Applied Geography 46 (2014) 21–34

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

A comparison of techniques for generating forest ownership spatial products

Brett J. Butler [a,*], Jaketon H. Hewes [b], Greg C. Liknes [c], Mark D. Nelson [c], Stephanie A. Snyder [c]

[a] USDA Forest Service, Northern Research Station, 160 Holdsworth Way, Amherst, MA 01003, USA
[b] USDA Forest Service/University of Massachusetts Amherst Family Forest Research Center, 160 Holdsworth Way, Amherst, MA 01003, USA
[c] USDA Forest Service, Northern Research Station, 1992 Folwell Avenue, St. Paul, MN 55108, USA

Keywords:
Thiessen polygon
Multinomial logit
Classification tree
United States
Forest inventory and analysis
Protected areas database

A B S T R A C T

To fully understand forest resources, it is imperative to understand the social context in which the forests exist. A pivotal part of that context is the forest ownership. It is the owners, operating within biophysical and social constraints, who ultimately decide if the land will remain forested, how the resources will be used, and by whom. Forest ownership patterns vary substantially across the United States. These distributions are traditionally represented with tabular statistics that fail to capture the spatial patterns of ownership. Existing spatial products are not sufficient for many strategic-level planning needs because they are not electronically available for large areas (e.g., parcels maps) or do not provide detailed ownership categories (e.g., only depict private versus public ownership). Thiessen polygon, multinomial logit, and classification tree methods were tested for producing a forest ownership spatial dataset across four states with divergent ownership patterns: Alabama, Arizona, Michigan, and Oregon. Over 17,000 sample points with classified forest ownership, collected as part of the USDA Forest Service, Forest Inventory and Analysis (FIA) program, were divided into two datasets, one used as the dependent variable across all of the models and 10 percent of the points were retained for validation across the models. Additional model inputs included a polygon coverage of public lands from the Conservation Biology Institute's Protected Areas Database (PAD) and data representing human population pressures, road densities, forest characteristics, land cover, and other attributes. The Thiessen polygon approach predicted ownership patterns based on proximity to the sample points in the model dataset and subsequent combining with the PAD ownership data layer. The multinomial logit and classification tree approaches predicted the ownership at the validation points based on the PAD ownership information and data representing human population, road, forest, land cover, and other attributes. The percentage of validation points across the four states correctly predicted ranged from 76.3 to 78.9 among the methods with corresponding weighted kappa values ranging from 0.73 to 0.76. Different methods performed slightly, but statistically significantly, better in different states Overall, the Thiessen polygon method was deemed preferable because: it has a lower bias towards dominant ownership categories; requires fewer inputs; and is simpler to implement.

Introduction

The United States is endowed with an estimated 304 million ha (751 million ac) of forest land covering 33 percent of the nation's land area (Smith, Miles, Perry, & Pugh, 2009). These forests provide a plethora of goods and services, including wood products, wildlife habitat, recreational opportunities, water purification, and carbon storage (Chopra & Dasgupta, 2008). Ownership has been shown to be an important determinant and factor in numerous natural resource issues (Jaimes, Sendra, Delgado, & Plata, 2010; Paudel & Thapa, 2004; Serra, Pons, & Saurí, 2008). The use and disposition

Abbreviations: FIA, Forest inventory and analysis program of the United States Department of Agriculture Forest Service; NLCD, National Land Cover Database; PAD, Protected Areas Database; USDA-FS, United States Department of Agriculture Forest Service.

* Corresponding author. Tel.: +1 413 545 1387; fax: +1 413 545 1860.
E-mail addresses: bbutler01@fs.fed.us (B.J. Butler), jhewes@eco.umass.edu (J. H. Hewes), gliknes@fs.fed.us (G.C. Liknes), mdnelson@fs.fed.us (M.D. Nelson), stephaniesnyder@fs.fed.us (S.A. Snyder).

0143-6228/$ – see front matter Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.apgeog.2013.09.020

of forest resources is dependent upon those who control it – the forest owners – operating within the biophysical constraints of the land and the economic, regulatory, and normative constraints of society. The forests of the United States are owned by a diversity of entities including: federal, state, and local governments; private corporations; families; individuals; Native American tribes; and other groups. Ownership goals, management practices, and applicable regulations can vary widely among these entities (Bengston, Asah, & Butler, 2011; U.S. Department of Agriculture Forest Service, 2011). Many of the private lands are under growing pressure due to parcellation, fragmentation, and development (Jin & Sader, 2006; White, Alig, & Stein, 2010).

Forest ownership patterns vary substantially across the U.S., at both coarse and fine scales. For example, at the coarsest spatial and thematic scales (i.e., broad ownership categories), the forests of the eastern U.S. are 81 percent *privately* owned versus the forests of the western U.S. which are 70 percent *publicly* owned (Smith et al., 2009). At finer scales, these patterns can still be highly variable with intermingling of ownership types and resulting implications for forest policy, industry, conservation, and society. From a policy perspective, it is important to understand which tools to use where and how these tools will interact (Harper et al., 2006) – for example, different policy tools are used to mitigate wildland fire depending on the ownership patterns. From a forest industry perspective, it is important to accurately predict the supply of raw materials, be it for lumber, biomass, or other end uses, and this depends, in part, on landowners' objectives and constraints (Butler, Ma, Kittredge, & Catanzaro, 2010; Polyakov, Wear, & Huggett, 2010). From a conservation perspective, it is important to know which lands are already protected, which are most threatened, and where the greatest opportunities for land conservation exist (Stein et al., 2005). For many private citizens, outdoor recreation is important (Cordell, Betz, & Green, 2008) and the accessibility of lands depends, to a large extent, on who owns it (Snyder & Butler, 2012; Snyder, Kilgore, Taff, & Schertz, 2008). All of these issues are strongly tied to forest ownership patterns and therefore, understanding and mapping forest ownership patterns can facilitate more informed decisions to help maintain forests and the social benefits they provide. Previous studies have focused on mapping of some social dimensions related to forests (Brown & Raymond, 2007; Sherrouse, Clement, & Semmens, 2011), but not ownership.

Like all land use patterns, land ownership patterns are not random. The specific use of a given piece of ground is a function of social, economic, political, historical, and environmental factors (van Kooten, 1993). It has been posited that land ownership patterns are the results of similar factors and a modified land rent theory can be applied (Hardie, Parks, Gottleib, & Wear, 2000; Wear, 2011). Land rent theory states that a given piece of land is allotted to its highest and best use based upon the demands from society (e.g., distance from population centers) and the characteristics of the land (e.g., suitability of the land for development) (van Kooten, 1993). Using analogous logic, forest land ownership patterns can also be thought of as being a function of the demands of society and the inherent characteristics of the land (Wear, 2012).

Land rent theory, while a useful construct, is, as with all theories, a simplification of the underlying processes. There are countless other factors that also impact the ultimate land use and ownership patterns, including historical factors. Banner (2011) provides an overview of the historical ownership patterns of the U.S. and discusses the country's adaptation of the British ownership system. As the Euro-Americans progressed across the U.S., they brought with them their ownership systems. From the meets and bounds measurements of the eastern U.S., a more systematic land division system, the Public Land Survey System, was authorized by the Land Ordinance of 1785 and first implemented in Ohio. This system resulted in the very ordered, checkerboard patterns of land ownership that exist across many parts of the country.

Homesteading acts, such as the Homestead Act of 1862, have played an important role in the legitimization and distribution of lands across much of the U.S. Individuals meeting certain requirements were able to claim lands, but if they were unable to "prove their claims" or became delinquent on taxes, the land, by default, reverted to public ownership. These lands were then sold or retained by the government, many of which became the basis for the federal holdings in the western U.S. The government, at federal-, state-, and local-levels, has also actively procured land. For example, the Weeks Act of 1911 and the Organic Act of 1916 set out to place large swaths of undeveloped land into the hands of the federal government, and as a result, millions of hectares of forest land were put into federal ownership. Land grants were used as a way to encourage the western expansion of railroads and funding for school systems. Many of these allotments were made over 100 years ago, but many still persist and the effects of others, e.g., the checkerboard ownership pattern, are still very apparent.

Private companies and individuals have acquired forest land through other mechanisms besides just land grants and homesteading; namely purchases and, for some individuals, inheritances. For forest lands, it was initially companies operating sawmills, pulp mills, and other wood consuming factories that acquired large acreages. These lands tended to be in more rural areas. In more recent times, there has been large-scale divestiture of forest lands by traditional, vertically-integrated forest companies with many of these acreages now being owned by timber investment management organizations and real estate investment trusts (Zhang, Butler, & Nagubadi, 2012). Although many large parcels of forest land are owned by these corporations and their investors, private citizens make up the vast majority of forest land owners in the U.S. Ownership of land by individuals has long been an American ideal and millions of Americans now own forest land, predominately for privacy, esthetic, and family legacy reasons (Butler, 2008).

Currently, forest ownership data is available in aspatial or coarse spatial formats, with fine-scale spatial data having limited availability. Tabular data are available from the USDA-FS (e.g., Smith et al., 2009), but lack spatial detail beyond state- or county-level summaries. The current spatial datasets are incomplete in extent and/or thematic detail (i.e., only report broad ownership categories, such as public versus private). Detailed, parcel-level maps in electronic formats, while available in some areas of the country, are not available for many locales. Moreover, the available datasets are often in different and/or inconsistent formats, are not in a centralized location, can cost substantial amounts of money to acquire, and can be difficult to work with. Some commercial sources have aggregated these data, but these sources can be expensive and still exclude large swaths of the U.S. The Protected Areas Database (PAD; Conservation Biology Institute (CBI), 2010) is a national geospatial database established to document the locations of protected lands across the U.S. and includes information on ownership of protected lands. PAD is freely available for public use, and some elements of PAD are included in this study. Ongoing efforts provide periodic updates to PAD, and references here pertain to PAD-US (CBI Edition) v1.1. PAD is built by assembling spatial data from public agencies from across the country; the accuracy of PAD varies among states depending on the quality of these input data. Private ownership categories are included only for private protected areas, which comprise a small minority of all private forest holdings in the U.S. A national database of conservation easements (www.conservationeasement.us) has been created, but this source focuses only on the small segment of the private land that is under easements and is not useful for mapping broad ownership patterns. A national database of land ownership parcels is not yet available, other data sources, such as PAD, do not provide the requisite resolution of ownership categories for private lands, and hence there is a need for forest ownership spatial data products.

There are at least two previous efforts that generated forest ownership maps for the conterminous U.S., both of which relied on PAD for delineating public lands. Butler (2008) used PAD to define public ownership and NLCD to define forest land to produce a map depicting two broad ownership categories, private and public. This product is available only as an image, i.e., no geospatial dataset was published. Nelson, Liknes, and Butler (2010b) expanded on this effort by differentiating between federal and other government forest lands on the public side and, using FIA data, depicting percent corporate ownership within 648 km^2 (250 mi^2) hexagons. A geospatial dataset for this products was published (Nelson, Liknes, & Butler, 2010a). These products could be improved by providing more detailed ownership categories and increased spatial resolutions.

This paper presents methods for producing spatial products that depict forest ownership patterns using FIA, PAD, and other data sources. FIA collects forest attributes, including ownership class, from a random set of points across the U.S. (Bechtold & Patterson, 2005). There are examples of techniques that take point-based estimates, often FIA data, and, combined with other data, create spatially continuous data layers of biophysical forest attributes. Ohmann and Gregory (2002) used canonical correspondence analysis and nearest-neighbor imputation to relate FIA plot attributes with topographic, geologic, climatic, satellite, and location variables and interpolate plot attributes across the landscape. Wilson, Lister, and Riemann (2012) used a similar approach. Blackard et al. (2008) used a regression tree approach to model biomass across the U.S. based on FIA sample point locations and ancillary data including satellite, topographic, climatic, soils, geologic, ecological, and social variables. A classification tree approach was also used to map forest types across the conterminous United States and Alaska (Ruefenacht et al., 2008). These approaches appear to work reasonably well for mapping biophysical forest attributes, but are untested for mapping social attributes such as ownership.

The primary objective of this paper is to present techniques for generating a forest ownership geospatial dataset for the U.S. The methods selected must be operationally feasible for application across the conterminous U.S. and result in forest ownership spatial products with finer spatial, thematic, and temporal resolutions than are currently available. This product will build upon existing efforts, such as PAD. The resulting forest ownership geospatial dataset should be able to address strategic level questions related to forest policy, business, and conservation. The current tabular forest ownership products do not provide sufficient information to meet the needs outlined above nor do the current geospatial products. The spatial products resulting from the techniques identified in this paper should be well suited for: visually depicting broad ownership patterns across national, regional, or state scales; incorporation into strategic forest assessments, such as state forest assessments, as mandated under the 2008 Farm Bill (P.L. 110–234 § 8002); inclusion in national and state forest statistical reports (e.g., Smith et al., 2009); and incorporation into other spatial analyses where these spatial and thematic scales are appropriate. A forest ownership map for the U.S. is not included as part of this paper. The purpose of this paper is to test different techniques for generated such a map. The full map will be published separately as a cartographic product

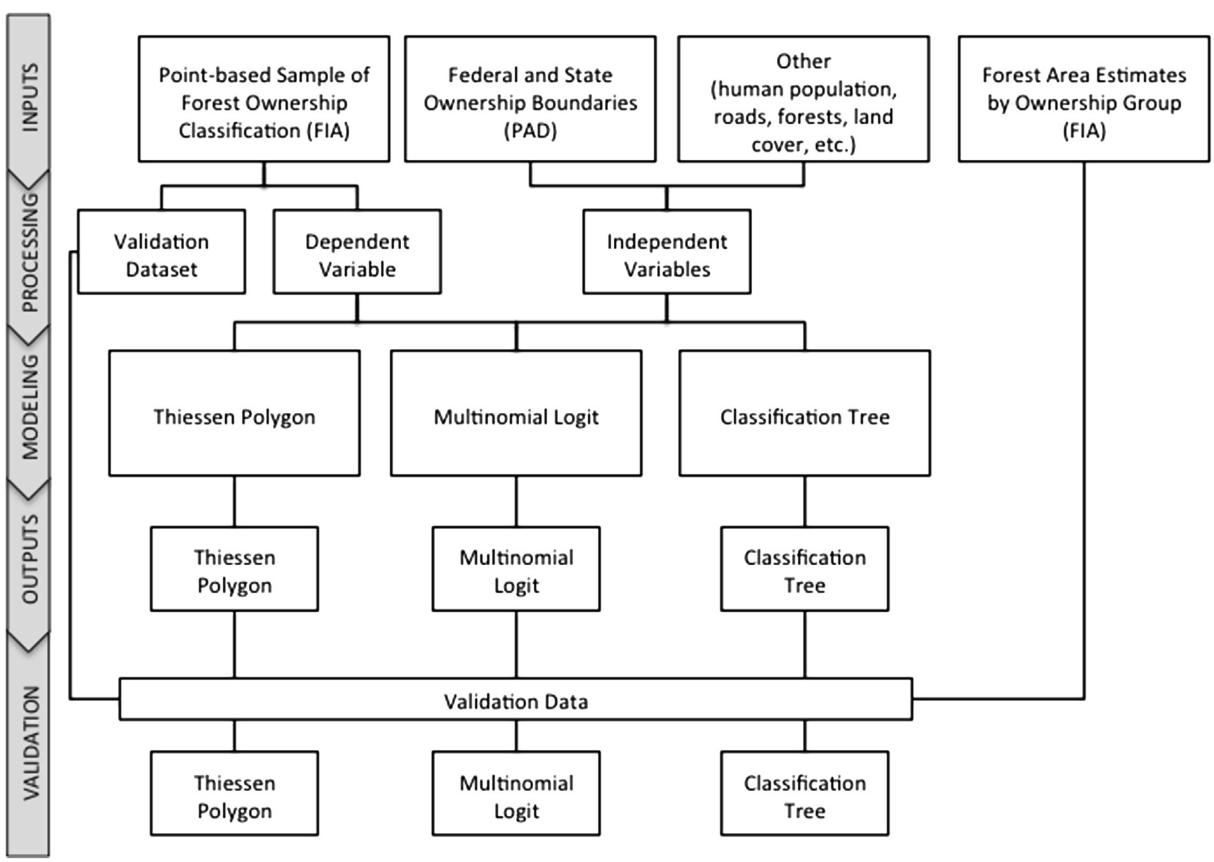

Fig. 1. Overview of approach used in this paper to compare techniques for generating a forest ownership spatial product.

(Hewes, Butler, Liknes, Nelson, & Snyder, 2013b) and a downloadable spatial dataset (Hewes, Butler, Liknes, Nelson, & Snyder, 2013a).

Methods

Using data from FIA and other sources, three methods were investigated as techniques for creating a forest ownership geospatial dataset: Thiessen polygons, multinomial logit, and classification tree. These methods were selected because of their previous applications in extrapolating FIA data and their diversity of approaches (i.e., simplistic to complex). Each method was tested using identical inputs and validation procedures in four states: Alabama, Arizona, Michigan, and Oregon. These states were selected because they represent different regions of the U.S. and a variety of ownership (e.g., public versus private) and land use (e.g., forest versus non-forest) patterns based on statistics published by the USDA-FS (Smith et al., 2009). An outline of the general approach used in this paper is provided in Fig. 1. After describing the data sources, each approach is discussed along with methods for comparing the results.

Data sources

FIA has established a set of permanent inventory plots across the U.S. (Bechtold & Patterson, 2005). A hexagonal sampling frame, with each hexagon approximately 2428 ha (6000 ac), was generated across the U.S. and within each hexagon a sample point was randomly located. Inventory plots at each sample point are remeasured every 5–7 years in the East and every 10 years in the West. FIA uses remote sensing and subsequent ground-truthing to determine if a plot is likely to contain any forest land. Forest is defined as "land at least 120 feet wide and 1 acre in size with at least 10 percent cover (or equivalent stocking) by live trees of any size, including land that formerly had such tree cover and that will be naturally or artificially regenerated" (Smith et al., 2009 p. 142) – a land use, rather than a land cover, definition. For the sample points that are deemed to be forested, the ownership is determined from public property tax records and classified into 1 of 16 categories (Table 1). A natural resource professional, such as a forester or ecologist, visits each forested plot to collect information on the species, size, and health of trees in addition to various environmental conditions. This information is used to make tabular estimates, including area by forest ownership category (e.g., Table 2 in Smith et al., 2009), and for the biophysical data, raster spatial products using interpolation techniques (Wilson et al., 2012).

The ownership information for the sample points that were determined to be forested formed the underlying data for the forest ownership modeling and validation datasets for this study. These sample points represent the centers of the FIA inventory plots. This study utilized: 3768 forested points from Alabama; 1716 from Arizona; 9274 from Michigan; and 2535 from Oregon with all points inventoried between 2003 and 2007. The differences in numbers of sample points per state are due to differences in forest area and differences in spatial and temporal sampling intensities which vary due to funding availability.

The other major source of ownership information was PAD (Conservation Biology Institute (CBI), 2010). This source is designed primarily to identify the locations of protected lands across the U.S. It also includes information on ownership, but only for those protected lands that are in the database, which are primarily publicly owned. There is a high degree of agreement between PAD and FIA for federally owned forest land, a moderate agreement for state owned forest land, and a low degree of agreement for forest land owned by local governments, but agreement varies considerably among states (Table 2). Combining the federal and state ownership data from PAD with the model outputs improved the accuracies and was therefore incorporated as described below.

Table 1
Forest ownership categories as defined by FIA (USDA-FS, 2010) and the broader categories used for generating a forest ownership geospatial product.

Ownership geospatial product category	FIA code	Description
Federal		
	11	National Forest
	12	National Grassland
	13	Other Forest Service
	21	National Park Service
	22	Bureau of Land Management
	23	Fish and Wildlife Service
	24	Departments of Defense/Energy
	25	Other Federal
State		
	31	State
Local		
	32	Local (County, Municipality, etc.)
	33	Other Non Federal Public
Corporate		
	41	Corporate (including private educational institutions)
Other private		
	42	Non-Governmental Conservation/Natural Resources Organization
	43	Unincorporated Partnerships/Associations/Clubs
	44	Native American (Indian) – within reservation boundaries
Family		
	45	Family/individual

Data representing human population pressures, accessibility, forest attributes, surrounding land cover, and other factors hypothesized to influence forest ownership categories were included in the multinomial logit and classification tree models (Table 3). All data used had to be spatially explicit, have national coverage, be readily available, and be hypothesized to influence ownership patterns. Population/housing density is hypothesized to be a predictor because there are few, if any, people living on public or corporate lands. As population pressures increase, there may be two different consequences. For those lands deemed to have high conservation and/or social benefit values, the population pressures may shift land towards public ownerships. In the absence of public ownership, the likelihood of an ownership category that maximizes land sale profits, e.g., family and individual ownerships, may be favored. How forests are managed is an indication of ownership objectives, especially for industrial forest ownerships, and therefore forest type and basal area variables were included. The landscape context can influence, or be a result of, ownership patterns and therefore fragmentation and land cover variables were included. Overall, land values are a function of population pressures and the amenity values (Kline & Alig, 2005; Snyder, Kilgore, Hudson, & Donnay, 2008), so variables representing proximity to water and proximity to public lands were also tested.

Thiessen polygons

The Thiessen polygons, also known as Voronoi tessellation, approach assumes that the underlying pattern is non-random and the value of any unknown point can be predicted, to some degree, by its proximity to a known point (Okabe, Boots, Sugihara, & Chiu, 2000). In creating Thiessen polygons, areas are built around each point in the model dataset such that any location within a given polygon is closer to its associated point than to any other input

Table 2
Comparison of FIA (USDA-FS, 2010) and PAD (CBI, 2010) forest ownership classifications at random forest sample points in Alabama, Arizona, Michigan, and Oregon. *Acc* = overall accuracy (sum of the major diagonal); κ = unweighted kappa statistic; n = number of observations.

		PAD				Total
A. Alabama (*Acc* = 96.7; κ = 0.73; n = 3768)						
		Federal	State	Local	Private	
FIA	Federal	3.9	0.0	0.0	0.5	4.4
	State	0.1	0.8	0.0	0.6	1.5
	Local	0.0	0.0	0.0	0.5	0.5
	Private	0.2	1.4	0.0	92.0	93.6
	Total	4.2	2.1	0.0	93.6	100.0
B. Arizona (*Acc* = 95.7; κ = 0.92; n = 1716)						
		Federal	State	Local	Private	
FIA	Federal	53.0	0.1	0.0	1.3	54.4
	State	0.3	8.8	0.0	0.7	9.8
	Local	0.0	0.0	0.1	0.1	0.1
	Private	0.8	1.0	0.0	33.8	35.6
	Total	54.1	9.9	0.1	35.9	100.0
C. Michigan (*Acc* = 95.4; κ = 0.92; n = 9274)						
		Federal	State	Local	Private	
FIA	Federal	15.1	0.1	0.0	0.4	15.6
	State	0.2	21.1	0.0	1.1	22.4
	Local	0.0	0.1	0.0	1.7	1.8
	Private	0.4	0.7	0.0	59.2	60.2
	Total	15.6	22.0	0.0	62.4	100.0
D. Oregon (*Acc* = 67.6; κ = 0.35; n = 2535)						
		Federal	State	Local	Private	
FIA	Federal	46.2	0.5	0.0	15.9	62.6
	State	0.5	0.3	0.0	2.4	3.2
	Local	0.2	0.0	0.0	0.3	0.4
	Private	11.7	1.0	0.0	21.1	33.8
	Total	58.5	1.9	0.0	39.6	100.0

point. This approach is essentially a nearest neighbor interpolation in which the polygons created take on the attributes of the parent point, including, in this case, ownership category. All processing was done using the ArcMap 10.0 geographic information system (ESRI, Redlands, CA).

For each state, the FIA points in the model dataset were used to create the Thiessen polygons. First, federal and state FIA points that were located within the boundaries of federal and state ownership areas contained in the PAD were dropped from the model dataset because these PAD polygons were later burned into the final spatial data product. All other sample points in the model dataset, including federal and state FIA points not within PAD polygons, were used as inputs for the Thiessen polygons. After the Thiessen polygons were generated, the results were burned in or "unioned" with the PAD federal and state ownership polygons with the PAD data given precedence where there was overlap. Experimenting with different techniques for handling sample points that fell within the PAD polygons and subsequent inclusion of the PAD data showed the approach outlined here to provide the highest accuracies.

Multinomial logit model

A multinomial logit approach (Allison, 1999) was developed to test the likelihood that a sample point (or pixel) was in one of the six ownership categories. The resulting model could then be used to model the ownership categories for all pixels across the area of interest. In contrast to the Thiessen polygon approach which assumed that ownership category was only influenced by a geographically neighboring sample point, this method models ownership patterns as the result of social, economic and biophysical factors.

Solution of a multinomial logit model requires one of the possible outcomes of the dependent variable to be chosen as the reference level. For all of the runs reported here, the family ownership category was chosen as the reference level, and the models were solved comparing the likelihood of membership in each of the remaining five ownership types relative to this one. The model was solved using the maximum likelihood estimation method with full model selection using SAS version 9.1 (SAS Institute, Cary, NC). Regression coefficients and significance levels are reported so that significant variables can be identified and the direction of their influence observed.

Separate models were developed for each of the four states to allow for state-specific patterns to be discerned. Prior to model construction, pairwise correlations were computed among the potential independent variables within each state. If partial correlations were higher than 0.35, the variable with the lower explanatory power was dropped from consideration in the corresponding model to avoid multicollinearity. Correlations among some of the predictor variables differed by state, so the variables retained in each state model varied slightly. In the Arizona and Oregon models, the state and local ownership categories were combined into a single category because there were too few local forest ownership sample points to create separate models. Although federal and state ownerships were included in the models, the relatively high accuracies and finer spatial resolution available in PAD led to further refinement of the predictions, similar to the approach used with PAD in the Thiessen polygon models. If there was a discrepancy between the model prediction and PAD data for federal and state ownerships, the PAD data took precedent and its value was assigned to the point.

Classification trees

Using the same predictor variables as the multinomial logit method, a classification tree approach was used to predict forest ownership categories within each state. Again, the resulting model could be used to predict the ownership categories for all pixels across the area of interest. Classification trees are a nonparametric tool for predicting a variable that aims to minimize variability within the final predicted nodes/bins and to maximize differences among the nodes/bins by segmenting the data based on predictor variables (Breiman, Friedman, Olshen, & Stone, 1984). These methods: allow the inclusion of categorical and continuous variables; can yield high classification accuracy with complex datasets; are relatively easy to apply and interpret; and permit high correlations among predictor variables.

This approach was implemented using the random forest algorithm in the R statistical computing environment (Liaw & Wiener, 2002; R Development Core Team, 2012). The random forests algorithm (Breiman, 2001) builds a series of classification trees by sequentially withholding observations and variables. Once the "forest of trees" has been constructed, an observation is passed through all of the trees, and the resulting class is assigned based on the most commonly predicted class. Based on classification accuracies, this method performed better than other classification tree algorithms, i.e., tree (Ripley, 2011) and rpart (Therneau, Atkinson, & Ripley, 2011), as is consistent with other research (Gislason, Benediktsson, &

Table 3
Descriptions of independent variables used to model forest ownership patterns. Values are means (and standard deviations).

Category/variable	Short name	Description	Data type[a]	Alabama	Arizona	Michigan	Oregon	Data source (and reference)
Population								
Population density	POP_DENS	Number of people per square kilometer within a Census block group	C	56.43 (138.83)	9.50 (52.00)	18.18 (68.32)	10.68 (51.73)	U.S. Census (GeoLytics, 2004)
Housing density	HOUSE_DENS	Number of houses per square kilometer within a Census block group	C	21.55 (55.06)	3.49 (20.99)	9.01 (31.12)	4.16 (21.65)	U.S. Census (GeoLytics, 2004)
Population gravity index (population pressure)	POP_GI	Average of the population divided by the distance (km) squared for the three most influential cities, towns, or places	C	306.29 (3072)	140.17 (1368)	173.41 (1921)	225.94 (2990)	U.S. Census (ESRI, 2002)
Roads								
Distance to roads	RD_DIST	Distance (km) to nearest road (limited access highway, highway/secondary road, or local road)	C	0.40 (0.37)	0.92 (1.16)	0.99 (6.60)	0.56 (0.92)	ESRI Streets (GDT, 2002)
Road density	RD_DENS	Density (km/km^2) of roads	C	1.20 (0.88)	0.75 (0.85)	1.26 (0.94)	1.32 (1.11)	ESRI Streets (GDT, 2002)
Forest attributes								
Forest type	HARDWOOD	Dummy variable indicating if the plot was classified as hardwood (1) or softwood (0)	B	0.53 (0.01)	0.23 (0.01)	0.74 (0.00)	0.07 (0.01)	USFS-FIA (Bechtold & Patterson, 2005)
Basal area	FOR_BA	Basal area of live trees over 2.5 cm dbh in m^2/ha	C	69.47 (26.07)	53.80 (43.08)	75.78 (38.67)	96.22 (73.17)	USFS-FIA (Bechtold & Patterson, 2005)
Core forest	FOR_CORE	Proportion of forest land within 1 km that is at least 120 m away from a nonforest edge	P	0.32 (0.22)	0.24 (0.33)	0.37 (0.28)	0.32 (0.29)	Morphological Spatial Pattern Analysis of Forest Cover (Riitters, 2011, 64 p.)
Land cover								
Forest	PROP_FOR	Proportion of land that is forested within a 1 km radius	P	0.70 (0.20)	0.44 (0.39)	0.74 (0.23)	0.63 (0.33)	U.S. National Land Cover Database (Homer et al., 2004)
Agriculture	PROP_AG	Proportion of land that is agricultural crop or pasture land within a 1 km radius	P	0.12 (0.15)	0.00 (0.03)	0.09 (0.18)	0.04 (0.15)	U.S. National Land Cover Database (Homer et al., 2004)
Developed	PROP_DEV	Proportion of land that is developed within a 1 km radius	P	0.04 (0.07)	0.01 (0.03)	0.06 (0.09)	0.02 (0.07)	U.S. National Land Cover Database (Homer et al., 2004)
Other	PROP_OTH	Proportion of land that is not forest, agriculture, or developed within a 1 km radius	P	0.12 (0.10)	0.55 (0.39)	0.09 (0.11)	0.30 (0.30)	U.S. National Land Cover Database (Homer et al., 2004)
Other								
Distance to water	WAT_DIST	Distance (km) to nearest stream, river, pond, lake, or other permanent water body	C	2.67 (2.03)	37.41 (37.64)	2.86 (2.28)	4.61 (4.88)	National Atlas of the United States (2005)
Water within 0.25 km	WAT_0_25	Dummy variable indicating whether there is a permanent water body within 0.25 km of actual plot location	B	0.06 (0.00)	0.00 (0.00)	0.06 (0.00)	0.05 (0.00)	National Atlas of the United States (2005)
Water within 1.0 km	WAT_1_0	Dummy variable indicating whether there is a permanent water body within 1.0 km of actual plot location	B	0.25 (0.01)	0.01 (0.00)	0.24 (0.00)	0.17 (0.01)	National Atlas of the United States (2005)
Water within 5.0 km	WAT_5_0	Dummy variable indicating whether there is a permanent water body within 5.0 km of actual plot location	B	0.86 (0.01)	0.09 (0.01)	0.84 (0.00)	0.66 (0.01)	National Atlas of the United States (2005)
Distance to public lands	PUB_DIST	Distance (km) to nearest public lands as defined by PAD	C	14.13 (12.47)	5.04 (12.22)	1.97 (3.37)	0.64 (1.45)	PAD (CBI, 2010)
Distance to private lands	PRIV_DIST	Distance (km) to nearest private lands as defined by PAD	C	0.04 (0.26)	1.97 (3.02)	0.29 (0.63)	2.27 (3.80)	PAD (CBI, 2010)
PAD inclusion	PAD_INCLU	Dummy variable indicating whether a point falls within the boundaries of a federal, state, or locally owned parcel according to the PAD	B	0.06 (0.00)	0.64 (0.01)	0.38 (0.01)	0.60 (0.01)	PAD (CBI, 2010)
PAD defined ownership type	PAD_CALL	Ownership type (Federal, State, Local, or Private) according to the PAD	D	–	–	PAD (CBI, 2010)		

[a] Variable type: B = binary/dummy; C = continuous; D = discrete/categorical; P = proportion.

Table 4
Weighting matrices used for calculating weighted kappa (κ_w) statistics for forest ownership confusion matrices with 5 and 6 ownership categories.

A.		Predicted				
		Federal	State/local	Family	Corporate	Other private
Observed	Federal	1.0	0.5	0.0	0.0	0.0
	State/local	0.5	1.0	0.0	0.0	0.0
	Family	0.0	0.0	1.0	0.5	0.5
	Corporate	0.0	0.0	0.5	1.0	0.5
	Other private	0.0	0.0	0.5	0.5	1.0

B.		Predicted					
		Federal	State	Local	Family	Corporate	Other private
Observed	Federal	1.0	0.5	0.5	0.0	0.0	0.0
	State	0.5	1.0	0.5	0.0	0.0	0.0
	Local	0.5	0.5	1.0	0.0	0.0	0.0
	Family	0.0	0.0	0.0	1.0	0.5	0.5
	Corporate	0.0	0.0	0.0	0.5	1.0	0.5
	Other private	0.0	0.0	0.0	0.5	0.5	1.0

Sveinsson, 2006). As with the other approaches, if a predicted ownership category differed from the federal or state ownership as identified in PAD, the point was assigned the PAD value.

Accuracy assessment

Model accuracies were assessed using a combination of site-specific and non-site specific approaches. Cross-validation was used to assess site-specific accuracies. Ten percent of the points in the initial FIA dataset were withheld as validation points. This subsampling was accomplished by randomly selecting 10 percent of the points from each ownership category within each state. This percentage is within the range used by Blackard et al. (2008) – a study that also relied on FIA sample points for modeling and validation. The same 1741 validation points were used to assess the accuracy of all methods. To assess the differences between the observed and predicted values, confusion matrices were created with the rows representing the observed ownership categories and the columns representing the predicted categories. Cell values reflect the percentages of the validation points that fell within the given observed/predicted combination. The overall accuracy is the sum of the major diagonals of each matrix. To test the hypothesis that the proportions correctly predicted were not significantly different (i.e., $p_1 = p_2$), the p-value was calculated based on $Z = |d|/\text{s.e.}$ where $d = \hat{p}_1 - \hat{p}_2$, $\text{s.e.} = \sqrt{(\hat{p}_1(1-\hat{p}_1) + \hat{p}_2(1-\hat{p}_2) - 2(\hat{p}_{1,2} - \hat{p}_1\hat{p}_2))/n_e}$, $\hat{p}_{1,2}$ = proportion in common, and n_e = effective sample size (Dorofeev & Grant, 2006).

Weighted kappa (κ_w) statistics (Cohen, 1968) were used to summarize each confusion matrix. A κ_w value of 1.0 represents perfect agreement between the predicted and observed values and a value of 0.0 represents full disagreement. The weightings (Table 4) take into account where the misclassifications are and here are used to differentiate between misclassifications within versus outside the public/private quadrants of the matrices. Ninety-five percent confidence intervals were calculated using the method described by Fleiss, Cohen, and Everitt (1969). κ_w values and

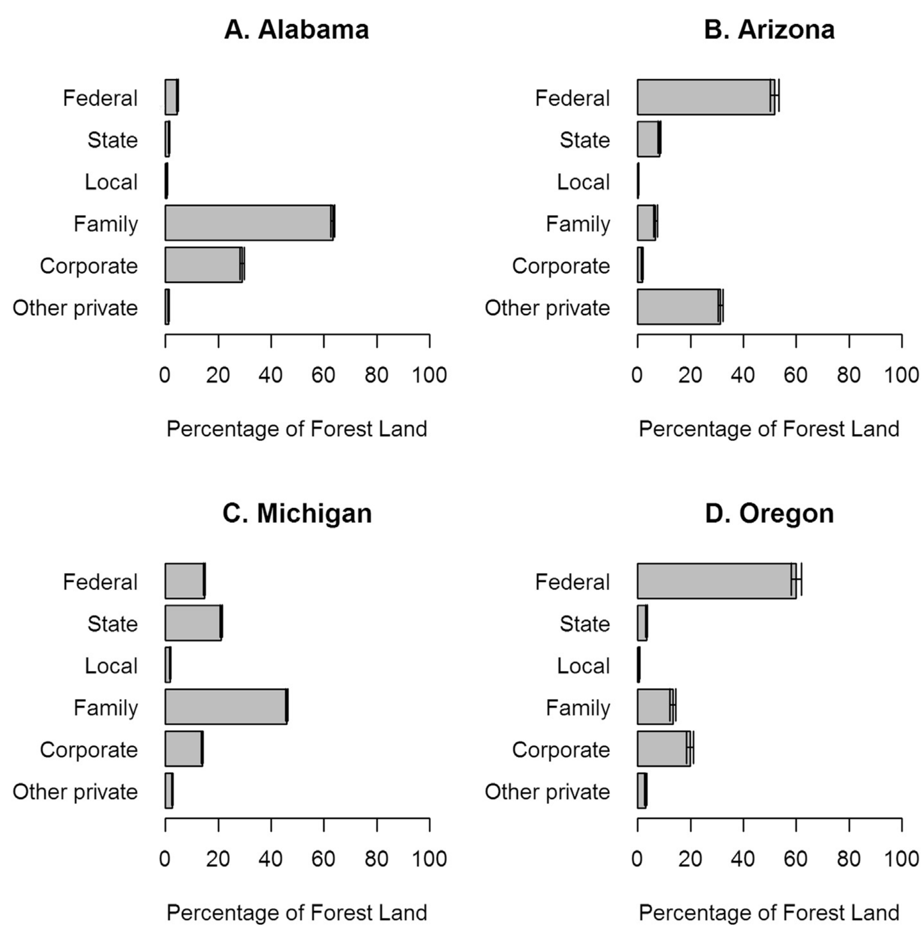

Fig. 2. Percentage of forest land by ownership based on FIA estimates for: A) Alabama, B) Arizona, C) Michigan, and D) Oregon. Error bars represent 95 percent confidence intervals.

associated confidence intervals were computed using the Cohen kappa function (Revelle, 2011) in the R statistical computing environment.

Non-site specific accuracies were assessed by comparing the predicted ownership percentages with those percentages based on the FIA samples for each state.

Results

Based on FIA estimates, the distribution of forest ownership varies considerably across the four states included in this analysis (Fig. 2). In Alabama, the forest ownership is dominated by family (63 percent) followed by corporate (29 percent). In Michigan, family is also dominant, but to a lesser extent (46 percent), and there are substantial holdings by state (21 percent), federal (15 percent), and corporate (14 percent) ownerships. In Oregon and Arizona, federal ownership dominates (60 and 52 percent, respectively), but corporate (20 percent) is the second most common category in Oregon and other private (31 percent), composed largely of Native American tribal lands, is the second most common category in Arizona.

Across the four study states, the Thiessen polygon method correctly predicts 77 percent of the validation points ($\kappa_w = 0.76$) (Table 5). The multinomial logit method correctly predicts 76 percent of the validation points ($\kappa_w = 0.73$). The classification tree approach correctly predicts 79 percent of the validation points ($\kappa_w = 0.76$). The only statistically significant differences are between the multinomial logit and classification tree approaches.

For Alabama, the percentage correctly predicted ranges from 65 to 71 across the methods (Fig. 3). The classification tree approach has a significantly (*p-value* = 0.04) higher percentage correctly predicted than the Thiessen polygon approach. The multinomial logit approach is not significantly different from either of the other approaches. The weighted kappas for the three methods are very similar (Fig. 4). The differences between the accuracies and the kappas are due to how the

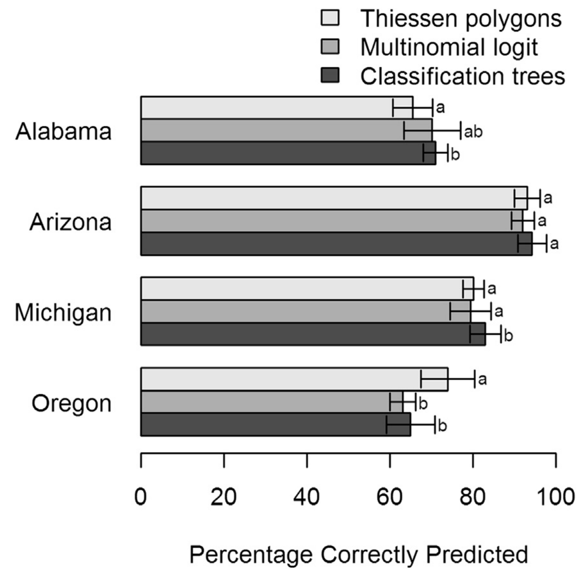

Fig. 3. Percentage of correctly predicted forest ownership validation points in Alabama, Arizona, Michigan, and Oregon using Thiessen polygons, multinomial logit, and classification tree approaches. Error bars represent 95 percent confidence intervals. Letters represent approaches that are not significantly different from each other within each state.

Table 5
Observed and predicted forest ownership using: A) Thiessen polygon, B) multinomial logit, and C) classification tree approaches. Values are percentages of 1741 validation points across Alabama, Arizona, Michigan, and Oregon. Acc = overall accuracy (sum of the major diagonal); κ_w = weighted kappa statistic; n = number of observations.

A. Thiessen polygons (*Acc* = 77.3; $\kappa_w = 0.76$; $n = 1741$)

		Predicted						
		Federal	State	Local	Family	Corporate	Other private	Total
Observed	Federal	22.9	0.2	0.0	0.2	0.5	0.0	23.8
	State	0.2	13.0	0.0	0.6	0.1	0.0	13.8
	Local	0.1	0.0	0.3	0.7	0.1	0.1	1.3
	Family	0.8	0.8	0.6	30.2	7.1	0.6	40.1
	Corporate	1.2	0.8	0.3	6.1	8.0	0.3	16.7
	Other private	0.2	0.1	0.0	1.0	0.2	2.9	4.4
	Total	25.4	14.9	1.3	38.8	15.9	3.8	100.0

B. Multinomial logit regression (*Acc* = 76.3; $\kappa_w = 0.73$; $n = 1741$)

		Predicted					
		Federal	State/local[a]	Family	Corporate	Other private	Total
Observed	Federal	23.0	0.1	0.5	0.3	0.0	23.8
	State/local[a]	0.6	12.8	1.6	0.1	0.0	15.0
	Family	1.4	0.6	36.6	1.3	0.2	40.1
	Corporate	2.2	0.6	12.1	1.8	0.1	16.7
	Other private	0.5	0.1	1.4	0.2	2.2	4.4
	Total	27.7	14.1	52.1	3.7	2.4	100.0

C. Classification tree (*Acc* = 78.9; $\kappa_w = 0.76$; $n = 1741$)

		Predicted						
		Federal	State	Local	Family	Corporate	Other private	Total
Observed	Federal	22.4	0.1	0.0	0.5	0.9	0.0	23.8
	State	0.3	12.8	0.0	0.7	0.1	0.0	13.8
	Local	0.2	0.0	0.1	0.8	0.2	0.0	1.3
	Family	0.9	0.5	0.0	35.3	3.3	0.1	40.1
	Corporate	1.6	0.5	0.0	8.9	5.7	0.0	16.7
	Other private	0.3	0.1	0.0	1.2	0.2	2.6	4.4
	Total	25.6	14.0	0.1	47.4	10.3	2.6	100.0

[a] Due to low numbers of points owned by local governments, Local was combined with State for the logistic regression method.

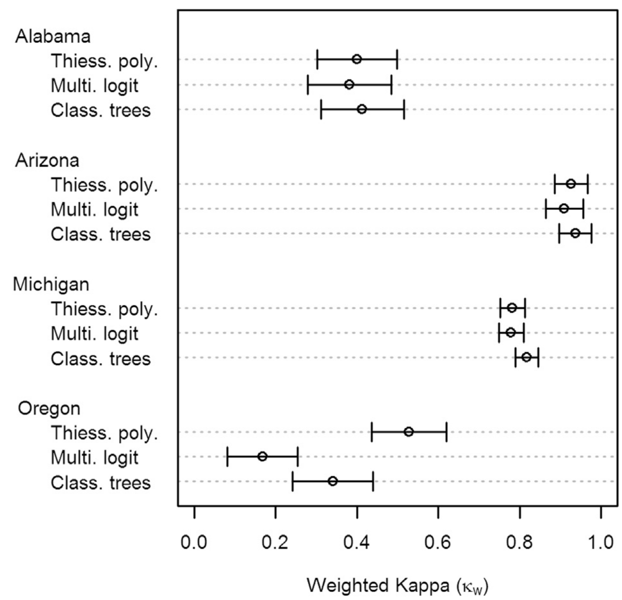

Fig. 4. Comparisons of state-level weighted kappa (κ_w) statistics using Thiessen polygons, multinomial logit, and classification tree approaches. Error bars represent 95 percent confidence intervals.

errors are propagated. The classification tree and multinomial logit approaches over-predict the dominant ownership category, family, and under-predict other categories, especially corporate (Fig. 5).

For Arizona, the methods correctly predict between 92 and 94 percent of the validation points (Fig. 3). The accuracies are not significantly different among the methods. The kappa values are likewise similar among the methods (Fig. 4). The predictions among the ownership categories are similar with the exception of corporate where the Thiessen polygon approach slightly over-predicts the FIA estimate and the other methods fail to predict any acres for this category (Fig. 5).

For Michigan, 79 to 83 percent of the validation points are correctly predicted across the methods (Fig. 3). The accuracy of the classification tree approach is significantly higher than the other methods, and the other two methods are not significantly different from each other. The weighted kappa statistics for the classification tree approach is slightly higher than the other approaches (Fig. 4). The estimates for federal and state ownership categories are similar across approaches and the Thiessen polygon approach more closely matches the FIA estimates for the other categories (Fig. 5).

Oregon has the greatest range in the model accuracies: 63 to 74 percent (Fig. 3). For this state, the Thiessen polygon approach is significantly more accurate than the other approaches and the other two approaches are not significantly different from each other. The weighted kappa for the Thiessen polygon approach is higher than classification tree approach which is higher than the multinomial logit approach (Fig. 4). All of the approaches over-estimate federal ownership and under-estimate family ownership, but the Thiessen polygon approach is closest to the FIA estimates (Fig. 5).

The multinomial logit and classification tree approaches allow for the investigation of what factors are correlated with forest ownership patterns. The significant variables in the multinomial logit models vary by ownership group and state (Table 6). The family ownership category was used as the reference condition in all states. As such, the model coefficients are interpreted relative to this ownership category. Housing density (HOUSE_DENS) is significant, with a negative sign, in at least one of the models for Arizona, Michigan, and Oregon. This suggests, for example, that points within the vicinity of higher housing density areas are less likely to be in federal versus family forest ownership. Road distance (RD_DIST) is significant in at least one model in Alabama, Arizona, and Michigan and had a positive sign. Forest basal area (FOR_BA) has a positive, significant coefficient in at least one model in each state. Proportion agricultural land (PROP_AG) is significant in at least one Alabama, Oregon, and Arizona model; the sign is negative in the Alabama and Oregon models and positive in the Arizona model. One of the water distance variables is significant in four of the Michigan models and one of the Arizona and Oregon models; the signs of the coefficients vary. Distance to public land (PUB_DIST) is significant in Alabama, Arizona, and Michigan models (due to multicollinearity it was dropped from the Oregon models) with varying signs. Distance to private land (PRIV_DIST) is significant in Arizona, Michigan, and Oregon models (it was dropped from the Alabama models) with positive, often large, signs. The population gravity index (POP_GI) is positive and significant in at least one of the Alabama and Michigan models. Forest type (HARDWOOD) is negative and significant in the corporate and other private models for Alabama and Arizona, respectively. The PAD variables are significant in multiple models for each state.

As with the multinomial logit model, the variables which are the most influential in the classification tree models vary considerably among the states (Fig. 6). The Arizona and Michigan classification tree models are the most similar in terms of relative importance values of variables. The ownership category from the PAD (PAD_CALL) and distance to public land (PUB_DIST) are the most influential variables in both of these models, but the difference between the relative importance values of these variables is much greater in Michigan. In the Oregon model, there are numerous variables that have high relative importance values including: the population gravity index (POP_GI); population density (POP_DENS); housing density (HOUSE_DENS); forest basal area (FOR_BA); distance to water (WATER_DIST); distance to road (RD_DIST); distance to public land (PUB_DIST); and density of the road network (DENS_RD). The Alabama model has many variables with high relative importance values including: distance to public land (PUB_DIST), which has by far the greatest relative significance; proportion agricultural land (PROP_AG); the population gravity index (POP_GI); distance to road (RD_DIST); and forest basal area (FOR_BA).

Discussion

The three methods correctly predicted the observed ownership values between 76 and 79 percent of the time (Table 5), values comparable to many spatial products derived from remotely sensed imagery (e.g., Homer, Huang, Yang, Wylie, & Coan, 2004). The multinomial logit regression approach, and to a lesser degree the classification tree approach, was prone to over-estimation of the dominant ownership categories, especially in Alabama, Michigan, and Oregon (Fig. 5). The accuracies of the methods vary across states and ownership categories with a subset of them being significantly different (Fig. 4). This variability is a result of forest ownership heterogeneity, spatial configurations, and other factors, such as historical land disposition events. In areas with lower heterogeneity and larger more contiguous ownership blocks (e.g., Arizona), accuracies are higher. Although the sampling intensity will not influence the expected values of the point estimates, higher sampling intensities do, all else being equal, produce tighter confidence intervals.

The Thiessen polygon method has accuracies that are similar to the other methods, requires fewer input variables, better corresponds to the FIA aspatial estimates (i.e., shows lower biases

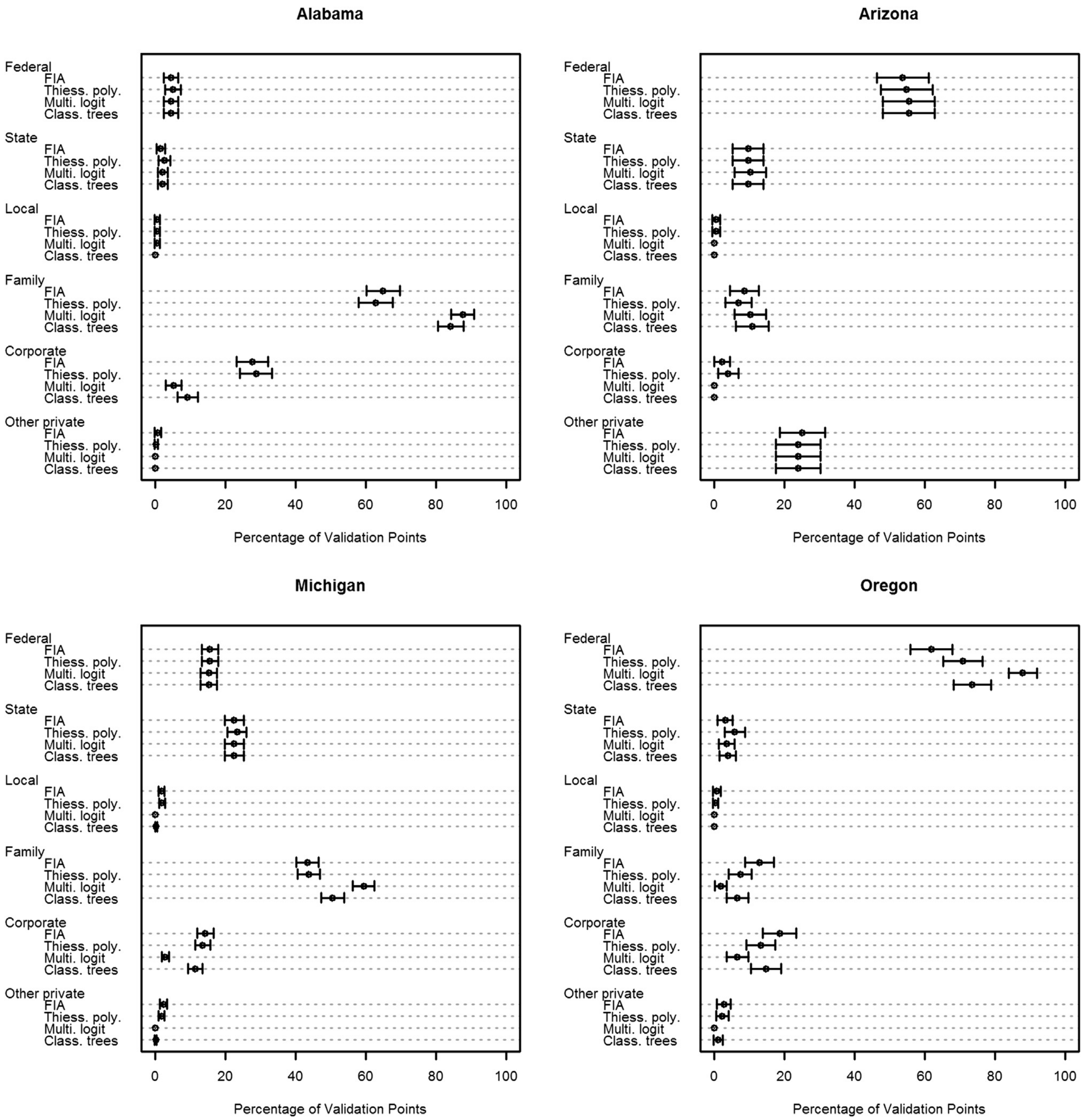

Fig. 5. Estimated percentages of forest area by ownership category using Thiessen polygons, multinomial logit, and classification tree approaches along with the corresponding percentage of FIA validation points. Error bars represent 95 percent confidence intervals.

towards dominant categories), and is much simpler to implement. This method can be used consistently for all states and across borders, whereas other methods appear to require development of separate models that are fit specifically for different states or at least regions. For these reasons, it was selected as the method for producing a forthcoming forest ownership geospatial dataset for the conterminous U.S. (Hewes et al., 2013b). Examples of maps created with this approach are in Fig. 7.

Although the multinomial logit and classification trees were not superior for generating a national geospatial dataset, these approaches do provide additional insights into factors that are correlated with forest ownership categories and patterns. Based on the multinomial logit models, variables related to remoteness and lower population pressures tended to indicate federal, state and/or corporate ownerships – ownerships that tend to have larger, more contiguous holdings. For example, as distance to road increases, the probability of forest land in Michigan being family owned decreases.

Public and private distance variables are important in many of the Arizona and Michigan models and at least one multinomial logit model for the other two states. The signs for the significant variables are as expected. Forest structure is correlated with forest ownership, especially corporate ownership patterns. Forest basal area (FOR_BA) is significant in 3 of 4 of the corporate models. As basal area increases, the probability of corporate ownership (versus family) increases. This may be an indication of more intensive forest

Table 6
Results from multinomial logit models of forest ownership models in Alabama, Arizona, Michigan, and Oregon. Numbers represent coefficients and asterisks represent significance levels.

Variable	Model[a]	Alabama	Arizona	Michigan	Oregon
INTERCEPT	FED	−3.94***	0.05	−1.53***	0.87***
INTERCEPT	STATE	−4.05***	−5.60***	−6.53***	−1.70***
INTERCEPT	LOCAL	−2.93**	–	−3.93***	–
INTERCEPT	CORP	0.09	−1.84***	−2.67***	−0.09
INTERCEPT	OTH_PRIV	−5.11***	−2.84***	−3.60***	−1.12***
HOUSE_DENS	FED	−0.01	0.01	−0.02*	−0.05***
HOUSE_DENS	STATE	0.00	−0.06*	0.00	0.00
HOUSE_DENS	LOCAL	0.00	–	0.00**	–
HOUSE_DENS	CORP	0.00	0.01	0.00	−0.02**
HOUSE_DENS	OTH_PRIV	0.00	−0.01	0.00	−0.01
RD_DIST	FED	−0.40	0.28	0.67***	−0.03
RD_DIST	STATE	0.54	0.18	0.63***	−0.48
RD_DIST	LOCAL	−1.20	–	−0.42	–
RD_DIST	CORP	0.34**	0.42	0.66***	−0.16
RD_DIST	OTH_PRIV	0.31	0.79**	0.68***	0.18
FOR_BA	FED	0.00	0.01	0.00	0.00
FOR_BA	STATE	0.00	−0.01	0.00	0.01**
FOR_BA	LOCAL	−0.02	–	0.00	–
FOR_BA	CORP	0.00**	0.01	0.02***	0.01***
FOR_BA	OTH_PRIV	0.01	0.02**	0.01**	0.00
PROP_AG	FED	−0.02	0.07	–	0.00
PROP_AG	STATE	−0.04*	0.13*	–	0.01
PROP_AG	LOCAL	−0.03	–	–	–
PROP_AG	CORP	−0.05***	−0.03	–	0.00
PROP_AG	OTH_PRIV	−0.01	−0.05	–	−0.03*
WAT_1_0	FED	0.36	–	0.33*	0.40*
WAT_1_0	STATE	−0.02	–	0.54**	0.31
WAT_1_0	LOCAL	0.29	–	0.50**	–
WAT_1_0	CORP	−0.06	–	−0.18*	0.31
WAT_1_0	OTH_PRIV	0.18	–	0.21	−0.32
WAT_5_0	FED	–	−0.47	–	–
WAT_5_0	STATE	–	−2.00**	–	–
WAT_5_0	LOCAL	–	–	–	–
WAT_5_0	CORP	–	−0.32	–	–
WAT_5_0	OTH_PRIV	–	0.45	–	–
PUB_DIST	FED	−0.01	−5.58***	−4.43***	–
PUB_DIST	STATE	−0.06*	0.64	0.15***	–
PUB_DIST	LOCAL	0.00		0.06**	–
PUB_DIST	CORP	0.00	0.07	0.00	–
PUB_DIST	OTH_PRIV	0.00	1.05***	−0.03	–
PRIV_DIST	FED	–	10.98***	4.77***	0.47***
PRIV_DIST	STATE	–	10.41***	4.34***	−0.12
PRIV_DIST	LOCAL	–	–	1.91	–
PRIV_DIST	CORP	–	−5.74	3.54***	0.17*
PRIV_DIST	OTH_PRIV	–	7.49**	−0.22	0.20
POP_GI	FED	0.00	–	0.00***	0.00
POP_GI	STATE	0.00	–	0.00	0.00
POP_GI	LOCAL	0.00**	–	0.00**	–
POP_GI	CORP	0.00	–	0.00	0.00
POP_GI	OTH_PRIV	0.00	–	0.00	0.00
HARDWOOD	FED	−0.14	−0.08	0.05	−0.34
HARDWOOD	STATE	0.24	0.52	−0.13	0.54
HARDWOOD	LOCAL	−0.59	–	0.38	–
HARDWOOD	CORP	−0.51***	−0.81	0.09	−0.08
HARDWOOD	OTH_PRIV	−0.20	−0.99*	−0.10	−0.62
PAD_INCLU	FED	6.90***	–	–	–
PAD_INCLU	STATE	4.30***	–	–	–
PAD_INCLU	LOCAL	−8.51	–	–	–
PAD_INCLU	CORP	1.78***	–	–	–
PAD_INCLU	OTH_PRIV	−10.05	–	–	–
PAD_CALL	FED	–	−1.06**	1.43***	0.26
PAD_CALL	STATE	–	3.42***	5.55***	−0.25
PAD_CALL	LOCAL	–	–	1.30**	–
PAD_CALL	CORP	–	0.46	−0.97**	−0.36
PAD_CALL	OTH_PRIV	–	0.14	−0.11	−1.18**

Significance levels: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.
[a] FED = Federal; STATE = State; LOCAL = Local; CORP = Corporate; OTH_PRIV = Other private. The reference level is Family.

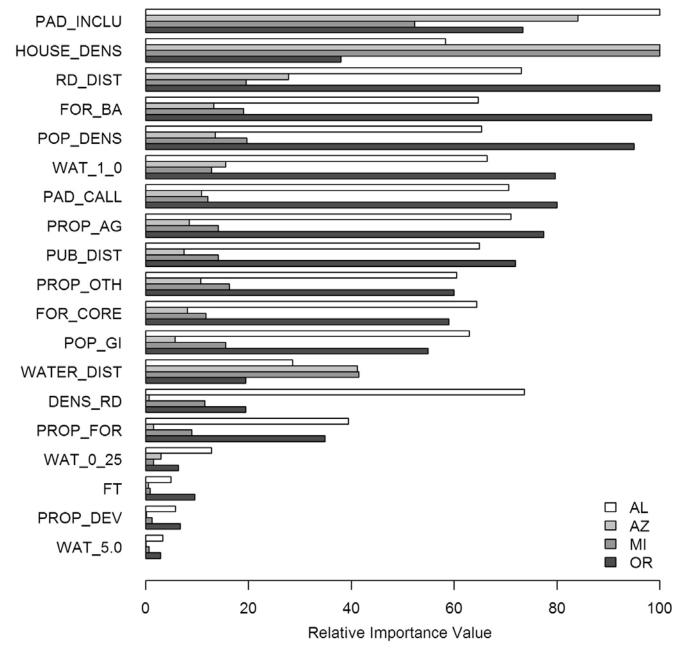

Fig. 6. Relative importance of variables from random forest classification tree models of forest ownership in Alabama, Arizona, Michigan, and Oregon. The values represent the importance values (Breiman, 2001) divided by the maximum importance value for a given state multiplied by 100.

management on corporate forest lands or the inclination of these ownerships to acquire lands with higher potential productivity relative to other ownership groups due to the greater emphasis on financial motivations of the corporate ownerships. These financial motivations are also likely the reason for the negative sign for the hardwood dummy variable (HARDWOOD) in the corporate model for Alabama where softwoods are the more important commercial species.

Water is primarily significant in the Michigan multinomial logit models. Points being close, i.e., within 1.0 km, to a water body increases the probability of the point being federal, state, or local and decreases the probability of them being corporate, as compared to the reference level, family. These water resources may be protected for the public good and/or the locations of these lands may be an artifact of the places where public ownerships have tended to acquire land.

Examining results from the classification tree model (Fig. 6), a slightly different set of important variables is observable. For Michigan and Arizona, the three top variables are related to PAD: PAD ownership call (PAD_CALL), distance to public land (PUB_DIST; which was measured using PAD), and whether the point is within one of the PAD identified public ownership boundaries (PAD_INCLU).

Although distance to public land (PUB_DIST) is the most important variable in the Alabama classification tree model, a majority of the most influential variables in the Alabama and Oregon models are population-related. In the models for the other states, population related variables are common among the second tier of predictive variables. The population gravity index (POP_GI) is highly influential in many of the models as are the population (POP_DENS) and housing (HOUSE_DENS) density variables. Public land holdings, particularly Federal and State, tend to be larger, contiguous pieces and, partially related to this, in more rural areas with lower population pressures. As compared to family, corporate holdings also tend to be in larger holdings and in more rural areas.

In summary, many of the predictor variables hypothesized to have an influence on forest ownership patterns did show some

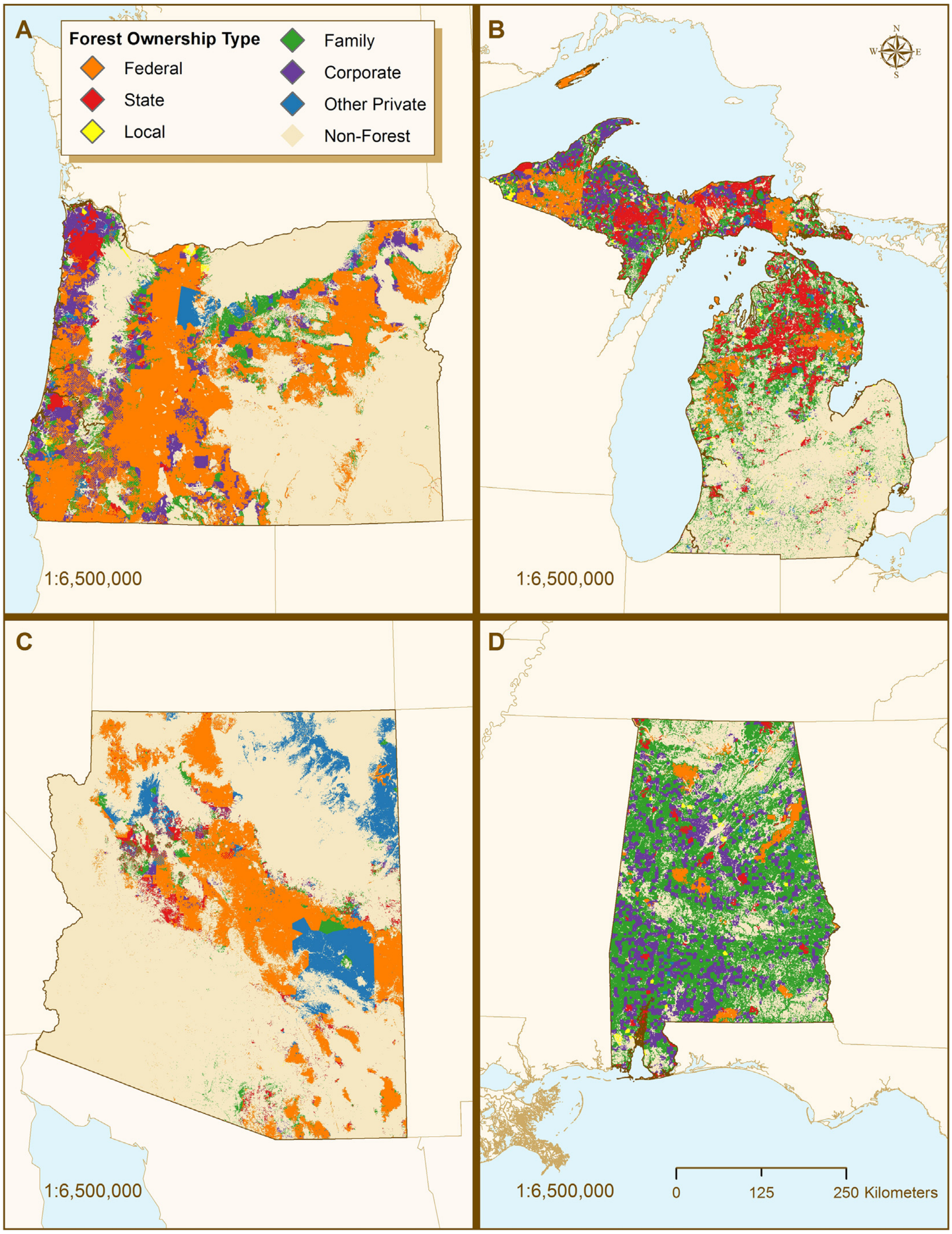

Fig. 7. Forest ownership maps for A) Oregon, B) Michigan, C) Arizona, and D) Alabama created using a Thiessen polygon approach. The forest/non-forest mask was developed by Blackard et al. (2008).

level of correlation with FIA-observed ownership category. However, their signs and magnitudes of influence vary considerably among the models. This suggests that different factors are influencing forest ownership patterns across the U.S. and applying the multinomial logit or classification tree approaches would require separate parameterization of the models for different states or at least regions.

Conclusions

The appropriate technique for creating a geospatial product, such as a forest ownership map, depends upon the intended uses of the end-product and the types and quality of the input data (Stewart et al., 2009; Zhang & Wimberly, 2007). Ideally, detailed forest ownership geospatial products could be derived from property tax records and plat maps (Donnelly & Evans, 2008). Unfortunately, these maps do not exist in electronic formats for all of the U.S. nor is there a central repository for those datasets that do exist. Until the time when comprehensive and compatible data sources of spatial ownership data are available for the country broad-scale forest ownership geospatial products must be created using modeling approaches.

Using the techniques outlined above, a geospatial product representing forest ownership patterns can be created that improves upon earlier products by providing increased spatial resolution, more refined ownership categories, and quantifying the accuracy. A forest ownership geospatial product for the conterminous U.S. is not included here because this paper is intended to examine just the techniques and a product of that extent cannot be satisfactorily displayed in a journal article. In addition, publishing the final product separately, with this paper referenced as the technical document, will allow for easier distribution of the final product and easier updating when new data, such as newer versions of PAD, become available. This dataset should be useful to national and state forestry agencies, policy analysts, researchers, and conservation organizations when conducting broad-scale, forest resource analyses. Such products are not intended for identification of specific ownerships or parcels. For this, plat maps would be the appropriate source. These are increasingly available in electronic formats, but they have yet to be consolidated for most states. As more FIA data become available and the PAD is updated/refined, the ownership geospatial products developed through the techniques outlined here can be further refined.

The goal of this paper was to test techniques for creating a national geospatial product of forest ownership, but the methodology should also be applicable to other social variables. For example, this approach, with appropriate modifications, may be useful for exploring techniques for generating spatial datasets that reflect size of forest holdings, management practices, and other factors that are important for analyzing forest resources. Likewise, this approach can be utilized to generate forest ownership geospatial products in other countries which may lack national spatial forest ownership data.

Acknowledgments

Funding for this project was provided, in part, by the USDA Forest Service, State and Private Forestry, Cooperative Forestry. Useful reviews and input were provided by Kai Henifin, Miranda Mockrin, Rachel Riemann, and Barry T. Wilson.

References

Allison, P. D. (1999). *Logistic regression using the SAS system: Theory and applications*. Cary, NC: SAS Institute.

Banner, S. (2011). *American property: A history of how, why, and what we own*. Cambridge, MA: Harvard University Press.

Bechtold, W. A., & Patterson, P. L. (2005). *The enhanced forest inventory and analysis program—national sampling design and estimation procedures*. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station.

Bengston, D. N., Asah, S. T., & Butler, B. J. (2011). The diverse values and motivations of family forest owners in the United States: an analysis of an open-ended question in the national Woodland owner survey. *Small-scale Forestry, 10*, 339–355.

Blackard, J. A., Finco, M. V., Helmer, E. H., Holden, G. R., Hoppus, M. L., Jacobs, D. M., et al. (2008). Mapping U.S. forest biomass using nationwide forest inventory data and moderate resolution information. *Remote Sensing of Environment, 112*, 1658–1677.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. G. (1984). *Classification and regression trees*. Belmont, CA: Wadsworth International Group.

Brown, G., & Raymond, C. (2007). The relationship between place attachment and landscape values: toward mapping place attachment. *Applied Geography, 27*, 89–111.

Butler, B. J. (2008). *Family forest owners of the United States, 2006*. Newtown Square, PA: U.S: Department of Agriculture, Forest Service, Northern Research Station.

Butler, B. J., Ma, Z., Kittredge, D. B., & Catanzaro, P. (2010). Social versus biological availability of woody biomass in the northern United States. *Northern Journal of Applied Forestry, 27*(4), 151–159.

Chopra, K., & Dasgupta, P. (2008). Assessing the economic and ecosystem service contributions of forests: issues in modelling, and an illustration. *International Forestry Review, 10*, 376–386.

Cohen, J. (1968). Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin, 70*(4), 213–220.

Conservation Biology Institute (CBI). (2010). *PAD-US 1.1* (CBI Edition). Corvallis, OR: Conservation Biology Institute. Available on-line at http://www.protectedareas.consbio.org Accessed on 22.12.11.

Cordell, H. K., Betz, C. J., & Green, G. T. (2008). Nature-based outdoor recreation trends and wilderness. *International Journal of Wilderness, 14*(2), 7–13.

Donnelly, S., & Evans, T. P. (2008). Characterizing spatial patterns of land ownership at the parcel level in south-central Indiana, 1928–1997. *Landscape Urban Plan, 84*, 230–240.

Dorofeev, S., & Grant, P. (2006). *Statistics for real-life sample surveys: Non-simple-random samples and weighted data*. Cambridge: Cambridge University Press.

Environmental Systems Research Institute Inc. (ESRI). (2002). *U.S. populated place points (ESRI Data & Maps)*. Redlands, CA: Environmental Systems Research Institute Inc.

Fleiss, J. L., Cohen, J., & Everitt, B. S. (1969). Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin, 72*, 323–327.

Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern Recognition Letters, 27*(4), 294–300.

Geographic Data Technology Inc. (GDT). (2002). *U.S. streets (ESRI Data & Maps)*. Redlands, CA: Environmental Systems Research Institute, Inc.

GeoLytics. (2004). *CensusCD 2000: Short form blocks*. E. Brunswick, NJ: GeoLytics.

Hardie, I., Parks, P., Gottleib, P., & Wear, D. N. (2000). Responsiveness of rural and urban land uses to land rent determinism in the U.S. South. *Land Economics, 76*(4), 659–673.

Harper, C., Crow, T., Cooksey, R., Hoover, A., Butler, B. J., Flather, C., et al. (2006). *Cooperating across boundaries: Partnerships to conserve open space in rural America*. Washington, DC: USDA Forest Service.

Hewes, J. H., Butler, B. J., Liknes, G. C., Nelson, M. D., & Snyder, S. A. (2013a). *Forest ownership across the conterminous United States: ForestOwn_v2 geospatial dataset*. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station (in press).

Hewes, J. H., Butler, B. J., Liknes, G. C., Nelson, M. D., & Snyder, S. A. (2013b). *Map of forest ownership across the conterminous United States*. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station (in press).

Homer, C., Huang, C., Yang, L., Wylie, B., & Coan, M. (2004). Development of a 2001 national landcover database for the United States. *Photogrammetric Engineering and Remote Sensing, 70*(7), 829–840.

Jaimes, N. B. P., Sendra, J. B., Delgado, M. G., & Plata, R. F. (2010). Exploring the driving forces behind deforestation in the state of Mexico (Mexico) using geographically weighted regression. *Applied Geography, 30*(4), 576–591.

Jin, S. M., & Sader, S. A. (2006). Effects of forest ownership and change on forest harvest rates, types and trends in northern Maine. *Forest Ecology and Management, 228*(1), 177–186.

Kline, J. K., & Alig, R. J. (2005). Forestland development and private forestry with examples from Oregon (USA). *Forest Policy and Economics, 7*, 709–720.

van Kooten, G. C. (1993). *Land resource economics and sustainable development*. Vancouver: UBC Press.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News, 2*(3), 18–22.

National Atlas of the United States. (2005). *Streams and waterbodies of the United States*. Reston, VA: National Atlas of the United States.

Nelson, M. D., Liknes, G. C., & Butler, B. J. (2010a). *Forest ownership in the conterminous United States: ForestOwn_v1 geospatial dataset*. Newtown Square, PA: U.S: Department of Agriculture, Forest Service, Northern Research Station.

Nelson, M. D., Liknes, G. C., & Butler, B. J. (2010b). *Map of forest ownership in the conterminous United States*. Newtown Square, PA: U.S: Department of Agriculture, Forest Service, Northern Research Station.

Ohmann, J. L., & Gregory, M. J. (2002). Predictive mapping of forest composition and structure with direct gradient analysis and nearest neighbor imputation in coastal Oregon, U.S.A. *Canadian Journal of Forest Research, 32*, 725–741.

Okabe, A., Boots, B., Sugihara, K., & Chiu, S. N. (2000). *Spatial tessellations: Concepts and applications of voronoi diagrams*. New York: Wiley.

Paudel, G. S., & Thapa, G. B. (2004). Impact of social, institutional and ecological factors on land management practices in mountain watersheds of Nepal. *Applied Geography, 24*(1), 35–55.

Polyakov, M., Wear, D. N., & Huggett, R. N. (2010). Harvest choice and timber supply models for forest forecasting. *Forest Science, 56*(4), 344–355.

R Development Core Team. (2012). In *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Revelle, W. (2011). *Psych: Procedures for personality and psychological research*. Evanston, IL. Available on-line at http://personality-project.org/r/psych.manual.pdf Accessed on 22.05.12.

Ripley, B. (2011). *Tree: Classification and regression trees*. R package version 1.0-29. Available on-line at http://CRAN.R-project.org/package=tree Accessed on.

Riitters, K. H. (2011). *Spatial patterns of land cover in the United States: A technical document supporting the Forest Service 2010 RPA Assessment*. Asheville, NC: Department of Agriculture Forest Service, Southern Research Station. Gen. Tech. Rep. SRS-136.

Ruefenacht, B., Finco, M. V., Nelson, M. D., Czaplewski, R. L., Helmer, E. H., Blackard, J. A., et al. (2008). Conterminous U.S. and Alaska forest type mapping using forest inventory and analysis data. *Photogrammetric Engineering and Remote Sensing, 74*, 1379–1388.

Serra, P., Pons, X., & Saurí, D. (2008). Land-cover and land-use change in a Mediterranean landscape: a spatial analysis of driving forces integrating biophysical and human factors. *Applied Geography, 28*(3), 189–209.

Sherrouse, B. C., Clement, J. M., & Semmens, D. J. (2011). A GIS application for assessing, mapping, and quantifying the social values of ecosystem services. *Applied Geography, 31*(2), 748–760.

Smith, W. B., Miles, P. D., Perry, C. H., & Pugh, S. A. (2009). *Forest resources of the United States, 2007*. Washington, DC: U.S. Department of Agriculture, Forest Service.

Snyder, S. A., & Butler, B. J. (2012). A national assessment of public recreational access on family forest lands in the United States. *Journal of Forestry, 110*(6), 318–327.

Snyder, S. A., Kilgore, M. A., Hudson, R., & Donnay, J. (2008). Influence of purchaser perceptions and intentions on price for forest land parcels: a hedonic pricing approach. *Journal of Forest Economics, 14*(1), 47–72.

Snyder, S. A., Kilgore, M. A., Taff, S. J., & Schertz, J. M. (2008). Estimating a family forest landowner's likelihood of posting against trespass. *Northern Journal of Applied Forestry, 25*(4), 180–185.

Stein, S. M., McRoberts, R. E., Alig, R. J., Nelson, M. D., Theobald, D. M., Eley, M., et al. (2005). *Forests on the edge: Housing development on America's private forests*. Portland, OR: U.S: Department of Agriculture, Forest Service, Pacific Northwest Research Station.

Stewart, S. I., Wilmer, B., Hammer, R. B., Aplet, G. H., Hawbaker, T. J., Miller, C., et al. (2009). Wildland-urban interface maps vary with purpose and context. *Journal of Forestry, 107*(2), 78–83.

Therneau, T. M., Atkinson, B., & Ripley, B. (2011). *Rpart: Recursive partitioning*. R package version 3.1-50.

U.S. Department of Agriculture Forest Service (USDA-FS). (2010). *Forest inventory and analysis national core field guide: Field data collection procedures for phase 2 plots*. Version 5.0. Unpublished information on file at http://www.fia.fs.fed.us/fia/data-collection/.

U.S. Department of Agriculture Forest Service. (2011). *National report on sustainable forests—2010*. Washington, DC: U.S: Department of Agriculture, Forest Service.

Wear, D. N. (2011). *Forecasts of county-level land uses under three future scenarios: A technical document supporting the forest service 2010 RPA assessment*. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station.

Wear, D. N. (2012). In *Personal communication*. Research Forester. USDA Forest Service, Southern Research Station.

White, E. M., Alig, R. J., & Stein, S. M. (2010). Socioeconomic changes and forestland development: commonalities and distinctions between the eastern and western United States. *Journal of Forestry, 108*(7), 329–337.

Wilson, B. T., Lister, A. J., & Riemann, R. I. (2012). A nearest-neighbor imputation approach to mapping tree species over large areas using forest inventory plots and moderate resolution raster data. *Forest Ecology and Management, 271*, 182–198.

Zhang, D., Butler, B. J., & Nagubadi, R. V. (2012). Institutional timberland ownership in the U.S. South: magnitude, location, dynamics, and management. *Journal of Forestry, 110*(7), 355–361.

Zhang, Y., & Wimberly, M. C. (2007). The importance of scale in using hierarchical census data to identify the wildland-urban interface. *Southern Journal of Applied Forestry, 31*(3), 138–147.