NC-4155
Staff: VMPR Problem: 02
Code: 1.11 Reprints: N

Ecosystems (2002) 5: 539–553
DOI: 10.1007/s10021-002-0167-8

ECOSYSTEMS
© 2002 Springer-Verlag

Mladenoff, David J.; Dahir, Sally E.; Nordheim, Eric V., *et al.* 2002. Narrowing historical uncertainty: probabilistic classification of ambiguously identified tree species in historical forest survey data. Ecosystems. 5: 539-553.

Narrowing Historical Uncertainty: Probabilistic Classification of Ambiguously Identified Tree Species in Historical Forest Survey Data

David J. Mladenoff,[1]* Sally E. Dahir,[1,2] Eric V. Nordheim,[1,3] Lisa A. Schulte,[1,4] and Glenn G. Guntenspergen[5]

[1]*Department of Forest Ecology and Management, University of Wisconsin–Madison, 1630 Linden Drive, Madison, Wisconsin 53706, USA;* [2]*Current address: Wisconsin Department of Natural Resources, Fitchburg, Wisconsin 53711, USA;* [3]*Department of Statistics, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA;* [4]*Current address: US Forest Service, North Central Research Station, Grand Rapids, Minnesota 55744, USA;* [5]*US Geological Survey, Biological Resources Division, Patuxent Wildlife Research Center, University of Wisconsin–Superior Sea Grant, Superior, Wisconsin 54880, USA*

Abstract

Historical data have increasingly become appreciated for insight into the past conditions of ecosystems. Uses of such data include assessing the extent of ecosystem change; deriving ecological baselines for management, restoration, and modeling; and assessing the importance of past conditions on the composition and function of current systems. One historical data set of this type is the Public Land Survey (PLS) of the United States General Land Office, which contains data on multiple tree species, sizes, and distances recorded at each survey point, located at half-mile (0.8-km) intervals on a 1-mi (1.6 km) grid. This survey method was begun in the 1790s on US federal lands extending westward from Ohio. Thus, the data have the potential of providing a view of much of the US landscape from the mid-1800s, and they have been used extensively for this purpose. However, historical data sources, such as those describing the species composition of forests, can often be limited in the detail recorded and the reliability of the data, since the information was often not originally recorded for ecological purposes. Forest trees are sometimes recorded ambiguously, using generic or obscure common names. For the PLS data of northern Wisconsin, USA, we developed a method to classify ambiguously identified tree species using logistic regression analysis, using data on trees that were clearly identified to species and a set of independent predictor variables to build the models. The models were first created on partial data sets for each species and then tested for fit against the remaining data. Validations were conducted using repeated, random subsets of the data. Model prediction accuracy ranged from 81% to 96% in differentiating congeneric species among oak, pine, ash, maple, birch, and elm. Major predictor variables were tree size, associated species, landscape classes indicative of soil type, and spatial location within the study region. Results help to clarify ambiguities formerly present in maps of historic ecosystems for the region and can be applied to PLS datasets elsewhere, as well as other sources of ambiguous historical data. Mapping the newly classified data with ecological land units provides additional information on the distribution, abundance, and associations of tree species, as well as their relationships to environmental gradients before the industrial period, and clarifies the identities of species formerly mapped only to genus. We offer some caveats on the appropriate use of data derived in this way, as well as describing their potential.

Received 9 July 2001; accepted 24 January 2002.
*Corresponding author: email: djmladen@wisc.edu

Key words: Northern Lakes States; Wisconsin; pre-European vegetation; logistic regression analysis; northern hardwood–conifer forest.

INTRODUCTION

Understanding North American ecosystem patterns and vegetation distribution from the pre-European settlement period is important to help assess the distribution and natural variability of ecosystems, a prerequisite for setting appropriate management strategies (Parsons and others 1999; Schulte and Mladenoff 2001). Such reconstructions can contribute to understanding biotic–abiotic relationships (Kline and Cottam 1979; Whitney 1986; Grimm 1984), disturbance patterns (Heinselman 1973), prior vegetation composition and structure (Davis 1981; Foster and others 1996; Radeloff and others 1999), and variability in landscape patterns (Mladenoff and Pastor 1993).

Pre-European settlement conditions have served as starting points for land-cover change analyses (White and Mladenoff 1994), as well as landscape simulations that examine hypotheses of dynamics and change for target ecosystems (He and Mladenoff 1999). For management, the information may be valuable for ecosystem and habitat restoration (Covington and others 1994; Baker 1994, 1995; Radeloff and others 2000; Moore and others 1999) and for sustainable management of ecosystems used for commodity production, particularly at broad landscape scales (Radeloff and others 2000; Cissel and others 1999).

Information may come from fossil data, such as pollen; direct sources, such as dendrochronology; or indirect sources, such as cultural archives or remnant ecosystems (Swetnam and others 1999). Depending on data source, evidence can be derived for past time periods spanning decades to millenia (Bürgi and others 2000; Foster and others 1990; Forman and Russell 1983). Historical data are most useful where several sources contribute to a common conclusion (Bürgi and others 2000; Swetnam and others 1999; Foster and others 1990).

One relevant question is whether such pre-European or preindustrial period vegetation reconstructions represent a relatively short time frame—that is, one of a century or less—or a much longer portion of the Holocene. In some regions, these presettlement reconstructions may represent the protohistoric period of great influence on the land by Native American populations. Other areas may reflect a period of reduced Indian influence due to postcontact depopulation. Still other regions, where native populations were typically low, would reflect a stronger influence of nonhuman effects (Vale 1998; Landres and others 1999). These issues concern the proper interpretation and use of such data in a given locale or region and should not be confused with their intrinsic value (Schulte and Mladenoff 2001). Other issues concern the source and reliability of historical data, which need to be assessed in using such information.

In the forested northern portion of the US Lake States, the Public Land Survey (PLS) data of the mid-late 1800s (largely 1830s–60s) are believed to represent a reasonable baseline for a period extending back 1000–3000 years before that time. This must be seen in a regional and relative sense. Variation of forest composition did occur at more local scales and at shorter time scales, and environmental gradients shifted with climate variability. Vegetation was affected locally due to Native American activities. But on a regional scale, anthropogenic impacts have been most severe during the past 150 years, with vegetation changes during the historical period exceeding that of the previous 1000 years by severalfold (Cole and others 1998). All tree species were present in the region by 3000 years before the present (Davis 1981). Recent archaeological evidence (Cleland 1992) and modern studies of fossil pollen and charcoal (Davis and others 1998, 1994; Cole and others 1998; Clark and Royall 1996) show that, with local exceptions, this region had a relatively stable fire disturbance regime over the 1000 years before the PLS. Major change in the abundance and distribution of forest types did not occur until after the PLS and the beginning of extensive Euro-American settlement and the industrial period. This reflects lower original Native American populations before contact and lower impact activities (that is, a more subsistence-based culture versus agriculture) than were prevalent in the prairie and savanna regions of southern Wisconsin (Cleland 1992). Northern Wisconsin is also more climatically limiting to fire, either of lightning or Native American origin, though both have occurred.

Knowledge of conditions that prevailed before the broad-scale and dramatic alterations of the landscape over the last 150 years of the industrial period is often particularly valuable. A primary source for the research and applications described above is the PLS data of the US General Land Office (GLO) (Stewart 1935). These data are contained in the field notebooks of the original US government survey, which was carried out largely by contract surveyors. The rectangular survey was in place from 1785 to 1910 and used from Ohio westward during the survey of the public domain, including Florida to Louisiana in the southern United States

(Stewart 1935). Detailed descriptions and evaluations of the data have been done many times (Bourdo 1956; Whitney 1994; Delcourt and Delcourt 1996; Manies and Mladenoff 2000). Potential problems concern (a) the consistency, expertise, and care of the surveyors who recorded the data; (b) the fact that data collection was not primarily for ecological purposes; and (c) occasional instances of outright fraud (Stewart 1935; Bourdo 1956). Instances of fraud were usually quickly identified and the surveys were redone. Surveyor variability and possible bias in the survey data exist, due to the methods used by the various individuals who carried out the data collection (Manies and others 2001). These factors can and should be assessed (Schulte and Mladenoff 2001), but they are often of reduced importance in large regional applications (Manies and Mladenoff 2000).

The PLS surveyors placed posts or other markers at section corners, dividing the land into a grid of 1 × 1 m (1.6 × 1.6 km) sections and townships containing 36 mi^2. Posts were also set at quarter section corners (0.5-mi [0.8-km] points on each section line) and meander points (locations where lines intersected navigable water bodies or lakes that were circumnavigated) (Stewart 1935). Changes occurred periodically in the instructions to surveyors, and these changes should be assessed in any application of the data. They can include instructions on the number, size, and selection criteria of trees, and the frequency of trees recorded along survey lines between corner points. Surveys in our study region were done from 1835 to 1891; most of the region was done from 1851 to 1863. Changes in the instructions that affected the data we use here were minor and do not affect our use (Onsrud 1979; Hawes 1882). The main data collected by the surveys of ecological use were the records of two to four trees, one in each quadrant of the compass surrounding each survey point. The species and diameter of each tree and its distance and compass bearing from the survey post were recorded. Varying amounts of descriptive notes were also recorded at each 1-mi point, which varied with changes in instructions as well as implementation by various surveyors. Typically, at each 1-mi point, the forest overstory and understory tree and shrub species were very briefly described. Natural disturbances, such as fire and windthrow, along with roads, trails, villages, or other cultural features, were noted at these locations, as well as when encountered along the survey line.

The quantitative PLS data can be used to construct regional tree species maps of regions or further aggregated and classified according to users needs (He and others 2000). The tree distance and bearing data may be used across an aggregation of points to calculate mean distance between trees, yielding mean area per tree, and converted to approximate density. Technically, because the survey instructions constrained the selection of trees, the data are not a truly random sample and may deviate from such a sample of the tree species density and diameters. Therefore, we recommend the use of relative measures derived from the data (Schulte and Mladenoff 2001). This method is identical to the distance sampling methods used for vegetation (Cottam and Curtis 1956). Various assumptions for statistical use of the data, and of any methods applied, should be tested and assessed by the user.

One significant problem that has typically arisen in such uses of the data has been the incidence of ambiguously identified tree species on the part of surveyors. For example, in the northcentral United States, and specifically in our northern Wisconsin study region (Figure 1), congeners such as white or yellow birch (*Betula papyrifera* or *B. alleghaniensis*), red or sugar maple (*Acer rubrum* or *A. saccharum*), and other such groups may be identified with ambiguous common names or only to genus (Table 1). In the case of birch in our region, *B. papyerifera* may be called "white birch", "paper birch", or "birch"; *B. alleghaniensis* may be called "yellow birch" or "birch". With this species pair, only the designation "birch" is ambiguous. However, the typical approach with such species has been to group them conservatively at the most general level (here "birch") for any analysis and mapping (for example, Finley 1976), unless the being area analyzed is small enough that it may be assumed, based on other factors, that all "birch" are white or yellow birch (White and Mladenoff 1994).

A similar problem occurs among the oak group in this region (Table 1). Oak trees were often described by surveyors as generic "oak", or more frequently as "black oak." There are four species of oak occurring in northern Wisconsin at varying abundances. Two are in the white oak group (*Lepidbalanus*): white oak proper (*Quercus alba*) and bur oak (*Q. macrocarpa*). These are of relatively minor occurrence in northern Wisconsin. There are also two in the black or red oak group (*Erythrobalanus*); red oak (*Quercus rubra*) and northern pin oak (*Q. ellipsoidalis*). Based on historical and current known tree ranges (Hansen 1992), we determined that black oak proper (*Q. velutina*) occurs only in southern and central Wisconsin. Species in the study region so designated must therefore be either red or northern pin oak, all of which are superficially similar in leaf type though they differ substantially in acorn form

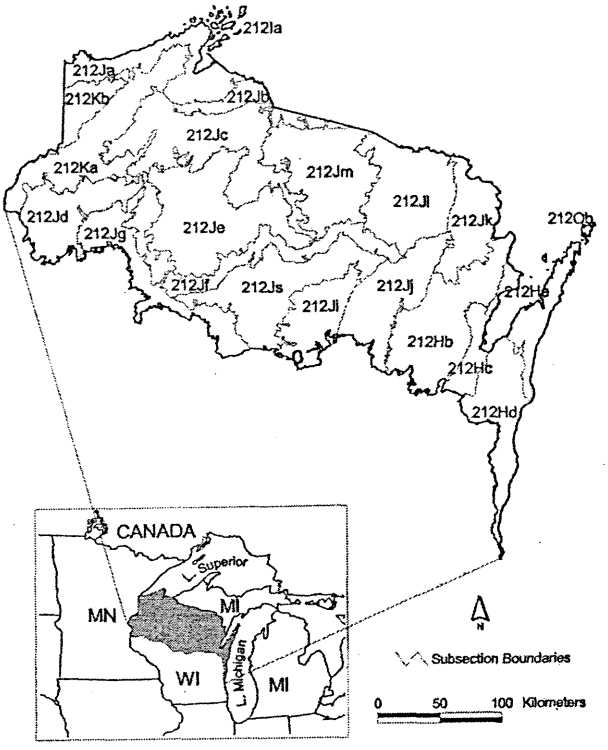

Figure 1. Map of study region in northern Wisconsin, USA. Area outlined is Province 212, the Laurentian Mixed Forest, based on USDA Forest Service hierarchical land classification (Keys and others 1995; WiDNR 1999). Province 212 is described as a region of transition between the northern boreal zone and the more southern broadleaf deciduous zone; it constitutes the northern forested region of the state. Subsection descriptions are in Table 3.

and other characters. The tree species of the substantial number of trees designated "oak" or "black oak" by the surveyors in northern Wisconsin are therefore unclear.

Clearly, ecological information is lost when ambiguity requires generalizing the species data to genus, such as "birch" or "oak", particularly where aggregated congeners differ ecologically and where a relatively high proportion of individual trees are undifferentiated. This is illustrated by separate maps of the distribution of tree species identified to the species (Figure 2a) or genus levels (Figure 2b) in the study region. There are abundant occurrences of major genera with two or more species occurring in the region (particularly birch, maple, and pine). Of the 194,629 trees in the database for our study, by major genera 10.4% of pine, 32.4% of maple, 80.4% of birch, 37.2% of oak, and 41.1% of ash were identified ambiguously. To create a map of the complete data as they exist, only overall genera can be mapped (Figure 3). Because the congenerics differ significantly in their ecological characteristics, a great deal of information is lost concerning past biogeography and range distribution, as well as information on more fine-grained patterns of species co-occurrence and abundance, and environmental and disturbance relationships. This generic mapping has typically been done, and it approximates the generic level available through most pollen data. At times, subjective interpretation and mapping has been performed by attempting to infer the species level by combining generic level data with maps of soils and other data (Finley 1976; Veatch 1928). This approach cannot usually be replicated consistently and often results in mapped classes that are inconsistent in their hierarchical level within a classification and of unknown accuracy.

Our purpose in this paper is to improve the use and utility of historical data and maximize information gain from pre-European vegetation mapping. To do so, we develop logistic regression models that use the tree information in the PLS and data from other sources to probabilistically classify these ambiguously identified trees to the species level. We then apply the reclassified data in paired maps to illustrate the ecological information gained in using this approach to map species distribution and abundance. We will particularly refer to details from the classification of the two birch species to illustrate the procedure and results.

Study Region

The study region is northern Wisconsin, USA, which is contained within Province 212 (Figure 1) of the US Forest Service hierarchical land classification system (Keys and others 1995; WiDNR 1999). Climate is continental with cool, short summers (July mean temperature, 20°C) and moisture distributed throughout the year (annual precipitation, 760–870 mm/y), and long, cold, and snowy winters (January mean temperature, −12°C). The area encompasses most of the northern forested region of the state, a glacially formed landscape of outwash plains, moraines, and till plains. Soils are sandy on outwash and loamy on the till landforms, with considerable heterogeneity within these subareas. Glacially derived wetlands and lakes are common and locally abundant on outwash. Fire was historically most important on the drier, sandy soils (Curtis 1959), as was windthrow in the deciduous and former hemlock-hardwood forests of till and moraines (Frelich and Lorimer 1991; Canham and Loucks 1984).

Currently pine (*Pinus*) and oak (*Quercus*) are common on the sandier soils, with aspen (*Populus*) and

Table 1. Ambiguously Identified Species in the US General Land Office's Public Land Survey Notes for the Northern Forested Region of Wisconsin

Tree Genus	Common Species Names used by Surveyors	Assumed Species	Generic Common Name
Acer			Maple
	Red maple, R maple	*A. rubrum*	
	Sugar, Sugar maple, Hrd maple, Rock maple	*A. saccharum*	
	Maple	ambiguous	
Betula			Birch
	Yellow birch, Y birch	*B. alleghaniensis*	
	White birch, W birch, Wht birch, Paper birch	*B. papyrifera*	
	Birch	ambiguous	
Fraxinus			Ash
	White ash, Wht ash	*F. americana*	
	Black ash, Blk ash, B ash, Brown ash	*F. nigra*	
	Ash	ambiguous	
Pinus			Pine
	Red pine, R pine, Norway pine, N pine, Y pine, Sugar pine	*P. resinosa*	
	White pine, W pine, Wht pine	*P. strobus*	
	Jack pine, J pine, Jk pine, Pitch pine, Black pine	*P. banksiana*	
	Pine	ambiguous	
Quercus			Oak
	Red oak, R oak	*Q. rubra*	
	White oak, W oak, Wht oak	*Q. alba*	
	Black oak, B oak, Blk oak	ambiguous	
	Jack oak, J oak, Jk oak, Pin oak, Spanish oak, Yellow oak	*Q. ellipsoidalis*	
	Bur oak, Br oak, Burr oak	*Q. macrocarpa*	
	Oak	ambiguous	
Ulmus			Elm
	White elm, W elm	*U. americana*	
	Red elm, Slippery elm	*U. rubra*	
	Elm	ambiguous	

paper birch (see Table 1 for a complete list). Deciduous species such as sugar maple now dominate on heavier soils, where eastern hemlock and yellow birch were also a major dominant before widespread logging in the past. From about the 1850s to the early 1900s, pine and hemlock were largely eliminated despite their former dominant roles. Aspen and birch also are common now on better soils in former disturbance areas. The region was entirely logged before and during the Euro-American settlement period, from the mid-1800s to the early 1900s. Although some areas remain in agriculture, most of the region is reforested. Public forest lands, private industrial forests, and individual ownership are all common (Mladenoff and Pastor 1993).

Our source of the PLS data was microfiche copies of the original surveyor notebooks, obtained from the State of Wisconsin Office of Public Lands. Data were entered into computer files, following a protocol developed to standardize data entry and provide for quality control and error checking (Manies and others 2001).

Methods and Approach

Model Formulation

Logistic regression has often been used to describe the relationship between a dichotomous dependent variable and multiple independent variables, both continuous and categorical (Ek and Monserud 1979; Teck and Hilt 1990; Hamilton and Edwards 1976; Dahir 1994; Mladenoff and others 1995). It has the advantages of restricting the response variable to the range of 0 to 1 and of being biologically interpretable. Logistic modeling is also relatively robust, performing equally well with multivariate normal or categorical independent variables (Press and Wilson 1978). The dichotomous dependent variable may describe the occurrence of an event

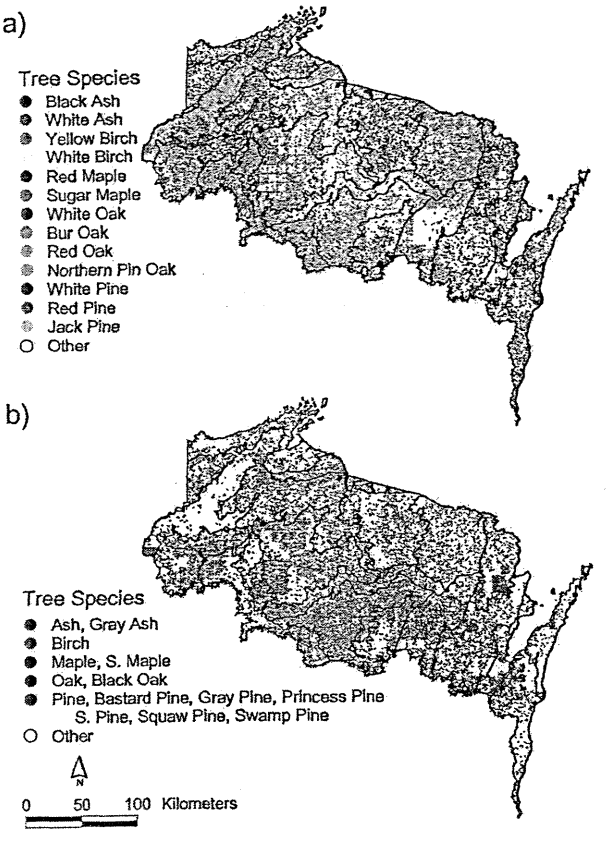

Figure 2. Maps of PLS data recorded at survey points for genera with multiple species occurring within the study region. (a) Data points of trees identified unambiguously to species level. (b) Data points with trees described ambiguously and identifiable only to genus. Species scientific names are in Table 1.

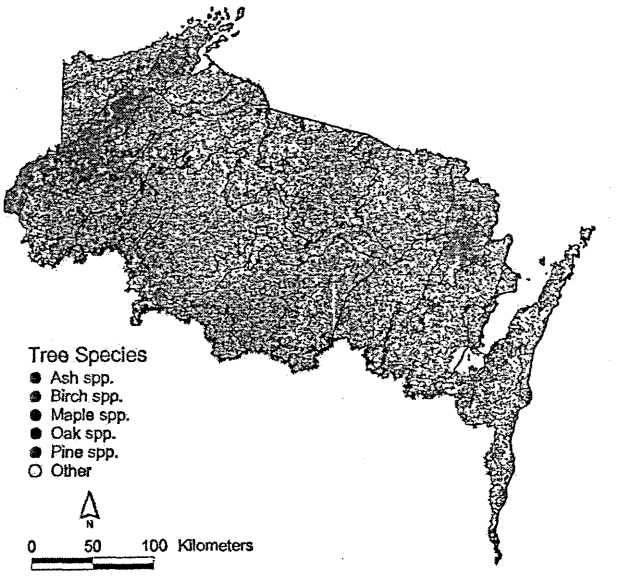

Figure 3. PLS data points combining ambiguous and unambiguous data (see Figure 2a and b) generalized to lowest common denominator (genus level). Although the resulting map is more complete spatially using all data, significant information is lost among congeners that differ ecologically.

Table 2. List of Variables Used in the Logistic Regression Models

Continuous Variables	Categorical Variables (P/A)
Diameter	Associated tree species at corner
Distance from survey corner	USFS[a] subsection (glacial landform)
Corner mean density	USFS LTA[b] (general soil type)
Corner mean tree diameter	
Geographic location (survey township/range; transformed)	

[a] *United States Forest Service.*
[b] *Land type association.*

(1 = occurrence, 0 = nonoccurrence), such as in forecasting the probability of tree mortality (Hamilton and Edwards 1976; Ek and Monserud 1979; Buchman 1983; Dahir 1985; Teck and Hilt 1990). It may also describe membership in one of two groups (Mladenoff and others 1995). The model takes the following form:

$$\pi(x) = e^{(x)}/(1 + e^{(x)}) \text{ where } x = \alpha + \beta_1 x_1 + \ldots + \beta_n x_n \quad (1)$$

where x = vector of predictor variables and coefficients estimated by the model α and $\beta_1 \ldots \beta_n$ = unknown parameters that are estimated from the data.

The data set that is used to calibrate the logistic model consists of observations, each having a known outcome or membership, and several predictive variables. Variables that are significant predictors of the outcome are selected, usually by forward stepwise regression. In this case, they are environmental variables and known tree species characteristics and associated species (Table 2). The model generated from this process is then used to categorize a second data set consisting of the same predictive variables but an unknown outcome or membership. The probability of membership or oc-

Table 3. Description of US Forest Service Subsections on Which Yellow and White Birch Predominantly Occur

Subsection	Name	Dominant Soils	Dominant Habitat Type	Major Tree Species	% Yellow Birch	% White Birch
212Js	Lincoln formation till plain	Silty loarn, loam	Mesic, wet mesic	Sugar maple, yellow birch, hemlock	100	0
212Je	NW & central WI loess plains	Silty soils over loamy till	Wet mesic, mesic	Sugar maple, yellow birch, hemlock	96	4
212Jf	Perkinstown end moraine	Sandy loams	Mesic, dry mesic	Sugar maple, yellow birch, hemlock	93	7
212Ka	Northwest WI sand plain	Excessively drained sands	Very dry to dry	Scrub oak, jack & red pine	11	89
212Jm	Northern Highlands pitted outwash	Sand, loamy sand	Very dry, dry to dry mesic	White & red pine, aspen	25	75
212Ja	Lake Superior clay plains	Clays, loamy clay	Wet, wet mesic	Spruce–fir, tamarack, aspen	30	70

currence is calculated for each observation in the unknown data set. This estimate is most often compared to a uniform random number between 0 and 1. If the estimate equals or exceeds the random number, the event being modeled is assumed to have occurred, or the entity is considered to be a member of the predictor group. In our analysis, logistic regression was used to determine whether the species of trees that had been classified only to genus could be predicted. For example, we classified generically identified "birch" to be either yellow birch (*B. alleghaniensis*) or paper birch (*B. papyrifera*) (Table 3).

For some species, there are more than two possible choices for species within a genus. For instance, in our region there are three dominant pine species, eastern white pine (*Pinus* strobus), red pine (*P. resinosa*), and jack pine (*P. banksiana*). Separate logistic models were run for each pair: (a) white pine versus red pine, (b) red pine, versus jack pine, and (c) white pine versus jack pine. To maintain a consistent relationship between the species, the same variables must be accepted into each model (that is, by setting the entry *P* value to 0.99). A combined estimate for each species could then be calculated from the relative likelihoods predicted by each pair of models, which must sum to one for any particular corner. For instance, the calculation of the probability that a given pine tree is a red pine (π_i, [RP]) is as follows:

$$\pi_i\,[\mathrm{WP}] + \pi_i\,[\mathrm{JP}] + \pi_i\,[\mathrm{RP}] = 1 \qquad (2)$$

$$\pi_i\,[\mathrm{WP}] = \pi_i\,[\mathrm{RP}]\left[\frac{\pi i\mathrm{WP}}{\pi_i\mathrm{RP}}\right] \qquad (3)$$

$$\pi_i\,[\mathrm{JP}] = \pi_i\,[\mathrm{RP}]\left[\frac{\pi i\mathrm{JP}}{\pi_i\mathrm{RP}}\right] \qquad (4)$$

$$\pi_i\,[\mathrm{RP}]\left[\frac{\pi i\mathrm{WP}}{\pi_i\mathrm{RP}}\right] + \pi_i\,[\mathrm{RP}]\left[\frac{\pi i\mathrm{JP}}{\pi_i\mathrm{RP}}\right] + \pi_i\,[\mathrm{RP}] = 1 \qquad (5)$$

$$\pi_i\,[\mathrm{RP}] = \frac{1}{\left(\left[\frac{\pi i\mathrm{WP}}{\pi_i\mathrm{RP}}\right] + \left[\frac{\pi i\mathrm{JP}}{\pi_i\mathrm{RP}}\right] + 1\right)} \qquad (6)$$

where $\left[\frac{\pi i\mathrm{JP}}{\pi i\mathrm{RP}}\right]$ = the probability that tree *i* is a jack pine based on the model of jack pine versus red pine for $i = 1 \ldots n$, and $\left[\frac{\pi i\mathrm{WP}}{\pi i\mathrm{RP}}\right]$ = the probability that tree *i* is a white pine based on the model of white pine versus red pine for $i = 1 \ldots n$.

Variables Used

The known data set consisted of trees that had been classified to species. Independent predictor variables included tree and sample point characteristics (diameter, mean tree diameter, and forest density), geographic location (here township and range), ecoregion (subsection and land type association) (Keys and others 1995; WiDNR 1999), as well as the species of neighboring trees and the target tree distance from the corner. The estimation of tree density at a point was based on the number and distance between all recorded trees at a corner (Manies and Mladenoff 2000; Anderson and Anderson 1975), here based on units as recorded by the GLO surveyors (1 ft = 30.48 cm; 1 link = 20.12 cm):

$$\text{Density} = (1/\text{MA})(107600\ \text{ft}^2/\text{ha}) \quad (7)$$

$$\text{MA} = [(\Sigma\ d_i/n)/c \cdot 0.66\ \text{ft/link}]^2 \quad (8)$$

where MA = mean area per tree in ft^2, d_i = distance of tree i from corner in links for $i = 1 \ldots n$, n = total number of trees at corner, and c = multiplier based on n: if $n = 1$, $c = 0.50$; if $n = 2$, $c = 0.66$; if $n = 3$, $c = 0.81$; if $n = 4$, $c = 1.00$.

The occurrence of neighboring tree species at a corner was coded as a dichotomous variable—that is, as presence or absence, regardless of the number of trees of each species present at the corner.

Geographic location is an important predictor for many tree species. Many have ranges that are restricted either to far northern or eastern parts of the state. Both township and range were entered in the regressions. Township, a continuous variable increasing from south to north (17N–53N), was entered without transformation, whereas range was converted to a continuous variable ranging from 30E (coded as 1) to 20W (coded as 50). The product of township and range could then be considered a measure of northwesterly location within the state. Subsection and land type association (LTA) variables were used as indicators both of geographic location and as site descriptors.

Spatial dependence or autocorrelation is common in spatial data; if present, it must be taken into account for the assumptions of classical regression analysis to be met. It is probable that spatial autocorrelation exists within the distribution of tree species in northern Wisconsin and in the historical data. We tested for autocorrelation using data from the unambiguously identified tree species. Omnidirectional and directional semivariograms (S-PLUS 2000 for windows; MathSoft Corporation, Seattle, WA, USA) were run for the dominant tree species (that is, white pine, jack pine, eastern hemlock, yellow birch, white birch, and American beech) at both the province and subsection scales. Resulting semivariograms lacked a clearly definable sill or range; patterns included straight lines, humps, and waves. We interpreted these patterns as a lack of autocorrelation within the data at the scale of the surveyor samples (points 0.5 km apart). Obvious clumping of species does occur at other scales, such as with jack pine on sand plains. This was not a scale that affects our models at the survey point scale. On the basis of this screening, spatial dependence was not built into the species models.

Model Assessment

The performance of the final model for each species group was assessed by two methods: first, by comparing observed and predicted values based on the entire data set; and second, by using randomly selected subsets of the data for model calibration and then assessing model fit on the remaining subset. Model fit was evaluated with the chi-square (χ^2) statistic and by regression of predicted on observed values (Hamilton and Edwards 1976; Ott 1988). In this study, we classified the data by subsection, a grouping that is broad enough to minimize the occurrence of categories with too few observations yet specific enough to be a meaningful basis for comparison. The observed value for a category was then the number of corners of the modeled species x in that subsection. The estimated value was equal to the sum of all estimates for species x (based on the model fit) within the subsection. The interpretation of test results, however, is quite different from the traditional chi square. In this case, the lower the χ^2 statistic—that is, the higher the P value—the greater the evidence in support of the null hypothesis of equality between observed and estimated numbers of trees per subsection and the greater the predictive ability of the model.

A statistic used by Buchman (1983) to evaluate logistic model results is based on a regression of the estimated number per class on the observed number. We have regressed the total number of trees in each subsection for both the dominant and subordinate species, resulting in twice the number of degrees of freedom. Buchman reported only the slope, but we have also reported the coefficient of determination as an indication of the amount of variance of individual estimates about the regression line.

Model Validation

A separate validation procedure was performed and repeated three times for each model. The purpose of this validation was to test the robustness of the final logistic model—that is, the model's predictive ability—against data not used in calibration. For each trial, the original data set of known tree species was randomly divided into two subsets. The first subset consisted of two-thirds of the species occurrences and was used for the calibration of a separate logistic model that was then used to classify the remaining smaller (one-third) data set. Because these were previously identified corners, the results of this estimated classification could be compared to the observed values. Again, the chi-square statistic and regression of estimated on observed values were used to assess fit. All logistic regression analyses were carried out using the SAS software (SAS Institute Inc.).

Table 4. Description of Models and Qualifying Statistics[a]

Model	Species	No. of Corners	No. of Trees	−2 LL (d.f.)[b]	Maximum % Correct Prediction	Chi-square *P* Value[c]	Regression[d] Slope	Regression[d] R^2
Ash	White	577	650	567 (29)	81.3	0.879	1.001	0.989
	Black	2216	2795					
Birch	Yellow	3423	4682	3436 (53)	85.6	1.000	0.999	1.000
	White	1631	2209					
Elm	American	78	83	28 (6)	81.9	0.955	1.012	0.973
	Red	22	22					
Maple	Sugar	12903	17824	1212 (21)	96.4	1.000	1.003	1.000
	Red	501	510					
Oak	Red	618	717	324 (12)	93.6	1.000	0.999	1.000
	Northern pin	210	272					
Pine	White	20160	28552	20278 (54)	80.8	0.980	0.998	0.997
	Red	10827	16327					
	White	20160	28552	8746 (45)	93.3	1.000	1.000	1.000
	Jack	7436	13508					
	Red	10827	16327	10789 (27)	82.6	0.995	1.005	1.000
	Jack	7436	13508					
Mean					86.9	0.976	1.002	0.995

[a]*The −2 Log Likelihood (−2 LL) is commonly reported in logistic regression output as an assessment of goodness of fit. The −2 LL increases with the number of observations as well as with the number of variables in the model and is therefore not comparable between models based on different data sets. The statistic follows a chi-square distribution with degrees of freedom equal to one less than the number of variables in the model plus the intercept. The maximum percent correct prediction is based on a comparison of the predicted value to successive probability levels.*
[b]*P values are all <0.0001.*
[c]*There were 19 degrees of freedom for all models except elm, which had 11 degrees of freedom.*
[d]*Regression of predicted numbers per subsection on observed with 39 degrees of freedom for all models except elm, which had 23 degrees of freedom.*

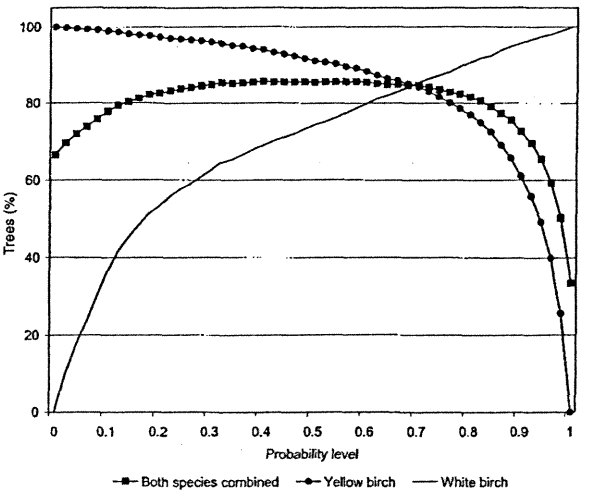

Figure 4. Classification values for white and yellow birch at various probability limits and both species combined.

Results and Discussion

Model Assessment

The average correct prediction for all species models is 86.9% using the full data set for each model (Table 4). This value generally exceeds levels considered acceptable and useful, for example, in the classification of forests using remote sensing imagery (Congalton 1991) and is within typical error rates in forest field inventory data (Hansen 1992). Results are illustrated by the birch model. At a probability level of 0.48, 85.6% of all birch corners are predicted correctly by the model (Figure 4). As the probability level increases, the percentage of the modeled species (here yellow birch) that is correctly predicted decreases and the percentage of the species not modeled (white birch) increases. This means that when all corners with an estimate equal to or greater than 0.48 are classified as yellow birch and all corners with estimates less than 0.48 are classified as white birch, 85.6% of the total number of corners agree with the observed value (Figure 4).

Highest predictions (93%–96%) occurred for the maples, oaks, and white/red pine models. Lowest predictions (81%–82%) occurred for the ash, elm, and red/jack pine models. For oak, only the red oak–northern pin oak model and validation is shown, since the numbers of ambiguous "oak" were too few in the study area (136) for validation across all possible oak outcomes.

Table 5. Results of Validation Procedure

Model	Trial	−2 Log Likelihood (d.f.)[a]	Maximum % Correct Prediction	Chi-square *P* Values[b]	Regression of Estimated on Observed	
					Slope	R^2
Ash	1	392 (23)	81.5	0.992	0.945	0.986
	2	339 (18)	80.6	0.998	1.043	0.986
	3	421 (23)	80.9	1.000	1.024	0.992
Birch	1	2109 (40)	85.3	0.956	0.998	0.996
	2	2099 (41)	85.1	0.999	1.031	0.998
	3	2058 (40)	85.8	0.599	1.032	0.982
Elm	1	31 (4)	87.1	[c]	0.879	0.968
	2	56 (4)	90.3	[c]	0.964	0.510
	3	30 (4)	85.5	[c]	1.133	0.839
Maple	1	782 (19)	96.4	1.000	1.000	1.000
	2	869 (21)	96.4	0.653	0.992	0.995
	3	840 (21)	96.5	1.000	0.999	1.000
Oak	1	440 (10)	95.4	0.999	0.954	0.963
	2	423 (10)	94.3	1.000	1.013	0.995
	3	485 (13)	94.9	1.000	0.983	0.996
Pine						
Red vs white	1	2591 (44)	80.7	0.997	0.984	0.996
	2	2481 (42)	79.8	0.994	0.991	0.996
	3	2346 (34)	79.6	0.994	0.991	0.996
White vs jack	1	5834 (44)	93.1	1.000	1.001	1.000
	2	5650 (35)	92.4	1.000	1.005	0.999
	3	5715 (42)	92.8	1.000	0.998	0.999
Red vs jack	1	2426 (27)	82.6	0.997	0.995	0.999
	2	2408 (27)	83.2	0.999	1.010	0.998
	3	2480 (28)	82.7	1.000	1.002	0.999

d.f., degrees of freedom
[a]*P values all <0.0001*
[b]*No more than five cells with fewer than five observations included in each analysis*
[c]*Fewer than five cells with more than four observations invalidating chi-square calculation*

The results of the chi-square test (Table 4) show that most *P* values were above 0.95 and half were equal to 1.000, indicating that there was no significant difference between the observed and predicted numbers for each subsection. Regressions of estimated on observed values for each subsection have an average slope of 1.002 and an average coefficient of determination of 0.995 for all models, again indicating a very good fit between observed and predicted values (Table 4). There are twice as many degrees of freedom as in the chi-square test because both species are included in the regression, whereas only the dominant species was analyzed in the chi square.

Model Validation

Validation results show that with the exception of elm, all slope values are within the range of 0.945 and 1.043 and all R^2 values are greater than 0.98 (Table 5). The elm species were a very uncommon group in the original data set. For all other species, the three trials yielded results within 1%–2% of each other, showing the high stability of the models.

Tree Distribution Patterns

The birch species, of moderate abundance and classification accuracy (Table 4), illustrate the results of the overall procedure. White birch and yellow birch are differentiated by the model predominantly based on associated species and site factors. Yellow birch is most associated with long-lived mesic species such as hemlock and sugar maple, and subsections with loamy as opposed to sandy soils. Conversely, white birch is associated with other early seral species such as aspen, as well as the pines, and

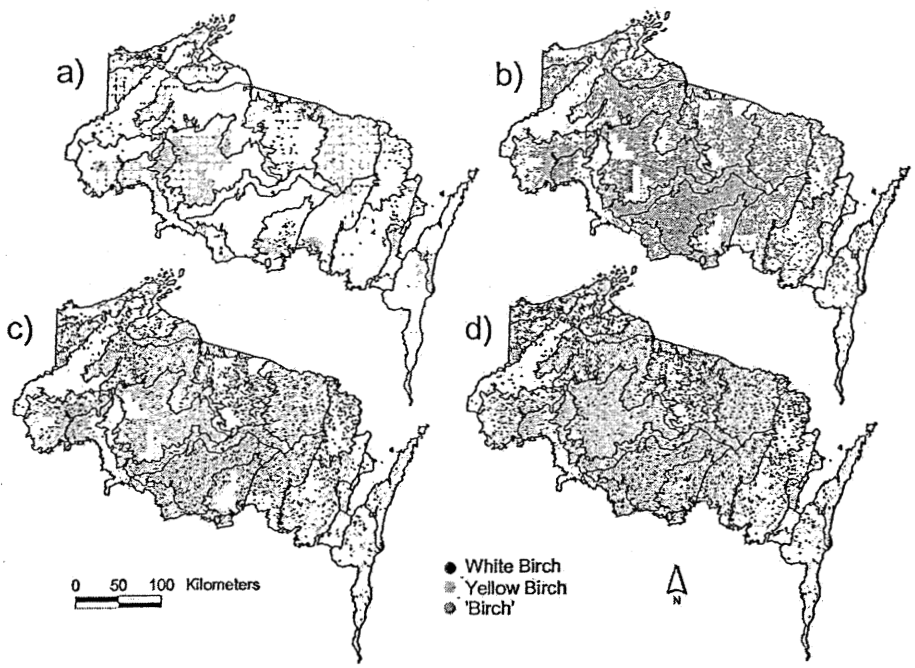

Figure 5. Locations of (a) known white and yellow birch corners in the study region, (b) unclassified ("birch") corners, (c) classified white and yellow birch corners based on the logistic model, and (d) locations of combined known and classified white and yellow birch corners (maps a, b).

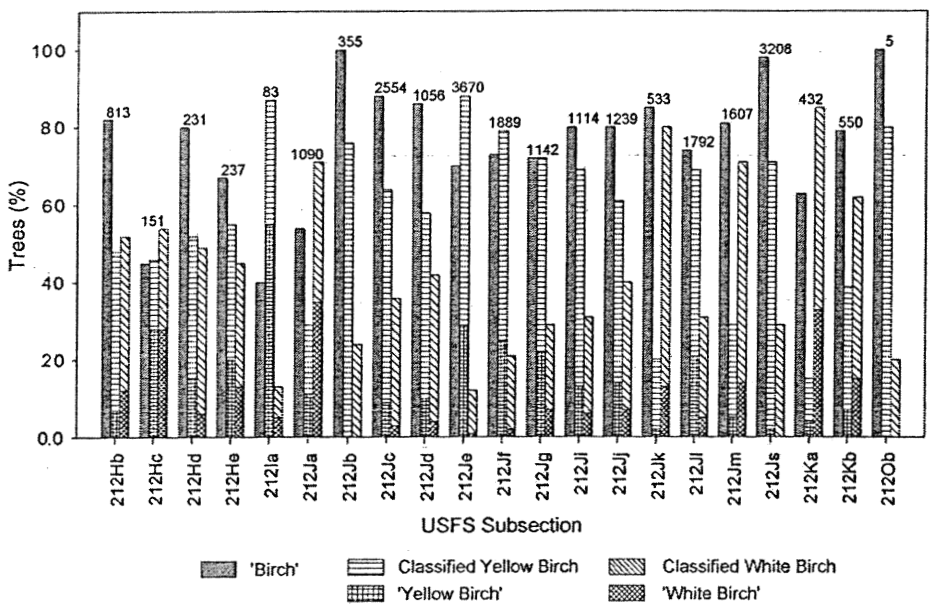

Figure 6. Proportions of birch tree species ambiguously identified by surveyors ("birch"), identified to species level by surveyors ("yellow birch," "white birch") and classified proportions based on the logistic regression model for yellow and white birch. Values at top of columns indicate total birch trees for each subsection.

characteristic of sandy soils on outwash, or following fire on mesic soils. Finer site breakdown within subsections by LTAs, indicative of finer soil differences within the glacially defined landforms of the subsections, proved helpful in refining the model further.

Maps of the birch species in the data set before and after classification (Figure 5) and the graphed summary by subsection (Figure 6) illustrate the decrease in ambiguity of the species distribution patterns and the increase in information available from the data. Based on the logistic model, the relationships of known white and yellow birch (Figure 5a) drive the relationships used to classify the generic "birch" (Figure 5b) into yellow and white birch (Figure 5c), resulting in the final combined map (Figure 5d). White birch dominates in different portions of the region, such as in the northwest

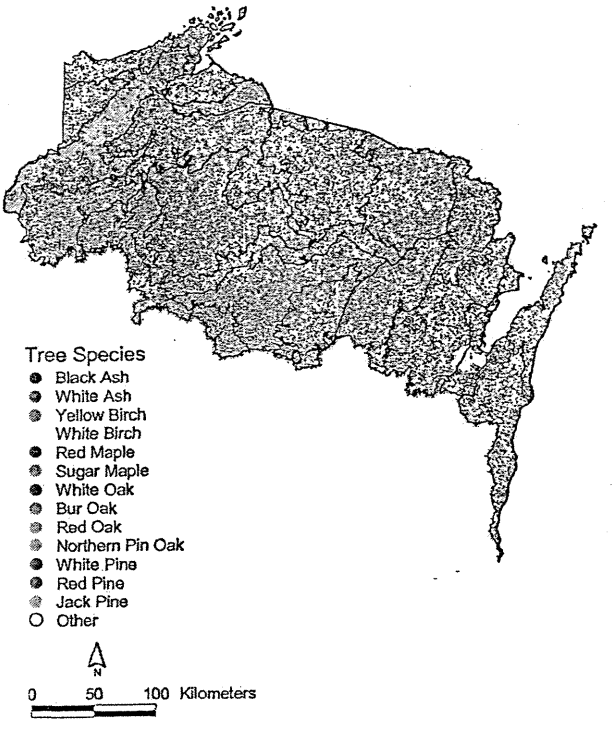

Figure 7. New combined map of differentiated and classified tree species. Species scientific names are in Table 1. For clarity, species that were not a subject of this analysis, such as eastern hemlock and American beech, are not shown on this map. However, hemlock was the leading dominant across northern Wisconsin on the mesic moraines and till plains that intervene among the sandy outwash plains.

(beyond the main contiguous, continental range of yellow birch), in sandy outwash plains in the north central and northeast portions of the region, and along the southwest and southern fringes of the region (Figure 5d). These latter areas are beyond the contiguous range and greatest abundance of yellow birch. The initial proportions of the three birch classes by subsection (Figure 6) reflect the mix of surveyors that worked across each subsection. The final classified proportions reflect tree distributional information gained through modeling, combined with knowledge of the ecological tolerances of the species.

The identification of known, differentiated species occurrences, such as that of white and yellow birch (Figure 5), is key to deriving the relationships that increase predictive accuracy in the logistic models. This is particularly true in subsections where both white and yellow birch occur and have been differentiated by a given surveyor (Figure 3a). Because surveyors would only rarely record any description of site, such as "swamp" or "barren", these classifications give a good, albeit general, indication of the soil, major forest cover, and glacial topography that could be expected at a specific corner. This is particularly valuable where species to be distinguished differ markedly in site preferences, as is the case with the two birch species (Table 3). For instance, yellow birch tends to grow best on well-drained fertile soils, silt loams, loams, and moderately well-drained sandy loams, and is most often associated with shade-tolerant species such as hemlock, sugar maple, and beech. Of the 20 subsections in northern Wisconsin on which yellow birch was identified to species, almost 40% of corners are found in two subsections in the central portion of the region, both dominated by rich silt-loam soils and northern hardwoods.

White birch, on the other hand, has a bimodal distribution (Curtis 1959); it is often more abundant on either drier or wetter, nutrient-poor sites and is most often associated with species such as aspen, white pine, and red pine. Again, almost 40% of white birch identified to species occurred either in the clay plains along Lake Superior, with poorly drained soils and dominated by aspen, spruce-fir, and pines, or on the pitted outwash located in the northcentral and eastern portions of the region with excessively drained sandy soils and dominated by pines and aspen. White birch also dominates in a band along the southwestern and southern margins of the study region. These latter may represent occurrences driven more by combined climatic and disturbance factors (that is, greater drought and fire) than in the north, where soil and substrate are better predictors. This distribution differs from that of the 20th century, where white birch and aspen increased dramatically across the north from 19th-century levels following widespread logging and fires.

The combined new map of all the classified and originally differentiated species (Figure. 7) reveals a clearer pattern of the relative abundance of the other congeners and the overall distribution of the presettlement forest dominants. The probabilistic and multivariate nature of the classification segregates species considering gradients and variables at a range of scales, from the broadest subsection classes, to species patterns of co-occurrence at individual points. Therefore, the relative abundance and differences in the spatial distribution of the three pine species are clarified on the three large sandy outwash plains. The poorest and coarsest sands occur in the northwest, indicated on the map by the great abundance of jack pine. Somewhat finer sands and more fertile outwash occur on the

pine plain of the northeast, still dominated by jack pine, but with increasing numbers of red and white pine. The northcentral outwash plain has the finest sands and is the most fertile of the three. It is dominated by white and red pine, but jack pine is also present. This gradient of soil texture and fertility on these three outwash plains also corresponds to a gradient of fire frequency, with highest fire occurrence where jack pine is most abundant (L. Schulte, Mladenoff DJ, Nordheim EV. subm. ms.).

Conclusions and Potential Use

The approach we have derived here has the potential to enhance the usability of historical data where alternatives, such as additional data collection, do not exist. The newly derived species distributions can be mapped individually, as shown for birch, as well as combined into an overall distribution map of the region. The resulting combined coverage can be rasterized and species dominance calculated at cell sizes that sample the region at varied scales, according to the required spatial resolution (He and others 2000). Similarly, the data can then be classified using multivariate methods, producing a hierarchical forest classification that can also be tailored to the classification resolution needed for research or management applications (L. Schulte, and other 2002). The differentiation of congeneric tree species can add considerably in efforts aimed at distinguishing patterns of wildlife habitat on the landscape. Maps resulting from such species differentiation can help refine the distribution of tree species, and help to literally fill in gaps of historical distribution patterns.

However, the contribution of the probabilistic differentiation must be understood and used at the proper scale, and it must be interpreted with species biology, as well as the nature of the data set, in mind. Due to their biology and response to environmental events, history, and current conditions, patterns of distribution vary among tree species. For example, in regions with long disturbance intervals, shade-intolerant, early successional species such as white birch may occur sparsely at broad scales, with local abundance in disturbance patches. Abundance on the landscape is greater on outwash plains where more frequent fire occurred. On the other hand, an important codominant like yellow birch, with greater shade tolerance, is more broadly distributed on the presettlement landscape. The nature of the data set is also important, because the number of known and ambiguous trees, their distribution across environments, and percent of the total will all influence the accuracy of resulting models differentiating unknown individuals. In most data sets, these numbers of individuals and other factors will vary among species. Therefore, differing accuracy levels will be contained within the final product. Additionally, because the method is probabilistic, any particular individual tree cannot be classified with certainty, even in a highly accurate model. We can have confidence in the derived identifications only in the aggregate and in the general pattern on the landscape. We have had success with logistic regression analysis, but other methods can be used in this type of analysis. Classification and regression tree analysis may be useful, especially in areas that are more species rich. In these cases more undifferentiated classes need to be included in a single model, such as with multiple oak species that may occur in more southern regions. Furthermore, the most reliable data set will result from pairing analyses that classify ambiguously identified tree species, as was done here with a quantitative and visual assessment of variability among surveyors within a given region (Manies and others 2001).

Careful thought must also be given to the ultimate use of the species classified using these methods. For example, if one wishes to examine the significance of species relationships with various environmental variables, the original portion of the data that are unambiguously identified must be used, to avoid circularity in the analysis; one cannot use environmental data to classify the species and then use the classified species to test for environmental relationships. This is another reason that argues for a conservative approach in an analysis to classify undifferentiated tree species. Using parsimonious models will allow other relationships to be tested among environmental variables that have not been used in model building, if they are suitably independent and high colinearity with variables used to build the models is not present (L. Schulte and others 2002). In regions with stronger environmental gradients than our study region, finer spatial classification may be possible. Regions with fewer species than are contained in our region may enhance the result using this process, or require different methods, depending on the relative abundance of identified and unidentified species and environmental gradients (J. Bolliger, E. V. Nordherm, and D. L. Mladenoff subm. ms.). It is unclear how these methods may apply to more species-rich regions. Based on our results, we speculate that our methods may be most successful under conditions where a statistically adequate number of individuals of each species are identified, this providing an adequate training set for model development, and

when a species niche can be well defined by obtainable variables, such as species associations, environmental preferences, or disturbance responses.

With these considerations in mind, the approach we have described here has significant potential to increase the value of historical data sets that were not collected for ecological purposes and do not have consistent species identification. The PLS data set is one common example of such data, but other similar information could also be used, such as colonial surveys (Foster and others 1998; Russell 1997, 1981). The PLS data have the advantage of existing for much of the United States. Careful analysis of the data can provide significant insight into past species composition and tree species relationships with environment, disturbance, and other factors.

ACKNOWLEDGMENTS

We thank the Forest Landscape Ecology Lab group for helpful comments and discussions on this work. The Wisconsin Department of Natural Resources has provided continuing funds for this project under the Pitmann-Robertson Program. Additional funding was provided by the USGS Biological Resources Division. Valuable technical assistance was provided by T. Sickley. Comments from J. Bolliger, H. Delcourt, and two anonymous reviewers improved both the analysis and the manuscript.

REFERENCES

Anderson RC, Anderson MR. 1975. The presettlement vegetation of Williamson County, Illinois. Castanea 40:345–63.

Baker WL. 1995. Long-term response of disturbance landscapes to human intervention and global change. Landscape Ecol 10:143–59.

Baker WL. 1994. Restoration of landscape structure altered by fire suppression. Conserv Biol 8:763–9.

Bourdo EA Jr. 1956. A review of the General Land Office survey and of its use in quantitative studies of former forests. J Ecol 37:754–68.

Buchman RG. 1983. Survival predictions for major Lake States tree species. USDA Forest Service Research Paper NC-233. St. Paul (MN). North Central Forest Experiment Station. 8 p.

Bürgi M, Russell WWB, Motzkin G. 2000. Effects of postsettlement human activities on forest composition in the northeastern United States: a comparative approach. J Biogeogr 27:1123–38.

Canham CD, Loucks OL. 1984. Catastrophic windthrow in the presettlement forests of Wisconsin. Ecology 65:803–9.

Cissel JH, Swanson FJ, Weisberg PJ. 1999. Landscape management using historical fire regimes; Blue River, Oregon, Ecol Appl 9:1217–31.

Clark JS, Royall PD. 1996. Local and regional sediment charcoal evidence for fire regimes in presettlement north-eastern North America. J Ecol 84:365–82.

Cleland CE. 1992. Rites of conquest: the history and culture of Michigan's Native Americans. Ann Arbor (MI): University of Michigan Press.

Cole KL, Davis MB, Stearns F, Guntenspergen GG, Walker K. 1998. Historical landcover changes in the Great Lakes region. In: Sisk T, editor. Perspectives on the land use history of North America: a context for understanding our changing environment. USGS/BRD/BSR-1998-0003. Springfield (VA): US Geological Survey, Biological Resources Division. p 43–50.

Congalton RG. 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens Environ 37:35–46.

Cottam G, Curtis JT. 1956. The use of distance measures in phytosociological sampling. Ecology 37:451–60.

Covington WW, Everett RL, Steele R, Irwin LL, Daer TA, Auclair AND. 1994. Historical and anticipated changes in forest ecosystems of the inland west of the United States. In: Sampson RN, Adams DL, editors. Assessing forest ecosystem health in the inland west. Binghamton (NY): Haworth Press. 13–63.

Curtis JT. 1959. The vegetation of Wisconsin: an ordination of plant communities. Madison (WI): University of Wisconsin Press. 657 p.

Dahir SE. 1994. Tree mortality and gap formation in old-growth hemlock–hardwood forests of the Great Lakes region [PhD dissertation]. Madison (WI): University of Wisconsin–Madison.

Davis MB. 1981. Quaternary history and the stability of forest communities. In: West DC, Shugart HH, Botkin DB, editors. Forest succession. New York. Springer-Verlag. p 132–53.

Davis MB, Calcote RR, Sugita S, Takahara H. 1998. Patchy invasion and the origin of a hemlock-hardwoods forest mosaic. Ecology 79:2641–59.

Davis MB, Sugita S, Calcote RR, Ferrari JB, Frelich LE. 1994. Historical development of alternative communities in a hemlock–hardwood forest in northern Michigan, USA. In: Edwards PJ, May R, Webb NR, editors. Large-scale ecology and conservation biology. Oxford (England): Blackwell. p 19–39.

Delcourt HR, Delcourt PA. 1996. Presettlement landscape heterogeneity: evaluating grain of resolution using General Land Office Survey data. Landscape Ecol 11:363–81.

Ek AR, Monserud RA. 1979. Performance and comparison of stand growth models based on individual tree and diameter-class growth. Can J For Res 9:231–44.

Forman RTT, Russell EWB. 1983. Evaluation of historical data in ecology. Bull Ecol Soc Am 64:5–7.

Foster DR, Motzkin G., Slater B. 1998. Land-use history as long-term, broad-scale disturbance: regional forest dynamics in central New England. Ecosystems 1:96–119.

Foster DR, Orwig DA, McLachlan JS. 1996. Ecological and conservation insights from reconstructive studies of temperate old-growth forests. Trends Ecol Evol 11:419–24.

Foster DR, Schoonmaker PK, Pickett STA. 1990. Insights from paleoecology to community ecology. Trends Ecol Evol 5:119–22.

Frelich LE, Lorimer CG. 1991. Natural disturbance regimes in hemlock-hardwood forests of the upper Great Lakes region. Ecol Monogr 61:145–64.

Grimm EC. 1984. Fire and other factors controlling the Big Woods vegetation of Minnesota in the mid-nineteenth century. Ecol Monogr 54:291–311.

Hamilton DA, Edwards BM. 1976. Modeling the probability of individual tree mortality. USDA Forest Service Research Paper

INT-185. Ogden (UT): Intermountain Forest and Range Experiment Station, 15 p.

Hansen MH. 1992. The Eastwide forest inventory database: user's manual. USDA Forest Service General Technical Report NC-151. St. Paul (MN): North Central Forest Experiment Station.

Hawes JH. 1882. Manual of United States surveying. Phildelphia: JB Lippincott.

He HS, Mladenoff DJ. 1999. Spatially explicit and stochastic simulation of forest-landscape fire disturbance and succession. Ecology 80:81–99.

He HS, Mladenoff DJ, Sickley T, Guntenspergen G. 2000. GIS interpolations of witness tree records (1839–1866) for northern Wisconsin at multiple scales. J Biogeogr 27:1031–42.

Heinselman ML. 1973. Fire in the virgin forests of the Boundary Waters Canoe Area, Minnesota. Quat Res 3:320–82.

Keys J Jr, Carpenter C, Hooks S, Koenig F, McNab WH, Russell W, Smith ML. 1995. Ecological units of the eastern United States—first approximation. Atlanta (GA): USDA Forest Service, 81 p.

Kline VM, Cottom G. 1979. Vegetation response to climate and fire in the Driftless Area of Wisconsin. Ecology 60:861–8.

Landres PB, Morgan P, Swanson FJ. 1999. Overview of the use of natural variability concepts in managing ecological systems. Ecol Appl 9:1179–88.

Manies KL, Mladenoff DJ, Nordheim EV. 2001. Assessing large-scale surveyor variability in the historic forest data of the original US public land survey. Can J For Res 31:1719–30.

Manies KL, Mladenoff DJ. 2000. Testing methods to produce landscape-scale presettlement vegetation maps from the U.S. public land survey records. Landscape Ecol 15:741–54.

Mladenoff DJ, Pastor J. 1993. Sustainable forest ecosystems in the northern hardwood and conifer forest region: concepts and management. In: Aplet GH, Johnson N, Olson JT, Sample VA, editors. Defining sustainable forestry. Washington (DC): Island Press. p 145–79.

Mladenoff DJ, Sickley TA, Haight RG, Wydeven AP. 1995. A regional landscape analysis and prediction of favorable gray wolf habitat in the northern Great Lakes region. Conserv Biol 9:279–94.

Moore MM, Covington WW, Fule PA. 1999. Reference conditions and ecological restoration: a southwestern ponderosa pine perspective. Ecol Appl 9:1266–77.

Onsrud HJ. 1979. A manual for resurvey of public land survey corners and sectionalized subdivision with the state of Wisconsin [dissertation]. Madison (WI): University of Wisconsin–Madison.

Ott, L. 1988. An introduction to statistical methods and data analysis. Boston: PWS-Kent.

Parsons DJ, Swetnam TW, Christensen NL. 1999. Use and limitations of historical variability concepts in managing ecosystems. Ecol Appl 9:1177–8.

Press SJ, Wilson S. 1978. Choosing between logistic regression and discriminant analysis. J Am Stat Assoc 73:699–705.

Radeloff VC, Mladenoff DJ, Boyce MS. 2000. A historical perspective and future outlook on landscape scale restoration in the northwest Wisconsin Pine barrens. Restor Ecol 8:119–26.

Radeloff VC, Mladenoff DJ, He HS, Boyce MS. 1999. Forest landscape change: the northwest Wisconsin Pine Barrens before European settlement and today. Can J For Res 29:1649–59.

Russell EWB. 1997. People and the land through time: linking ecology and history. New Haven, (CT): Yale University Press.

Russell EWB. 1981. Vegetation of northern New Jersey before human settlement. Am Midl Nat 105:1–12.

Schulte LA, Mladenoff DJ. 2001. The original US Public Land Survey records: their use and limitations in reconstructions of pre-European settlement vegetation. J For 99:5–10.

Schulte LA, Mladenoff DJ, Nordheim EV. 2002. Quantitative classification of a historic northern Wisconsin (USA) landscape: mapping regional forest types and their spatial uncertainty. Can J For Res XX:XXX–XX.

Stewart LO. 1935. Public land surveys—history, instructions, methods. Ames (IA): Collegiate Press. 202 p.

Swetnam TW, Allen CD, Betancourt JL. 1999. Applied historical ecology: using the past to manage for the future. Ecol Appl 9:1189–206.

Teck RM, Hilt DE. 1990. NE-TWIGS: an individual-tree probability of survival model. Durham (NH): USDA Forest Service Northeastern Forest Experiment Station. 10 p.

Vale TR. 1998. The myth of the humanized landscape: an example from Yosemite National Park. Nat Areas J 18:231–6.

Veatch JO. 1928. Reconstruction of forest cover based on soil maps. Q Bull Mich Agri Exp Stat 10:116–26.

White MA, Mladenoff DJ. 1994. Old-growth forest landscape transitions from pre-European settlement to present. Landscape Ecol 9:191–205.

Whitney GG. 1994. From coastal wilderness to fruited plain: a history of environmental change in temperate North America from 1500 to the present. Cambridge (England): Cambridge University Press. 451 p.

[WiDNR] Wisconson Department of Natural Resources. 1999. Ecological units of Wisconsin - first approximation, including Land Type Associations (LTAs). Madison (WI): Wisconsin Department of Natural Resources.