USING CONSERVATION GIS TO BUILD A PREDICTIVE MODEL FOR OAK SAVANNA ECOSYSTEMS IN NORTHWEST OHIO

Marcus Enrico Ricci

A Thesis

Submitted to the Graduate College of Bowling Green State University

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May, 2006

Committee:

Helen J. Michaels, Co-Advisor

Karen V. Root, Co-Advisor

Enrique Gomezdelcampo

ABSTRACT

Helen J. Michaels & Karen V. Root, Co-Advisors

The Oak Openings Region in Northwest Ohio is one of the few remaining remnants of oak savanna and oak barrens, or "oak savanna complex." It is a 33,670 ha complex of globally-significant ecosystems and has more listed species than any other similarly-sized region in the state. Agriculture, drainage and fire suppression have reduced its area by half, underscoring the need to locate and prioritize appropriate habitat for acquisition and conservation. Land managers often have difficulty in implementing regional conservation efforts due to a lack of detailed ecological knowledge or habitat quality data. I used ArcGIS 9.1 to build a predictive geographic model (PGM) to detect oak savanna complex remnants and restorable patches by determining significant ecological variables from known remnant patches. Software and data used was constrained to readily available sources and ecological variables investigated included soil type, elevation, slope, topographic position and aspect. This research used predictive modeling in a new way by using it to predict areas of high probability of a rare ecosystem, rather than its typical use for creating predictive habitat models for individual taxa, multiple taxa or vegetative communities. The resulting model succeeded in locating potential remnants and restorable patches at the landscape level, as well as creating a suitability index to rank the probability of accurately predicting oak savanna complex presence at the landscape level. Both simple statistics and regression analysis were used to determine the significant predictors of oak savanna complex presence: suitable soil types; mean elevation and topographic position. Single-variable predictive models reduced the county-wide search area as much as 93% with a predictive accuracy of 87 – 100% . However, combining these models into a multi-variable model reduced the search area as much as 99%. Regression analysis determined that the model explaining the highest amount of variance used only two ecological variables: suitable soils and mean elevation. Validation of this two-variable model on a randomly-generated data set proved it was 90% accurate in locating high-probability areas of oak savanna complex. This research produces a scientifically robust predictive ecosystem model that more simply and systematically locates and prioritizes conservation at a landscape scale.

This work is dedicated to many.

First, to my partner, Jeannie, for her unflagging love, support and belief in my ability.

Second, to my advisors and my lab mates, for their insight and ever-ready critique and input.

Finally, to the people who have inspired me with their love of the land, the naturalist's naturalists: Kim High, Bob Jacksy and Pam Menchaca.

“A thing is right when it tends to preserve the integrity, stability and beauty of the biotic community. It is wrong when it tends otherwise.”

- Aldo Leopold, Sand County Almanac and Sketches from Here and There, 1949

“The last word in ignorance is the man who says of an animal or plant: "What good is it?" If the land mechanism as a whole is good, then every part is good, whether we understand it or not. If the biota, in the course of aeons, has built something we like but do not understand, then who but a fool would discard seemingly useless parts? To keep every cog and wheel is the first precaution of intelligent tinkering.”

- Aldo Leopold, A Sand County Almanac: with essays on conservation from Round River, 1949

ACKNOWLEDGEMENTS

This research would not have been possible without those resources that all conservation biology projects require: guidance; innovation; data; and funding. I would like to thank my advisors, Drs. Karen Root and Helen Michaels, and my committee member, Dr. Enrique Gomezdelcampo, for their cooperation and willingness to both teach and learn as the project progressed. The spark of the utility of GIS was Dr. Norman Levine; the kindling of conservation biology and landscape ecology was Dr. Kimberly With; the new way to use predictive modeling was Dr. Karen Root. Ecological data was graciously provided by the Lucas County Auditor's Office and the Ohio Department of Natural Resources Division of Geology. The data set of study sites was provided by Gary Haase of The Nature Conservancy and John Jaeger and Tim Schetter of the MetroParks of the Toledo Area. Statistical assistance was provided by Nancy Boudreau of the Bowling Green State University (BGSU) Statistical Consulting Center. Finally, funding for this project was provided by the Department of Biological Sciences, BGSU, and the Ohio Biological Survey Small Grants Program.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

I. INTRODUCTION

Oak Savannas and Barrens: Biodiversity Hotspots in Peril

Oak savannas and barrens, once abundant, are now globally rare ecosystems with less than 20,000 ha remaining globally. Locally, they are listed as critically imperiled with less than 1,000 ha in the entire state of Ohio (The Nature Conservancy 2000; Stein et al. 1995. This rarity, coupled with their importance as habitat for locally rare species, makes it imperative that their conservation and restoration become a priority for land managers, planners, conservation organizations and the general public. Human activities – agricultural conversion, fire suppression, and hydrological alteration – have reduced their extent in the Midwest by over 99% (Conner et al. 2001; Nuzzo 1986), placing savannas and other grasslands at the top of a federal "critically endangered" ecosystem list (Noss et al. 2003). In Ohio, only three known savanna remnants still exist, two in the northwestern Oak Openings Region and one in central Madison County (Fig. 1).

The Oak Openings Region, shown in Fig. 2, is actually a complex of six globally rare ecosystems including oak savanna and oak barrens, hereinafter referred to as oak savanna complex (Brewer & Vankat 2004). This diverse range of ecosystems provides a wide range of habitats for over 177 federal- and state-listed species: over 1/3 of Ohio's listed species and more listed species than any other similarly-sized region in the state. Species of note include the federally-listed Skinner's false foxglove, *Agalinis skinneriana*, and Karner blue butterfly, *Lycaeides melissa samuelis* (Ohio Department of Natural Resources Division of Natural Areas and Preserves 2004; The Nature Conservancy 1997). More species are being identified annually – only 113 had been identified by 2001 (The Nature Conservancy 2000). Historically, the region supported diverse fauna including bison, elk, squirrel, fox, lynx, wolverine, mountain lion, bobcat, grey wolves, black bear, beaver and porcupine (Mayfield 1976). Reported bird species include sandhill cranes, swallow-tailed kites, golden-winged warblers, greater prairie chickens, wild turkeys, eastern ruffed grouse, Henslow's sparrows and lark sparrows (Mayfield 1976). Various insects, especially butterflies and moths, depend on the unique savanna and wet prairie ecosystems for nectar- and host-plants (The Nature Conservancy 2000). For these reasons, The Nature Conservancy (TNC) has identified the Oak Openings Region as one of its “Last Great Places” (Green Ribbon Initiative 2004).

The losses of thousands of hectares of these rare ecosystems and species is doubly important because these losses decrease regional biodiversity. This, in turn, decreases the stability of the environment and its ability to recover from anthropogenic or natural stressors such as land conversion, drought or fire (MacDougall 2005; Meffe & Carroll 1997). As commercial and residential development continue to spread out from the City of Toledo, Ohio, these losses will continue, unless land managers, scientists, politicians and the public quickly identify remaining oak savanna complex patches and target them for acquisition and conservation efforts. Applying the principles of conservation biology and landscape ecology towards prioritizing conservation efforts and analyzing the relationships amongst ecosystems and ecological processes at landscape scales will greatly improve the ability of land managers to respond quickly and effectively to threats to these “great places” (Turner et al. 2001; Meffe & Carroll 1997). This project is designed to fulfill this need by designing a method to accurately locate and prioritize oak savanna complex patches, based on ecological variables.

Oak Savanna Complex Composition and Extent

The extent of oak savanna in North America prior to European settlement was estimated as approximately 50 million hectares (Conner et al. 2001). This habitat was located in two north-south bands, one along the Pacific Coast and one east of the Great Plains. The latter zone was a transition band between the prairie grasslands to the west and the deciduous forests to the east. The Midwestern oak savanna region – the northern half of the central transition band – was estimated as more than 11 million hectares (Nuzzo 1986). Because of its co-existence with prairies and its nature as an extension of the Great Plains grassland complex, Transeau (1935) called this area the "Prairie Peninsula. Due to fire suppression, agriculture conversion and hydrological alteration, less than 0.02% of the historic extent of Midwest oak savannas remains (Conner et al. 2001; Nuzzo 1986). This highlights the crisis that conservationists face: how do we find exactly where the remnants are located and how do we prioritize acquisition and restoration efforts?

The first step in conservation is identifying and quantifying the area of interest. Oak savannas are a terrestrial ecosystem generally described as prairie-like with an herbaceous groundcover of grasses and forbs on a sandy plain and interspersed with groves or individual oak trees (Brewer & Vankat 2004; Noss et al. 2003; Meisel et al. 2002;Leach & Givnish 1999; Packard & Mutel 1997; Dunevitz 1993; Haney & Apfelbaum 1993; Apfelbaum & Haney 1990; Nuzzo 1986; Nuzzo 1985; Hehr 1970; Gordon 1969;Transeau 1935; Sears 1926). They are usually considered an intermediate phase or ecotone between prairies and oak woodlands and not as an independent ecosystem, shown in Fig. 3 (Packard & Mutel 1997). However, the defining ecological variables of oak savannas vary among researchers, disciplines and agencies, which is reflected in the plethora of names for the ecosystem, e.g., oak openings, oak barrens, scrub prairie, brush prairie, and savanna (Nuzzo 1986; Gordon 1969; Sears 1926). Attempts to quantify the density of tree cover of oak savannas have yielded different opinions. Curtis (1959) stated that it was "the most variable of all characteristics of the oak openings" and set a threshold of 50% canopy between savanna and forest "[p]urely for convenience"; he was unable to find any such threshold between savanna and prairie. In summary, each state in the Midwest has a different set of criteria for defining oak savanna (Nuzzo 1986). Anderson's definition, which is the definition accepted by the Ohio Department of Natural Resources for the State of Ohio is "a community in which oaks and occasionally other tree species 'comprise a prominent yet partial overstory' with canopy levels of 10-100% 'above a prairie understory" (Nuzzo 1986). The lack of specificity and great latitude of both variables and quantified values in this definition made it less than useful for *a priori* determination of the significant ecological predictors of oak savanna presence. One option for building a predictive model was to build a deductive model based on the assertions of significant ecological variables through either expert opinion or expert literature. This method has been shown to produce predictive models of lower predictive accuracy and transferability (Corsi et al. 2000; Guisan & Zimmermann 2000). Therefore, I decided to build an inductive model based on known sites of oak savanna complex presence, based on statistically significant ecological variables.

Ecological Processes as Ecological Variables

Oak savanna complex is both created and maintained by three primary ecological processes: soil development from glaciation and other geological activity; periodic fire disturbance; and well-draining surface and groundwater hydrology. Knowledge of the importance of these processes guided my investigation into both the determination of significant ecological variables and construction of the predictive model.

Each of these ecological processes is itself driven by interacting biotic and abiotic factors. For example, soil development is driven by five factors: *climate* and *living organisms* (biota) acting on *parent material* over *time* under the influence of *topography.* The most influential factors are climate, topography and the parent material itself (Brady & Weil 1999). If any one of these factors changes for any of the primary processes, the resulting ecosystem and vegetative community may change, depending on the influence of that particular factor, both individually and synergistically with the other factors (Forsyth 1970). More importantly, a certain combination of variables in one space or time may yield the same ecosystem that results from a different combination of variables in a different space or time. This property of ecosystem flux limits the transferability of any predictive model to a region dictated by both the spatial and geographic extent of the model's training data set. Transferability of the model is also dictated by the extent of the model's explanation of causality versus its function as merely descriptive, as well as the nature of the ecological variables upon which it is built (Corsi et al. 2000; Guisan & Zimmermann 2000).

My investigation for the significant predictors of oak savanna was guided by this knowledge of primary ecological processes – glacial activity, fire, hydrology – and their interacting factors – climate, topography, parent material. These processes resulted in the formation of the Oak Openings Region in the following manner: the global climate around 12,000 years before present (b.p.) created glacial processes that reworked the surface and subsurface strata of North America, resulting in a landscape of alternating sandy ridges, glacial outwash lakebeds and subsurface clay strata that were the future parent material of the Oak Openings Region (Weiher 2003; Meisel et al. 2002; Leach & Givnish 1999;Forsyth 1998, 1968, 1959). Two other ecological processes – fire and hydrology – then reworked the resulting landscape to create the resulting complex of oak savanna, wet prairie and floodplain forests.

Other ecological processes that influence oak savanna structure include activity by vegetation and animals. Digital data on the actual biotic activity was not available, so surrogate measures of productivity and biomass were pursued, including light-availability, canopy-cover and tree-density (Weiher 2003, Leach & Givnish 1999). For all of these landscape-scale ecological processes, quantifying the actual process was not possible. Therefore, I focused on quantifying their results: soil type; elevation; slope; and aspect.

It is also possible to use individual plant or animal taxa as species indicators of oak savanna, but this concept has drawbacks. First, correlation between presence or absence of the species and the ecosystem is not perfect, e.g., because the species is not there does not mean the ecosystem is not there, nor vice versa. Second, ranking of plants and animals according to their conservatism, i.e., level of obligation or facultation to a particular ecosystem, has been performed, but more often for wetlands than uplands. Third, those species most likely to be obligative indicators of oak savanna are also most likely to be rare or listed species, such as the Karner blue butterfly, Lark Sparrow, or Dotted Horsemint. The data on locations and abundances of these species are not openly disclosed to the public and require approval by the ODNR Division of Natural Areas and Preserves and, therefore, would not be readily available information to the typical land manager (Ohio Department of Natural Resources Division of Natural Areas and Preserves 2005).

I investigated the effects of the glacial geological processes by analyzing soil type, elevation and aspect. The geologic ontogeny of the region resulted in a series of high ridges of sandy soils that were the beaches of Glacial Lake Warren roughly 9,700 years b.p. (Forsyth 1959). These sandy ridges, piled higher by winds into dunes before vegetation eventually stabilized them, were meters above the surrounding wetter prairies and floodplain forests and were the sites of oak savanna complex establishment and current remnants.

I also investigated slope and topographic position because they influence the historic and future tendency for fire disturbance, either environment- or human-initiated, by indirectly affecting soil moisture availability, fuel availability and wind speed and direction. Fire is commonly accepted as the primary disturbance factor in maintaining oak savanna health by burning the organic ground layer; warming soil temperatures; releasing soil nutrients from ash; stimulating microbial activity; promoting sod formation which inhibits woody seedling establishment; and killing fire-intolerant herbaceous and woody plants (Brewer & Vankat 2004; Weiher 2003; Meisel et al. 2002; The Nature Conservancy 2000; Packard & Mutel 1997; Haney & Apfelbaum 1993; Apfelbaum & Haney 1990; Gordon 1969; Curtis 1959).

Soil type, slope and topographic position were important because they also influence, or reflect the action of, ground water hydrologic processes. The perched water table caused by the interspersion of the sandy soils and subsurface clay created a combination of dry and saturated soils that resulted in xeric ecosystems in the dunal savannas and mesic and flooded areas in the interdunal wet prairies and their associated vegetative communities (The Nature Conservancy 2000; Mayfield 1976).

Additionally, increases in relative elevation have been linked to higher species richness (Nally et al. 2003). So, even though individual species will not be used as ecological variables, higher elevations may serve as a proxy or indirect indicator of the high species richness associated with oak savannas.

Predictive Modeling and GIS: Expanding the Toolbox

Land managers must often make expedited decisions on planning, conservation and restoration that require scientific knowledge. This often involves identifying areas suitable or unsuitable for both for natural area management as well as for human development. When quantifiable data is lacking, many decisions are made on anecdotal knowledge or based on scientifically unsupported assumptions (Johnson & Gillingham 2004; Clevenger et al. 2002).

Predictive modeling (PM) can determine the suitability of areas for conservation efforts more accurately than merely relying on expert opinion (Vaughan & Ormerod 2003; Corsi et al. 2000; Guisan & Zimmermann 2000). Predictive habitat models are used to locate suitable habitat for selected species. Similarly, predictive geographic models (PGM) are used to locate geographic areas meeting specific ecological criteria. Regardless of the target, these models use one of two general methodologies: deductive or inductive modeling. Deductive models use *a priori* assumptions of the ecological variables associated with the presence of the species or geographic area to train the model. Inductive models use known sites of presence or absence of the species or geographic area of interest and train the model on the ecological variables of these known sites to locate similar areas with a high probability of additional populations or similarity.

PMs are useful because they greatly reduce the amount of area that must be physically surveyed for either new populations or suitability for conservation, a trait that makes them invaluable to conservationists with limited funds, time and personnel. They are frequently used for single taxa such as mammals (Johnson & Gillingham 2004; Treves et al. 2004; Allen et al. 2001; Carroll et al. 2001), birds (Seoane et al. 2005; Larson et al. 2004), butterflies (Nally et al. 2003; Luoto et al. 2002, 2001) and even amphibians (Lipps 2005). They are occasionally used to predict single plant species but more often for predicting plant communities (Cohen & Goward

2004; Hong et al. 2004; Kupfer & Franklin 2000; Franklin 1995). Multi- and cross-taxa PMs are also used as the foundation of regional conservation plans (Armenteras et al. 2003; Root et al. 2003; Kautz & Cox 2001; Hoctor et al. 2000). Land managers seeking to preserve oak savanna complex in Lucas County have an area of almost 89,000 ha to search or at least 33,000 if the search area is limited to the Oak Openings Region. A PM that successfully locates oak savanna complex would prove invaluable for conservation and restoration efforts, especially if it could reduce the search area to less than 5%, or 4,500 ha.

No single model, regardless of its complexity, can accurately predict nature's form – physiography, species, processes – across both time and space, due to the individual heterogeneity and complexity of time and space as well as their combined complexity (Levins 1966). This is true for predictive models, especially those which try to accurately and precisely reflect the real world while attempting to make broad generalizations. Simpler models based on fewer variables are often more generalizable but unable to model the complex interactions found in nature. More complex models based on a higher number of variables may be able to model these interactions but have a greater chance of being restricted in their spatial or temporal applicability as well as higher risk of having accepted an incorrect model assumption (Green et al. 2005; Guisan & Zimmermann 2000). Therefore, when creating a model, the research objectives should be taken into account when deciding which two of three characteristics the model will maximize: realism; precision; or generalism. The resulting models will be, respectively, more empirical/phenomenological; analytical/theoretical or mechanistic/process-based. This can be viewed as a triangular continuum, wherein approaching two of the objectives requires sacrificing the third, as shown in Fig. 4 (Guisan & Zimmermann 2000; Levins 1966). As my PGM was descriptive and not causal, and seeks to model the existing ecosystem as realistically and precise as possible via a large dataset of training sites, it is more empirical and analytical as opposed to mechanistic.

Model variables represent one of three types of environmental gradients: resource; direct; or indirect (Vaughan & Ormerod 2003; Guisan & Zimmermann 2000; Franklin 1995). Resource gradients reflect the matter and energy directly consumed by the organism in question: water; light; nutrients (plants) or food (non-plants). Direct gradients reflect items that are not consumed but are physiologically important parameters such as pH or temperature such as pH and temperature. Indirect gradients are not physiologically relevant but are often correlated with species distributions and thought to be surrogate measures for singular or multiple resource or direct gradients, e.g., slope, aspect, topographic position, ecosystem type, or geology. The choice of gradient type depends on the model type, its spatial and temporal scales and its intention to account for disturbance and/or climate change. Resource and direct gradients are the most appropriate variables for generalistic-mechanistic models as they are the most ecologically- and biologically-relevant. However, they are the most difficult gradient type on which to obtain digital data. Indirect gradients are appropriate for non-mechanistic models, have the most abundant data resources, but their distance from the actual causal factors places constraints on the spatial and temporal transferability of the model (Vaughan & Ormerod 2003; Guisan & Zimmermann 2000; Franklin 1995). Because my model was descriptive, not causal, and is not designed to be transferred to areas or times outside the spatial or temporal extent of the training sites, the most appropriate choice was to use indirect gradients.

In addition to model type and variable type, PMs can be classified according to their logical approach. Deductive models start with *a priori* assumptions of the habitat preferences or requirements of the organism or area of interest – typically ecological variable requirements such as vegetation, moisture, temperature – and then finds areas that meet these criteria. Inductive models do not start with *a priori* assumptions. Rather, they start with sites of known presences and absences, determine the values and significance of various ecological variables at the different sites, and then finds areas that meet these criteria. Because they are based on fewer assumptions, inductive models have a lower potential for predictive error. Because of the subjectivity of current definitions and the lack of consensus on measurable ecological variables associated with oak savanna, plus the existence of a large data set of remnants available for model training sites, I chose to build an inductive model.

Many predictive models are created using Geographic Information Systems (GIS). GIS is routinely used for government and commercial planning and management but is increasingly being use for planning and management of natural areas and other natural resources (Cohen & Goward 2004; Nally et al. 2003; Clevenger et al. 2002; Luoto et al. 2002; Clark & Slusher 2000; Corsi et al. 2000; Breininger et al. 1998; Lathrop & Bognar 1998). GIS is well-adapted to processing the vast amounts of presence/absence and ecological variable data and can then present the resulting predictive models as easily understood maps (Chang 2004; Theobald 2003). My research will use a GIS-based predictive model to locate patches of oak savanna complex and prioritize them according to their probability of success for supporting conservation or restoration efforts. In summary, this was a descriptive, empirical and analytical PGM for oak savanna complex which maximized both the realistic and precise representation of the target ecosystem. It used inductive modeling techniques based on a large training data set and indirect environmental gradients as the ecological variables.

Research Objectives and Approach

My research goal was to minimize the uncertainty in land management decision-making by creating a predictive model that was both realistic and precise and was based on the fewest assumptions. To minimize assumptions and risk of predictive error, I built an inductive model that would still be able to generalize to the necessary spatial and temporal scale. Additionally, because this model described the current environment and did not attempt to explain the environment in terms of causal factors, it was an empirical model, not a mechanistic model.

Based on the information regarding the importance of glaciation, fire disturbance and ground water on oak savanna creation and maintenance, I hypothesized that a model using sandy soil types, higher elevation; moderate slope and canopy cover of 10-30% can predict oak savanna complex with a high probability. Although this research focused on the first three questions, long-term research will address all of the following questions:

1. "What are the significant ecological predictors of oak savanna complex?"
2. "Where are the areas that have the highest probability of presence of oak savanna?"
3. "How can we prioritize areas for expenditure of limited conservation resources?"
4. "How can we optimize conservation efforts by connecting existing protected areas?"

Conservation Implications

Currently, when conservationists, environmental planners and land managers plan conservation and restoration efforts, they are limited to personal field knowledge or non-quantifiable estimates of the location of remnant or potentially restorable areas (Johnson & Gillingham 2004). Although ecosystems and vegetation cover types are often mapped and occasionally even characterized, they have not been the focus of inductive predictive models

(Franklin 1995). My research is innovative because it applies two proven techniques in a new way by using predictive modeling and GIS to determine significantly predictive ecological variables and using them to build a predictive model for a rare ecosystem – oak savanna complex. Additionally, it broadens the scale of conservation efforts from a small region surrounding currently protected areas in a few clustered townships to the landscape level of Lucas County.

The methodology of this research is invaluable because it is simple and straightforward and can be replicated by land managers, planners and scientists alike. The method also produces testable hypotheses and generates statistically analyzable data. The model concept itself is valuable because it can be applied to other ecosystems simply by training the model using presence and absence sites of that ecosystem. Finally, its ability to handle new environmental or anthropogenic conditions such as land alteration, ownership changes or new threats, through the routine input of updated or improved data make it perfect for conservationists implementing adaptive management plans.

II. METHODOLOGY

General Approach

My objective was to use GIS to build an inductive PHM for oak savanna complex based on a data set of oak savanna and oak barrens remnants and significantly predictive ecological variables. I used ArcGIS 9.1 to process the oak savanna complex presence/absence data and the data on ecological variables and to produce my predictive habitat maps (ESRI 1999-2002). My access to a large data set of oak savanna training sites allowed me to use readily available statistical software, SAS 8.01, to perform robust tests to determine which ecological variables were significant predictors: suitable soils and elevation (SAS Institute Inc. 2000). My modeling methodology required:

1. collating readily available methods and data;
2. compiling a data set of functioning and degraded oak savanna and oak barrens remnants;
3. identifying the ecological variables that are significant predictors of the presence of high-quality and/or degraded oak savannas and related oak barrens (oak savanna complex), based on characteristics of known presence sites;
4. building a predictive geographic model (PGM) built on these known presence sites that locates areas with high probability for the presence of oak savanna complex;
5. evaluating the various candidate PGMs for relative accuracy using an reserved portion of the known presence sites; and
6. validating the final PGM by truthing a set of validation sites selected from its predicted area of high probability of oak savanna complex presence across the entire county.

The following sections describe in detail the focus area and study site selection; modeling process; and model conception, training, calibration, evaluation, and validation.

Focus Area and Study Site Selection

My focus area was the Oak Openings Region of Lucas County, part of a 193-km sand belt extending from Napoleon, Ohio, to Detroit, Michigan. In Gordon's 1966 map, "The Natural Vegetation of Ohio at the Time of the Earliest Land Surveys," the area was listed as "oak savannas," a subtype of mixed oak forests, surrounded by beech forests, elm-ash forests and prairie grasslands (Fig. 5). The original extent of the area was approximately 77,000 ha (Gordon 1969). Timbering, agricultural conversion and construction drainage took their toll on the natural landscape and reduced the region to its current extent of approximately 33,670 ha. This remnant was extensively surveyed by Edwin Moseley in his classic publication, "Flora of the Oak Openings" (Moseley 1928).

As stated previously, the region is actually a complex of globally and locally rare ecosystems. The upland sandy areas supported oak savanna, formally known as Black Oak/Lupine Barrens (G3/S1), oak barrens, formally known as Midwest Sand Barrens (G2/G3), Mesic Sand Tallgrass Prairie (G2), and Black Oak-White Oak/Blueberry Forest (G4) ecosystems (Noss et al. 2003; Stein et al. 1995; The Nature Conservancy 1997). Historically, the upland beach ridges and dunes supported 24,300 hectares of oak savanna and barrens, hereafter called "oak savanna complex," with an average tree density of 2.4 trees per hectare. I chose the Brewer & VanKat delineation of the Oak Openings Region over Moseley's delineation because it was based on land surveys of 1817-1832, placing its estimate closer to the time of presettlement. It also provided finer-scale ecosystem subcategories, and was available as a digital data layer easily imported into ArcGIS 9.1 (Brewer & Vankat 2004).

My research objective was to create a predictive geographic model (PGM) for oak savanna complex usable by land managers without extensive ecological knowledge using a methodology adaptable to identify and locate other ecosystems of interest. Therefore, the methodology itself had to be simple and had to use readily available data and technology. Because I did not find conclusive agreement in the relevant literature on the significant ecological variables that can predict oak savanna complex presence, I chose to use an inductive model which would generate the significant ecological variables from a data set of known presence sites (Weiher 2003; Leach & Givnish 1999; Nuzzo 1994; Haney & Apfelbaum 1993; Apfelbaum & Haney 1990; Curtis & McIntosh 1951). Because I did not have personal field knowledge of these sites, I used expert opinion to compile this data set but not to solicit any opinions regarding selection of ecological variables for the model. Recent research indicates that use of expert opinion for model variable selection actually decreases a model's predictive accuracy without improving transferability of the model to other areas (Seoane et al. 2005; Johnson & Gillingham 2004; Clevenger et al. 2002). For purposes of standardization, land managers categorized their areas according to the NatureServe classification scheme. Oak savannas and oak barrens most closely matched NatureServe's "Black Oak/Lupine Barrens" and "Midwest Sand Barrens," respectively (NatureServe 2005).

I consulted land managers from The Nature Conservancy (TNC) and the MetroParks of the Toledo Area (Gary Haase, pers. comm.; John Jaeger, pers. comm.; Tim Schetter, pers. comm.) and asked them to identify sites with oak savanna habitat. Paper maps provided by TNC were scanned and digitized into a non-topological vector file using ArcGIS 9.1 (ESRI 1999-2002). MetroParks provided its data as vector files created in ArcView 3.2 (ESRI 1992-1999) which were readily imported into ArcGIS 9.1. The land managers were asked to differentiate between functioning (high-quality) and degraded oak savannas. Functioning oak savannas were defined as those more closely meeting the NatureServe ecosystem criteria with a higher percentage of oak savanna-associated species and a physiography (topography and landforms) more closely resembling a functioning oak savanna. Degraded oak savannas were further from meeting the NatureServe criteria with fewer oak savanna-associated species and a less savanna-like physiography. Degraded sites were those that had not been the object of restoration efforts or had been only recently acquired. Only thirty-seven oak savanna areas were identified, totaling 241.09 hectares, as either functioning or degraded oak savanna (Table 1).

Table 1. Composition and characteristics of oak savanna and oak barrens study sites, categorized by quality and ecosystem subtype.

Ecosystem type	No. of sites	Area (ha)
functional oak savanna	15	71.81
functional oak barrens	17	36.35
degraded oak savanna	22	169.28
degraded oak barrens	8	24.96
total	62	302.40
mean		4.88
variance		44.88
range		0.10 – 33.84

Because the accuracy and utility of PGMs are directly related to the size of the model's training data set, I decided to increase the number of study sites by including the closely-related oak barrens sites. This larger training data set improved the model by allowing a greater number of model variables while minimizing over-fit of the model. Current research suggests that no more than $m/10$ variables be included in the final model, where m equals the number of observations (Vaughan & Ormerod 2003; Guisan & Zimmermann 2000; Harrell et al. 1996). The additional study sites also facilitated internal evaluation of the various candidate models to rate their relative predictive accuracy.

Increasing the number of training sites provided other benefits. First, it facilitated later exclusion of some of the smaller sites if their size or spatial configuration placed constraints on the analysis. Second, it provided flexibility to land managers who may require a larger selection of acquisition or restoration sites than afforded by the oak savanna data set. Finally, inclusion of the second ecosystem provided enough study sites to allow creation of individual “functioning” and “degraded” study site models for future analyses. An additional twenty-five oak barrens study sites were located, for an additional 61.31 hectares. The combined total was 62 sites encompassing 302.40 hectares and covering a range of sizes and configurations (locations are shown in Fig. 6). These sites formed the foundation of the PGM. Table 2 lists the various data sets used in the model and their method of generation. To evaluate the created models, I randomly divided the study sites into a training data set (75% subset) and an evaluation data set (25% subset), following current statistical methods (Seoane et al. 2005; Guisan & Zimmermann 2000). Training and evaluation site characteristics and locations are included in Appendix B., Tables 14 and 15. The accuracy of the final predictive model was validated against a data set of sites randomly selected from within the area of the highest probability of presence of oak savanna complex, i.e., model validation set.

Table 2. Characteristics of data sets of study, training, simulated, evaluation and validation sites. Number of sites; total and mean areas, and method of data set generation are indicated.

Data set	Sites	Total area (ha)	Mean area (ha)	Generation method
study sites (presence)	62	302.40	4.88	expert opinion
training (presence sites)	47	220.91	4.70	randomly selected from study site data set using Hawth's Tools extension in ArcGIS 9.1
simulated (absence sites)	47	229.36	4.88	randomly generated within Lucas County, outside of study site data set, using Hawth's Tools extension
model evaluation (internal)	15	81.49	10.19	remaining study sites
model validation (external)	10	point data	point data	randomly generated within area of highest probability using Hawth's Tools extension

Modeling Process

I followed the general four-step modeling process, shown in Fig. 7: (1) model conception/formulation; (2) model training/calibration; (3) model evaluation; and (4) model validation (Corsi et al. 2000; Franklin 1995). Model conception involved reviewing the literature for candidate ecological variables, consulting experts to build a data set of known oak savanna complex sites, generating a data set of simulated oak savanna complex absence sites to compare the training data set against, and collection of digital data on the candidate ecological variables. Model training involved determining the ecological variables that were statistically significant predictors of oak savanna complex by comparing the known presence sites to the absence sites, and building the various candidate predictive models. Model evaluation entailed comparing the predictive accuracy of the various candidate models by determining the following for each: the number of evaluation sites correctly predicted; the total evaluation data set area correctly predicted; and the amount of reduction in search area provided. Model validation was accomplished by generating a set of validation sites within the area of highest oak savanna complex probability predicted by the PHM, and then analyzing the immediate and adjacent land use and land cover of these sites using high-resolution aerial photography to determine their potential for oak savanna complex presence or restoration.

Model Conception

Digital data on soil type, elevation, slope and aspect were obtained or generated from various sources. Data availability, resolution and currency were comparable. All data were free to the public either online or upon submission of a research proposal (Table 3). Format of the data was either in GIS-importable non-topological vector files or raster data sets.

Model Training: data collection and processing

Soil type data was available as a raster data set from the Ohio Department of Natural Resources Geographic Information Management System (2005) and as a non-topological vector file from the Lucas County Auditor in its 2005 Auditor's Real Estate Information System (AREIS) update (Kaczala 2005). Both are based on a soil surveys conducted between 1973 and 1976 (Stone et al. 1980). I chose the vector file because it was more current and of a higher standard (national-level SSRGO versus state-level STATSGO) (Larry Kaczala, pers. comm.). I then converted it to a raster data set (Fig. 8) using the Spatial Analyst extension in ArcGIS 9.1 (ESRI 1999-2002). There were 62 different soil types in Lucas County, grouped into pedologically-relevant soil

Table 3. Metadata for ecological variable data sets. Information includes source of data, when the data was created (not posted), its spatial resolution and GIS format (vector or raster), where or how the data was obtained, and the site statistic related to the variable.

Ecological variable	Source	Date	Resolution & data type	Availability	Site statistic analyzed
soil type	Ohio Department of Natural Resources, Geographic Information Management Service	2002	30m, raster	http://www.dnr.state.oh.us/gims/	percent area of each soil type
	Lucas County Auditor	2003	30m, non-topological vector	on CD	
elevation	US Geologic Survey, Earth Resources and Observation Science	2003	30m, raster	http://seamless.usgs.gov/	mean elevation
slope	generated from elevation raster using Spatial Analyst extension	2003	30m, raster	self-generated; Spatial Analyst extension	mean slope
topographic position	generated from elevation and slope raster	2003	30m, raster	self-generated; TPI extension	
aspect	generated from elevation raster using Spatial Analyst extension	2003	30m, raster	self-generated; Spatial Analyst extension	

associations, Although the categories of soil type were considered nominal-level data, the statistical analysis was based on the area of each soil type contained in each of the model presence and absence sites. This ratio-level data allowed for a one-way nonparametric analysis of variance (ANOVA) using SAS 8.01 to determine which soil types were significant predictors of oak savanna complex presence.

Topographically-related ecological variables were based on a digital elevation model (DEM) obtained from the U.S. Geological Survey Earth Resources and Observation and Science (EROS) Center's online Seamless Data Distribution System (US Geological Survey 2005). The mean elevations for the training sites were compared to those of the simulated sites to determine if mean elevation was a significant predictor of oak savanna complex presence. From this DEM, I derived a slope data set and an aspect data set using the Spatial Analyst extension. Mean slopes of the training sites were compared to those of the simulated sites to determine if mean slope was significant predictor of oak savanna complex presence. The aspect raster data set, Fig. 11, included many flat areas with no orientation to the sun and non-flat areas whose orientation ranged from 0° to 360°. Because flat areas have no aspect, it was necessary to perform the analysis in two steps. First, the mean percent of flat areas of the sites was analyzed to determine if this characteristic was a significant predictor. Next, the mean aspect for the remaining non-flat areas was analyzed to determine if orientation to the sun of these areas was a significant predictor.

I converted the slope raster data set into a topographic position index (TPI) raster data set using the Topographic Position Index (TPI) extension (Jenness 2005) within ArcView 3.2 (ESRI 1992-1999). This extension first created a TPI raster data set that indicated each pixel's relative position on the slope within a user-defined neighborhood, ranging from ridges at the top to valleys at the bottom with slopes of various steepness in between. I used two different classification schemes – 4-class and 6-class – to determine whether a finer resolution of analysis affected the determination of significances of ecological variables. These schema are based on the TPI and slope values for each pixel (Table 4). The 6-class system is based on a relative TPI using standard deviations (SD) from the mean TPI value of the entire data set and a slope of 5% (Weiss 2001). The 4-class system is based on an absolute TPI value of 8 and a slope of 6% (Dickson & Beier 2002).

I also used two different sizes of analysis neighborhoods for TPI, either 100 m- and 200 m-radius, to determine the appropriate scale at which to analyze the sites (Turner et al. 2001). Increasing the neighborhood distance causes the analysis to look at each pixel at a broader scale: what may be a valley at a 100 m window may actually turn out to be a depression in a ridge top at a 200 m window (Fig. 12). Topographic position is a nominal-level datum, so I analyzed each of the twenty classes separately: two different neighborhoods of both the six-class system (twelve classes) and of the four-class system (eight classes). For each of the twenty classes, I compared the mean percent area of the training sites to those of the simulated sites. This ratio-level data allowed me to run a one-way nonparametric ANOVA to determine which topographic positions were significant predictors of oak savanna complex presence. The six-class, 100 m-neighborhood topographic position data set is shown in Fig. 13.

Model Calibration: data analysis

Each ecological variable was analyzed to determine its significance in predicting the presence or absence of oak savanna complex by comparing the training sites of known presence to the simulated sites of presumed absence (Table 2). The statistics analyzed were either the percentage of area of each site exhibiting a particular trait – a certain soil type or a certain topographic position – or the mean physical attribute of each site – mean elevation, mean slope, mean percent flat or non-flat area, mean orientation of non-flat area – following standard

Table 4: Coarse- and fine-scale topographic position classification systems based on relative or absolute Topographic Position Index (TPI) and slope criteria (Jenness 2005).

Classes	TPI	Slope
SIX CLASS (Weiss 2001)		
1. ridge	≤ 1 SD	-
2. upper slope	1 SD to 0.5 SD	-
3. middle slope	0.5 SD to -0.5 SD	≥ 5°
4. flat slope	0.5 SD to -0.5 SD	≤ 5°
5. lower slope	-0.5 SD to ≤ -1 SD	-
6. valley	> -1 SD	-
FOUR CLASS (Dickson & Beier 2002)		
1. ridgeline	≥ 8	-
2. steep slope	-8 to 8	≥ 6°
3. gentle slope	-8 to 8	< 6°
4. canyon bottom	≤ -8	-

predictive modeling statistical analysis (Franklin 1995; Guisan & Zimmermann 2000; Luoto et al. 2001; Nally et al. 2003). The Hawth's Tools extension randomly selected these simulated sites equal in number (47) and size (4.88 ha) to the training set (Beyer 2004). The simulated sites were at least 1.00 km apart and were set randomly within the county except for within a training or evaluation site. The Hawth's Tools extension's Thematic Raster Summary tool was used to derive this mean percent area data (Beyer 2004).

Most of the ecological variable data sets did not exhibit a normal distribution when normality was analyzed in SAS 8.01 (Shapiro-Wilks, $p < 0.05$). Therefore, all variables were analyzed using one-way nonparametric ANOVA in SAS 8.01. This test returns the same values as the Wilcoxon test, also known as the Mann-Whitney U test, which is the statistical test often used for analyzing significance of ecological variables (Treves et al. 2004; Nally et al. 2003;

Luoto et al. 2001). I accepted significance when there was a 95% probability that the difference in the means of the training and simulated sites was not due to random sampling ($\alpha = 0.05$). This corresponded to a p-value of less than 0.05 ($p < 0.05$). I used Satterthwaite's corrected p-values for unequal variances regardless of the equality of variances between the training and simulated data sets because it provides a more conservative p-value with less chance of committing a Type I error of commission (SAS Institute Inc. 2000). I determined the variable's relationship – direct or inverse – by determining whether the variable was present more or less often in the training sites than in the simulated sites. For the construction of the predictive models, I selected a suitable range for significant variables at a 95% confidence interval around the mean, consistent with my acceptance of a Type I error of 5% or less.

Model Evaluation: single- and multi-variable models

Each of the ecological variables determined to be a significant predictor of oak savanna complex presence was rated as to its predictive accuracy by conducting an internal evaluation using the set of evaluation sites (Table 2). Each variable's data set was reclassified to "1" as suitable or "0" as unsuitable. Suitable was considered to be the sets of individual values for soil type, topographic position or flat aspect that were significant predictors of oak savanna complex, or the 95% confidence interval of the mean elevation, slope or non-flat aspect orientation if they were significant predictors of oak savanna complex. I then examined each single-variable models' predicted suitable areas and determined the following using the Hawth's Tools Summary Statistics function: the number of evaluation sites predicted accurately by each model as oak savanna complex; the total percent area of the evaluation sites accurately predicted; and the reduction of the search area, or how much of the county was eliminated as unsuitable for oak savanna complex.

I then investigated whether or not combining individual variables into a single multi-variable model would provide greater predictive accuracy. I created a simple boolean model by combining the individual maps of the significant ecological variables using Spatial Analyst (Corsi et al. 2000; Guisan & Zimmermann 2000; Franklin 1995). Creating a boolean additive model provides those without the advanced statistical knowledge of regression analysis to quantitatively analyze multiple data sets. This created a new map with a data set that ranged in value from zero to the sum of the number of added maps. The value of any pixel in the map was its suitability index score for the probability of presence of oak savanna complex. A suitability index score of zero indicated that a particular area was considered unsuitable for each of the significant ecological variables criteria while a maximum score indicated that it met all of the criteria. Any intermediate score indicated it met some of the criteria for oak savanna complex presence without indicating which particular criteria it met. The areas for each particular index value were then considered as a multi-variable PGM and evaluated to determine the predictive accuracy provided by a PGM containing that number of ecological variables. The internal evaluation was conducted in the same fashion as it was for the single-variable PGMs: the number of evaluation sites predicted accurately as oak savanna complex; the total percent area accurately of the evaluation sites accurately predicted; and the reduction of the search area image, or how much of the county was eliminated as unsuitable for oak savanna complex.

Model Validation

I created an external validation site data set to estimate the predictive accuracy of the multi-variable PGM (Table 2). The multi-variable predictive model's raster file was converted to a non-topological vector file using Spatial Analyst. Hawth's Tools Generate Random Points function selected a set of ten random points within the area predicted to meet the maximum suitability index of 4 but outside the area of any training or evaluation sites (Beyer 2004). These ten points were then visually inspected using an orthorectified high-resolution (6”) aerial photograph taken in 2003 in ArcGIS 9.1 to determine their immediate and adjacent land use and land cover (Kaczala 2005). The higher the fraction of sites with those characteristics compatible with oak savanna complex conservation and restoration – rural residential, agricultural or forested land use; barrens to scrub to appropriate deciduous land cover – the greater the predictive accuracy and utility of the model.

III. RESULTS

Soil Type

A one-way nonparametric ANOVA showed that five soil types were significant predictors of oak savanna complex presence – OtB (Ottokee), Gr (Granby), OaB and OaC (Oakville) and TdA (Tedrow). Thirteen soil types were significant indicators of absence of oak savanna complex: DcA (Del Rey-Urban), Uo (Udorthents), So (Sloan), Ur (Urban), ByA (Bixler-Urban), Co (Colwood), NnA (Nappanee), HnA (Haskins), Ho (Hoytville), Mf (Mermill), To (Toledo), Lc (Latty), FuA (Fulton). Forty-four soil types were not significant (Table 5).

Table 5. Soil types significantly associated with oak savanna complex presence (one-way nonparametric ANOVA, $n = 94$, $p < 0.0001$)

Soil Type	Training site mean area (%)	Training site area variance	Simulated site mean area (%)	Simulated site variance
Ottokee (OtB)	35.85	910.01	6.75	336.79
Granby (Gr)	27.59	997.55	2.98	95.97
Oakville (OaB)	12.76	540.03	0.00	0.00
Tedrow (TdA)	10.99	424.25	1.20	37.41
Oakville (OaC)	6.81	227.70	0.00	0.00

Elevation

In Lucas County, elevation ranged from 170 m to 218 meters above sea level (masl). A one-way nonparametric ANOVA showed that the mean of the training site elevations (205 masl) was a significant indicator of oak savanna presence when compared to that of the simulated sites (191 masl) with a $p < 0.0001$ (Table 6). The range of suitable elevations was set conservatively at 202 – 209 masl.

Table 6. Significance of mean elevation (one-way nonparametric ANOVA, n = 94, p < 0.0001)

Elevation (masl)	Mean	Variance	CI lower	CI upper
training sites	205	4.13	201.41	209.53
simulated sites	191	160.80	165.90	216.625

Slope

In Lucas County, slope ranged between 0% to 28%. A one-way nonparametric ANOVA showed that the mean slope of the training data set (1.415%) was not a significant predictor of oak savanna complex presence ($p = 0.1206$) (Table 7).

Table 7. Significance of mean slope (one-way nonparametric ANOVA, n = 94, p = 0.1206)

Slope (%)	Mean	Variance	CI lower	CI upper
training sites	1.42	0.79	-0.37	3.20
simulated sites	1.32	1.99	-1.50	4.14

Topographic position

A one-way nonparametric ANOVA showed that none of the eight 4-class topographic positions was a significant predictor of oak savanna complex presence ($p > 0.50$) regardless of neighborhood size. Of the twelve 6-class topographic positions, only the ridge position was a significant predicator with a $p < 0.0001$ for both the 100 m- 200 m-radius neighborhood (Table 8). Conversely, flat slope was a significant predictor of oak savanna absence ($p = 0.0002$) for both the 100 m- and 200-m neighborhoods. Because of the ability to account for the effects of larger landform interactions, the suitable topographic position was considered the ridge class of the 6-class, 200 m-radius neighborhood.

Table 8 . Topographic positions associated with oak savanna complex presence (one-way nonparametric ANOVA, n = 94, p < 0.0001)

Topographic position (% area)		Mean	Variance
100 m, 6-class: ridge	training	16.43	315.04
	simulated	5.91	43.26
200 m, 6-class: ridge	training	28.39	780.91
	simulated	8.76	141.53

Aspect

Two qualities of aspect were analyzed – mean percent of flat areas and mean orientation of non-flat areas. A one-way nonparametric ANOVA showed that the mean percent of flat area of the training sites at 28.22% was significantly lower than that of the simulated sites (47.03%, p = 0.0053) and that flat areas were a significant predictor of oak savanna complex *absence.* Therefore, non-flat areas would be a significant predictor for oak savanna *presence*. The mean aspect orientation of the non-flat areas of the training sites of 172.22˚ was not a significant predictor of oak savanna complex when compared to that of the simulated site (157.19˚, p = 0.2527) (Table 9).

Table 9 . Significance of flat aspect (one-way nonparametric ANOVA, n = 94, p = 0.0053) and non-flat aspect orientation (one-way nonparametric ANOVA, n = 81, p = 0.2527)

Aspect		Mean	Variance
flat (% area)	training	28.22	1004.61
	simulated	47.03	1215.28
orientation (degrees)	training	172.23	3212.95
	simulated	157.19	3925.21

Single-variable Predictive Geographic Models

Maps of the predicted areas of high probability of presence of oak savanna complex were created for each of the significant ecological predictors: soil type, mean elevation, topographic position and aspect (Figs 14-17). The range of suitable values for each variable is shown in Table 10. The percent of suitable area was the amount of the county that met that variable's criteria. The single-variable models which provided the greatest search area reduction were topographic position at 8.59% suitable area, followed by soil type, elevation and aspect.

Table 10. Suitable areas predicted by single-variable PGMs

Characteristic	Suitable values	Suitable area (%)	:Unsuitable area (%)	Suitable county area (ha)
soil type	OtB, OaC, Gr, OaB, TdA	21.90	78.10	197.08
elevation	202-209m	23.33	76.67	209.96
topographic position	ridge, 200 m	8.59	91.41	77.35
aspect	not flat	49.50	50.50	445.46

Multi-variable Predictive Geographic Models

I was able to further reduce the search area by combining the four individual significant ecological variables into a single boolean overlay map. The values of the data set creating a suitability index which ranged from 0 to 4. A suitability index score of "0" indicated the 30 m-pixel met none of the four criteria and a score of "4" indicated it met all four of the criteria for predicting oak savanna presence. Areas meeting fewer criteria were assigned the respective number of criteria; all criteria were considered of equal importance. The one drawback to this model was that it did not directly show which criteria were met for any given pixel; one would have to refer back to the single-variable maps to determine whether a given variable was suitable or unsuitable. This complication was later rectified by performing regression analysis. This multi-variable PGM substantially reduced the search area in a range from 63.03% for one criterion up to 98.70% for all four criteria (Table 11). The combined map is shown in Figure 18.

Table 11. Suitable areas predicted by multi-variable PGMs using a suitability index

Suitability index score	Suitable area (%)	Unsuitable area (%)	Suitable county area (ha)
4	1.30	98.70	1120.95
3	5.82	94.18	5038.11
2	17.81	82.19	15,407.55
1	38.10	61.90	32,962.95
0	36.97	63.03	31,987.89

Internal Model Evaluation

All of the single- and multi-variable models predicted the location of evaluation sites very well, i.e., 87 – 100% accuracy, except for the 1-criterion multi-variable model (Table 12). Evaluation area predicted correctly ranged from 100% for the 3-criteria multi-variable model and 93.1% for the soil type model, down to 35.9% for the topographic position model and 4% for the 1-criterion multi-variable model. Search area reduction ranged from 98.7% for the 4-criteria multi-variable model and 91.4% for the topographic position model, down to 61.9% for the 1-criterion model and 50.5% for the non-flat aspect model.

To qualitatively determine which models provided the most utility, I ranked each model on a scale of 0-3 for three evaluation criteria: number of evaluation sites predicted correctly; amount of evaluation site area predicted correctly; and search area reduction. Those models fulfilling more criteria better had higher utility scores. Based on this analysis, the soil type model was the most accurate and most useful single-variable PGM because it located all of the evaluation sites, had a 93.1% correct evaluation site-area prediction and 78.1% search area reduction. The second best single-variable PGM was the elevation model. The best multi-variable PGM was the 3-criteria model because it located all of the evaluation sites, had a 100% correct evaluation site-area prediction and 94.2% search area reduction.

Table 12. Internal evaluation of single- and multi-variable PGMs

Model	Sites predicted (# / %)	Site area predicted (%)	Search area reduction (%)	Utility (0 – 9)
single-variable:				
soil	15 / 100%	93.10	78.10	9 = 3+3+3
elevation	13 / 87%	74.46	76.67	7 = 2+2+3
topographic position	15 / 100%	35.88	91.41	6 = 2+2+3
aspect	15 / 100%	81.45	50.50	6 = 3+2+1
multi-variable:				
4 criteria	13 / 87%	19.63	98.70	5 = 2+0+3
3 criteria	15 / 100%	49.41	94.18	7 = 3+2+3
2 criteria	13 / 87%	27.51	82.19	5 = 2+1+2
1 criterion	5 / 33%	3.44	61.90	3 = 1+0+2

External Model Validation

I examined the immediate and adjacent land use and land cover of the ten validation sites predicted by the 4-criteria multi-variable PGM. Two sites were near residential development with large areas of agricultural or woods nearby that may still permit savanna restoration at a small scale. Otherwise, the sites had land use (agricultural, forested, rural residential) and land cover (agricultural, grassland barrens, scrub-shrub, woodland) compatible with oak savanna acquisition and conservation. Due to two of the sites posing some savanna restoration complications, the 4-criteria model's validation was estimated at 90% predictive accuracy.

Regression Analysis

The boolean overlay map allowed analysis of multiple data sets to determine how many criteria any one pixel met without showing *which* criteria were met. Regression analysis highlighted *which* ecological variables explained the most variance in the presence or absence of the dependent variable, e.g. presence of oak savanna complex. I performed a backward-stepwise logistic regression on the significant ecological predictors – suitable soil types, mean elevation, topographic position, mean area of non-flat aspect. The most useful model was chosen based on four factors: an acceptable amount of variance explanation (max-rescaled r^2 approaching 1.00); the entire equation being a significant predictor (χ^2-value, likelihood ratio, $p > 0.05$); each of the constituent variables being a significant predictor ($p < 0.05$); and an acceptable goodness-of-fit (Hosmer & Lemeshow test, minimal χ^2-value, p approaching 1.00) (N. Boudreau, pers. comm.). This data is shown in Table 13; statistics not meeting criteria are shown with an asterisk. The regression analysis confirms that the "best model" is a two-criteria model using soil and elevation. The related regression equation (Eq. 1) was:

Eq.1 $y = -50.2339 + 0.0656*(\text{suitable soil area}) + 0.2280*(\text{mean elevation})$

Table 13 . Backward-stepwise logistic regression analysis of significant ecological variables. Models indicate which significant ecological variables are included, max rescaled r^2 indicates model's explanation of variance, $\chi 2$ indicates significance of model as predictor of oak savanna complex.

Model	Max-rescaled r^2	χ^2, df p-value	Constituent p-values	Goodness-of-fit df
Soil-Elevation-Topographic position-Aspect	0.8836	102.15 df = 4 p < 0.0001	soil, p < 0.0001 elev, p = 0.0086 topo, p = 0.5244* asp, p = 0.2440*	p = 0.0979* df = 5
Soil-Elevation-Aspect	0.8812	101.66 df =3 p < 0.0001	soil, p < 0.0001 elev, p = 0.0115*	p = 0.0872* df = 6
Soil-Elevation	0.8612	97.58 df = 2 p < 0.0001	soil, p < 0.0001 elev, p = 0.0167	p = 0.5086 df = 7

IV. DISCUSSION

This research successfully produced a simple, but statistically robust, model that accurately predicted current and potential oak savanna habitat in Lucas County, Ohio. First and foremost, I quantitatively verified four groups of ecological variables that both significantly predicted presence of oak savanna complex and substantially reduced the search area of the county: five soil types; mean elevation; ridge topographic position; and non-flat aspect. Second, I assembled a collection of accurate, current, readily available digital data on various ecological variables. Third, I created a layer of remnant high-quality and degraded oak savannas and oak barrens large enough to accommodate robust statistical analysis. This training site data will be invaluable for future research and land management decisions. The split-sample use of the study site data set into training and evaluation sites allowed me to internally evaluate my candidate single- and multi-variable models. The data set was large enough that, even after retaining 25% of the sites as an evaluation set, it permitted up to five ecological variables for the multi-variable PGM which allowed for maximal model complexity with minimal model over-fit (Vaughan & Ormerod 2003; Guisan & Zimmermann 2000; Harrell et al. 1996).

The single-variable PGMs accurately predicted 87-100% of the evaluation sites and 36-93% of the evaluation site area. The multi-variable PGMs had variable site predictive accuracy of 53-100% and were less predictive of the total site area with an accuracy of 5% (4 criteria) to 83% (2 criteria). Additionally, the models substantially reduced search area for oak savanna complex presence and were very accurate in predicting potential sites of current or restorable oak savanna complex in the validation data set. The “suitable soil model” had the highest utility because it reduced the search area substantially (78.10%) while predicting 100% of the evaluation sites and 93.10% of the evaluation site area. The two-criteria model, e.g., any two of the four criteria were met, had the highest utility of the multi-variable models because it reduced the search area by 80% while predicting all of the evaluation sites and 82% of the their area. Regression analysis confirmed that a two-criteria model, which included suitable soils and elevation, explained the most variance in the oak savanna complex presence data set. External validation estimated this multi-variable model's predictive accuracy at 90% because nine of the ten validation sites had land cover and land use compatible with oak savanna complex acquisition and conservation (agricultural, barrens, scrub or forested, rural residential land use, and appropriate grassland or woody plant cover).

Finally, this model creation and calibration methodology are easily adapted to develop a prediction for a different ecosystem within the same landscape, or readily updated with new or more accurate information. To create a new model, e.g., for wet prairies of the Oak Openings Region, the major step would to locate a training data set of existing high-quality and/or degraded wet prairies. This data set could then be used with the same ecological variable data sets to determine variables that were significant predictors of the new ecosystem of interest. Conversely, if a newer soil type data set or a new parcel division data set became available, one would simply replace the old data set with the new one and rerun the analysis. This approach is quantitative, flexible, and statistically robust.

Strengths and limitations of predictive models

The accuracy of predictive models depends on the availability of an adequate amount of precise, accurate data. A small sample size of observations or locations will decrease both the types and power of statistical analyses that can be performed as well as the predictive power of the resulting model (Vaughan & Ormerod 2003; Corsi et al. 2000; Guisan & Zimmermann 2000; Franklin 1995). Although it is possible to generate a useful PM with a small data set, such as is likely for a rare or cryptic species, it will likely have a lower predictive accuracy and may require substantial evaluation and validation. For example, a PHM for green salamanders (*Aneides aeneus*) was built for a three-county area in southern Ohio from only seven training sites. Even with this small data set, the model reduced the search area for this rare and cryptic species by 99.30% and helped located five new records for the species (Lipps 2005).

For multi-variable PMs, the size of the data set also determines the number of variables that can be included in the model. A ratio of 10:1 of the number of observations to the number of model variables maintains an acceptable amount of error while maximizing the model complexity and minimizing model over-fitting (Vaughan & Ormerod 2003; Harrell et al. 1996). The size of the data set available for this PHM allowed incorporation of up to five ecological variables into the model while retaining enough sites for an independent evaluation data set.

In addition to sample size, accuracy and precision of the ecological data from which the model is built and calibrated determine how well the model's output will predict areas similar that same environment. This research had access to a large data set of sites to both train and evaluate the model. A smaller data set would have reduced the number of candidate predictive variables. Digital data was not readily available for all of the ecological variables. Some data had to be specially requested while other data was not available at all. Canopy-cover or available-light data – a characteristic commonly accepted as significant to oak savanna complex presence – was not available north of the southern tier of Ohio counties (Weiher 2003; Leach & Givnish 1999). Another unavailable layer was sand-strata thickness. Unlike soil type, which only indicates ground materials up to 1.5m below the ground's surface, the sand strata in areas of the Oak Openings Region is up to 30 m thick and may significantly influence both ecological processes and vegetative communities (Forsyth 1959).

A second facet of predictive model accuracy is the extent of their transferability to areas outside the extent of their training and calibration data set. Although properly-built mechanistic models may be transferred to larger spatial or temporal areas than empirical or analytical models, they are still limited to areas where the causal actions and interactions of the model hold true. If the model is transferred to an area where the causal factors are different then the model's accuracy decreases (Seoane et al. 2005; Guisan & Zimmermann 2000; Forsyth 1970). As the predictions were constrained to the same region as the original training data set, this model does not risk errors due to inappropriate application of the model.

The final limitation of predictive models, and a limitation of GIS models in general, is the resolution of available data. All of the available digital data had a standard 30 m resolution. Because this research was conducted at a landscape level and sought to predict large areas of ecological variables which do not change abruptly over a small distance across the landscape, this limitation did not negatively impact my study. It cannot be ruled out that finer-resolution data could provide a better predictor of oak savanna presence. For example, LIDAR (LIght Detection And Ranging) data has a horizontal resolution of 30-50 cm and a vertical resolution of 8-10 cm, compared to the standard DEM horizontal and vertical resolutions of 30 m and 10 m. However, LIDAR data is currently very costly and not available for all areas. A land manager would have to weigh the opportunity costs of acquiring this data against the use of these funds toward land acquisition, conservation or additional research.

Recommended refinements to methodology

There are few foreseeable methodological refinements. First, the validation process should be ground-truthed to determine whether the 2003 high-resolution aerial photography accurately reflects the validation sites' land use and land cover. Unlike most PGMs, this model may not substantially benefit from a larger data set as it already produced a very useful four-variable model from the available training data set. Preliminary regression analysis showed that the best-fit model was actually composed of only two significant predictive variables – suitable soils and mean elevation – which the current data set was more than adequate to support. This research confirms previous studies that soil type and elevation are ecological variables that are important either to savanna maintenance or to maintaining high biodiversity (Nally et al. 2003; Leach & Givnish 1999). Additional ecologically relevant characteristics – canopy-cover/light-availability, bedrock topography, depth to water, depth of glacial till, depth to bedrock, spatial configuration/connectivity of site, and means and seasonality of both temperature and precipitation – may provide greater predictive power to the model. However, the cost to obtain data that is not currently available would likely outweigh additional predictive benefits.

Additional knowledge may be gained by separating the training data set into its high-quality and degraded categories and into the oak savanna and oak barrens categories. Creation of models for these more refined categories may provide vital information for land managers that wish to identify specific habitat quality levels or ecosystems, either for data collection or conservation purposes, or identify sites that are good candidates for ecological restoration. Separating the different quality sites would also allow a test of the assumption that the aggregation of sites of differing habitat quality and or ecosystem type produces a model that accurately predicts the union of both of these model components. If validated, this may aid in the creation and use of predictive models for ecosystems with few available training sites.

Implications for conservation

Efficient and effective conservation depends on proper planning and expenditure of money, effort and other resources. It is critical to the short-term implementation success and long-term sustainability of natural areas that conservation resources be focused on areas that are most conducive to the desired ecosystem. This research has two major positive benefits for conservation. First, my analysis determined which ecological variables are significant to the presence of oak savanna complex, e.g., suitable soils and appropriate elevation. Not only can this guide acquisition efforts to areas more likely to result in a successful oak savanna complex, but it could also guide restoration efforts in recreating aspects of the environment in areas that may have differentiated from an oak savanna complex successional trend due to anthropogenic or other external factors such as fire suppression, hydrological alteration, woody plant succession or non-native species encroachment. Second, the suitability index and map indicate areas that currently, or in the recent past, possess(ed) all, or some, of the these significant ecological variables. This ranked index will help managers to prioritize their conservation and restoration efforts by focusing on areas with a higher oak savanna complex suitability score, which, due to their historic and current similarity to oak savanna complex will presumably require reduced effort require to move them towards a higher-quality oak savanna complex status.

Future research

Areas of future research fall into two broad categories: landscape ecology; and gap analysis and reserve design. Landscape ecology focuses on three fundamental concepts: scale; heterogeneity; and stochasticity (Turner 2005; Turner et al. 2001; Akcakaya 2000; Burke 2000; With et al. 1997; Schumaker 1996). In addition to simple ecological process – ecosystem presence relationships, it would be useful to determine the relationships between ecosystem quality, patch size and functional or physical connectivity. Understanding the relative importance of patch size versus connectivity to individual patch quality would help guide conservation decisions regarding acquisition and management, such as the SLOSS question

(Single-Large Or Several-Small) – "should we purchase a few large parcels far apart or several small parcels close together?" I would like to analyze the sizes and spatial configuration of my oak savanna complex study sites to determine any relationships between the quality of the sites with either the their size or their proximity to each other. This could be accomplished using the various tools in ArcGIS 9.1's Spatial Analyst. I will also consider investigating the transferability of the model to the other areas of the Oak Openings Region including Fulton and Henry County and southern Michigan. It may also be useful to determine how it performs in the savannas outside the Oak Openings Region in northeast and south-central Ohio.

Gap analysis is the process of comparing the total area of a species' habitat or ecosystem of interest to the areas that are currently protected where the unprotected areas are the "gaps" in the network (Root et al. 2003; Allen et al. 2001; Jennings 2000). Reserve design uses this knowledge as a starting point and designs a system of protected areas (Cowling et al. 2003; Kautz & Cox 2001; Akcakaya 2000; Clark & Slusher 2000; Hoctor et al. 2000). These may be as simple as a set of protected single- or multiple-use areas close enough that the species or ecological process of interest can disperse or function. They may also be complex networks with strictly-protected core areas, multiple-use buffer areas and physical or functional corridor-linkages that actually promote and facilitate organism/process dispersal and gene flow. In all of this, locating the ecosystem or species of interest is the first step. I plan to create a data set of natural areas currently protected in the county, including oak savannas and oak barrens. Combining this with the data set of high probability areas of oak savanna complex produced by this current research will reveal the gaps of remnant or restorable areas that should be targeted for acquisition or restoration. I will also collect landscape-level data on current anthropogenic variables such as land use, current and proposed infrastructure (roads, water and sewer lines), and zoning restrictions to determine current and future development pressures in the county. By combining these data sets, I can create a data set of anthropogenic "costs" to oak savanna acquisition or restoration. Using least-cost path algorithms, GIS can determine which areas to acquire to connect currently protected areas via corridors of low development pressure.

Conclusions

The use of GIS models to increase the efficiency and accuracy of locating species of interest is well-accepted. This research expands the current range of predictive models from taxa-oriented – plant, mammal, bird, insect – to a greater level of scale: ecosystems or landforms. With a high degree of accuracy, the newly developed, simple, quantitative, model (with suitable soils and elevation) predicted areas where oak savannas are, or could potentially be, on the landscape of Northwest Ohio. This GIS model utilized standardized concepts, freely available digital data and straightforward statistical analysis to create a tool that can be utilized by both land managers and academics, and will prove invaluable to conservation efforts to a globally rare oak savanna ecosystem.

V. BIBLIOGRAPHY

Akcakaya, H. R. 2000. Conservation and management for multiple species: Integrating field research and modeling into management decisions. Environmental Management 26:S75-S83.

Allen, C. R., L. G. Pearlstine, and W. M. Kitchens. 2001. Modeling viable mammal populations in gap analyses. Biological Conservation 99:135-144.

Apfelbaum, S. I., and A. Haney. 1990. Structure and dynamics of midwest oak savannas. Pages 20-30 in J. M. Sweeney, editor. Management of dynamic ecosystems: Proceedings of a symposium; 1989 December 5. North Central Section, The Wildlife Society, Springfield, IL.

Armenteras, D., F. Gast, and H. Villareal. 2003. Andean forest fragmentation and the representativeness of protected natural areas in the eastern Andes, Colombia. Biological Conservation 113:245-256.

Beyer, H. L. 2004. Hawth's Analysis Tools for ArcGIS. http://www.spatialecology.com/htools. accessed: December 1, 2005.

Brady, N. C., and R. R. Weil 1999. The nature and properties of soils. Prentice Hall, Upper Saddle River, NJ.

Breininger, D. R., V. L. Larson, B. W. Duncan, and R. B. Smith. 1998. Linking habitat suitability to demographic success in Florida scrub-jays. Wildlife Society Bulletin 26:118-128.

Brewer, L. G., and J. L. Vankat. 2004. Description of vegetation of the Oak Openings of northwestern Ohio at the time of Euro-American settlement. Ohio Journal of Science 104:76-85.

Burke, V. J. 2000. Landscape ecology and species conservation. Landscape Ecology. 15:1-3.

Carroll, C., R. F. Noss, and P. C. Paquet. 2001. Carnivores as focal species for conservation planning in the rocky mountain region. Ecological Applications 11:961-980.

Chang, K. 2004. Introduction to Geographic Information Systems. McGraw-Hill, New York, NY.

Clark, F. S., and R. B. Slusher. 2000. Using spatial analysis to drive reserve design: A case study of a national wildlife refuge in Indiana and Illinois (USA). Landscape Ecology 15:75-84.

Clevenger, A. P., J. Wierzchowski, B. Chruszcz, and K. Gunson. 2002. GIS-generated, expert-based models for identifying wildlife habitat linkages and planning mitigation passages. Conservation Biology 16:503-514.

Cohen, W. B., and S. N. Goward. 2004. Landsat's role in ecological applications of remote sensing. Bioscience 54:535-545.

Conner, R., A. Seidl, L. VanTassell, and N. Wilkins. 2001. United States Grasslands and Related Resources: An economic and biological trends assessment. Texas A&M University. https://landinfo.tamu.edu/presentations/grasslands.cfm. accessed: March 26, 2005.

Corsi, F., J. D. Leeuw, and A. Skidmore. 2000. Modeling Species Distribution with GIS. Pages 389-434 in L. Boitani, and T. K. Fuller, editors. Research Techniques in Animal Ecology: Controversies and Consequences. Columbia University Press, New York, NY.

Cowling, R. M., R. L. Pressey, M. Rouget, and A. T. Lombard. 2003. A conservation plan for a global biodiversity hotspot: The Cape Floristic Region, South Africa. Biological Conservation 112:191-216.

Curtis, J. T. 1959. The Vegetation of Wisconsin: An Ordination of Plant Communities. The University of Wisconsin Press., Madison, WI.

Curtis, J. T., and R. P. McIntosh. 1951. An Upland Forest Continuum in the Prairie-Forest Border Region of Wisconsin. Ecology 32:476-496.

Dickson, B. G., and P. Beier. 2002. Home-range and habitat selection by adult cougars in southern California. Journal of Wildlife Management 66:1235-1245.

Dunevitz, H. 1993. Characterization of Midwestern Oak Savannas. Pages 1-5 in F. Stearns, and K. Holland, editors. 1993 Midwest Oak Savanna Conference. US Environmental Protection Agency, Chicago, IL.

ESRI. 1992-1999. ArcView GIS. Environmental Systems Research Institute, Redlands, California.

ESRI. 1999-2002. ArcGIS. Environmental Systems Research Institute, Redlands, California.

Forsyth, J. L. 1959. "The Beach Ridges of Northern Ohio." Pages 1-10. Ohio Division of Geological Survey, Information Circular, Columbus, OH.

Forsyth, J. L. 1968. "A Summary of Toledo's Geologic Setting." Report prepared for the Toledo Regional Area Plan for Action, Regional report 8.2, Toledo, Ohio. Pages printed from report, 7-16 of 111. Pages 304-309 in R. L. Stuckey, editor. Linking Ohio Geology and Botany: Papers by Jane L. Forsyth, 2003. RLS Creations, Columbus, OH.

Forsyth, J. L. 1970. "A Geologist Looks at the Natural Vegetation Map of Ohio." Ohio Journal of Science. May, 70(3):180-191. Pages 147-159 in R. L. Stuckey, editor. Linking Ohio Geology and Botany: Papers by Jane L. Forsyth, 2003. RLS Creations, Columbus, OH.

Forsyth, J. L. 1998. "Ohio's Speedy Glacier." The Compass (Sigma Gamma Epsilon). Fall, 73(1):22-25. Pages 37-40 in R. L. Stuckey, editor. Linking Ohio Geology and Botany: Papers by Jane L. Forsyth, 2003. RLS Creations, Columbus, OH.

Franklin, J. 1995. Predictive vegetation mapping. Progress in Physical Geography 19:474-499.

Gordon, R. B. 1969. Natural Vegetation of Ohio in Pioneer Days. Bulletin of the Ohio Biological Survey, New Series III:2: iv-107.

Green, J. L., A. Hastings, P. Arzberger, F. J. Ayala, K. L. Cottingham, K. Cuddington, F. Davis, J. A. Dunne, M.-J. Fortin, L. Gerber, and M. Neubert. 2005. Complexity in ecology and conservation: Mathematical, statistical, and computational challenges. Bioscience 55:501-510.

Green Ribbon Initiative. 2004. The Nature Conservancy designates the Oak Openings Region One of America's "Last Great Places". http://www.oakopen.org/news/display.asp?id=209. accessed: June 1, 2005.

Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135:147-186.

Haney, A., and S. I. Apfelbaum. 1993. Characterization of Midwestern Oak Savannas. Pages 1-9 in F. Stearns, and K. Holland, editors. Midwest Oak Savanna Conference. US Environmental Protection Agency, Chicago, IL.

Harrell, F. E., Jr., K. L. Lee, and D. B. Mark. 1996. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in Medicine 15:361-387.

Hehr, D. W. 1970. A comparative study of the composition of the pre-settlement vegetation and the characteristic geologic substrate of the Oak Openings and surrounding area in northwestern Ohio. Page 90. Biological Sciences. Bowling Green State University, Bowling Green, Ohio.

Hoctor, T. S., M. H. Carr, and P. D. Zwick. 2000. Identifying a linked reserve system using a regional landscape approach: The Florida Ecological Network. Conservation Biology 14:984-1000.

Hong, S.-K., S. Kim, K.-H. Cho, J.-E. Kim, S. Kang, and D. Lee. 2004. Ecotope mapping for landscape ecological assessment of habitat and ecosystem. Ecological Research 19:131-139.

Jenness, J. 2005. Topographic Position Index (tpi_jen.avx) extension for ArcView 3.x. Jenness Enterprises. http://www.jennessent.com/arcview/tpi.htm. accessed: November 1, 2005.

Jennings, M. D. 2000. Gap analysis: concepts, methods, and recent results. Landscape Ecology 15:5-20.

Johnson, C. J., and M. P. Gillingham. 2004. Mapping uncertainty: sensitivity of wildlife habitat ratings to expert opinion. Journal of Applied Ecology 41:1032-1041.

Kaczala, L. A. 2005. Auditor's Real Estate Information System. Lucas County Auditor, Toledo, Ohio.

Kautz, R. S., and J. A. Cox. 2001. Strategic Habitats for Biodiversity Conservation in Florida. Conservation Biology 15:55-77.

Kupfer, J. A., and S. B. Franklin. 2000. Evaluation of an ecological land type classification system, Natchez Trace State Forest, western Tennessee, USA. Landscape and Urban Planning 49:179-190.

Larson, M. A., F. R. T. III, J. J. Millspaugh, W. D. Dijak, and S. R. Shifley. 2004. Linking population viability, habitat suitability, and landscape simulation models for conservation planning. Ecological Modelling 180:103-118.

Lathrop, J., Richard G., and J. A. Bognar. 1998. Applying GIS and landscape ecological principles to evaluate land conservation alternatives. Landscape and Urban Planning 41:27-41.

Leach, M. K., and T. J. Givnish. 1999. Gradients in the Composition, Structure, and Diversity of Remnant Oak Savannas in Southern Wisconsin. Ecological Monographs 69:353-374.

Levins, R. 1966. The strategy of model building in population ecology. American Scientist 54:421–431.

Lipps, G. J., Jr. 2005. A Framework for Predicting the Occurrence of Rare Amphibians: a Case Study with the Green Salamander. Pages: 75. Biological Sciences. Bowling Green State University, Bowling Green, OH.

Luoto, M., M. Kuussaari, H. Rita, J. Salminen, and T. von Bonsdorff. 2001. Determinants of distribution and abundance in the clouded apollo butterfly: A landscape ecological approach. Ecography 24:601-617.

Luoto, M., M. Kuussaari, and T. Toivonen. 2002. Modelling butterfly distribution based on remote sensing data. Journal of Biogeography 29:1027-1037.

MacDougall, A. S. 2005. Responses of diversity and invasibility to burning in a northern oak savanna. Ecology (Washington D C) 86:3354-3363.

Mayfield, H. 1976. The Changes in the Natural History of the Toledo Regions Since the Coming of the White Man. reprinted from the Jack Pine Warbler, Vol. 40:2 and The Northwest Ohio Quarterly, Vol. 34.

Meffe, G. K., and C. R. Carroll 1997. Principles of Conservation Biology. Sinauer Associates, Inc., Sunderland, MA.

Meisel, J., N. Trushenski, and E. Weiher. 2002. A gradient analysis of oak savanna community composition in western Wisconsin. Journal of the Torrey Botanical Society 129:115-124.

Moseley, E. L. 1928. Flora of the Oak Openings. The Ohio Academy of Science VIII:Part 3 Special Paper No. 20.

Nally, R. M., E. Fleishman, J. P. Fay, and D. D. Murphy. 2003. Modeling butterfly species richness using mesoscale environmental variables: model construction and validation for mountain ranges in the Great Basin of western North America. Biological Conservation 110:21-31.

NatureServe. 2005. NatureServe Explorer: An online encyclopedia of life [web application]. NatureServe, Arlington, Virginia. http://www.natureserve.org/explorer. accessed: January 6, 2006.

Noss, R. F., E. T. L. III, and J. M. Scott. 2003. Endangered Ecosystems of the United States: A Preliminary Assessment of Loss and Degradation. USGS Biological Resources. http://biology.usgs.gov/pubs/ecosys. accessed: May 9, 2005.

Nuzzo, V. 1986. Extent and status of Midwest oak savanna: presettlement and 1985. Natural Areas Journal 6:6-36.

Nuzzo, V. 1994. Extent and status of Midwest oak savanna: presettlement and 1985. Page 22. North American Conference on Savannas and Barrens.

Nuzzo, V. A. 1985. Extent and status of Midwest oak savanna: presettlement and 1985. Page 104. Landscape Architecture. University of Wisconsin - Madison, Madison, WI.

Ohio Department of Natural Resources 2005. Geographic Information Management Systems. David Crecelius. http://www.dnr.state.oh.us/gims/. accessed: March 1, 2005.

Ohio Department of Natural Resources Division of Natural Areas and Preserves. 2004. Natural Heritage Database, Kitty Todd Preserve. http://www.ohiodnr.com/dnap/location/kittytodd.htm. accessed: October 22, 2005.

Ohio Department of Natural Resources Division of Natural Areas and Preserves. 2005. Heritage Database. http://www.dnr.ohio.gov/dnap/heritage/default.htm. accessed: June 1, 2005.

Packard, S., and C. F. Mutel, editors. 1997. The Tallgrass Restoration Handbook: for prairies, savannas, and woodlands. Island Press, Washington, DC.

Root, K. V., H. R. Akcakaya, and L. Ginzburg. 2003. A multispecies approach to ecological valuation and conservation. Conservation Biology 17:196-206.

SAS Institute Inc. 2000. SAS System for Windows. SAS Institute Inc., Cary, NC.

Schumaker, N. H. 1996. Using landscape indices to predict habitat connectivity. Ecology (Washington D C) 77:1210-1225.

Sears, P. B. 1926. The natural vegetation of Ohio. II. The Prairies. Ohio Journal of Science 26:128-146.

Seoane, J., J. Bustamante, and R. Diaz-Delgado. 2005. Effect of expert opinion on the predictive ability of environmental models of bird distribution. Conservation Biology 19:512-522.

Stein, B. A., L. L. Master, L. E. Morse, L. S. Kutner, and M. Morrison. 1995. "Status of U.S. Species: Setting Conservation Priorities" in "Our Living Resources - A Report to the Nation on the Distribution, Abundance, and Health of U.S. Plants, Animals, and Ecosystems." U.S. Department of the Interior, National Biological Service. http://biology.usgs.gov/status_trends/reptspubs.html. accessed: December 2, 2005.

Stone, K. L., E. H. McConoughey, and G. D. Bottrell. 1980. Soil Survey of Lucas County, Ohio. Page 187. USDA Soil Conservation Service & ODNR Division of Lands and Soil & Ohio Agricultural Research and Development Center.

The Nature Conservancy. 1997. Oak Openings Site Conservation Plan. Page 34, Dublin, OH.

The Nature Conservancy. 2000. Site Management Plan - Kitty Todd Preserve. Page 10, Dublin, OH.

Theobald, D. M. 2003. GIS concepts and ArcGIS methods. Conservation Planning Technologies, Fort Collins, Colo.

Transeau, E. N. 1935. The Prairie Peninsula. Ecology 16:423-437.

Treves, A., L. Naughton-Treves, E. K. Harper, D. J. Mladenoff, R. A. Rose, T. A. Sickley, and A. P. Wydeven. 2004. Predicting human-carnivore conflict: a spatial model derived from 25 years of data on wolf predation on livestock. Conservation Biology 18:114-125.

Turner, M. G. 2005. Landscape ecology in North America: Past, present, and future. Ecology (Washington D C) 86:1967-1974.

Turner, M. G., R. H. Gardner, and R. V. O'Neill 2001. Landscape Ecology in Theory and Practice: Pattern and Process. Springer-Verlag, New York, NY.

US Geological Survey. 2005. Earth Resources Observation and Science (EROS) Data Center Seamless Data Distribution Systems. US Geological Survey, Sioux Falls, South Dakota. http://seamless.usgs.gov/. accessed: September 1, 2005.

Vaughan, I. P., and S. J. Ormerod. 2003. Improving the quality of distribution models for conservation by addressing shortcomings in the field collection of training data. Conservation Biology 17:1601-1611.

Weiher, E. 2003. Species richness along multiple gradients: Testing a general multivariate model in oak savannas. Oikos 101:311-316.

Weiss, A. 2001. Topographic Position and Landforms Analysis [poster]. ESRI User Conference, San Diego, California.

With, K. A., R. H. Gardner, and M. G. Turner. 1997. Landscape connectivity and population distributions in heterogeneous environments. Oikos 78:151-169.

VI. APPENDIX A: FIGURES

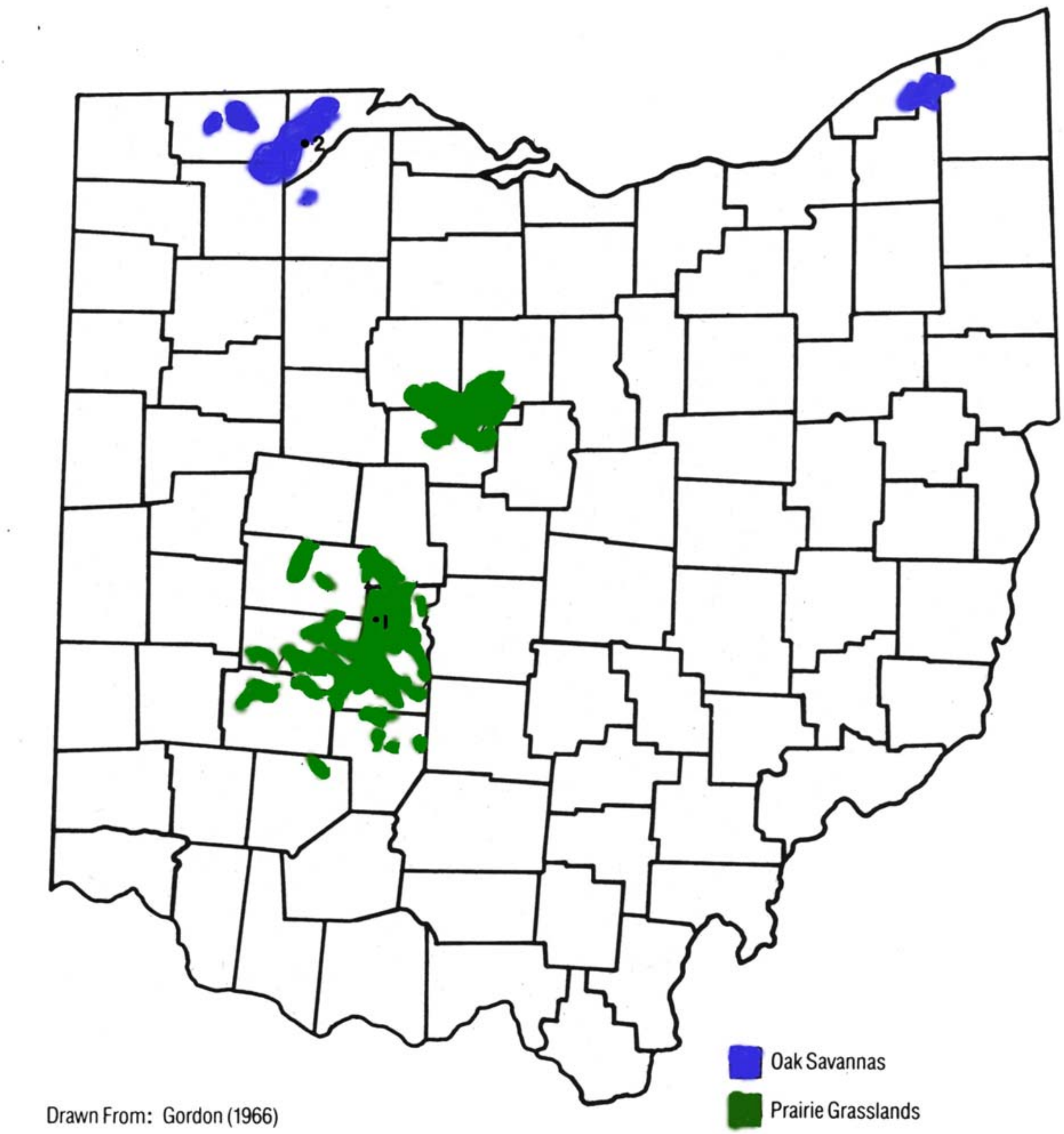

Figure 1. Pre-settlement extent and remnants of oak savanna and prairies in Ohio. From Gordon (1966) in Nuzzo (1986). Blue areas are oak savannas and green areas are prairies. Remnants are located in (1) the Oak Openings Region and (2) Union County.

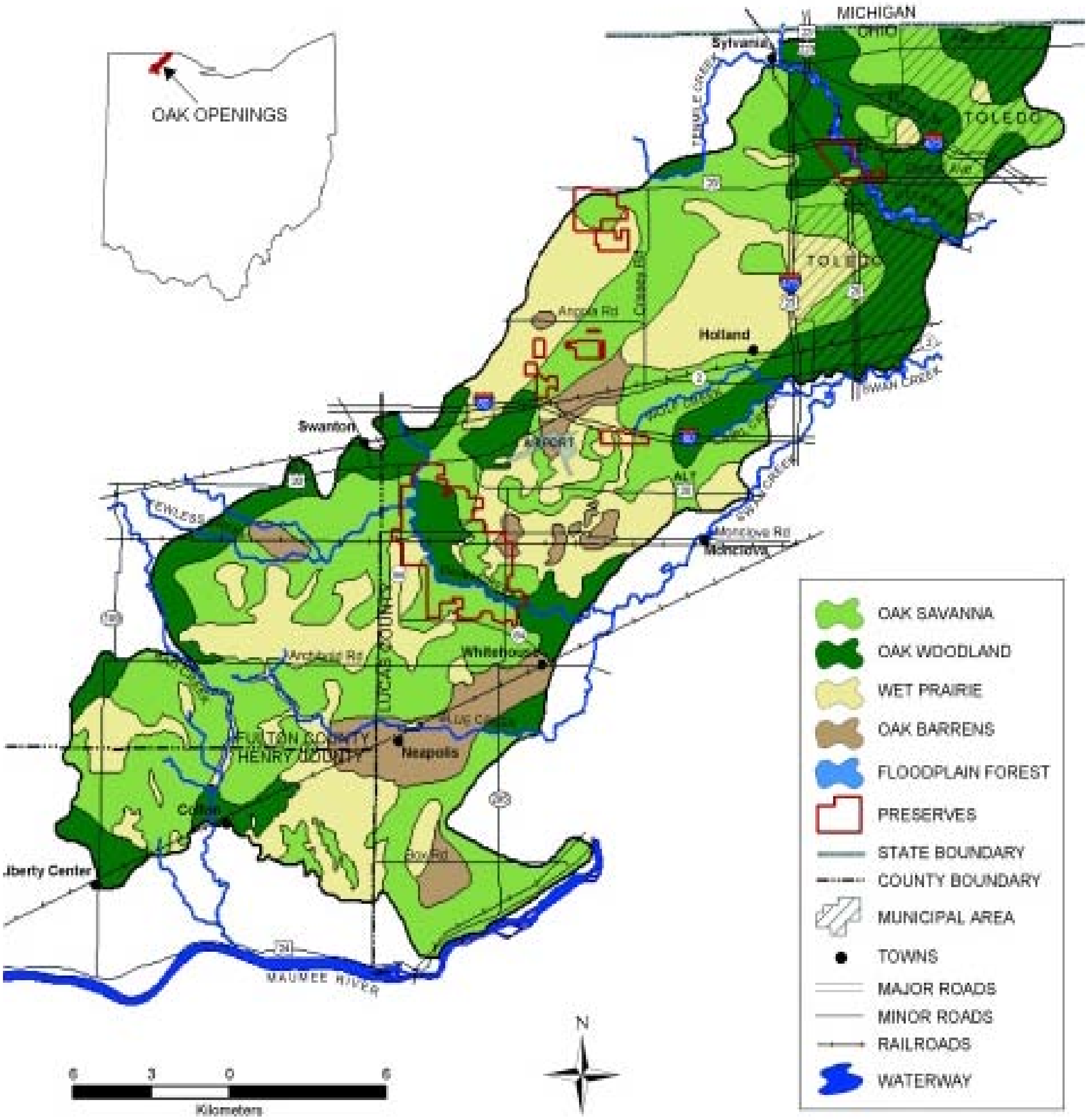

Figure 2. Vegetation of the Oak Openings at the Time of Euro-American Settlement (Brewer & Vankat 2004). This delineation was based on a compilation of General Land Office Surveys of 1785 which showed the dominant vegetation communities during 1817-1832.

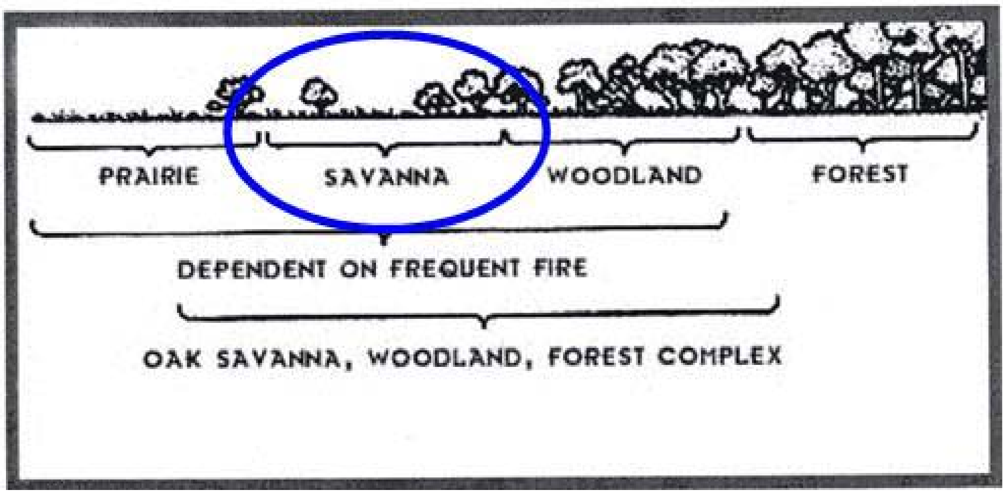

Figure. 3. Savanna as ecotone between prairie and woodland. After (Packard & Mutel 1997). Savanna has lower tree density than oak woodland but more than prairie, and is dependent on fire to maintain this structure and composition.

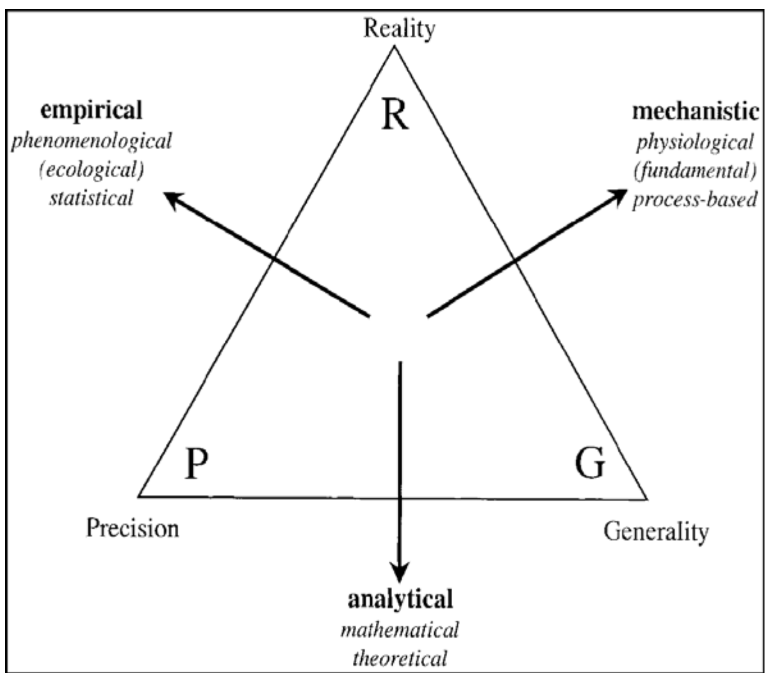

Figure. 4. A classification of models based on their intrinsic properties. After Levins). This diagram illustrates the three general properties of predictive models – realism, precision and accuracy – and the resulting model types: empirical; mechanistic; analytical. Increasing any two properties requires sacrificing the third.

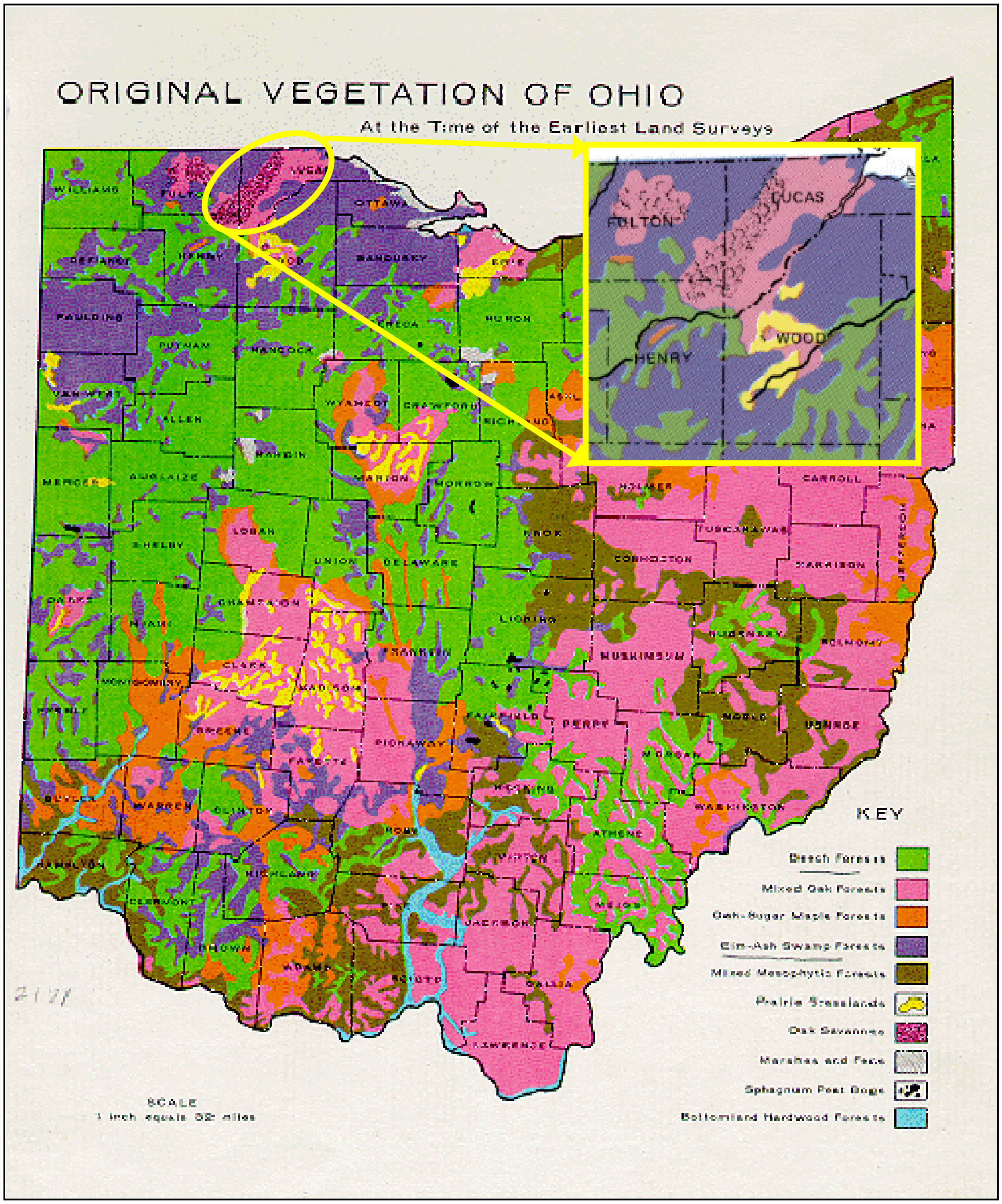

Figure 5. Location of study area: Oak Openings Region of Lucas County, Ohio. After Gordon (1966). This map shows the study area as “oak savannas” (pink w/stippling) within a larger complex of “mixed oak forests” (pink), “beech forests” (green) and “elm-ash swap forests” (purple) with “prairie grasslands” (yellow) to the south.

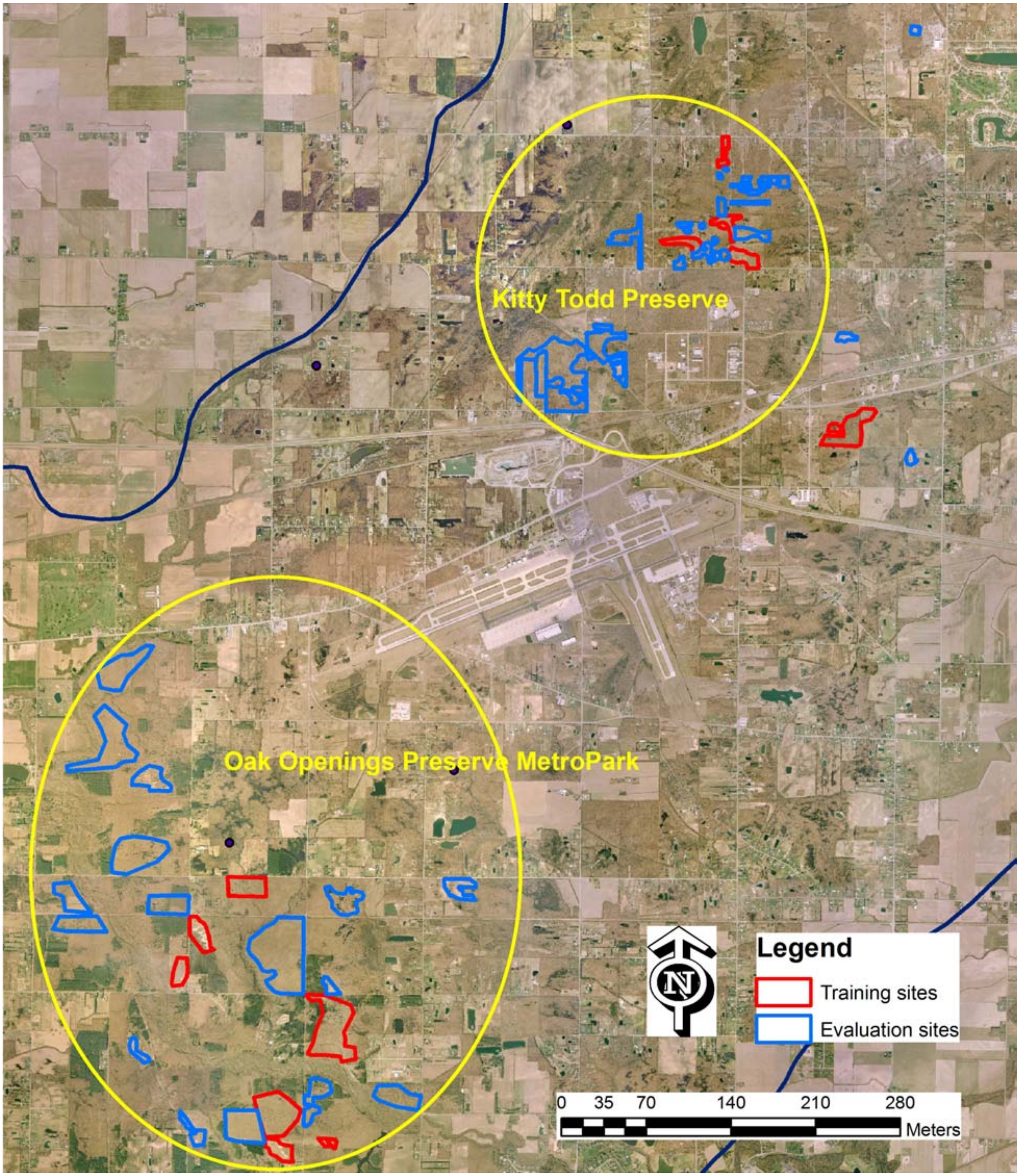

Figure 6. Study Site Locations. Training sites (47) are shown in red; evaluation sites (15) are shown in blue. The large complex between the two preserves is Lucas County Airport. The roads are Ohio State Route 2 and the I-80/I-90 toll road.

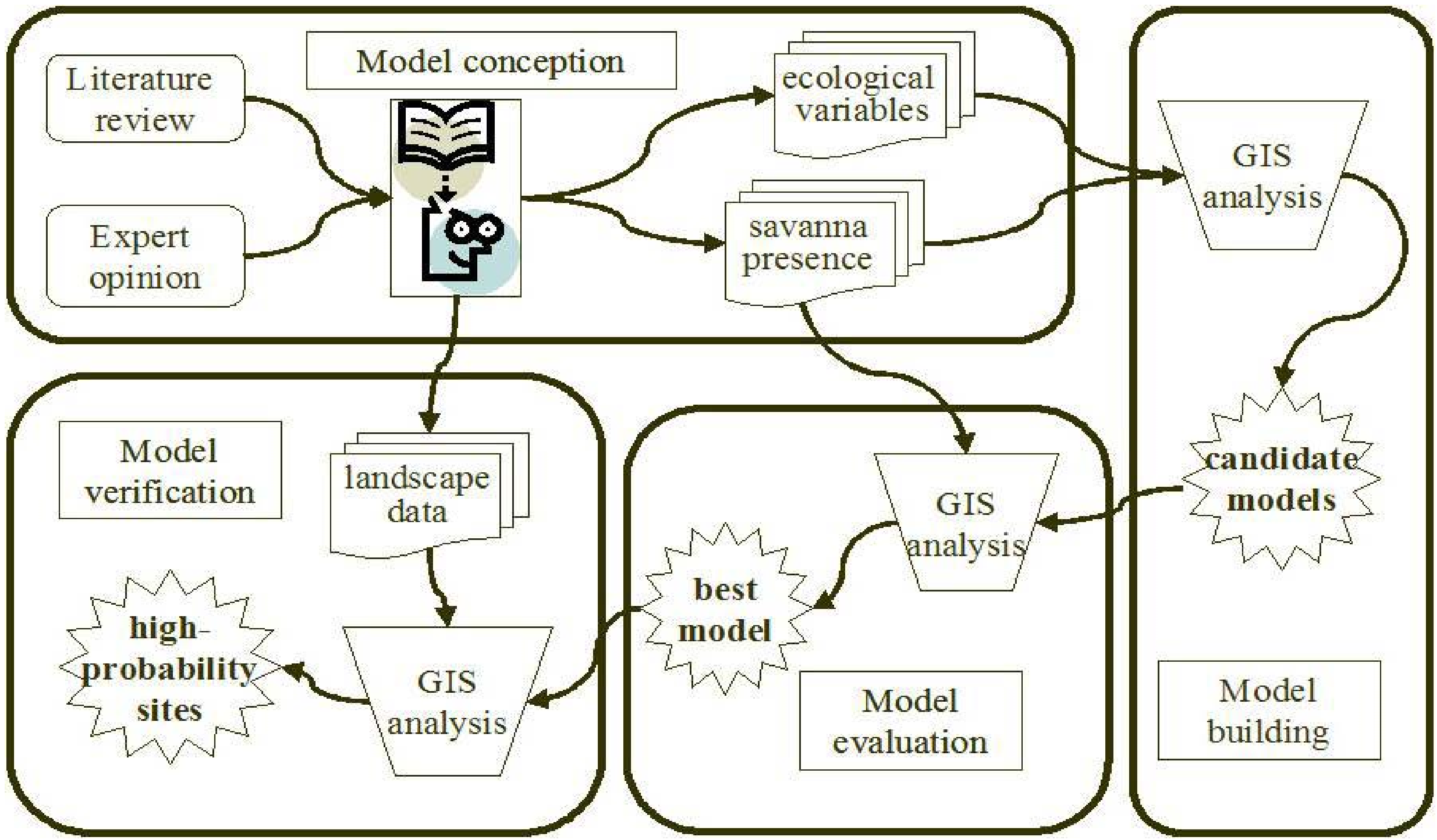

Figure 7. Oak Savanna Predictive Modeling Process. Follows the general four-step process outlined by Corsi, et al. (2000) and Guisan & Zimmerman (2000): model conception; model building/calibration; model evaluation; and model validation.

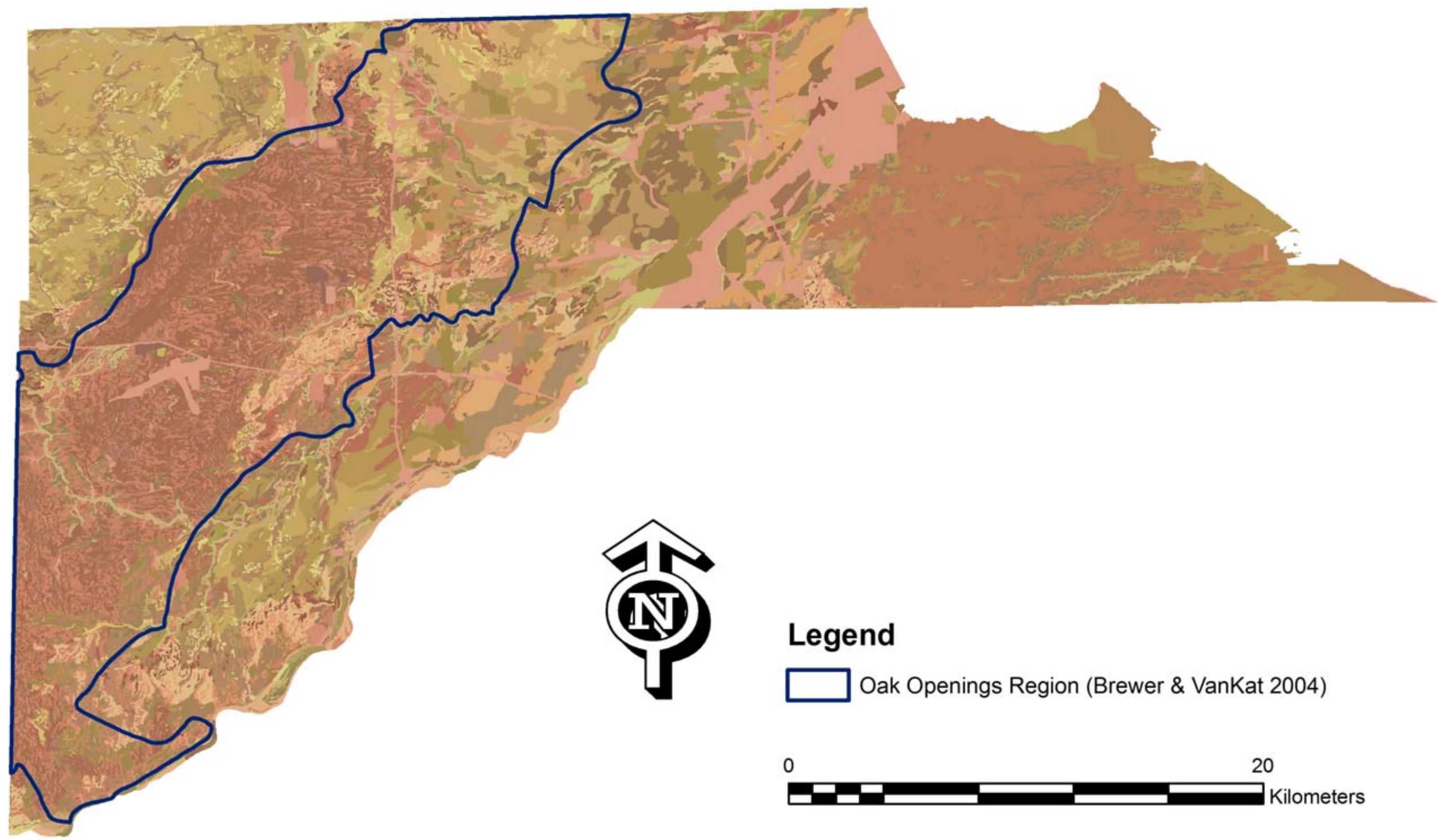

Figure 8. Soil Types of Lucas County, Ohio. After Stone (1980). Rasterized non-topological vector data, 62 unique soil types, 30 m resolution, available from Lucas County Auditor. Based on 1966 field soil surveys.

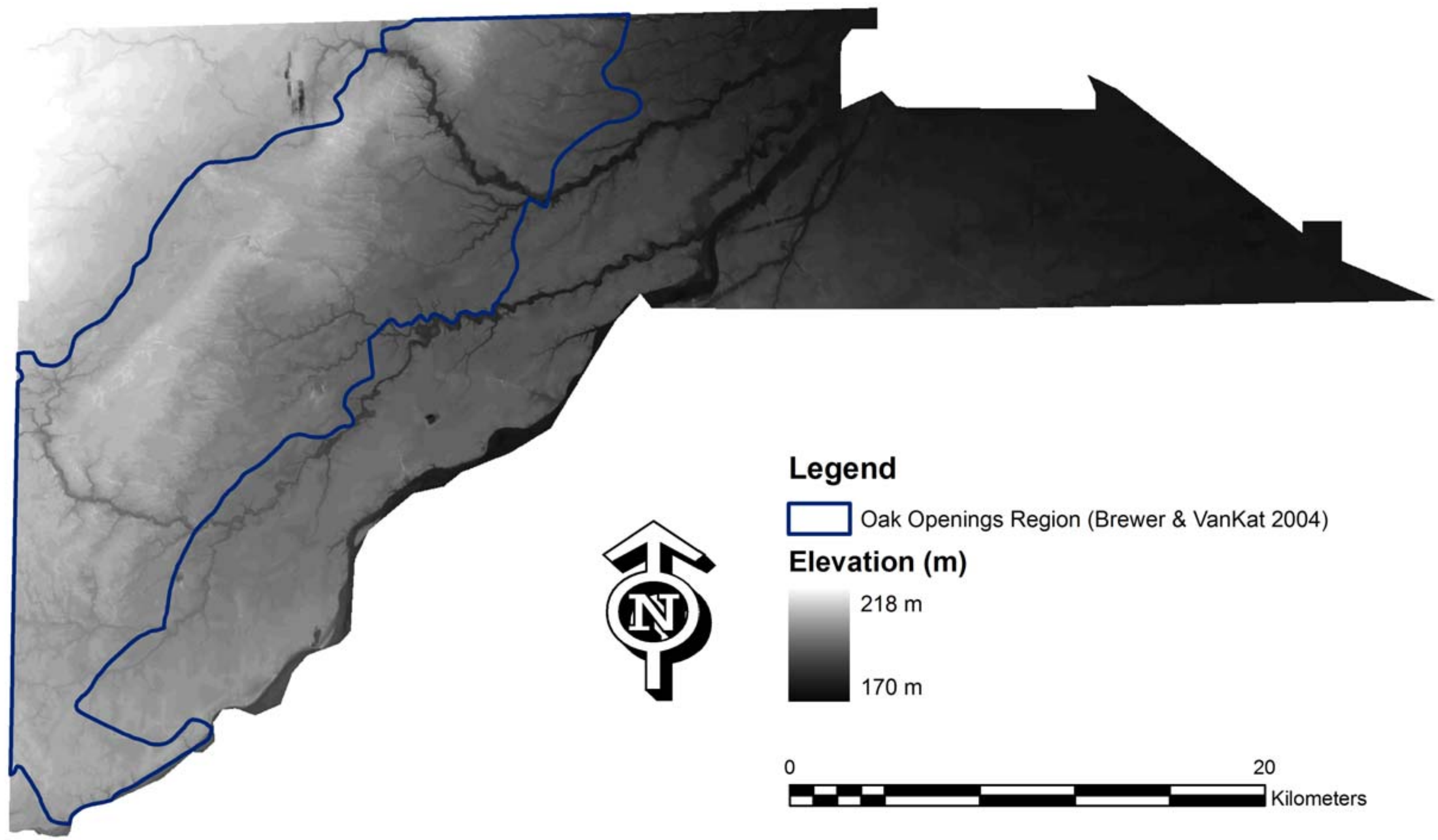

Figure 9. Elevations (m) of Lucas County, Ohio. Raster data from 30 m DEM, available from USGS Seamless Data Base. Elevations ranged from 170-218 masl.

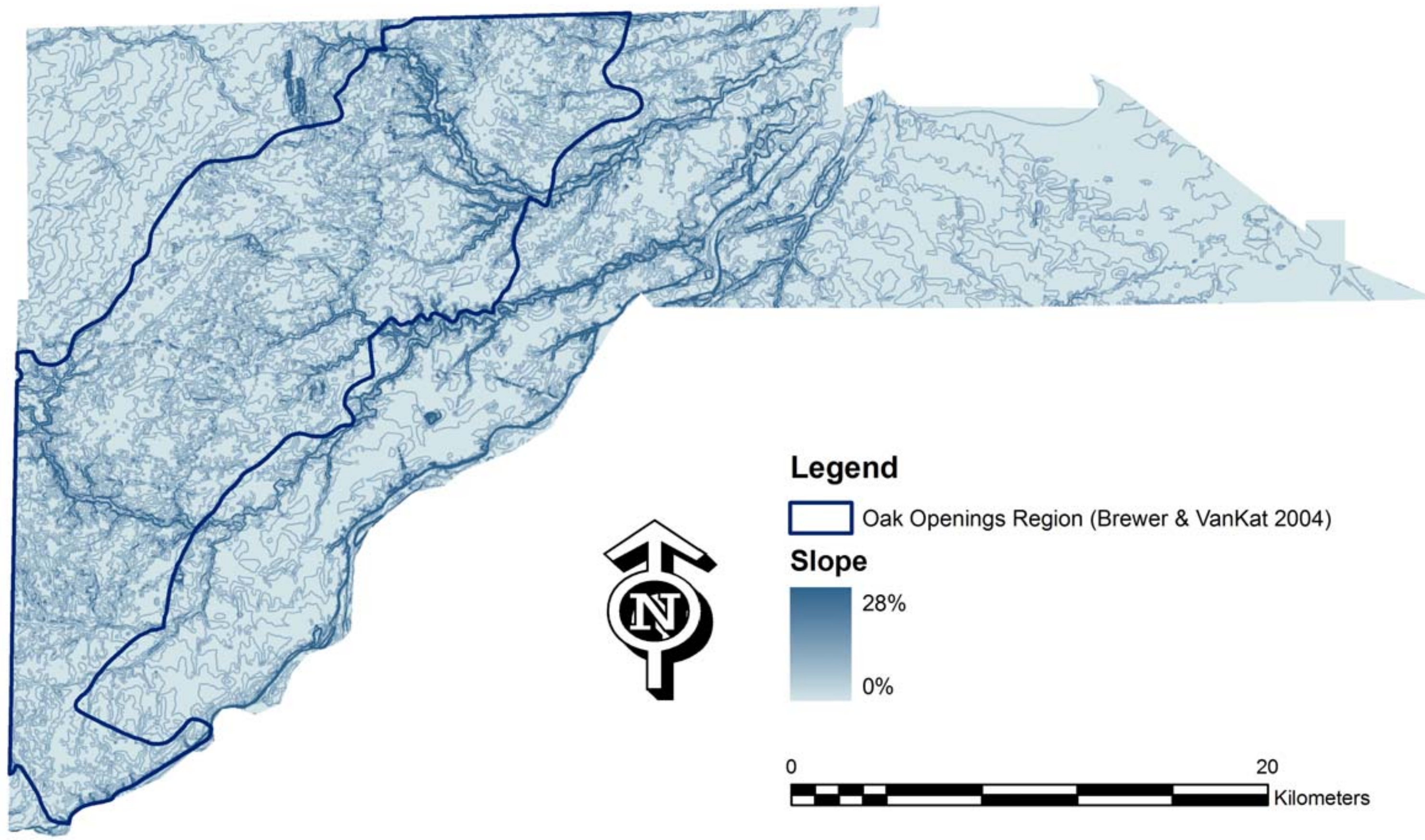

Figure 10. Slopes (%) of Lucas County, Ohio. Raster data, 30m , derived from elevation data using Spatial Analyst. Slopes ranged from 0% to 28%

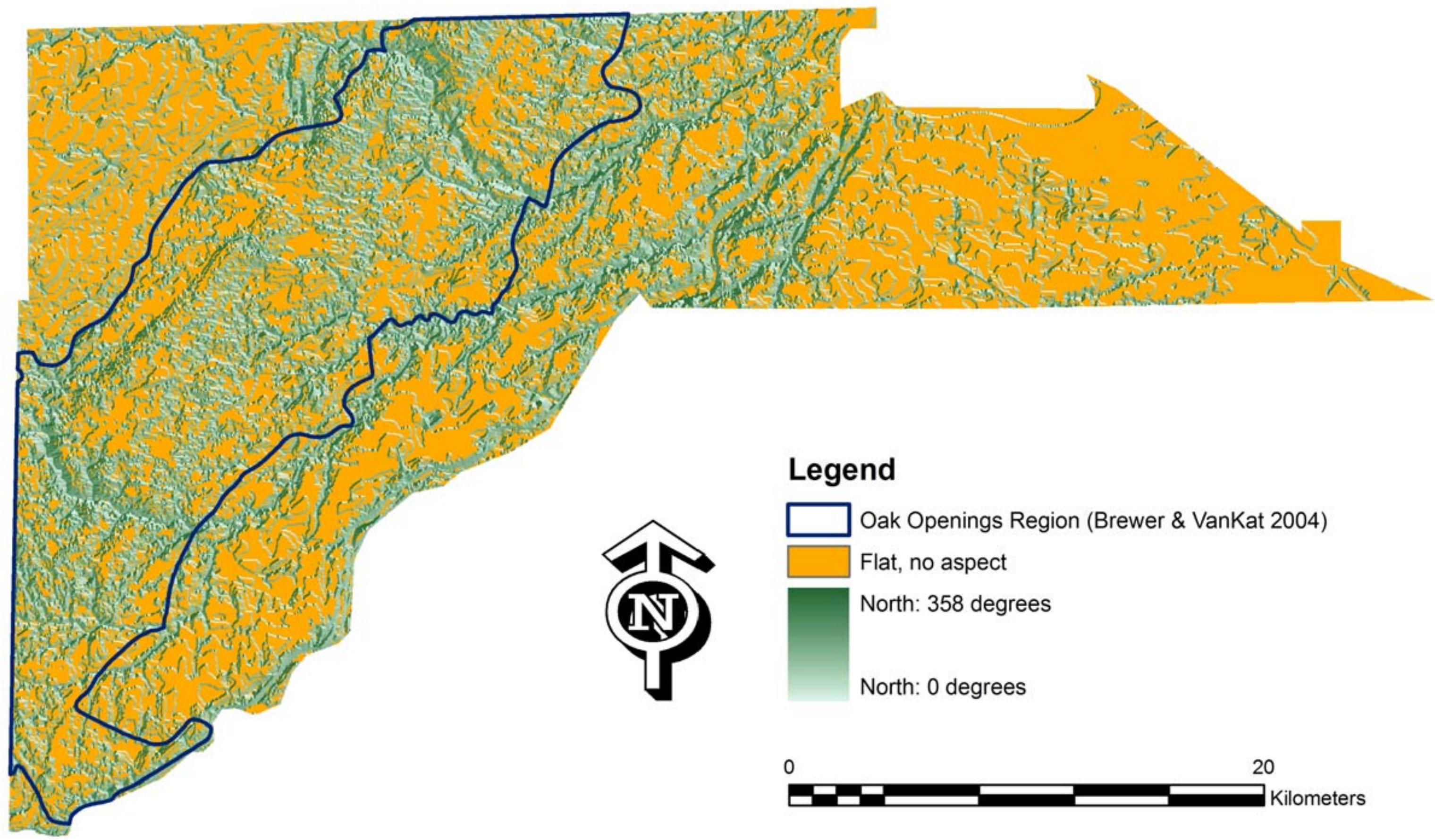

Figure 11. Flat aspect and non-flat aspect orientation of Lucas County, Ohio. Raster data, 30m, derived from elevation and slope data using Spatial Analyst.

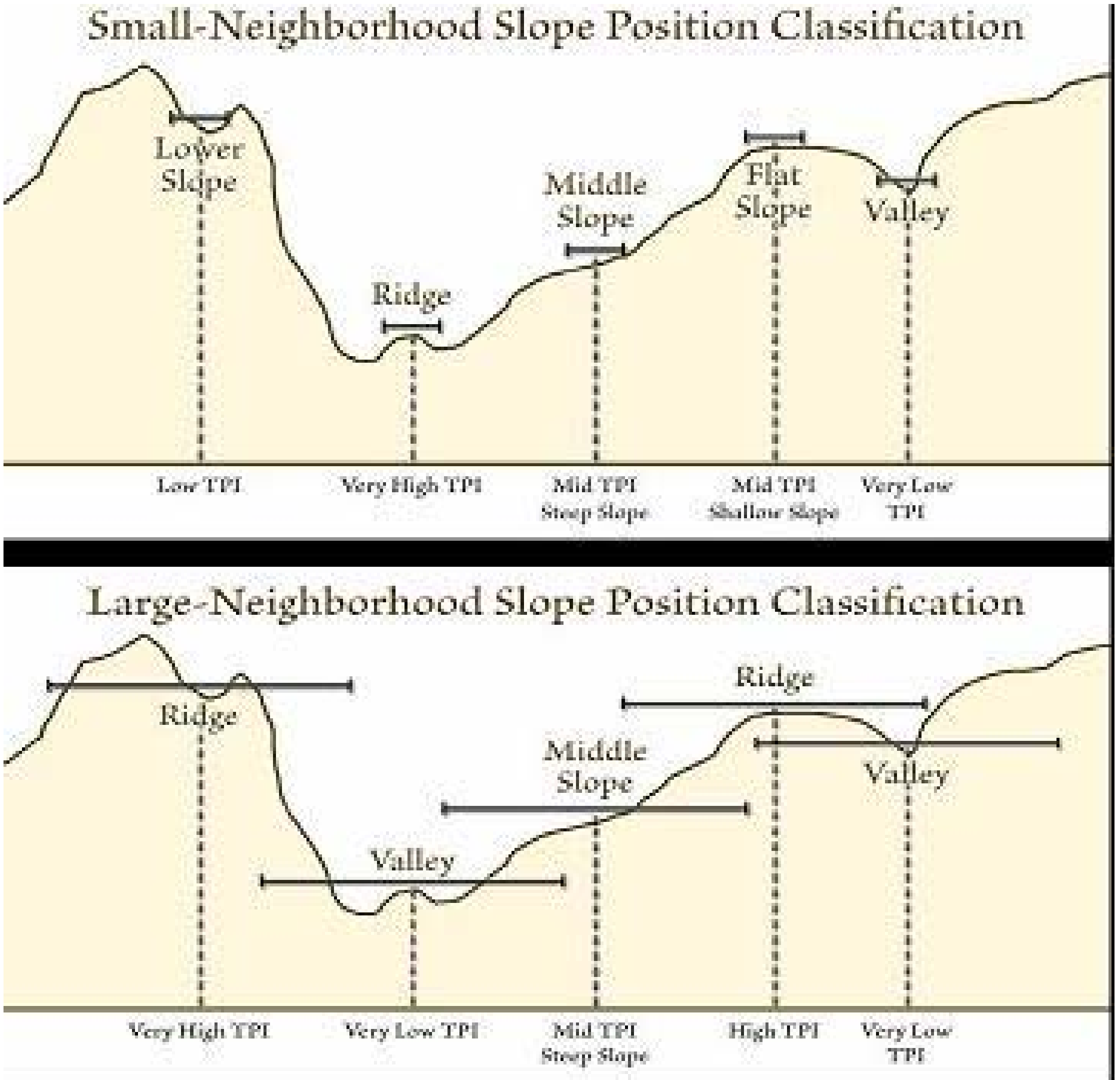

Figure 12. Effects of scale on topographic position analysis (Jenness 2005). Analyzing a given area using a smaller neighborhood may result in inaccurate classification of a landform. For example, at the smaller scale (upper box), the leftmost area is identified as a lower slope. At the larger scale (lower box), it is revealed that the landform is actually located at a small-scale depression on the top of a prominent ridge.

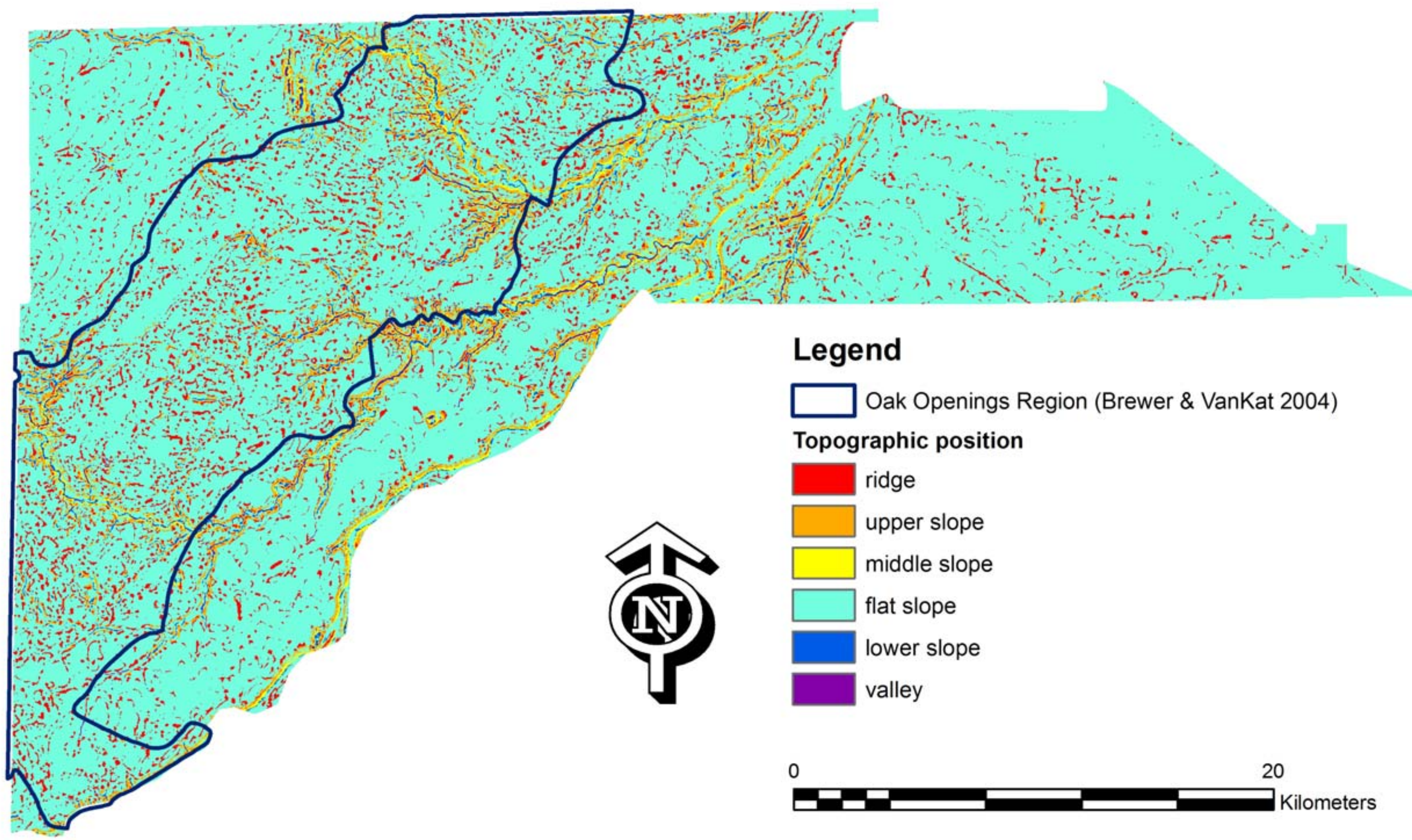

Figure 13. Topographic positions (six-class, 200 m neighborhood) of Lucas County, Ohio. Raster data, 30 m. Generated from elevation and slope data using TPI extension. Most of the county is flat-sloped with a few middle-slope river valleys. Extreme areas of ridge and valley are rare.

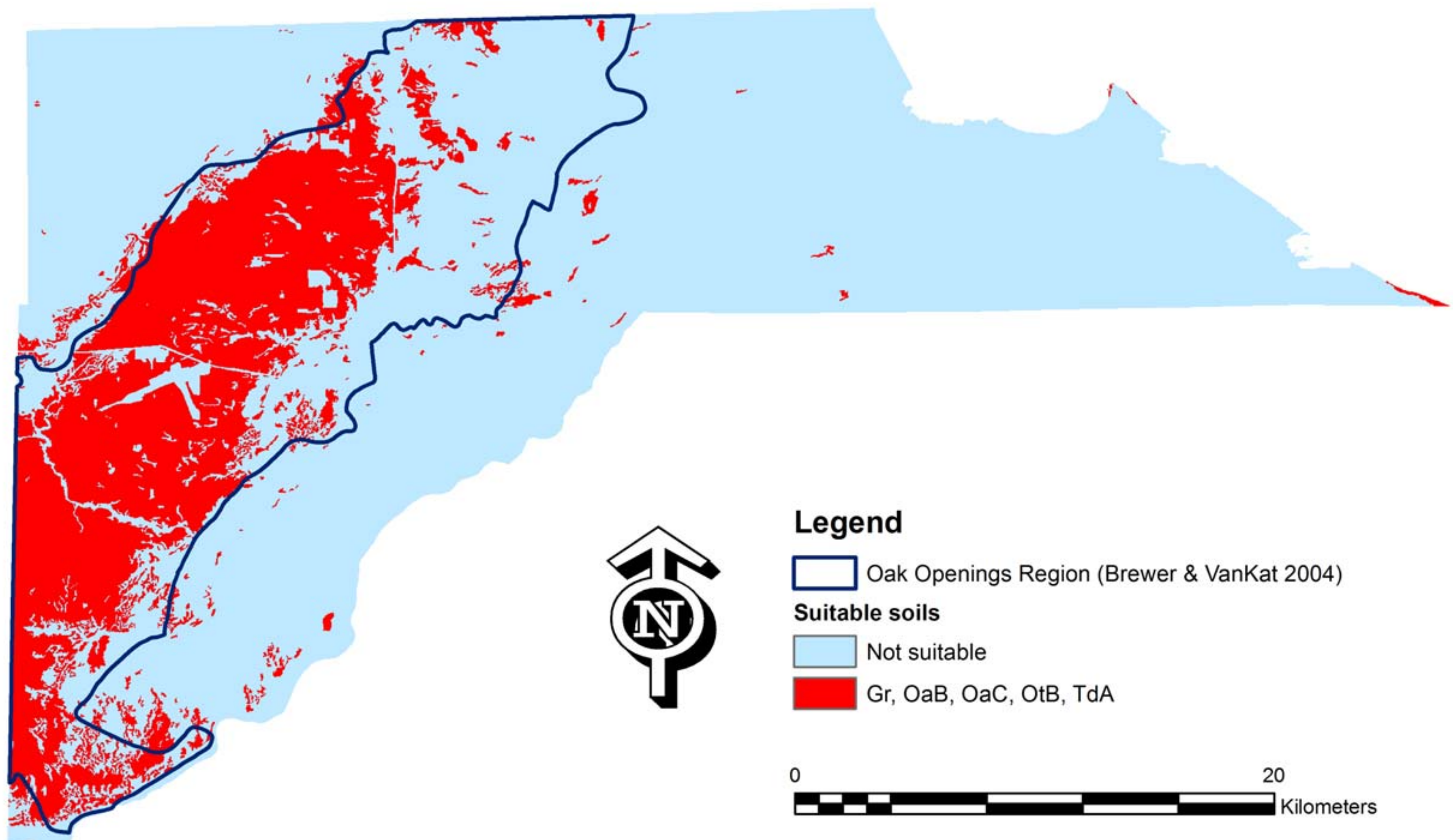

Figure 14. Areas of suitable oak savanna complex soils in Lucas County, Ohio. Suitable soils (OaB, OaC, TdA, Gr, OtB) are shown in red. The other 57 soil types are shown in blue.

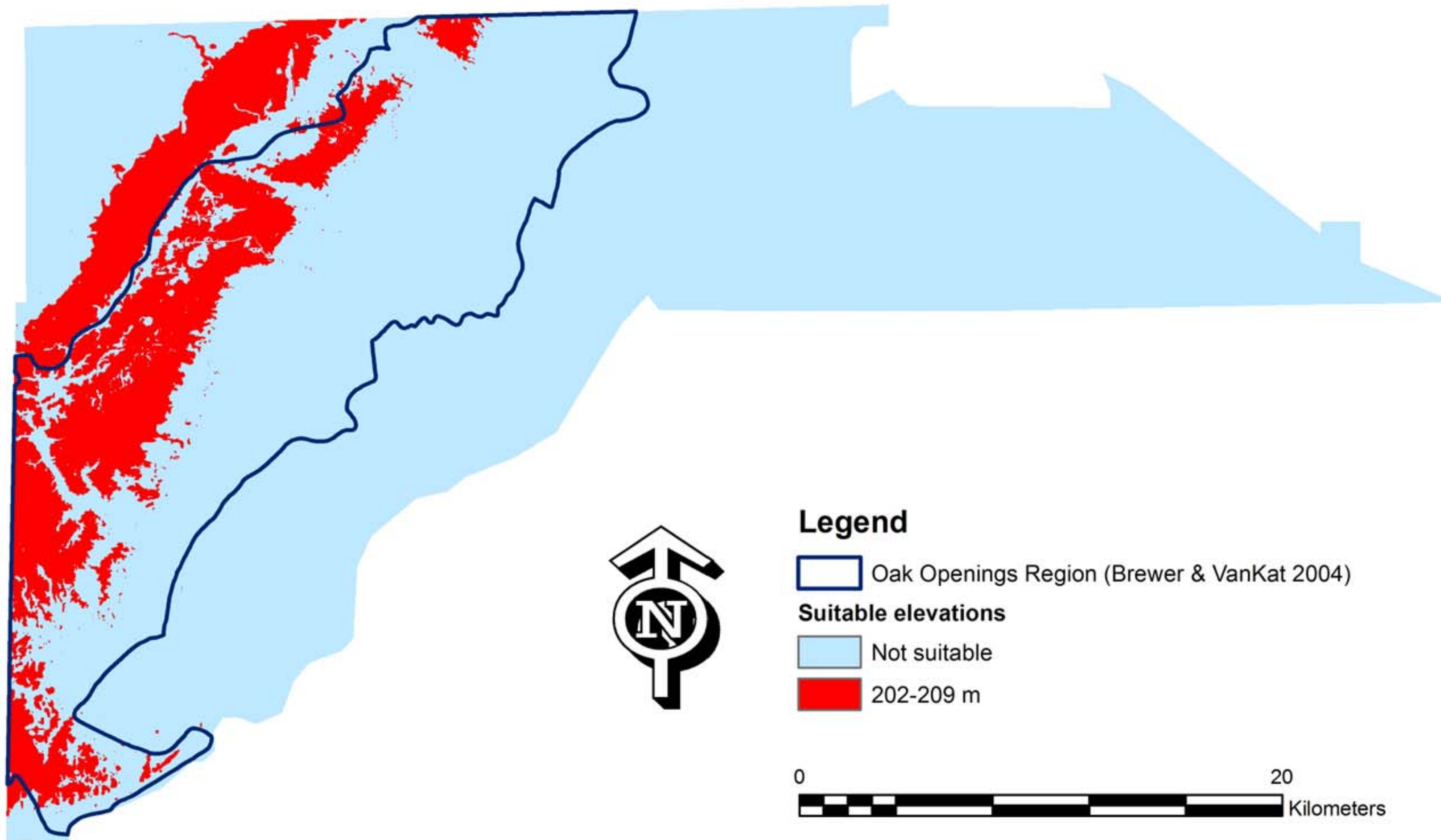

Figure 15. Areas of suitable oak savanna complex elevations in Lucas County, Ohio. Suitable elevations of 202-209 m are shown in red. All other elevations are shown in blue.

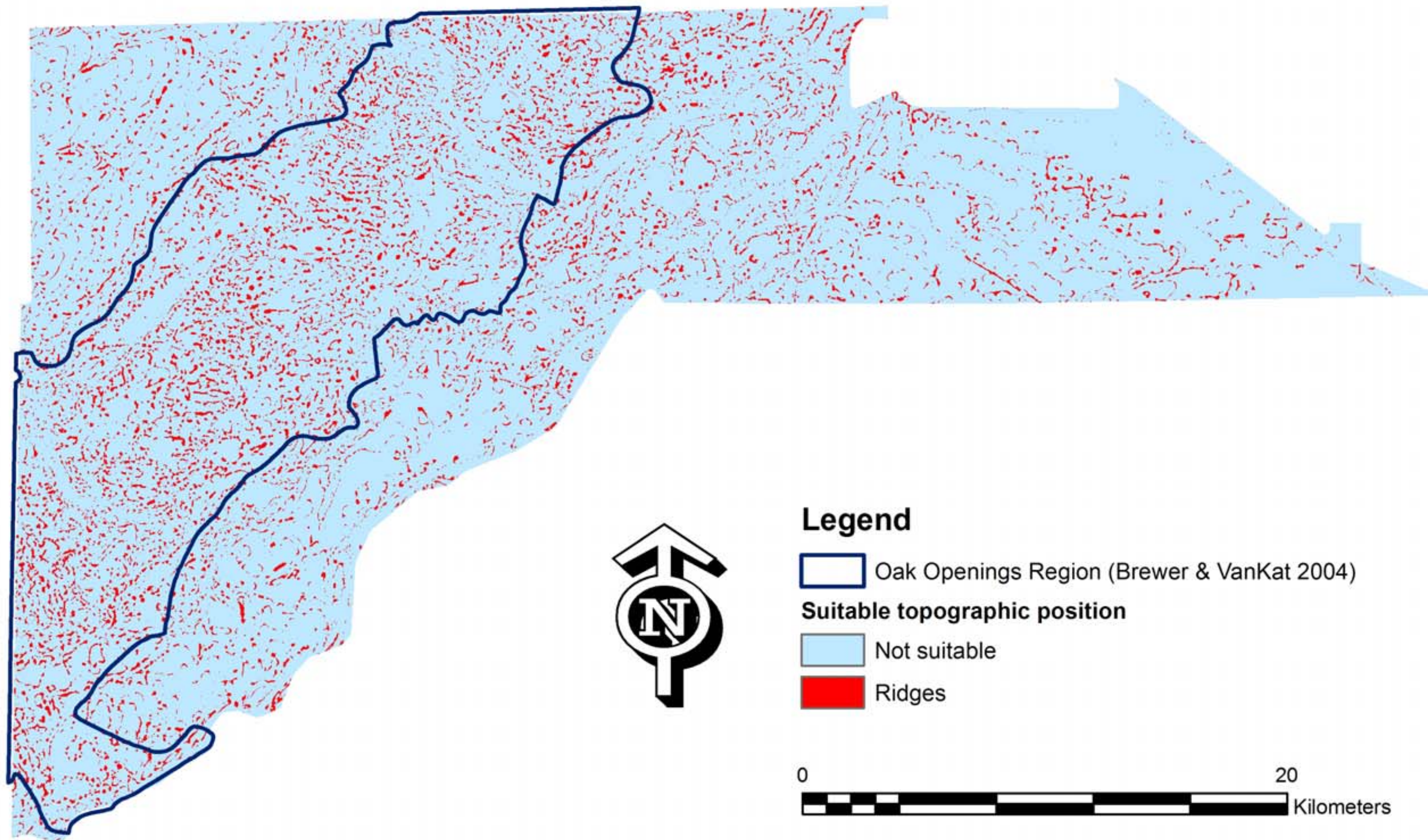

Figure 16. Areas of suitable oak savanna complex topographic position in Lucas County, Ohio. Suitable ridge topographic positions are shown in red. The other 18 position are shown in blue.

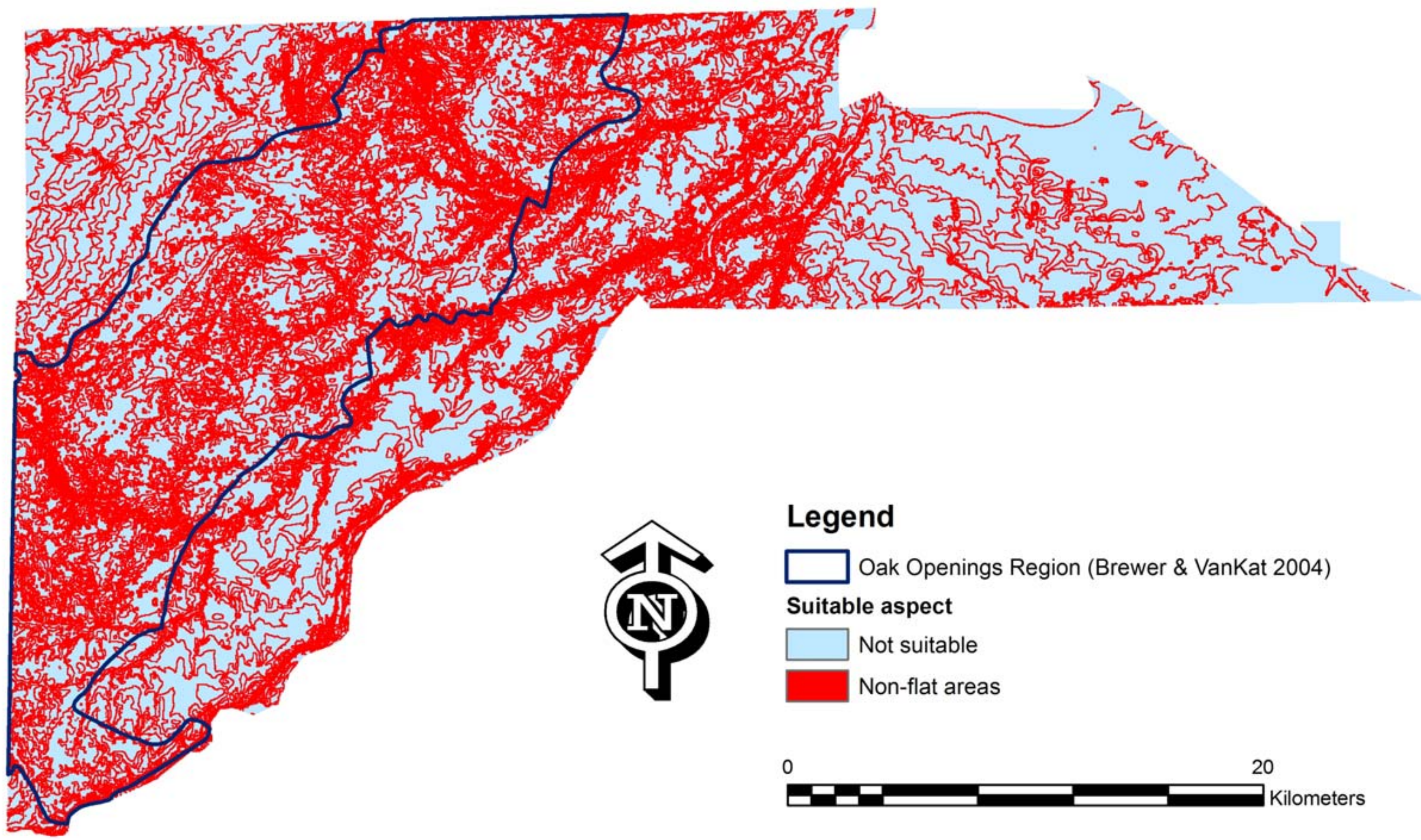

Figure 17. Areas of suitable oak savanna complex non-flat aspect in Lucas County, Ohio. Suitable areas of non-flat aspect are shown in red. Flat aspect areas are shown in blue.

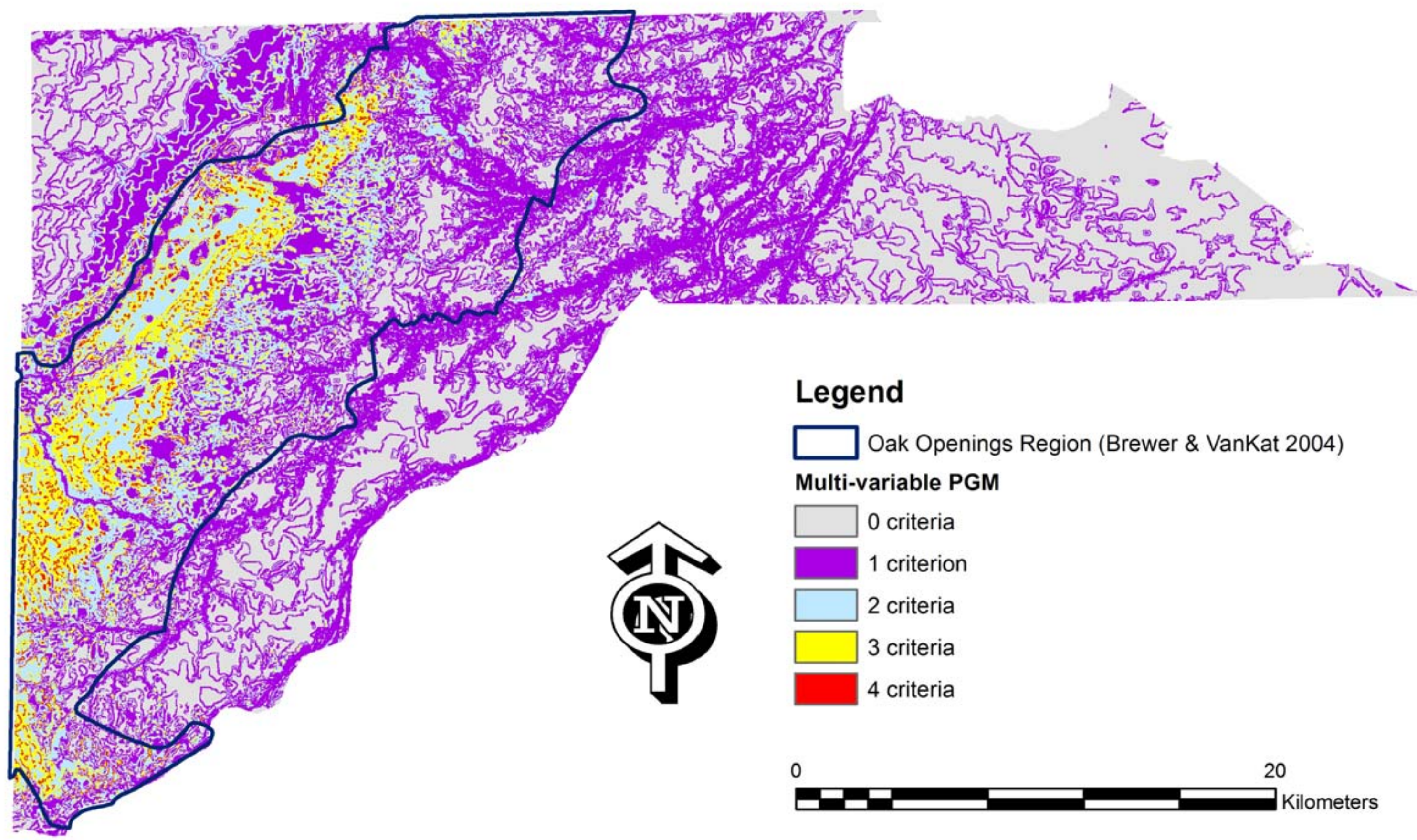

Figure 18. Multi-variable PGM for oak savanna complex in Lucas County, Ohio. Areas meeting all four criteria – suitable soils, elevation, topographic position and aspect are shown in red. Areas meeting fewer criteria are shown in yellow, blue, purple and grey.

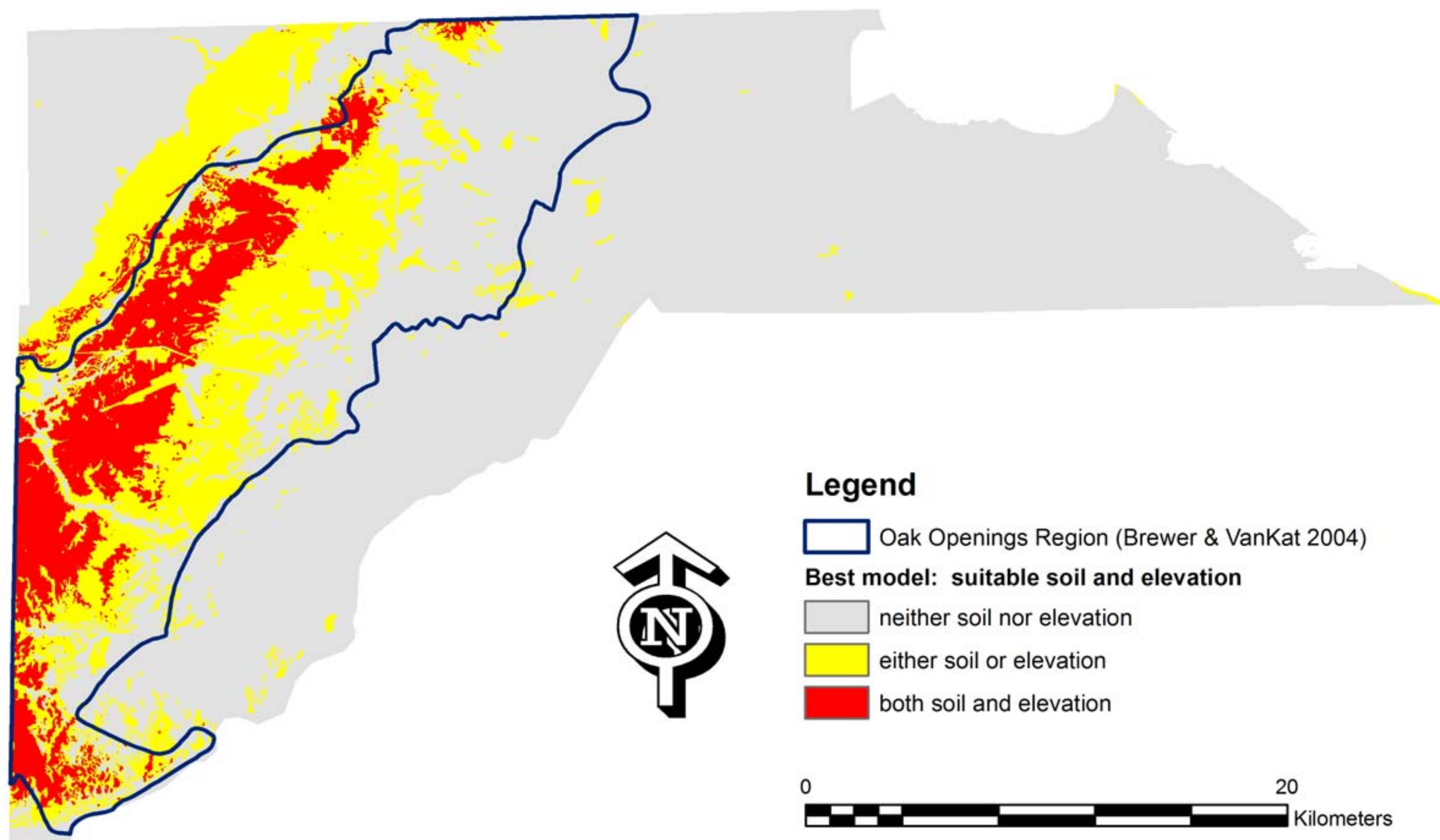

Figure 19. Soil-Elevation predictive model for oak savanna complex. Regression analysis revealed that a soil-elevation model explained most of the variance in the data set. Areas meeting both criteria are shown in red; meeting one criteria, yellow; meeting no criteria, grey.

VII. APPENDIX B: SITE DATA

Note: the following applies to Tables 15 and 16 in this appendix:

1. Ecosystem subtype:

   sb = sand barrens, aka Midwest sand barrens or “oak barrens”

   lb = lupine barrens, aka Black Oak/Lupine barrens or “oak savannas”

   dsb = degraded sand barrens

   dlb = degraded lupine barrens

2. Owner-ID:

   mpta = MetroParks of the Toledo Area

   tnc = The Nature Conservancy

3. Area in hectares

4. Geographic location

Table 14. Training site statistics and locations.

Ecosystem subtype	Owner-ID	Area (ha)	Latitude (°N)	Longitude (°W)
sb	mpta3	2.85	-83.8528	41.5361
dsb	mpta4	1.87	-83.8600	41.5436
dsb	mpta7	6.50	-83.8588	41.5689
dsb	mpta8	1.66	-83.8363	41.5499
dsb	mpta9	6.83	-83.8347	41.5579
dsb	mpta10	4.64	-83.8205	41.5593
dsb	mpta11	1.12	-83.7654	41.5999
dsb	mpta13	1.35	-83.7736	41.6111
sb	mpta14	0.67	-83.7656	41.6398
lb	mpta15	11.93	-83.8467	41.5368
lb	mpta17	7.84	-83.8564	41.5573
dlb	mpta19	16.38	-83.8605	41.5620
lb	mpta21	6.58	-83.8380	41.5395
dlb	mpta23	33.84	-83.8424	41.5529
lb	mpta24	8.13	-83.8278	41.5397
lb	mpta25	6.20	-83.8686	41.5578
dlb	mpta26	7.29	-83.8675	41.5554
lb	mpta27	12.28	-83.8628	41.5792
dlb	mpta28	18.01	-83.8641	41.5722
lb	tnc1	0.91	-83.8037	41.6117
dlb	tnc2	7.77	-83.8034	41.6098
sb	tnc3	2.07	-83.8018	41.6081
dlb	tnc4	0.40	-83.8053	41.6090
sb	tnc5	4.63	-83.8113	41.6072
lb	tnc6	0.14	-83.8059	41.6066
lb	tnc7	1.20	-83.8074	41.6052
dlb	tnc8	4.48	-83.8008	41.6200
dlb	tnc9	0.20	-83.7896	41.6259
dlb	tnc10	0.10	-83.7888	41.6266
dlb	tnc12	1.33	-83.7893	41.6230
dlb	tnc13	0.77	-83.7855	41.6245
dlb	tnc14	0.44	-83.7863	41.6257
dlb	tnc15	0.99	-83.7857	41.6235
dlb	tnc16	0.52	-83.7869	41.6251
sb	tnc17	0.94	-83.7944	41.6177
lb	tnc19	0.20	-83.7900	41.6198
sb	tnc20	1.87	-83.7913	41.6188
sb	tnc21	0.23	-83.7916	41.6213
lb	tnc22	0.63	-83.7938	41.6214
sb	tnc23	0.37	-83.7936	41.6209
dlb	tnc26	27.74	-83.8091	41.6072
sb	tnc27	0.53	-83.7897	41.6185
dlb	tnc28	0.78	-83.7888	41.6186
sb	tnc30	3.19	-83.7857	41.6206
dlb	tnc31	0.95	-83.7839	41.6254
dlb	tnc32	1.04	-83.7816	41.6255
sb	tnc33	0.52	-83.7828	41.6256

Table 15. Evaluation site statistics and locations.

Ecosystem subtype	Owner-ID	Area (ha)	Latitude (° N)	Longitude (° W)
sb	mpta1	4.36	-83.8423	41.5346
sb	mpta2	0.58	-83.8367	41.5354
sb	mpta5	3.68	-83.8549	41.5511
sb	mpta6	5.28	-83.8525	41.5546
dsb	mpta12	0.99	-83.7748	41.6026
lb	mpta16	0.56	-83.83578	41.5354
dlb	mpta18	7.73	-83.8468	41.5591
dlb	mpta20	15.53	-83.8427	41.5380
dlb	mpta22	21.14	-83.8359	41.5460
lb	mpta29	9.38	-83.7728	41.6025
dlb	tnc11	1.86	-83.7891	41.6282
lb	tnc18	3.01	-83.7937	41.6197
lb	tnc24	2.81	-83.7885	41.6215
sb	tnc25	0.51	-83.7890	41.6213
sb	tnc29	4.06	-83.7864	41.6183

Table 17. Verification site characteristics and locations. Data includes immediate and adjacent land uses and tree densities, structure density, distance to nearest structure and geographic coordinates.

Pt.	Land use	Adjacent land use	Tree density	Adjacent tree density	Structure density	Distance to structure	Latitude (°N)	Longitude (°W)
0	rural residential	rural residential, forested	thin deciduous	none to thin deciduous	light	45 m	-83.8270	41.5573
1	forested	forested, residential, commercial	thin deciduous	thin deciduous	medium	120 m	-83.7900	41.6040
2	forested	forested, agricultural, rural residential	thick deciduous	none to thick deciduous	very light	60 m	-83.8105	41.6286
3	rural residential	agricultural, forested, rural to medium residential	none (rural residential)	none to thin deciduous	light to medium	7 m	-83.8450	41.5969
4	forested	forested, agricultural, rural residential	thin deciduous	thin deciduous	very light	284 m	-83.8208	41.5704
5	agricultural	agricultural, forested, rural residential	none (agricultural)	none to thick deciduous	very light	87 m	-83.8760	41.4809
6	forested	forested, agricultural, recreational, rural residential	thin deciduous	thin to thick deciduous	very light	487 m	-83.7886	41.6648
7	agricultural	agricultural, forested, rural residential	none (agricultural)	none to thin deciduous	very light	139 m	-83.8503	41.5856
8	agricultural	agricultural, forested, rural residential, commercial	none (agricultural)	none to thick deciduous	very light	126 m	-83.8303	41.6185
9	residential	agricultural, residential, forested	thin deciduous	none to thin deciduous	light	47 m	-83.8633	41.4914