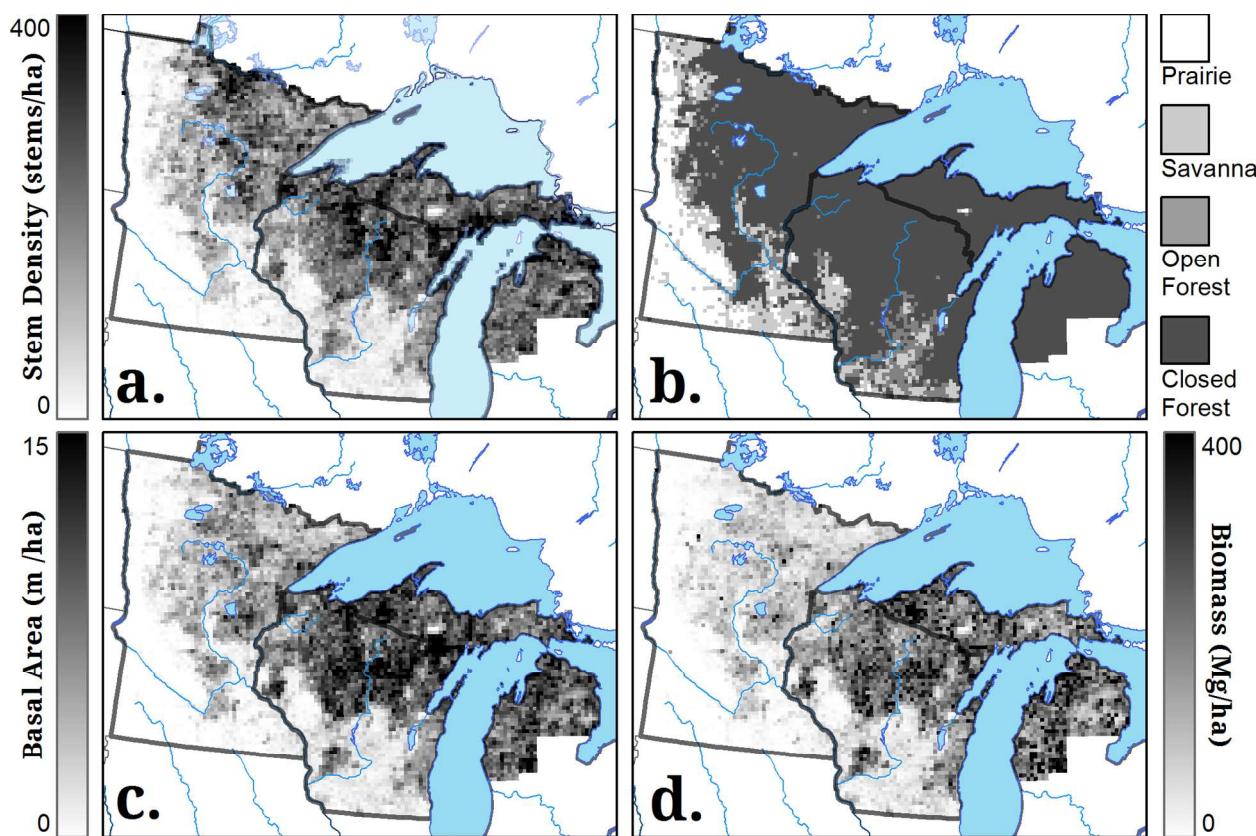


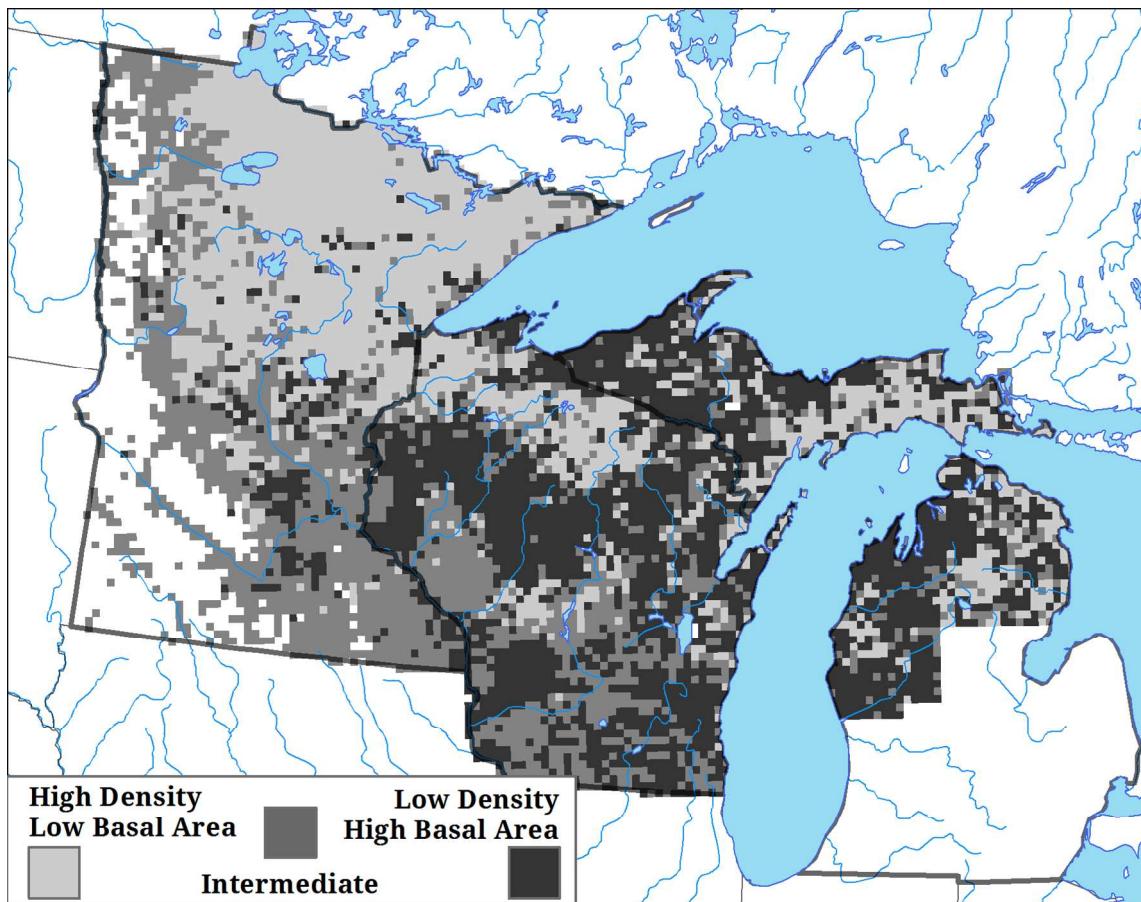


**Changes in Forest Composition, Stem Density, and Biomass
from the Settlement Era to Present in for the Upper
Midwestern United States**

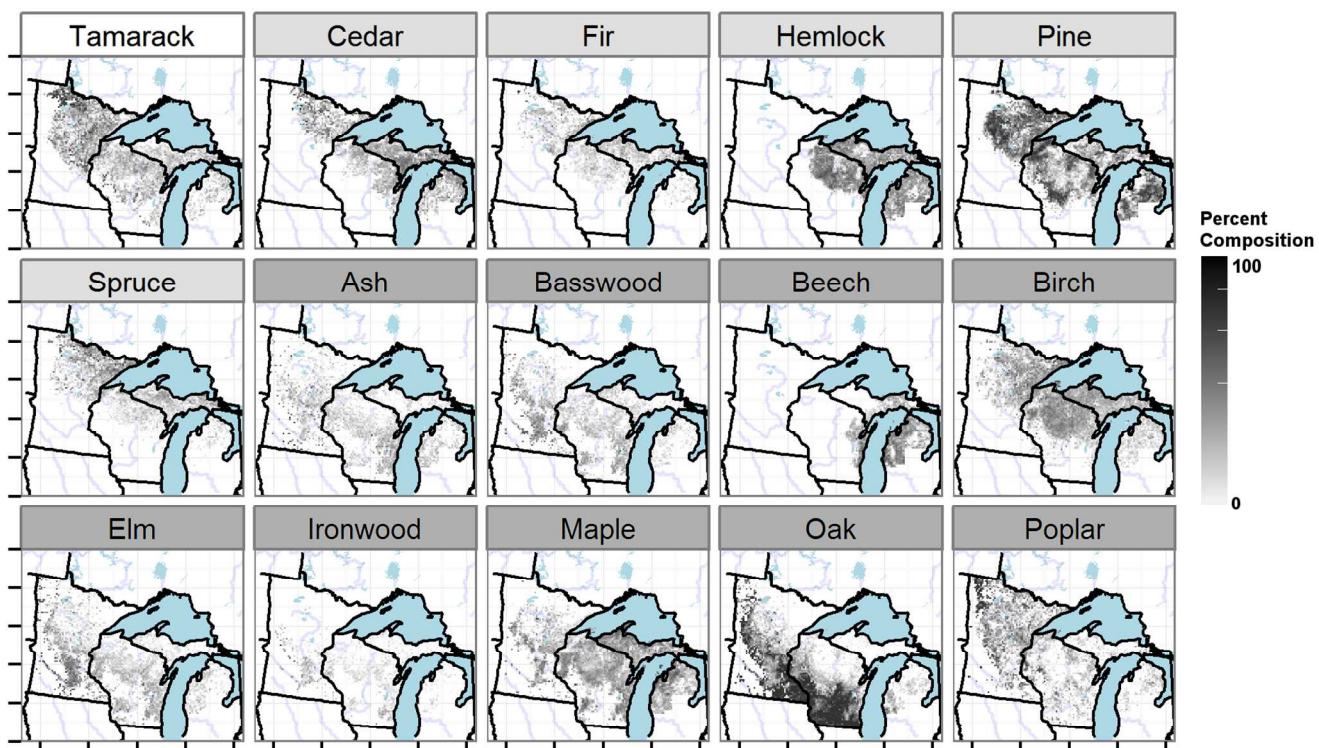
Journal:	<i>Journal of Biogeography</i>
Manuscript ID:	Draft
Manuscript Type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Goring, Simon; University of Wisconsin, Department of Geography Williams, Jack; University of Wisconsin, Geography Mladenoff, David; University of Wisconsin, Department of Forest and Wildlife Ecology Cogbill, Charles; Harvard University, Harvard Forest Record, Sydne; Bryn Mawr College, Department of Biology Paciorek, Christopher; University of California, Berkeley, Department of Statistics Jackson, Stephen; Department of the Interior , Southwest Climate Science Center; University of Arizona, Department of Geosciences Dietze, Michael; Boston University, Department of Earth and Environment Matthes, Jaclyn; Boston University, Department of Earth and Environment McLachlan, Jason; Notre Dame, Department of Biological Sciences
Key Words:	biomass, stem density, euro-american settlement, potential vegetation, land use, historic vegetation, non-analogue vegetation
Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.	
Supplement2.rar	



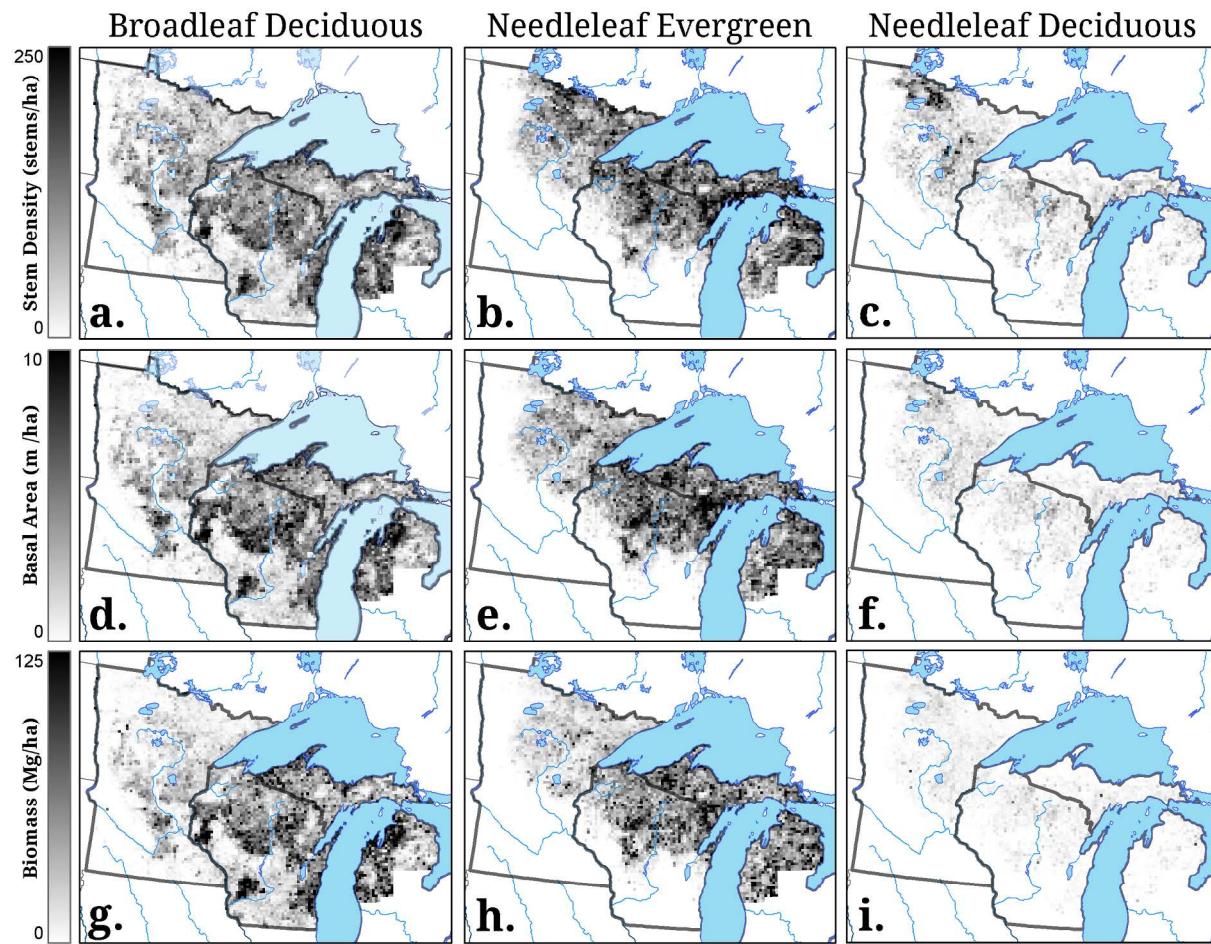


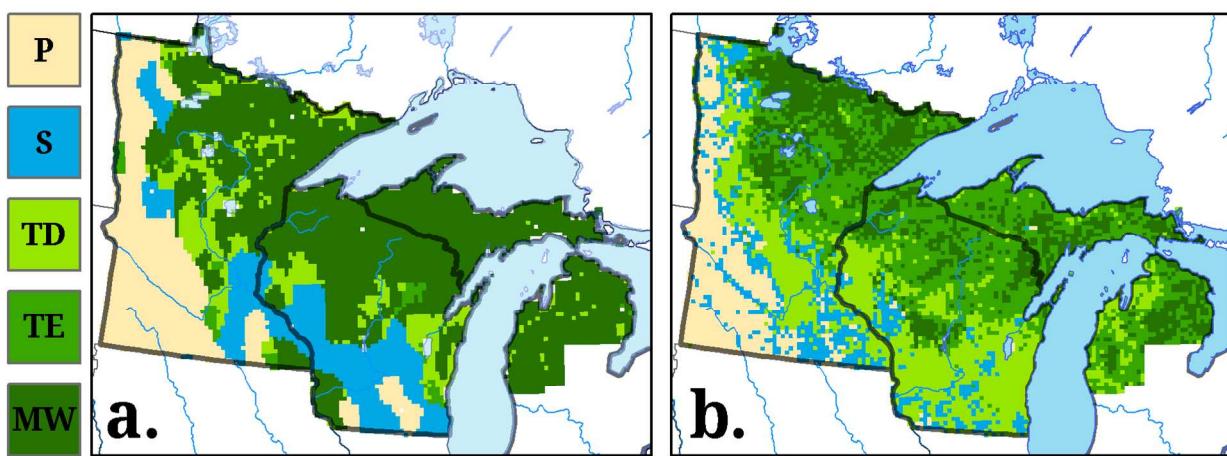


Review

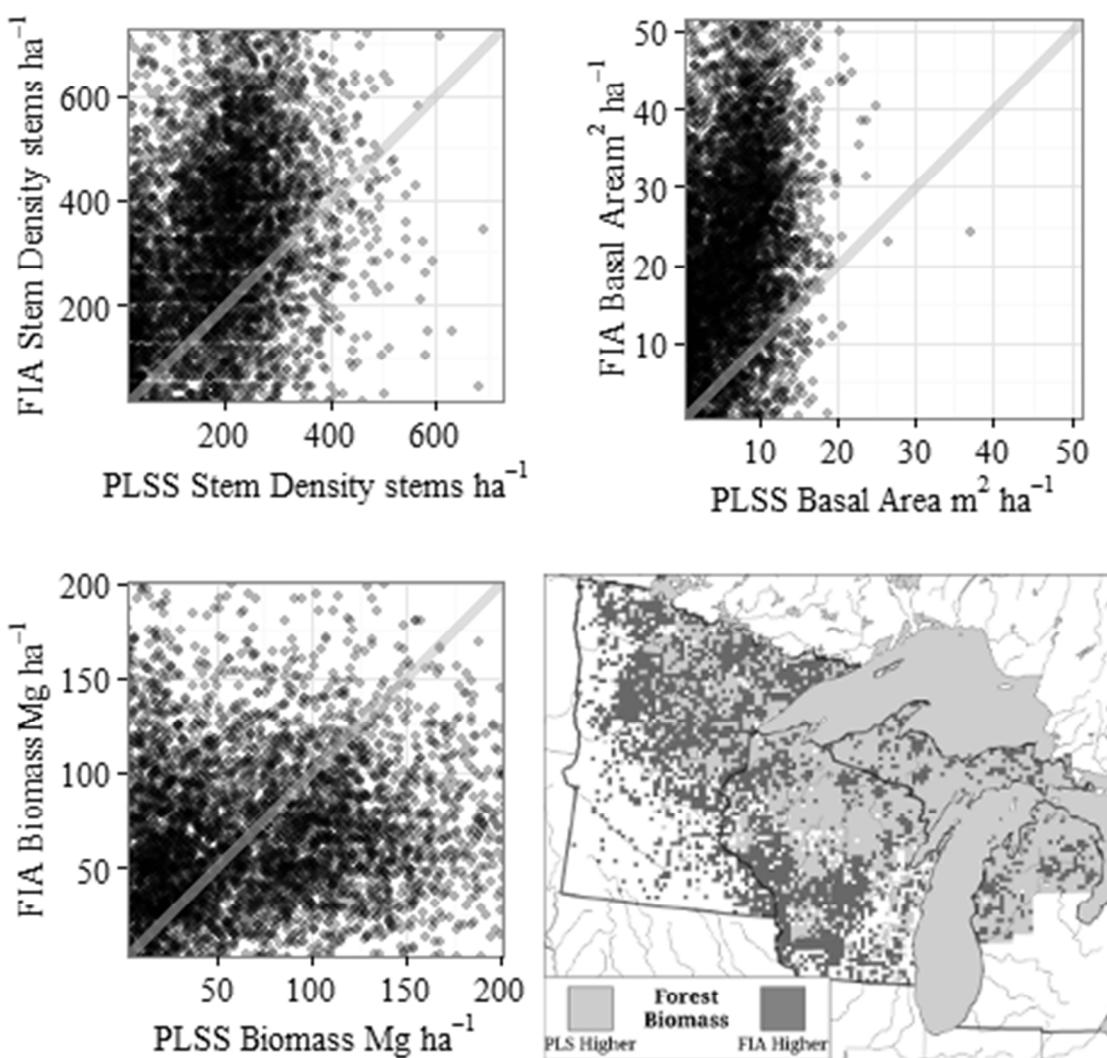


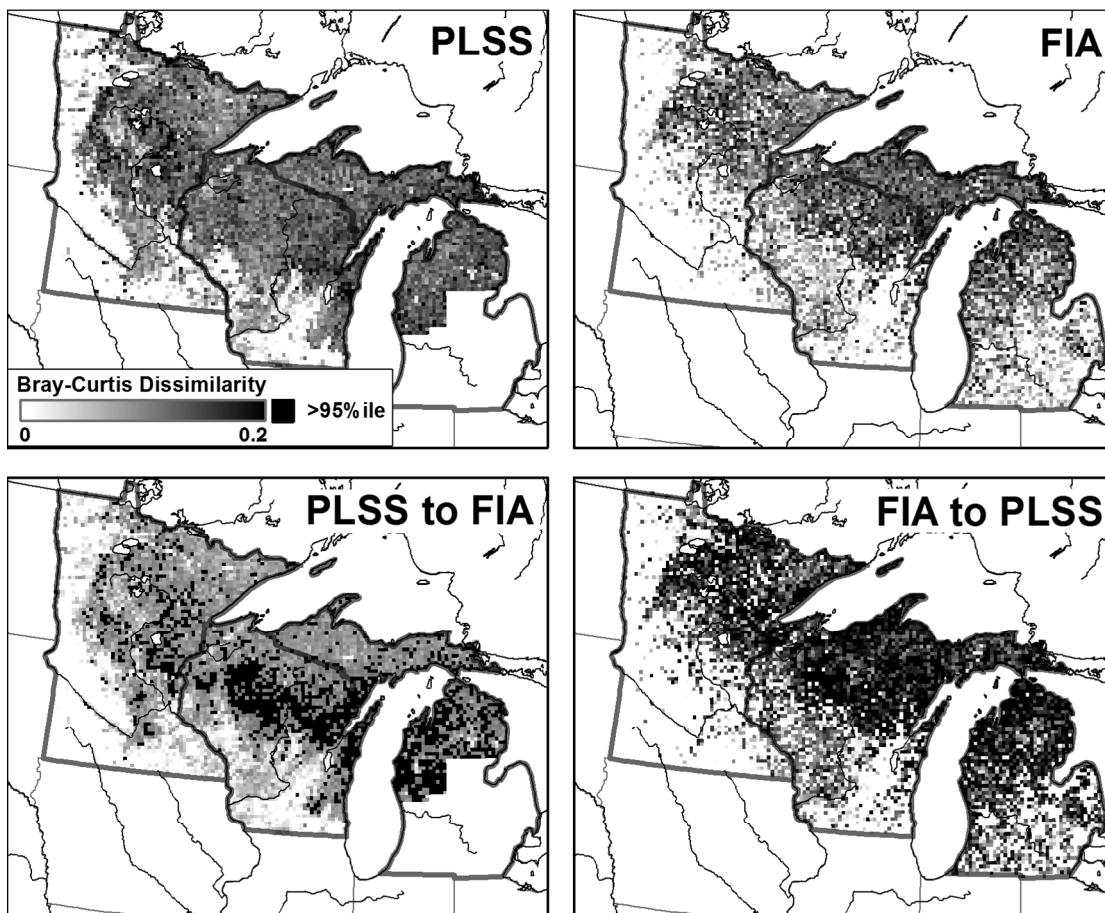
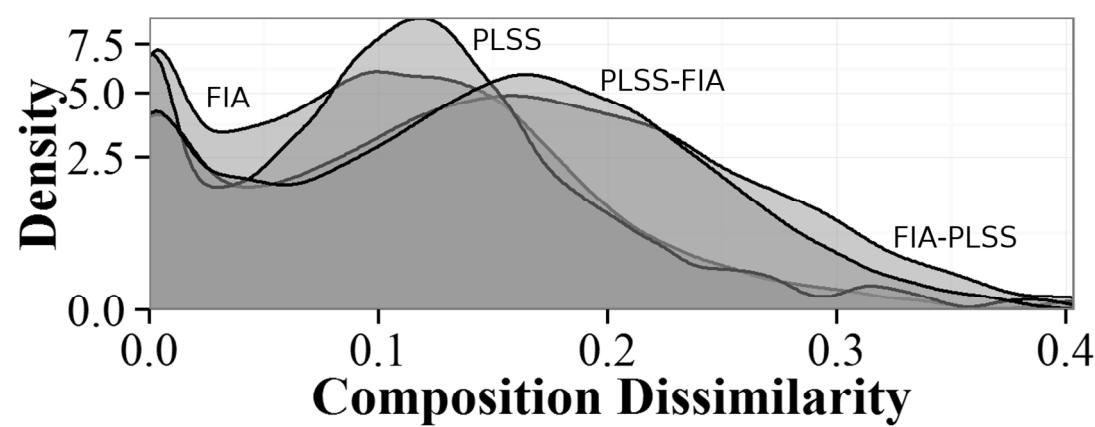
Review





For Peer Review





1
2
3 1 Original Article:
4 2 Changes in Forest Composition, Stem Density, and Biomass from the Settlement Era to Present
5 3 in for the Upper Midwestern United States
6
7 4

8 5 Simon J. Goring¹, John W. Williams^{1,2}, David J. Mladenoff³, Charles V. Cogbill⁴, Sydne
9 6 Record⁵, Christopher J. Paciorek, Stephen T. Jackson, Michael C. Dietze, Jaclyn Hatala Matthes,
10 7 Jason S. McLachlan
11 8

12 9 1Department of Geography, University of Wisconsin, Madison, 550 N Park St, Madison WI
13 10 53706

14 11 2Center for Climatic Research, University of Wisconsin, Madison, 1225 W Dayton St, Madison
15 12 WI 53706

16 13 3Department of Forest and Wildlife Ecology, University of Wisconsin-Madison, 1630 Linden
17 14 Dr, Madison WI 53706

18 15 4Harvard Forest, Harvard University, 324 N Main St, Petersham MA 01366

19 16 5Department of the Interior Southwest Climate Science Center, 1955 E. Sixth St. Tucson, AZ
20 17 85719; School of Natural Resources and the Environment and Department of Geosciences,
21 18 University of Arizona, Tucson AZ 85721

22 19 6Department of Statistics, University of California, Berkeley, 367 Evans Hall, Berkeley CA
23 20 94720

24 21 7Department of Earth and Environment, Boston University, 685 Commonwealth Ave, Boston,
25 22 MA 02215

26 23 8Department of Biological Sciences, University of Notre Dame, 100 Galvin Life Sciences
27 24 Center, Notre Dame, IN 46556
28
29

30 26 Corresponding Author:

31 27 Simon Goring

32 28 Department of Geography

33 29 University of Wisconsin

34 30 550 N Park St

35 31 Madison WI 53706

36 32 goring@wisc.edu

37
38 35 Running Title: Pre-settlement forest change in the upper Midwestern USA
39
40 36 Word Count: 7500 words
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 **37 Abstract**

4
5
6 **38 Aim:** Reconstructing pre-settlement vegetation for the Upper Midwestern United States to
7
8 **39** improve understanding of post-settlement regional change and to provide gridded data layers for
9
10 **40** vegetation modeling.

11
12
13 **41 Location:** The Upper Midwestern United States (Minnesota, Wisconsin and Michigan)

14
15
16 **42 Methods:** Estimates of stem density, basal area and biomass come from Public Land Survey
17
18 **43** (PLS) data and Forest Inventory and Analysis (FIA) data products. Data is gridded to an 8x8km
19
20 **44** grid. Comparisons are made between data layers and to potential vegetation from Ramankutty
21
22 **45** and Foley (1999). Analogues between PLS and FIA forests are estimated using Bray-Curtis
23
24 **46** compositional dissimilarity.

25
26
27
28
29 **47 Results:** Stem density and basal area are lower in the PLS (mean 190 stems/ha) than in FIA
30
31 **48** (mean data, but biomass is comparable, with higher FIA biomass in the northwest and higher
32
33 **49** PLS biomass in the east. PLS data has higher heterogeneity and less extensive mixed wood
34
35 **50** forest than potential vegetation maps indicating the importance of hemlock (*Tsuga canadensis*)
36
37 **51** in pre-settlement forests of Wisconsin and Michigan.

38
39
40 **52 Main Conclusions:** Almost a quarter of modern forests have no analogue in pre-settlement
41
42 **53** forests, while 20% of PLS forests have no modern analogue. PLS forests also show strong
43
44 **54** differences with potential vegetation, meaning regional climate and vegetation models may
45
46 **55** suffer from poor models of forest structure and composition, with serious implications for
47
48 **56** predictions of nutrient flow and future development.

1
2
3 57 Keywords: biomass, stem density, euro-american settlement, potential vegetation, land use
4
5 58 change, dynamic vegetation models
6
7
8 59
9
10

11 60 **Introduction:**
12
13

14 61 The composition, demography, and structure of forests in eastern North America have changed
15
16 62 continuously over the last millennium, driven by human land use (Foster et al., 1998;
17
18 63 Ramankutty and Foley, 1999; Ellis and Ramankutty, 2008; Thompson et al., 2013; Munioz et al.,
19
20
21 64 in press) and climate variability during the Little Ice Age and Medieval Climate Anomaly (Booth
22
23 65 et al., 2012; Hotchkiss et al., 2007; Umbanhowar et al., 2004; Pederson et al., 2005). Forests in
24
25 66 the upper Midwestern United States (US: Minnesota, Wisconsin and Michigan) now have
26
27 67 decreased species richness and functional diversity relative to forests prior to Euro-American
28
29 68 settlement (ca. mid-late 1800s, EAS; Schulte et al., 2007) due to near complete logging. But
30
31 69 upper Midwestern forests are in a state of regrowth, continuing to sequester carbon (Rhemtulla et
32
33 70 al. 2009a) as a consequence of these extensive land use and land cover conversions and
34
35 71 subsequent recovery.
36
37

38 72 Pockets of primary forest originating before EAS exist in the upper Midwest, but these are an
39
40 73 incomplete and unrepresentative sample of past vegetation because of extensive and strongly
41
42 74 selective land use. Legacies of past land use have been shown to persist in the upper Midwest
43
44 75 (Grossman & Mladenoff, 2008). Hence, observed ecological processes reflect both
45
46 76 anthropogenic and natural influences at decadal to centennial scales. These dual influences pose
47
48 77 the challenge that natural processes may be masked, or heavily modified by anthropogenic
49
50 78 effects. More broadly, realized niches for many taxa may be contingent on past and present land
51
52
53
54
55
56
57
58
59
60

1
2
3 79 use. If realized niches are strongly biased by land use change, then species distribution modeling
4
5 80 may have much weaker ability to interpret species responses to climate change than model
6
7 81 diagnostics might indicate.
8
9

10
11 82 Modern forest structure and composition data (e.g., from the USDA Forest Service's Forest
12
13 83 Inventory and Analysis National Program, FIA; Woudenberg et al., 2010) play a ubiquitous role
14
15 84 in forest management, carbon accounting, and basic research on forest ecosystems and
16
17 85 community dynamics. These surveys can be extended with longer-term historical data to
18
19 86 understand how composition has changed since EAS. The Public Land Survey (PLS) provides
20
21 87 broad coverage of the US prior to EAS (e.g., Almendinger, 1996; Liu et al., 2011). Better
22
23 88 understanding of settlement-era forest composition and post-settlement trajectories can serve
24
25 89 multiple purposes. For example, distributional models of tree species often rely upon FIA or
26
27 90 other contemporary observational data to build species-climate relationships that can be used to
28
29 91 predict potential range shifts (Iverson and McKenzie, 2013). Changes in forest composition since
30
31 92 EAS have altered species-climate relationships, affecting our ability to understand individual or
32
33 93 community responses to climate change (Thompson et al., 2013). So, are we measuring
34
35 94 contemporary forests that have no past analogs to predict a future with no contemporary analogs?
36
37
38

39
40 95 Reconstructions of historical forest structure and composition are also important to terrestrial
41
42 96 ecosystem modelers seeking to understand how fast- and slow-acting processes (Luo et al., 2011)
43
44 97 drive carbon cycling and vegetation dynamics. Models that lack dynamic vegetation often use
45
46 98 'potential vegetation' maps to represent vegetation prior to major land use change as a static
47
48 99 vegetation type (Ramankutty and Foley, 1999). Similarly, such maps are used as the land
50
51 100 surface in atmospheric models (e.g. Goosse et al. 2012, Notaro and Gutzler 2012, Bagley et al
52
53 101 2012, Loutre et al 2011, Urban and Keller 2010). Potential vegetation defines initial conditions
54
55
56
57
58
59
60

1
2
3 102 for dynamic vegetation models, and is critical to estimate ecosystem pools and fluxes for carbon
4
5 103 and other biologically active elements. Flawed maps of initial vegetation can strongly affect
6
7 104 model results, particularly when vegetation estimates are coarsely resolved relative to the
8
9 105 underlying environmental gradients (Hurt et al., 2010). Models are now able to make much
10
11 106 greater use of information about vegetation structure, composition, and disturbance history at
12
13 107 regional scales than earlier models (Moorecroft et al 2001, Smith et al. 2002), with species-level
14
15 108 estimates of variables such as stem density, diameter at breast height (dbh) and basal area.
16
17
18
19
20 109 Here we use survey data from the original PLS in the upper Midwest to derive estimates of pre-
21
22 110 EAS forest composition, basal area, stem density, and biomass and compare these to a) previous
23
24 111 maps of potential vegetation (Ramankutty and Foley 1999) and b) late-20th-century estimates of
25
26 112 forest composition, tree stem density, and biomass from FIA data. Digitization and classification
27
28 113 of PLS data has previously been performed for Wisconsin (Manies and Mladenoff, 2000, Schulte
29
30 114 et al., 2002) and for parts of Minnesota (Hanberry et al., 2012; Friedman and Reich, 2005). Most
31
32 115 prior PLS-based reconstructions are for individual states, smaller extents (among others: Duren
33
34 116 et al., 2012; Hanberry et al., 2012a; Rhemtulla et al., 2009b; Friedman and Reich, 2005) or with
35
36 117 coarser spatial aggregations (Schulte et al., 2007) so a key advance is to integrate and cross-
37
38 118 harmonize the PLS data at a regional scale. Here we aggregate point based estimates of stem
39
40 119 density, basal area and biomass to an 8 x 8km grid and classify forest types in the upper Midwest
41
42 120 using both genera and plant functional types (PFTs) to facilitate finer scale comparisons between
43
44 121 FIA and PLS data, and to provide data at spatial scales that can facilitate evaluations of dynamic
45
46 122 terrestrial ecosystem models (Hatala et al., submitted). We assess how the interpretations of
47
48 123 forest structure and composition varies among these three metrics and with the taxonomy used
49
50 124 for classification. Finally, we compare these datasets to potential vegetation that are simulated by
51
52
53
54
55
56
57
58
59
60

1
2
3 125 vegetation models, and modern inventory data to better understand the regional ecology of the
4
5 126 upper Midwestern US.
6
7
8
9 127 **Methods:**
10
11
12 128 **Public Land Survey Data: Assembly, and Standardization**
13
14
15 129 The PLS was undertaken by the US Government from 1785 to 1907 CE to facilitate the division
16
17 130 and sale of land from Ohio westward and south, replacing earlier town proprietor surveys (TPS)
18
19 131 used for the northeastern states (Thompson et al., 2013; Cogbill et al., 2003). The PLS created a
20
21 132 1 mi² (2.56 km²) grid (sections) on the landscape. At each section corner, a stake was placed as
22
23 133 the official location marker. To mark these survey points, PLS surveyors recorded tree stem
24
25 134 diameters, measured distances and azimuths of, generally, two to four trees near sample points
26
27 135 and identified tree taxa using common (and often regionally idiosyncratic) names along with
28
29 136 secondary site information (Schulte and Mladenoff 2005). PLS data thus represent measurements
30
31 137 by hundreds of surveyors over multiple decades, with changing sets of instructions (Stewart,
32
33 138 1979). However, potential uncertainties exist within the PLS and township level dataset (Bourdo,
34
35 139 1956). Uncertainty arises from the spatial scale of the dataset (fixed sampling every 1 mile),
36
37 140 uncertainty in the species taxonomy (Mladenoff et al., 2002) and potential surveyor variability
38
39 141 and bias during sampling (Bourdo, 1956; Schulte and Mladenoff, 2001; Manies et al., 2001; Liu
40
41 142 et al., 2011). Even given the limitations, PLS records remain the best, and only, source of data
42
43 143 about both forest composition and structure in the US prior to EAS.
44
45
46
47
48
49
50
51 144 This analysis builds upon and merges prior state-level efforts to digitize and database the point-
52
53 145 level PLS data for Wisconsin, Minnesota (Almendinger, 1996) and the Upper Peninsula and
54
55 146 upper third of the Lower Peninsula of Michigan (Michigan Department of Natural Resources).
56
57
58
59
60

1
2
3 147 These datasets were combined using spatial tools in R (R Core Team, 2014; package *rgdal*: Keitt
4 et al., 2012) to form a common dataset for the upper Midwest (Figure 1) using the Albers Great
5 Lakes and St Lawrence projection (see code in Supplement 1, file: *step_one_clean_bind.R*;
6
7 149 proj4: *+init:EPSG:3175*) that is gridded using an 8 x 8 km raster to create a dataset that has
8 sufficient numerical power for spatial statistical modeling and sufficient resolution for regional
9 scale analysis (Thurman et al. in prep.). This scale is smaller than the 100km² gridded scale used
10 in Freidman and Reich (2005), and provides a uniform metric grid similar to the 5 arc-minute
11
12 151 in Friedman and Reich (2005), and provides a uniform metric grid similar to the 5 arc-minute
13 scale analysis (Thurman et al. in prep.). This scale is smaller than the 100km² gridded scale used
14 in Freidman and Reich (2005), and provides a uniform metric grid similar to the 5 arc-minute
15 scale analysis (Thurman et al. in prep.). This scale is smaller than the 100km² gridded scale used
16 in Freidman and Reich (2005), and provides a uniform metric grid similar to the 5 arc-minute
17 scale analysis (Thurman et al. in prep.). This scale is smaller than the 100km² gridded scale used
18 in Freidman and Reich (2005), and provides a uniform metric grid similar to the 5 arc-minute
19 scale analysis (Thurman et al. in prep.). This scale is smaller than the 100km² gridded scale used
20 in Freidman and Reich (2005), and provides a uniform metric grid similar to the 5 arc-minute
21 scale analysis (Thurman et al. in prep.). This scale is smaller than the 100km² gridded scale used
22 in Freidman and Reich (2005), and provides a uniform metric grid similar to the 5 arc-minute
23 scale analysis (Thurman et al. in prep.). This scale is smaller than the 100km² gridded scale used
24 in Freidman and Reich (2005), and provides a uniform metric grid similar to the 5 arc-minute
25 scale analysis (Thurman et al. in prep.). This scale is smaller than the 100km² gridded scale used
26 in Freidman and Reich (2005), and provides a uniform metric grid similar to the 5 arc-minute
27 scale analysis (Thurman et al. in prep.). This scale is smaller than the 100km² gridded scale used
28 in Freidman and Reich (2005), and provides a uniform metric grid similar to the 5 arc-minute
29 scale analysis (Thurman et al. in prep.). This scale is smaller than the 100km² gridded scale used
30 in Freidman and Reich (2005), and provides a uniform metric grid similar to the 5 arc-minute
31 scale analysis (Thurman et al. in prep.). This scale is smaller than the 100km² gridded scale used
32 in Freidman and Reich (2005), and provides a uniform metric grid similar to the 5 arc-minute
33 scale analysis (Thurman et al. in prep.). This scale is smaller than the 100km² gridded scale used
34 in Freidman and Reich (2005), and provides a uniform metric grid similar to the 5 arc-minute
35 scale analysis (Thurman et al. in prep.). This scale is smaller than the 100km² gridded scale used
36 in Freidman and Reich (2005), and provides a uniform metric grid similar to the 5 arc-minute
37 scale analysis (Thurman et al. in prep.). This scale is smaller than the 100km² gridded scale used
38 in Freidman and Reich (2005), and provides a uniform metric grid similar to the 5 arc-minute
39 scale analysis (Thurman et al. in prep.). This scale is smaller than the 100km² gridded scale used
40 in Freidman and Reich (2005), and provides a uniform metric grid similar to the 5 arc-minute
41 scale analysis (Thurman et al. in prep.). This scale is smaller than the 100km² gridded scale used
42 in Freidman and Reich (2005), and provides a uniform metric grid similar to the 5 arc-minute
43 scale analysis (Thurman et al. in prep.). This scale is smaller than the 100km² gridded scale used
44 in Freidman and Reich (2005), and provides a uniform metric grid similar to the 5 arc-minute
45 scale analysis (Thurman et al. in prep.). This scale is smaller than the 100km² gridded scale used
46 in Freidman and Reich (2005), and provides a uniform metric grid similar to the 5 arc-minute
47 scale analysis (Thurman et al. in prep.). This scale is smaller than the 100km² gridded scale used
48 in Freidman and Reich (2005), and provides a uniform metric grid similar to the 5 arc-minute
49 scale analysis (Thurman et al. in prep.). This scale is smaller than the 100km² gridded scale used
50 in Freidman and Reich (2005), and provides a uniform metric grid similar to the 5 arc-minute
51 scale analysis (Thurman et al. in prep.). This scale is smaller than the 100km² gridded scale used
52 in Freidman and Reich (2005), and provides a uniform metric grid similar to the 5 arc-minute
53 scale analysis (Thurman et al. in prep.). This scale is smaller than the 100km² gridded scale used
54 in Freidman and Reich (2005), and provides a uniform metric grid similar to the 5 arc-minute
55 scale analysis (Thurman et al. in prep.). This scale is smaller than the 100km² gridded scale used
56 in Freidman and Reich (2005), and provides a uniform metric grid similar to the 5 arc-minute
57 scale analysis (Thurman et al. in prep.). This scale is smaller than the 100km² gridded scale used
58 in Freidman and Reich (2005), and provides a uniform metric grid similar to the 5 arc-minute
59 scale analysis (Thurman et al. in prep.). This scale is smaller than the 100km² gridded scale used
60 in Freidman and Reich (2005), and provides a uniform metric grid similar to the 5 arc-minute

22
23 155 Steps used to standardize the dataset are preserved in in the supplementary R code (Supplement
24 1: *R/step_one_clean_bind.R*). We exclude trees not located at section or quarter corners because
25 of the variability of, and potential for surveyor biases (Liu et al., 2011). We used only the closest
26 two trees at each corner point because of biases with respect to species composition and diameter
27 of the third and fourth furthest trees (Liu et al., 2011). Corner points were used only if there were
28 at least two trees in different quadrants (defined by the cardinal directions) with valid azimuths
29 and diameters (numeric, non-zero).

30
31 162 Surveyor species-level identifications can be ambiguous. Statistical models do exist to predict
32 the identity of ambiguous species (Mladenoff et al., 2002), but, given the regional scale of the
33 analysis, and the inherent uncertainty in the survey data itself, we chose to retain the more certain
34 genus level identification (Supplement 1, *data/input/conversion_table.csv*). Taxa were also
35 aggregated to PFTs using the classes Broadleaf Deciduous (BD), Needleleaf Evergreen (NE),
36 and Needleleafed Deciduous (ND). In areas of open prairie or other treeless areas, e.g.
37 southwestern Minnesota, surveyors recorded distances and bearings to 'Non Tree' objects.

1
2
3 169 To reduce errors associated with transcription from PLS notes to digital databases we exclude
4
5 170 sites for which multiple large trees have a distance of 1 link (20.12 cm) to plot center, trees with
6
7 171 very large diameters ($dbh > 100"$; 254 cm), plots where the azimuth to the tree is unclear, and
8
9 172 plots where the tree is at plot center but has a recorded azimuth. All removed plots are
10
11 173 documented in the code used for analysis (Supplement 1: *R/step_one_clean_bind.R*).
12
13
14
15

16 174 **Stem Density, Basal Area and Biomass**
17
18

19 175 Many efforts over the years have assessed and corrected for the idiosyncrasies of the original
20
21 176 surveyor data and minimize sampling bias (Manies et al. 2001, Bouldin 2008, Williams and
22
23 177 Baker 2011; Hanberry 2011, 2012a, b; Liu et al 2011, Kronenfeld and Wang, 2007; Cogbill et al.
24
25 178 in prep.). Survey sampling instructions changed throughout the implementation of the PLS in
26
27 179 this region and differed between section and quarter section points, and between internal and
28
29 180 external points within townships (Stewart, 1979; White, 1983; Liu et al. 2011). The changing
30
31 181 plot geometry across the region thus precludes uniform aggregative methods (e.g. Hanberry et al.
32
33 182 2011) because there is significant variability between sampling geometries among plots (Cogbill
34
35 183 et al., in prep). Our approach allows for spatial variation in surveyor methods.
36
37
38

39 184 We estimate stem density based on a modification of the distance-to-tree measurements for the
40
41 185 two closest trees at each point (stems/m^2 , Morisita 1954) using explicit and spatially varying
42
43 186 correction factors, modeled after the Cottam correction factor (Cottam and Curtis, 1956). This
44
45 187 method is described (and code is provided) in Supplement 2.
46
47
48

49 188 Basal area is calculated by multiplying the point-based stem density estimate by the average stem
50
51 189 basal area at the point. Aboveground dry biomass (Mg/ha) is calculated using the USFS FIA tree
52
53
54
55
56
57
58
59
60

1
2
3 190 volume and dry aboveground biomass equations for the US (Jenkins et al., 2004). The form of
4
5 191 the equation and group assignments are reported in Supplement 2.
6
7
8
9 192 **Forest Classification**
10
11
12 193 To reconstruct forest classes we use the global potential vegetation categories defined by
13
14 194 Haxeltine and Prentice (1996) and used in Ramankutty and Foley (1999), which are widely used
15
16 195 by terrestrial ecosystem modelers to set initial vegetation states (Table 1). We use the stem
17
18 196 density thresholds of Anderson and Anderson (1975) to discriminate prairie, savanna, and forest.
19
20
21
22 197 **FIA Stem Density, Basal Area and Biomass**
23
24
25 198 The US Forest Service has monitored forests through the FIA Program since 1929, with an
26
27 199 annualized state inventory system implemented in 1998 (Woudenberg et al., 2010). On average
28
29 200 there is one permanent FIA plot per 2,428 ha of land in the US classified as forested. Each FIA
30
31 201 plot consists of four 7.2m fixed-radius subplots in which measurements are made of all trees
32
33 202 >12.7cm dbh (Woudenberg et al. 2010). We used mean basal area (m^2/ha), stem density
34
35 203 (stems/ha), diameter at breast height (cm), mean biomass (Mg/ha) for all live trees with dbh
36
37 204 greater than 20.32cm (8in) from the most recent full plot inventory (2007-2011). The FIA plot
38
39 205 inventory provides a median of 3 FIA plots per cell. Biomass calculations used the same set of
40
41 206 allometric regression equations as for the PLS data (Jenkins et al. 2004: Supplement 1:
42
43 207 *data/input/plss.pft.conversion.csv*).
44
45
46
47
48
49
50 208 **Statistical Modelling**
51
52
53 209 Forest composition data is based on the number of individuals of each taxon (genera or PFTs)
54
55 210 within a cell. Stem density, basal area and biomass (calculated below) are calculated as the mean
56
57
58
59
60

1
2
3 211 point values within the cell. Stem density estimates are highly sensitive to trees close to the plot
4 center. To accommodate these points we carry these values through our analysis, but exclude the
5
6 212 top 2.5%ile when reporting means and standard deviations in our analysis. The base raster and
7
8 213 all rasterized data are available in Supplement 1 (folder: *data/output*).
9
10
11

12
13 215 Within-cell variance is expected to be high for the PLS point data, but spatial patterns are
14 expected to be robust at the cell level. A relatively low density of FIA plots per cell (median n =
15
16 216 3) in typically heterogeneous forest will result in high within-cell variance and relatively high
17
18 217 between-cell variability for the FIA data. Thus, fundamentally the two datasets are capturing
19 different patterns and scales of variability on the landscape, but between-dataset comparisons
20 should show clear regional patterns due to our taxonomic aggregation and since many
21
22 218 compositional and structural processes in the region operate on large spatial scales.
23
24
25

26 222 Standard statistical analysis, including correlations and regression, was carried out in R (R Core
27 Team, 2013), and is documented in supplementary material that includes a subset of the raw data
28 to allow reproducibility. Analysis and presentation uses elements from the following R packages:
29
30 223 ggplot2 (Wickham, 2009), gridExtra (Auguie, 2012), igraph (Csardi and Nepusz, 2006), mgcv
31
32 224 (Wood, 2011), plyr (Wickham, 2011), raster (Hijmans, 2014), reshape2 (Wickham, 2007), rgdal
33
34 225 (Bivand et al., 2013), rgeos (Bivand & Rundel, 2013), sp (Pebesma and Bivand, 2005; Bivand et
35
36 226 al., 2013), and spdep (Bivand, 2013).
37
38
39

40 229 Differences in composition between and within gridded PLS and FIA datasets are examined
41 using Bray-Curtis dissimilarity (vegdist in vegan; Oksanen et al., 2013) for proportional basal
42 area composition, reporting the minimum distance between a focal cell and its nearest
43
44 231 compositional neighbor within or between datasets. The distribution of closest analog distances
45
46 232 within datasets can give us information about forest heterogeneity, while the search for closest
47
48 233

1
2
3 234 analogs between datasets can tell us whether contemporary forests lack analogs in pre-settlement
4
5 235 forests, and vice versa.
6
7
8
9 236 **Results:**
10
11
12 237 **Data Standardization**
13
14
15 238 The Midwestern PLS dataset contains 519,000 corner points (Wisconsin: 171,000; Minnesota:
16
17 239 252,000; Michigan: 96,000) aggregated to 7,970 raster cells. Standardizing data and accounting
18
19 240 for potential outliers removed approximately 1.5% points from the dataset. Each cell contains
20
21 241 between 1 and 94 corner points, with a mean of 65 ($\delta= 15$) and a median of 69 corners. Cells
22
23 242 with a low number of points were near water bodies or along political boundaries such as the
24
25 243 Canada/Minnesota border.
26
27
28
29
30 244 **Spatial Patterns of Settlement-Era Forest Composition: Taxa and PFTs**
31
32
33
34 245 **Stem Density, Basal Area and Biomass**
35
36
37 246 Mean stem density for forested stands (prairie and savanna excluded) in the region is 190
38
39 247 stems/ha (Figure 2) with a 95%ile range from 0 to 400 stems/ha and cell-level standard
40
41 248 deviations between 0 and 430 stems/ha. Basal area in the domain (Figure 4c) has a 95%ile range
42
43 249 between 0 and 15 m^2/ha , a mean of 6.7 m^2/ha and cell level standard deviations between 0 and 18
44
45 249 m^2/ha . Biomass ranged from 0 to 200 Mg/ha (Figure 4d), with cell level standard deviations
46
47 250 between 0 and 590 Mg/ha.
48
49
50
51 252 Highest stem densities are in north-central Minnesota and in north-eastern Wisconsin (Figure
52
53 253 2a). Stem density is lowest in the western and southwestern prairie and savanna region (Figure
54
55 254 4b, Table 1). The interplay between broad-scale climate and local scale hydrological controls on
56
57
58
59
60

1
2
3 255 forest composition and density is evident along the Minnesota River in south-western Minnesota,
4
5 256 where a corridor of savanna is sustained in a region mostly occupied by prairie (Figure 2b).
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

257 Forest structure prior to EAS can be understood in part by the relationships between stem density
258 and biomass. Regions in northern Minnesota have low biomass and high stem densities (Figure
259 3, light gray). This indicates the presence of young, small diameter, even-ages stands, due to
260 frequent stand-replacing fire disturbance prior to EAS. This area also has shallower soils, colder
261 climate, and resulting lower productivity. Higher biomass values relative to stem density (Figure
262 3, darker gray) occur in southern Wisconsin, and throughout Michigan, regions with greater
263 proportions of deciduous species and higher diameters than northern Minnesota, due to higher
264 productivity soil and climate and much rarer stand-replacing disturbance.

265 Taxon composition is highly heterogenous across the region (Figure 4). Oak is dominant, with
266 high proportions in savanna and prairie regions. Pine also has high proportions across the
267 region. Maple, birch, tamarack and hemlock make up the rest of the major taxa.

268 Species distributions are broadly associated with the distributions of the PFTs to which they
269 belong (e.g., pine distributions occupy the extent of the NE PFT; Figures 4). Tamarack, the only
270 ND taxon, is primarily a wetland species, and it is most dense in the far northwestern extent of
271 the Upper Midwest. Hemlock was the dominant NE taxon across much of this region (Wisconsin
272 and Michigan), and represents a mesic taxon with habitat characteristics very different from
273 other NE taxa in this region. Thus NE in Wisconsin and Michigan can be attributed to mesic sites
274 dominated by Hemlock, rather than dry upland pine (Figures 4 & 5). Pine distributions represent
275 three dominant taxa, *Pinus strobus*, *P. resinosa* and *P. banksiana*. These species have
276 overlapping but ecologically dissimilar distributions, occurring in close proximity in some
277 regions, such as central Wisconsin, and typically associated with sandy outwash plains with low

1
2
3 278 water availability. Spruce represents two species (*Picea glauca*, *P. mariana*) with close
4
5 279 distributional overlap, but strongly different site preferences (dry upland and moist sites
6
7 280 respectively). Both cedar (*Thuja occidentalis*) and fir (*Abies balsamea*) are mono-specific.
8
9
10 281 The BD taxa can be divided into three classes. Northern hardwoods, such as birch (*Betula*
11
12 282 *alleghaniensis*, *B. paperifera*) and sugar maple (*Acer saccharum*), and beech (*Fagus*
13
14 283 *grandifolia*), are less common in the lower peninsula of Michigan, and southern Wisconsin,
15
16 284 except along Lake Michigan. Birch is extensive in the north, likely reflecting high pre-settlement
17
18 285 proportions of yellow birch (*Betula alleghaniensis*) on mesic soils, and paper birch on sandy fire-
19
20 286 prone soils and in northern Minnesota (birch proportions reach upwards of 34% in the northeast).
21
22
23 287 Hardwoods in southern and eastern portions of the region, such as oak, elm, basswood and
24
25 288 beech, are most typically mono-specific groupings, with the exception of oak, which comprises
26
27 289 seven species. These taxa are located primarily along the savanna and southern forest margins, or
28
29 290 in southern temperate deciduous forests. Finally, maple and poplar (aspen) have a broad regional
30
31 291 distribution, occupying nearly the entire wooded domain. Poplar comprises 4 species in the
32
33 292 region, while maple comprises five. Thus overlap among PFT distributions (Figure 5) emerges
34
35 293 from the changing composition among BD taxa.

41
42
43 294 The distribution of PFTs shows the well-known ecotone between forest and prairie in western
44
45 295 Minnesota and the mosaic of prairie, savanna, and forest in southwestern Wisconsin. BD taxa are
46
47 296 distributed across the entire region, with highest densities in the south, while ND and NE taxa are
48
49 297 limited to the central and northern portions of the region. BD proportions are high in any cell
50
51 298 with presence ($60.1 \pm 32.7\%$), while NE is lower when present ($12.1 \pm 15.6\%$ when present).
52
53 299 BD is more abundant when both PFTs are present ($x_{BD} = 45\%$, $x_{NE} = 12$, $t_{6779} = 70.6$, $p < 0.001$),
54
55 300 indicating widespread BD presence, even in the north.

1
2
3 301 **Settlement-Era Vegetation Types**

4
5
6 302 Differences between PLS-based classification and the model-based simulation of potential
7
8 303 vegetation types are substantial (Figure 5). Heterogeneity is substantially higher in the PLS
9
10 304 classification than for the RF potential vegetation map, with correspondingly smaller patch sizes
11
12 305 and higher proportions of edge cells (Table 2). Prairie, savanna and mixedwood forests occupy
13
14 306 much less space in the PLS-based estimates than in the RF map of potential vegetation (Table 2).
15
16 307 This difference can be accounted for by higher proportions of BD and NE classes at the expense
17
18 308 of the Mixedwood class in particular.

19
20
21
22
23 309 **Mensuration, Allometric Scaling and Diameter**

24
25
26 310 **Comparison to FIA Composition**

27
28
29
30 311 Comparing PLS and FIA estimates for stem density, basal area and biomass indicates that the
31
32 312 modern forests (FIA) have higher stem densities and total basal areas, but overall, comparable
33
34 313 biomass between the PLS and FIA data (Figure 7). The similarity in biomass despite lower stem
35
36 314 density and total basal area in the PLS data is surprising. Two likely factors are shifts in
37
38 315 allometric scaling associated with changes in species composition, or a higher mean diameter for
39
40 316 PLS trees (Figure 7d). The FIA appears to have higher average diameters in northern regions,
41
42 317 where the relationship between increases in diameter and increases in biomass is expected to be
43
44 318 lower because of allometric scaling. The higher average diameters in the Mixedwood and
45
46 319 Temperate Deciduous forests of the east (Figure 7d) may be the mechanism by which low
47
48 320 density and basal area produce roughly equivalent biomass estimates between the FIA and PLS.
49
50
51 321 Differences between FIA and PLS data in sampling design are unlikely to be a factor; these
52
53 322 differences are expected to affect how these datasets sample local- to landscape-scale

1
2
3 323 heterogeneity, but should not affect the overall trends between datasets. Differences in variability
4
5 324 introduce noise into the relationship, but given the large number of samples used here, the trends
6
7 325 should be robust.
8
9

10
11 326 Both the PLS- and FIA-era compositional data show similar patterns of dissimilarity within
12
13 327 datasets (Figure 8a), which supports the assertion that differences in within and between cell
14
15 328 variability between the datasets do not affect regional-scale vegetation analyses. In the PLS data
16
17 329 (Fig. 8a), the highest dissimilarities occur in central Minnesota, and northwestern Wisconsin,
18
19 330 most commonly associated with high proportions of birch, maple and fir (Figure 4).
20
21

22
23 331 Dissimilarities for the FIA data are more commonly associated with high proportions of
24
25 332 hemlock, cedar and fir. The higher grid-scale heterogeneity apparent in the FIA data likely is the
26
27 333 result of lower sampling density and the presence of both managed and semi-natural stands.
28
29 334 However, there is a strong correlation between datasets ($r=0.90$, $p < 0.001$), suggesting that this
30
31 335 heterogeneity does not obscure the general patterns of compositional dissimilarity between the
32
33 336 FIA and PLS maps (Figure 8a,b).
34
35

36
37 337 Identical pairs represent 6.8% of PLS cells and 9% of FIA cells. Identical PLS cells are largely
38
39 338 located along the southern margin and 75% are composed entirely of oak. Cells in the FIA with
40
41 339 identical neighbors are commonly pure oak (26%), pure poplar (24%), or pure ash (14%).
42
43

44
45 340 We define no-analog communities as those whose nearest neighbour is beyond the 95%ile for
46
47 341 dissimilarities within a particular dataset. In the PLS dataset, forests that have no modern analogs
48
49 342 are defined here as lost forests, while forest types in the FIA data with no past analogs are
50
51 343 defined as novel. More than 36% of PLS sites have no analogue in the FIA dataset, while 38% of
52
53 344 FIA sites have no analogue in the PLS data (Figure 8c & d). For the PLS dataset the no-analog
54
55 345 region roughly follows the spatial distribution of the Wisconsin "Tension Zone" (cf. Curtis 1959)
56
57
58
59
60

1
2
3 346 where deciduous forest grades into hemlock-dominated mixed forests (Figure 6). The
4
5 347 distribution of no-analog sites in the FIA dataset is much more spatially diffuse and appears to be
6
7 348 shifted north of the PLS distribution. The taxa associated with lost forest types in the PLS data
8
9 349 are beech and hemlock, while novel FIA forests are associated with ash, maple and fir.

10
11
12
13 350 **Discussion**

14
15
16 351 Distinct features of the current work include 1) the systematic standardization of PLS data to
17
18 352 enable mapping at broad spatial extent and high spatial resolution, 2) the use of spatially varying
19
20 353 correction factors to accommodate variations among surveyors in sampling design, 3)
21
22 354 comparison of stem density, basal area, and biomass between the PLS and FIA datasets, and 4)
23
24 355 analog analyses to identify novel and lost forest types. Results show clear influences of local
25
26 356 features, for example riparian corridors in the prairie regions (Figure 2) that are not found in
27
28 357 studies that aggregate PLS data to coarser scales, but are seen in finer scale analysis (Danz et al.,
29
30 358 2011; Rayburn and Schulte, 2009). The combination of regional-scale extent and high spatial
31
32 359 grain, supported by a publicly available dataset and open source analytic code will further
33
34 360 support analysis of ecological patterns across the region. Our results support the consensus that
35
36 361 robust estimates of pre-settlement forest composition and structure can be obtained from PLS
37
38 362 data (e.g., Wisconsin: Schulte et al., 2002; California: Williams and Baker, 2011; Iowa: Rayburn
39
40 363 and Schulte, 2009; Oregon: Duren et al., 2012). Patterns of density, basal area and biomass are
41
42 364 roughly equivalent to previous estimates (Schulte et al., 2007, Rhemtulla et al., 2009a). Our
43
44 365 results for stem density are lower than those estimated by Hanberry et al. (2012a) for eastern
45
46 366 Minnesota, but density and basal area are similar to those in the northern Lower Peninsula of
47
48 367 Michigan (Leahy and Pregitzer, 2003) and biomass estimates are in line with estimates of
49
50 368 aboveground carbon for Wisconsin (Rhemtulla et al., 2009a).

1
2
3 369 Anthropogenic shifts in forest composition over decades and centuries seen here and elsewhere
4
5 370 (Cogbill, 2002; Thompson et al., 2012) are embedded within a set of interacting systems operate
6
7 371 on multiple scales of space and time (macrosystems, *sensu* Heffernan et al., 2013). Combining
8
9 372 regional historical baselines, long term ecological studies and high frequency analyses can reveal
10
11 373 complex responses to climate change at local and regional scales (Groffman et al., 2012).
12
13 374 Estimates of pre-settlement forest composition and structure are significant because they
14
15 375 represent a snapshot of the landscape prior to EAS. For these reason, the widespread loss of
16
17 376 regional PLS forest associations, and the rapid rise of novel forest assemblages have important
18
19 377 implications for our ability to understand ecological responses to changing climate. The loss of
20
21 378 these forests implies that the modern understanding of forest cover, climate relationships,
22
23 379 realized and potential niches and species associations may be strongly biased toward a single
24
25 380 state, even though 24% of the total regional cover is novel relative to forests only two centuries
26
27 381 ago.

33
34
35 382 A broad goal of this work is to provide high-quality and high-resolution maps of settlement-era
36
37 383 forest composition and structure that can be used as a baseline for terrestrial ecosystem models.
38
39 384 For vegetation modelers the development of global potential vegetation maps such as
40
41 385 Ramankutty and Foley (1999) was a major step forward. It provided a baseline for modeling
42
43 386 efforts and spurred the development of a new generation of coupled atmospheric-vegetation
44
45 387 models. However, as terrestrial ecosystem models have improved they have become significantly
46
47 388 more data hungry (Hartig et al., 2012) and have become more sensitive to data quality. The
48
49 389 development of accurate and high-resolution vegetation maps that can provide baseline data to
50
51 390 terrestrial ecosystem models is becoming increasingly important.

1
2
3 391 The distributions of forest, savanna, and prairie are broadly similar between the PLS-based maps
4
5 392 and the potential vegetation distributions simulated by Ramankutty and Foley (1999). Some
6
7 393 apparent differences are sensitive to the choice of classification thresholds and do not reflect
8
9 394 major differences in actual composition (e.g., the sensitivity of the placement of the
10
11 395 forest/savanna threshold) (Table 1, Fig. 6).

14
15
16 396 PLS-based maps of forest structure and composition are more spatially heterogeneous than
17
18 397 modeled vegetation (Figure 6). The higher-resolution PLS maps are able to resolve the general
19
20 398 Mixedwood forest type of RF into distinct patches of mixed stands of deciduous and evergreen
21
22 399 forests in a matrix of northern forests that is generally dominated by NE taxa (Figure 6). The
23
24 400 finer scale identification to genera also makes clear that "Mixedwood" and "Temperate
25
26 401 Evergreen" may stretch the 'useful fiction' of potential vegetation (*sensu* Jackson, 2013), given
27
28 402 the likely differences in ecological function between mixed pine/aspen forest in northern
29
30 403 Minnesota and sugar maple-hemlock-yellow birch forest in Michigan and northern Wisconsin.
31
32
33
34

35
36 404 The heterogeneity, distribution, and control of savanna-forest boundaries (Staver et al., 2011) is
37
38 405 of particular interest to ecologists and modelers given the ecological implications of current
39
40 406 woody encroachment on savanna ecosystems (Ratajczak et al., 2012). Increases in landscape
41
42 407 heterogeneity may also strongly affect ecosystem models. Recent work using the FLUXNET
43
44 408 tower network has shown that energy budgets are strongly related to landscape measures of
45
46 409 heterogeneity (Stoy et al., 2013). Because flux measurements are so critically important to the
47
48 410 development and calibration of many ecosystem models (e.g., Richardson et al., 2010); failure to
49
50 411 accurately represent past heterogeneity can have significant impact on accurate hindcasting.
51
52
53

54
55 412 These maps of settlement-era forest composition and structure can also provide a useful
56
57 413 calibration dataset for pollen-based vegetation reconstructions for time periods prior to the
58
59
60

1
2
3 414 historic record. Calibration datasets comprised of modern pollen samples are used to build
4
5 415 transfer functions for inferring past climates and vegetation from fossil pollen records (Sugita,
6
7 416 2007a, b; Goring et al., 2009; Paciorek and McLachlan, 2009). However, modern pollen datasets
8
9 417 are potentially confounded by recent land use, which can alter paleoclimatic reconstructions (St.
10
11 418 Jacques et al. 2008). By linking pollen and vegetation at modern and historical periods we
12
13 419 develop capacity to provide compositional datasets at broader spatio-temporal scales, providing
14
15 420 more data for model validation and improvement. Ultimately, it should be possible to assimilate
16
17 421 these empirical reconstructions of past vegetation with dynamic vegetation models in order to
18
19 422 infer forest composition and biomass during past climate changes. Data assimilation, however,
20
21 423 requires assessment of observational and model uncertainty in the data sources used for data
22
23 424 assimilation. Spatiotemporal models of uncertainty are being developed for the compositional
24
25 425 data (Thurman et al., in prep.) and biomass data (Feng et al. in prep.).
26
27
28
29
30
31
32 426 Ultimately the pre-settlement vegetation data present an opportunity to develop and refine
33
34 427 statistical and mechanistic models of terrestrial vegetation that can take multiple structural and
35
36 428 compositional forest attributes into account. The future development of uncertainty estimates for
37
38 429 the data remains an opportunity that can help integrate pre-settlement estimates of composition
39
40 430 and structure into a data assimilation framework to build more complete and more accurate
41
42 431 reconstructions of past vegetation dynamics, and to help improve predictions of future vegetation
43
44 432 under global change scenarios.

49
50 433 **Acknowledgements**
51
52
53 434
54
55
56
57
58
59
60

435 **References**

- 436 Almendinger, J.C. (1996) *Minnesota's Bearing Tree Database*. Minnesota Department of
437 Natural Resources Biological Report No. 56.
- 438 Anderson, R.C. & Anderson, M.R. (1975) The presettlement vegetation of Williamson County,
439 Illinois. *Castanea*, **40**, 345-363.
- 440 Auguie B. (2012) *gridExtra: functions in Grid graphics*. R package version 0.9.1.
441 <http://CRAN.R-project.org/package=gridExtra>
- 442 Bagley JE, Desai AR, Dirmeyer PA, Foley JA. (2012) Effects of land cover change on moisture
443 availability and potential crop yield in the world's breadbaskets. *Environmental Research
Letters*. **7**:014009.
- 445 Bivand R.S. (2013) *spdep: Spatial dependence: weighting schemes, statistics and models*. R
446 package version 0.5-68. <http://CRAN.R-project.org/package=spdep>
- 447 Bivand, R.S., Keitt T. & Rowlingson B. (2013). *rgdal: Bindings for the Geospatial Data
Abstraction Library*. R package version 0.8-14. <http://CRAN.R-project.org/package=rgdal>
- 449 Bivand, R.S., Pebesma, E. & Gomez-Rubio, V. (2013) *Applied spatial data analysis with R*,
450 Second edition. Springer, NY. <http://www.asdar-book.org/>
- 451 Bivand R. & Rundel C. (2013) *rgeos: Interface to Geometry Engine - Open Source (GEOS)*. R
452 package version 0.3-2. <http://CRAN.R-project.org/package=rgeos>
- 453 Booth, R. K., Jackson, S. T., Sousa, V. A., Sullivan, M. E., Minckley, T. A., & Clifford, M. J.
454 (2012). Multi-decadal drought and amplified moisture variability drove rapid forest community
455 change in a humid region. *Ecology*, **93**(2), 219-226.

- 1
2
3 456 Bouldin, J. (2008). Some problems and solutions in density estimation from bearing tree data: a
4
5 457 review and synthesis. *Journal of Biogeography*, **35**, 2000-2011.
6
7
8 458 Bourdo Jr, E.A. (1956) A review of the General Land Office survey and of its use in quantitative
9 studies of former forests. *Journal of Ecology*, **37**, 754–68.
10
11
12 460 Cogbill C.V., Burk J. & Motzkin G. (2003) The forests of presettlement New England, USA:
13 spatial and compositional patterns based on town proprietor surveys. *Journal of Biogeography*,
14
15 461 **29**, 1279-1304.
16
17
18 463 Cottam, G. & Curtis, J.T.. (1956) The use of distance measures in phytosociological sampling.
19
20
21 464 *Ecology*, **37**, 451–460.
22
23
24 465 Csardi, G., & Nepusz, T. (2006) *The igraph software package for complex network research*,
25
26
27 466 InterJournal, Complex Systems 1695. <http://igraph.sf.net>
28
29
30 467 Curtis, JT. 1959. *The vegetation of Wisconsin: an ordination of plant communities*. The
31 University of Wisconsin Press. pp. 657.
32
33
34 469 Danz, N. P., Reich, P. B., Frelich, L. E., & Niemi, G. J. (2011). Vegetation controls vary across
35 space and spatial scale in a historic grassland-forest biome boundary. *Ecography*, **34**, 402-414.
36
37
38 471 Duren, O. C., Muir, P. S., & Hosten, P. E. (2012). Vegetation Change from the Euro-American
39
40
41 472 Settlement Era to the Present in Relation to Environment and Disturbance in Southwest Oregon.
42
43
44 473 *Northwest Science*, **86**, 310-328.
45
46
47 474 Ellis, E. C., & Ramankutty, N. (2008). Putting people in the map: anthropogenic biomes of the
48 world. *Frontiers in Ecology and the Environment*, **6**, 439-447.
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 476 Foster, D. R., Motzkin, G., & Slater, B. (1998). Land-use history as long-term broad-scale
4
5 477 disturbance: regional forest dynamics in central New England. *Ecosystems*, **1**, 96-119.
6
7
8 478 Friedman, S. K., & Reich, P. B. (2005). Regional legacies of logging: departure from
9
10 presettlement forest conditions in northern Minnesota. *Ecological Applications*, **15**, 726-744.
11
12
13 480 Goosse, H., E. Crespin, S. Dubinkina, M.F. Loutre, M.E. Mann, H. Renssen, Y. Sallaz-Damaz,
14
15 481 & Shindell, D. (2012) The role of forcing and internal dynamics in explaining the "Medieval
16
17
18 482 Climate Anomaly". *Climate Dynamics*. **39**, 2847-2866.
19
20
21
22 483 Goring S., Pellatt, M.G., Lacourse T., Walker, I.R., Mathewes, R.W. (2009) A new methodology
23
24 for reconstructing climate and vegetation from modern pollen assemblages: an example from
25
26
27 485 British Columbia. *Journal of Biogeography*. **36**, 626-638.
28
29
30 486 Groffman, P.M., Rustad, L.E., Templer, P.H., Campbell, J.L., Christenson, L.M., Lany, N.K.,
31
32 487 Socci, A.M., Vadeboncoeur, M.A., Schaberg, P.G., Wilson, G.F., Driscoll, C.T., Fahey, T.J.,
33
34 488 Fisk, M.C., Goodale, C.L., Green, M.B., Hamburg, S.P., Johnson, C.E., Mitchell, M.J., Morse,
35
36 489 J.L., Pardo, L.H. & Rodenhouse, N. L. (2012). Long-term integrated studies show complex and
37
38 490 surprising effects of climate change in the northern hardwood forest. *BioScience*, **62**, 1056-1066.
39
40
41
42 491 Grossman, E.B. & Mladenoff, D.J. (2008) Farms, fires, and forestry: Disturbance legacies in the
43
44 soils of the Northwest Wisconsin (USA) Sand Plain. *Forest Ecology and Management*, **256**, 827-
45
46 493 836.
47
48
49
50 494 Hanberry, B. B., Fraver, S., He, H. S., Yang, J., Dey, D. C., & Palik, B. J. (2011) Spatial pattern
51
52 495 corrections and sample sizes for forest density estimates of historical tree surveys. *Landscape
53
54
55 496 ecology*, **26**(1), 59-68.
56
57
58
59
60

- 1
2
3 497 Hanberry, B. B., Palik, B. J., & He, H. S. (2012a) Comparison of historical and current forest
4 surveys for detection of homogenization and mesophication of Minnesota forests. *Landscape*
5
6 498
7
8 499 *ecology*, 27(10), 1495-1512.
9
10
11 500 Hanberry, B. B., Yang, J., Kabrick, J. M., & He, H. S. (2012b) Adjusting forest density estimates
12
13 501 for surveyor bias in historical tree surveys. *The American Midland Naturalist*, 167, 285-306.
14
15
16 502 Hartig, F., Dyke, J., Hickler, T., Higgins, S. I., O'Hara, R. B., Scheiter, S., & Huth, A. (2012)
17
18 503 Connecting dynamic vegetation models to data—an inverse perspective. *Journal of*
19
20 504 *Biogeography*, 39, 2240-2252.
21
22
23
24 505 Haxeltine A, Prentice IC. (1996) BIOME3: An equilibrium terrestrial biosphere model based on
25
26 ecophysiological constraints, resource availability, and competition among plant functional
27
28 507 types. *Global Biogeochemical Cycles*, 10, 693-709.
29
30
31
32 508 Heffernan J, Soranno P, Angilletta M, Buckley L, Dodds W, Harms T, Keitt T, Knapp A,
33
34 509 Kominoski J, McDowell B, Powell H, Xiao J, Gruner D, Goring S, Koenig L, Kellner J,
35
36 510 Richardson A, Rocha A, Stow C, Vargas R, Weathers K. What is macrosystems
37
38 511 ecology? *Frontiers in Ecology and the Environment*. 12, 5-14.
39
40
41
42 512 Hijmans, R.J. (2014) *raster: raster: Geographic data analysis and modeling*. R package version
43
44 513 2.2-12. <http://CRAN.R-project.org/package=raster>
45
46
47
48 514 Hotchkiss, S. C., Calcote, R., & Lynch, E. A. (2007). Response of vegetation and fire to Little
49
50 515 Ice Age climate change: regional continuity and landscape heterogeneity. *Landscape*
51
52
53 516 *Ecology*, 22, 25-41.
54
55
56
57
58
59
60

- 1
2
3 517 Hurtt, G. C., Fisk, J., Thomas, R. Q., Dubayah, R., Moorcroft, P. R., & Shugart, H. H. (2010).
4
5 518 Linking models and data on vegetation structure. *Journal of Geophysical Research:*
6
7 519 *Biogeosciences (2005–2012)*, **115**(G2).
8
9
10 520 Iverson, L. R., & McKenzie, D. (2013) Tree-species range shifts in a changing climate:
11
12 521 detecting, modeling, assisting. *Landscape Ecology*, **28**, 879-889.
13
14
15 522 Jenkins JC, Choknacky DC, Heath LS, Birdsey RA. (2004) *Comprehensive database of*
16
17 523 *diameter-based biomass regressions for North American tree species*. Newton Square, PA:
18
19 524 USDA Forest Service. GTR-NE-319.
20
21
22
23
24 525 Keitt TH, Bivand R, Pebesma E, Rowlingson B. (2012) *rgdal: Bindings for the Geospatial Data*
25
26 526 *Abstraction Library*. R package version 0.7-18. <http://CRAN.R-project.org/package=rgdal>
27
28
29
30 527 Kronenfeld, B. J., & Wang, Y. C. (2007) Accounting for surveyor inconsistency and bias in
31
32 estimation of tree density from presettlement land survey records. *Canadian Journal of Forest*
33
34 529 *Research*, **37**, 2365-2379.
35
36
37
38 530 Leahy, M. J., & Pregitzer, K. S. (2003). A comparison of presettlement and present-day forests
39
40 in northeastern lower Michigan. *The American Midland Naturalist*, **149**, 71-89.
41
42
43 532 Liu, F., Mladenoff, D.J., Keuler, N.S., & Moore, L.S. (2011) Broadscale variability in tree data
44
45 533 of the historical Public Land Survey and its consequences for ecological studies. *Ecological*
46
47 534 *Monographs*, **81**, 259-275.
48
49
50
51 535 Loutre, M.F., Mouchet, A., Fichefet, T., Goosse, H., Goelzer, H. & Huybrechts, P. (2011)
52
53 536 Evaluating climate model performance with various parameter sets using observations over the
54
55 537 recent past. *Climate of the Past*, **7**, 511-526.
56
57
58
59
60

- 1
2
3 538 Manies, K.L. & Mladenoff, D.J. 2000. Testing methods to produce landscape-scale presettlement
4
5 539 vegetation maps from the U.S. public land survey records. *Landscape Ecology*, **15**, 741-754.
6
7
8 540 Manies, K.L., Mladenoff, D.J. & Nordheim, E.V. 2001. Assessing large-scale surveyor
9
10 variability in the historic forest data of the original U.S. Public Land Survey. *Canadian Journal
11*
12
13 541 *of Forest Research*, **31**, 1719-1730.
14
15
16 543 Mladenoff, D.J., Dahir, S.E., Nordheim, E.V., Schulte, L.A. & Guntenspergen, G.G. (2002)
17
18 544 Narrowing historical uncertainty: Probabilistic classification of ambiguously identified tree
19
20 species in historical survey data. *Ecosystems*, **5**, 539-553.
21
22
23 546 Moorcroft, P.R., Hurtt, G.C. & Pacala, S.W. (2001) A method for scaling vegetation dynamics:
24
25 547 The Ecosystem Demography model (ED). *Ecological Monographs*, **71**, 557-586.
26
27
28 548 Morisita, M. (1954) Estimation of population density by spacing method. *Kyushu
29*
30
31 549 *University Faculty Science Memorial Series E*, **1**, 187-197.
32
33
34 550 Munoz, S.E., Mladenoff, D.J., Schroeder, S. & Williams, J.W. (In review). Defining the spatial
35
36 patterns of land use associated with the indigenous societies of eastern North America. *Journal
37*
38
39 552 *of Biogeography*.
40
41
42 553 Notaro, M. & Gutzler, D. (2011). Simulated impact of vegetation on climate across the North
43
44 American monsoon region in CCSM3.5. *Climate Dynamics*, **38**, 795-814.
45
46
47 555 Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson,
48
49
50 556 G.L., Solymos, P., Stevens, M.H.H. & Wagner, H. (2013) *vegan: Community Ecology Package*.
51
52
53 557 R package version 2.0-10. <http://CRAN.R-project.org/package=vegan>
54
55
56
57
58
59
60

- 1
2
3 558 Paciorek, C.P. & McLachlan, J.S. (2009) Mapping ancient forests: Bayesian inference for spatio-
4 temporal trends in forest composition using the fossil pollen proxy record. *Journal of the*
5
6 559
7
8 560 *American Statistical Association*. **104**, 608-622
9
10
11 561 Pebesma, E.J. & Bivand R.S. (2005) Classes and methods for spatial data in R. *R News*, **5**, 9-13.
12
13
14 562 Pederson, D.C., Peteet, D.M., Kurdyla, D. & Guilderson, T. (2005). Medieval Warming, Little
15
16 563 Ice Age, and European impact on the environment during the last millennium in the lower
17
18 564 Hudson Valley, New York, USA. *Quaternary Research*, **63**, 238-249.
19
20
21
22 565 R Core Team (2013) *R: A language and environment for statistical computing*. R Foundation for
23
24 566 Statistical Computing, Vienna, Austria.
25
26
27 567 Ramankutty, N. & Foley, J.A. (1999) Estimating historical changes in global land cover:
28
29 568 croplands from 1700 to 1992. *Global Biogeochemical Cycles*, **13**, 997-1027.
30
31
32 569 Ratajczak, Z., Nippert, J.B. & Collins, S.L. (2012) Woody encroachment decreases diversity
33
34 570 across North American grasslands and savannas. *Ecology*, **93**(4), 697-703.
35
36
37
38 571 Rayburn, A. P. & Schulte, L. A. (2009) Integrating historic and contemporary data to delineate
39
40 572 potential remnant natural woodlands within Midwestern agricultural landscapes. *Natural Areas
41
42
43
44
45
46
47 573 Journal*, **29**, 4-14.
48
49
50
51 574 Rhemtulla, J.M., Mladenoff, D.J. & Clayton, M.K. (2009a). Historical forest baselines reveal
52
53
54
55
56
57
58
59
60
575 potential for continued carbon sequestration. *Proceedings of the National Academy of Sciences*,
576 **106**, 6082-6087.

- 1
2
3 577 Rhemtulla, J.M., Mladenoff, D.J. & Clayton, M.K. 2009b. Legacies of historical land use on
4 regional forest composition and structure in Wisconsin, USA (mid-1800s to 1930s to 2000s).
5
6 578
7
8 579 *Ecological Applications*, **19**, 1061-1078.
9
10
11 580 Richardson, A.D., Williams, M., Hollinger, D.Y., Moore, D.J., Dail, D.B., Davidson, E.A., Scott,
12
13 581 N.A., Evans, R.S., Hughes, H., Lee, J.T., Rodrigues, C. & Savage, K. (2010). Estimating
14 parameters of a forest ecosystem C model with measurements of stocks and fluxes as joint
15 constraints. *Oecologia*, **164**, 25-40.
16
17
18 583
19
20
21 584 St. Jacques, J.M., Cumming, B.F., & Smol, J.P. (2008). A pre-European settlement pollen–
22
23 climate calibration set for Minnesota, USA: developing tools for palaeoclimatic
24 reconstructions. *Journal of Biogeography*, **35**, 306-324.
25
26
27
28
29 586 Schulte, L.A. & Mladenoff, D.J. (2001) The original US Public Land Survey Records: Their use
30
31 588 and limitations in reconstructing presettlement vegetation. *Journal of Forestry*. **99**, 5-10.
32
33
34 589 Schulte, L.A. & Mladenoff, D.J. (2005) Severe wind and fire regimes in northern forests:
35
36 590 Historical variability at the regional scale. *Ecology* **86**, 431-445
37
38
39
40 591 Schulte, L.A., Mladenoff, D.J., Crow, T.R., Merrick, L.C. & Cleland, D.T. (2007)
41
42 592 Homogenization of northern U.S. Great Lakes forests due to land use. *Landscape Ecology*, **22**,
43
44 593 1089-1103.
45
46
47
48 594 Schulte, L.A., D.J. Mladenoff & E.V. Nordheim. (2002) Quantitative classification of a historic
49
50 595 northern Wisconsin (U.S.A.) landscape: mapping forests at regional scales. *Canadian Journal of
51
52 596 Forest Research*, **32**, 1616-1638.
53
54
55
56
57
58
59
60

- 1
2
3 597 Smith, B., Prentice, I.C. & Sykes, M.T. (2002) Representation of vegetation dynamics in the
4 modeling of terrestrial ecosystems: comparing two contrasting approaches within European
5
6 598 climate space. *Global Ecology and Biogeography*, **10**, 621-637.
7
8 599
9
10 600 Staver, A.C., Archibald, S. & Levin, S.A. (2011). The global extent and determinants of savanna
11 and forest as alternative biome states. *Science*, **334**, 230-232.
12
13 601
14
15 602 Stewart, L. O. (1979) *Public land surveys*. Arno Press.
16
17 603 Stoy, P.C., Mauder, M., Foken, T., Marcolla, B., Boegh, E., Ibrom, A., Arain, M.A., Arneth, A.,
18
19 604 Aurela, M., Bernhofer, C., Cescatti, A., Dellwik, E., Duce, P., Ganelle, D., van Gorsel, E.,
20
21 605 Kiely, G., Knohl, A., Margolis, H., McCaughey, H., Merbold, L., Montagnani, L., Papale, D.,
22
23 606 Reichstein, M., Saunders, M., Serrano-Ortiz, P., Sottocornola, M., Spano, D., Vaccari, F. &
24
25 607 Varlagin, A. (2013). A data-driven analysis of energy balance closure across FLUXNET
26
27 research sites: The role of landscape scale heterogeneity. *Agricultural and Forest Meteorology*,
28
29 608
30 609 **171**, 137-152.
31
32
33
34 610 Sugita, S., (2007a) Theory of quantitative reconstruction of vegetation. I. Pollen from large sites
35
36 611 REVEALS regional vegetation. *The Holocene*, **17**, 229–241.
37
38
39 612 Sugita, S. (2007b) Theory of quantitative reconstruction of vegetation. II. All you need is LOVE.
40
41 613 *The Holocene*, **17**, 243–257.
42
43
44 614 Thompson, J. R., Carpenter, D. N., Cogbill, C. V., & Foster, D. R. (2013) Four centuries of
45
46 615 change in northeastern United States forests. *PloS one*, **8**, e72540.
47
48
49 616 Umbanhowar CE (2004) Interaction of fire, climate and vegetation change at a large landscape
50
51 617 scale in the Big Woods of Minnesota. *Holocene*, **14**, 661–676
52
53
54
55
56
57
58
59
60

1
2
3 618 Urban, N.M. & Keller K. (2010) Probabilistic hindcasts and projections of the coupled climate,
4 carbon cycle, and Atlantic meridional overturning circulation systems: A Bayesian fusion of
5
6 619 century-scale observations with a simple model. *Tellus A*, **62**, 737-750.
7
8 620
9
10

11 621 White, C. A. (1983) *A history of the rectangular survey system*. US Department of the Interior,
12
13 622 Bureau of Land Management.
14
15

16 623 Wickham, H. (2007) Reshaping data with the reshape package. *Journal of Statistical Software*,
17
18 624 **21**, 1-20.
19
20

21 625 Wickham, H. (2009) *ggplot2: elegant graphics for data analysis*. Springer New York.
22
23

24 626 Wickham, H. (2011) The split-apply-combine strategy for data analysis. *Journal of Statistical
25 Software*, **40**, 1-29.
26
27
28

29 628 Williams, M. A. & Baker, W. L. (2011) Testing the accuracy of new methods for reconstructing
30 historical structure of forest landscapes using GLO survey data. *Ecological Monographs*, **81**, 63-
31
32 629
33
34 630 88.
35
36
37

38 631 Wood, S.N. (2006) *Generalized Additive Models: An Introduction with R*. Chapman and
39
40 632 Hall/CRC
41
42

43 633 Woudenberg, S.W., Conkling, B.L., O'Connell, B.M., LaPoint, E.B., Turner, J.A., & Waddell,
44
45 634 K.L. (2010) *The forest inventory and analysis database description and user's manual version
46 4.0 for phase 2*. General Technical Report RMRS-GTR-245. US Department of Agriculture,
47
48 635
49
50 636 Forest Service, Rocky Mountain Research Station, Fort Collins, CO.
51
52
53
54
55
56
57
58
59
60

1
2
3 637 **Biosketch:**
4
5
6 638 Simon Goring is a postdoctoral researcher at the University of Wisconsin studying interactions
7
8 639 between vegetation, climate and land use change in the near-modern and Holocene. PaleON
9
10 640 (the PaleoEcological Observatory Network, <http://www.paleonproject.org>) is an interdisciplinary
11
12 641 team of paleoecologists, ecological statisticians, and ecosystem modelers. Our goal is to
13
14 642 reconstruct forest composition, fire regime, and climate in forests across the northeastern US and
15
16 643 Alaska over the past 2000 years and then use this to drive and validate terrestrial ecosystem
17
18 644 models.
19
20
21 645
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

1
 2
 3 646 **Table 1.** Forest classification scheme used in this paper for comparison between pre-
 4 settlement forests and the Haxeltine and Prentice (1996) potential vegetation classes
 5 represented in Ramankutty and Foley (1999). Plant functional types (PFTs) for grasslands
 6 (CG, grassland; Non-Tree samples in the PLS), broad leafed deciduous taxa (BD) and
 7 needleleaded evergreen taxa (NE) are used, but leaf area index (LAI) used in Haxeltine and
 8 Prentice (1996) is replaced by stem density classes from Anderson and Anderson (1975).
 9
 10

Forest Class	Haxeltine & Prentice Rules	Current Study
Prairie	Dominant PFT CG, LAI > 0.4	Stem dens. < 1 stem/ha
Savanna	Dominant PFT CG, LAI > 0.6	1 < Stem dens. < 19 stems/ha
Temperate Deciduous	Dominant PFT BD, LAI > 2.5	Stem dens. > 19 stems/ha, BD > 70% composition
Temperate Conifer	Dominant PFT (NE + ND), LAI > 2.5	Stem dens. > 19 stems/ha, NE + ND > 70% composition
Mixedwood	Both BD (LAI > 1.5) & NE (LAI > 2.5) present	Stem dens. > 19 stems/ha, BD & NE both < 70% composition

24
 25 652
 26
 27 653 **Table 2.** Classification proportions, patch size and edge cell proportion for various
 28 classification schemes used with the Public Land Survey data. All patch size estimates are in
 29 1000s of km².
 30

Classification Metric	Prairie	Savanna	Temperate Deciduous	Temperate Evergreen	Mixed Wood	Mean Patch Size (km ²)	Edge Cells (%)
Ramankutty and Foley (1999)	86	63	9	50	241	184	27.5
PLS Data	57	49	117	150	101	34	72.3

42 656
 43
 44

45 657
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60

1
2
3 658 **Figure 1.** The domain of the Public Land Survey investigated in this study. The broad domain
4 659 includes Minnesota, Wisconsin and the upper two thirds of Michigan state.
5
6
7 660 **Figure 2.** Total stem density (a) in the Upper Midwest, along with forest type classification
8 661 (b) based on Anderson and Anderson (1975; Table 2). Fine lines represent major rivers. Basal
9 662 area (c) and biomass (d) show similar patterns to stem density.
10
11 663 **Figure 3.** Variations in the relationship between biomass and stem density can be used to
12 664 understand forest structure. Regions with high stem density to biomass ratios (light gray)
13 665 indicate dense stands of smaller trees, while regions with low stem density to biomass ratios
14 666 (darkest gray) indicate larger trees with wider spacings. Only cells greater than 1 standard
15 667 deviation from the ratio mean are classified as low or high.
16
17 668 **Figure 4.** Forest composition as a percent for the 15 most abundant tree taxa. The scale is
18 669 drawn using a square-root transform to emphasize low abundances. Shading of the bar above
19 670 individual taxon maps indicates plant functional type assignments (white: needleleafed
20 671 deciduous; light gray: needleleafed evergreen; dark gray: broadleafed deciduous).
21
22 672 **Figure 5.** Proportional distribution of Plant Functional Types (PFTs) in the upper Midwest
23 673 from PLS data, for temperate broadleaved deciduous trees (BD), needleleafed deciduous trees
24 674 (ND), and needleleafed evergreen trees (NE). Distributions are shown as proportions relative
25 675 to total basal area, total biomass, and composition, but do not indicate the magnitude of each
26 676 of these estimates. The grassland PFT is determined by non-tree counts with the assumption
27 677 that if trees were available surveyors would have sampled them.
28
29
30
31 678 **Figure 6.** Forest classification using Ramankutty and Foley (1999) estimates and those using
32 679 baseline PLS data (aggregated to PFTs for equivalence) with the Haxeltine and Prentice
33 680 (1996) classification scheme (Table 2). Colors represent forest classes for the region as
34 681 defined by Ramankutty and Foley (1999): (Pr) Prairie; (Sa) Savanna; (TD) Temperate
35 682 deciduous; (TE) Temperate evergreen; (MW) Mixedwood forest.
36
37
38 683 **Figure 7.** The relationship between average stem density, total basal area and biomass values
39 684 in the PLS and FIA datasets. Stem density and total basal area are higher in the FIA than in
40 685 the PLS, however average cell biomass is higher in the PLS.
41
42
43 686 **Figure 8.** Minimum dissimilarity maps. Cells with a high minimum dissimilarity lack close
44 687 compositional analogs to other cells in the reference dataset. The top two panels show
45 688 compositional heterogeneity within the PLS and FIA data, while the bottom two panels
46 689 identify PLS locations with no close analogs in the FIA data (PLS to FIA) and FIA locations
47 690 with no close PLS analogs (FIA to PLS). More than half (55%) of points in the PLS have
48 691 minimum dissimilarities to the FIA data that are greater than the 95%ile for minimum
49 692 dissimilarities within either the FIA or PLS datasets.
50
51
52
53
54
55
56
57
58
59
60

1
2
3 The ZIP file containing the code and raw data is too large for upload.
4
5 The link to this file is here:
6 <https://www.dropbox.com/s/osgggtt59arg0f8r/Supplement1.rar>
7
8 The file is 299Mb in size. RAR files can be extracted using most common
9 software applications, including WinRAR (for Windows).

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review