Plant Ecol (2011) 212:207–217
DOI 10.1007/s11258-010-9815-9

Patterns from the past: modeling Public Land Survey witness tree distributions with weights-of-evidence

Todd D. Fagin · Bruce W. Hoagland

Received: 18 March 2010 / Accepted: 10 July 2010 / Published online: 15 August 2010

Abstract The Public Land Survey (PLS) witness tree data provide one of the few quantitative data sets of pre-and early-European settlement composition and structure of the forests and woodlands in the western United States. However, quantifying the areal extent of individual woody species from PLS records has proven difficult due to the coarse sampling structure of the data. Several attempts have been made to convert the discrete PLS witness tree data into continuous distributions through the use of various interpolation techniques. While these methods may adequately represent the spatial patterns of individual species over large areas, they fail to consider the numerous environmental covariates that can influence the distribution of individual tree species at finer scales. A more statistically rigorous method calls for combining species–environment relationships to estimate the areal extent of individual species from point data. In this study, we utilize weights-of-evidence (WofE), a discrete multivariate method, to estimate the probable historical distribution of six important woody plant taxa of the cross timbers of south-central Oklahoma. We successfully created posterior probability distribution maps for *Quercus stellata*, *Q. marilandica*, *Q. velutina*, *Carya texana*, *C. illinoinensis*, and *Juniperus* spp. Each posterior probability map was classified into four predictive categories, thereby enabling better estimations of the historical distribution of individual taxon from coarse-resolution PLS data. Model validation indicated that the WofE method effectively estimated the posterior probabilities of all taxa under consideration.

Electronic supplementary material The online version of this article (doi:10.1007/s11258-010-9815-9) contains supplementary material, which is available to authorized users.

T. D. Fagin (✉)
Department of Geography, University of Oklahoma,
100 E. Boyd SEC 684, Norman, OK 73019, USA
e-mail: tfagin@ou.edu

B. W. Hoagland
Oklahoma Biological Survey, Department of Geography,
University of Oklahoma, 111 E. Chesapeake St,
Norman, OK 73019, USA

T. D. Fagin
Department of Geography, Oklahoma State University,
360 Murray Hall, Stillwater, OK 74078, USA

Keywords Public Land Survey · Witness trees · Presettlement forest · Weights-of-evidence · Cross timbers · Oklahoma

Introduction

The structure and composition of North American forests at the time of European settlement have received considerable attention in recent years (e.g., Wang 2005; DeWeese et al. 2007). Since past disturbance regimes have been shown to effect the current composition of an ecosystem (Dupouey et al.

2002), these historical vegetation reconstructions typically serve as baselines from which subsequent changes in ecosystems can be evaluated (Bahre 1991; Fralish et al. 1991); provide insight into the contemporary composition of landscapes (Dupouey et al. 2002); and are valuable tools in restoration ecology (Radeloff et al. 1999). A number of resources are available to researchers interested in historical vegetation reconstructions, among them the records of the Public Land Survey System (PLS) (Fagin and Hoagland 2002; Wang 2007).

Public Land Survey data provide one of the few quantitative data sets of pre-and early-European settlement vegetation for the western United States (Whitney and DeCant 2001). As surveyors partitioned the land into 93.24 km^2 (36 $mile^2$) townships and further subdivided each township into 2.59 km^2 (1 $mile^2$) sections, they created township plats on which they mapped land cover types and locations of prominent physical and man-made features (Hutchinson 1988). Surveyors also recorded quantitative information related to the so-called witness trees encountered along the survey lines: at the intersection of section lines and at each quarter section point (0.8 km along a section line), surveys noted the nearest tree in each of the adjoining sections, recording its identification and diameter at breast height (DBH), as well as the compass direction and distance from the corner or quarter section point.

Public Land Survey records have been used to evaluate vegetation dynamics (Bahre 1991; DeWeese et al. 2007), composition and structure of historical forest and woodland communities (Anderson and Anderson 1975), species–environment interactions (Cowell 1995; Wang 2007), and distribution and abundance of individual species (Abrams 2001; Wang and Larsen 2006). As per the latter, quantifying the areal extent of select woody species from PLS records has proven difficult due to the coarse sampling structure—tree data were only collected along section lines at 0.8 km (0.5 mile) intervals. In addition, bias in tree selection has been demonstrated, with tree size, longevity, and/or economic value often influencing witness tree selection (Bourdo 1956). As a result of these biases, insufficient data often exist, which makes it difficult to estimate the areal extent of select species.

Nonetheless, several attempts have been made to convert discrete PLS point data into continuous data using kriging and other interpolation methods (e.g., Batek et al. 1999; Wang and Larsen 2006; Wang 2007). While these methods may adequately represent the spatial patterns of individual species over large areas (Wang and Larsen 2006), these methods typically fail to consider the numerous covariates, such as edaphic conditions or topographic position, which can influence the distribution of individual species at finer scales. Instead, these models treat witness tree data as numeric values (typically 1 for present, 0 for absent) that can be interpolated without consideration of underlying ecological processes (He et al. 2007).

A more statistically rigorous method calls for combining species–environment relationships to estimate the areal extent of individual species from point data (He et al. 2007). One such method that shows potential is weights-of-evidence (WofE). WofE is a discrete, data-driven multivariate method originally developed for the purpose of medical diagnosis (Bonham-Carter et al. 1989), and later adapted for spatial predictions (Agterberg et al. 1993). WofE uses a log-linear form of Bayes' rule to measure the spatial association between maps of independent variables and dependent variable point data (Bonham-Carter et al. 1989; Bonham-Carter and Agterberg 1999).

The objective of this study is, therefore, to test the efficacy of WofE modeling in the estimation of the potential pre- and early-European distribution of select woody plant taxa from discrete PLS witness tree data. Specifically, we analyzed recorded occurrences of six important woody plant taxa (*Quercus stellata*, *Q. marilandica*, *Q. velutina*, *Carya texana*, *C. illinoinensis*, and *Juniperus* spp.) with six environmental covariates (soils, geological substrate, elevation, slope, aspect, and historical land cover) to calculate the posterior probability of their historical occurrence in the Arbuckle Mountains, Oklahoma. These estimates can then be used as a baseline from which subsequent changes in woody plant distributions can be gauged and to ascertain whether past land use practices and other anthropogenic disturbance regimes have influenced the distribution of individual taxon.

Materials and methods

Study area

The Arbuckle Mountains in south-central Oklahoma are a spatially heterogeneous region covering an area of approximately 215,000 ha (Fig. 1). The Arbuckle Mountains are a topographically low plateau, rising a few hundred meters above the surrounding prairie, sloping from an elevation of 411 m (1,350 ft) in the west to 229 m (750 ft) in the east (Dale 1956). Structurally, the Arbuckle Mountains consist of extensive faulting and folding which has exposed late Cambrian to middle Mississippian limestone and late Mississippian and Pennsylvanian sedimentary rocks (Suneson 1997). The surface geology is characterized mostly by outcrops of carbonate rocks (Ham 1969), though one also finds granitic outcrops surrounded by limestones, conglomerates, sandstones, shales, cherts, and other types of rocks (Dale 1956; Suneson 1997).

The Arbuckles lay within a region of vegetation known as the cross timbers, a mosaic of forest, woodland, and prairie vegetation types (Hoagland et al. 1999). The woodland communities of the Arbuckle Mountains vary considerably with soil type and moisture availability, with *Q. stellata* and *Q. marilandica* as the most important species on dry, upland soils. *C. texana* and *Q. buckleyi* are important secondary species in mesic to xeric upland sites, respectively. Important bottomland species include *Q. muehlenbergii*, *Celtis laevigata* var. *laevigata*, *C. laevigata* var. *reticulata*, *Platanus occidentalis*, *Ulmus americana*, *U. rubra*, *Carya illinoensis, Juglans nigra*, *Salix nigra*, and *Populus deltoides* (Rice and Penfound 1959; Hoagland and Johnson 2001). Ecologically, the cross timbers reside at the periphery of the eastern deciduous forest and are a transition zone into the mid-latitude grasslands. In addition, the cross timbers represent both the western and eastern limits of the ranges of a number of woodland and prairie taxa, respectively (Hoagland et al. 1999).

Data sources

Weights-of-evidence modeling proceeds in several phases: development of a spatial database, extracting predictive evidence for the phenomena under investigation, calculating weights for each predictive map (evidential layer), combining the weights from each evidential layer to predict occurrence potential, and model evaluation (Kemp et al. 1999). The spatial database includes the identification of sites (each represented by a single *x,y* coordinate pair) in which the spatial phenomenon under investigation is known to have occurred (the dependent variable). In this study, the points are historical woody plant occurrences. The series of independent variables used for the prediction of other occurrences of the phenomena under investigation is also defined. In WofE modeling, the predictor variables typically take the form of GIS layers consisting of two or more classes (Bonham-Carter and Agterberg 1999).

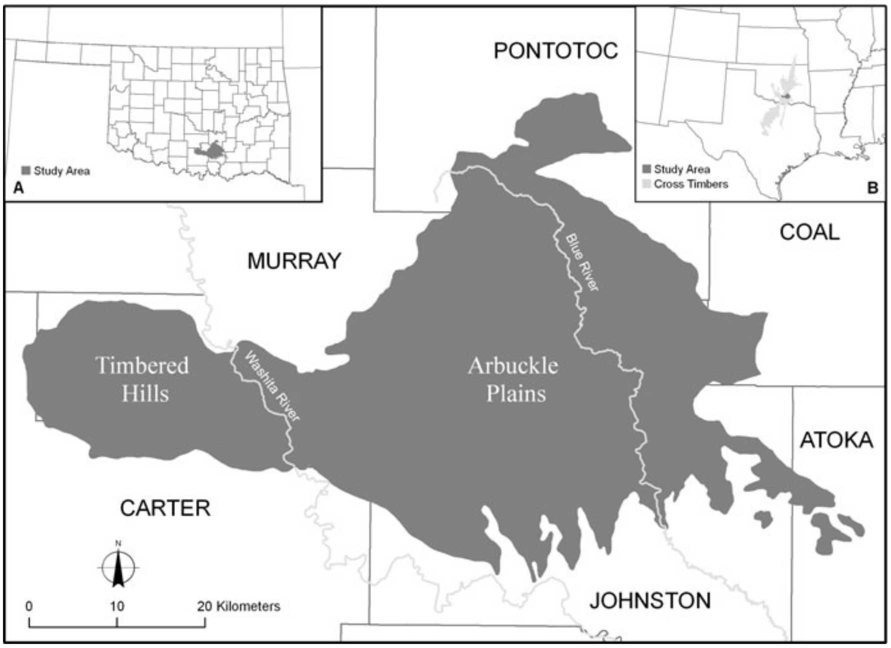

Fig. 1 Study area location: (*A*) The Arbuckle Mountains in relationship to the state of Oklahoma, USA, and (*B*) the cross timbers ecoregion in south-central United States

For the dependent variable in each of our models, we used PLS witness tree occurrence data for select taxa. The General Land Office (GLO) conducted two separate, complete surveys in the study area. The first lasted from 1870 to 1872, and the second from 1897 to 1898 (for a discussion of the two separate surveys conducted within the study area, see Hoagland 2006). We plotted historical occurrences of individual witness trees based on textual descriptions in the GLO field notes (bearing and distance from known corner and quarter section points). Preliminary analysis of these data indicated that *Q. stellata*, *Q. marilandica*, *Q. velutina*, *C. texana*, and *C. illinoinensis* were among the most important woody taxa in the study area (Table 1). In addition, during the past century, *Juniperus* spp. have increased in abundance and dominance throughout Oklahoma, primarily due to fire suppression and other land use practices (Rice and Penfound 1959; Engle et al. 1997). Knowledge of the historical distribution of this subsequently important taxon may have utility in woody plant encroachment studies. We, therefore, used the witness tree records of these six taxa as the occurrence data in our WofE models.

A number of previous studies have utilized PLS data to analyze species–environmental relationships. Common covariates identified in these studies include edaphic factors (e.g., Veatch 1925; Wang 2007; He et al. 2007), topographic position (e.g., Whitney and Steiger 1985; Batek et al. 1999; Dyer 2001); and parent material (e.g., Whitney and Steiger 1985; Sears 1925; Batek et al. 1999). We identified six environmental layers to use as our predictor variables. Three criteria went into the selection of the independent variables: previous PLS literature on species–environmental relationships; factors known to influence the distribution of the selected taxa within the study area and data availability at both the spatial and temporal scale under investigation. Data selected included those features believed to adequately represent the spatial heterogeneity of the study area, while maintaining relative consistency from the time of surveys and the time these data were acquired. The covariates selected were substrate (parent material), soil type, elevation, slope, aspect, and historical land cover. Slope and aspect were combined into a single composite "moisture availability index" layer after Batek et al. 1999 and served as a proxy for microclimate. Table 2 lists the covariates used, the sources of each, and the processing steps to prepare each for WofE modeling.

Calculating weights

The WofE method is based on a log-linear form of Bayes' Theorem (Bonham-Carter and Agterberg 1999) and involves the following steps (Bonham-Carter et al. 1989): (1) estimation of the prior probability ($P\{D\}$) of the occurrence under investigation; (2) calculation of positive (W^+) and negative (W^-) weights for each evidential layer class; (3) calculation of the contrast (C), i.e., the difference between W^+ and W^-, and studentized contrast (C_s); (4) generalization of multiclass evidential layers to several classes based on C_s values (see Romero-Calcerrada and Luque 2006 for generalization thresholds used); (5) calculation of the posterior probability (P_k) and total confidence ($P_k/\sigma_{\text{Total}}$) for each unique overlap condition of combinations of evidential layers; and (6) test of conditional independence (Agterberg and Cheng 2002). We used the ArcSDM extension (Sawatzky et al. 2009) for ArcGIS 9.3 (ESRI 2008) for all WofE calculations. For further discussion of each weight calculation, see Bonham-Carter (1994) and the online supplemental material.

Predictive map

A final predictive map representing probable habitat for each taxon is created by classifying the output into four predictive categories based on the magnitude of the ratio of P_k to $P\{D\}$ and total confidence (Romero-Calcerrada and Luque 2006):

(1) High probability: $P_k/P\{D\} > 5$ and $P_k/\sigma_{\text{Total}} > 1.5$
(2) Moderate probability: $5 > P_k/P\{D\} > 1$ and $P_k/\sigma_{\text{Total}} > 1.5$
(3) Low probability: $1 > P_k/P\{D\}$ and $P_k/\sigma_{\text{Total}} > 1.5$
(4) High uncertainity: $P_k/\sigma_{\text{Total}} < 1.5$

Model validation

We used the split-sample approach in which the occurrences of each taxon are divided into two randomly generated sets, a model building set and a validation set, to evaluate each of the models (Carranza and Hale 2002; Neuhäuser and Terhorst 2007). Each of the model building and validation sets is combined with the probability map to determine

Table 1 Comparison of frequency (no. of trees) and importance value (I.V.) for all the recorded taxa from PLS data, 1870s and 1890s

Taxon	1870s		1890s	
	No. of trees	I.V.	No. of trees	I.V.
Quercus stellata	1234	47.117	1242	41.005
Quercus velutina	529	17.981	73	2.479
Ulmus spp.	328	21.009	474	18.540
Carya texana	118	2.779	69	1.441
Quercus alba	81	2.422	57	1.326
Carya illinoinensis	56	1.702	123	4.793
Fraxinus spp.	45	1.364	69	1.820
Quercus falcata	37	1.021	184	5.944
Celtis laevigata	24	0.673	58	1.603
Juglans nigra	23	0.693	42	1.822
Quercus palustris	22	0.531	27	0.677
Populus deltoides	19	0.740	6	0.201
Quercus macrocarpa	18	0.678	19	0.602
Quercus marilandica	6	0.309	315	11.556
Platanus occidentalis	6	0.138	6	0.244
Diospyros virginiana	5	0.228	17	0.479
Juniperus spp.	4	0.105	7	0.219
Cercis canadensis	4	0.079	–	–
Morus rubra	3	0.072	–	–
Quercus spp.	3	0.065	164	4.552
Madura pomifera	3	0.067	11	0.360
Sideroxylon lanuginosum	2	0.050	7	0.129
Prunus spp.	2	0.041	–	–
Acernegundo	2	0.051	2	0.045
Malus ioensis	1	0.020	–	–
Gymnocladus dioicus	1	0.022	1	0.019
Salix spp.	1	0.020	1	0.017
Crataegus spp.	1	0.024	–	–
Quercus nigra	–	–	5	0.107
Sapindus saponaria	–	–	1	0.018

The GLO conducted two separate surveys in the study area (see Hoagland 2006). Selection of dependent variables was based primarily on historical importance value or increases in importance since historic times. With the exception of *Juniperus* spp., taxon identified only to the generic level were excluded from analysis. Due to possible taxonomic uncertainly in the 1870s data related to *Quercus marilandica* and *Q. velutina*, 1890s data were used for these two taxa (after Fagin 2009)

the overall predictivity of the model. However, in cases with a small number of occurrences, such an approach is impractical because each set would be too small of generate robust results (Carranza 2004). An independent set of validation data is, therefore, necessary. However, since we are working with historical data, no other independent data set was available. Instead, in those cases, we did not split the model into two sets and based the model performance on overall predictivity of the model building set.

We then assessed each model using area-adjusted frequency (AAF) and the continuous Boyce index (Hirzel et al. 2006). The AAF is the predicted-to-expected ratio of evaluation points for each output class, where predicted frequency is the number of points within a class divided by the total number of points, and the expected frequency is the relative area (area of class/area of total study area) of each class (Boyce et al. 2002; Hirzel et al. 2006).

Model runs

We ran six models, one for each taxon under investigation. Owing to variability in data availability and/or quality for each taxon, parameters for each model varied. For *Q. stellata*, *C. texana*, and

Table 2 Data sources and the processing steps of the covariates used in the six Wofe models

Covariate	Source data	Processing steps
Surficial geology	1:250,000 vector layer (Cederstrand 1996)	Converted from vector to raster
Soil association	1:250,000 STATSGO vector layer (NRCS 2007)	Converted from vector to raster
Elevation	1 arcsec raster layer (USGS 2008)	Reclassified into 80 ft (~25 m) classes
Moisture availability index	1 arcsec raster layer (USGS 2008)	Combined slope and aspect layer after Batek et al. (1999)
Land cover	Scanned and digitized PLS township plats (Fagin 2009)	Converted from vector to raster

Substrate data were extracted from a preexisting 1:250,000 scale digital data set of surficial geology (Cederstrand 1996). General soil association data were obtained from the 1:250,000 U.S. General Soil Map (STATSG02) Database (USDA NRCS 2007). The terrain data (elevation, slope, and aspect) were derived from the National Elevation Dataset (NED) 1 arcsec (approximately 30 m) digital elevation model (USGS 2008). Elevation data were reclassified into 80 ft (~25 m) elevation classes, while slope and aspect were combined into a single composite layer after Batek et al. (1999) to create a moisture availability index layer. Land cover data were obtained from a map consisting of digitized PLS plats (Fagin 2009)

C. illinoinensis, we used PLS witness tree data from the 1870s surveys. However, there was a limited number of *Q. marilandica* occurrences in the 1870s survey (see Table 1) and we, therefore, used the 1890s PLS occurrence data. In addition, *Q. velutina* occurrence data from the 1870s are higher than subsequent surveys of the region, but consistent with data from the 1890s (e.g., Dale 1956; Rice and Penfound 1959). Thus, we used the 1890s PLS point data for *Q. velutina*. Finally, despite the dramatic increase in abundance during the last century, the *Juniperus* spp. records from both the 1870s and 1890s were too small to create an effective model, and so it was necessary to combine the 1870s and 1890s *Juniperus* spp. occurrence data into a single data set. All the six models used the same evidential layers.

Results

Calculated weights

A total of 619 occurrence points representing six different taxa were combined with the evidential layers to calculate weights and produce six posterior probability maps of occurrence, one for each taxon under investigation. Tables showing calculated weights (W^+ and W^-), the contrast (C), and studentized contrast (C_s) for each taxon are available in the online supplement material.

The contrast (C) represents a measure of spatial association between occurrences and classes of an evidential layer, while C_s provides a measure of confidence (Bonham-Carter 1994). A C_s value greater than 1.96 indicates that the hypothesis that $C = 0$ can be rejected at $\alpha = 0.05$ (Bonham-Carter et al. 1989). The calculated weights thus serve as a valuable indicator of those environmental factors that show the greatest spatial association to each layer, as well as form the basis for layer generalization and, ultimately, the posterior probability computations.

Predictive maps

Each posterior probability map was classified into four classes (high probability, moderate probability, low probability, and high uncertainty) based on $P_k/P\{D\}$ and $P_k/\sigma_{\text{Total}}$ values, thereby showing probable historical distributions of each taxon (Figure 2). Figure 3 shows the area occupied by each predictive class, illustrating potential habitat for the taxa under consideration. Areal values were also used to validate each of the models using AAF. Moreover, computed conditional independence (CI) for each predictive map (Agterberg and Cheng 2002) indicated that all the CI values were within acceptable ranges.

Model validation

In addition, 1,188 points representing the six different taxa were used to validate the models (Table 3). A model in which AAF increases as the suitability class increases is deemed a good model (Hirzel et al. 2006). A low predictive class should contain fewer

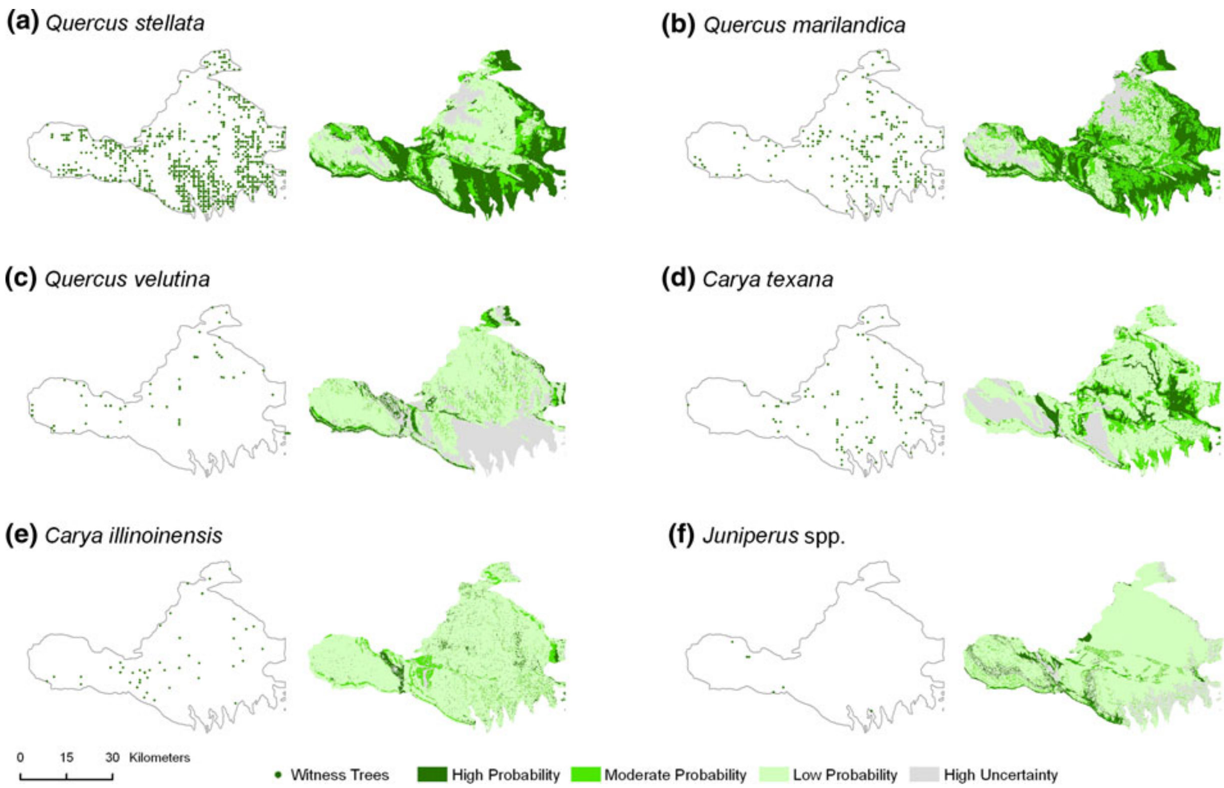

Fig. 2 Discrete PLS point distributions and continuous probability surfaces for the six taxa based on Wofe model results. Probability classes based on $P_k/P\{D\}$and $P_k/\sigma_{\mathrm{Total}}$

predicted than expected points (AAF <1), while each successive probability class should have AAF values increasingly higher than 1. Based on computed AAF, the models accurately predicted the distributions of all the taxa under consideration (Figure 4).

Each of our models demonstrated a monotonic increase with each successive predictive class, with Spearman's ρ values (plot of AAF value against mean posterior probability value, i.e., the Boyce index (Boyce et al. 2002)) ranging from 0.8 to 1 for both the model building and validation sets. A positive Boyce index value indicates that the predictions are consistent with the known occurrences (Hirzel et al. 2006). However, because the Boyce index is sensitive to the number of classes (Boyce et al. 2002), we also calculated the continuous Boyce index, which uses a moving window rather than fixed classes. All the continuous Boyce index values were positive, ranging from 0.439 to 1, demonstrating that WofE effectively predicted the posterior probability of occurrence of each taxon under investigation.

Discussion and conclusions

Weights-of-evidence belongs to a growing body of research techniques that can be used to predict species distribution from point occurrence data (see Guisan and Zimmermann 2000; Elith et al. 2006 for reviews of similar methods). WofE has been used successfully by geoscientists (e.g., Bonham-Carter et al. 1988; Porwal et al. 2001), archeologists (e.g., Diggs and Brunswig 2006), geomorphologists (e.g., Neuhäuser and Terhorst 2007; Bui et al. 2008), hydrologists (e.g., Arthur et al. 2007; Masetti et al. 2007), and ecologists (Romero-Calcerrada and Luque 2006; MacNally 2007). Our results indicate that WofE can also be used to create probabilistic maps of the historic distribution of woody plant taxa from discrete PLS data. As such, WofE may serve as a valuable tool for restoration and historical plant ecology studies.

For instance, quantitative studies of the historical vegetation of the cross timbers are limited (e.g., Shutler

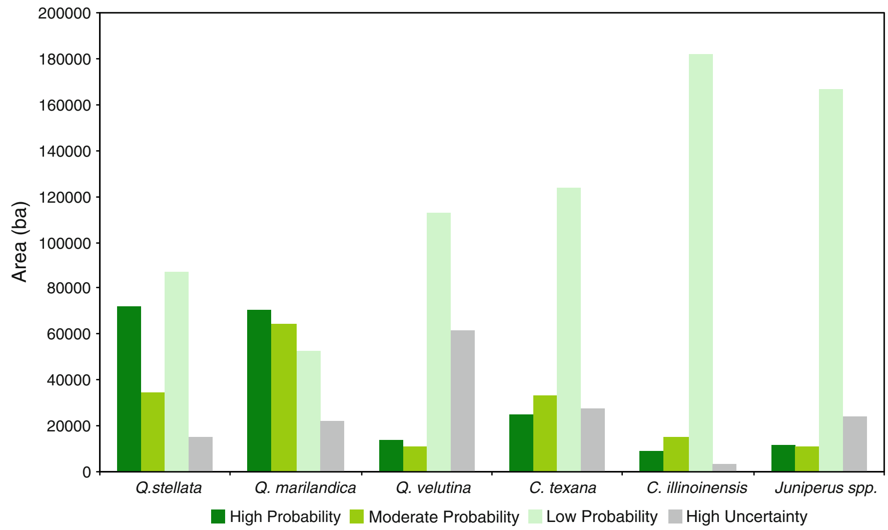

Fig. 3 Estimated area of each probabilistic class for the six taxa. Class area and the number of individual points within each class (Table 3) were used for area-adjusted frequency computations for model validation

Table 3 Observed frequency (count and percentage) and calculated area-adjusted frequency of model building and validation points for each predictive class for each model run. The predicted frequency (observed frequency/total frequency) is divided by the relative area of each class (see Fig. 3) to compute the area-adjusted frequency (Fig. 4)

Taxon	High			Moderate			Low			Uncertain			
	Count	%	AAF	Count	%	AAF	Count	%	AAF	Count	%	AAF	
Q. stellata	176	71.84	2.08	48	19.59	1.18	19	7.76	0.186	2	0.82	0.111	Model
	696	70.37	2.04	173	17.49	1.05	108	10.92	0.262	12	1.21	0.166	Validation
Q. marilandica	99	48.29	1.43	61	29.76	0.97	41	20	0.798	4	1.95	0.186	Model
	46	41.82	1.24	37	33.64	1.09	23	20.91	0.834	4	3.64	0.347	Validation
Q. velutina	19	41.3	6.02	6	13.04	2.33	17	36.96	0.653	4	8.7	0.28	Model
	8	29.63	4.32	2	7.41	1.32	15	55.56	0.983	2	7.41	0.239	Validation
C. texana	40	52.63	12.14	17	22.37	3.13	15	19.74	0.227	4	5.26	3.43	Model
	12	28.58	6.58	13	30.95	4.32	16	38.1	0.438	1	2.38	1.55	Validation
C. illinoinensis	11	30.56	7.05	4	11.11	1.55	20	55.56	0.639	1	2.78	1.81	Model
	6	30	6.59	4	20	2.79	10	50	0.574	0	0	0	Validation
Juniperus spp.	9	81.82	14.82	1	9.09	1.77	1	9.09	0.116	0	0	0	Model
	0	0	0	0	0	0	0	0	0	0	0	0	Validation

2001; Shutler and Hoagland 2004). Nonetheless, many believe that the arborescent communities of the region were less widespread prior to Euro-American settlement (e.g., Rice and Penfound 1959; Engle et al. 2006). Evidence suggests that fire suppression and other land use practices, such as grazing, have contributed to increases in dominant overstory *Quercus* species (Engle et al. 2006). Moreover, there is sufficient evidence that, in the period since widespread Euro-American settlement, *Juniperus* spp. have encroached in former grasslands and woodlands throughout the region, resulting in the conversion of the former to woodlands and the latter to closed canopy forest (Rice and Penfound 1959; Engle et al. 1997; Hoagland and Johnson 2001).

Because these changes often proceed at rates that exceed the availability of quantitative data, estimating changes in woody plant distribution since historic times is problematic (Briggs et al. 2002). Moreover, the few quantitative historical data sets available, such as PLS data, typically have resolutions too coarse for ecological analysis (Delcourt and Delcourt

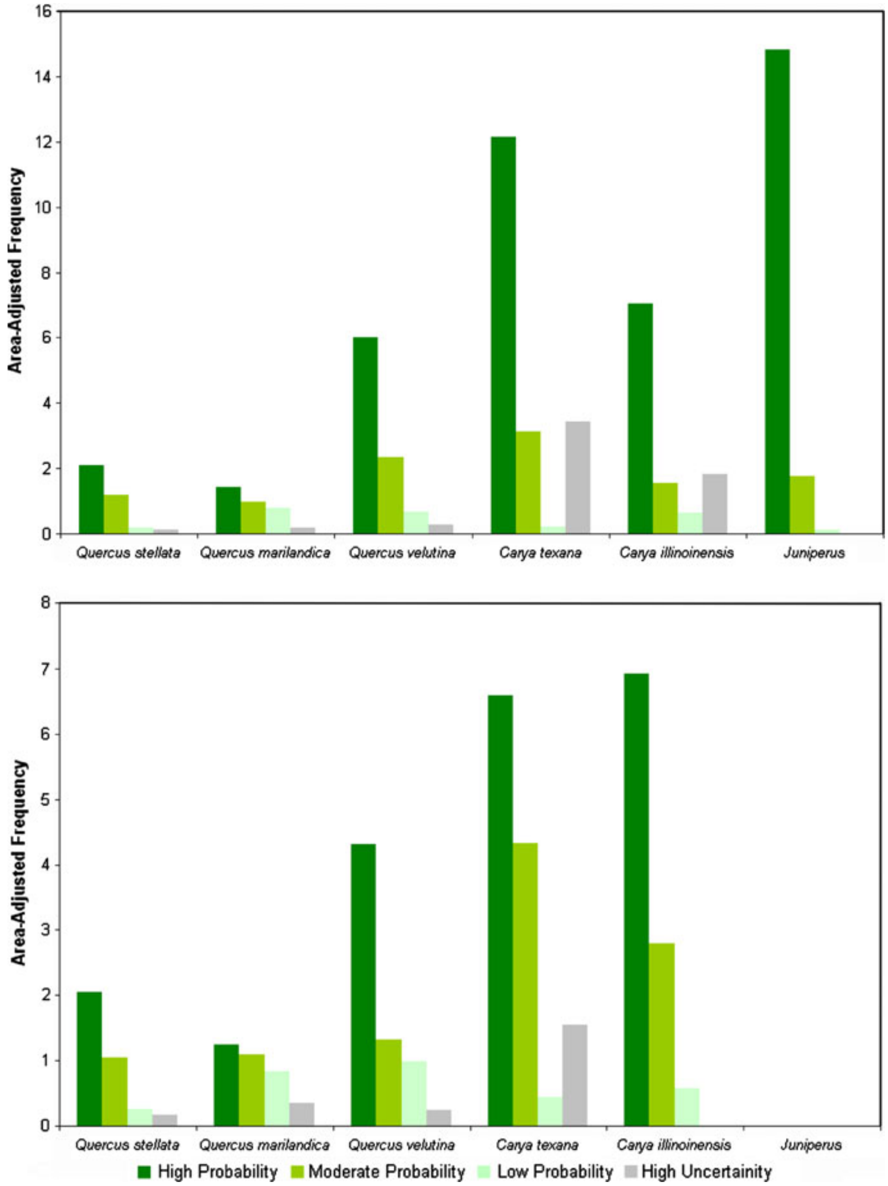

Fig. 4 Calculated area-adjusted frequency (observed frequency/expected frequency) values for both the model building and validation occurrence points

1996; Manies and Mladenoff 2000). The predictive maps generated with WofE analysis, though, can be used to estimate the probable historical distributions of individual woody species. In the instance of this study, statistically validated probable historical distribution of six important western cross timber taxa have been produced and can be further used as baselines from which to compare subsequent distributions of these taxa.

The calculated association between a taxon and the environmental covariates served as the basis for the WofE models in our study. A rich body of PLS research, dating from Veatch (1925) and Sears (1925) to the present (e.g., Wang 2007), has utilized PLS data to analyze species–environment relationships. An underlying assumption of these analyses is that the PLS data portray distributions prior to widespread human disturbance (Fagin and Hoagland 2002). Historic species–environment relationships, therefore, often serve as the basis of restoration efforts (Whitney and DeCant 2001). In WofE, the calculated weights represent a measure of the spatial association between a taxon and environmental variables (Kemp et al. 1999). Because an individual taxon's distribution is often influenced by a several environmental factors, such as edaphic conditions, topographic position, and parent material, the combined weights represented by the calculated posterior probability provide a statistical basis for identifying the historic site requirements of an individual taxon.

The results of this study should, therefore, be placed within the broader context of historical species–environment relationships. In the instance of the western cross timbers, the output of our models can be used to better quantify the degree of woody plant encroachment and distributional changes of important woody taxa since Euro-American settlement and serve as a tool for guiding restoration along this important prairie-forest ecotone.

References

Abrams MD (2001) Eastern white pine versatility in the presettlement forest. BioScience 51:967–979

Agterberg FP, Cheng Q (2002) Conditional independence test for weights-of-evidence modeling. Nat Resour Res 11: 249–255

Agterberg FP, Bonham-Carter GF, Cheng Q, Wright DF (1993) Weights of evidence and weighted logistic regression for mineral potential mapping. In: Davis JC, Herzfeld UC (eds) Computers in geology, 25 years of progress. Oxford University Press, Oxford, pp 13–32

Anderson RC, Anderson MR (1975) The presettlement vegetation of Williamson County, Illinois. Castanea 40: 345–363

Arthur JD, Wood HAR, Baker AE, Cichon JR, Raines GL (2007) Development of a Bayesian-based aquifer vulnerability assessment in Florida. Nat Resour Res 16: 93–107

Bahre CJ (1991) A legacy of change: historic human impact on vegetation in the Arizona borderlands. The University of Arizona Press, Tuscon

Batek MJ, Rebertus AJ, Schroeder WA, Haithcoat TL, Compas E, Guyette RP (1999) Reconstruction of early nineteenth-century vegetation and fire regimes in the Missouri Ozarks. J Biogeogr 26:397–412

Bonham-Carter GF (1994) Geographic information systems for geoscientists: Modelling with GIS. Pergamon, Oxford

Bonham-Carter GF, Agterberg FP (1999) Arc-WofE: a GIS tool for statistical integration of mineral exploration datasets. Bull Int Stat Inst 52 Session, Helsinki, Finland

Bonham-Carter GF, Agterberg FP, Wright DF (1988) Integration of geological datasets for gold exploration in Nova Scotia. Photogramm Eng Remote Sens 54:1585–1592

Bonham-Carter GF, Agterberg FP, Wright DF (1989) Weights of evidence modelling: a new approach to mapping mineral potential. In: Agterberg FP, Bonham-Carter GF (eds) Statistical applications in the geoscience. Geological Survey of Canada Paper 89-9. Geological Survey of Canada, Ottawa, pp 171–183

Bourdo EA (1956) A review of the General Land Office Survey and of its use in quantitative studies of former forests. Ecology 37:754–768

Boyce MS, Vernier PR, Nielsen SE, Schmiegelow FKA (2002) Evaluating resource selection functions. Ecol Model 157:281–300

Briggs JM, Hoch GA, Johnson LC (2002) Assessing the rate, mechanisms, and consequences of the conversion of tallgrass prairie to *Juniperus virginiana* forest. Ecosyst 5:578–586

Bui HB, Nguyen QP, Nguyen VT (2008) GIS-based weights-of-evidence modeling for landslide susceptibility mapping at Jaechon area, Korea. In: International symposium on geoinformatics for spatial infrastructure development in earth and allied sciences, Hanoi, Vietnam. http://wgrass.media.osaka-cu.ac.jp/gisideas08/viewpaper.php?id=253. Accessed 14 March 2009

Carranza EJM (2004) Weights of evidence modeling of mineral potential: a case study using small numbers of prospects, Abra, Philippines. Nat Resour Res 13:173–187

Carranza EJM, Hale M (2002) Where are porphyry copper deposits spatially localized? A case study in Benguet Province, Philippines. Nat Resour Res 11:45–59

Cederstrand, JR(1996) Digital geologic map of Ardmore and Sherman quadrangles, South-Central Oklahoma. U.S. Geological Survey Open-File Reports 96-370

Cowell CM (1995) Presettlement piedmont forests: patterns of composition and disturbance in central Georgia. Ann Assoc Am Geogr 85:65–83

Dale EE (1956) A preliminary survey of the flora of the Arbuckle Mountains, Oklahoma. Tex J Sci 8:41–73

Delcourt HR, Delcourt PA (1996) Presettlement landscape heterogeneity: evaluating grain resolution using general land office survey data. Landsc Ecol 11:363–381

DeWeese GG, Grissino-Mayer HD, Lam N (2007) Historical land-use/land-cover change in a bottomland hardwood forest, Bayou Fountain, Louisiana. Phys Geogr 28:345–359

Diggs, DM, Brunswig, RH (2006) Modeling Native American sacred sites in Rocky Mountain National Park. In: ESRI international conference proceedings, San Diego

Dupouey JL, Dambrine E, Laffite JD, Moares C (2002) Irreversible impact of past land use on forest soils and biodiversity. Ecology 83:2978–2984

Dyer JM (2001) Using witness trees to assess forest change in southeastern Ohio. Can J Forest Res 31:1708–1718

Elith J, Graham CH, Anderson RP, Dud′k M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM, Peterson AT, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Sobero′n J, Williams S, Wisz MS, Zimmermann NE (2006) Novel methods improve predictions of species' distributions from occurrence data. Ecography 29:129–151

Engle DM, Bidwell TG, Moseley ME (1997) Invasion of Oklahoma rangelands and forests by eastern Redcedar and Ashe juniper. Circular E-947. Oklahoma Cooperative Extension Service, Stillwater

Engle DM, Bodine TN, Stritzke JF (2006) Woody plant community in the cross timbers over two decades of brush treatments. Range Ecol Manage 59:153–162

ESRI (2008) ArcGIS: Release 9.3 [software]. Environmental Systems Research Institute, Redlands

Fagin TD (2009) In search of the forest primeval: data-driven approaches to mapping historic vegetation. Dissertation, University of Oklahoma

Fagin TD, Hoagland BW (2002) In search of the forest primeval: the use of land survey records in reconstructing past landscapes and evaluating human impact. N Am Geogr 4:1–20

Fralish JS, Crooks FB, Chambers JL (1991) Comparison of presettlement, second-growth, and old-growth forest on six site types in the Illinois Shawnee Hills. Am Midl Nat 125:294–309

Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. Ecol Model 135:147–186

Ham WE (1969) Regional geology of the Arbuckle Mountains, Oklahoma. In: Oklahoma Geological Survey guide book 17. Oklahoma Geological Survey, Norman

He HS, Dey DC, Fan X, Hooten MB, Kabrick JM, Wikle CK, Fan Z (2007) Mapping pre-European settlement vegetation at fine resolutions using a hierarchical Bayesian model and GIS. Plant Ecol 191:85–94

Hirzel AH, Le Lay G, Helfer V, Randin C, Guisan A (2006) Evaluating the ability of habitat suitability models to predict species presence. Ecol Model 199:142–152

Hoagland BW (2006) Township & range survey system. In: Goins CR, Goble D (eds) Historical atlas of Oklahoma. University of Oklahoma Press, Norman, pp 114–115

Hoagland BW, Johnson FL (2001) Vascular flora of the Chickasaw National Recreation Area, Murray County, Oklahoma. Castanea 66:383–400

Hoagland BW, Butler IH, Johnson FH, Glenn S (1999) The cross timbers. In: Anderson RC, Fralish JS, Baskin JM (eds) Savannas Barrens and rock outcrop plant communities of North America. Cambridge University Press, New York, pp 231–245

Hutchinson M (1988) A guide to understanding, interpreting, and using Public Land Survey field notes in Illinois. Nat Areas J 8:245–255

Kemp LD, Bonham-Carter GF, Raines GL (1999) WofE: Arcview extension for weights of evidence mapping. http://www.ige.unicamp.br/wofe/project.htm. Accessed 14 March 2009

MacNally R (2007) Consensus weightings of evidence for inferring breeding success in broad-scale bird studies. Austral Ecol 32:479–484

Manies KL, Mladenoff DJ (2000) Testing methods to produce landscape-scale presettlement vegetation maps from the U.S. Public Land Survey Records. Landsc Ecol 15:741–754

Masetti M, Poli S, Sterlacchini S (2007) The use of the weights-of-evidence modeling technique to estimate the vulnerability of groundwater to nitrate contamination. Nat Resour Res 16:109–119

Neuhäuser B, Terhorst B (2007) Landslide susceptibility assessment using "weights-of-evidence" applied to a study area at the Hurassic escarpment (SW-Germany). Geomorphology 86:12–24

Porwal A, Carranza EJM, Hale M (2001) Extended weights-of-evidence modelling for predictive mapping of base metal potential in Aravalli Province, western India. Explor Min Geol 10:273–287

Radeloff VC, Mladenoff DJ, He HS, Boyce MS (1999) Forest landscape change in the northwestern Wisconsin pine barrens from pre-European settlement to the present. Can J For Res 29:1649–1659

Rice EL, Penfound WT (1959) The upland forests of Oklahoma. Ecology 40:593–607

Romero-Calcerrada R, Luque S (2006) Habitat quality assessment using weights-of-evidence based GIS modelling: the case of *Picoides tridactylus* as species indicator of the biodiversity value of the Finnish forest. Ecol Model 196:62–76

Sawatzky DL, Raines GL, Bonham-Carter GF, Looney CG (2009) Spatial Data Modeller (SDM): ArcMap 9.3 geoprocessing tools for spatial data modelling using weights of evidence, logistic regression, fuzzy logic, and neural networks. http://www.ige.unicamp.br/sdm/ArcSDM93/source/ReadMe_ArcSDM2009.pdf

Sears PB (1925) The natural vegetation of Ohio I: a map of the virgin forest. Ohio J Sci 25:139–149

Shutler A (2001) Change in the distribution of forest and grasslands: a landscape-lavel analysis of Carter County, Oklahoma, 1871 and 1897. Thesis, University of Oklahoma

Shutler A, Hoagland BW (2004) Vegetation patterns in Carter County, Oklahoma. 1871. Proc Okla Acad Sci 84:19–26

Suneson NH (1997) The geology of the eastern Arbuckle Mountains in Pontotoc and Johnston Counties, Oklahoma: an introduction and field-trip guide. Oklahoma Geological Survey Open-File Report, Norman

United States Geological Survey (USGS) (2008) National elevation dataset. EROS Data Center. http://ned.usgs.gov/. Dec 2008

USDA Natural Resources Conservation Service (USDA NRCS) (2007) State Soil Geographic (STATSGO2) data base: data use information. US Department of Agriculture, Natural Resources Conservation Service, Washington

Veatch JT (1925) Soil maps as a basis for mapping original forest cover. Mich Acad Sci 15:267–273

Wang Y-C (2005) Presettlement land survey records of vegetation: geographic characteristics, quality, and modes of analysis. Progr Phys Geog 28:568–598

Wang Y-C (2007) Spatial patterns and vegetation-site relationships of the presettlement forests in western New York, USA. J Biogeogr 34:500–513

Wang Y-C, Larsen PS (2006) Do coarse resolution U.S. presettlement land survey records adequately represent the spatial pattern of individual tree species? Landscape Ecol 21:1003–1017

Whitney GG, DeCant JP (2001) Government Land Office Surveys and other early land surveys. In: Egan D, Howell EA (eds) Historical ecology handbook. Island Press, Washington, pp 147–176

Whitney GG, Steiger JR (1985) Site-factor determinants of the presettlement prairie-forest border areas of north-central Ohio. Bot Gaz 146:421–430