Oak Savannas in Western New York State, Circa 1795: Synthesizing Predictive Spatial Models and Historical Accounts to Understand Environmental and Native American Influences

Stephen J. Tulowiecki,* David Robertson,* and Chris P. S. Larsen†

*Department of Geography, SUNY Geneseo
†Department of Geography, University at Buffalo

This study models and maps oak savanna distribution ca. 1795 and determines how environmental conditions and Native American land use (NALU) shaped its distribution. Historical research has analyzed early landscape accounts to assess how Native Americans modified forests throughout the eastern United States. Predictive spatial modeling has sought to quantify anthropogenic and environmental drivers of forest conditions and to predict locations where NALU changed forest composition. Yet, studies have not rigorously synthesized these two methods. This research focused on oak savannas in western New York State (27,617 km^2). We trained models of oak savanna distribution from historic vegetation data in relation to environmental predictors and NALU proxies. We then mapped historical accounts of oak savannas and NALU at European-American arrival and compared them to model predictions. Results suggest that 2 to 17 percent (depending on modeling technique) of the study area contained oak savanna, with a favored estimate of 3 to 6 percent. Synthesis of models and accounts suggests that oak savannas were attributable to NALU and dry environmental conditions but that NALU (specifically burning) was present at most oak savanna locations. Models of oak savanna distribution that considered proximity to Native American settlement had higher predictive performance and better predicted locations of historical oak savanna accounts, including those with descriptions of Native American burning. This study suggests that former oak savannas in the study area can be largely attributed to NALU. Furthermore, this study's methodology and results contribute to a larger body of geographical literature on savanna landscapes. *Key Words: anthropogenic burning, forests, modeling, Native Americans, oak savanna.*

本研究模式化并绘製橡木稀树草原大约在公元 1795 年时的分佈，并判定环境条件与美洲原住民的土地使用（NALU）如何形塑此一分佈。历史研究已分析早期的地景描述，以评估美洲原住民如何改变横贯美国东部的森林。预测的空间模式化，业已寻求量化森条件的人类世与环境驱力，并预测 NALU 改变森林组成的地点，但尚未有研究精确地综合上述两种研究方法。本研究聚焦纽约州西部的橡木稀树草原（27,617 平方公里）。我们培养历史植栽数据中的橡木稀树草原分佈相较于环境预测指标和 NALU 邻近性的模型。我们接着绘製橡木稀树草原和 NALU 在欧裔美国人抵达时的历史数量，并将其与模型预测相互比较。研究结果显示，百分之二到十七（取决于模式化技术）的研究区域包含橡木稀树草原，而较适宜的评估是百分之三到六。综合模型与说明显示，橡木稀树草原可归因于 NALU 和乾燥的环境条件，但 NALU（特别是焚烧）则存在于大部分的橡木稀树草原地点之中。考量与美洲原住民部落的邻近性的橡木稀树草原分佈模型，具有较高的预测表现和较佳的历史橡木稀树草原纪录的地点预测，该纪录包含有关美洲原住民焚烧的描述。本研究主张，研究区域中从前的橡木稀树草原能够主要归因于 NALU。本研究的研究方法和结果，更进一步对于有关稀树草原地景的广泛地理学文献做出贡献。*关键词：人类世焚烧，森林，模式化，美洲原住民，橡木稀树草原。*

Este estudio modela y mapea la distribución de las sabanas de robledales en ca. 1795 y determina el modo como las condiciones ambientales y el uso del suelo por nativos americanos (NALU) configuraron esa distribución. En la investigación histórica se han analizado antiguas descripciones del paisaje para evaluar el modo como los nativos americanos modificaron los bosques a través de los Estados Unidos orientales. El modelado espacial predictivo ha buscado cuantificar los controles antropogénicos y ambientales de las condiciones forestales, y predecir las localidades donde el NALU cambió la composición del bosque. No obstante, los estudios no han sintetizado rigurosamente estos dos métodos. La investigación para este artículo

Annals of the American Association of Geographers, 110(1) 2020, pp. 184–204 © 2019 by American Association of Geographers
Initial submission, December 2018; revised submissions, March and April 2019; final acceptance, April 2019
Color versions of one or more figures in the article can be found online at www.tandfonline.com/raag
Published by Taylor & Francis, LLC.

se centró en las sabanas de robledales de la parte occidental del Estado de Nueva York (27.617 km^2). Los modelos de distribución de la sabana de robledales los surtimos con datos históricos de la vegetación en relación con los predictores ambientales y proxis del NALU. Luego cartografiamos los relatos históricos de las sabanas de robles y el NALU en la época de llegada de los euroamericanos, comparando todo con las predicciones del modelo. Los resultados sugieren que entre el 2 y el 17 por ciento (dependiendo de la técnica del modelado) del área de estudio estaba cubierta con sabana de robledal, con un cálculo preferido del 3 al 6 por ciento. Las síntesis de los modelos y de los relatos sugieren que las sabanas de robledal eran atribuibles al NARU y a condiciones ambientales secas, pero que el NALU (específicamente abrasador) estuvo presente en la mayoría de las localidades de sabana de robledal. Los modelos de distribución de este tipo de sabana que tomaron en cuenta la proximidad de asentamientos de nativos americanos tuvieron un desempeño predictivo más alto, y predijeron mejor las localizaciones de los relatos históricos de las sabanas de robledal, incluso las que incluían descripciones de quemas por nativos americanos. El presente trabajo sugiere que las antiguas sabanas de robledal localizadas en el área de estudio en gran medida pueden atribuirse al NALU. Por lo demás, la metodología y los resultados del estudio contribuyen al cuerpo más extenso de la literatura geográfica sobre paisajes de sabana. *Palabras clave: bosques, indígenas norteamericanos, modelado, quema antropogénica, sabana de robledal.*

Biogeographical and related research has sought to understand the spatial and temporal dimensions of past Native American land use (NALU) and its effects on forests of eastern North America prior to European-American settlement in the seventeenth to nineteenth centuries (Munoz et al. 2014). NALU practices that modified forests and promoted advantageous plant and animal resources were numerous and diverse (Smith 2011). One practice was burning, which cleared and thinned forests along travel routes (Pyne 1982), created habitats favored by game (Engelbrecht 2003), maintained open landscapes as hunting grounds (Cronon 1983; Stewart 2002), and promoted nut-producing trees like oak (*Quercus* spp.) and hickory (*Carya* spp.; Black and Abrams 2001; Black, Ruffner, and Abrams 2006; Tulowiecki and Larsen 2015). This body of research informs historical geographical debates concerning early landscapes of North America (Denevan 1992; Vale 2002), guides ecological restoration practices incorporating knowledge of past fire regimes (Ryan, Knapp, and Varner 2013), and advances understanding of forest dynamics (Matlack 2013). It has also generated debate over whether changes in forest composition and structure since European-American settlement, particularly forest mesophication and oak decline, is due to the loss of NALU (Nowacki and Abrams 2008; McEwan, Dyer, and Pederson 2011).

Eastern oak savanna (hereafter oak savanna) is a rare and ecologically important vegetation type for which biogeographical research has investigated the relative importance of past environmental conditions and NALU on its structure and composition. Spanning the midwestern United States and portions of the eastern United States and Canada (USGS Gap Analysis Program et al. 2018), oak savannas contain a mix of predominantly oak and other xeric-site tree species with a graminoid ground cover in a mosaic of open forest and grassland. Research on oak savannas dates back many decades (e.g., Gleason 1922) and has generally attributed their past existence to conditions that limit tree growth such as dry climate, poor soils, and anthropogenic burning, the relative importance of which varies geographically (Anderson, Fralish, and Baskin 1999). Fire suppression and land use conversion have contributed to oak savanna declines since European-American settlement (Nowacki and Abrams 2008). Research into factors influencing oak savanna distribution is part of a body of literature investigating environmental and anthropogenic influences on savannas, where the unique codominance of tree and grass species has presented a "conundrum" (House et al. 2003): Trees and grasses typically occupy different niches, have traits that confer competitive advantages under different conditions, and have differential responses to disturbance. This conundrum is potentially explained by complex interactions between climate, soil, and disturbance that includes fire (House et al. 2003).

Researchers have used varied methods to understand how NALU modified the structure and composition of forests and oak savannas. Environmental historians and historical geographers have used primary-source archival documents containing eyewitness accounts of Native American forest modification, including burning, to assess environmental and NALU influences on sparsely timbered landscapes. For example, Sauer (1927) studied early maps and firsthand accounts to reason that Native American burning shaped the structure of the Kentucky barrens. Rostlund (1957) used similar sources to study past prairie distribution in Alabama. Day (1953) and Stewart (2002) compiled early scientific literature and historical accounts of NALU practices to reveal connections between past NALU, environmental conditions, and forest structure and composition. Pyne (1982) and Cronon (1983) used archival sources documenting instances of Native American burning to assess the role of fire in eastern U.S. forests. This literature generally attributes sparsely timbered landscapes like oak savanna to burning alongside dry environmental conditions.

Other researchers have used quantitative methods to determine the relative importance of NALU and the spatial extent of its impacts on past forests. Black, Ruffner, and Abrams (2006) applied statistical methods with environmental data, Native American settlement data, and vegetation data from original land survey records (OLSRs) to study distributions of oak, hickory, and chestnut (*Castanea dentata*) in Pennsylvania ca. 1800. Their study revealed that proximity to Native American settlement best predicted the abundance of such taxa. In New York State, Tulowiecki and Larsen (2015) applied species distribution models trained from OLSR data, also discovering that models including proximity to former Native American settlement (termed Native American variables [NAVs]) best predicted the distribution of those taxa. Both studies concluded that NALU had an effect on forest composition within 10 to 15 km from settlements. These studies, however, focused on NALU impacts on composition rather than on forest structure. Others have used spatially explicit dynamic models of disturbance and forest succession with simulated Native American burning to assess whether fire or climatic changes caused vegetation change, both structurally (Bean and Sanderson 2008) and compositionally (Klimaszewski-Patterson et al. 2018).

The methods of historical analysis and modeling each have limitations, yet their strengths are complementary. Although providing descriptions of NALU and its effects on forests, historical accounts represent anecdotal data that might be biased (McClenachan et al. 2015) and not spatially resolvable (Wieczorek, Guo, and Hijmans 2004). Statistical and modeling approaches have quantified aspects of NALU and its effects on forests using systematically collected vegetation data (i.e., OLSRs), but NAVs used in these approaches were proxies that did not incorporate actual knowledge of the spatial extent or intensity of NALU. Using these methods in isolation, it is unclear where and how much forest NALU modified.

The use of historical analysis and modeling together to understand eastern oak savannas thus holds potential to offset their respective limitations and to corroborate findings, thereby improving knowledge of environmental and human influences on oak savannas. This synthesis would also serve as a response to long-standing calls for greater integration of human and physical geography research made from various subfields of geography, particularly in savanna and forest research. Outside of North America, biogeographers have called for greater acknowledgment of the full range of environmental and human factors influencing current savanna distributions (Duvall 2011) and for "humanizing" quantitative models of savanna distribution (Laris 2011). Elsewhere, research has pointed out the growing use of qualitative and quantitative data sets from human and physical geography to understand the legacies of past human land use on forested landscapes (Robertson, Larsen, and Tulowiecki 2018).

This study maps oak savannas observed ca. 1795 in western New York, United States, and analyzes the relative influence of NALU and environmental conditions on their distribution. This study synthesizes methods from biogeography to produce predictive spatial models of oak savanna distribution and assess their causes and from historical geography to discover historical accounts of oak savannas to corroborate and add explanatory depth to modeling results. To our knowledge, no study has synthesized these methodologies to study past eastern oak savannas.

Study Area

The study area is defined by the Holland Land Company, Morris Reserve, and Phelps and Gorham purchases (Figure 1). Bounding the study area are Lake Ontario (north), Lake Erie (west), and the New York–Pennsylvania border (south). The study

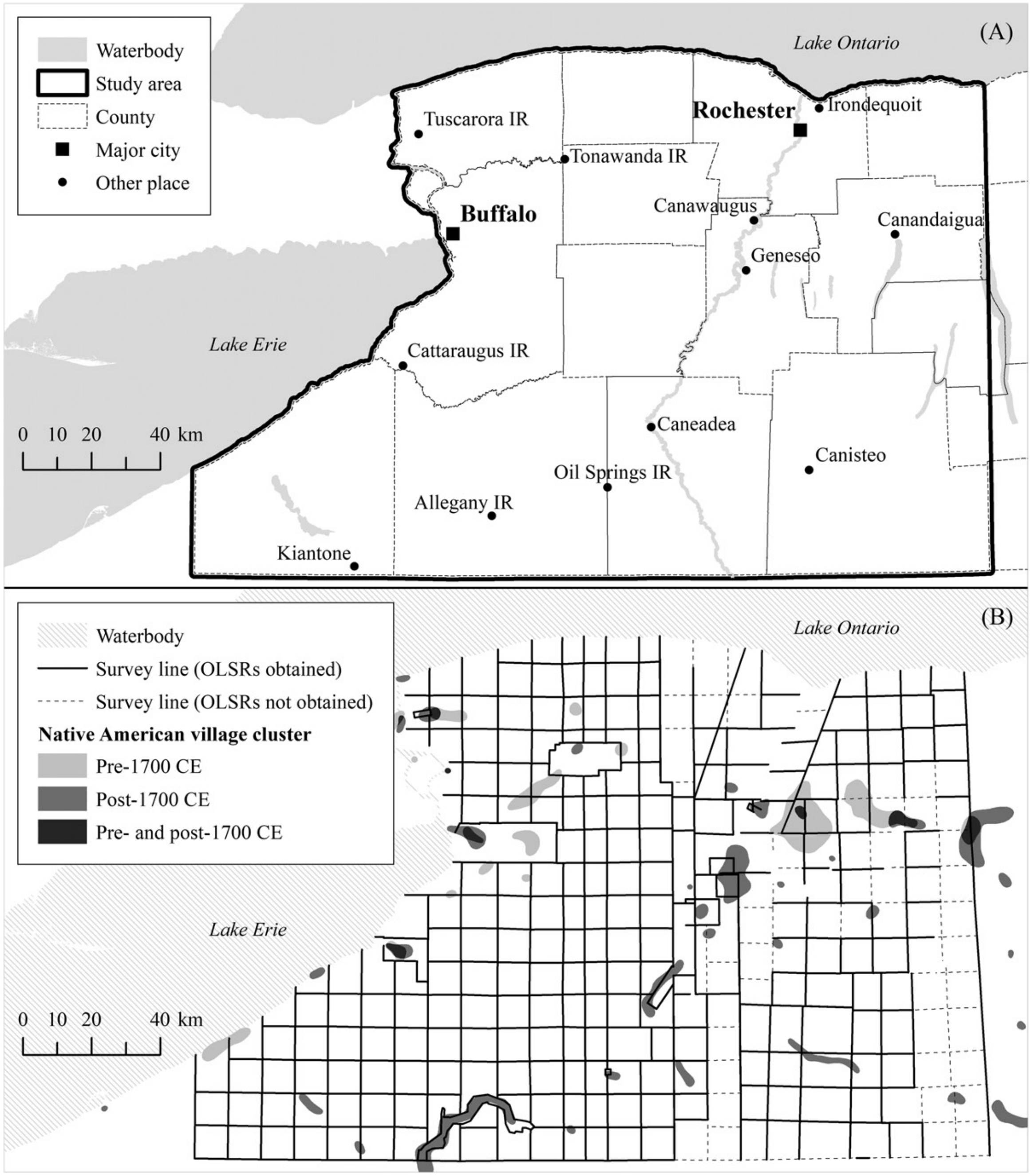

Figure 1. (A) Overview of the study area, including present-day IRs and selected places with Native American names. (B) Survey lines with OLSR data that were obtained for this study, along with Native American village clusters. IR = Indian Reservation; OLSR = original land survey record.

area spans the warmer, drier Erie–Ontario Lowlands and cooler, moister Allegheny Plateau (Fenneman 1938). Mean annual precipitation varies spatially from 79.5 to 131.2 cm and mean annual temperature ranges from 6.2 °C to 9.7 °C (PRISM Climate Group 2013). Mesic-site tree species such as beech (*Fagus grandifolia*) and sugar maple (*Acer saccharum*) dominated forests historically, with xerophytic taxa such as oak and pine (*Pinus* spp.) in drier sites and taxa such as black ash (*Fraxinus nigra*) and American elm (*Ulmus americana*; Wang 2007) in poorly drained sites.

In New York, oak savannas contain a mix of grassland and low-density woodland composed predominantly of oak species, embedded within oak-hickory forests (Edinger et al. 2014). They are associated with dry, coarse-textured, shallow soils and limestone bedrock (Shanks 1966; Edinger et al. 2014). Shanks (1966) attributed past oak savannas in the study area to soil conditions and possibly Native American burning, although he claimed that there was no evidence of widespread burning. Keister (1998) quoted historical accounts of oak savannas and Native American burning within the study area but did not map or extensively reference them. Seischab and Orwig (1991) and Seischab (1992) noted that OLSRs described grasslands, oak, and "thinly timbered" areas near Native American settlement but did not rigorously analyze environmental or anthropogenic correlates. Oak savannas are currently listed as an imperiled ecosystem globally and have five or fewer occurrences in the state (Edinger et al. 2014). Due to discrepancies between vegetation classification schemes (Anderson, Fralish, and Baskin 1999) and ambiguities in the historical record (Chapman and Brewer 2008), we use *oak savanna* to encompass sparsely timbered, oak-dominated environments also described historically as oak openings, oak barrens, or oak plains.

Native American groups inhabited the study area before European-American arrival, most notably the Seneca nation of the Haudenosaunee (Iroquois) Confederacy (Figure 1B; Snow 1996; Engelbrecht 2003). From ca. 1400 to 1700, the Seneca inhabited the eastern portion of the study area (Figure 1B) near the Finger Lakes and Genesee Valley, whereas the Neutral, Wenro, and Erie inhabited western and southwestern portions. These groups practiced traditional subsistence strategies such as horticulture within clearings, hunting in semiforested areas, and foraging of wild plants including mast (large nuts) from oak, hickory, chestnut, and black walnut (*Juglans nigra*). From ca. 1700 to 1800, European-American settlement displaced Native Americans from traditional lands in the study area. The Seneca and other groups were resettled on reservation lands in western New York (Figure 1B) or migrated to Ontario, Ohio, and Oklahoma (Morgan 1901; Snow 1996; Engelbrecht 2003).

Data and Methods

This study consisted of three main steps. First, predictive spatial models of oak savannas were developed to predict their distribution and to quantify the relative importance of environmental factors and NALU. We then retrieved, analyzed, and mapped historical accounts of oak savannas. Finally, we synthesized and compared model predictions and historical accounts. All tasks requiring geographic information systems (GIS) software were performed using ArcMap 10.5.1 (ESRI 2017).

Modeling Oak Savanna Distribution

Developing models of oak savanna distribution required four steps. Step 1 created dependent and independent variables for the models. Step 2 parameterized and trained models of oak savanna, and Step 3 created spatial predictions of oak savanna distribution. Step 4 evaluated the predictive performance and other characteristics of models, including by comparing the predictive performance of models that did versus did not include NAVs. Each step is detailed next.

Creating Variables: Oak Savanna Locations, Environmental Conditions, and NAVs. To map oak savannas, we acquired OLSRs (Figure 1B) of the 1780 s and 1790s. OLSRs are records of surveys that subdivided lands prior to sale and European-American settlement (Whitney 1996; Wang 2005). OLSRs contain lists of bearing trees marking survey corners and line descriptions recording vegetation and soil characteristics along survey lines. Studies have used OLSRs to assess relationships between vegetation and NALU (Black and Abrams 2001; Black, Ruffner, and Abrams 2006; Tulowiecki and Larsen 2015). Land companies in the study area generally surveyed townships using a 9.7×9.7 km (6×6 mile) grid, although irregularities occurred (e.g., due to Native American reservations and water bodies). Tulowiecki, Larsen, and Wang (2015) prepared Holland Land Company survey records in GIS format used in this study. Transcriptions of Morris Reserve records by Cogbill and Guderian (unpublished) from the New Jersey Historical Society were mapped specifically for this study. We also transcribed and mapped Phelps and Gorham survey records provided by the Ontario County Historical Society and the Steuben County Clerk. Additional lot-scale survey records from county clerk's offices were transcribed and mapped where township surveys were missing. All OLSR survey lines were digitized as lines in GIS software, using tax parcel data

Table 1. Predictors used in models of oak savanna distribution

Category	Predictor	Unit	Predictor code	Source
Climate	Mean precipitation, May–September	cm	climate_precip0509	PRISM Climate Group
	Mean annual temperature	°C	climate_tempann	PRISM Climate Group
Geology	Recent alluvium	% land area	geology_al	NYSM
	Colluvial diamicton	% land area	geology_cd	NYSM
	Kame deposits	% land area	geology_k	NYSM
	Kame moraine	% land area	geology_km	NYSM
	Lacustrine beach	% land area	geology_lb	NYSM
	Lacustrine sand	% land area	geology_ls	NYSM
	Lacustrine silt and clay	% land area	geology_lsc	NYSM
	Outwash sand and gravel	% land area	geology_og	NYSM
	Swamp deposits	% land area	geology_pm	NYSM
	Bedrock	% land area	geology_r	NYSM
	Till	% land area	geology_t	NYSM
	Till moraine	% land area	geology_tm	NYSM
	Undifferentiated stratified drift assemblage	% land area	geology_usd	NYSM
Native American	Distance to post-1700 Native American settlement	km	nav_dist_to_post	(various; see Table 2)
	Distance to pre-1700 Native American settlement	km	nav_dist_to_pre	(various; see Table 2)
Soil	Soil available water supply, 0 to 150 cm	cm	soil_aws150cm	NRCS
	Soil bulk density	g/cm	soil_bulkdensity	NRCS
	Compound topographic index	—	soil_cti	USGS
	Depth to soil restrictive layer	cm	soil_depthrestrictive	NRCS
	Ranked soil drainage class	—	soil_drainageclass	NRCS
	Soil erodibility factor	—	soil_kffact	NRCS
	Saturated hydraulic conductivity (soil permeability rate)	mm/hr	soil_ksat	NRCS
	Soil organic matter	% weight	soil_organicmatter	NRCS
	Soil clay	% weight	soil_percentclay	NRCS
	Soil sand	% weight	soil_percentsand	NRCS
	Degree of acidity or alkalinity	pH	soil_ph	NRCS
	Soil, passing sieve no. 10 (coarse)	% weight	soil_sieve10	NRCS
	Soil, passing sieve no. 200 (fine)	% weight	soil_sieve200	NRCS
Topography	Mean terrain slope angle	°	topography_slope	USGS

Notes: Predictor codes represent abbreviations for predictors appearing in other figures of this study. NYSM = New York State Museum; NRCS = Natural Resources Conservation Service Soil Survey Geographic Database; USGS = U.S. Geological Survey.

and historical maps to locate and trace locations of OLSRs (Tulowiecki, Larsen, and Wang 2015).

Oak savannas described in OLSRs were mapped by interpreting these descriptions as oak savanna: barrens, clear land, meadows, open areas, open woods, openings, plains, scarce timber, scattered trees, scattering timber, or thinly timbered areas where oak was listed first in timber descriptions. We designated oak savannas as present if more than 50 percent of the surveyed line passing through a grid cell (see later) recorded oak savannas and absent otherwise. Excluded from analysis were surveyed areas containing no timber descriptions, because it was unclear whether this absence was due to treeless areas or surveyor omissions.

NAVs and environmental variables were then created (Table 1). To create NAVs, we calculated distance from ca. 1500 to 1700 (pre-1700) Native American village clusters and distance from ca. 1700 to 1800 (post-1700) village clusters (Figure 1B). Village locations were mapped and cross-referenced using eight historical and archaeological sources (Table 2). We created NAVs based on distance to village clusters, rather than individual villages, to accommodate discrepancies across sources over the exact number and locations of villages. Furthermore, these two periods generally represented the finest temporal resolution into which village clusters could be classified. We additionally acquired and processed data on climate (PRISM Climate Group 2013), geology (New York State Museum 2018), soil (Natural Resources Conservation Service 2014), and topography (U.S. Geological Survey 2013) to create twenty-nine environmental predictors (Table 1). All variables were resampled and aggregated to the resolution of the coarsest resolution climate predictors ($\approx 800 \times 800$ m).

Table 2. Sources used to map Native American village clusters for developing Native American variables

Sources	ca. 1500–1700 village sites?	ca. 1700–1800 village sites?	Geographic regions of the study area covered
Cappon (1976)	No	Yes	Entire area
Grumet (1995)	Yes	Yes	Entire area
Hays and Post (1999)	No	Yes	Southeastern portion
Jennings and Fenton (1995)	Yes	Yes	Entire area
Jones (2010)	Yes	No	East/northeastern portion
Morgan (1901)	No	Yes	Entire area
Parker (1920)	Yes	Yes	Entire area
White (1978)	Yes	No	Western/northwestern portion

Developing Models of Oak Savanna Distribution. We trained models with different combinations of training data, modeling techniques, and predictors (Table 3). Given uncertainties in the meaning of oak savanna absence in OLSRs, modeling techniques using either presence-only or presence–absence data were used. Oak savanna presence–absence data were reserved to form a test data set for some models, to assess their predictive ability. Models were also trained with and without NAVs to assess whether changes in model performance and quality resulted with their inclusion. Additional details on modeling are provided in the supplemental materials.

We trained models using MaxEnt (Phillips, Anderson, and Schapire 2006), a technique utilizing presence-only data, by treating oak savanna locations from OLSRs as presences and other locations with OLSRs as environmental background data (Table 3). MaxEnt version 3.4.1 was used for model development (Phillips, Dudik, and Schapire 2017), and model tuning and variable selection were performed using the "MaxentVariableSelection" package in R (R Development Core Team 2011; Jueterbock 2018). Various measures quantified model quality (i.e., Akaike information criterion corrected for small sample sizes [AIC_c]), goodness of fit (i.e., model gain), and predictive performance in the final MaxEnt models, including the area under the receiver operating characteristic measure (AUC) applied to training and test data sets. Various measures quantified predictor variable importance (Phillips 2010), and response curves were created to describe the relationship between oak savannas and each predictor.

Models were also trained using boosted regression trees (BRT), a technique using presence–absence data, using the oak savanna presence–absence locations from OLSRs (Table 3). Functions within the "dismo" (Hijmans et al. 2017) package in R were used to train BRT models, specifically model parameterization and variable selection. The AUC measure quantified predictive performance, evaluated on training and test data sets. We calculated relative variable importance measures and generated response curves showing the relationship between oak savannas and each predictor using partial dependence plots (Elith, Leathwick, and Hastie 2008; Hijmans et al. 2017).

Creating Predictive Surfaces of Oak Savanna Distribution. We mapped oak savanna distribution by converting outputted model probabilities into binary presence–absence predictions. Various thresholds were selected in the probabilities to produce different estimates of oak savanna extent ca. 1795. We used training presences and absences to determine appropriate thresholds using the following three procedures. The "maximize kappa" threshold maximized the kappa statistic, which is a measure of agreement between two raters, in this case actual oak savanna locations in the training data and predicted oak savanna locations. The "equal sensitivity and specificity" threshold achieved equal sensitivity (true positive rate) and specificity (true negative rate). This threshold achieved binary model predictions that correctly classified presence or absence of oak savannas at the same rate. The "maximize sensitivity plus specificity" threshold achieved the highest sum of sensitivity and specificity values.

Comparing Models. We evaluated model characteristics to determine the usefulness of NAVs in predicting oak savanna distribution. Table 4 summarizes the measures and plots used and the manner in which they assessed the validity of including NAVs when modeling oak savanna distribution.

Table 3. A list of models used throughout this study

Model	Modeling technique	Training data set	Predictors considered
1	MaxEnt	100% of presence data	Environmental
2	MaxEnt	100% of presence data	Environmental, Native American
3	MaxEnt	Random 75% of presence data	Environmental
4	MaxEnt	Random 75% of presence data	Environmental, Native American
5	BRT	100% of presence–absence data	Environmental
6	BRT	100% of presence–absence data	Environmental, Native American
7	BRT	Random 75% of presence–absence data	Environmental
8	BRT	Random 75% of presence–absence data	Environmental, Native American

Notes: For Models 3, 4, 7, and 8, the remaining 25 percent of the data were reserved for later model evaluation. For models that split the data set into training and test data sets, the same training and test data sets were used for all models. BRT = boosted regression tree.

Discovering, Analyzing, and Mapping Historical Accounts of Oak Savannas

Historical accounts of oak savannas were sought for comparison with model predictions. We downloaded 117 historical documents in .txt format from the Internet Archive (1996), a free-access digital library containing user-uploaded media. We sought historical documents from the study area previously cited in scholarly literature on pre-European-American forest conditions (Day 1953; Stewart 2002). Additional historical documents were discovered using the Internet Archive search features. Publication dates of documents ranged from 1793 to 1921 with most published in the mid-nineteenth century. The most common types acquired were county histories and descriptive gazetteers (thirty-eight total): Locally written in the nineteenth and early twentieth centuries, these documents provide county-extent coverage of history and geography. The acquired county histories represented approximately 70 percent of public-domain county histories in the study area listed in a bibliography by Filby (1985); seven additional county histories acquired were unlisted. Additional documents acquired were sixteen town histories, fourteen traveler accounts, ten village or city histories, ten archaeological or anthropological texts, eight regional histories, six state-level gazetteers, six citizen journals, four development district histories, three resource inventories, and two military journals. Although portions of the study area had greater document coverage (e.g., cities, early travel routes), we obtained at least one local history document and at least one county history for each of the sixteen counties in the study area.

We searched for accounts of forest clearings, including oak savannas, in the acquired documents. A Python script modified from Tulowiecki (2018) extracted potentially relevant paragraphs with historical accounts based on the presence of keywords. The script received .txt format historical documents as input and outputted a .txt file with all paragraphs containing at least one keyword, along with the count of each keyword in each paragraph. We searched for keywords or phrases related to landscape features (e.g., "oak opening"), Native Americans (e.g., "Indian"), and fire (e.g., "burned"). Various approaches were used to filter and prioritize extracted paragraphs for inspection; for example, paragraphs explicitly referencing oak openings or those with a high number of unique keywords were inspected. Using these methods, we read approximately 1,150 paragraphs of text for accounts of forest clearings. From these paragraphs we compiled accounts of oak savannas, judged by whether they specifically described oak openings or whether they implied their presence (i.e., a description of low tree density and oak). Also recorded were author explanations for the cause(s) of oak savannas including fire (e.g., Native American burning) or soil conditions. Additional accounts of forest clearings like agricultural fields, NALU, or both were also collected.

Two methods were employed to map historical accounts of oak savannas with sufficient locational information. First, we used a modified point-radius method (Wieczorek, Guo, and Hijmans 2004) to map historical accounts as circles conveying positional uncertainty (Figure S.1 in the supplemental materials). Second, the centroids of uncertainty circles served as "best guess" point representations of the approximate locations of the historical accounts used in latter analyses. Further details on mapping historical accounts are provided in the supplemental materials.

Table 4. Methods used to assess validity or importance of including NAVs in models of oak savanna distributions

Category	Modeling technique	Method	Applicable models
Variable selection	MaxEnt and BRT	See whether models that considered NAVs as predictors selected NAVs as predictors	Performed for four models: 2, 4, 6, 8
Variable importance	MaxEnt	Calculate percentage contribution, permutation importance, and jackknife test importance of NAVs	Performed for two models: 2, 4
Variable importance	BRT	Calculate relative variable importance of NAVs	Performed for two models: 6, 8
Response curves	MaxEnt and BRT	Examine relationships between oak savanna and NAVs using response curves	Performed for four models: 2, 4, 6, 8
Relative quality	MaxEnt	Compare AIC_c values between models that consider vs. do not consider NAVs during model and variable selection	One comparison made: selection process when including vs. intentionally excluding NAVs
Goodness of fit	MaxEnt	Compare model gain between models that include vs. exclude NAVs	Two comparisons made: models 1 vs. 2, 3 vs. 4
Goodness of fit and predictive performance	MaxEnt and BRT	Compare AUC values evaluated on training data between models that include vs. exclude NAVs	Four comparisons made: models 1 vs. 2, 3 vs. 4, 5 vs. 6, 7 vs. 8
Predictive performance	MaxEnt and BRT	Compare AUC values evaluated on test data between models that include vs. exclude NAVs	Two comparisons made: models 3 vs. 4, 7 vs. 8

Notes: See Table 3 for explanations of models. NAV = Native American variables; BRT = boosted regression tree; AIC_c = Akaike information criterion corrected for small sample sizes; AUC = area under the receiver operating characteristic curve.

Synthesizing Model Predictions and Historical Accounts

Model predictions and historical accounts of oak savannas were synthesized in two ways. First, we assessed whether models that included NAVs were better able to predict locations where historical accounts described oak savannas. Two test data sets were generated for this purpose, representing the presence and absence of oak savanna accounts. For both, unique point locations of historical accounts served as "best guess" presence locations for oak savannas. The test data sets differed in how they represented absence of oak savanna accounts. In one data set, absences were an equal number of randomly generated points. In the other data set, absence points were historical accounts of "heavily timbered" or "wilderness" forest conditions prior to European-American settlement. We then evaluated model performance with the AUC measure using these two test data sets.

The second synthesis sought to assess possible consensus between models and historical accounts regarding the causes of oak savannas. We used models (with NAVs) to determine the relative effect of environmental conditions and NALU on oak savannas, by creating an index of how influenced oak savannas were by environment conditions versus NALU. Creating this index involved three steps. First, we converted model predictions (i.e., continuous probabilities) into a binary presence–absence map of oak savannas using thresholds described previously. Second, we simulated gradually decreasing pressure from NALU by applying a uniform 10-km increase in distance from village clusters for both NAVs, assuming that increasing the distance from village clusters would represent less NALU. Simulations were performed for each 10-km interval from 10 km to 60 km. At each interval the same threshold from the first step was applied to predict presence and absence of oak savannas. Third, binary predictions of oak savannas from the seven binary maps (i.e., from each of the six simulations, plus the original binary prediction) were summed. The output was a map interpreted as an ordinal index (1–7) of oak savanna influence by environmental or NALU variables. Higher values indicated areas with oak savannas more resilient under decreasing pressure from NALU and thus more environmentally driven, whereas lower values indicated areas with oak savannas that "disappeared" without NALU. Comparisons between this map and the locations of oak savanna accounts were then made.

Results

Models of Oak Savanna Distribution

A total of 182 out of 5,989 points (3.0 percent) were designated as oak savanna presences for model training based on the OLSR data obtained (Figure 1B). Terms most often associated with oak savannas in OLSRs were thinly timbered (74.7 percent), followed by scattering timber (6.6 percent), plains (6.0 percent), open woods (3.3 percent), open (2.2 percent), and openings (2.2 percent). Remaining points (4.9 percent) were associated with combinations of these terms or other unique terms.

For MaxEnt models that considered NAVs (i.e., Models 2 and 4; Table 3), both NAVs were selected using variable selection procedures (Table S.1 in the supplemental materials). Based on variable importance measures, distance to post-1700 Native American village clusters was the most or second-most important predictor of oak savanna distribution in these models. The probability of oak savannas being observed diminished rapidly with increasing distance from these clusters (from 0 to 43 km; Figure 2B). Distance to pre-1700 Native American village clusters was the third- to fifth-most important predictor of oak savanna distribution, and the probability of oak savanna being observed consistently diminished with increasing distance from these clusters (from 0 to 94 km; Figure 2C). NAVs possessed a combined importance of 38.2 (Model 2) to 37.6 (Model 4) percent using the percentage contribution measure and 62.5 (Model 2) to 52.1 (Model 4) percent using the permutation importance measure.

For BRT models that considered NAVs (i.e., Models 6 and 8; Table 3), both NAVs were selected using variable selection procedures (Table S.2 in the supplemental materials). Distance to pre-1700 Native American village clusters was the most important predictor of oak savanna distribution in both BRT models that considered NAVs. As judged from partial dependence plots, areas within approximately 40 km of pre-1700 village clusters were associated with a higher probability of oak savannas (Figure 3A). Distance from post-1700 Native American village clusters was the third-most important in these models, and areas within approximately 15 km of these clusters were associated with a higher probability of oak savannas (Figure 3C). NAVs possessed a combined relative importance of 39.0 (Model 5) to 38.3 (Model 7) percent.

Mean growing season precipitation was the most important environmental predictor of oak savanna distribution, being the most to third-most important across all MaxEnt models (Table S.1) and the most or second-most important across all BRT models (Table S.2). Lower precipitation amounts were associated with oak savanna presence (Figures 2A and 3B). Six other environmental predictors were selected for all MaxEnt models: from most to least important they were ranked soil drainage class, soil pH, percentage sand, mean terrain slope angle, lacustrine silt and clay, and depth to soil restrictive layer (Figure 2). Four other environmental predictors were selected for all BRT models: from most to least important they were mean annual temperature, ranked soil drainage class, soil pH, and mean terrain slope angle (Figure 3).

Measures of model relative quality, goodness of fit, and predictive performance showed that including NAVs improved models. AUC, a measure of predictive performance calculated for MaxEnt and BRT models, ranged from 0.863 to >0.999 when models were evaluated on training data, and from 0.819 to 0.937 when models were evaluated on test data (Table 5). In nearly all comparisons between models, including NAVs increased AUC regardless of modeling technique or data used (i.e., training or test data) to calculate AUC.

Using models that included NAVs, continuous model predictions were converted into binary predictions of oak savanna distribution (Figure 4). Estimates varied by technique and by threshold used to convert continuous probabilities into a binary presence–absence of oak savannas. For Model 2 (a MaxEnt model), estimates of land area covered by oak savannas prior to European-American settlement ranged from 1,586 to 4,714 km^2, or 5.8 to 17.4 percent of the study area (Figure 4). For Model 6 (a BRT model), estimates ranged from 624 to 1,549 km^2, or 2.3 to 5.7 percent of the study area. Estimates were calculated excluding areas without

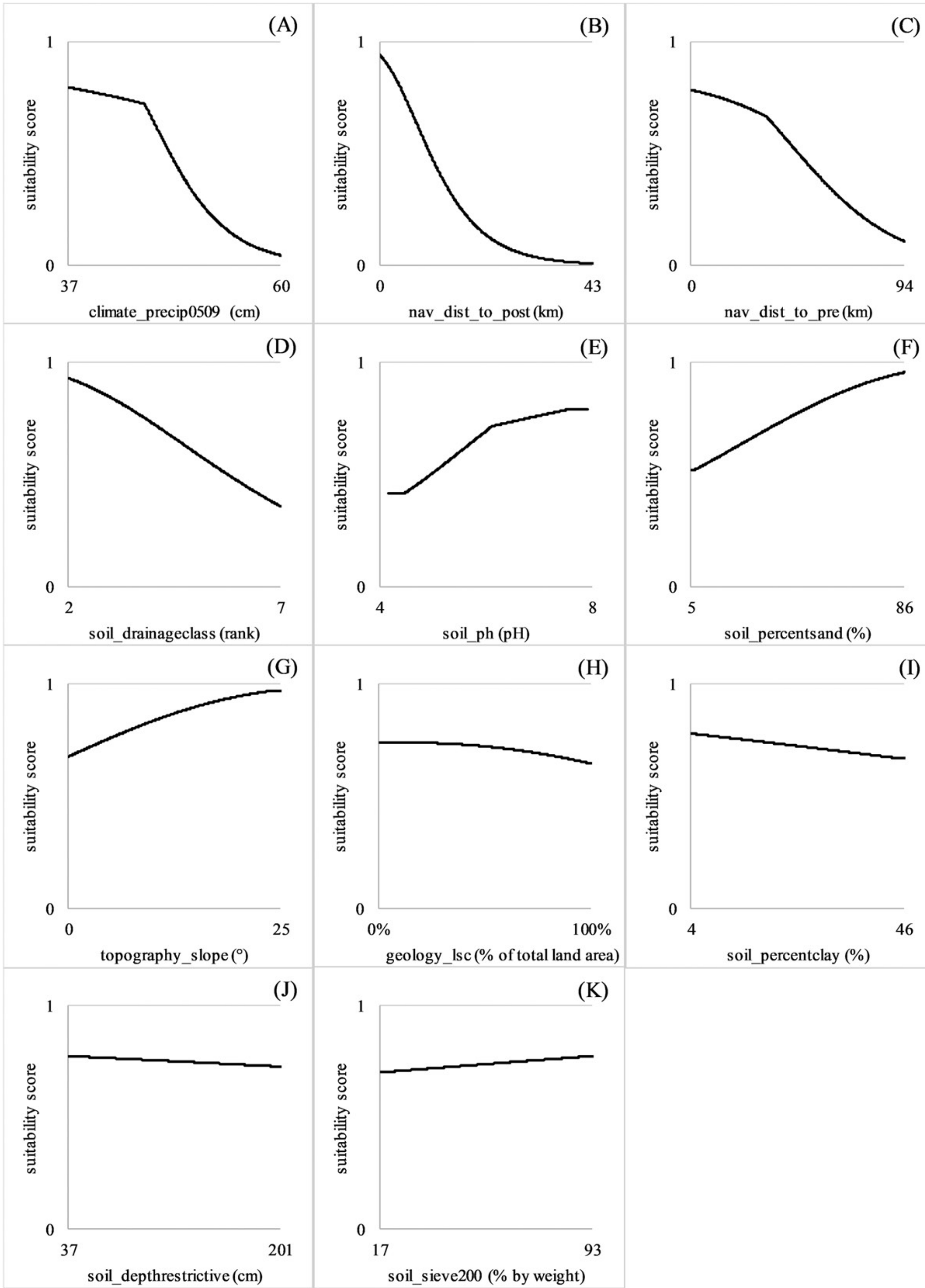

Figure 2. Curves showing the relationship between predictors and oak savanna probability for Model 2 (a MaxEnt model), in order from most to least important predictors. For predictor explanations, see Table 1.

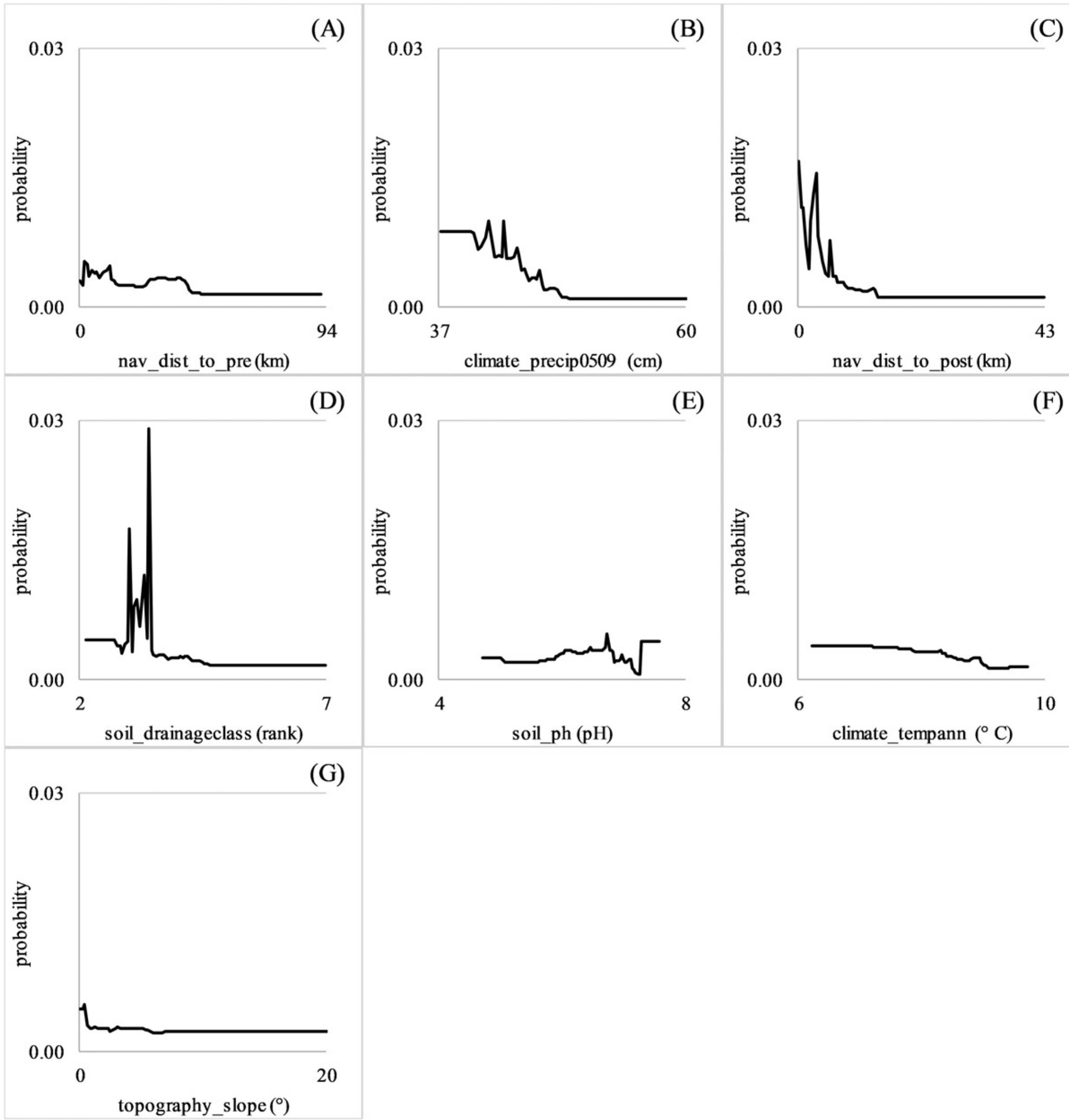

Figure 3. Partial dependence plots showing the relationship between the predictors and predicted probabilities for oak savanna for Model 6 (a boosted regression tree model). For predictor explanations, see Table 1.

soil data and therefore without model predictions (e.g., cities of Buffalo and Rochester, water bodies, and the Tuscarora Indian Reservation; Figure 1A). The median estimate across all modeling techniques and thresholds used was 5.8 percent of the study area.

Analysis of Historical Accounts of Oak Savanna Distribution

From the historical documents, we discovered 245 accounts describing forest clearings, of which 184 could be mapped (Figure 5). The number of historical accounts per unique locality for all accounts of forest clearings ranged from one (for sixty-four localities) to twenty-one (the Genesee Valley). Of the 184 mapped accounts, sixty-two described oak savannas across forty-seven unique localities (Figure 5B). Detail in the accounts varied: Some mentioned forest clearings within a narrative (e.g., "[he] traveled … to the oak openings east of Buffalo"), whereas others provided lengthy descriptions of oak savannas. The number of historical accounts identifying oak savannas per unique locality ranged from one (for

Table 5. AUC measures for all models developed in this study, calculated using OLSR data and historical accounts as independent test data

			Calculated using OLSR data		Calculated using historical accounts of oak savannas	
Model	Modeling technique	NAVs included?	AUC, training data	AUC, test data	AUC, with random points as absence locations	AUC, with "wilderness" accounts as absence locations
1	MaxEnt	No	0.863	N/A	0.796	0.879
2	MaxEnt	Yes	0.919	N/A	0.883	0.887
3	MaxEnt	No	0.881	0.819	0.811	0.878
4	MaxEnt	Yes	0.930	0.886	0.892	0.884
5	BRT	No	0.998	N/A	0.912	0.840
6	BRT	Yes	0.999	N/A	0.900	0.830
7	BRT	No	0.994	0.896	0.879	0.809
8	BRT	Yes	>0.999	0.937	0.922	0.849

Notes: AUC is calculated differently for modeling techniques that use presence-only (i.e., MaxEnt) versus presence–absence (i.e., BRT) data. Refer to Table 3 for model explanations. AUC = area under the receiver operating characteristic curve; OLSR = original land survey record; NAVs = Native American variables; BRT = boosted regression tree.

thirty-eight localities) to five (the Town of Oakfield). Of the sixty-two oak savanna accounts, twenty-nine were firsthand eyewitness accounts, ten were secondhand accounts (accounts by eyewitness observers but summarized by an author), and twenty-three were of unknown origin. County histories, traveler accounts, and development district histories together produced 77 percent of all oak savanna accounts.

Some accounts offered no descriptions of the origin of forest clearings, whereas others had detailed accounts of their possible causes. Of the 184 mapped accounts of forest clearings, 113 (61.4 percent) were associated with NALU: The details varied, but phrases like "Indian clearings" and "Indian meadows" were used to describe clearings, and thirty-one accounts connected forest clearings to Native American burning (Figure 5D). Of the sixty-two accounts identifying oak savannas, forty-four attributed or implied at least one cause, and NALU was implicated in twenty-six of these accounts (59.1 percent; Figure 5E and F). Causes ascribed in these forty-four oak savanna accounts were eleven to Native American fire use and soils, eleven to soils only (i.e., dry or rocky conditions), seven to Native American fire use only, seven to NALU only but without mentioning fire, seven to fire only but without mentioning Native Americans, and one to soils and NALU but without mentioning fire. A portion of oak savanna accounts associated them with generally good soils (Figure 5E). Of the twenty-six oak savanna accounts that attributed at least partial cause to Native Americans, nine were firsthand eyewitnesses, two were secondhand, and fifteen were of unknown origin. Of the eighteen oak savanna accounts that attributed at least partial cause to Native American burning, six were firsthand eyewitnesses and twelve were of unknown origin.

Comparison of Models with Historical Accounts

A correspondence existed between locations of oak savanna accounts and model predictions of oak savanna (Figure 6). The largest areas that models predicted to contain oak savannas also corresponded with accounts of Native American burning. Conversely, in areas without NALU accounts or with accounts of NALU but not burning, models typically predicted the absence of oak savannas. Models that included NAVs as predictors were generally better at predicting the "best guess" locations (i.e., centroids) of oak savanna accounts than models without NAVs (Table 5). When examining oak savanna causes, areas with oak savannas most resilient to a simulated decrease in pressure from NALU (Figure 4) generally corresponded to those locations where oak savanna accounts described poor soils: a linear region stretching from east of modern-day Buffalo through the Tonawanda Indian Reservation, Canawaugus, and Canandaigua (Figures 1A and 5F). Many of these same areas were also associated with NALU described in oak savanna accounts, however (Figure 5E). Oak savannas least resilient to a simulated decrease in NALU were those generally south

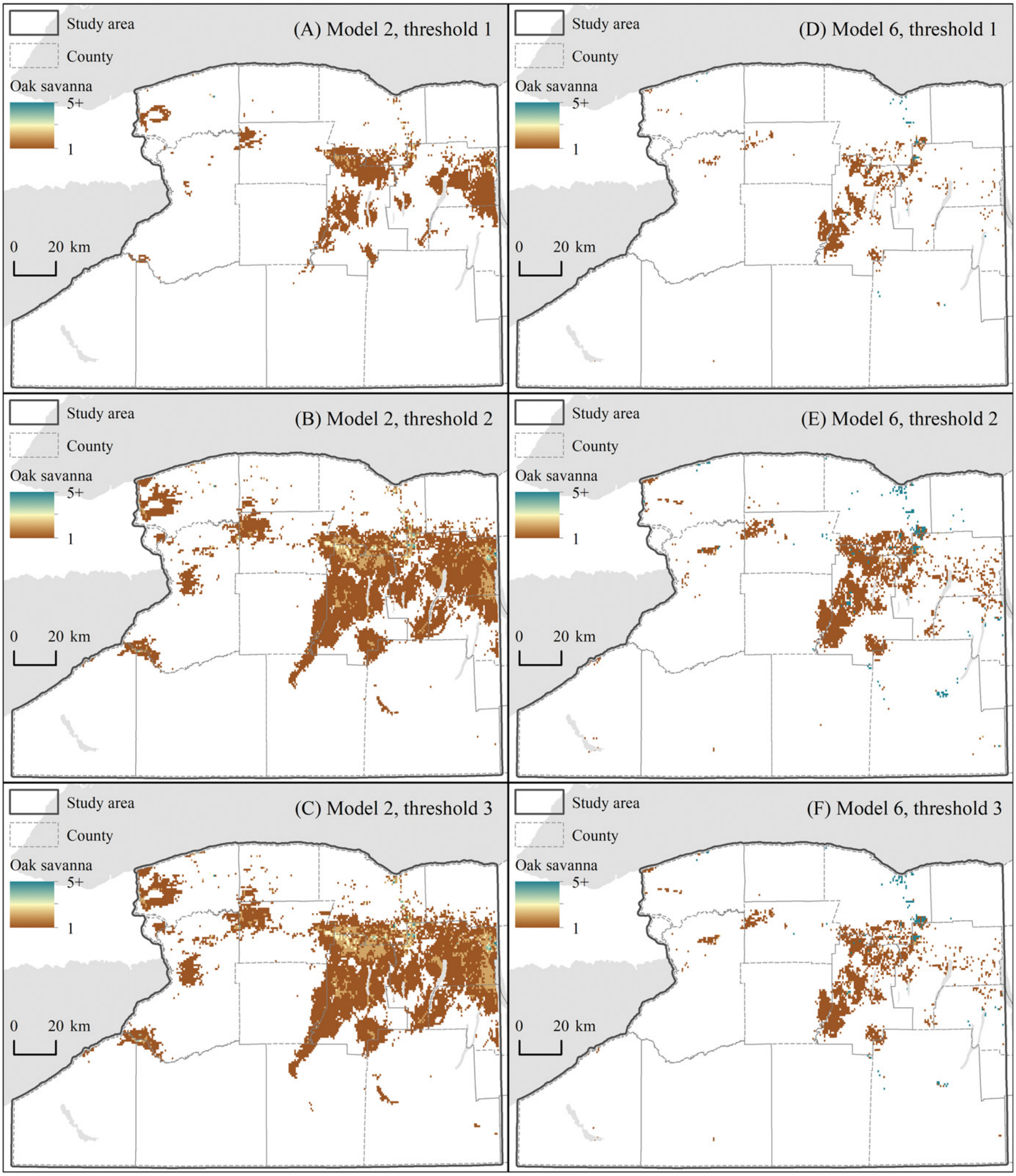

Figure 4. Predictions of oak savanna distribution. Model predictions on the left are from Model 2 (a MaxEnt model), and those on the right are from Model 6 (a boosted regression tree model). See Table 3 for model descriptions. Threshold 1 is maximize kappa, Threshold 2 is equal sensitivity and specificity, and Threshold 3 is maximize sensitivity plus specificity. Symbolized is the index estimating the degree to which oak savannas were environmentally driven versus anthropogenic (i.e., due to NALU); low values indicate oak savannas more attributable to NALU. NALU = Native American land use.

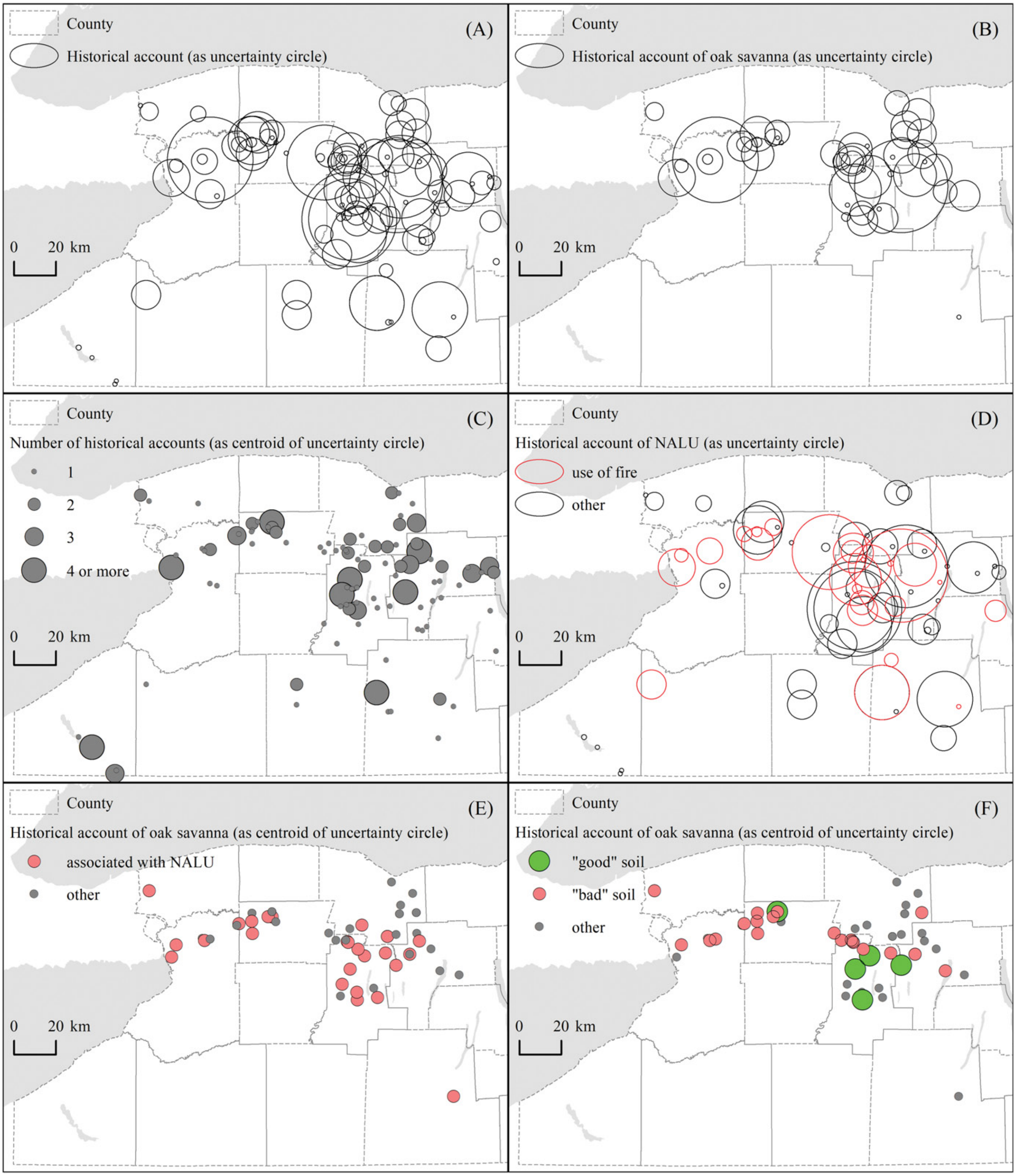

Figure 5. Distribution of historical accounts: (A) all accounts, (B) oak savanna, (C) number of accounts per unique location, (D) NALU, (E) oak savannas associated with NALU, and (F) oak savannas described as occurring on generally “good” or “bad” soils. NALU = Native American land use.

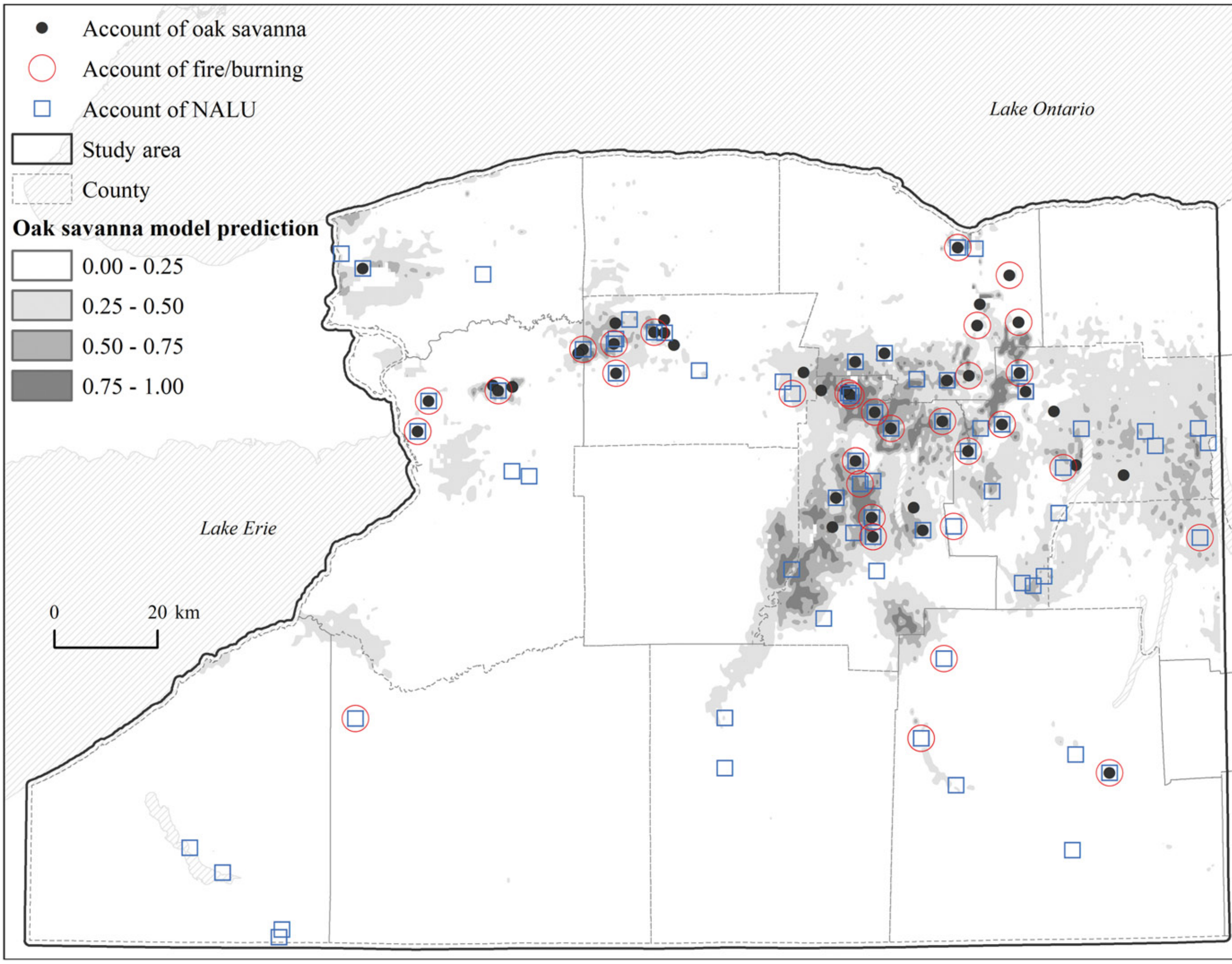

Figure 6. Comparison between model predictions (i.e., mean of Models 2 and 6) and historical accounts of oak savannas, burning, or NALU. Depending on its characteristics, one account can be symbolized up to three times (i.e., an account describing oak savanna, NALU, and fire/burning). Accounts are mapped using "best guess" locations. NALU = Native American land use.

of the band of poor soils described earlier (Figure 4) and in areas near Native American settlements at the fringes of oak savannas on poorer soils.

Discussion

Oak Savannas Resulted from Native American Burning and Dry Environmental Conditions

Results suggest that oak savannas covered sizable portions of western New York and that Native Americans influenced their distribution through land-use practices including burning. Models selected NAVs as predictors of oak savannas, assigned high importance to NAVs, and exhibited increased predictive performance when NAVs were included. NAVs accounted for roughly one third to one half the variable importance as judged from different variable importance measures and modeling techniques (Tables S.1 and S.2).

The most important NAV for MaxEnt models was distance to post-1700 village clusters, but for BRT models it was distance to pre-1700 village clusters. Distance to post-1700 village clusters might be most important because land-use practices like burning might have maintained oak savannas close to settlements until shortly before European-American settlement. Tulowiecki and Larsen (2015) reached similar interpretations regarding the influence of more recent versus older NALU on forest patterns. Distance to pre-1700 village clusters might be more important, however, given that these villages were associated with more populated and longer occupied cultural centers with more intensive NALU across time and space. Interpretation of these results requires further consideration of place-specific cultural and historical factors.

Models generally agreed on the nature of the relationship between oak savannas and proximity to pre- or post-1700 settlement: Oak savannas were found more diffusely, and at further distances, from pre-1700 village clusters, yet were more concentrated near post-1700 village clusters (Figures 2 and 3). These results reflect the history of Native American settlement in the area: As Native Americans were displaced by European-Americans, their settlements became more dispersed and their land use more localized (Engelbrecht 2003). As inferred from models, oak savannas occurred within approximately 15 km of village sites. This distance is similar to previous research suggesting that areas within 10 km (Black, Ruffner, and Abrams 2006) to 15 km (Tulowiecki and Larsen 2015) of Native American village sites exhibited increased frequency of mast-producing tree species. The modeled peak (in BRT models) in oak savanna distribution around 25 to 40 km from pre-1700 village clusters (Figure 3A) might imply hunting and fire use far from villages, but it could also stem from model overfitting to extensive oak savannas southwest of Geneseo shaped by post-1700 settlement (Figures 1B and 6).

Although this study showed that NALU maintained oak savannas, it also suggested the importance of environmental conditions. In all models, drought-related predictors (i.e., mean growing-season precipitation and soil drainage) were among the most important predictors of oak savannas, with drier conditions favoring oak savannas. In the study area, somewhat excessively drained and well-drained soils are generally associated with coarse-textured or rocky soils. These results are consistent with previous literature on historical drivers of oak savanna distribution in this area (Shanks 1966; Edinger et al. 2014). The absence of accounts describing Native American burning in areas with higher precipitation suggests that burning was not attempted in moister environmental conditions. Furthermore, model simulations of Native American absence (Figure 4) suggest that areas with more environmentally influenced oak savannas bordered areas with NALU-caused oak savannas, implying that Native Americans encouraged and expanded oak savannas in these areas.

The spatial pattern of historical oak savanna accounts reflected the predicted locations of oak savannas (Figure 6), and the causes of oak savannas (either stated or speculated on) within historical accounts support the causes inferred from model simulations (Figures 4 and 5). For example, areas northwest of Canawaugus (Figures 1A and 4) typify the combined effects of 1700's Native American burning and poor soils (Figure 5E and F): A traveler wrote that he "passed into a plain, which probably in former times had been annually ravaged by fires" but also noticed that the "soil [was] chiefly sand and gravel" (Thomas 1819). This area corresponds to locations of model consensus regarding the presence of oak savannas and where models predicted oak savannas would be somewhat resilient in the absence of NALU (Figure 4). As another example, the Geneseo area (Figures 1A and 4) demonstrates oak savannas maintained and potentially created by Native American burning, with less environmental influence. An early European-American settler here observed that "[t]he openings grew up to a tall red grass, which was generally burnt over every fall by the Indians. … [T]he land was considered poor. … [T]hen the land seemed to come right up. … [F]armers could then raise as good crops as we of the valley" (Doty 1905). This area corresponds to locations with model consensus on oak savanna distribution and where models simulated the loss of oak savannas without NALU.

This study produced varying model estimates of the spatial extent of oak savannas due to different modeling techniques and different thresholds that converted raw model outputs to map oak savanna distribution (Figure 4). Considering both OLSR-derived and model-derived estimates, oak savannas covered 2 to 17 percent of western New York around 1795, more area than is covered by oak savannas in all of the state today (Edinger et al. 2014). Due to characteristics of the OLSR data, modeling techniques, and study area, coupled with the OLSR-derived estimate of 3.0 percent, we favor lower estimates made by BRT models of 3 to 6 percent and also favor variable importance measures (Table S.2) and response curves (Figure 3) associated with this technique. Techniques like MaxEnt are often used to predict potential distributions of geographic phenomena (e.g., species distributions; Franklin and Miller 2009) using more limited presence-only data, in contrast to more systematically collected data that characterize OLSR data used in this study. We favor BRT model estimates higher than the OLSR-derived estimate, partially because surveyors delineated Indian reservations and potentially enclosed oak savannas instead of surveying through them, thereby lowering the OLSR-derived estimate. This interpretation is supported by accounts from travelers who visited Indian reservations (e.g., Tonawanda Indian Reservation; Dwight 1823) and described oak savannas within them.

Although models and historical accounts suggest the maintenance and expansion of oak savannas, their formation is difficult to determine from this study. Assessing the creation of oak savannas requires study over longer temporal scales, such as by analyzing paleoecological data sets. Szeicz and MacDonald (1991), for example, reasoned from pollen and charcoal in lake sediments that oak savannas in nearby southern Ontario were established 6,000 to 4,000 years BP and predated Native American settlement but that Native American burning maintained oak savannas until European-American settlement.

Synthesis of Methodological Approaches Strengthens the Study of Oak Savannas

This study shows that synthesizing results from historical analysis and modeling enhances the study of NALU and its effects on forests and oak savannas. Combining these two commonly used but heretofore not synthesized approaches offsets limitations of each. Ample historical accounts of NALU in areas that our models predicted should have NALU-influenced oak savannas strongly support this study's modeling results. Conversely, models supported historical accounts through quantifying the importance of NALU and the spatial extent of NALU-driven impacts, results not obtainable through analysis of historical accounts alone. Furthermore, agreement between model predictions and historical accounts supports the interpretation of various descriptions (e.g., "thinly timbered" oak landscapes) as oak savanna in OLSRs.

The addition of historical accounts gives confidence to findings in the face of common modeling issues. For instance, overfitting of models to training data is a problem that might reduce the transferability of models to make accurate predictions in other geographic areas (Radosavljevic and Anderson 2014). Differences between AUC measures when calculated using training versus test data suggest modest overfitting in this study (Table 5). The inclusion of historical accounts, however, helped assuage concerns regarding overfitting. For example, historical accounts of oak savannas occurred near present-day Tuscarora and Tonawanda Indian Reservations (Figures 1A and 6), two locations where models predicted oak savannas even where few or no observed presences were derived from OLSRs.

Although we found few cases in which accounts of oak savannas existed where models did not predict them, some locations that models predicted to have oak savannas did not possess any historical accounts. These differences could be attributable to four reasons. First, although many historical documents were examined, accounts of oak savannas might still have been biased toward early travel routes and centers of European-American settlement and away from more remote areas. Second, accounts of oak savannas might not exist in certain areas due to a time lag between Native American and European-American settlement during which oak savannas could have reverted to closed forests. Third, as discussed previously, MaxEnt predictions might represent potential distribution and overestimate oak savanna distribution in this study. Finally, NAVs were developed for models by treating all village clusters as having equal weight, potentially ignoring differences between historical NALU at different locations due to varying Native American populations and durations of occupation.

Conclusion

This study advances larger bodies of geographical research investigating Native American influence on forests at the time of European-American settlement and research into savanna landscapes more broadly. It offers a new approach that synthesizes methods in historical analysis and predictive modeling for disentangling the complex human and environmental factors shaping past oak savanna distribution. This study provides evidence of the importance of anthropogenic burning in historic oak savannas, thereby informing land managers seeking to maintain or restore oak savannas of ecological or cultural importance through prescribed burning.

Acknowledgments

The authors thank Brendan Cullen, Noah Haber, and Peter Scilla for transcribing and mapping historical data for this study and Charles V. Cogbill and Greg Guderian for providing transcriptions of Morris Reserve OLSRs.

Funding

This research received support from the National Science Foundation under Grant No. 1660388.

Supplemental Material

Supplemental data for this article can be accessed on the publisher's website.

ORCID

Stephen J. Tulowiecki http://orcid.org/0000-0002-6787-0962

References

Anderson, R. C., J. S. Fralish, and J. M. Baskin. 1999. Introduction. In *Savannas, barrens, and rock outcrop plant communities of North America*, ed. R. C. Anderson, J. S. Fralish, and J. M. Baskin. New York: Cambridge University Press.

Bean, W. T., and E. W. Sanderson. 2008. Using a spatially explicit ecological model to test scenarios of fire use by Native Americans: An example from the Harlem Plains, New York, NY. *Ecological Modelling* 211 (3–4):301–8. doi:10.1016/j.ecolmodel.2007.09.014.

Black, B. A., and M. D. Abrams. 2001. Influences of Native Americans and surveyor biases on metes and bounds witness-tree distribution. *Ecology* 82 (9):2574–86. doi:10.1890/0012-9658(2001)082[2574:IONAAS.2.0.CO;2]

Black, B. A., C. M. Ruffner, and M. D. Abrams. 2006. Native American influences on the forest composition of the Allegheny Plateau, northwest Pennsylvania. *Canadian Journal of Forest Research* 36 (5):1266–75. doi:10.1139/x06-027.

Cappon, L. J., ed. 1976. *Atlas of early American history: The revolutionary era, 1760–90*. Princeton, NJ: Princeton University Press. doi:10.1086/ahr/82.2.432-a.

Chapman, K. A., and R. Brewer. 2008. Prairie and savanna in southern lower Michigan: History, classification, ecology. *The Michigan Botanist* 47:1–48.

Cronon, W. 1983. *Changes in the land: Indians, colonists, and the ecology of New England*. New York: Hill and Wang.

Day, G. M. 1953. The Indian as an ecological factor in the northeastern forest. *Ecology* 34 (2):329–46. doi:10.2307/1930900.

Denevan, W. M. 1992. The pristine myth: The landscape of the Americas in 1492. *Annals of the Association of American Geographers* 82 (3):369–85. doi:10.1111/j.1467-8306.1992.tb01965.x.

Doty, L. R. 1905. *History of Livingston County, New York: From its earliest traditions to the present, together with early town sketches*. Jackson, MI: W. J. Van Deusen.

Duvall, C. S. 2011. Biocomplexity from the ground up: Vegetation patterns in a West African savanna landscape. *Annals of the Association of American Geographers* 101 (3):497–522. doi:10.1080/00045608.2011.560061.

Dwight, T. 1823. *Travels in New-England and New-York*. London: W. Baynes and Ogle.

Edinger, G. J., D. J. Evans, S. Gebauer, T. G. Howard, D. M. Hunt, and A. M. Olivero. 2014. *Ecological communities of New York State: A revised and expanded edition of Carol Reschke's ecological communities of New York State*. 2nd ed. Albany: New York Natural Heritage Program, New York State Department of Environmental Conservation.

Elith, J., J. R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* 77 (4):802–13. doi:10.1111/j.1365-2656.2008.01390.x.

Engelbrecht, W. 2003. *Iroquoia: The development of a native world*. Syracuse, NY: Syracuse University Press.

ESRI. 2017. *ArcGIS 10.5.1*. Redlands, CA: Esri.

Fenneman, N. M. 1938. *Physiography of eastern United States*. New York: McGraw-Hill.

Filby, P. W. 1985. *A bibliography of American county histories*. Baltimore, MD: Genealogical Publishing Company.

Franklin, J., and J. A. Miller. 2009. *Mapping species distributions: Spatial inference and prediction*. New York: Cambridge University Press.

Gleason, H. A. 1922. The vegetational history of the Middle West. *Annals of the Association of American Geographers* 12 (1):39–85. doi:10.2307/2560590.

Grumet, R. S. 1995. *Historic contact: Indian people and colonists in today's northeastern United States in the sixteenth through eighteenth centuries*. Norman: University of Oklahoma Press.

Hays, J., and C. F. Post. 1999. *Journey on the forbidden path: Chronicles of a diplomatic mission to the Allegheny country, March–September, 1760*, ed. R. S. Grumet, J. A. Rementer, and B. L. Pearson. Philadelphia: American Philosophical Society.

Hijmans, R. J., S. Phillips, J. Leathwick, and J. Elith. 2017. *Dismo: Species distribution modeling* (R package version 1.1-4). http://CRAN.R-project.org/package=dismo.

House, J. I., S. Archer, D. D. Breshears, R. J. Scholes, and NCEAS Tree-Grass Interactions Participants. 2003. Conundrums in mixed woody-herbaceous plant systems. *Journal of Biogeography* 30 (11):1763–77. doi:10.1046/j.1365-2699.2003.00873.x.

Internet Archive. 1996. Internet Archive. Accessed September 10, 2017. https://archive.org/.

Jennings, F., and W. N. Fenton. 1995. *The history and culture of Iroquois diplomacy: An interdisciplinary guide to the treaties of the six nations and their league*. Syracuse, NY: Syracuse University Press.

Jones, E. E. 2010. An analysis of factors influencing sixteenth and seventeenth century Haudenosaunee (Iroquois) settlement locations. *Journal of Anthropological Archaeology* 29 (1):1–14. doi:10.1016/j.jaa.2009.09.002.

Jueterbock, A. 2018. *MaxentVariableSelection: Selecting the best set of relevant environmental variables along with the optimal regularization multiplier for Maxent niche modeling*. R package version 1.0-3. https://cran.r-project.org/web/packages/MaxentVariableSelection/index.html.

Keister, M. 1998. *Rush Oak Openings unit management plan*. Bath, NY: New York State Department of Environmental Conservation.

Klimaszewski-Patterson, A., P. J. Weisberg, S. A. Mensing, and R. M. Scheller. 2018. Using paleolandscape modeling to investigate the impact of Native American-set fires on pre-Columbian forests in the southern Sierra Nevada, California, USA. *Annals of the American Association of Geographers* 108:1635–54.

Laris, P. 2011. Humanizing savanna biogeography: Linking human practices with ecological patterns in a frequently burned savanna of southern Mali. *Annals of the Association of American Geographers* 101 (5):1067–88. doi:10.1080/00045608.2011.560063.

Matlack, G. R. 2013. Reassessment of the use of fire as a management tool in deciduous forests of eastern North America. *Conservation Biology* 27 (5):916–26. doi:10.1111/cobi.12121.

McClenachan, L., A. B. Cooper, M. G. McKenzie, and J. A. Drew. 2015. The importance of surprising results and best practices in historical ecology. *BioScience* 65 (9):932–39. doi:10.1093/biosci/biv100.

McEwan, R. W., J. M. Dyer, and N. Pederson. 2011. Multiple interacting ecosystem drivers: Toward an encompassing hypothesis of oak forest dynamics across eastern North America. *Ecography* 34 (2):244–56. doi:10.1111/j.1600-0587.2010.06390.x.

Morgan, L. H. 1901. *League of the Ho-De'-No-Sau-Nee, or Iroquois*, ed. H. M. Lloyd. New York: Dodd, Mead.

Munoz, S. E., D. J. Mladenoff, S. Schroeder, and J. W. Williams. 2014. Defining the spatial patterns of historical land use associated with the indigenous societies of eastern North America. *Journal of Biogeography* 41 (12):2195–2210. doi:10.1111/jbi.12386.

Natural Resources Conservation Service. 2014. Description of SSURGO database. Accessed August 16, 2017. http://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/?cid=nrcs142p2_053627.

New York State Museum. 2018. Geographic Information System (GIS). Accessed July 13, 2018. http://www.nysm.nysed.gov/research-collections/geology/gis.

Nowacki, G. J., and M. D. Abrams. 2008. The demise of fire and "mesophication" of forests in the Eastern United States. *Bioscience* 58 (2):123–38. doi:10.1641/B580207.

Parker, A. C. 1920. *The archeological history of New York*. Albany: The University of the State of New York.

Phillips, S. J. 2010. A brief tutorial on Maxent. *Lessons in Conservation* 3:108–35.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190 (3–4):231–59. doi:10.1016/j.ecolmodel.2005.03.026.

Phillips, S. J., M. Dudik, and R. E. Schapire. 2017. Maxent software for modeling species niches and distributions. Accessed April 27, 2017. http://biodiversityinformatics.amnh.org/open_source/maxent/.

PRISM Climate Group. 2013. Gridded climate data for the contiguous USA. Accessed May 21, 2017. http://www.prism.oregonstate.edu/.

Pyne, S. J. 1982. *Fire in America: A cultural history of wildland and rural fire*. Princeton, NJ: Princeton University Press. doi:10.1086/ahr/88.3.759.

R Development Core Team. 2011. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. http://www.R-project.org.

Radosavljevic, A., and R. P. Anderson. 2014. Making better Maxent models of species distributions: Complexity, overfitting and evaluation. *Journal of Biogeography* 41 (4):629–43. doi:10.1111/jbi.12227.

Robertson, D., C. P. S. Larsen, and S. J. Tulowiecki. 2018. Forest land-use legacy research exhibits aspects of critical physical geography. In *The Palgrave handbook of critical physical geography*, ed. R. Lave, C. Biermann, and S. N. Lane, 227–48. Cham, Switzerland: Palgrave Macmillan.

Rostlund, E. 1957. The myth of a natural prairie belt in Alabama: An interpretation of historical records. *Annals of the Association of American Geographers* 47 (4):392–411. doi:10.1111/j.1467-8306.1957.tb01552.x.

Ryan, K. C., E. E. Knapp, and J. M. Varner. 2013. Prescribed fire in North American forests and woodlands: History, current practice, and challenges. *Frontiers in Ecology and the Environment* 11 (Suppl. 1):e15–24. doi:10.1890/120329.

Sauer, C. O. 1927. *Geography of the Pennyroyal: A study of the influence of geology and physiography upon industry, commerce and life of the people*. Louisville, KY: Lost Cause Press.

Seischab, F. K. 1992. Forests of the Holland Land Company in western New York, circa 1798. In *Late eighteenth century vegetation of central and western New York State on the basis of original land survey records. New York State Museum, Bulletin 484*, ed. P. L. Marks, S. Gardescu, and F. K. Seischab, 36–53. Albany: New York State Museum, University of the State of New York.

Seischab, F. K., and D. Orwig. 1991. Catastrophic disturbances in the presettlement forests of western New York. *Bulletin of the Torrey Botanical Club* 118 (2):117–22. doi:10.2307/2996851.

Shanks, R. E. 1966. An ecological survey of the vegetation of Monroe County, New York. *Proceedings of the Rochester Academy of Science* 11 (3):105–255.

Smith, B. D. 2011. General patterns of niche construction and the management of "wild" plant and animal resources by small-scale pre-industrial societies. *Philosophical Transactions of the Royal Society B-Biological Sciences* 366 (1566):836–48. doi:10.1098/rstb.2010.0253.

Snow, D. R. 1996. *The Iroquois*. Malden, MA: Blackwell.

Stewart, O. C. 2002. *Forgotten fires: Native Americans and the transient wilderness*, ed. H. T. Lewis and K. Anderson. Norman: University of Oklahoma Press.

Szeicz, J. M., and G. M. MacDonald. 1991. Postglacial vegetation history of oak savanna in southern Ontario. *Canadian Journal of Botany* 69 (7):1507–19. doi:10.1139/b91-195.

Thomas, D. 1819. *Travels through the western country in the summer of 1816*. Auburn, NY: David Rumsey.

Tulowiecki, S. J. 2018. Information retrieval in physical geography: A method to recover geographical information from digitized historical documents. *Progress in Physical Geography: Earth and Environment* 42 (3):369–90. doi:10.1177/0309133318770972.

Tulowiecki, S. J., and C. P. S. Larsen. 2015. Native American impact on past forest composition inferred from species distribution models, Chautauqua County, New York. *Ecological Monographs* 85 (4):557–81. doi:10.1890/14-2259.1.

Tulowiecki, S. J., C. P. S. Larsen, and Y.-C. Wang. 2015. Effects of positional error on modeling species distributions: A perspective using presettlement land survey records. *Plant Ecology* 216 (1):67–85. doi:10.1007/s11258-014-0417-9.

U.S. Geological Survey. 2013. The National Map Viewer and download platform. Accessed February 22, 2018. http://nationalmap.gov/viewer.html.

USGS Gap Analysis Program, Ecological Society of America, USGS, and NatureServe. 2018. The U.S. National Vegetation Classification. Accessed September 1, 2018. http://usnvc.org/.

Vale, T. R. 2002. The pre-European landscape: Pristine or humanized? In *Fire, native peoples, and the natural landscape*, ed. T. R. Vale, 1–40. Washington, DC: Island.

Wang, Y.-C. 2005. Presettlement land survey records of vegetation: Geographic characteristics, quality and modes of analysis. *Progress in Physical Geography* 29 (4):568–98. doi:10.1191/0309133305pp463ra.

Wang, Y.-C. 2007. Spatial patterns and vegetation-site relationships of the presettlement forests in western New York, USA. *Journal of Biogeography* 34 (3):500–513.

White, M. E. 1978. Erie. In *Handbook of North American Indians*, ed. B. G. Trigger and W. C. Sturtevant, 412–17. Washington, DC: Smithsonian Institution.

Whitney, G. G. 1996. *From coastal wilderness to fruited plain: A history of environmental change in temperate North America, 1500 to the present*. New York: Cambridge University Press.

Wieczorek, J., Q. G. Guo, and R. J. Hijmans. 2004. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *International Journal of Geographical Information Science* 18 (8):745–67. doi:10.1080/13658810412331280211.

STEPHEN J. TULOWIECKI is an Assistant Professor in the Department of Geography at SUNY Geneseo, Geneseo, NY 14454. E-mail: tulowiecki@geneseo.edu. His research interests include historical forest conditions, species distribution modeling, and Native American impacts on past landscapes.

DAVID ROBERTSON is a Professor in the Department of Geography at SUNY Geneseo, Geneseo, NY 14454. E-mail: robertsd@geneseo.edu. His research interests include cultural and historical geography with a focus on the cultural and environmental implications of landscape change.

CHRIS P. S. LARSEN is an Associate Professor in the Department of Geography at the University at Buffalo, State University of New York, Buffalo, NY 14261. E-mail: larsen@buffalo.edu. His research interests include wildfire, historical ecology, and anthropogenic impacts on plant communities.