DEPARTMENT OF TRANSPORTATION

Historic Vegetation Model for Minnesota

MnModel Phase 4

Elizabeth Hobbs

June 24, 2019

MnModel was financed with Transportation Enhancement and State Planning and Research funds from the Federal Highway Administration and a Minnesota Department of Transportation match.

Contents

Introduction

Minnesota is fortunate to have detailed descriptions of its landscape and vegetation immediately prior to extensive Euro-American settlement. These are contained in the records of the General Land Office's Public Lands Survey. These records were used to develop an historic vegetation model for Minnesota using statistical modeling. The resultant high-resolution historic vegetation map in GIS format was used as input to the Minnesota Department of Transportation's (MnDOT) MnModel Phase 4 archaeaological predictive models (Hobbs 2019).

Need for Project

Marschner Map

In 1930, Francis J. Marschner produced a map of Minnesota vegetation complied from the Public Land Survey notes (Marschner 1974). This 1:500,000 scale map was later digitized by the Minnesota Department of Natural Resources (MnDNR) and used as the source of vegetation variables for MnModel Phase 3 (Hudak et al. 2002). The digitized Marschner map has several problems as a data source for modeling. First, it is very generalized. A number of features are mentioned in the surveyors' notes and illustrated on their plat maps that do not appear on the Marschner map. Second, Marschner's methods were not documented, and his vegetation classification scheme is not ideal for our purposes. Finally, and most important, the map does not register well with terrain. Most conspicuously, lakes and wetlands do not overlay their basins.

MnModel Phase 4

The purpose of MnModel Phase 4 was to update the archaeological predictive model developed twenty years previously using better data. These better data included more, and more accurately mapped, archaeological site and survey locations, higher resolution terrain, soils and geomorphic data, and a model of historic and prehistoric surface hydrography (Hobbs 2019; Hobbs and Brown 2019). Vegetation variables are important predictor variables for archaeological predictive models. Vegetation diversity is an important indicator of nearby available resources. Vegetation types may indicate which specific resources are present locally. In Phase 3 of MnModel, the variable 'vegetation diversity within one kilometer' figured into ten of 22 models (Hobbs et al. 2002). Consequently, MnDOT needed something better than the Marschner map to represent historic vegetation distributions for Phase 4. The new vegetation layer needed to be higher resolution, to better represent the scale of vegetation patterning in the landscape, and to better 'fit' into the terrain (i.e. lakes and wetlands needed to be within their basins).

Project Goals

Improve on Marschner

The primary goal of this project was simply to produce something better than our previous historic vegetation model, the Marschner map. The new model needed to be both higher resolution and have greater locational accuracy. To evaluate the quality of the new model, we have the Public Land Survey (PLS) plat maps for

comparison. These are not perfect, but they do show approximate boundaries between forest and prairie as observed by the surveyors. They also show small vegetation polygons, such as wetlands, stands of trees, and patches of prairie. Although the boundaries are not perfect, they do tend to be accurate where they cross section lines, as these are the portions actually observered and recorded in line ntoes. If the model approximates these plat map patterns, we can assume that it is doing a reasonable job.

Prove Concept of Statistical Vegetation Modeling

Statistical modeling is an efficient way to analyze and model large quantities of data. Moreover, the model results can be used to create high resolution GIS raster layers. To the best of our knowledge, General Land Office vegetation data have not been used previously as input to statistical models of vegetation distributions. If we can prove that this is possible and produces reasonable results, we can develop better models in the future by improving the data.

Limitations

The quality and quantity of the vegetation data vary across Minnesota. It is unrealistic to expect that vegetation models will be equally successful everywhere. In the northern half of Minnesota, MnDNR has transcribed the surveyors' line notes into GIS format, as attributes of the surveyed lines. These provide their verbatim descriptions of the vegetation observed, as well as notes indicating where they entered or left specific types of vegetation. Such transcriptions would be a valuable addition to the southern half of the state, where it was necessary to base the vegetation classifications only on MnDNR's generalized vegetation categories assigned to section and quarter-section corners and to the bearing trees mapped at the corners.

Moreover, the predictor variables we have available for modeling are not a complete list of all factors that control vegetation distributions. In particular, we have no way to re-create the disturbance factors that acted on the vegetation prior to each survey. We know that some of the vegetation types are typical of past disturbance by fire. If you remove fire from the equation, the vegetation pattern would look very different. Yet without knowing how many years since each sample point burned, we cannot include that variable. This undoubtedly confuses the model – since Oak Woodland, Oak Savanna, and Oak Forest may be different temporal expressions in the same location.

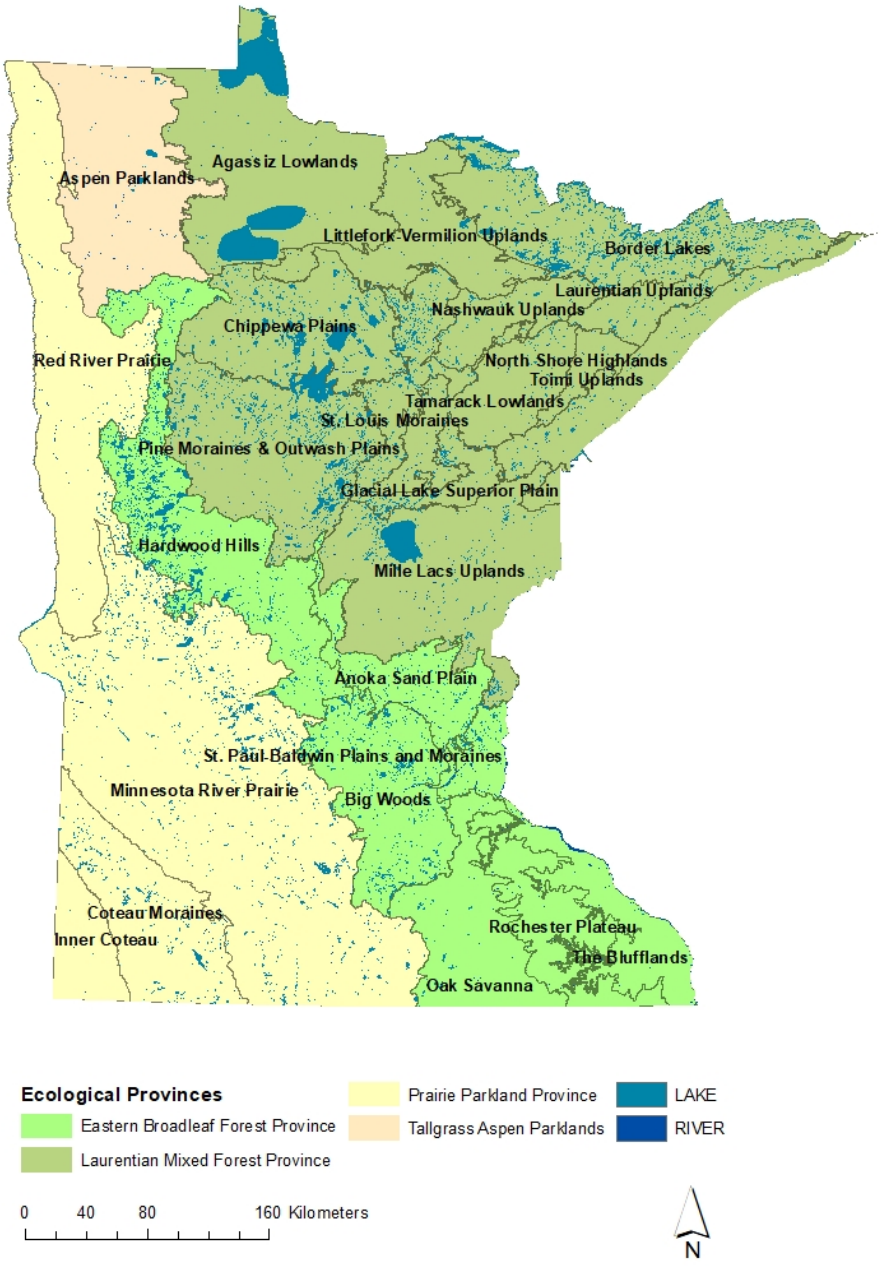
Finally, the resulting model is simply a view of the potential vegetation distributions at one point in time. Minnesota's climate and vegetation has changed more or less rapidly since ice sheets retreated about 10,000 years ago. However, we believe that the vegetation patterns represented in that survey are reasonable estimates of vegetation for the historic and late pre-historic periods.

Minnesota's Historic Vegetation

Minnesota is characterized by several distinct vegetation zones reflecting gradients of decreasing temperature from south to north and of decreasing rainfall from east to west. The Ecological Classification System (ECS) for Minnesota (Cleland et al. 1997; Hanson and Hargrave 1996) maps four ecological provinces (Figure 1). Drier regions are dominated by prairie, wetter regions by trees. Forests in the south are dominated by deciduous

trees, while northern forests are dominated by conifers. Ecotones of mixed vegetation (trees and grasses, mixed coniferous and deciduous forests) occupy the transition zones, which may be locally broad or narrow. These broad features have been more or less stable for nearly 3,000 years (Gibbon et al. 2002). ECS provinces are further subdivided into ten sections and 26 subsections. The subsections have been used by MnModel since Phase 3 (1998) to define regions for archaeological predictive modeling and were used to regionalize this historic vegetation model (Figure 1).

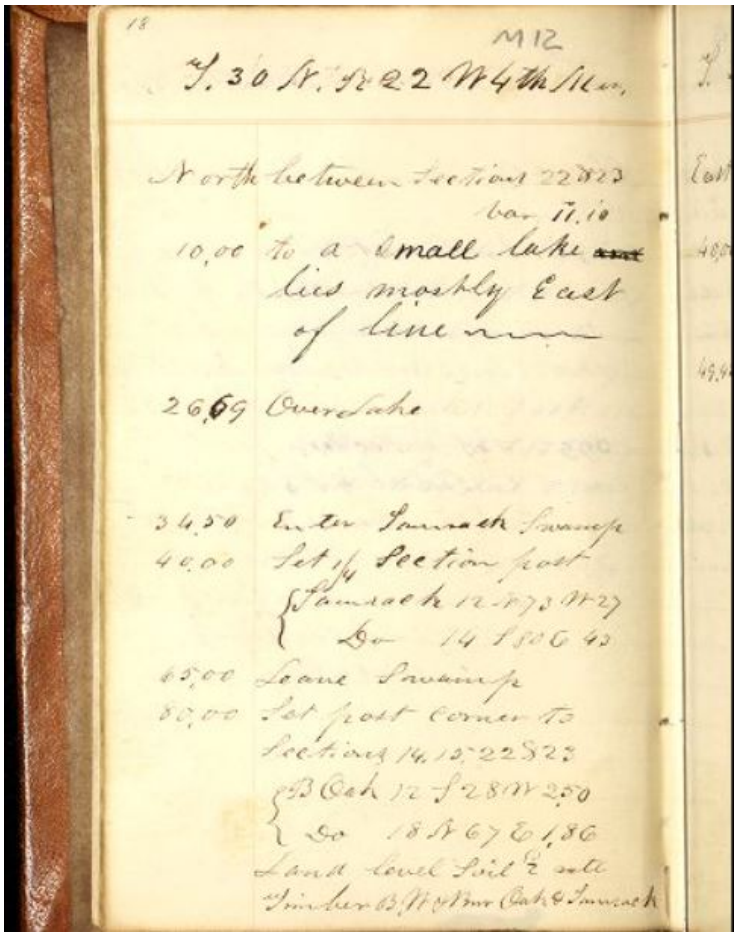
Figure 1: Ecological Provinces and Subsections of Minnesota



General Land Office Survey Records

The best primary source for the distribution of vegetation types in Minnesota prior to extensive Euro-American settlement is the Public Land Survey conducted by the U.S. Surveyor General's Office. These surveys were conducted in Minnesota between 1848 and 1907. Surveyors recorded their observations according to specific instructions they were given at the time (Stewart 1935). The instructions varied over time, and some surveyors were more diligent note-takers than others, but in general their notes (Figure 2) describe the vegetation they observed as they walked each section line, mentions of when they entered and left distinctive vegetation types, and the species and diameter of two to four trees at each corner (including the distance and direction to each tree). These notes have been scanned and are available from a U.S. Bureau of Land Management web site.

Figure 2: Example of Public Land Survey Line Notes

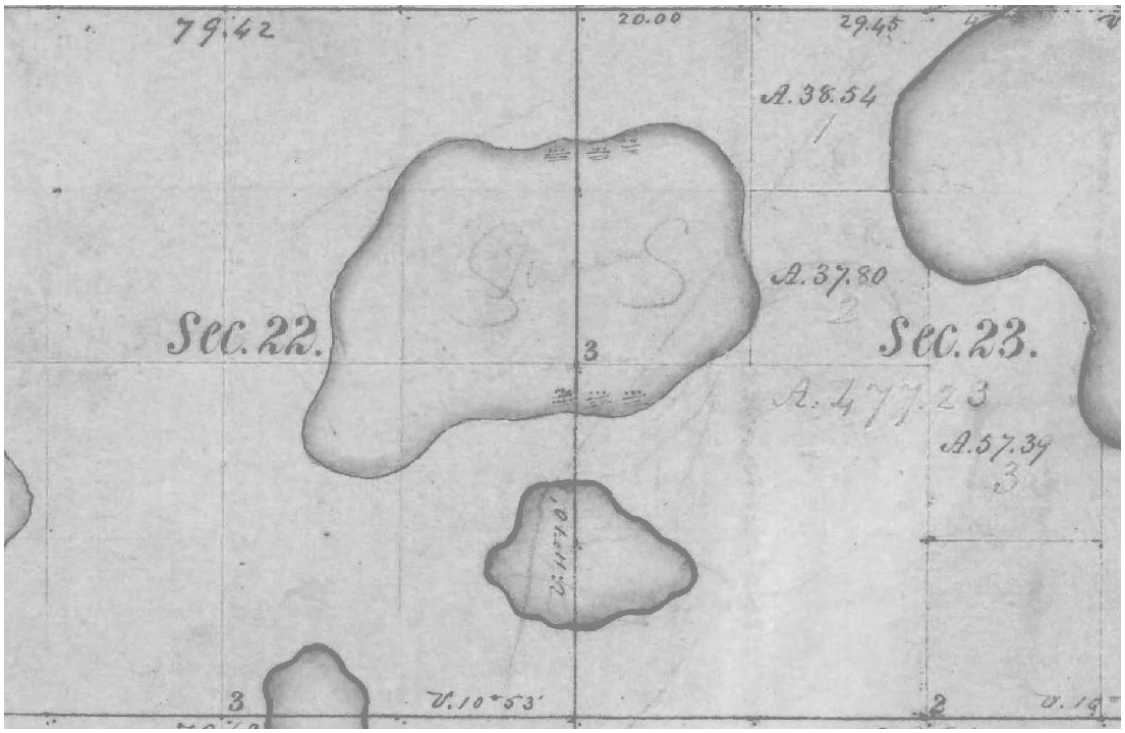


The notes shown in Figure 2 are for Township 30 North, Range 22 West of the Sixth Meridian, in Ramsey County, Minnesota. The surveyor is walking north between sections 22 and 23. All distances are recorded in chains (1 chain = 66 feet, or about 20 meters). He notes that at 10 chains he encounters a small lake, which is mostly east of the line. He records the opposite side of the lake at 26.69 chains. He enters a tamarack swamp at 36.5 chains. He sets his quarter section post at 40 chains, still within the swamp. He marks two tamaracks as bearing trees for the quarter section. At 65 chains he leaves the swamp, then sets his section corner at 80 chains. At the

section corner, he marks two bearing trees, both bur oaks. He describes the line as level, with second rate soil, oak timber, and tamarack.

The surveyors' line notes and field sketch maps were used as the source material for the creation of the Public Land Survey plat maps. These are maps of each township surveyed showing section lines, water bodies, wetlands, and other features observed. In general, the locations of features are accurate along the section lines where they correspond to the line note. Figure 3 shows the section of plat map for the notes in Figure 2. The small lake and tamarack swamp are correctly located along the line, but their shapes and extents away from the line may not be reliable.

Figure 3: Surveyor's Plat Map for Line Described in Figure 2 Example



Public Land Survey records have been used to reconstruct historic vegetation in Minnesota (Marschner 1974; Grimm 1984), Wisconsin (Bolliger et al. 2004; Hanron 1981; Vogl 1964), Michigan (Delcourt and Delcourt 1996; Brown 1998; Manies and Mladenoff 2000), and Illinois (Anderson and Anderson 1975). Most of these maps are for relatively small areas, and vegetation types are mapped simply as points or using interpolation techniques. They have also been used to study disturbance by catastrophic windthrow (Canham and Loucks 1984) and fire (Grimm 1984; Heinselman 1973; Spurr 1954). Biases in the selection of trees have been noted (Bourdo 1956), as has uncertainty in the designation of species (Mladenoff et al. 2002).

Marschner Map

In 1930, Francis J. Marschner produced a map of Minnesota vegetation complied from the Public Land Survey notes (Marschner 1974). This 1:500,000 scale map was later digitized by MnDNR and used as the source of vegetation variables for MnModel Phase 3. The digitized Marschner map has several problems as a data source

for modeling. First, it is very generalized. Many features are mentioned in the surveyors' notes and illustrated on the plat maps but do not appear on the Marschner map. Second, Marschner's methods were not documented, and his vegetation classification scheme is not ideal for our purposes. There is no distinction, for example, between deciduous forests dominated by oaks and those dominated by maple and basswood. Finally, and most important, the map does not register well with either the digital PLS plat maps (Figure 4) or the terrain. Most conspicuously, lakes and wetlands do not overlay their basins (Figure 5).

Figure 4: Digital Marschner Map Overlaid on PLS Plat Map, Showing Poor Registration of Lakes and Wetlands

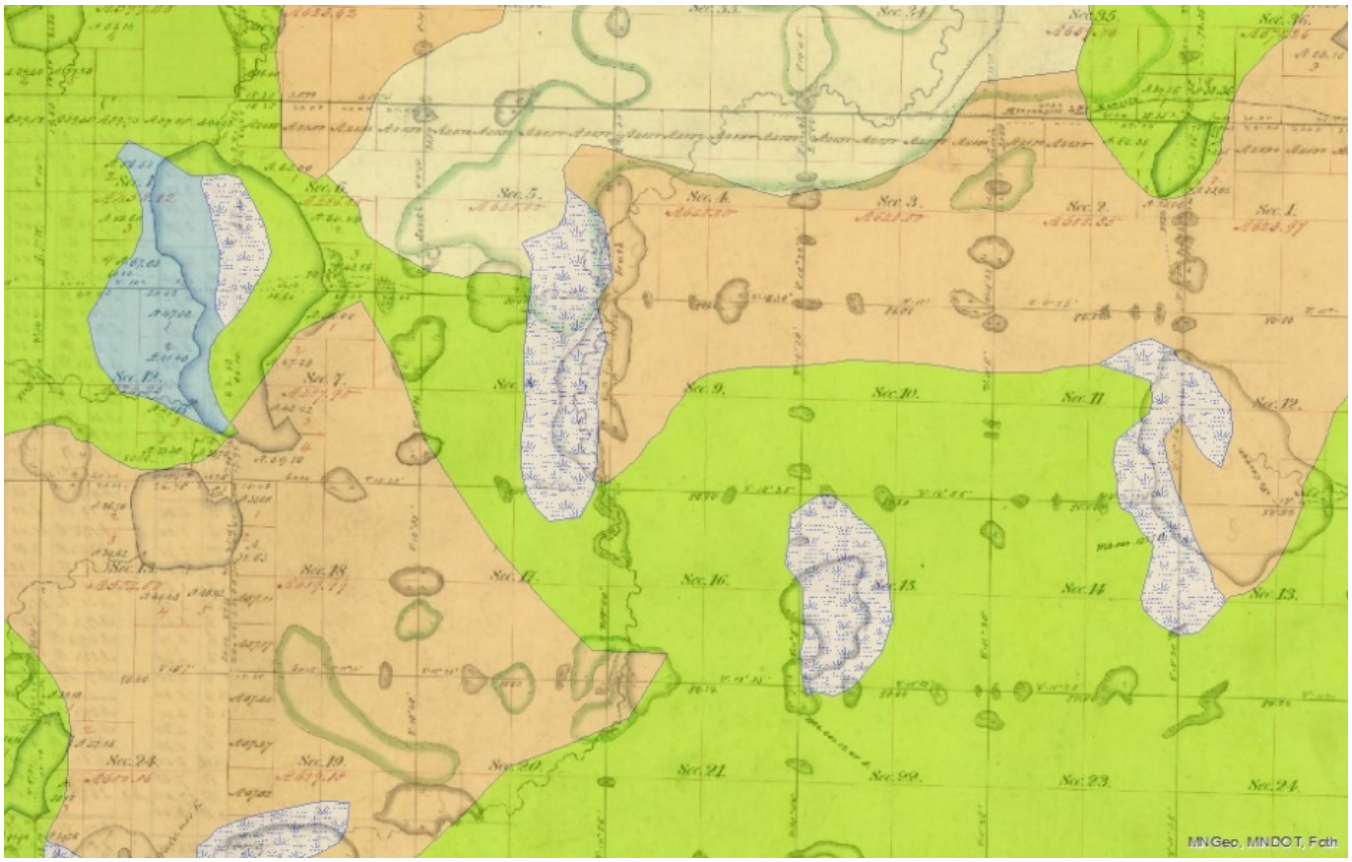
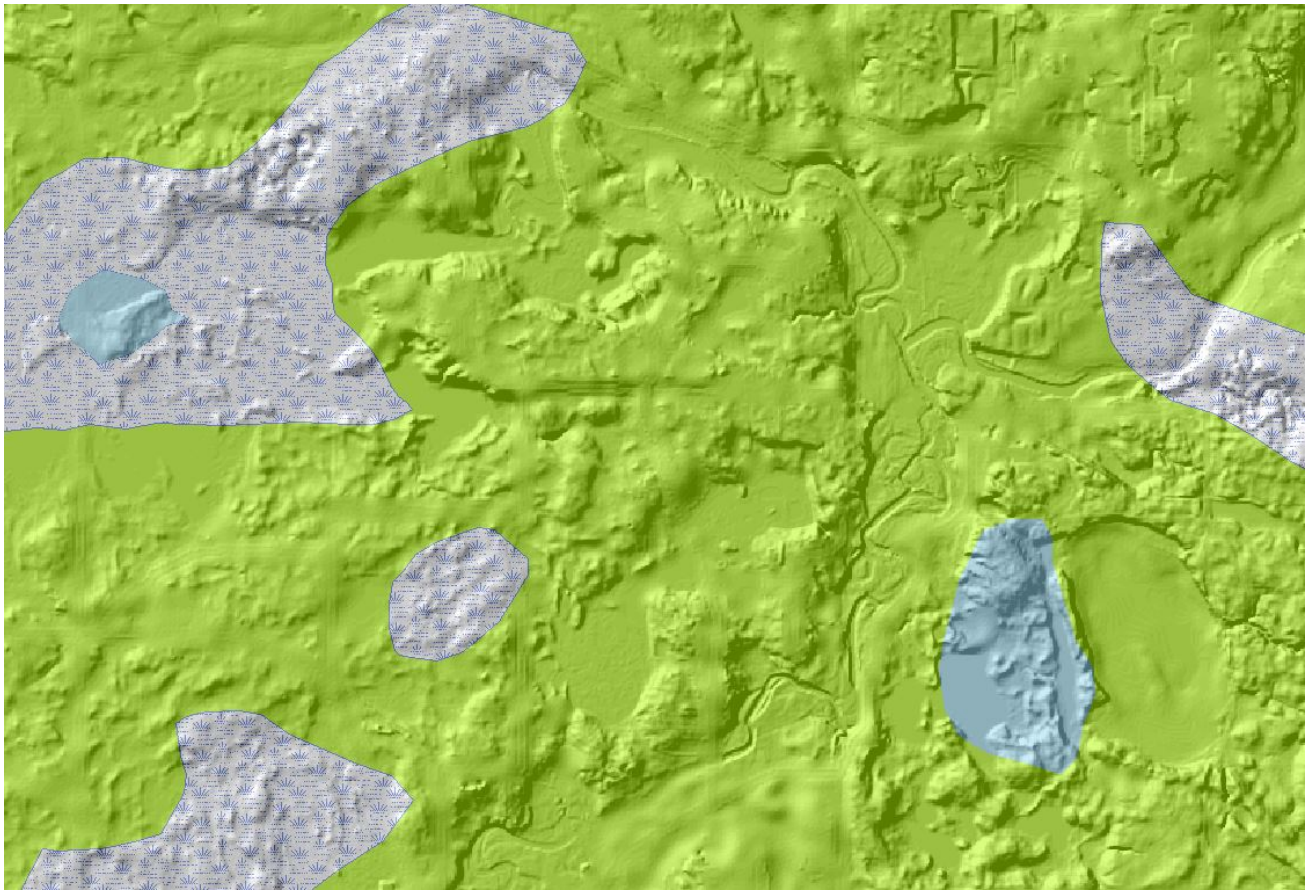


Figure 5: Lack of Correspondence between Marschner Lakes and Wetlands and their Basins



Methods

The statistical modeling procedures used require both 'observation points' where the vegetation type is known and 'prediction points' where the vegetation type has not been observed and will be predicted (Landrum and Hobbs 2019). The observation points for this model were derived from the Public Land Survey data. The prediction points are a set of points generated from a 30 meter grid of each region modeled (Brown et al. 2019). The prediction points are ultimately converted into 30 meter raster cells to create the GIS version of the vegetation model.

Only the observation points will include a 'vegetation type' variable. Both the observation and prediction points must include a suite of 'predictor' variables thought to be associated with vegetation type. For predicting vegetation type, we used a combination of terrain, geomorphic, and soil variables thought to be important to species and vegetation distributions. Preparation of the vegetation data and environmental predictor variables is discussed below.

Public Land Survey Vegetation Data

Public Land Survey data, as described above, are the primary source of information about historic vegetation in Minnesota (Table 1). Surveyors' vegetation observations and bearing trees were extracted to section and quarter section corner points and published by MnDNR in 1997. In a separate effort, John Almendinger (MnDNR) extracted survey notes to section lines for the northern half of the state and made these available to MnDOT. In 2013, MnDOT created a statewide mosaic of the scanned and georeferenced Public Land Survey plat maps and digitized polygons of hydrographic and vegetation features.

Table 1: Available Digital Historic Vegetation Data

Dataset	Source	Geography	Extent	Number of Features
Vegetation Points	MN DNR	Section corners; quarter section corners where on section lines	Statewide, except for large lakes, Fort Snelling, and isolated townships in Cook, Norman, Polk, and Wilkin Counties	251,656
Bearing Trees	MN DNR	Section corners; quarter section corners where on section lines	Statewide (with exceptions noted above) where trees present	357,468
Line Notes	MN DNR	Section lines	Complete for 38 counties and parts of 5 others	727,777
GLO Plat Maps	MnGeo/MnDOT	Statewide	Statewide	94,040

Data Quality

Surveyors walked north-south and east-west trending section lines, which are organized in a one mile grid, so the survey is not particularly 'high resolution.' Vegetation observations, aside from notations of entering or leaving a particular type of vegetation, are recorded at section corners and are generalizations for the entire line traversed. Except where no trees are near, bearing trees are recorded at section corners and often at quarter-section corners as well. Consequently, the total number of points observed is substantial (Table 1).

Vegetation Points

The MnDNR data referred to here as 'vegetation points' consists of point data at section corners and quarter-section corners that coincide with section lines. At each of these points, MnDNR extracted information from surveyors' line notes including:

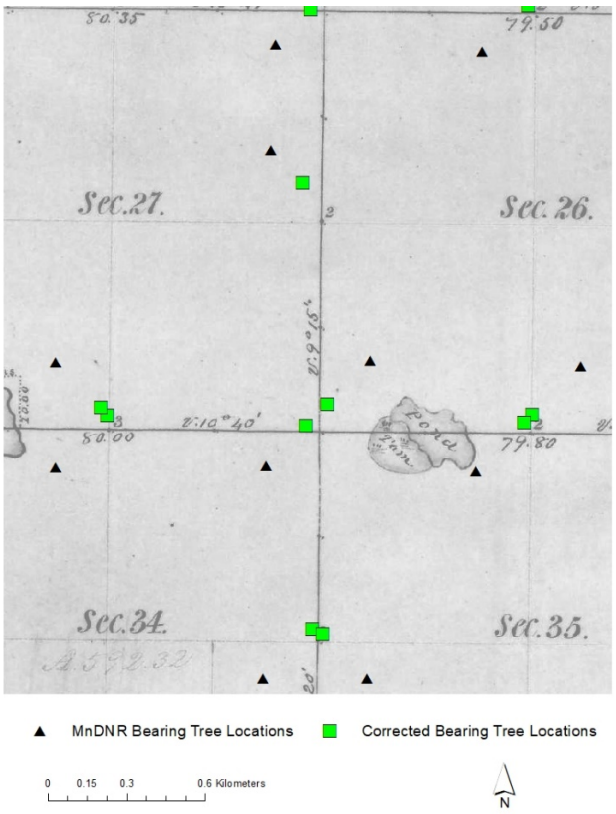
- A generalized vegetation type. This 'vegetation type' is essentially a shorthand for the surveyor's description. Their classification did not distinguish between types of forest but included three categories that could be considered 'oak savanna' ('Oak Barrens', 'Oak Openings', and 'Scattering Oak, Scattering Timber'). Several types of disturbance (fire, windthrow) were recorded without indicating anything about the vegetation that had been disturbed. Some categories were descriptions of the terrain rather than the vegetation ('Bottom', 'Dry Land', 'Dry Ridge', 'Island', 'Valley, Ravine').
- The species, diameter, direction, and distance to each bearing tree recorded at the corner.

The classification system used by MnDNR was inadequate for this project, so MnDOT reclassified the data (see below). In the process, MnDOT found additional systemic problems with the data. For example, though a section corner might be within a swamp, the vegetation might be identified as 'forest' because the majority of the line was forested. Thus users of these data should realize that the vegetation classifications may refer to the section line and not to the point itself. Moreover, some vegetation categories are clearly wrong. Either the person who recorded the data was looking at the wrong survey notes for that location, could not read the surveyor's handwriting, or entered the wrong code.

Bearing Trees

The MnDNR bearing trees feature class consists of individual points for each tree recorded with the attributes species, diameter, direction from corner, and distance from corner. The points are offset from the section corners at which they were recorded by a standard distance. MnDOT used the distance and direction data in the attribute table to move the points to their reported locations (Figure 6). In some cases, trees moved closer to the corners. In other cases, they move farther away. This allowed us to better evaluate local vegetation patterns, particularly with respect to terrain, soils, and polygons defined on the PLS plat maps.

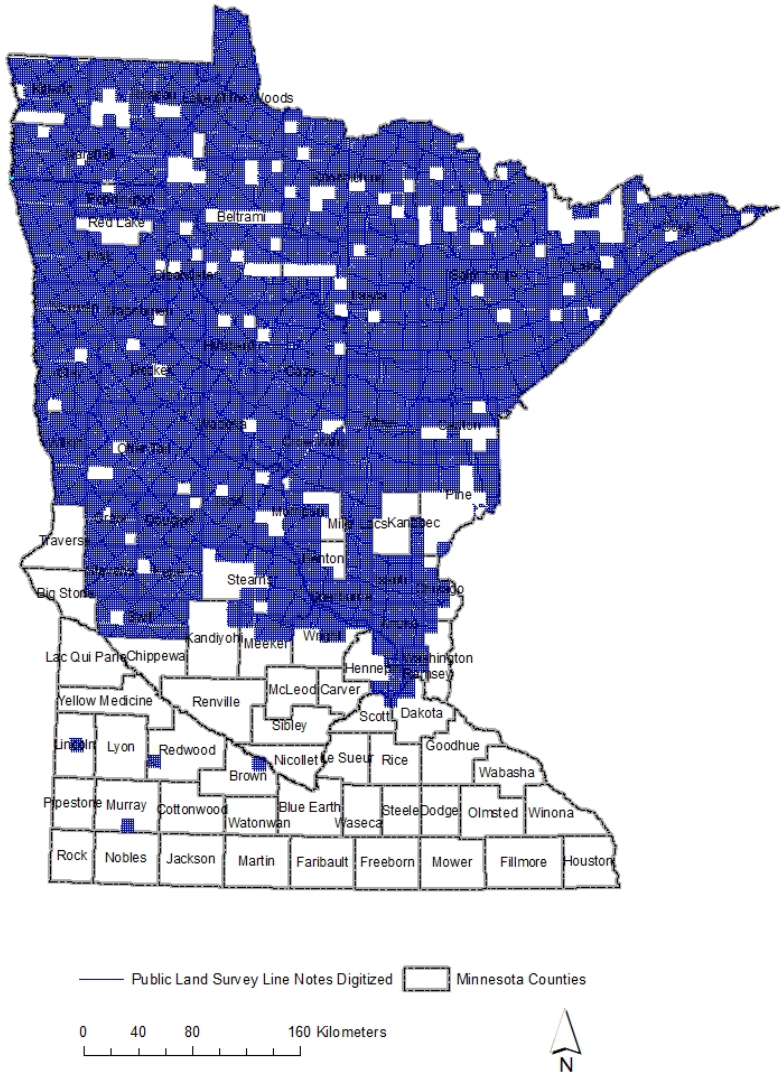
Figure 6: Locations of Bearing Trees Before and After Correction



Line Notes

The digitized line notes provided by MnDNR are invaluable for the northern part of the state (Figure 7). Each line segment is coded with a note representing an observation along the line (for example 'Over low level land through timber and dense brush' or simply 'Cor to Secs 6 31 & 32'). They also include the surveyor's summary of the entire line (for example, 'All flat and wet nearly all overflowed swamp and marsh'). This summary is attached to all segments that make up the same section line. The data also record a generalized vegetation type, disturbance code, bearing tree species, diameter,direction, and distance, two fields for 'note trees' (tree species mentioned in the line notes), seven fields for 'summary trees' (tree species mentioned in the surveyor's summary), and three fields for understory trees mentioned by the surveyor.

Figure 7: Extent of Digitized Line Note Data



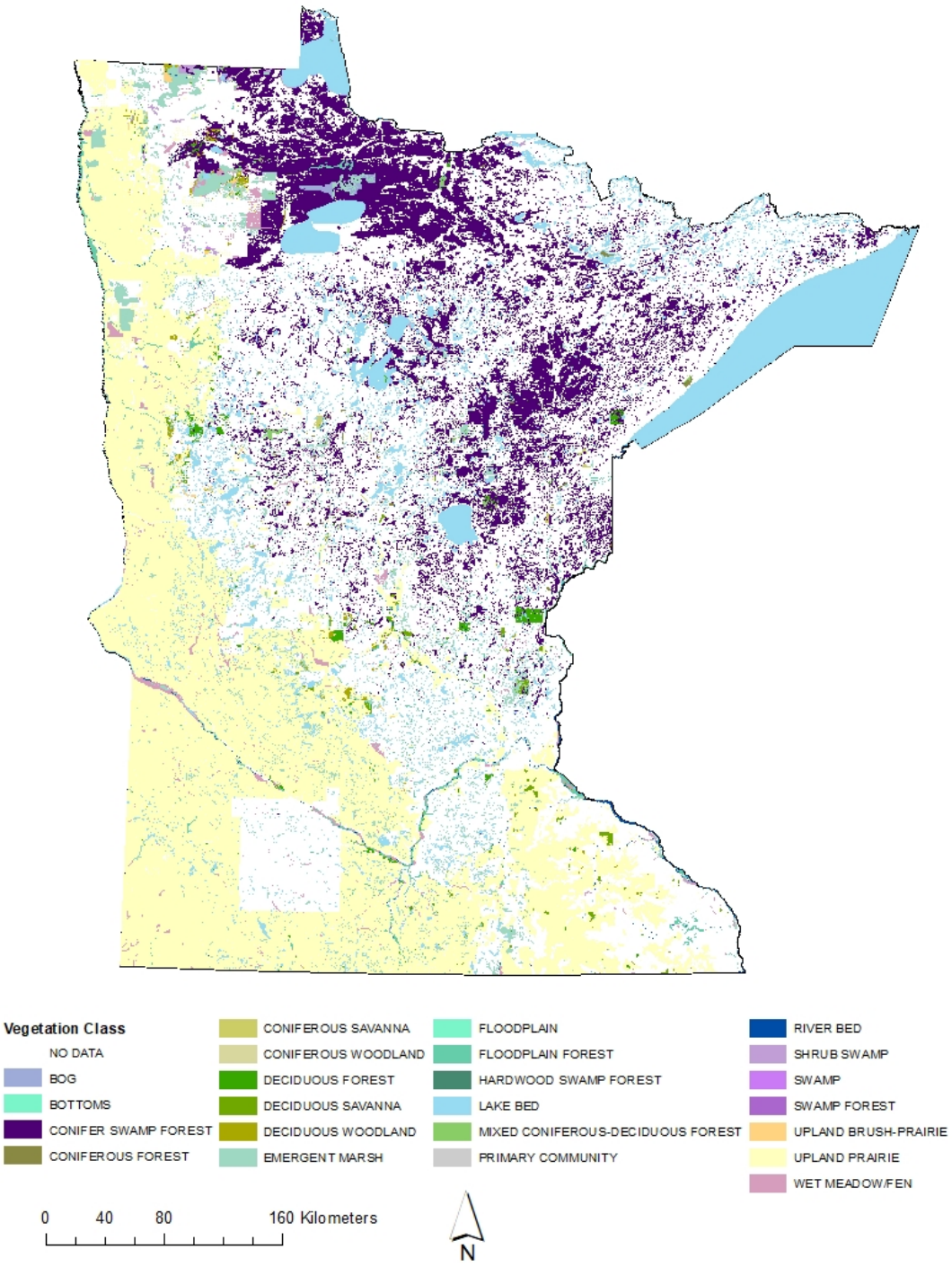
These notes are sometimes a challenge to interpret. For example, the note 'Entering marsh' may be attached to the line segment just prior to the marsh or to the segment within the marsh. It is useful to overlay the line data with the plat maps to verify the direction the surveyor was traveling. In some cases, the line notes and plat maps differ. In some cases, either the MnDOT recorder or the cartographer drafting the plat map misinterpreted the direction the surveyor was traveling. In other cases, it appears that notes for the wrong line were recorded. On the whole, though, the data are extremely useful.

GLO Plat Maps

Public Land Survey plat maps were scanned and georeferenced by the Minnesota Geospatial Information Office and mosaicked by MnDOT for this project. Georeferencing of the individual maps is tied to township corners

only. MnDOT georeferenced some northern Minnesota townships to section corners prior to mosaicking as they were otherwise too distorted internally to be digitized. Anyone using these maps should be aware of the georeferencing problem and compare the maps' features to terrain for verification. Also, note that locations of features along section lines tend to be accurate while features farther from section lines are imaginative.

Figure 8: Digitized Polygons from PLS Plat Maps



MnDOT also digitized vegetation and hydrographic features from the digital mosaic (Figure 8). These features were classified, to the extent possible, using either information displayed on the maps themselves or by overlaying and interpreting the vegetation point and line data described above. Many large 'background' polygons could not be classified as they contained multiple vegetation types.

Preparation of Vegetation Points

The vegetation point data from MnDNR were the primary data source used for modeling. All other data layers were used to assist in the interpretation and classification of the vegetation points. Had more time been available, it would have been desirable to extract data from the line notes to include in the modeling database. In particular, the line notes could be a valuable source of data for locations of rare vegetation types that were seldom intersected at section corners.

The primary task necessary to prepare the vegetation points for modeling was to apply a standard and meaningful vegetation classification scheme. The vegetation classification system adopted was developed by MnDNR (Aaseng 1993). Though not their most recently published system, it provided an appropriate level of detail, with some modification, for the available data. It is a hierarchical classification system of plant communities. At the top level, communities are categorized by system (terrestrial, palustrine, lacustrine, and riverine). The next level, vegetation class (see Figure 8), considers water regime, vegetation physiognomy, the life form of the dominant species, and the associated soils and landforms. Even these categories were too specific in some cases. For example, surveyors often note that they entered a 'floodplain', 'bottoms', or 'swamp' without further description of the vegetation. These categories were added to this classification level. The lowest classification level is the vegetation type, based primarily on species composition. The final categories applied are outlined in Appendix A.

Points were classified based on surveyors' descriptions as recorded in the digitized point, line, and plat map data. Bearing tree species assemblages at section and quarter section corners were used to distinguish between forest types. Most of these determinations were made by issuing queries for specific combinations of tree species in the point data. Where surveyors did not specify swamp types, bearing trees could sometimes be used to make that determination. Where bearing trees were absent or their assemblages inconclusive or ambiguous, surveyors' notes recorded in the line data were consulted if they were available. Data from plat maps were sometimes used for identifying, clarifying, or correcting point values. Distinctions between forest types with overlapping species assemblages (for example, Maple-Basswood Forest and Lowland Hardwood Forest) were made by consulting soil and geomorphic data.

Environmental Variables

Environmental variables are the 'predictor' variables in the statistical model. They were selected on the basis of their presumed effects on plant growth.

Data Sources

Terrain

Terrain variables were derived from the MnModel Phase 4 Digital Terrain Model (DTM) using ArcGIS. This 10 meter resolution DTM was derived from one-meter resolution LiDAR data and conditioned to restore the DTM to something approximating a pre-modern surface (Hobbs 2019; Hobbs et al. 2019). The terrain variables used for vegetation modeling were:

- Aspect Range: For this variable, aspect was divided into categories so that the sunniest locations have the highest values and the north facing slopes have the lowest values.
- Surface Curvature: Positive curvature values indicate that the land surface is upwardly convex at the cell. Negative values indicate that the surface is concave at the cell. Curvature is expected to effect surface water retention.
- Elevation: Elevation in feet.
- Relative Elevation within 90 Meters: This is a measure of a cell's height above the lowest point within 90 meters. If the cell itself is the lowest point, the value is zero. This measure is also used as an input to the Surface Roughness calculation.
- Surface Roughness within 90 Meters: This measure of roughness suggested by Hammer (1993) is calculated as (RGH = ((Elevation * 0.3048) + (Slope * 6) + (Relative Elevation * 0.6096)) / 2)
- Shelter Index: This index is designed to measure how 'sheltered' or 'exposed' a cell is with respect to the surrounding landscape (Kvamme and Kohler 1988). More sheltered locations have lower values.
- Percent Slope: The slope of the land surface.
- Topographic Position within 90 Meters: The Topographic Position Index (TPI) is intended to elucidate whether a terrain cell is situated on a ridge, within a valley, or on a side-slope. Calculations are based on a method developed by Guisan et al. (1999). Positive TPI values indicate locations higher than their neighborhood surroundings; near zero values indicate flat areas or areas of constant slope; and negative values are lower than their surroundings.
- Topographic Wetness Index: The Topographic Wetness Index (TWI) is a function of slope and the upstream contributing area orthogonal to the flow direction. Values are estimate of water accumulation and will be high in flat or depressed areas and low on slopes.

Geomorphology

The MnModel Landscape Model is the result of the MnModel Phase 4 project's reclassification and mosaicking of MnDNR, Minnesota Geological Survey (MGS), and MnDOT derived regional and local surficial geology and geomorphic data. Two variables were extracted from this model to use as predictors:

- Landform: Landforms are the smallest geomorphic unit mapped. There are 89 unique landforms defined by the Landscape Model.

- Landscape: In this hierarchical model, landscapes are the next level above landforms. There are eighteen unique landscapes mapped in Minnesota. The most extensive are the Stagnant Ice and Glaciolacustrine landscapes, while the rarest are the Tributary Fans and Meltwater Trough Fans.

Soils

All soil variables for MnModel Phase 4 were extracted from 2017 gSSURGO data. These data are available for most of Minnesota from the Natural Resources Conservation Service (NRCS). Even where soils data are present, there are many gaps in coverage. These include missing variable values within water bodies, disturbed areas (e.g. gravel pits or mines), and urban areas. Some variables simply were not reported for all map units. In some cases, missing data can be extracted from map unit names or other text fields. However, the extent of missing attribute data affected which variables we could use for modeling. We supplemented the gSSURGO data with drainage and productivity indices provided by Michigan State University (Schaetzl et al. 2009).

The gSSURGO database provides a mapunit table that aggregates selected soil attributes by soil mapunit. Many more attributes are not aggregated, but are presented in tables by soil components and soil horizons that require many-to-one joins to the mapunit table (and hence to the GIS data). We developed Python tools to aggregate these data by determining the values occupying the largest percentage of the mapunit.

The variables extracted for vegetation modeling are associated with soil drainage, water storage, fertility, and chemistry.

- AWS150: Available water storage (cm) in the top 150 cm of soil.
- CACO3: Calcium Carbonate in the surface horizon, expressed as a weight percentage of the < 2mm size fraction.
- CEC7: Cation Exchange Capacity (electrical conductivity), at pH 7.0, of the surface horizon.
- CLAY: Percentage of clay in the surface horizon.
- DRAIN: Dominant drainage class for the mapunit. Numeric values were assigned with a low value of '1' indicating very poorly drained soil and a high value of '7' indicating excessively drained soil.
- FFD_R: Number of frost-free days per year. The range is from 85 to 100.
- FLDFRQD: Flooding frequency of the mapunit. Numeric values were assigned to the classes ranging from '0' (None) to '5' (Very frequent).
- GRTGRP: Taxonomic Great Group.
- GYPSUM: Gypsum (hydrated calcium sulfate) in the surface horizon, expressed as the percent by weight in the < 20 mm fraction of soil.
- HYDGRPDCD: The Hydrologic Group, a grouping of soils with similar runoff potential under similar storm and land cover conditions. Low values indicate soils with low runoff potential when wet. High values indicate soils with high runoff potential when wet.
- HZDEP: Depth of the surface horizon (cm).
- OM: Percentage of organic matter in the surface horizon.

- PI: Productivity Index. Numeric values range from 0 (water, rocks, pits, urban land) to 18 (very rich mesic mollisols).
- REG_RICH: Regime Richness. This variable was created by extracting the regime richness values from Michigan State University's more complex REGIME (Ecological Class) index. Values range from very poor (10) to very rich (50), with water assigned '58'.
- REG_WET: Regime Wetness. This variable was created by extracting the regime wetness values from the more complex REGIME index. Values range from very dry (1) to very wet (7).
- SAND: Percentage of sand in the surface horizon.
- SILT: Percentage of silt in the surface horizon.

Preparation for Modeling

Regionalization

Because Minnesota is a large state (218,601 km^2 or 85,254 mi^2) with considerable environmental variation, it is necessary to model by smaller, relatively homogeneous regions then mosaic the regional models into a statewide model. Boundaries based on Ecological Classification System (ECS) subsections (Hanson and Hargrave 1996) were adopted for this purpose. Minnesota's ECS is part of a hierarchical national system of classification (Cleland et al. 1997) based on climate, geomorphology, terrain, soils, and vegetation. As subsection sizes vary some of the smaller subsections were combined with each other or with adjacent, larger subsections.

Initially, the intention was to model the same regions that would be used for the archaeological predictive models (Figure 9). However, deviations were required. Attempting to model the largest regions (AGLV and MNRP) exceeded the RAM of the available computer. The AGLV region was divided into its two component ECS subsections, AGLW (Agassiz Lowlands) and LFVU (Littlefork-Vermilion Uplands), for modeling. MNRP (Minnesota River Prairie) was divided into three subregions based on its component Ecological Land Type Associations. Finally, two modeling regions, ICOT (Inner Coteau) and COTM (Coteau Moraines), were combined in an attempt to assemble enough forested sample points to include in the model (Table 2). Even with this effort, the model did a poor job of predicting forest and savanna in the locations where those types appear on the plat maps.

Figure 9: MnModel Phase 4 Modeling Regions

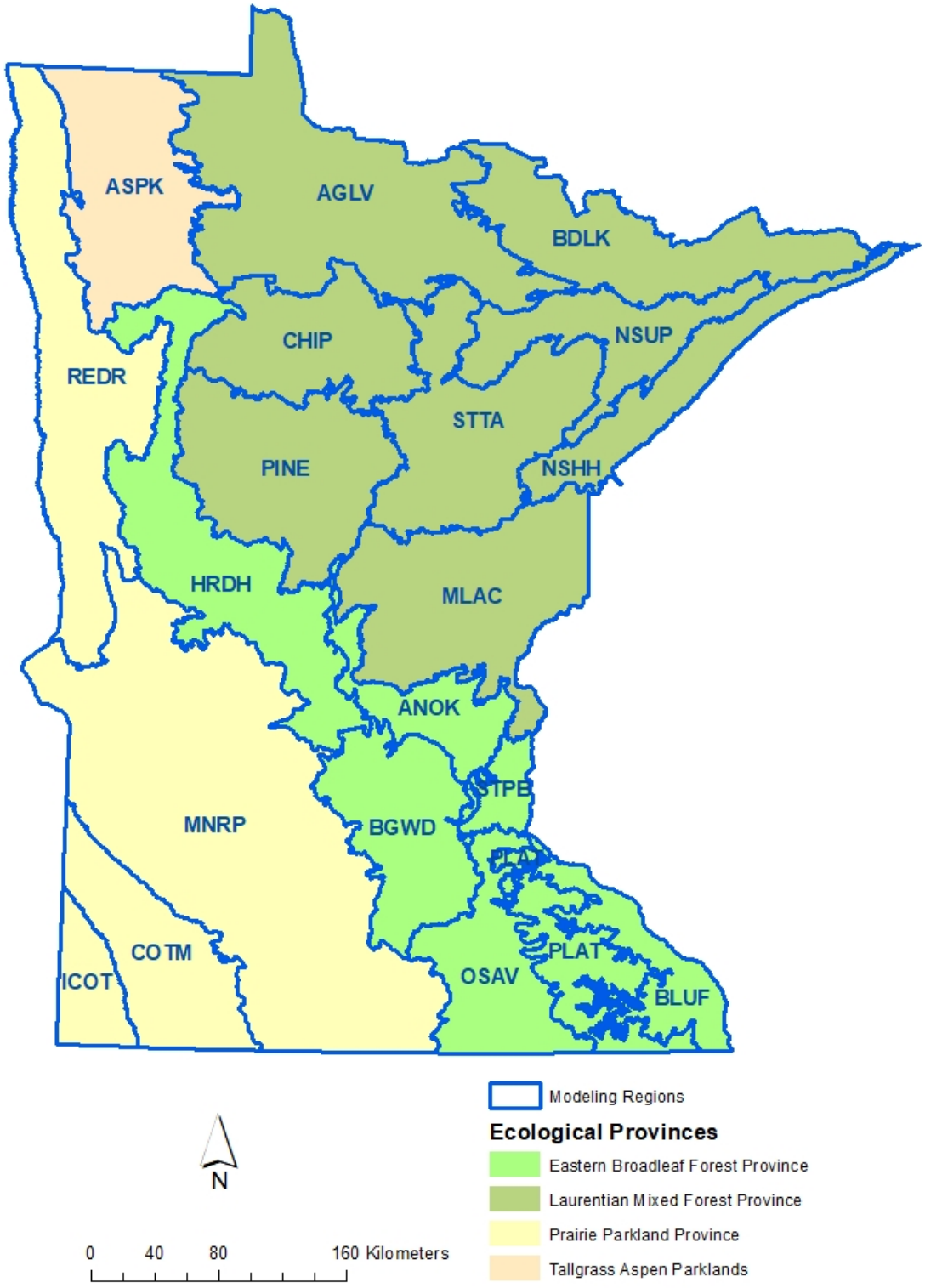


Table 2: Forest/Savanna Sample Points in ICOT and COTM

Type	ICOT	COTM	Combined	Reclassified As
Floodplain Forest	6	19	25	N/A
Oak Forest	3	25	28	N/A
Oak Savanna	2	6	8	Oak Forest
Lowland Hardwood Forest	0	2	2	Maple-Basswood Forest
Maple-Basswood Forest	0	11	11	N/A
Aspen Woodland	0	1	1	Oak Forest
Oak Woodland	0	1	1	Oak Forest

Buffers

All models were derived for their defined regions plus a surrounding 10 km buffer zone. This buffer zone is necessary to derive accurate measures of vegetation diversity for cells near the borders of the regions. Buffer zones for some regions necessarily extended outside of Minnesota. Vegetation types predicted outside of the state are necessarily based on vegetation/environment relationships within Minnesota. However, we did have soil and terrain data for surrounding states and terrain data for Canada to contribute to the models for these areas.

Missing Data

Statistical procedures cannot handle missing data. Unfortunately, we had missing data for two of our datasets in some locations. Soils data (gSSURGO) are available for most of Minnesota and all of the surrounding states but not for Canada. Yet even where soils data are available, there are many gaps in coverage. These include missing variable values within water bodies, disturbed areas (e.g. gravel pits or mines), and urban areas. In addition, some soil variable values were not reported for all map units. We were able to reconstruct some missing data from map unit names or other text fields, but were still left with considerable gaps. We also had no geomorphic data for Canada and the surrounding states. This became a problem only when modeling regions on the state border.

To compensate for the missing data, we developed procedures for multiple sets of models. The first, without soils variables (Model A), covered an entire region, The second, with soils variables (Model B), covered only areas with soils data. These models could be run for any area within Minnesota. A third model, one without soils data and geomorphology (Model C), could be run to model a border region plus buffers in neighboring states and Canada. A fourth model, with soils but not geomorphology (Model D), could be run for a region and its buffer if the buffer included a neighboring state. Our final models consisted of Model B where both soils and geomorphic data were available, Model A where soils data were absent but geomorphic data were available, Model D where soils data were available but geomorphic data absent, and Model C for buffer zones within Canada and areas where soils data were missing from neighboring state buffers.

A modeling mask (Brown et al. 2019) was created to facilitate selection of data and prediction points for Models B and D. For the statistical software to recognize missing data, all NULL values were calculated to '-999'.

GIS/R Interface

We tested several methods for interfacing between the GIS software (ArcGIS) and the statistical software (R), including R-Bridge. We found the simplest and most efficient procedure was to export point attribute tables from ArcGIS as text files (.csv format), import these files into R for analysis, export the models from R as text files, then create model rasters in ArcGIS from those exported files. To implement this required creating a 30-m grid of 'prediction points' containing x,y coordinates and using these to sample the same environmental variables sampled by the vegetation data points. Tools for creating these prediction points are documented in Brown et al. (2019).

Sampling

Two sets of point data are required for each model. The first, the 'training data,' consists of the coded vegetation points and the values of the predictor variables at each point. The second, the 'prediction points,' consists of a 30 m grid of points with only x,y coordinates and predictor variable values. The training data are used to develop the model, as these data include both vegetation type values and their associated environmental characteristics. After the model is developed, it is applied to the predictor points, where only the values of the predictor variables are known beforehand.

All vegetation and prediction points within each region and its 10 m buffer zone were used to sample the group of variables needed for Model C. The same points were used for Model A (all variables except soils), provided the region did not border a neighboring state or Canada. In this case, points outside of Minnesota were removed prior to sampling.

The modeling mask (Brown et al. 2019) was used to create versions of the vegetation and prediction points for Models B and D, where points within areas missing soils data were removed. For Model D, all variables except geomorphology were sampled. For Model B, points outside of Minnesota were also removed, if present, as these would have been missing geomorphic data. The remaining points were then used to sample all variables including soils.

For most regions, only Model A and Model B were needed, so only two vegetation point and two prediction point files were created. Where a region bordered a neighboring state, vegetation and prediction points files

were created for Model C (all variables except geomorphology). Where a region bordered Canada, files for Model D (only terrain variables) were created. All sampling was done using the ArcGIS Spatial Analyst 'Extract Multi Values to Points' tool.

Software Platform

Two software platforms were considered for vegetation modeling, MaxEnt and R. Both software platforms are public domain. The modeling team determined that R would be the preferred software platform for this project because it would be more efficient. Procedures available in R allow all vegetation types to be modeled in a single model. MaxEnt procedures would have required a separate model for each vegetation type, with the final model consisting of a composite of the individual models. The MaxEnt procedures might have advantages over the R models. For example, MaxEnt might do a better job of modeling rare vegetation types. However, there was not sufficient time in this project to implement and test both methods. It should be noted that separate models for each vegetation type can also be implemented in R and can also use Random Forest procedures.

Preliminary Models

A test vegetation model was run for one region (BGWD) using preliminary data and procedures. Initially, there was some consideration that we might be able to transfer vegetation type data to soil polygons for mapping. Correlations between the vegetation and soil variables, among others, were assessed using the Spearman Rank correlation coefficients and mixed Principal Component Analysis (PCA). The Spearman Rank correlations between vegetation types and environmental variables (soil, terrain, landscape) proved generally weak, albeit significant ($p<0.05$). In general, vegetation classes showed stronger correlations (also significant) with environmental variables than vegetation types. For the vegetation type mixed PCA, the first two PCs explained close to 30% of the total variation. For the vegetation class PCA, the first two PCs explained close to 34% of the total variation. The Mixed PCA and the Spearman Rank Correlation results suggest that vegetation class is a more apt response variable than vegetation type for predictive modeling. Additionally, the Spearman Rank and mixed PCA results indicated no strong relationship between vegetation type/class and the environmental predictor variables, including soils. Because of the weak association with soils, the thought of assigning vegetation categories to soil polygons was discarded.

Initially, two predictive models were explored. The first is multinomial logistic regression (MLR), which is like logistic regression but predicts multi-class features (e.g. 14 vegetation classes) instead of binary the features of the archaeological predictive models (e.g. site versus no site). MLR is sensitive to skewed data distributions. Variables were transformed if they exhibited high skewness, which is typical for most environmental variables.

The MLR performance proved poor. Consequently an ensemble method called Random Forest (RF) was explored. RF is like bagging, which is the technique Gary Oehlert suggested as the preferred predictive statistical model for MnModel Phase 4 (Oehlert and Shea 2007). Based on the model performance criteria, the random forest model was selected for a 'first go' effort at predicting vegetation classes across the BGWD Region. The random forest model was applied to the predictor variables sampled at a 10x10 square meter grid resolution to predict BGWD vegetation classes. Provided the large amount of data, the computing time for running the predictive model was demanding and the modeling team determined not to use this surface for further modeling.

Though the preliminary models performed better for vegetation classes than for types, the modeling team determined that some distinction between classes would be necessary for the final models. The vegetation classification system adopted for modeling (Appendix A) combines wetland and prairie vegetation types into classes but maintains types for forests, woodlands, and savannas. This classification scheme may be reconsidered for future modeling, based on the results of the current models.

Statistical Analysis

Statistical procedures for creating this vegetation model are thoroughly documented in Landrum and Hobbs (2019). The key steps are summarized here.

Refine Dataset

Most multivariate models cannot handle observations with missing ('NA' or 'NULL') information. As such they simply ignore any record with a NULL observation, even if this record contains useful information for multiple other predictor variables. This can potentially result in a measurable reduction in database size, and therefore, useful information. As such, any predictor variable with a NULL frequency above 5% was removed; the 5% criterion is a 'rule of thumb' and can be adjusted. Predictor variables were also removed if they exhibited measurable collinearity and a near zero variance.

Categorical variables, such as landform and landscape, can also cause problems in the analysis. Because there are many more prediction points than vegetation points in every region, it is inevitable that some rare features represented in the prediction point data will be missing or poorly represented in the training data. Moreover, if specific categorical values (i.e. specific landforms) are rare in the training data, the statistical analysis will not have enough information to assess their association with vegetation types. To reduce these problems, we developed procedures to reclassify categorical variables within R to reduce observed imbalances.

Exploratory Data Analysis

The statistical procedures included a suite of standard analysis to describe the data. Environmental variables were summarized by vegetation types, histograms were constructed, and collinearity was measured using Spearman's Rand Correlation Coefficients and Variance Inflation Factor (VIF). Pearson's chi-square tests were performed to identify correlations between categorical variables and vegetation types.

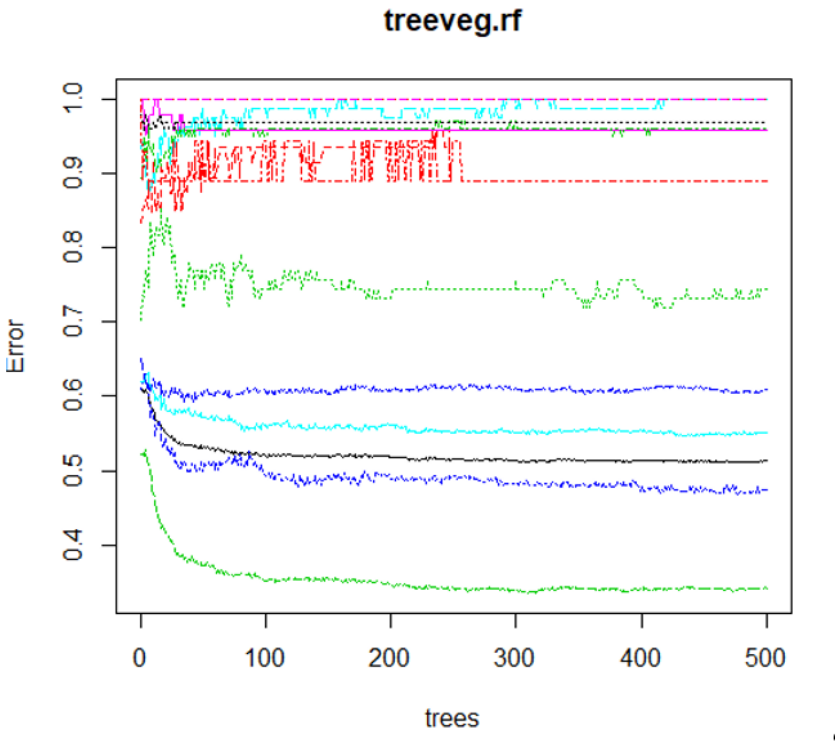
Modeling

After preliminary steps to refine and describe the data, the vegetation points were randomly divided into a training dataset (75 percent of the points) and a testing dataset (25 percent of the points). A random forest model was fit to the training dataset. The Random Forest procedure creates multiple tree models for the data set and calculates predictions based on the results of all the trees. For the vegetation models, the default value of 500 trees was used.

Model output included a graph of error vs. the number of trees built. An example of such a graph is provided in Figure 10. In this graph, results for each vegetation type in the modeled region are graphed in different colors. It is not possible to determine which lines represent which vegetation types. However, after examining the

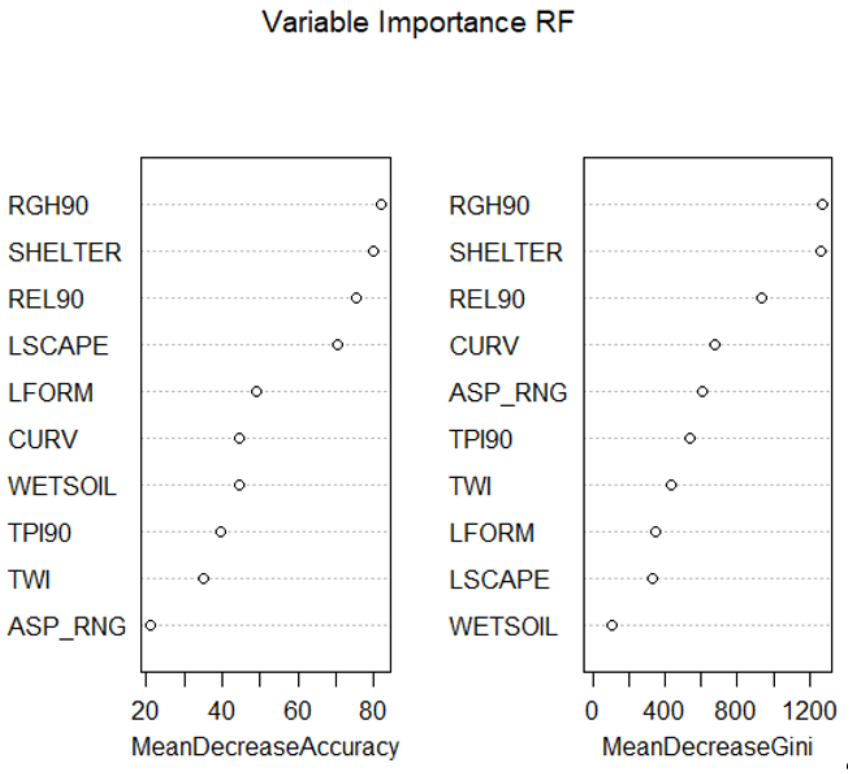
model evaluation results (see below), it is apparent that the vegetation types with the lowest error rates are those that are more dominant in the region, while the rare vegetation types have very high rates of modeling error.

Figure 10: Error vs. Number of Trees in a Random Forest Model



The second piece of very useful information produced by running the random forest model is the evaluation of variable importance, which is provided in both tabular and graphic formats. The table provides values for each predictor variable for each vegetation type in the model as well as values for the entire model. The graph (Figure 11) shows only the summary values for the model for two measures, the mean decrease in accuracy if the variable is removed and the mean decrease in 'node purity' if the variable is removed. In the example in Figure 11, the top four variables clearly play an important role in predicting vegetation distributions for the region.

Figure 11: Variable Importance Graphs from Random Forest



Model Evaluation

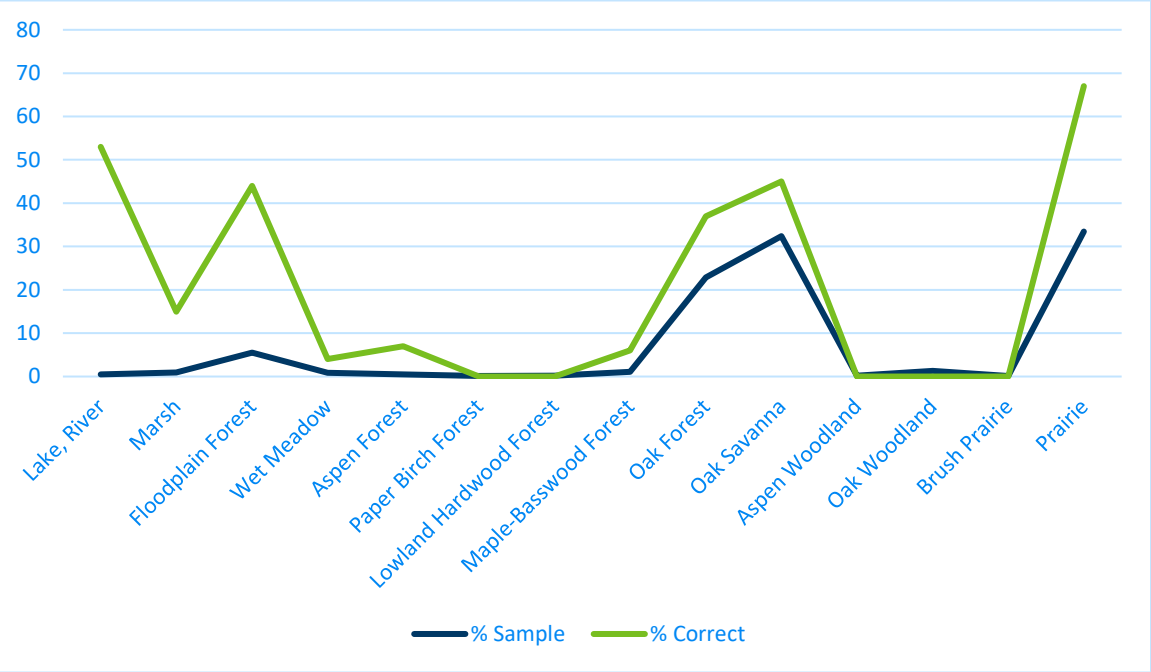
The performance of the random forest model is evaluated by applying the model built on the training data to the test data. This allows the construction of a confusion matrix (Figure 12). This matrix shows the predicted vegetation type for each actual ('Reference') vegetation type in the testing dataset.

Figure 12: Confusion Matrix Example

Prediction \ Reference	150	230	240	270	361	363	364	365	367	372	381	382	391	392
150	8	0	3	1	0	0	0	0	0	0	0	0	0	0
230	1	4	1	1	0	0	0	0	0	0	0	0	0	1
240	3	9	76	9	0	1	2	5	12	24	0	1	0	13
270	0	0	1	2	0	0	0	0	0	1	0	0	0	3
361	0	0	0	0	1	0	0	0	0	0	0	0	0	0
363	0	0	0	0	0	0	0	0	0	1	0	0	0	0
364	0	0	0	0	0	0	0	0	0	0	0	0	0	0
365	0	0	0	0	0	0	0	2	0	0	0	0	0	0
367	0	1	20	0	4	1	2	10	267	245	1	8	1	89
372	3	3	44	3	4	2	2	15	284	455	2	12	3	240
381	0	0	0	0	0	0	0	0	0	0	0	0	0	0
382	0	0	0	0	0	0	0	0	0	0	0	0	0	0
391	0	0	0	0	0	0	0	0	0	0	0	0	0	0
392	0	10	27	10	6	0	0	1	150	283	4	21	1	697

This matrix is easier to interpret if the correct predictions are converted to percentages of the actual vegetation types and compared to the proportion of each vegetation type in the test population (Figure 13). The total number of sample points is 3,118, which is quite a large test population. However, it is quite unbalanced. While it includes 1,043 Prairie points (33 percent of the test population), it has only four Paper Birch Forest points (0.12 percent). We can expect that these proportions were similar in the training data. Percentages of correct predictions for individual vegetation types range from 0 (Paper Birch Forest, Lowland Hardwood Forest, Aspen Woodland, Oak Woodland, Brush Prairie) to 67 (Prairie). The percentage of correct predictions closely tracks the percentage of the vegetation type in the population, with the exception of Lakes & Rivers, Marshes, and Floodplain Forests, which seem to be in sufficiently distinctive landscape positions in this region to be more readily predicted.

Figure 13: Percent of Correct Responses Compared to Percent Representation in Sample



The confusion matrix is used to calculate several performance measures for the model as a whole.

- Overall accuracy is defined as the percentage of points in the test population that are accurately predicted by the model. In the case of the example in Figures 12 and 13, the overall accuracy of the model is 0.4851, with a 95 percent confidence that it is between 0.4674 and 0.5028.
- The 'No Information Rate' (NIR) is the error rate when the input and output are independent. In the example above, the NIR is 0.3346
- Ideally, the NIR should be lower than the accuracy estimate (ACC). In the example above, the probability (P-Value) that ACC is greater than NIR is less than 2.2e-16.
- Cohen's Kappa: Kappa measures how well the model performed compared to how well it would have performed by chance. Kappa should be high if there is a large difference between accuracy and the NIR. In the above example, the Kappa value is low (0.2731).

We can see from these statistics that the accuracy of this model is greater than the error rate, but that the model as a whole is rather weak.

Apply Model to Prediction Points

After a random forest model is fit to the training data and evaluated by the testing data, the next step is to fit the model to the prediction points. In this step, the predicted vegetation types will be attached to the prediction points on the basis of the predictor variable values at each point. The predicted values and the x,y coordinates of the prediction points are then exported to a comma-delimited text file (.csv).

Import Model to ArcGIS

Each exported model file created from the prediction points (.csv format) is imported into a raster in ArcGIS using a customized tool (Landrum and Hobbs 2019). The resulting raster will contain only the modeled vegetation value for each cell.

Create Composite Models

There will be two to four different raster models for each region. Only one will have points for the entire region. One will be missing points where soil variable values are NULL. One may have all points except those outside of the state, as those outside the state will be missing geomorphic data. Finally, one may have only points inside the state that also have valid soils data. These must be combined into a composite model that gives priorities to values from the model created using the full suite of predictor variables and fills the gaps in that model with values from the other models as needed.

Incorporate Lakes and Rivers from Historic Hydrographic Model

The final step is to insert lakes and rivers from the MnModel Phase 4 historic hydrographic model into the composite vegetation model. This insures that the vegetation model and hydrographic model are consistent with one another. It also insures that the lake and river outlines in the vegetation model are based primarily on the PLS plat maps and not on predictions. Any cells predicted as lakes or rivers in the vegetation model that were not lakes and rivers in the hydrographic model are reclassified as 'wet land.'

Evaluate Statewide Model

Because the final models for each region are composites of several models, it is impossible to evaluate their overall accuracy based on the evaluation measures provided by R. To determine how well the final models perform, we mosaicked them to create a statewide model, sampled the statewide model with the original vegetation points, and created a confusion matrix to record the results.

Results

There are many ways to evaluate the results of this model. One is simply to consider whether the vegetation patterns appear to be reasonable, based on our understanding of Minnesota vegetation and the evidence of the PLS plat maps. Another is to compare the model to our previous reference for historic vegetation, the

Marschner Map (Marschner 1974). The primary goal of this project was to produce a model that better fit the terrain and displayed vegetation at a higher resolution than Marschner. We can also evaluate models for individual regions using the statistics produced by R. Using these statistics, we can compare the performance of the four types of models run in each region. Finally, we can consider the ability of the model to predict individual vegetation types as documented in the vegetation point sample.

Visual Evaluation of Historic Vegetation Model

A visual evaluation of this model is necessarily scale-dependent. On a general level, this is a reasonable model. The statewide patterns of vegetation are as we would expect (Figure 14). Prairies dominate the southern and western portions of the state. Coniferous forests dominate the northeast. A discontinuous band of deciduous forest separates these two zones. Swamps, primarily conifer swamps, are conspicuous in the north and northeast.

When viewed in detail (Figure 15), it is apparent that the vegetation model is very sensitive to terrain. In heavily urbanized areas, artifacts of the regular street grid are apparent. Elsewhere, wetlands are well-fit into their basins and terrestrial vegetation types clearly occupy the uplands. Still, even where the land is not urban, artifacts attributable to roads and other disturbance are apparent, as in the northeast quadrant of the right-hand map in Figure 15, where the boundary between 'swamp' and 'marsh' follows two roads.

This sensitivity to terrain is likely the cause of the high degree of 'pixelization' of the model. Isolated cells or small groups of cells of one or more vegetation types may be imbedded within larger areas dominated by a single consistent type. An extensive jack pine forest in Beltrami County, for example, is dotted with inclusions of red pine forest and mixed-pine-hardwood forest. Each of these inclusions may consist of only one to 20 cells. All bearing trees recorded in this area are jack pines. However, red pines and hardwoods exist in other parts of the region and the model has identified small areas within the jack pine forest that are more similar, with respect to their environmental variables, to these two vegetation types. Tiny inclusions like this could be removed from the model by merging them into the dominant surrounding vegetation. We decided, however, to leave them in as they may provide additional information of interest with respect to the availability of different habitat types. Moreover, such generalization is likely to cause the loss of some of the rare vegetation types where they should remain.

Figure 14: MnModel Phase 4 Historic Vegetation Model for Minnesota

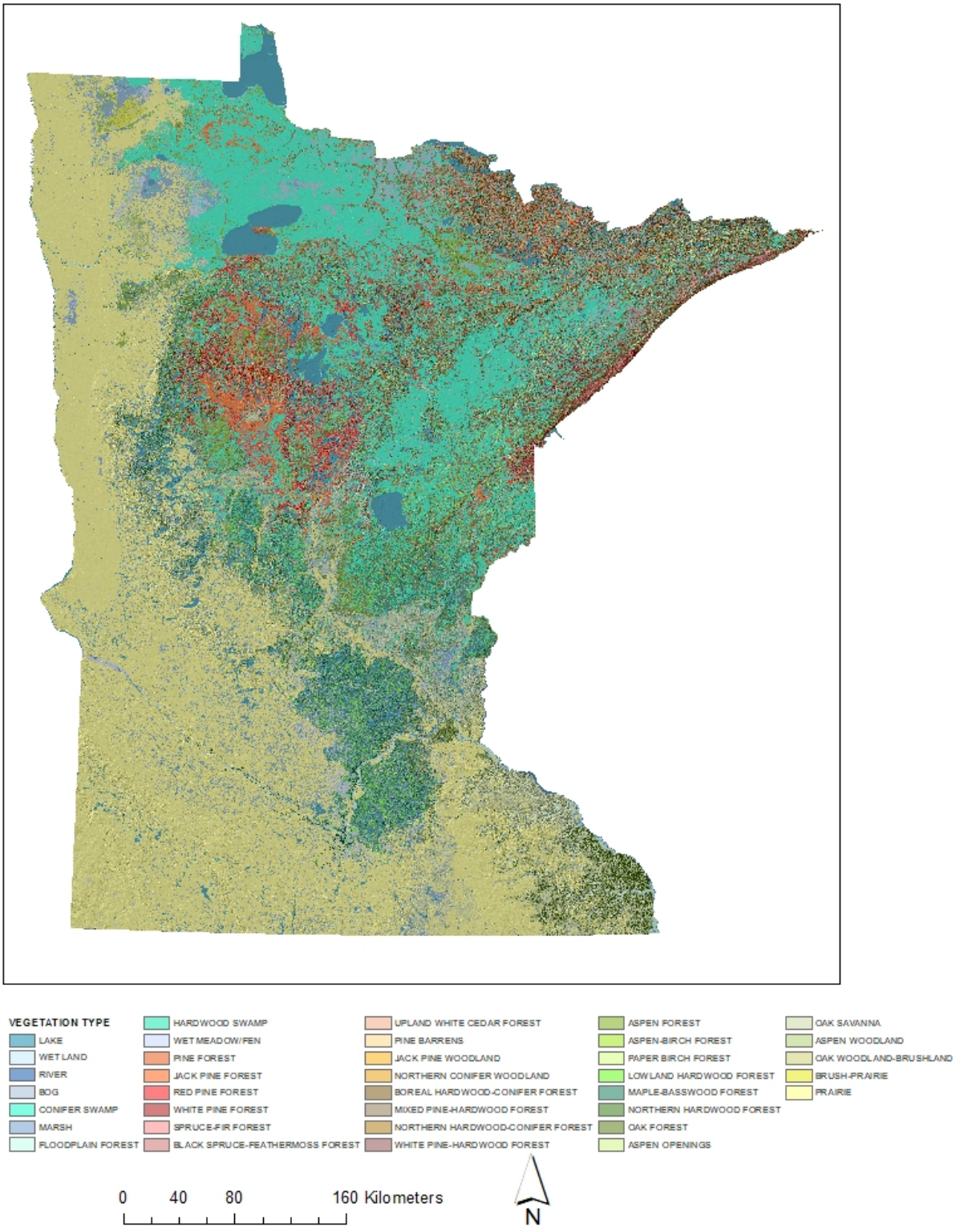
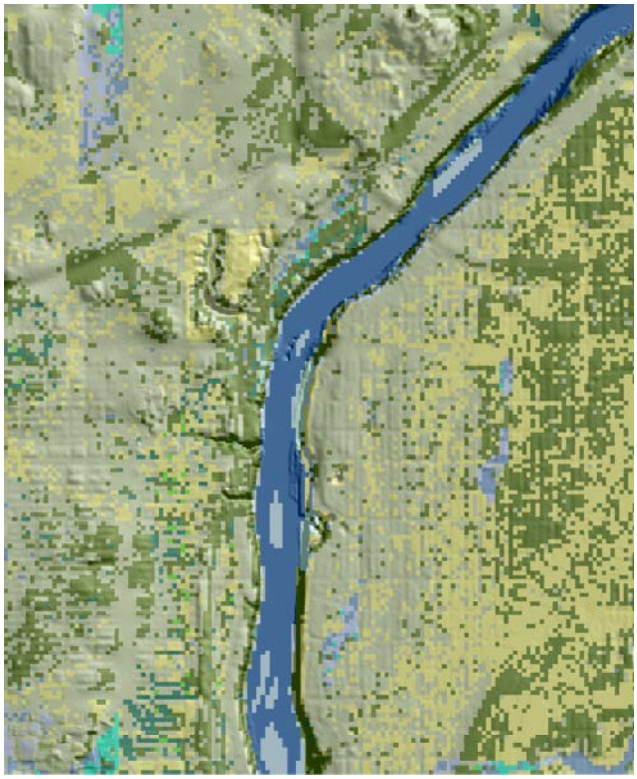
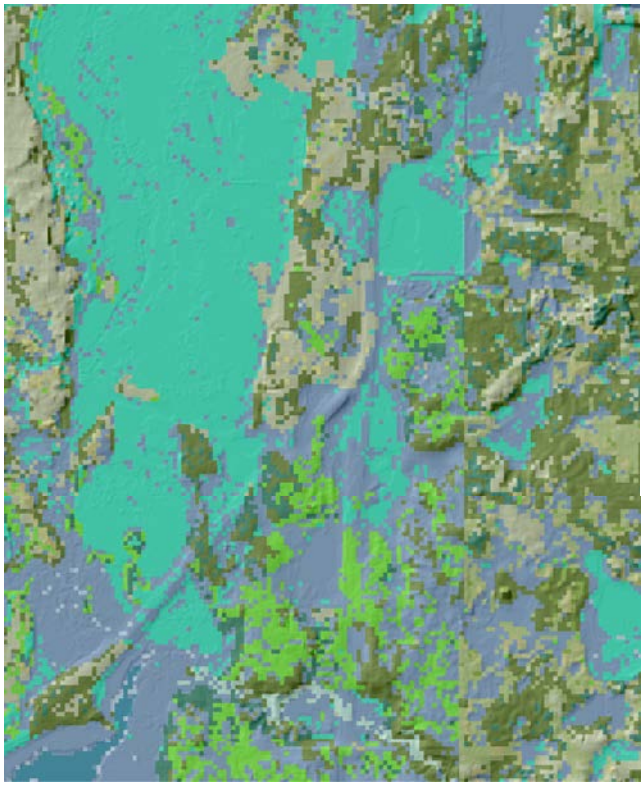


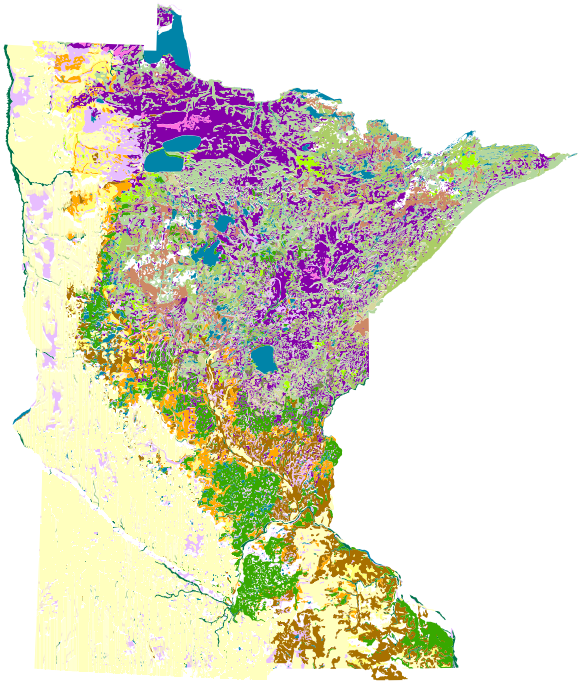
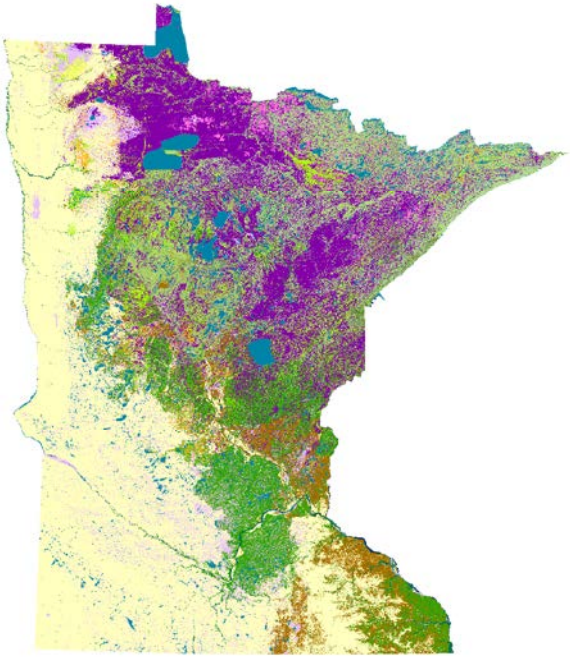
Figure 15: Sensitivity of Vegetation Model to Terrain

Urban Area: Hennepin and Ramsey Counties	Wetland/Forest: Anoka and Washington Counties
	

Comparison to Marschner Map

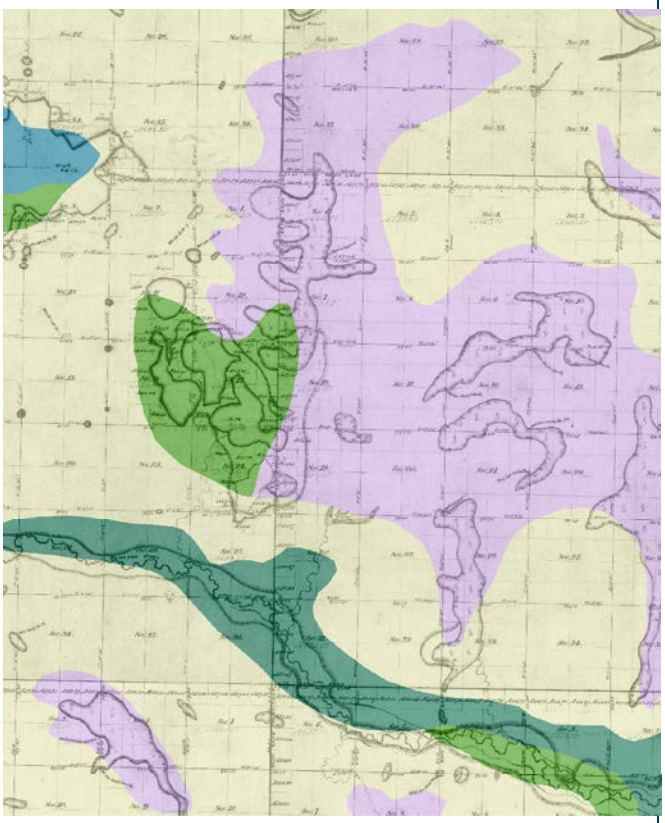
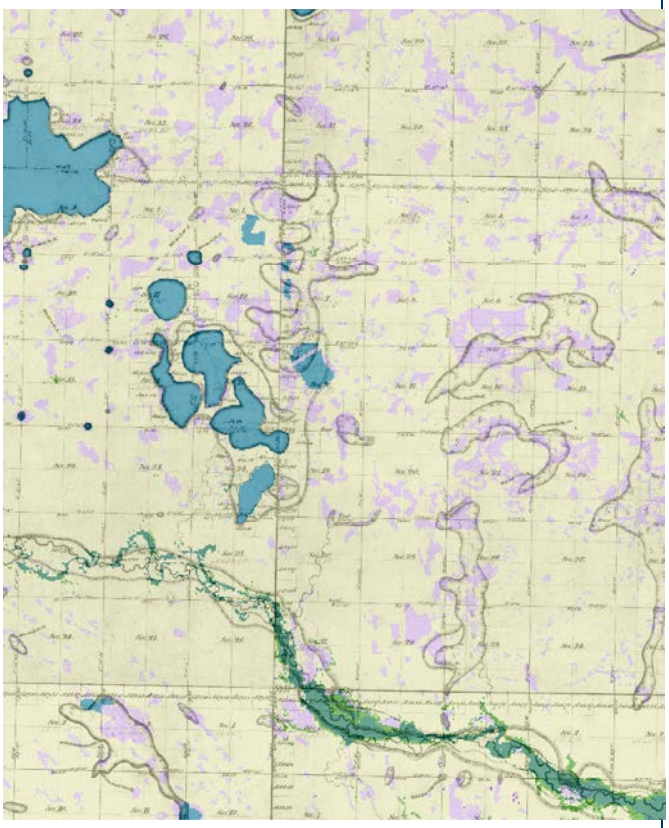
The primary goal for this model was to improve on the Marschner Map. We can visually compare the model to Marschner (Figure 16) by reclassifying the vegetation model to the same categories that Marschner used. This in itself is tricky, since Marschner has no coniferous forest categories that are not dominated by pine. By overlaying the classified vegetation points and Marschner, it was determined that most of the points MnDOT classed as non-pine coniferous forest types were classified by Marschner as 'Aspen-birch (conifers)', though many received other classes as well.

Figure 16: Comparison of the Vegetation Model to the Marschner Map

The Original Vegetation of Minnesota (after Marschner 1974)	MnModel Phase 4 Historic Vegetation Model
	

A side-by-side comparison of the two maps (Figure 16) show broad-scale similarities. When examined more closely, wet praries, marshes, and sloughs mapped by Marschner are larger than those mapped by the model, but not necessarily more realistic (Figure 17). Plat maps polygons for wetlands are obvious generalizations. Surveyors observed the wetlands only along the section lines, then generalized them into the sections, usually making efforts to connect wetlands that may not have actually had any connection. Marschner further compounded this problem by grouping individual wetlands into even larger polygons. The vegetation model, on the other hand, maps smaller, more discrete wetlands, though many more of them.

Figure 17: Comparison of Wetland Mapping to Public Survey Plat Maps

The Original Vegetation of Minnesota (after Marschner 1974)	MnModel Phase 4 Historic Vegetation Model
	

The vegetation model wetlands are a better fit for the topography (Figure 18). Not only do the marshes and small lakes fit their basins, but the floodplain forest occupies the floodplain and not the bluffs above the river.

Figure 18: Comparison of Wetland Mapping to Terrain

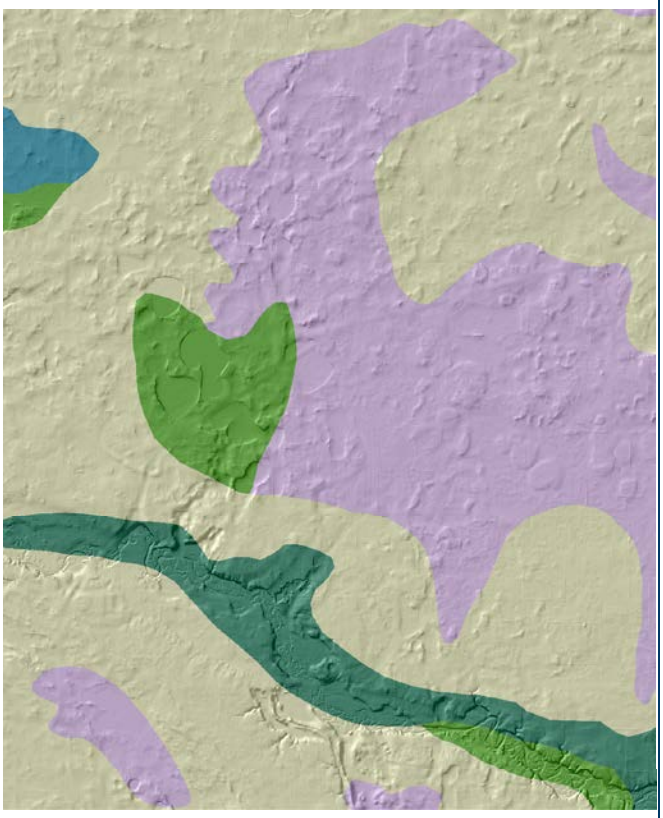
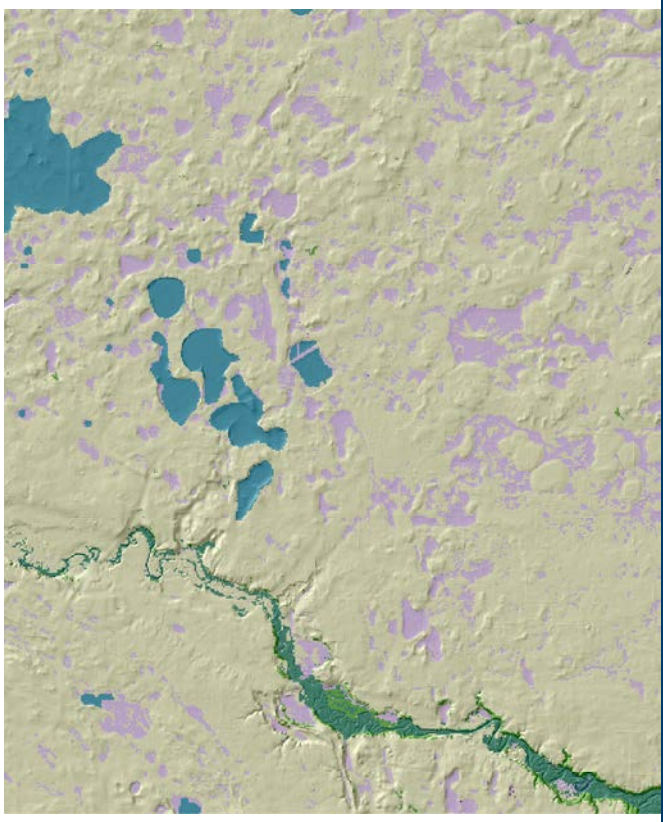
The Original Vegetation of Minnesota (after Marschner 1974)	MnModel Phase 4 Historic Vegetation Model
	

Figure 18 also illustrates one of the flaws of the vegetation model. It does a poor job of modeling rare vegetation types. Although it did suggest a narrow band of 'Big Woods' vegetation near the river, it completely misses the wooded land around the cluster of lakes. Marschner, on the other hand, may exaggerate these.

This difference is emphasized in Table 2, where we see the total area of each of Marschner's vegetation classes in the state, as mapped by the two models. Several 'rare' types (less than five percent of the state according to

Marschner) are even more rare in the vegetation model. These include aspen-oak woodland, brush-prairie, jack pine barrens and openings, mixed hardwood and pine forest, river bottom forest, and white pine forest.

Table 2: Percent Area of Comparable Vegetation Classes, Marschner vs. MnModel Phase 4 Historic Vegetation Model

Vegetation Class	Marschner	MnModel
Aspen-birch (hardwoods)	0.86	3.58
Aspen-birch (conifers)	11.89	7.58
Aspen-oak woodland	3.4	1.05
Big Woods	7.34	9.13
Brush-prairie	2.6	0.20
Conifer bogs and swamps	12.98	17.58
Jack pine barrens and openings	3.31	0.10
Mixed hardwood and pine	1.39	0.41
Oak openings and barrens	6.72	4.87
Open muskeg	0.26	1.21
Pine flats	0.09	0
Prairie	29.37	35.14
River-bottom forest	1.54	0.83
Water	4.51	6.09

Vegetation Class	Marschner	MnModel
Wet prairies, marshes and sloughs	7.58	6.67
White and Norway pine	5.88	4.47
White pine	0.29	1.09

On the other hand we also see (Table 2) the tendency of the vegetation model to overestimate the extent of dominant vegetation types (in particular, conifer bogs and swamps in the north and prairie in the south and west). This is an artifact of the statistical procedures, whereby the very large number of data points of dominant vegetation types overwhelm the analysis.

Other differences between Marschner's map and the vegetation model are attributable to the difficulty of fitting the vegetation model types into Marschner's classes. These are apparent in both Figure 16 and Table 2. In particular, Marschner's categorization of two types of 'aspen-birch' forest, one dominated by hardwoods and one by conifers, is difficult to reconcile with our classification scheme. We were forced to include all coniferous forest dominated by spruce or cedar in the 'aspen-birch (conifers)' category. Even with that, we do not approach the extent of this category that Marschner mapped. Since our area classified as 'aspen-birch (hardwoods)' is larger than Marschner's, it is likely we interpreted some of the area he associated with conifers as hardwood forest. On the other hand, the larger extent the vegetation model maps to 'open muskeg' is attributable to the fact that we did not separate open sphagnum bogs from black spruce bogs for modeling.

Performance of Models by Regions

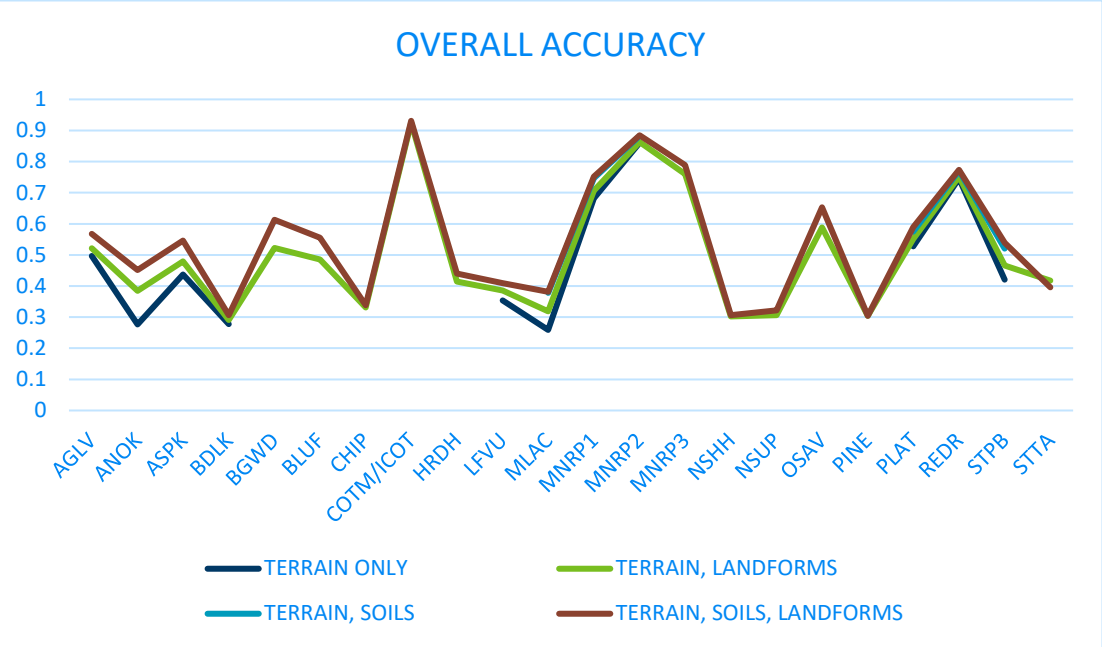
It is informative to look at the model evaluation measures provided by R for two reasons. First, these are the only evaluations conducted using a testing dataset that was not used to build the models. Second, they allow us to evaluate the relative performance of the four types of models.

Overall Accuracy

The number of models run varied between regions, depending primarily on whether the region included a state border. Model accuracy varies considerably between regions, ranging from 0.26 to > 0.93. Accuracy increases with more variables added to the models (Figure 19). When only terrain variables were used, accuracy was consistently lower than when terrain and landform variables were used. Likewise, when terrain, landform, and soil variables were used accuracy was highest. However, this difference within regions does not approach the degree of difference in accuracy between regions.

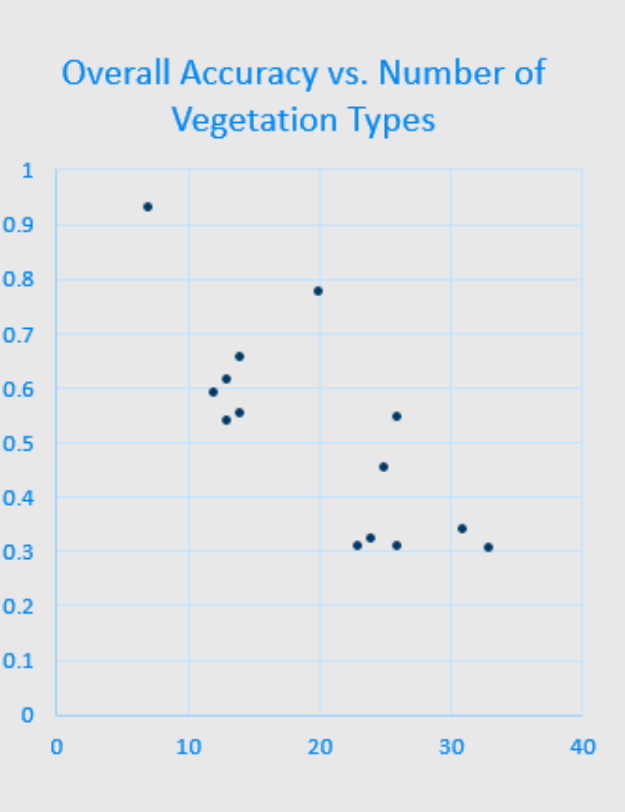
Accuracy varies more between regions than between individual models within a single region (Figure 19). The highest overall accuracy (0.93) was achieved in the combined COTM/ICOT region where prairie was very strongly dominant. MNRP and REDR models were also highly accurate, apparently for the same reason.

Figure 19: Model Accuracy by Modeling Region



There seems to be an inverse relationship between accuracy and the number of vegetation types modeled, particularly when the number of variables is controlled for. However, this relationship appears to be weak (Figure 20), and imbalance between classes is likely to be more important. Where one vegetation type is represented by many more data records than any other, it dominates the analysis and the resulting prediction. There are several ways to mitigate this problem in the future, as discussed below.

Figure 20: Relationship of Overall Accuracy to Number of Modeled Vegetation Types



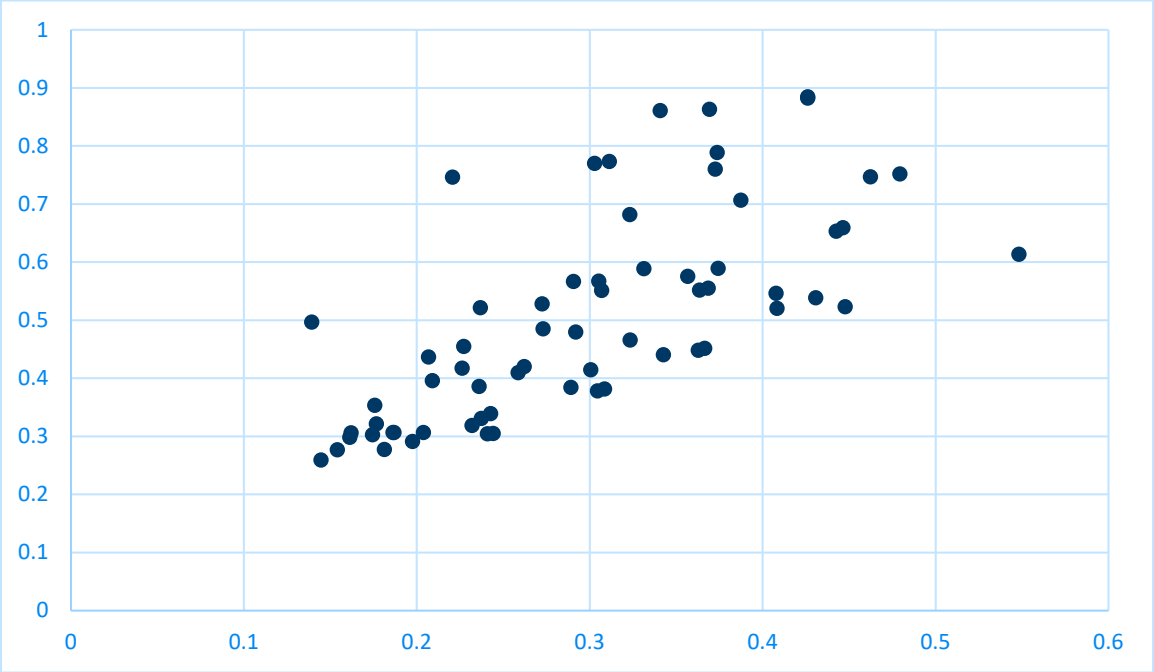
No Information Rate

The No Information Rate (NIR) or error rate on these models ranges from 0.1432 to 0.9216. The NIR is always lower than the overall accuracy, as it should be, and this difference is always significant.

Kappa

Cohen's Kappa for these models ranges from 0.1392 to 0.5482. These values are rather low, indicating that the models are performing only a bit better than by chance. There is a general tendency for more accurate models to have higher Kappa values (Figure 21), as should be expected.

Figure 21: Relationship Between Overall Accuracy and Kappa Values



Statewide Performance by Vegetation Types

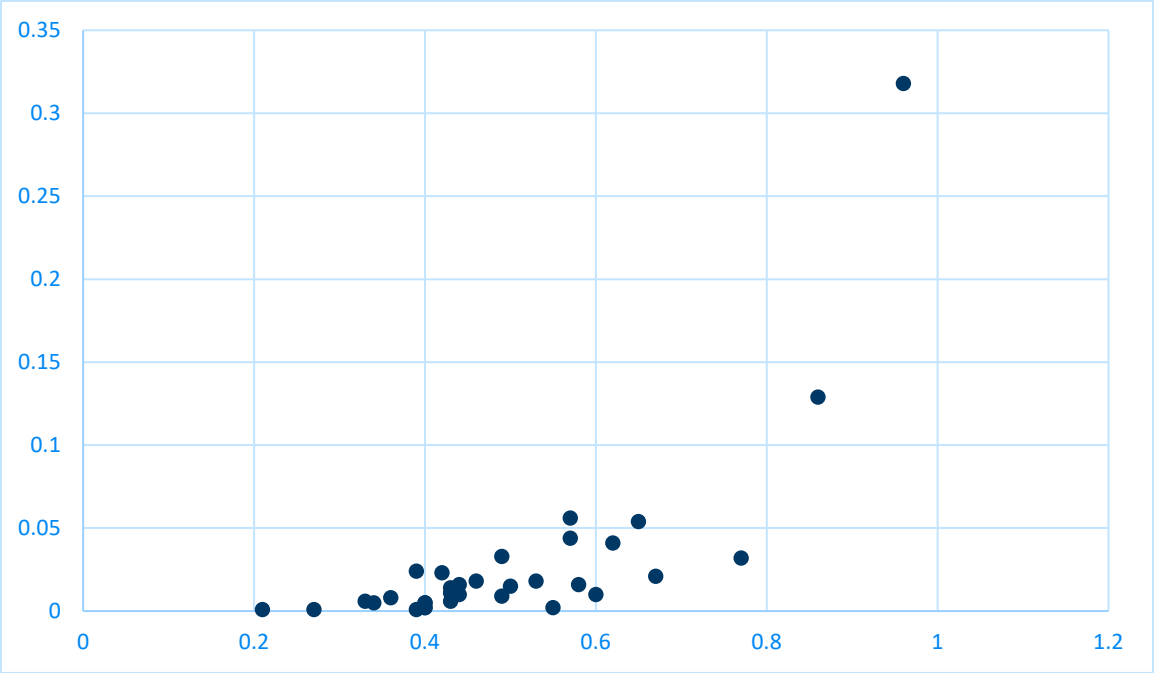
Performance measures for individual vegetation types are detailed in Appendix B of this report. These are summarized in this section.

Accuracy of Predictions

For our purposes, we are defining accuracy as the percentage of vegetation points of a given type that were accurately predicted by the model. This was determined by sampling the statewide model values with the vegetation point data used to build the model and counting the number of points of each vegetation type that had the same value in both the original data and the model. Accuracy values ranged from 21% (Aspen Openings) to 96% (Prairie). Lakes also have high accuracy values (93%), but this should be attributed to the fact that lake polygons are displayed on the plat maps, providing corroborating evidence for the classification of the points. 'Wet Land' has an accuracy of 0%, but that is only because it is not a modeled vegetation type (i.e. one represented in the data) but a classification after the fact of cells predicted to be lakes or rivers but not coinciding with lakes and rivers on the plat maps. With lakes, rivers, and 'wet land' excluded, 72 percent of all vegetation points statewide are correctly predicted. In comparison, only 49 percent of our vegetation point classifications were predicted by the Marschner map.

Figure 22 illustrates how accuracy (x axis) increases with the percent representation of the vegetation type in the sample (y axis). Prairie is strongly dominant in the sample (32 percent of the data) and has by far the highest prediction rate. The next most dominant vegetation type is Conifer/Shrub Swamp (13 percent of the data), and it has the second highest accuracy value (86 percent). All other vegetation types are each represented by less than 10 percent of the sample and are correspondingly less well predicted.

Figure 22: Relationship Between Percent of Total Sample and Accuracy for Modeled Vegetation Types (Excluding Lakes and Rivers)



Confidence in Model Classification

Interestingly, the percentage of vegetation points of each type that are accurately predicted may not be the best measure of confidence in the model classifications. It may be more relevant for the end user to consider the percentage of the model cells of each type that are accurately classified. We can estimate this by calculating the percentage of correct predictions of model cells intercepted by sample points. Overall, 73 percent of the sampled cells correctly predicted the vegetation type.

Although the statewide confidence value is quite similar to model accuracy, the values can be quite different for individual vegetation types because proportions of the predicted vegetation types in the model differ from the proportions of the vegetation types in the sample points. Strongly dominant vegetation types (prairie and conifer/shrub swamp) are overrepresented in the model. This increases the number of incorrect predictions for these classes. While data points known to be prairie are accurately predicted 93 percent of the time, only 88 percent of the points predicted to be prairie were actually prairie. Likewise, while conifer/shrub swamp was accurately predicted 86 percent of the time, only 66 percent of the points predicted to be swamp were actually swamp.

On the positive side, confidence in rare vegetation types is more likely to be higher than accuracy. Bog, for example, has an accuracy of only 42 percent but a confidence value of 72 percent. Accuracy and confidence values for all vegetation types are presented in Appendix B.

Key Variables

Table 3 summarizes the performance of the environmental variables. The measure of a variables' performance is the percent increase in mean square error (%IncMSE) that would be observed if the variable were to be removed from the model. Another measure of performance would simply be the number of models in which a variable appears, since not all variables are used in all models.

Table 3: Performance of Environmental Variables in Models Using All Variables

Variable	Definition	# Models	Average %IncMSE	Maximum %IncMSE
ASP_RNG	Aspect Range	22	25.3	41.88
AWS150	Available Water Storage	22	51.5	86.46
CAC03	Calcium Carbonate	6	22.6	32.24
CEC7	Cation Exchange Capacity	4	35.4	52.47
CLAY	Percent Clay	6	31.4	52.65
CURV	Surface Curvature	20	36.5	59.14
DRAIN	Soil Drainage	14	29.5	55.38
ELEV	Elevation	21	92.3	175.85
FFD_R	Frost-Free Days	22	44.7	91.64
FLDFRQD	Flooding Frequency	4	25.5	26.38
GRTGRP	Great Group	20	47.5	100.11
GYPSUM	Gypsum	0	0	0
HYDGRPDCD	Hydrologic Group	22	40.6	242.81

Variable	Definition	# Models	Average %IncMSE	Maximum %IncMSE
HZDEP	Depth of the Surface Horizon	14	39.1	60.66
LFORM	Landform	22	54.1	93.93
LSCAPE	Landscape	22	53.1	103.52
OM	Percent Organic Matter	14	29.3	40.76
PI	Productivity Index	15	32.9	54.19
REG_RICH	Regime Richness	12	21.6	43.57
REG_WET	Regime Wetness	9	34.2	59.19
REL90	Relative Elevation within 90 Meters	16	52.0	69.39
RGH90	Surface Roughness within 90 Meters	4	59.7	70.22
SAND	Percent Sand	5	38.2	51.65
SHELTER	Shelter Index	22	58.8	83.52
SILT	Percent Silt	2	32.7	34.13
SLOPE	Percent Slope	15	58.2	76.88
TPI90	Topographic Position within 90 Meters	22	33.5	52.25
TWI	Topographic Wetness Index	22	39.1	67.09

Improving the Model

With 72 percent accuracy and 73 percent confidence in the model predictions, this historic vegetation model performs very well. However, there are several ways that it can be improved.

Improve the Data

Improving the data used to build the model would be the most time-consuming way to improve the model. The first step would be to make better use of the transcribed line notes that have been digitized for the northern part of the state. These should be used more consistently to verify and correct values of the section corner points used for modeling. Unfortunately, these notes have not been transcribed for the southern part of the state. Doing so would improve the model by providing additional information about wetlands and rare vegetation types.

Ideally, we should improve the georeferencing of the digital GLO Plat Map. For most of Minnesota, the plat maps were georeferenced using only township corners. Section corners and lines are not always in their true locations. This causes offsets in the lakes, wetlands, and other polygons mapped. If this can be accomplished, naturally the digitized polygons would also need to be corrected. This, however, would be very time-consuming and may not have a great effect on the model itself.

Analysis of Bearing Tree Distributions

We need to make better use of the bearing tree data. It would be possible to analyze species distributions, species associations, and tree spacing. Having a better understanding of these aspects of tree distributions could help make decisions about vegetation classes.

Balance Vegetation Classes

We need to create more 'balanced' classes for statistical analysis. The statistical software cannot do a good job of predicting rare vegetation types, and vegetation classes that are exceedingly dominant (for example, prairie in southwestern Minnesota) completely swamp the analysis. Several steps can be taken to achieve more balance:

- Create more points for rare vegetation types from the line notes and from GLO map polygons.
- Reduce the number of points of dominant vegetation types.
- Combine vegetation types into larger categories (somewhere between TYPE and CLASS). It may help to refer to the confusion matrix results to determine which vegetation types are most often confused for each other in each region.

Remove Lakes and Rivers

Since lakes and rivers will be added to the model from the historic hydrographic model, we do not need to 'predict' their locations. Without the 'lake' and 'river' categories, we should then get wetland or other vegetation types assigned to the areas that are now classified as 'wet land', which we know from the plat maps were not lakes or rivers.

Remove 10 km Buffer

The 10 km buffer around each region was needed so that vegetation diversity could be calculated for the archaeological predictive model. However, the buffer introduces different ecological types into the analysis that may confuse the model. Modeling without the 10 km buffer may produce better models since the vegetation patterns within the region are assumed to be more alike than the vegetation patterns outside the region.

Implement Jack-knife Procedures

Only one model of each type was run for each region. This model used only 75% of the data, and the other 25% was reserved for testing the model. We should run four models, each using a different 75% of the data. These models would be tested with the reserved data. Those tests would provide information about model performance. Finally, we should construct a model using 100% of the data. We can expect that model to perform at least as well, and probably better, than the first four models. However, we would not be able to test the model in the same way.

Conclusions

This was a first attempt to use statistical tools and PLS data to model historic vegetation in Minnesota. The two goals of the project were to improve upon the Marschner map and to prove the concept of statistical vegetation modeling using PLS data. Both were successful. The historic vegetation model is more than adequate for our current needs. It improves on Marschner in its correspondence to terrain and its accuracy. The statistical modeling procedures worked well and are an efficient way to create this type of map. Moreover, the results show where we have opportunities to improve the model.

References

Aaseng, Norman E. et al.
1993 *Minnesota's Native Vegetation: A Key to Natural Communities. Version 1.5*. Minnesota Department of Natural Resources, Natural Heritage Program. St. Paul, MN.

Anderson, Roger C. and M. Rebecca Anderson
1975 The presettlement vegetation of Williamson County, Illinois. *Castanea* 40(4): 345-363.

Bolliger, Janine, Lisa A. Schulte, Sean N. Burrows, Theodore A. Sickley, and David J. Mladenoff
2004 Assessing ecological restoration potentials of Wisconsin (U.S.A.) using historical landscape reconstructions. *Restoration Ecology* 12(1): 124-142.

Bourdo, Eric A. Jr.
1956 A review of the General Land Office Survey and of its use in quantitative studies of former forests. *Ecology* 37(4): 754-768.

Brown, Andrew, Alexander Anton, Luke Burds, and Elizabeth Hobbs
2019a *Tool Handbook*. Appendix C in *MnModel Phase 4 User Guide*, by Carla Landrum et al. Minnesota Department of Transportation. St. Paul, MN.

Brown, Daniel G.
1998 Mapping historical forest types in Baraga County Michigan, USA as fuzzy sets. *Plant Ecology* 134(1): 97-111.

Canham, Charles D. and Orie L. Loucks
1984 Catastrophic windthrow in the presettlement forests of Wisconsin. *Ecology* 65(3): 803-809.

Cleland, D.T., P.E. Avers, W.H. McNab, M.E. Jensen, R.G. Bailey, T. King, and W.E. Russell
1997 National Hierarchical Framework of Ecological Units. In *Ecosystem Management Applications for Sustainable Forest and Wildlife Resources,* edited by M.S. Boyse and A. Haney, pp. 181-200. Yale University Press, New Haven, CT.

Delcourt, Hazel R. and Paul A. Delcourt
1996 Presettlement landscape heterogeneity: Evaluating grain of resolution using General Land Office Survey data. *Landscape Ecology* 11(6): 363-381.

Gibbon, Guy E., Craig M. Johnson, and Elizabeth Hobbs
2002 Minnesota's Environment and Native American Culture History. Chapter 3 in *Mn/Model Final Report Phases 1-3*. Minnesota Department of Transportation. St. Paul, MN.

Grimm, Eric C.
1984 Fire and other factors controlling the Big Woods vegetation of Minnesota in the mid-nineteenth century. *Ecological Monographs* 54(3): 291-311.

Guisan, A., S.B., Weiss, and A.D. Weiss
1999. GLM versus CCA spatial modeling of plant species distribution. *Plant Ecology* 143: 107-122.

Hammer, John
1993 A New Predictive Site Location Model for Interior New York State. *Man in the Northeast* 45:39-76.

Hanron, Kevin F.
1981 *A Reconstruction of the Black Woods of Wisconsin, 1847-1849*. M.A. Thesis, Department of Geography, University of Minnesota. Minneapolis.

Hanson, D.H. and B.C. Hargrave
1996 Development of a Multilevel Ecological Classification System for the State of Minnesota. *Environmental Monitoring and Assessment* 39:75-84.

Heinselman, Miron L.
1973 Fire in the virgin forests of the Boundary Waters Canoe Area, Minnesota. *Quaternary Research* 3: 329-382.

Hobbs, Elizabeth
2019 *MnModel Phase 4: Project Summary and Statewide Results*. Minnesota Department of Transportation. St. Paul, MN.

Hobbs, Elizabeth, Andrew Brown, Alexander Anton, and Luke Burds
2019 *Historic/Prehistoric Hydrographic Models for Minnesota: MnModel Phase 4*. Minnesota Department of Transportation. St. Paul, MN.

Hobbs, Elizabeth, Jeffrey Walsh and Curtis M. Hudak
2019 *Environmental Variables: MnModel Phase 4*. Minnesota Department of Transportation. St. Paul, MN.

Hobbs, Elizabeth, Craig M. Johnson, Guy E. Gibbon, Carol Sersland, Mark Ellis, and Tatiana Nawrocki.
2002 Model Results and Interpretations. Chapter 8 in *Mn/Model Final Report Phases 1-3*. Minnesota Department of Transportation. St. Paul, MN.

Hudak, G. Joseph, Elizabeth Hobbs, Allyson Brooks, Carol Ann Sersland, and Crystal Phillips, Eds.
2002 *Mn/Model Final Report Phases 1-3*. Minnesota Department of Transportation. St. Paul, MN.

Kvamme, Kenneth L. and Timothy A. Kohler
1988 Geographic Information Systems: Technical Aids for Data Collection, Analysis and Display. In *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling,* edited by W. James Judge and Lynne Sebastian, 493-548. U.S. Government Printing Office, Washington, D.C.

Landrum, Carla and Elizabeth Hobbs
2019 *Vegetation Modeling User's Guide: MnModel Phase 4*. Minnesota Department of Transportation. St. Paul, MN.

Manies, Kristen L. and David J. Mladenoff
2000 Testing methods to produce landscape-scale presettlement vegetation maps from the U.S. public land survey records. *Landscape Ecology* 15: 741-754.

Marschner, Francis J.
1974 *The Original Vegetation of Minnesota.* Compiled from U.S. General Land Office Survey notes. North Central Forest Experiment Station, Forest Service, U.S. Department of Agriculture.

Mladenoff, David J., Sally E. Dahir, Eric V. Nordheim, Lisa A. Schulte, and Glenn G. Guntenspergen
2002 Narrowing historical uncertainty: probabilistic classification of ambiguously identified tree species in historical forest survey data. *Ecosystems* 5: 539-553.

Oehlert, Gary W. and Brian Shea
2007 *Statistical Methods for MnModel Phase 4: Final Report*. Research Services Section, Minnesota Department of Transportation. St. Paul, MN.

Schaetzl, Randall J., Frank J. Krist, Jr., Kristine Stanley, and Christina M. Hupy
2009 The natural soil drainage index: an ordinal estimate of long-term soil wetness. *Physical Geography* 30:383-409.

Spurr, Stephen H.
1954 The forests of Itasca in the nineteenth century as related to fire. *Ecology* 35(1): 21-25.

Stark, Stacey L., Patrice M. Farrell, and Susan C. Mulholland
2008 *Methods to Incorporate Historic Surface Hydrology Layer in Mn/Model [Phase 4] Using Existing Geographic Information System Data*. Minnesota Department of Transportation. St. Paul, MN.

Stewart, Lowell O.
1935 *Public Land Surveys: History, Instructions, Methods*. Collegiate Press, Inc. Ames, Iowa.

Vogl, Richard J.
1964 Vegetational history of Crex Meadows, a prairie savanna in northwestern Wisconsin. *The American Midland Naturalist* 72(1): 157-175.

Appendices

Appendix A: Vegetation Classification

Table A1 outlines the vegetation classification scheme used in MnModel Phase 4. The first three columns (SYSTEM, CLASS, and TYPE) were used to classify vegetation point data prior to modeling. Because the Public Land Surveyor's vegetation descriptions and the subsequent MnDNR classification of vegetation from those descriptions do not conform to this classification scheme (Aaseng et al. 1993), a considerable amount of reinterpretation was required.

- All points falling in plat map lakes and ponds were classified as the type 'LAKE BED'. In 60 cases, MnDNR recorded adjacent terrestrial or palustrine vegetation types at corners that were within lakes. MnDNR also classified 24 points as lakes that were not within lakes. Since lakes and ponds are readily identified on the plat maps, it was possible to verify and correct the point values using a simple overlay procedure in ArcGIS.
- Of 1,123 points coded as rivers by MnDNR, 329 were reclassified to lacustrine, palustrine, or terrestrial categories. Most of these were on floodplains, though some were on stream terraces and uplands. Only eight points not coded as rivers by MnDNR intersected major rivers on the plat maps.
- 'PRAIRIE' was rather straightforward. Most points previously classified as such were also at corners lacking bearing trees. Prairie points were reclassified only if bearing trees were present or if the point fell into a mapped wetland.
- Distinctions between savanna and forest were based, first, on the terms used by the surveyor in the line notes. If the line notes were not available, the spacing and species composition of the bearing trees were considered.
- Distinctions between different types of forest were based on combinations of bearing and line note tree species.
- Several types of forest may contain the combination of maple, basswood, and elm. Distinctions were made based on soils and geomorphic data as follows:
    - Floodplain Forest: On a floodplain and flooding frequency was not 'None'.
    - Lowland Hardwood Forest: Not on a floodplain but with a high water table.
    - Maple-Basswood Forest: On well-drained soils.
- If described as a 'thicket' by surveyors or classified as 'thicket' by MnDNR (where no line notes available), points were classified as a type of woodland depending on the dominant species present.

- Wetland types were based on surveyors' descriptions where available as well as any bearing or line note trees present. Ultimately, though, swamp types were generally combined for modeling, as shrub swamp and hardwood swamp in particular tended to be rather rare.

MODTYPE (Table A1) indicates the vegetation type used as model input. In most cases, these are the same as the originally classified type. However, it was necessary to combine some rare vegetation types into larger categories for statistical analysis. VALUE is the numeric code assigned to the MODTYPE value, as the statistical software can only interpret numeric values.

Table A1: MnModel Phase 4 Vegetation Classification System

SYSTEM	CLASS	TYPE	MODTYPE	VALUE
LACUSTRINE	LAKE BED	LAKE BED	LAKE (from HYDMOD)	100
RIVERINE	RIVER BED	RIVER BED	RIVER (from HYDMOD)	200
PALUSTRINE	WET LAND	WET LAND	WET LAND (predicted as lakes & rivers by VEGMOD but not by HYDMOD)	150
PALUSTRINE	BOG	BLACK SPRUCE BOG	BOG	210
PALUSTRINE	BOG	OPEN SPHAGNUM BOG	BOG	210
PALUSTRINE	CONIFER SWAMP FOREST	BLACK SPRUCE SWAMP	CONIFER SWAMP	220
PALUSTRINE	CONIFER SWAMP FOREST	TAMARACK SWAMP	CONIFER SWAMP	220
PALUSTRINE	CONIFER SWAMP FOREST	WHITE CEDAR SWAMP	CONIFER SWAMP	220

SYSTEM	CLASS	TYPE	MODTYPE	VALUE
PALUSTRINE	EMERGENT MARSH	MARSH	MARSH	230
PALUSTRINE	FLOODPLAIN FOREST	FLOODPLAIN FOREST	FLOODPLAIN FOREST	240
PALUSTRINE	HARDWOOD SWAMP FOREST	BLACK ASH SWAMP	HARDWOOD SWAMP (usually combined with CONIFER SWAMP)	250 (220)
PALUSTRINE	HARDWOOD SWAMP FOREST	MIXED HARDWOOD SWAMP	HARDWOOD SWAMP (usually combined with CONIFER SWAMP)	250 (220)
PALUSTRINE	SHRUB SWAMP	ALDER SWAMP	SHRUB SWAMP (usually combined with CONIFER SWAMP)	260 (220)
PALUSTRINE	SHRUB SWAMP	SHRUB SWAMP	SHRUB SWAMP (usually combined with CONIFER SWAMP)	260 (220)
PALUSTRINE	SHRUB SWAMP	WILLOW SWAMP	SHRUB SWAMP (usually combined with CONIFER SWAMP)	260 (220)
PALUSTRINE	WET MEADOW/FEN	WET BRUSH-PRAIRIE	WET MEADOW/FEN	270
PALUSTRINE	WET MEADOW/FEN	WET MEADOW	WET MEADOW/FEN	270
PALUSTRINE	WET MEADOW/FEN	WET MEADOW/FEN	WET MEADOW/FEN	270
PALUSTRINE	WET MEADOW/FEN	WET PRAIRIE	WET MADOW/FEN	270

SYSTEM	CLASS	TYPE	MODTYPE	VALUE
PALUSTRINE	WET MEDOW/FEN	FEN	WET MEADOW/FEN	270
TERRESTRIAL	CONIFEROUS FOREST	PINE FOREST	PINE FOREST	310
TERRESTRIAL	CONIFEROUS FOREST	JACK PINE FOREST	JACK PINE FOREST	311
TERRESTRIAL	CONIFEROUS FOREST	RED PINE FOREST	RED PINE FOREST	312
TERRESTRIAL	CONIFEROUS FOREST	WHITE PINE FOREST	WHITE PINE FOREST	313
TERRESTRIAL	CONIFEROUS FOREST	SPRUCE-FIR FOREST	SPRUCE-FIR FOREST	321
TERRESTRIAL	CONIFEROUS FOREST	BLACK SPRUCE-FEATHERMOSS FOREST	BLACK SPRUCE-FEATHERMOSS FOREST	322
TERRESTRIAL	CONIFEROUS FOREST	UPLAND WHITE CEDAR FOREST	UPLAND WHITE CEDAR FOREST	323
TERRESTRIAL	CONIFEROUS SAVANNA	PINE BARRENS	PINE BARRENS	330
TERRESTRIAL	CONIFEROUS WOODLAND	JACK PINE WOODLAND	JACK PINE WOODLAND	341
TERRESTRIAL	CONIFEROUS WOODLAND	NORTHERN CONIFER WOODLAND	NORTHERN CONIFER WOODLAND	342
TERRESTRIAL	MIXED CONIFEROUS-DECIDUOUS FOREST	BOREAL HARDWOOD-CONIFER FOREST	BOREAL HARDWOOD-CONIFER FOREST	351

SYSTEM	CLASS	TYPE	MODTYPE	VALUE
TERRESTRIAL	MIXED CONIFEROUS-DECIDUOUS FOREST	MIXED PINE-HARDWOOD FOREST	MIXED PINE-HARDWOOD FOREST	352
TERRESTRIAL	MIXED CONIFEROUS-DECIDUOUS FOREST	NORTHERN HARDWOOD-CONIFER FOREST	NORTHERN HARDWOOD-CONIFER FOREST	353
TERRESTRIAL	MIXED CONIFEROUS-DECIDUOUS FOREST	WHITE PINE-HARDWOOD FOREST	WHITE PINE-HARDWOOD FOREST	354
TERRESTRIAL	DECIDUOUS FOREST	ASPEN FOREST	ASPEN FOREST	361
TERRESTRIAL	DECIDUOUS FOREST	ASPEN-BIRCH FOREST	ASPEN-BIRCH FOREST	362
TERRESTRIAL	DECIDUOUS FOREST	PAPER BIRCH FOREST	PAPER BIRCH FOREST	363
TERRESTRIAL	DECIDUOUS FOREST	LOWLAND HARDWOOD FOREST	LOWLAND HARDWOOD FOREST	364
TERRESTRIAL	DECIDUOUS FOREST	MAPLE-BASSWOOD FOREST	MAPLE-BASSWOOD FOREST	365
TERRESTRIAL	DECIDUOUS FOREST	NORTHERN HARDWOOD FOREST	NORTHERN HARDWOOD FOREST	366
TERRESTRIAL	DECIDUOUS FOREST	OAK FOREST	OAK FOREST	367
TERRESTRIAL	DECIDUOUS SAVANNA	ASPEN OPENINGS	ASPEN OPENINGS	371
TERRESTRIAL	DECIDUOUS SAVANNA	OAK SAVANNA	OAK SAVANNA	372

SYSTEM	CLASS	TYPE	MODTYPE	VALUE
TERRESTRIAL	DECIDUOUS WOODLAND	ASPEN WOODLAND	ASPEN WOODLAND	381
TERRESTRIAL	DECIDUOUS WOODLAND	OAK WOODLAND-BRUSHLAND	OAK WOODLAND	382
TERRESTRIAL	UPLAND BRUSH-PRAIRIE	MESIC BRUSH-PRAIRIE	BRUSH-PRAIRIE	391
TERRESTRIAL	UPLAND PRAIRIE	PRAIRIE	PRAIRIE	392
TERRESTRIAL	UNKNOWN	UNKNOWN	NO DATA	-999

Appendix B: Statewide Model Results

This evaluation of statewide model results is based on a sample of 251,377 points. The columns in Table B1 are defined as follows:

- MODTYPE: The vegetation type in the sample population. These are the 'actual' values of the vegetation points, as interpreted by MnDOT.
- SAMPLE: The percentage of the total population of vegetation points classified by the MODTYPE value.
- ACCURACY: The percentage of the vegetation points that are correctly predicted by the model.
- MODEL: The percentage of points the model predicts to be the vegetation type.
- CONFIDENCE: The percentage of predictions that are correct, which may be interpreted as the confidence we can have that a cell classified as a vegetation type is actually that vegetation type. This differs from accuracy because the portions of each vegetation type in the model differ from the portions of those vegetation types in the sample population.
- ALT_TYPE: The most common incorrect prediction(s) for points of MOD_TYPE. The value in parentheses is the percentage of points of MODTYPE that are incorrectly predicted to be ALT_TYPE.

Table B1: Evaluation of Final Statewide Historic Vegetation Model

MODTYPE	VALUE	SAMPLE	ACCURACY	MODEL	CONFIDENCE	ALT_TYPE
LAKE	100	4.0%	93%	4.1%	88%	WET LAND (3%)
RIVER	200	0.2%	50%	0.3%	34%	WET LAND (23%)
WET LAND	150	0%	0%	0.6%	0%	N/A
BOG	210	2.3%	42%	1.4%	72%	CONIFER/SHRUB SWAMP (52%)
CONIFER SWAMP/SHRUB SWAMP	220	12.9%	86%	16.9%	66%	BOREAL HARDWOOD-CONIFER FOREST (3%)
MARSH	230	5.6%	57%	4.4%	72%	PRAIRIE (20%)
FLOODPLAIN FOREST	240	1%	60%	0.9%	72%	CONIFER/SHRUB SWAMP (8%) PRAIRIE (8%)
HARDWOOD SWAMP	250	0.2%	55%	0%	0%	MARSH (19%)
WET MEADOW/FEN	270	2.4%	39%	1.4%	69%	PRAIRIE (42%)
PINE FOREST	310	0.6%	43%	0.3%	74%	CONIFER/SHRUB SWAMP (13%) BOREAL HARDWOOD-CONIFER FOREST (10%)

MODTYPE	VALUE	SAMPLE	ACCURACY	MODEL	CONFIDENCE	ALT_TYPE
JACK PINE FOREST	311	2.1%	67%	2.5%	55%	CONIFER/SHRUB SWAMP (9%) BOREAL HARDWOOD-CONIFER FOREST (10%)
RED PINE FOREST	312	1.6%	58%	1.6%	58%	JACK PINE FOREST (11%) BOREAL HARDWOOD-CONIFER FOREST (8%)
WHITE PINE FOREST	313	1.5%	50%	1.2%	61%	CONIFER/SHRUB SWAMP (11%) BOREAL HARDWOOD-CONIFER FOREST (15%)
SPRUCE-FIR FOREST	321	1.6%	44%	1.2%	62%	CONIFER/SHRUB SWAMP (22%) BOREAL HARDWOOD-CONIFER FOREST (18%)
BLACK SPRUCE-FEATHERMOSS FOREST	322	0.6%	33%	0.3%	77%	CONIFER/SHRUB SWAMP (19%) BOREAL HARDWOOD-CONIFER FOREST (22%)
UPLAND WHITE CEDAR FOREST	323	0.005%	40%	0.3%	77%	CONIFER/SHRUB SWAMP (22%) BOREAL HARDWOOD-CONIFER FOREST (22%)
PINE BARRENS	330	0.2%	40%	0.1%	77%	JACK PINE FOREST (19%)
JACK PINE WOODLAND	341	0.2%	40%	0.1%	80%	JACK PINE FOREST (14%)

MODTYPE	VALUE	SAMPLE	ACCURACY	MODEL	CONFIDENCE	ALT_TYPE
NORTHERN CONIFER WOODLAND	342	0.1%	27%	0.05%	78%	CONIFER/SHRUB SWAMP (28%) BOREAL HARDWOOD-CONIFER FOREST (14%)
BOREAL HARDWOOD-CONIFER FOREST	351	4.1%	62%	5.4%	47%	CONIFER/SHRUB SWAMP (17%)
MIXED PINE-HARDWOOD FOREST	352	1.0%	44%	0.6%	71%	CONIFER/SHRUB SWAMP (10%) BOREAL HARDWOOD-CONIFER FOREST (11%)
NORTHERN HARDWOOD-CONIFER FOREST	353	0.5%	40%	0.2%	77%	CONIFER/SHRUB SWAMP (15%) BOREAL HARDWOOD-CONIFER FOREST (15%)
WHITE PINE-HARDWOOD FOREST	354	0.1%	39%	0.1%	78%	CONIFER/SHRUB SWAMP (11%) BOREAL HARDWOOD-CONIFER FOREST (8%)
ASPEN FOREST	361	3.3%	49%	2.5%	63%	CONIFER/SHRUB SWAMP (13%)
ASPEN-BIRCH FOREST	362	0.8%	36%	0.4%	79%	CONIFER/SHRUB SWAMP (16%) BOREAL HARDWOOD-CONIFER FOREST (14%)
PAPER BIRCH FOREST	363	1.8%	46%	1.4%	61%	CONIFER/SHRUB SWAMP (14%) BOREAL HARDWOOD-CONIFER FOREST (17%)

MODTYPE	VALUE	SAMPLE	ACCURACY	MODEL	CONFIDENCE	ALT_TYPE
LOWLAND HARDWOOD FOREST	364	1.8%	53%	1.4%	67%	CONIFER/SHRUB SWAMP (10%)
MAPLE-BASSWOOD FOREST	365	3.2%	77%	3.7%	66%	LOWLAND HARDWOOD FOREST (3%) OAK FOREST (4%)
NORTHERN HARDWOOD FOREST	366	0.9%	49%	0.7%	60%	CONIFER/SHRUB SWAMP (11%) BOREAL HARDWOOD-CONIFER FOREST (9%) MAPLE-BASSWOOD FOREST (8%)
OAK FOREST	367	4.4%	57%	3.7%	68%	MAPLE-BASSWOOD FOREST (8%) OAK SAVANNA (12%) PRAIRIE (11%)
ASPEN OPENINGS	371	0.1%	21%	0.02%	83%	OAK SAVANNA (12%) PRAIRIE (28%)
OAK SAVANNA	372	5.4%	65%	5.0%	69%	OAK FOREST (7%) PRAIRIE (15%)
ASPEN WOODLAND	381	1.4%	43%	0.9%	69%	PRAIRIE (19%)
OAK WOODLAND	382	1.1%	43%	0.6%	74%	OAK SAVANNA (11%) PRAIRIE (20%)
BRUSH-PRAIRIE	391	0.5%	34%	0.3%	61%	PRAIRIE (49%)

MODTYPE	VALUE	SAMPLE	ACCURACY	MODEL	CONFIDENCE	ALT_TYPE
PRAIRIE	392	31.8%	96%	35.3%	86%	MARSH (1%) OAK SAVANNA (1%)