, Merrick LC, Cleland DT. Homogenization of northern

1122   US Great Lakes forests due to land use. Landscape Ecology. Springer; 2007;22: 1089–1103.

1123   17. Hanberry BB, Palik BJ, He HS. Comparison of historical and current forest surveys for

1124   detection of homogenization and mesophication of Minnesota forests. Landscape Ecology.

1125   Springer; 2012;27: 1495–1512.

1126   18. Li D-J, Waller DM. Drivers of observed biotic homogenization in pine barrens of central

1127   Wisconsin. Ecology. Eco Soc America; 2014;

1128   19. Rhemtulla JM, Mladenoff DJ, Clayton MK. Historical forest baselines reveal potential for

1129   continued carbon sequestration. Proceedings of the National Academy of Sciences. National

1130   Acad Sciences; 2009;106: 6082–6087.

1131   20. Foster D, Swanson F, Aber J, Burke I, Brokaw N, Tilman D, et al. The importance of land-

1132   use legacies to ecology and conservation. BioScience. Oxford University Press; 2003;53:

1133   77–88.

1134   21. Wolkovich E, Cook B, McLauchlan K, Davies T. Temporal ecology in the anthropocene.

1135   Ecology letters. Wiley Online Library; 2014;17: 1365–1379.

1136   22. Rhemtulla JM, Mladenoff DJ, Clayton MK. Legacies of historical land use on regional

1137   forest composition and structure in wisconsin, USA (mid-1800s-1930s-2000s). Ecological

1138   Applications. Eco Soc America; 2009;19: 1061–1078.

1139   23. Cramer VA, Hobbs RJ, Standish RJ. What's new about old fields? Land abandonment and

1140   ecosystem assembly. Trends in Ecology & Evolution. Elsevier; 2008;23: 104–112.

1141   24. Bellemare J, Motzkin G, Foster DR. Legacies of the agricultural past in the forested

1142   present: An assessment of historical land-use effects on rich mesic forests. Journal of

1143   Biogeography. Wiley Online Library; 2002;29: 1401–1420.

1144   25. Murcia C, Aronson J, Kattan GH, Moreno-Mateos D, Dixon K, Simberloff D. A critique of

1145   the 'novel ecosystem'concept. Trends in ecology & evolution. Elsevier; 2014;29: 548–553.

1146   26. Hobbs RJ, Higgs ES, Harris JA. Novel ecosystems: Concept or inconvenient reality? A

1147   response to murcia et al. Trends in ecology & evolution. Elsevier; 2014;29: 645–646.

1148   27. Menz MH, Dixon KW, Hobbs RJ. Hurdles and opportunities for landscape-scale

1149   restoration. Science. American Association for the Advancement of Science; 2013;339:

1150   526–527.

1151   28. Hobbs RJ, Higgs E, Hall CM, Bridgewater P, Chapin III FS, Ellis EC, et al. Managing the

1152   whole landscape: Historical, hybrid, and novel ecosystems. Frontiers in Ecology and the

1153   Environment. Eco Soc America; 2014;12: 557–564.

1154   29. Matthes JH, Goring S, Williams JW, Dietze MC. Historical vegetation reconstruction

1155   benchmarks CMIP5 pre-colonial land-climate feedbacks across the upper Midwest and

1156   northeastern United States. Global Change Ecology. in revision;

1157   30. Radeloff VC, Mladenoff DJ, Boyce MS. A historical perspective and future outlook on

1158   landscape scale restoration in the northwest Wisconsin pine barrens. Restoration Ecology.

1159   Wiley Online Library; 2000;8: 119–126.

1160   31. Fritschle JA. Reconstructing historic ecotones using the Public Land Survey: The lost

1161   prairies of Redwood National Park. Annals of the Association of American Geographers.

1162   Taylor & Francis; 2008;98: 24–39.

1163   32. Knoot TG, Schulte LA, Tyndall JC, Palik BJ. The state of the system and steps toward

1164   resilience of disturbance-dependent oak forests. Ecology and Society. 2010;15: 5.

1165   33. Gimmi U, Radeloff VC. Assessing naturalness in northern Great Lakes forests based on

1166   historical land-cover and vegetation changes. Environmental management. Springer;

1167   2013;52: 481–492.

1168   34. Gray AN, Brandeis TJ, Shaw JD, McWilliams WH, Miles PD. Forest Inventory and

1169   Analysis database of the United States of America (FIA). In: Dengler J, Oldeland J, Jansen F,

1170   Chytry M, Ewald J, Finckh M, et al., editors. Vegetation databases for the 21st century–

1171   Biodiversity & Ecology. 2012. pp. 255–264.

1172   35. Stewart LO. Public Land Surveys: History, instructions, methods. Ames, Iowa, USA:

1173   Collegiate Press, Incorporated; 1935. p. 202.

1174   36. White CA. A history of the rectangular survey system. Washington D.C., USA: US

1175   Department of the Interior, Bureau of Land Management; 1983. p. 774.

1176   37. Almendinger JC. Minnesota's bearing tree database. Minn. Department of Natural

1177   Resources; 1996.

1178   38. Liu F, Mladenoff DJ, Keuler NS, Moore LS. Broadscale variability in tree data of the

1179   historical Public Land Survey and its consequences for ecological studies. Ecological

1180   Monographs. Eco Soc America; 2011;81: 259–275.

1181   39. Williams MA, Baker WL. Testing the accuracy of new methods for reconstructing

1182   historical structure of forest landscapes using GLO survey data. Ecological Monographs. Eco

1183   Soc America; 2011;81: 63–88.

1184   40. Tomscha SA, Gergel SE. Historic land surveys present opportunities for reconstructing

1185   frontier settlement patterns in North America. Landscape Ecology. Springer; 2014;30: 203–

1186   213.

1187   41. Iverson LR, Prasad AM. Predicting abundance of 80 tree species following climate

1188   change in the eastern United States. Ecological Monographs. Eco Soc America; 1998;68:

1189   465–485.

1190   42. Iverson LR, McKenzie D. Tree-species range shifts in a changing climate: Detecting,

1191   modeling, assisting. Landscape Ecology. Springer; 2013;28: 879–889.

1192   43. Manies KL, Mladenoff DJ. Testing methods to produce landscape-scale presettlement

1193   vegetation maps from the US Public Land Survey records. Landscape Ecology. Springer;

1194   2000;15: 741–754.

1195   44. Schulte LA, Mladenoff DJ, Nordheim EV. Quantitative classification of a historic northern

1196   Wisconsin (USA) landscape: Mapping forests at regional scales. Canadian Journal of Forest

1197   Research. NRC Research Press; 2002;32: 1616–1638.

1198   45. Friedman SK, Reich PB. Regional legacies of logging: Departure from presettlement

1199   forest conditions in northern Minnesota. Ecological applications. Eco Soc America;

1200   2005;15: 726–744.

1201   46. Duren OC, Muir PS, Hosten PE. Vegetation change from the Euro-American settlement

1202   era to the present in relation to environment and disturbance in southwest Oregon.

1203   Northwest Science. BioOne; 2012;86: 310–328.

1204   47. Kronenfeld BJ, Wang Y-C, Larsen CP. The influence of the "Mixed Pixel"? Problem on the

1205   detection of analogous forest communities between presettlement and present in western

1206   New York. The Professional Geographer. Taylor & Francis; 2010;62: 182–196.

1207   48. Cogbill CV, Burk J, Motzkin G. The forests of presettlement New England, USA: Spatial

1208   and compositional patterns based on town proprietor surveys. Journal of Biogeography.

1209   Wiley Online Library; 2002;29: 1279–1304.

1210   49. Schulte LA, Mladenoff DJ. Severe wind and fire regimes in northern forests: Historical

1211   variability at the regional scale. Ecology. Eco Soc America; 2005;86: 431–445.

1212   50. Bourdo EA. A review of the General Land Office survey and of its use in quantitative

1213   studies of former forests. Ecology. JSTOR; 1956; 754–768.

1214   51. Mladenoff DJ, Dahir SE, Nordheim EV, Schulte LA, Guntenspergen GG. Narrowing

1215   historical uncertainty: Probabilistic classification of ambiguously identified tree species in

1216   historical forest survey data. Ecosystems. Springer; 2002;5: 539–553.

1217   52. Schulte LA, Mladenoff DJ. The original US Public Land Survey records: Their use and

1218   limitations in reconstructing presettlement vegetation. Journal of Forestry. Society of

1219   American Foresters; 2001;99: 5–10.

1220   53. Manies KL, Mladenoff DJ, Nordheim EV. Assessing large-scale surveyor variability in the

1221   historic forest data of the original US Public Land Survey. Canadian Journal of Forest

1222   Research. NRC Research Press; 2001;31: 1719–1730.

1223   54. Terrail R, Arseneault D, Fortin M-J, Dupuis S, Boucher Y. An early forest inventory

1224   indicates high accuracy of forest composition data in pre-settlement land survey records.

1225   Journal of Vegetation Science. Wiley Online Library; 2014;25: 691–702.

1226   55. Kronenfeld BJ, Wang Y-C. Accounting for surveyor inconsistency and bias in estimation

1227   of tree density from presettlement land survey records. Canadian Journal of Forest

1228   Research. NRC Research Press; 2007;37: 2365–2379.

1229   56. Kronenfeld BJ. Validating the historical record: A relative distance test and correction

1230   formula for selection bias in presettlement land surveys. Ecography. Wiley Online Library;

1231   2014;

1232   57. Cogbill CV, Goring SJ, Thurman A. Estimation of robust correction factors for Public

1233   Land Survey Data. in prep;

1234   58. Hanberry BB, Yang J, Kabrick JM, He HS. Adjusting forest density estimates for surveyor

1235   bias in historical tree surveys. The American Midland Naturalist. BioOne; 2012;167: 285–

1236   306.

1237   59. Bouldin J. Some problems and solutions in density estimation from bearing tree data: A

1238   review and synthesis. Journal of Biogeography. Wiley Online Library; 2008;35: 2000–2011.

1239   60. Hanberry BB, Fraver S, He HS, Yang J, Dey DC, Palik BJ. Spatial pattern corrections and

1240   sample sizes for forest density estimates of historical tree surveys. Landscape Ecology.

1241   Springer; 2011;26: 59–68.

1242   61. Team RC. R: A language and environment for statistical computing (version 3.1. 0).

1243   vienna, Austria: R Foundation for Statistical Computing. 2014.

1244   62. Bivand R, Keitt T, Rowlingson B. rgdal: Bindings for the Geospatial Data Abstraction

1245   Library [Internet]. 2014. Available: http://CRAN.R-project.org/package=rgdal

1246   63. Paciorek C, Goring SJ, Thurman A, Cogbill C, Williams J, Mladenoff D, et al. Statistically-

1247   estimated tree composition for the northeastern United States at the time of Euro-

1248   American settlement. PLoS One. in review;

1249   64. Morisita M. A new method for the estimation of denisty by the spacing method

1250   applicable to non-randomly distributed populations (in Japanese). Physiology and Ecology

1251   (Kyoto). 1957;7: 134–144.

1252   65. Persson O. The robustness of estimating density by distance measurements. In: Patil GP,

1253   Pielou EC, Waters WE, editors. Statistical Ecology: Sampling and Modeling Biological

1254   Populations and Population Dynamics. University Park, Pennsylvania: Pennsylvania

1255   University Press; 1971.

1256   66. Morisita M. Estimation of population density by spacing method. Memoirs of the Faculty

1257   of Science Kyushu University, Series E. 1954;1: 187–197.

1258   67. Cottam G, Curtis JT. The use of distance measures in phytosociological sampling.

1259   Ecology. JSTOR; 1956; 451–460.

1260   68. Jenkins JC, Chojnacky DC, Heath LS, Birdsey RA, others. Comprehensive database of

1261   diameter-based biomass regressions for North American tree species. US Department of

1262   Agriculture, Forest Service, Northeastern Research Station Newtown Square, PA; 2004;

1263   69. Anderson RC, Anderson MR. The presettlement vegetation of Williamson county,

1264   Illinois. Castanea. JSTOR; 1975; 345–363.

1265   70. Woudenberg SW, Conkling BL, O'Connell BM, LaPoint EB, Turner JA, Waddell KL. The

1266   Forest Inventory and Analysis database: Database description and users manual version

1267   4.0 for phase 2. US Department of Agriculture, Forest Service, Rocky Mountain Research

1268   Station; 2010;

1269   71. Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. Cluster: Cluster analysis basics

1270   and extensions. 2014.

1271   72. Wickham H. ggplot2: Elegant graphics for data analysis. Springer; 2009.

1272   73. Wickham H. ggplot2: Elegant graphics for data analysis [Internet]. Springer New York;

1273   2009. Available: http://had.co.nz/ggplot2/book

1274   74. Auguie B. GridExtra: Functions in Grid graphics [Internet]. 2012. Available:

1275   http://CRAN.R-project.org/package=gridExtra

1276   75. Csardi G, Nepusz T. The igraph software package for complex network research.

1277   InterJournal. 2006;Complex Systems: 1695. Available: http://igraph.org

1278   76. Wood S. Fast stable restricted maximum likelihood and marginal likelihood estimation

1279   of semiparametric generalized linear models. Journal of the Royal Statistical Society (B).

1280   2011;73: 3–36. doi:10.1111/j.1467-9868.2010.00749.x

1281   77. Wickham H. The split-apply-combine strategy for data analysis. Journal of Statistical

1282   Software. 2011;40: 1–29. Available: http://www.jstatsoft.org/v40/i01/

1283   78. Hijmans RJ. raster: Geographic data analysis and modeling [Internet]. 2014. Available:

1284   http://CRAN.R-project.org/package=raster

1285   79. Wickham H. Reshaping data with the reshape package. Journal of Statistical Software.

1286   2007;21: 1–20. Available: http://www.jstatsoft.org/v21/i12/

1287   80. Bivand R, Rundel C. rgeos: Interface to geometry engine - open source (GEOS)

1288   [Internet]. 2014. Available: http://CRAN.R-project.org/package=rgeos

1289   81. Pebesma E, Bivand R. Classes and methods for spatial data in R. R News. 2005;5.

1290   Available: http://cran.r-project.org/doc/Rnews/

1291   82. Bivand R, Pebesma E, Gomez-Rubio V. Applied spatial data analysis with R [Internet].

1292   2nd ed. Springer; 2013. Available: http://www.asdar-book.org/

1293   83. Bivand R. spdep: Spatial dependence: Weighting schemes, statistics and models

1294   [Internet]. 2014. Available: http://CRAN.R-project.org/package=spdep

1295   84. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, et al. vegan:

1296   Community ecology package [Internet]. 2014. Available: http://CRAN.R-

1297   project.org/package=vegan

1298   85. Curtis JT. The vegetation of Wisconsin: An ordination of plant communities. University

1299   of Wisconsin Press; 1959.

1300   86. Bolliger J, Schulte LA, Burrows SN, Sickley TA, Mladenoff DJ. Assessing ecological

1301   restoration potentials of Wisconsin (USA) using historical landscape reconstructions.

1302   Restoration Ecology. Wiley Online Library; 2004;12: 124–142.

1303   87. Haxeltine A, Prentice IC. BIOME3: An equilibrium terrestrial biosphere model based on

1304   ecophysiological constraints, resource availability, and competition among plant functional

1305   types. Global Biogeochemical Cycles. Wiley Online Library; 1996;10: 693–709.

1306   88. Schulte LA, Mladenoff DJ, Burrows SN, Sickley TA, Nordheim EV. Spatial controls of pre–

1307   Euro-American wind and fire disturbance in northern Wisconsin (USA) forest landscapes.

1308   Ecosystems. Springer; 2005;8: 73–94.

1309   89. Record S, Fitzpatrick MC, Finley AO, Veloz S, Ellison AM. Should species distribution

1310   models account for spatial autocorrelation? A test of model projections across eight

1311   millennia of climate change. Global ecology and biogeography. Wiley Online Library;

1312   2013;22: 760–771.

1313   90. Heffernan JB, Soranno PA, Angilletta Jr MJ, Buckley LB, Gruner DS, Keitt TH, et al.

1314   Macrosystems ecology: Understanding ecological patterns and processes at continental

1315   scales. Frontiers in Ecology and the Environment. Eco Soc America; 2014;12: 5–14.

1316   91. Groffman PM, Rustad LE, Templer PH, Campbell JL, Christenson LM, Lany NK, et al.

1317   Long-term integrated studies show complex and surprising effects of climate change in the

1318   northern hardwood forest. Bioscience. Oxford University Press; 2012;62: 1056–1066.

1319   92. Hartig F, Dyke J, Hickler T, Higgins SI, OâHara RB, Scheiter S, et al. Connecting dynamic

1320   vegetation models to data–an inverse perspective. Journal of Biogeography. Wiley Online

1321   Library; 2012;39: 2240–2252.

1322   93. Staver AC, Archibald S, Levin SA. The global extent and determinants of savanna and

1323   forest as alternative biome states. Science. American Association for the Advancement of

1324   Science; 2011;334: 230–232.

1325   94. Ratajczak Z, Nippert JB, Collins SL. Woody encroachment decreases diversity across

1326   North American grasslands and savannas. Ecology. Eco Soc America; 2012;93: 697–703.

1327   95. Radeloff VC, Williams JW, Bateman BL, Burke KD, Carter SK, Childress ES, et al. The rise

1328   of novelty in ecosystems. Ecological Applications. Eco Soc America; in press;

1329   96. Martinuzzi S, Gavier-Pizarro GI, Lugo AE, Radeloff VC. Future land-use changes and the

1330   potential for novelty in ecosystems of the United States. Ecosystems. Springer; 2015; 1–11.

1331   97. Faith DP, Minchin PR, Belbin L. Compositional dissimilarity as a robust measure of

1332   ecological distance. Vegetatio. Springer; 1987;69: 57–68.

1333   98. Ewing HA. The influence of substrate on vegetation history and ecosystem

1334   development. Ecology. Eco Soc America; 2002;83: 2766–2781.

1335   99. Lynch EA, Calcote R, Hotchkiss SC, Tweiten M. Presence of lakes and wetlands decreases

1336   resilience of jack pine ecosystems to late-holocene climatic changes. Canadian Journal of

1337   Forest Research. NRC Research Press; 2014;44: 1331–1343.

1338   100. Matthes JH, Goring SJ, Williams JW, Dietze MC. Benchmarking historical cMIP5 land-

1339   climate feedbacks across the upper midwest and northeastern United States. Journal of

1340   Geophysical Research. in review;

1341   101. Seddon PJ. From reintroduction to assisted colonization: Moving along the

1342   conservation translocation spectrum. Restoration Ecology. Wiley Online Library; 2010;18:

1343   796–802.

1344   102. Dunn CP, Guntenspergen GR, Dorney JR. Catastrophic wind disturbance in an old-

1345   growth hemlock-hardwood forest, wisconsin. Canadian Journal of Botany. NRC Research

1346   Press; 1983;61: 211–217.

1347   103. Johnson WC. Estimating dispersibility of acer, fraxinus and tilia in fragmented

1348   landscapes from patterns of seedling establishment. Landscape Ecology. Springer; 1988;1:

1349   175–187.

1350   104. Nyamai PA, Goebel PC, Hix DM, Corace RG, Drobyshev I. Fire history, fuels, and

1351   overstory effects on the regeneration-layer dynamics of mixed-pine forest ecosystems of

1352   eastern upper michigan, USA. Forest Ecology and Management. Elsevier; 2014;322: 37–47.

1353   105. Fahey RT, Lorimer CG, Mladenoff DJ. Habitat heterogeneity and life-history traits

1354   influence presettlement distributions of early-successional tree species in a late-

1355   successional, hemlock-hardwood landscape. Landscape ecology. Springer; 2012;27: 999–

1356   1013.

1357   106. He HS, Mladenoff DJ. The effects of seed dispersal on the simulation of long-term forest

1358   landscape change. Ecosystems. Springer; 1999;2: 308–319.

1359   107. Rayburn AP, Schulte LA. Integrating historic and contemporary data to delineate

1360   potential remnant natural woodlands within Midwestern agricultural landscapes. Natural

1361   Areas Journal. BioOne; 2009;29: 4–14.

1362   108. Hanberry BB, Kabrick JM, He HS. Densification and state transition across the Missouri

1363   Ozarks landscape. Ecosystems. Springer; 2014;17: 66–81.

1364   109. Leahy MJ, Pregitzer KS. A comparison of presettlement and present-day forests in

1365   northeastern lower Michigan. The American Midland Naturalist. BioOne; 2003;149: 71–89.

1366   110. Dawson A, Paciorek C, McLachlan J, Goring S, Williams JW, Jackson S. Calibrating the

1367   pollen vegetation relationship. Quaternary Science Reviews. in review;

1368   111. Birks HB, Heiri O, Seppä H, Bjune AE. Strengths and weaknesses of quantitative

1369   climate reconstructions based on late-Quaternary biological proxies. Open Ecology Journal.

1370   Bentham Science Publishers B. V., P. O. Box 1673 Hilversum 1200 BR Netherlands; 2010;3:

1371   68–110.

1372   112. Paciorek CJ, McLachlan JS. Mapping ancient forests: Bayesian inference for spatio-

1373   temporal trends in forest composition using the fossil pollen proxy record. Journal of the

1374   American Statistical Association. Taylor & Francis; 2009;104: 608–622.

1375   113. Jacques J-MS, Cumming BF, Smol JP. A pre-European settlement pollen–climate

1376   calibration set for Minnesota, USA: Developing tools for palaeoclimatic reconstructions.

1377   Journal of Biogeography. Wiley Online Library; 2008;35: 306–324.

1378   114. Goring S, Pellatt MG, Lacourse T, Walker IR, Mathewes RW. A new methodology for

1379   reconstructing climate and vegetation from modern pollen assemblages: An example from

1380   British Columbia. Journal of Biogeography. Wiley Online Library; 2009;36: 626–638.