Modeling the Potential Distribution of Kittentails
A Pilot Study on Element Distribution Modeling Techniques

Prepared for
Natural Heritage and Nongame Research Program
Minnesota Department of Natural Resources

By

Marcel Jouseau, Principal
SOMANAL
St. Paul, MN 55104

Contract A72783

June 30, 2005

INTRODUCTION

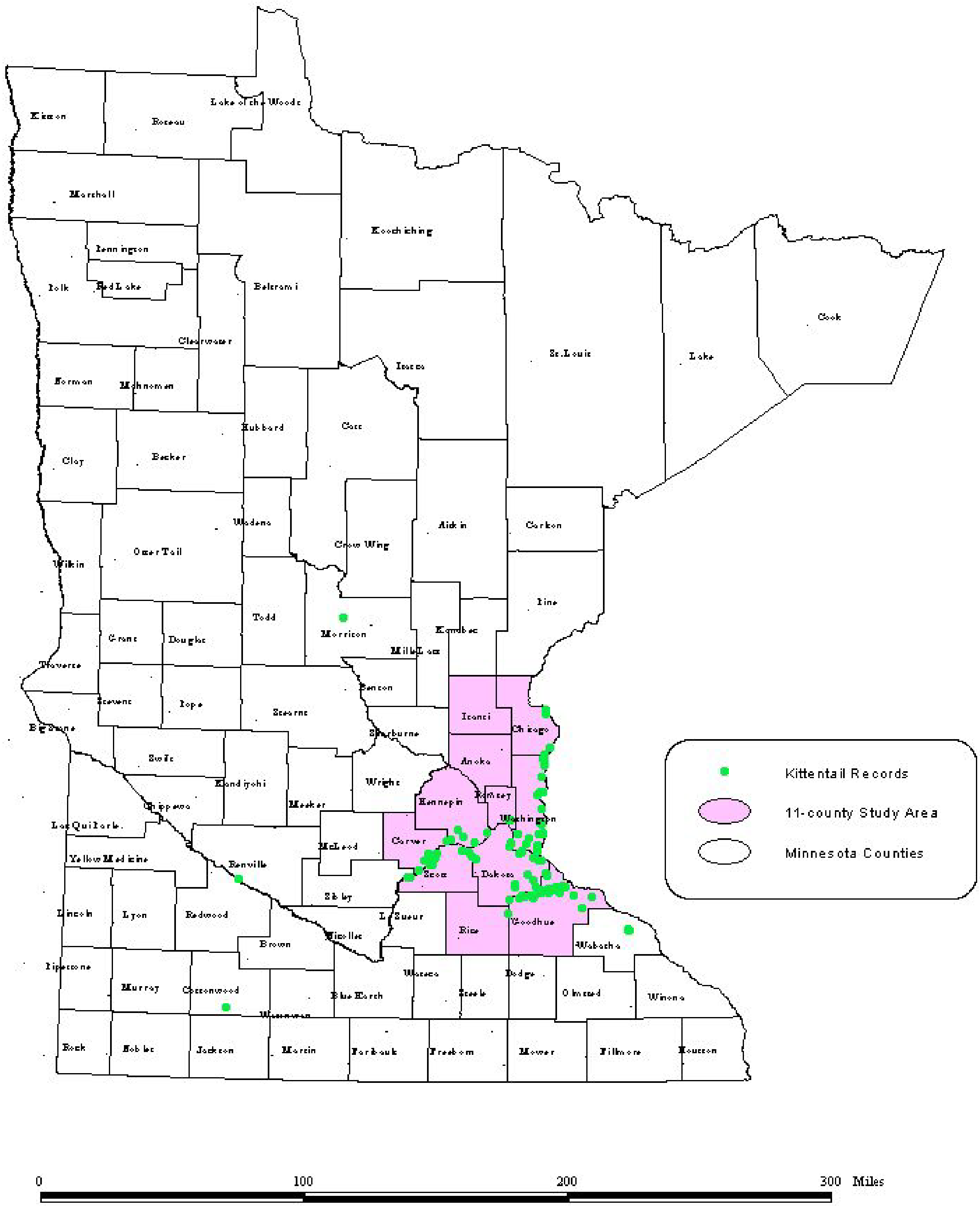
The State of Minnesota Natural Heritage and Nongame Research Program is charged with the tasks of identifying biological resources of significance to the state of Minnesota and generally overseeing the protection of endangered, threatened species and species of concern. In addition the Program manages significant habitats owned by the State for the purpose of scientific research and the protection of significant species and communities. In the general national and local climates of fiscal austerity the Program, like many other agencies, must developed means and approaches to discharging their responsibilities that permit greater efficiencies to maximize the reduced level of financial resources allocated to them.

Staff of the Minnesota Natural Heritage and Nongame Research Program (Natural Heritage) proposed to investigate the usefulness of Element (Species) Distribution Modeling as an efficient means of mapping the potential of and surveying for occurrences of endangered, threatened and species of concern by undertaking a pilot project.

Element Distribution Modeling (EDM) has become a standard tool in the mapping of distribution of species in various parts of the World such as Australia, New Zealand and the state of Wyoming in the United States. Recognizing the usefulness of EDM, NatureServe, an international natural heritage conservancy organization, recently organized a series of training workshops to expend the use of this tool. Techniques for predicting the presence-absence of species have been extensively documented in the scientific literature starting in the mid 1980s (Tzilkowski et al., 1986; Slovan et al., 1996; Franklin, 1998; Ozemi & Ozemi, 1999; Guisan & Zimmerman, 2000; Guisan et al., 2002; Elith, 2002; Park et al., 2003). Practitioners such as Fertig applied the techniques to model the distribution of sensitive and threatened and endangered species in Wyoming (Fertig & Thurston, 2003; Jouseau applied logistic regression model and brought to light 14 new populations of *Aquilegia jonesii* in the Bighorn National Forest (Jouseau, 2005). EDM has been reported to result in greater efficiency in the delineation of species distribution (Hernandez, 2005) through assigning probabilities of occurrence of a species or community to every parcel of land of a modeled area thereby focusing field survey efforts on areas with high probability of occurrence, thus reducing the cost of surveying.

For the purpose of demonstrating the use of EDM, Natural Heritage staff selected the kittentail (*Besseya bullii*), a threatened species in Minnesota. *Besseya bullii* is unique to the Midwest where it has been recorded in Ohio, Michigan, Indiana, Illinois, Wisconsin, Minnesota and Iowa. The species is said to have been extirpated in Ohio and to now have minimal footholds in Michigan, Indiana, Illinois and Iowa. According to herbarium records for Minnesota, the known distribution of the species is centered on the seven counties of the Twin Cities Metropolitan Area and the adjacent counties of Chisago, Goodhue and Rice, in addition to Morrison County about 90 miles northwest of Saint Paul and Renville and Cottonwood counties about 100 miles west-southwest of the Twin Cities (see Fig. 1). The presence of the taxon in Iowa and Wisconsin and the records northwest and south-southwest of the Twin Cities suggest that the species may have had a foothold in at least the southern 2/3 of Minnesota. The geography of the pilot study encompasses the seven metropolitan counties and the counties of Goodhue, Rice, Chisago and Isanti (see Fig. 1).

Fig. 1. Pilot Study Area
Kittentail Distribution Modeling Study



METHODS

Advances in speed, memory and storage capacity of computers, progress in the development of geographic databases, the availability of species records from herbaria and museums, together with advances in computerized statistical analysis methods have made possible the use of predictive modeling to identify potential habitats for species. A survey of the last 10 years of such journals as Ecological modeling, Journal of Conservation Biology, Journal of Applied Ecology and the Journal of Biogeography will identify numerous articles on different modeling approaches. Numerous techniques have been used among which are: logistic regression, classification and regression trees, artificial neural networks, principal component analysis, maximum entropy, generalized linear models and generalized additive models.

Statistical Modeling Methodology

It is not the purpose of this pilot study to review and compare the various modeling approaches or to describe them. The journals referred to above contain such reviews of the techniques and explanations of the various approaches. However, it is appropriate to briefly describe the two techniques applied during this pilot study. For this pilot project two techniques were selected: logistic regression and classification trees.

The logistic regression model allows the formulation of a statistical relationship between the binary (presence/absence) occurrence of a species and environmental conditions e.g. geology, slope steepness, temperature, rainfall, solar radiation, land cover that are presumed to describe the habitat of that species. This relationship is transparent and easily interpretable (Guisan & Zimmerman, 2002). In other words, this approach allows one to calculate a linear relationship between the presence and absence of kittentails (the dependent variable) and an array of environmental variables (the independent variables), which together may constitute the required habitat for kittentails to establish themselves and persist. The difference between logistic regression and the standard linear regression approach that all introductory statistic courses introduce is that logistic regression is applied to situation where the dependent variable has a binary value (0-1, yes-no, present-absent). Furthermore the value of the dependent variable is limited to "0" or "1", as we cannot have half of a no or yes. As a result a graph of the function shows an "S" curve rather than the straight line of the standard linear regression where the dependent variable can take any value positive or negative as dictated by the constant and the set of independent variables in the equation. The values of the independent variables or predictors can be continuous such as topography expressed in feet or meters, or categorical such as slopes greater than 15 percent and slopes less than 15 percent. The logistic regression technique has been amply documented in the literature and in several books, the most comprehensive of which is likely by Hosmer and Lemeshow (Hosmer and Lemeshow, 1989). One of the most appealing characteristics of logistic regression models is the calculation of a probability of the presence of the species for every cell of the grid of the geographic area represented. Additionally, the logistic regression technique is fully integrated with a number of geographic information systems software packages such as ArcView, GRASS and IDRISSI to name only three broadly used software packages. The basic formula with which the probability of occurrence is computed in logistic regression is as follows:

$$\mathrm{Py} \mid (\mathrm{x}) = \frac{e^{(\alpha + \beta 1X1 + \beta 2X2 + \beta 3X3 + ... + \beta nXn)}}{1 + e^{(\alpha + \beta 1X1 + \beta 2X2 + \beta 3X3 + ... + \beta nXn)}}$$

The probability of the presence of the kittentail given a set of variables X is the ratio of "e" which is equal to 2.17 raised to the power of the regression with variables that are considered to constitute the habitat of the species to the sum of 1+"e" raised to the power of the regression of the habitat variables.
X1 can be slope, X2 aspect, X3 annual monthly average rainfall, Xn sand and gravel, etc. When the model is integrated into a geographic information system with data formatted in a grid cell format, it is then possible to compute a probability of occurrence of the species in every grid cells for which environmental data are available.

Classification trees on the other hand use a partitioning algorithm that subdivides variables associated with the presence of a species in a dichotomous way, much like a botanical key does e.g. leaves pubescent or not pubescent, into increasingly smaller and homogenous classes associated with a species response. Most classification algorithms allow the results of the classification to be represented graphically as a tree with branches representing environmental variables and various values of the variables associated with a binary response of the species and the number of species records associated with the subdivision of a variable. The graphic representation of the classification makes the classification tree a very easily interpreted model. The disadvantage of the technique is that it is generally not well integrated with GIS techniques and the geographic representation of the potential habitat is arrived at by adding maps representing values of the variables that are associated with the presence of the species e.g. gravelly soils, slopes greater than 30 percent, barren land cover with a western exposure). Also no probability regarding the presence-absence of the species is associated with a geographic area delineated, although it is possible to construct a scale that accounts for the number of variables present at each delineated geographic site.

Environmental Data

Environmental variables that were considered for inclusion as predictor variables are listed in Table 1. Most of the variables are surrogates for habitat conditions important to the germination, growth, reproduction and survival of the species. For example, monthly precipitation and soil hydrologic groups are surrogates for water availability for which no data are readily available on the geographic scale of the study area.

Four main classes of environmental variables are represented: climate, land cover, topography and substrate. The data for these variables are either numerical representation of a continuum of values such as topography, air temperature and slope percentage, or numerical representation of categorical or discrete classes such as surficial geology, soil units or vegetation types (prairie, oak savanna, conifer plantation).

Table 1. Environmental Data Layers considered in modeling

Continuous Variables	**Units**	**Code**
Elevation	feet	Elev
Slopes	percent	Slope
Aspect	degrees	Aspect
Monthly minimum temperature (January-Dec. each month)	Degree celsius	Tmin1- Tmin 12
Monthly mean temperature (January-Dec. each month)	Degree celsius	Tmean1-Tmean12
Monthly maximum temperature (January-Dec. each month)	Degree celsius	Tmax1- Tmax 12
Annual Monthly minimum Temperature	Degree celsius	Tmin13
Annual Monthly mean Temperature	Degree celsius	Tmean13
Annual Monthly maximum Temperature	Degree celsius	Tmax13
Monthly mean rainfall (January-Dec. each month)	mm	Precip1-Precip12
Monthly mean snowfall (January-Dec. each month)	mm	Snow1-Snow12
Categorical Variables		**Code**
Soil classification		HydrG
Sand and Gravel formation		Sand
MLCCS Landcover		MLCCS
GAP landcover		GAP
Pre-development vegetation (Marshner's Map)		PrVeg
Minnesota Geomorphology		Geomor

While the 11-county region studied is very small it was impossible to find environmental data layers with the same degree of resolution throughout the area. The exception is the set of climate data available to the author from previous research, which is uniformly gridded over the entire United States. This data set, produced by Climate Source LLC. http://www.climatesource.com, includes monthly average minimum, mean and maximum temperatures for each month, monthly average rainfall and monthly average snowfall. The grid size is one mile square. The data set was resampled to match the 10-meter grid size for the study area.

Two digital elevation models (DEM) were obtained from the Minnesota Department of Natural Resources Data Deli http://deli.dnr.state.mn.us/ : a 30-meter DEM available for the entire state of Minnesota and a finer grain 10-meter DEM developed for the Metropolitan Council and extending over an area slightly larger than the 7-county metropolitan area. The 30-meter DEM was re-sampled to a 10-meter grid. The area of the 7-county metro area was then cut out and replaced with the 10-meter DEM available for the metropolitan area. Slopes and aspects were computed from the created 10-meter DEM for the 11-county study area.

A uniform statewide STATSGO soil survey was downloaded from the data server of the Natural Resources Conservation Service (NRCS). Unfortunately this survey is made up of aggregated soil units and much information on characteristics of individual soil units is lost in this aggregation. For several counties a new digital survey is available (SSURGO soil survey) these individual county soil surveys were downloaded from the NRCS. A third type of digital soil data set is available for the 7-county metropolitan area providing information at the soil unit level. Goodhue and Isanti counties do not have a digital soil survey, a SSURGO soil survey will soon be available for Goodhue. Part of Goodhue County was digitized in raster form, possibly at the University of Minnesota but several townships were not digitized. As a result a soil map for the 11-county area was constructed using a combination of survey resolutions SSURGO soil survey for Rice and Chisago, STATSGO and digitized raster data for Goodhue County, STATSGO soil survey for Isanti County and the digitized metropolitan area survey for the seven metropolitan counties.

Surficial and bedrock geology maps for the state of Minnesota were downloaded from the web site of the Minnesota Geological Survey. Maps of sand and gravel deposits were obtained from the websites of the Metropolitan Council for the 7-county metropolitan area and from the Minnesota Department of Natural Resources Division of Minerals for adjacent counties. Areas described in the NRCS soil surveys as sandy soils or gravelly soils was added to the information on the sand and gravel maps downloaded from the DNR and Metropolitan Council.

Land cover and vegetation data were obtained from the Minnesota Department of Natural Resources Data Deli and from the Metropolitan Council. Specifically the Minnesota Presettlement Vegetation map was obtained from the Data Deli. The presettlement vegetation of Minnesota is based on Marschner's original analysis of Public Land Survey notes and landscape patterns. Marschner compiled his results in map format, which was subsequently captured in digital format. This layer was thought to be useful as it would represent vegetation as it likely existed when *Bessaya bullii* established itself in the area. Marshner's pre-development vegetation map covers the entire study area and beyond. The digital 2002 land cover map for the 7-county metropolitan area was obtained from the Metropolitan Council. This map is derived from the interpretation of Landsat imagery prepared at the University of Minnesota Remote Sensing Laboratory. The map is at a 30-meter pixel resolution. The Minnesota Department of Natural Resources provided a copy of the Minnesota GAP Land Cover 1.2, 1-acre MMU Arc GRID dataset for the entire state on CD. A digital, statewide map of the native plant communities prepared by the Minnesota County Biological Survey was downloaded from the Data Deli. Additionally a copy of the digital map of Minnesota Land Cover Classification System (MLCCS) version 5.4 was obtained from the Minnesota Department of Natural Resources Metro Region Office. The MLCCS map covers most of the 7-county metropolitan area but substantial geographic gaps exist and furthermore there is no similar information for the remaining four counties in the study area. One of the serious issues with the land cover and vegetation data available i.e. the Marshner pre-development vegetation, the MLCCS, GAP Land cover 1.2 and the Metropolitan Area 2002 land cover map is the lack of correspondence between the data throughout these vegetation and land cover maps. Table 2 lists the land cover types that occur on each of the Marshner, MLCCS and GAP maps at the point of occurrence of the known kittentail records. A second serious issue is the discontinuity in the maps because surveys are not yet completed or surveys have purposely focused on high quality plant communities rather than all plant communities.

All data were reprojected when necessary to the Universal Tranverse Mercator (UTM) Zone 15 and vertical datum NAD 1983. All vector format data were converted to a grid format and all grids were uniformly converted to a 10-meter pixel size. The statewide data sets were first trimmed to the 11-county study area. Later, as the modeling process proceeded, because of concern with data gaps and lack of uniformity the statewide data and other data were trimmed to the 7-county metropolitan area which has a more uniform quality of data with the exception of the MLCCS map which presently covers only about 80 percent of the 7-county metropolitan area.

Pre-modeling evaluation of the climatological data was performed with Arcview 3.3 (ESRI) GIS software package. All presence data for kittentails were mapped and a map in decimal degrees and datum NAD 1983 was produced. This map was overlaid on statewide map of average minimum temperatures, average mean temperatures, average maximum temperatures, average rainfall and average snowfall for each month of the year, in addition to average annual monthly minimum, mean and maximum temperatures, average annual rainfall and average annual snowfall. Kittentails data was overlaid on a total of 65 climatological maps for the purpose of determining whether there were discernable patterns in the climate between where kittentails are present and those areas from which they are not known. This analysis demonstrated that there were no discernable climatological patterns that would influence the distribution of kittentails at the local scale of the study area. Other researchers have demonstrated that at the continental or national scale climatological data show a significant relationship to the distribution pattern of a species or plant community (Loehle and LeBlanc, 1996). However, the climatological data needs to be collected at a very fine scale to be useful in correlating with species distribution at the local scale. These detailed climatological data were simply not available for this study.

As a result of the lack of relationship between the climate variables and the presence of kittentails shown in the pre-modeling analysis, the climate variables were not further used in the modeling.

Presence-Absence Data

In order to model species distribution with most types of model it is necessary to have spatial geographic location data with coordinates, or to which can be assigned coordinates, for locations where a species has been found (presence data) but also where surveys did not record the species (absence data). Absence data is rarely routinely collected. As a result, absence data must be derived from other data sets by making assumptions about past data collection efforts. For example the County Biological Survey (CBS) native plant community data is the result of fairly intensive fieldwork. When *Besseya bullii* was found during the course of the fieldwork for the county biological survey a record of its presence was noted as a member of the plant community it was found in. For this modeling purpose it was assumed that if the CBS data did not record the species in a plant community polygon, any randomly selected geographic point in that community could be presumed to represent an absence of the species. For the modeling effort about 800 points were randomly selected across the 11-county study area. Later two subsets of absence data were created: one with 299 absence points for the portion of the 7-county metropolitan area covered by MLCCS data; the other with 403 absence points throughout the 7-county metropolitan area. The random absence data points were selected from all plant communities and land cover classes. Absence data points that coincided or were at close proximity to recorded presence data were deleted. Efforts were made to maintain a minimum distance of 1000 meters between absence points.

Presence data for Besseya bullii was obtained from the Natural Heritage database after signing a license agreement. About 110 usable presence points are available for the study area. The dates of the records vary from 1884 to 2003; however, almost 30 percent of the records were last observed prior to 1990. Additional concerns with the presence data are the horizontal positional accuracy and the destruction of sites by the rapid urban development in the 11-county study area. In previous modeling efforts (Jouseau, 2005) relocated with a GPS the position of known records of *Aquilegia jonesii* prior to modeling the distribution of that species. Because of the four-month time limit of the present project, part of which was during a period of snow cover, it was not possible to undertake a survey with GPS to ascertain the existence of the species population and its positional accuracy. During the limited field verification efforts on model results, it became obvious that some known presence sites no longer appeared to be valid because of housing development at the site.

Table 2. Land cover types as recorded on three maps at locations of *Besseya bullii* records

Marshner Pre-development vegetation	**GAP Land Cover 1.2, 1-acre MMU**	**MLCCS Land Cover**
Aspen-Oak Land	1905- Barren	Coniferous trees
Big Woods	1907- Grassland	Perennial grasses & sparse trees
Oak openings and Barrens	1918- Red Pine	Deciduous trees
Prairie	1936- Red Oak	Mixed coniferous/deciduous trees
River bottom forest	2007- Grassland	Deciduous forests
Wet Prairie	2037- Northern Pin Oak	Non-native mixed woodland
	2302- High Intensity Urban	Oak savanna
	2303- Low Intensity Urban	Prairie
	2306- Cropland	Grassland & sparse conifer or mixed deciduous/coniferous trees
	2307- Grassland	Lowland hardwood forest
	2309- Upland Shrubs	Maple-Basswood forest
	2335- Bur/White Oak	Mixed hardwood swamp
	2336- Red Oak	Mixed Pine hardwood forest
	2338- Maple/Basswood	Oak forest
	2343- Lowland Deciduous	Short grasses and mixed trees
	2606- Cropland	Short grasses on upland soils
	2607- Grassland	White Pine hardwood forest
	2636- Red Oak	Undefined
	2638- Maple/Basswood	Pavement 91-100% impervious cover
	2706- Cropland	
	2734- White/Red Oak	
	2736- Red Oak	

Table 3. Example of prepared variables and presence-absence data for the models

KITTENTAIL	SANDVAL	MLCCSVAL	ASPECVAL	HYDRGVAL	SURFAVAL	SLOPEVAL	PVEGVAL
0	0	0	0.0000	1	1	0.0000	1
0	0	0	0.0000	0	1	0.0000	0
0	1	0	0.0000	0	0	0.0000	1
0	0	0	0.0003	0	0	2.5400	0
1	0	0	0.7252	0	1	10.0338	1
1	1	0	3.3665	0	1	19.4646	0
0	0	0	3.8142	0	0	1.9092	0
1	0	1	4.1277	0	1	12.3510	1
0	0	0	8.1296	0	1	0.8980	0
0	0	0	8.9726	0	1	9.7716	1
0	1	0	9.7821	0	1	3.7373	0
1	1	0	10.0081	1	1	4.3847	1
0	0	0	11.3108	0	1	0.6476	1
0	0	0	12.9950	0	0	3.3888	1
0	0	0	26.5668	0	1	0.5680	0
0	0	1	27.6460	1	1	3.0107	1
1	1	1	28.4957	0	1	15.1732	1
1	1	1	28.6108	1	1	3.1826	0
0	0	0	29.0547	0	1	1.3075	0
1	1	0	29.7449	0	1	4.0956	1
0	1	0	35.5376	0	1	2.1850	1
0	1	0	35.5383	0	1	1.0925	0
0	0	0	37.0565	0	0	7.7978	0
0	1	0	40.0303	1	1	8.2930	0
0	0	0	41.5315	0	0	5.9378	0
0	0	0	45.0000	0	1	0.3592	0
1	0	0	45.0000	0	1	19.3974	1
0	0	0	45.0000	1	0	0.5388	0
0	1	0	45.0000	0	0	2.8737	1
0	1	0	45.0000	0	1	0.5388	1
0	0	0	45.0000	1	1	0.1796	0
1	0	0	54.8657	0	1	4.1929	1
0	0	0	56.3102	0	0	0.9158	0
1	1	0	59.0363	1	1	7.4053	1
1	0	0	60.7512	0	1	29.1115	0
0	0	0	63.4344	0	1	0.5680	0
0	0	0	71.5650	0	0	4.0161	1
0	0	0	71.5650	0	1	0.4016	0
1	0	0	73.1094	1	1	32.7832	1
1	0	0	73.7910	0	0	22.7482	1
1	0	0	74.1289	0	1	26.9348	1
0	0	0	75.2564	1	1	2.4952	1
0	1	0	90.0000	0	1	0.5080	1
1	1	0	95.6307	1	1	27.1822	1
0	0	0	98.1295	0	1	0.8980	0
0	0	0	98.1303	0	0	1.7961	0

Logistic Regression Models

The development of a logistic regression model requires that the presence-absence data set must be populated with the value for each environmental variable that intervenes at the location of each presence-absence data point; see Table 3 for an example for the 7-county metropolitan area. The presence-absence data and the constructed environmental data set are imported into a commercial statistical package such as SAS, S-Plus, or SPSS, STATS, MiniTab, or any other favorite package. SPSS 12.01 was used to construct the models. The presence-absence data for kittentail is entered as the dependent variable (KITTENTAIL), where "1"represents presence and "0" absence of the species. The environmental variables such as sand-gravel (SANDVAL), slope steepness (SLOPEVAL), slope aspect (ASPECVAL), surficial geology (SURFAVAL), soil permeability (HYDRGVAL), MLCCS land cover (MLCCSVAL), or pre-development vegetation (PVEGVAL) are entered as independent variables either as continuous or categorical type of variable as follows

SANDVAL	Categorical variable, 1= presence of sandy/gravelly soils and 0= absence of those soils.
SLOPEVAL	Continuous variable, slope steepness values range from 0-100 percent.
ASPECVAL	Continuous variable, values range from 0 to 360 of the compass values of the orientation of the slopes. A few models were informally built in which ASPECVAL was used as a categorical variable with "1" repreenting orientations which seemed favorable to the presence of the species (Northeast to West) and "0" as orientations without records.
SURFAVAL	Categorical variable with "1"favorable surficial/geomorphologic features such as terraces and "0" unfavorable features such as floodplain deposit, silt and clay deposits.
HYDRGVAL	Categorical variable "1" representing well drained soils (NRCS hydrologic soil group A) and "0" soils with very poor to moderately drained soils (hydrologic soil groups B, C, D).
MLCCSVAL	Categorical variable "1" representing favorable land cover classes including all prairie types and oak savanna (land cover categories as determined from the habitat information available on the labels of the herbarium specimens for the records that make up the presence file. "0" represents land classes for which there were no kittentail records such as tamarack swamp, wetlands, cultivated fields, urban development greater than 15 percent imperviousness.
PVEGVAL	Categorical variable "1" represents prairie and oak opening and barrens and "0" unfavorable vegetation categories.
GAPVAL	Categorical variable "1" represents barrens, grassland, white/red oak and bur/white oak and "0" for unfavorable land cover categories.

Models with GAPVAL provide a very poor response because of the broad array of land cover classes in the GAP land classification apparently associated with kittentails. This is likely the result of classification errors when classifying the land cover from satellite imagery. For examples kittentail records occur in three classes of cropland, four classes of grassland, five classes of red oaks, as well as lowland deciduous. The presence of kittentails in these classes of land cover is counter-intuitive or does correspond to the experts' opinion on what constitutes the habitat type for the species.

Variables entered in the models and presence/absence prediction successes for each of 14 of the 18 or so models that were tested are provided in Table 4. Numerous other models were tested and immediately discarded because of lack of relationship.

Table 4. Summary of logistic regression models

Note: Models 1 through 14 were run both with logistic regression software and classification tree software. In the last column of the table are the percentage of success in classifying presence (top number) and the percentage of success at classifying absence (bottom number). In other words e.g. model 1 classified correctly 41 percent of the presence data and 90 percent of the absence data. The "X" in a column shows that the variable was used in the model. For each training model presence/absence data points were selected with the use of random numbers; data points which were not selected for training were available for model validation.

Model	Area Covered	Sand	Slope	Aspect	Hydrogroup	Pre-Dev. Vegetation	MLCCS	Surficial Geology	Score
1	11-county	X	X	X	X	X		X	41/90
2	11-county	X	X			X			43/98
3	7-county	X	X	X	X	X		X	70/97
4	7-county	X	X		X	X			65/97
5	7-county	X	X	X	X	X		X	60/98
6	7-county	Slope + Pre-dev-vegetation+ Sand + (Sand* Slope) + (Sand * Aspect)							61/97
7	7-county	X	X	X	X		X	X	77/98
8	7-county	X	X				X		75/98
9	11-county	X	X	X	X	X		X	37/97
10	11-county	X	X			X		X	46/98
11	11-county	X	X	X	X	X		X	44/97
12	11-county	X	X			X		X	50/96
13	11-county	X	X			X		X	40/92
14	11-county	X	X			X		X	29/96

The most successful models according to the presence/absence success rates shown in table 4 above are models 7 and 8. The equations for the logistic regression version of models 7 and 8 are listed below:

Model 7 = -6.567 + (sand*2.833) + (slope*0.141) + (aspect*0.002) - (hydrogroup*0.023) +(mlccs*1.498) + (surfigeo*0.331)

Model 8 = -6.034 + (sand*2.896) + (slope*0.141) + (mlccs*1.510)

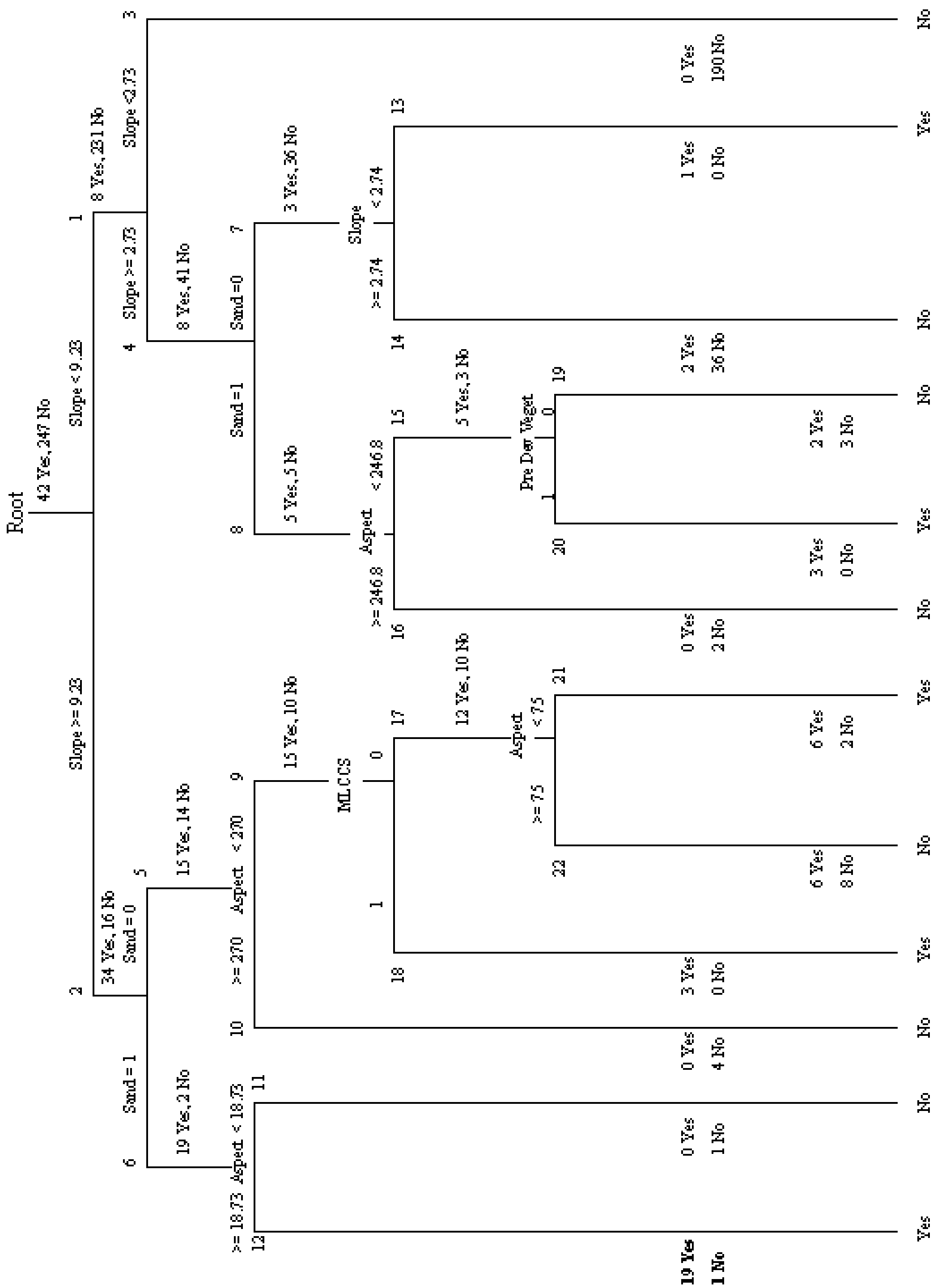
Classification Tree Model

As with the logistic regression model the presence-absence data set must be populated with the value for each environmental variable that intervenes at the location of each presence-absence data point see Table 3 for an example for the 7-county metropolitan area. To create a classification tree model the constructed environmental data set is imported into a commercial statistical package such as SAS, S-Plus, or SPSS, or freeware packages like R, or QUEST, or LOTUS that has a module for classification trees. QUEST version 1.9.2, a freeware downloadable from the University of Wisconsin Department of Statistics http://www.stat.wisc.edu/~loh/quest.html was used to develop the classification trees. The presence-absence data for kittentail is entered as the dependent variable and the environmental variables such as sand-gravel, slope steepness, slope aspect, surficial geology, soil permeability, MLCCS land cover, GAP Land cover, or pre-development vegetation are entered as independent variables either as continuous or categorical type of variable. Following the example of one model shown in Graph 1 for illustration, the software algorithm begins with the complete presence-absence set at level 1 or "root " as it is known in classification tree terminology (here 42 presence and 247 absence cases). The algorithm then continuously splits the data into pairs of subsets (nodes) of the data along the variable values that explain the greatest difference between presence and absence. The splitting continues along each branch using different variables and values at each bifurcation or nodeuntil a pre-determined maximum number of nodes is reached, or a pre-determined minimum number of cases is reached, or the presence-absence data point on a branch is "pure" that is either presence only, or absence only.

In Graph 1., the model uses the slope steepness variable as the first explanatory variable. The first split occurs at the value 9.23 percent of the slope variable. Slopes less than 9.23 percent (Node 1) explain 231 absences and 8 presences while slopes greater than or equal to 9.23 percent (Node 2) explain 34 presences and 16 absences. The subset at Node 1 is then split further with the slope variable at a value of 2.73 percent. Slopes less than 2.73 percent (Node 3) account for 190 absences and no presence; as a result of the split this subset is "pure" and does not need to be further split. At the other end of the branch, Node 4 (Slopes greater than 2.73 percent) still contains a mixture of presence and absence points and therefore needs further splitting and the results are Nodes 7, 8, 13, 14, 15, 16, 19 and 20). Splitting at Nodes 14 and 19 is interrupted, as there are no additional benefits to be derived from an additional split; additional splits would simply clutter the graph and therefore are pruned using rules such as percentage of "purity". On the other side of the graph, Node 2 is split according to the presence of sand or gravel on those steep slopes. At Node 6 the subset is split according to the orientation of the slopes (aspect) with aspect greater than 18.7 degrees or south of N-NE (Node 12) accounting for 19 presences of kittentail and one absence. This node is almost "pure" and not further split. "Yes" and "No" labels at the end of a branch are known as the "Leaf"; they indicate whether the leaf represent presence or absence, respectively.

Numerous classification tree models were run in this fashion with data for the 11-county region, or data for the 7-county area or just the area for which MLCCS data were available. In addition models were built with as few as two independent variables and as many as seven variables.

The percentages of overall correctly classified presence-absence data points indicate that the classification tree models appeared to perform reasonably well. However, results from the classification need to be scrutinized further with the help of any one of three statistical analysis tools explained in the section on model selection.



Graph 1. Classification tree model of Besseya bullii for 7-county metropolitan area

MODEL SELECTION AND RESULTS

The results of the various models can be validated by applying the resulting equations or classification trees rules to a subset of presence-absence data not used in the construction of the models. Total success rate is measured by dividing the total of properly classified presence and absence points by the total number of presence-absence points used in the model. One can also examine the rate of success at predicting presence as well as the rate of success at predicting absence of the species. Misclassification is comprised of false positive errors and false negative errors. All the information for interpretation of the success of the model is contained in the confusion matrix (a 2 by 2 matrix).

Table 5. Confusion matrix

	Model Present (Predicted)	**Model Absent (Predicted)**
Known Present (Observed)	Classified Correctly (Present Success Rate) **(a)**	Misclassified (False Negative Error) **(b)**
Known Absent (Observed)	Misclassified (False Positive Error) **(c)**	Classified Correctly (Absent Success Rate) **(d)**

The information provided in the four cells in Table 5. is then used to calculate success coefficients such as Cohen's Kappa and the Normalized Mutual Information (NMI), a very robust indicator of the validity of the model and the most conservative measure of performance of models (Forbes, 1995). The confusion matrix and Kappa and NMI coefficients are calculated for both the success of the training models built, as well as the results from the validation models.

The NMI coefficient is calculated with the equation listed below. The letters "a, b, c, d" in the equation refers to the numeric value in the respective cells of the confusion matrix; "N" is the sum total of all presence and absence data or "a+b+c+d" and "ln" is the natural log.

$$NMI = 1 - \frac{-a\ln a - b\ln b - c\ln c - d\ln d + (a+b)*\ln(a+b) + (c+d)*\ln(c+d)}{N\ln N - [(a+b)*\ln(a+c) + (b+d)*\ln(b+d)]}$$

The confusion matrix for the classification tree shown in Graph 1. is shown below. An analysis of the data in the confusion matrix for this training model shows that 76 percent of the presence data and 99 percent of the absence data are properly classified. The overall classification success rate is 95 percent.

Table 6. Confusion matrix for classification tree model shown in Graph 1.

	Model Present (Predicted)	Model Absent (Predicted)
Known Present (Observed)	32 **(a)**	10 **(b)**
Known Absent (Observed)	3 **(c)**	244 **(d)**

While the Cohen's Kappa coefficient of 0.8 is very good, the NMI coefficient is lower at 0.63 but within the acceptable range of values.

When the rules of this classification tree model are applied to a set of presence-absence data not used in building the training model, 67 percent of the presence data is properly classified and 96 percent of the absence data is properly classified. On the surface this appears to be a good model. However, the Cohen's Kappa coefficient on this test model is lower at 0.62 but still acceptable; on the other hand the NMI coefficient is only 0.36 and definitely in the poor range.

The confusion matrix information percent of success with presence data and absence data as well as the Cohen's Kappa and NMI coefficient for eight of the logistic regression models are given in Table 7 below. Models 3, 7 and 8 have very respectable Cohen's Kappa coefficients, however the NMI coefficients show that these models are not as robust as the classification success percentages and Kappa coefficient make them to be and only model 7 and 8 are acceptable.

Table 7. Presence-absence classifications and Kappa and NMI coefficients for 14 of the logistic regression models.

	a	b	c	d	Presence success	Absence success	Cohen's Kappa	NMI
Model 1	46	62	16	719	42.6	97.8	0.49	0.32
Model 3	54	23	10	350	70.1	97.2	0.72	0.49
Model 4	50	27	10	393	64.9	97.5	0.69	0.47
Model 5	46	31	6	354	59.7	98.3	0.67	0.48
Model 6	47	30	10	350	61.0	97.2	0.65	0.43
Model 7	59	18	8	395	76.6	98.0	0.79	0.58
Model 8	58	19	8	395	75.3	98.0	0.78	0.57
Model 9	40	68	19	715	37.0	97.4	0.42	0.25
Model 10	38	45	12	499	45.8	97.6	0.52	0.33
Model 11	48	60	15	476	44.4	96.9	0.49	0.30
Model 12	40	40	13	354	50.0	96.5	0.53	0.31
Model 13	10	15	19	205	40.0	91.5	0.29	0.09
Model 14	8	20	5	119	28.6	96.0	0.30	0.15

FIELD VERIFICATION

Element distribution modeling serves two purposes. First it provides an understanding of the potential range of the distribution of a species and in so doing provides information that may be useful to biologists interested in habitat restoration for the purpose of re-introducing the species or strengthening the occurrence of the species. The second purpose of the modeling is aimed at finding new populations of the species being modeled. Since models such as the logistic regression type provide a grid with a probability value of the suitability of anyone cell as a habitat for the species, it is a matter of determining a cut off point on acceptable probability values and survey the areas with a probability equal to or greater than set as the cut off point. The statistical analyses done with the information in the confusion matrix are very useful and must be undertaken to ascertain the validity of the models. The presence-absence success rates must be further analyzed using Cohen's Kappa analysis, Receiver Operating Characteristics Curve (ROC), or Normalized Mutual Information analysis; preferably either ROC or NMI, the last one being the most robust analysis. However, while rarely done by researchers because absence of the plant in the range can be the result of many conditions unrelated to the performance of the model, field verification of model results is the ultimate test: can the species be found where predicted.

For the purpose of verifying the predictions of the models about 50 sites in Carver County (Figures 8 & 9), Dakota County (Figure 11) and Washington County (Figure 11) were selected from areas the models showed to have a high probability of occurrence of the species. Small aerial photos with the sites, site identification numbers and coordinates were prepared. A handheld Global Positioning System (GPS) Magellan Meridian Gold instrument was used to ascertain the location in the field and that one was in fact at the intended sites.

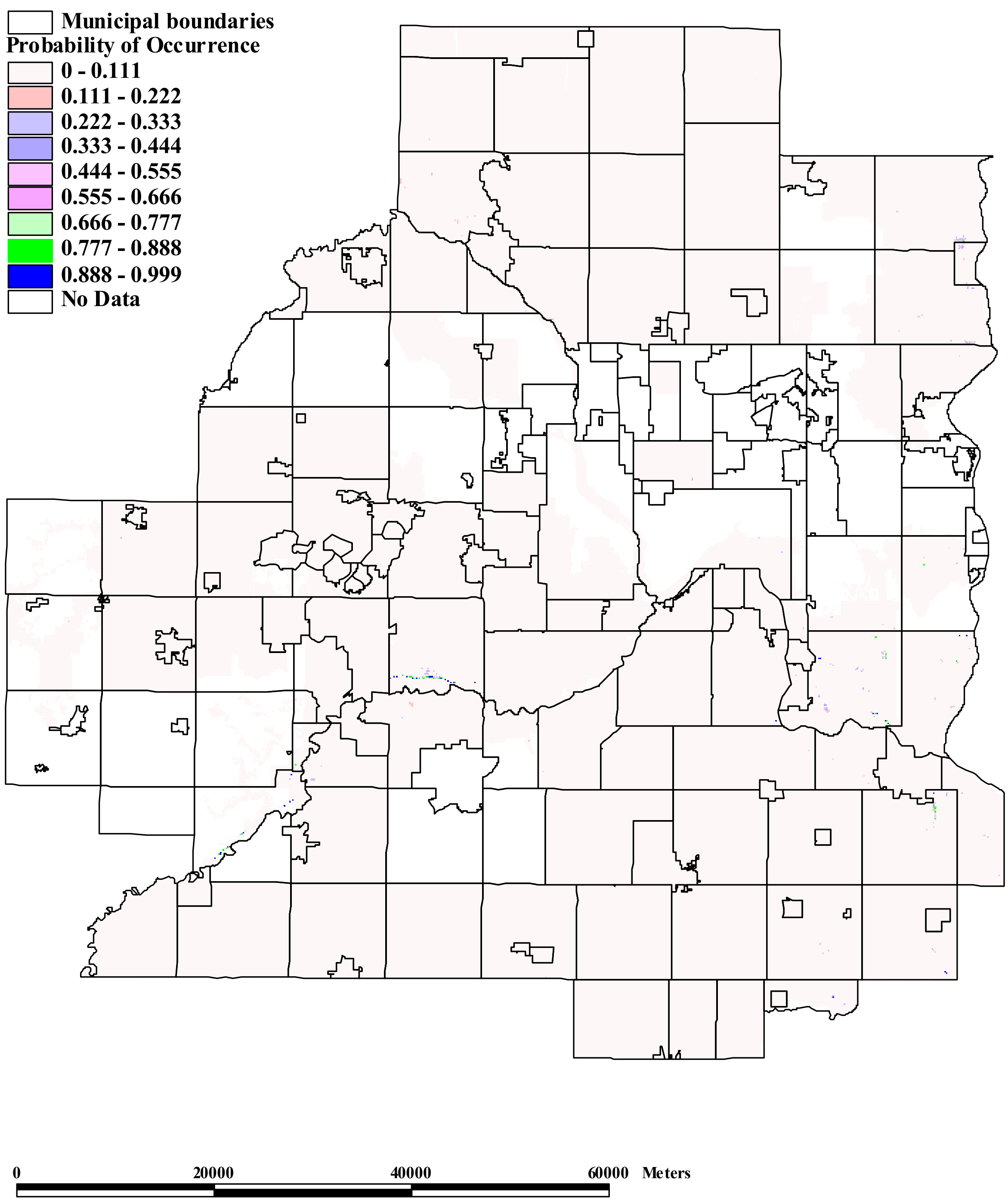
The Carver County sites in San Francisco Township had all the characteristics desired as known to this author for the sites to support *Besseya bullii*. The sites are gravel prairies, typically with steep slopes, sand and gravel substratum, excellent exposure to sunlight, tree cover --generally oaks -- at the periphery of these small prairies. Not only were the environmental conditions conducive to providing a good habitat for kittentails but the various plant species present at the sites are known to be present at sites where kittentails are known to grow. Prairie plants typically in bloom at those sites at the time of the visits were: *Delphinium virescens, Campanula rotundifolia, Lithospermum canescens, Penstemon affinis, Penstemon grandiflora*. Plants of *Dalea sp., Liatris sp.* were also seen, though not in bloom at the time. However, despite an extensive search of more than 20 of the 33 sites for the modeled species, the species could not be found. Of note *Heuchera richardonii*, a species often associated with *Besseya bullii*, was not found at any of those sites either.

The sites at Pine Bend, Dakota County are also prairie remnants. While they exhibited a much higher degree of degradation than the Carver County sites with fewer prairie species, nonetheless typical prairie species were also present at the Pine Bend sites. Again no kittentails were found during the search.

The Washington County sites were different from the others as they were identified as oak savanna. However, it became rapidly clear to this author that they likely to support substantial populations of kittentails. All the sites were far too densely wooded to support *Besseya bullii*. The crowns of trees were intermingled, heavy scrub under story and a thick layer of humus and decomposing leaves was covering the ground. Steep slopes were present at all the sites. The search was totally unsuccessful.

At many of the sites, urban encroachment is evident. Carver county sites 17 through 20 next to a gravel road, had become the back yard of several homes. Carver county sites 29 to 33 immediately east of County Rd 45 have homes on the sites and manicured landscape have encroached on the prairie remnants. Across the road on the west side of County Rd 45 there is a known element occurrence. These known occurrences of kittentail have likely disappeared because of the extensive sand and gravel mining operation. The coordinates for one of the element occurrences obtained from the Department of Natural Resources put the species smack in the middle of a very large gravel stock pile. The other element on the eastern edge of the pit was searched for but could not be found.

Kittentail Logistic Regression Model # 8



June 2005

Figure 2

DISCUSSION

As stated earlier, many researchers involved in element distribution modeling have taken the position that absence of an element for an area for which the species was predicted to occur by the model does not lead necessarily to the conclusion that the model is ineffective, or bad. Researchers view these models as a means of delineating the potential range of a species. The fact that a species modeled cannot be found has lead researchers to surmise that it could be the result of the lack of dispersion of seeds to the site; the nearest neighbor sites being too far for seeds to reach the area, given the characteristics of the seeds of the species. Additional reasons could be that a site is visited during the wrong season and that the species is not yet growing above ground at the time of the visit, or that climatologic conditions during the year have been unfavorable to the development of the species and that it has remained dormant or underdeveloped.

Of course other reasons may be the cause. These can range from the inaccuracies in the location points of the element occurrences; this translates then in the wrong environmental data variables being selected and confusion in the model. Time and resource constraints did not allow for field verification of the presence-absence data to address accuracy issues. Environmental data maps can also be too coarse or uneven in their specificity to lead to proper identification, or definition of the correct habitat. The 11-county area study is definitely lacking in uniform, quality data. Additionally, it is plausible that some requirements for a specific environmental condition is not reflected in a model because of the lack of knowledge of the species requirements, or the unavailability of the data layer.

Moreover, many species and particularly prairie species were dependent on fairly regular wild fires caused by thunderstorms burning the prairie and thereby controlling woody species and other species that might be invading. In an urban environment such as the 11-county study area, urban encroachment has precluded the use of fire to control invading remnant prairies. The sites in Washington County exemplify the situation as tree crowns are too dense to let the light in, scrubs under the trees provide additional attenuation of light and the dense humus and leaf layer have changed the environmental conditions provided by the substratum.

The lack of success in finding kittentails at the various sites searched caused the author to go to two extensive populations of kittentails to re-acquaint himself with the vegetative characteristics of the species. One of those sites, the Cannon River Terrace SNA, the author had over the course of 2004 mapped the specific locations of populations the species with a GPS and produced a map of populations in Arcview GIS. In 2004, several hundreds of plants of *Besseya bullii* were found. In May and early June 2004 they bloomed extensively and throughout the summer spikes bearing seed capsules were visible; in fact these dry inflorescences were still visible in the early Spring 2005, after the snow melt. Visits of the sites in May and June 2005 provided only a very scarce indication of the presence of the plant. Whereas in 2004 hundreds of plants were visible at the River Terrace SNA, in 2005 only three small groups of flowering plants were found; two were comprised of one individual plant, the other had approximately 12 plants. A more intensive search, square foot by square foot of the sites, crisscrossing longitudinally and laterally the slopes of the areas known to have exhibited the species lead to finding just a handful of very small, immature plants.

Climatologic conditions during the winter 2004-05 were very harsh in that the area was subjected to sub-zero temperatures (Farenheit) while there was a total lack of snow. Could the weather conditions of the winter have caused the species to remain largely dormant? The author has been unable to locate any published research on the ecology and life cycle of this species of kittentails.

While the harsh conditions of the winter 2004-05 may have affected the growth of *Besseya bullii* and therefore the ability to find the species in the field, it is clear that a number of problems seriously impeded the modeling efforts and these issues had a large impact on generally the poor quality of the model results as shown in Table 6. The more serious issues are discussed below:

- Accuracy of the location of the element occurrence records

  As expressed earlier in the report, the duration and timing of the study did not allow the author to ascertain the accuracy of the coordinates of the location in the field of the records of the species. Some records of the species when overlaid on landcover or vegetation maps appear to be locate d in coniferous swamp, flood plain forest, or several types of cropland. If the location is inaccurate, very likely the environmental conditions associated with the location will give the wrong definition of the "preferred" environment or habitat for the species. The results will be totally erroneous models or models grossly overpredicting.

- Continued presence of the element occurrence

  Element occurrence data is also of concern, especially in an urbanizing area, as land development affects the longevity of a record. A small number of known occurrences of kittentails were visited; alas, in some instances lawn was established where a prairie previously existed and the element seemed to have been destroyed. In another instance, the record puts an element smack in the center of a very active sand and gravel mining operation. The two records at that location were not found in June 2005. A quick check of the dates on which element occurrences of kittentails were last observed shows that 25 percent of the records were last observed prior to 1990 and 75 percent were last observed more than 10 years ago. If the sites have been destroyed or the land cover has been severely altered, this can lead to the formulation of an habitat layer that is erroneous and that will result in poor model being developed and poor model performance.

  These two issues relating to accuracy and presence of the element occurrence records can be remedied by including a period for field checking prior to the start of the actual model development.

- Quality of environmental data layers

  The quality of the environmental data layers is of concern both in terms of the grain and inconsistent quality, The soil layer for the 11-county study area had to be produced from four different sources and levels of soil unit definition. For some counties only STATSGO surveys were available. This is a very coarse grain definition of soils as it shows only soil associations leaving out much detail where soils are grouped together. Some other counties have SSURGO surveys, a modern, more precise and refined delineation of individual soil units. The 7-county metropolitan area has yet another digitized survey, though with the level of definition going across the whole metropolitan area.

As the Natural Resources Conservation Service proceeds with completing and publishing the SSURGO surveys, the much improved soil layer will provide valuable information for modeling.

Similar issues exist with the four landcover/vegetation layers available. The three statewide layers (Marshner's pre-development vegetation map, GAP landcover version 1.2 and the County Biological Survey) and the MLCCS available only for a portion of the metropolitan area present issues of accuracy and lack of concordance between layers. For example, for the metropolitan area GAP contains six grassland classes none of which is specifically identified as prairie and they include some croplands, golf courses, lawns and likely prairies as well. MLCCS, on the other hand, has numerous categories for prairie and there is no concordance or correspondence between the GAP and MLCCS data.

The completion of the MLCCS layer for the 7-county metropolitan area will provide a substantial resource and improve the ability to model with land cover/vegetation data.

Model 8 performed well, even though it did not lead to successful verification efforts. Model 8 relied on the more precise, though incomplete, mapping of land cover of the MLCCS, on the much higher accuracy of the definition of slopes afforded by the 10-meter digital elevation model for the metropolitan area and on good mapping of sand and gravel deposits. The sites generated by model 8 that were inspected (San Francisco Twp., Figures 8 and 9) were very credible.

It is hoped that the Natural Heritage and Nongame Research Program will seek another opportunity to apply EDM for the purpose of modeling habitat and finding new populations of a particular species. In applying EDM techniques to a new study area it would be advantageous to pay attention to the following points:

- Focus on an area not subject to rapid changes;
- Area should have a reasonably consistent data quality throughout;
- Time must be provided to ascertain location accuracy, continued presence of the records and accuracy of the habitat description, as good element records are the foundation of good models.
- Availability of published information on the ecology of the species would be a plus.

These points would improve the ability to develop useful models.

LITERATURE CITED

Elith, J. 2002. Predicting the Distribution of Plants.Ph.D. dissertation. Scool of Botany, University of Melbourne, Australia. pp.304.

Fertig, W. and R.Thurston. 2003. Modeling the potential distribution of BLM sensitive and USFWS threatened and endangered plant species in Wyoming.. Report prepared for the Bureau of Land Management, Wyoming State Office., pp. 43.

Forbes, D. 1995. Classification-algorithm evaluation: five performance measures based on confusion matrices. Journal of Clinical Monitoring, 11(3), 189-206.

Franklin, J. 1995. Predictive vegetation mapping: geographic modeling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, 19 (4), 474-499

Guisan , A. and Zimmerman, N.E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling*, 135 (2-3), 147-186.

Guisan, Antoine, T.C. Edwards, T. Hastie. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling,* Vol. 157(1), p. 89-10l.

Hernandez, P. 2004. Spatially explicit distribution models for predicting species occurrences. Thesis Master degree of Science. University of Toronto, pp.77.

Hosmer, D.W. and S. Lemeshow. 1989. Applied Logistic Regression. John Wiley & Sons, New York.

Jouseau, M. 2005, *Aquilegia jonesii* threatened or is it?, *Rock Garden Quarterly*, In Press, due July 2005. North American Rock Garden Society.

Loehle, C. and D. LeBlanc. 1996. Model-based assessments of climate change effects on forests: acritical review. *Ecological modeling*, 90(1), 1-31.

Ozesmi, S.L. and Ozesmi, U. 1999. An artificial neural network approach to spatial habitat modeling with interspecific interaction. *Ecological modelling*, 116 (1), 15-31.

Park, Young-Seuk, R. Celreghino, A. Compin, S. Lek. 2003. Applications of neural networks for patterning and predicting aquatic insect species richness in running waters. *Ecological Modelling*, Vol. 160(3), p. 265-280.

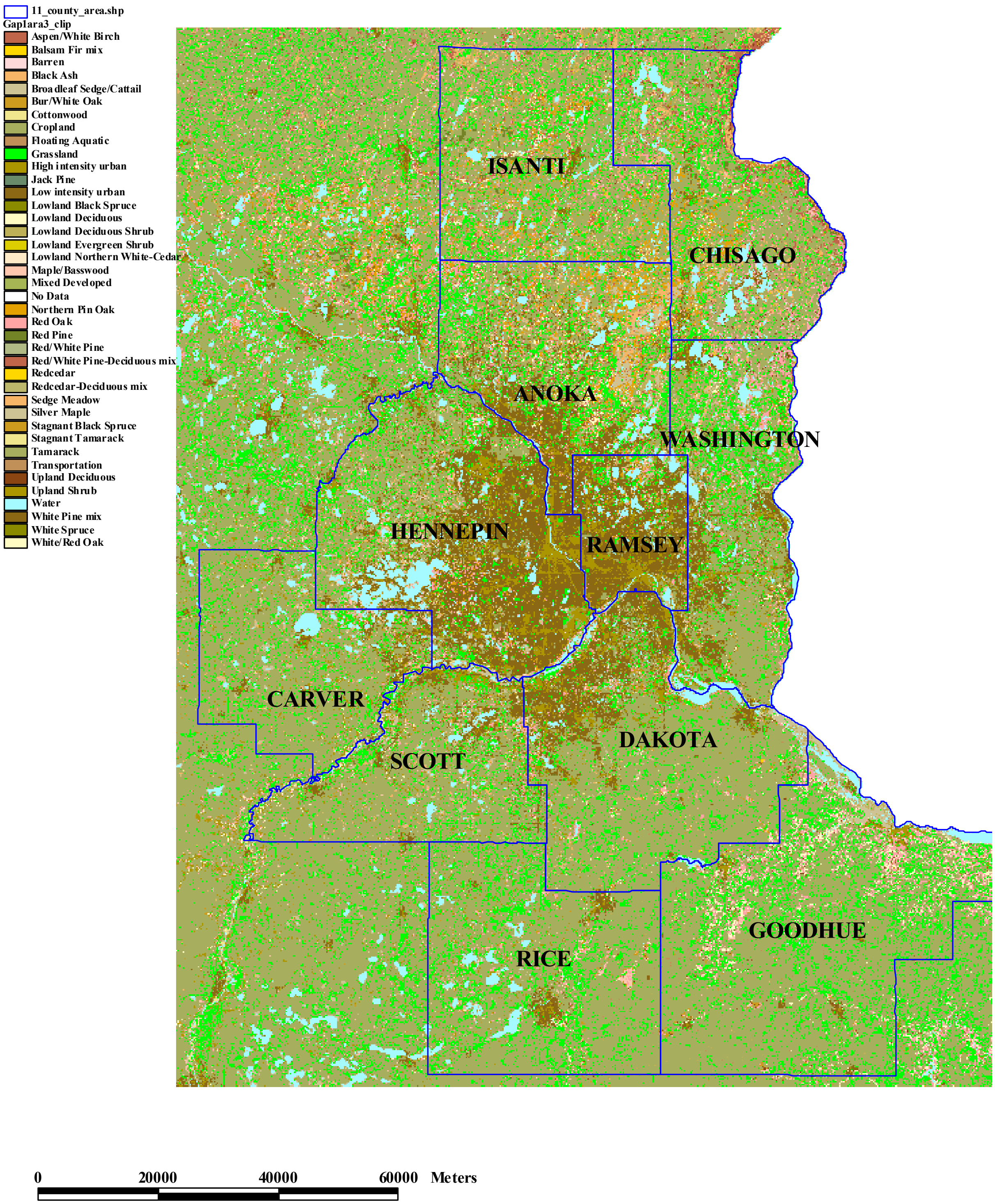
Slovan, Lek, M. Delacoste, P. Baran, I. Dimopoulos, J. Langa, S. Aulagnier. 1996. Application of Neural Networks to modeling non-linear relationships in ecology. *Ecological Modelling,* Vol. 90(1), p. 39-52.

Tzilkowski, W.M., J.S. Wakeley, L.J. Morris. 1986. Relative use of municipal street trees by birds during summer in State College, Pennsylvania. *Urban Ecology,* Vol. 9(3-4), p. 387-398.

Appendix

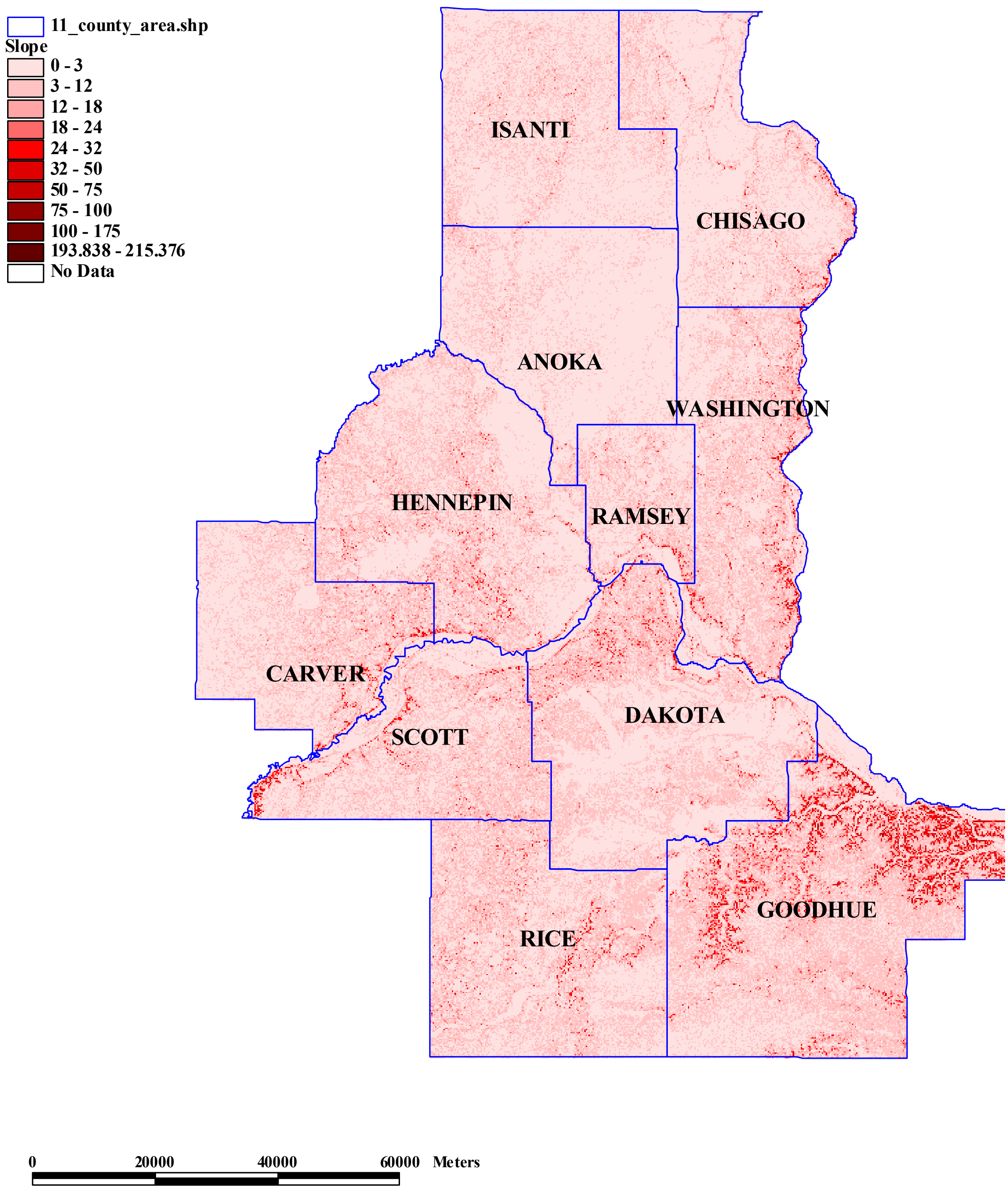
The appendix contains a few samples of some of the data used in the modeling, as well as a few maps resulting from the calculation of the probability of kittentails occupying a site. Moreover four maps that show sites that were part of the model field verification efforts are included. Literally hundreds of maps were produced in this project but there was no justification to make them all part of this report. Maps are contained on the DVDs on which the GIS project has been transcribed.

Figure 3. Gap Data Used in Modeling Kittentail



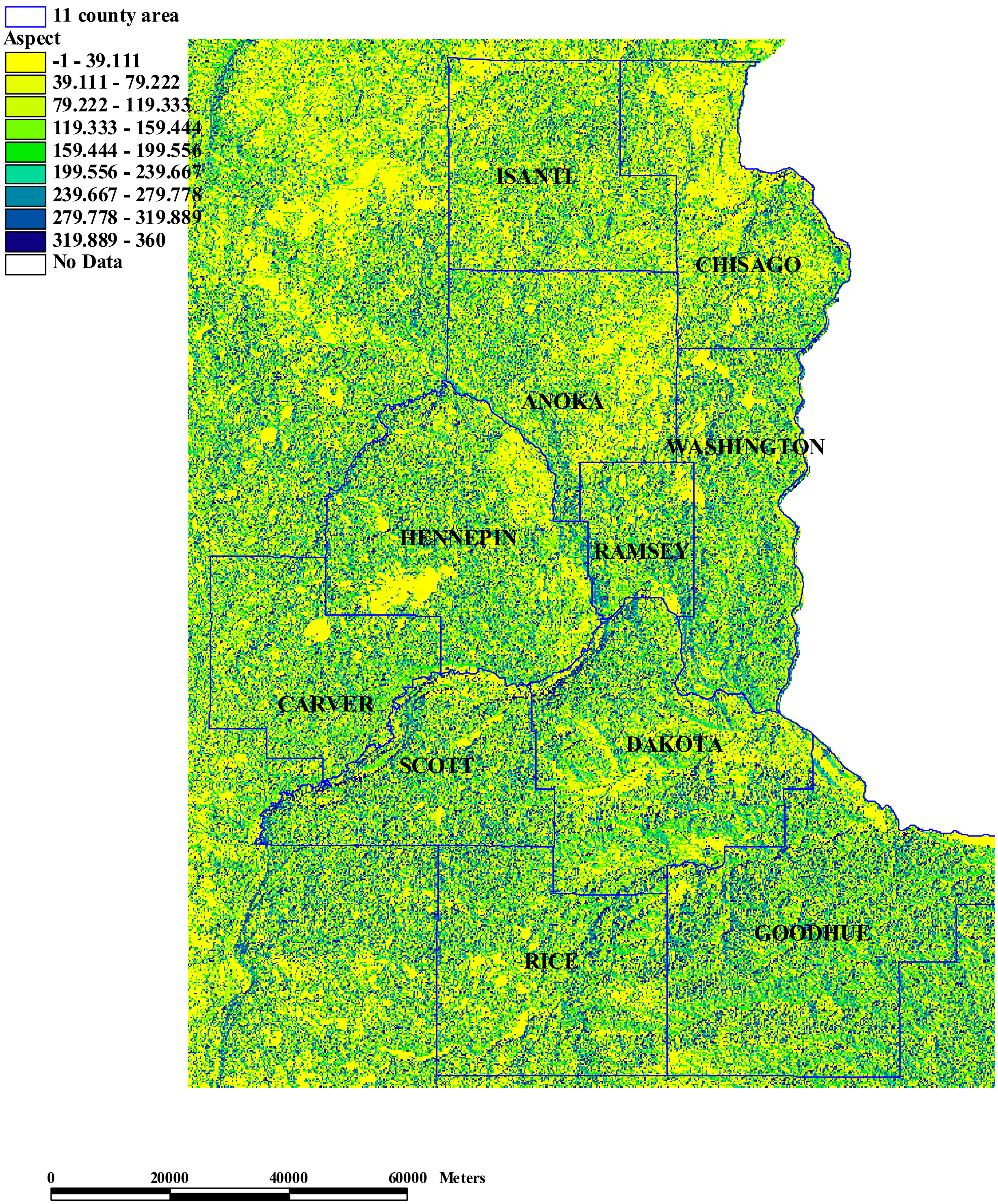
June 2005

Figure 4. Slope Data Used in Modeling Kittentail



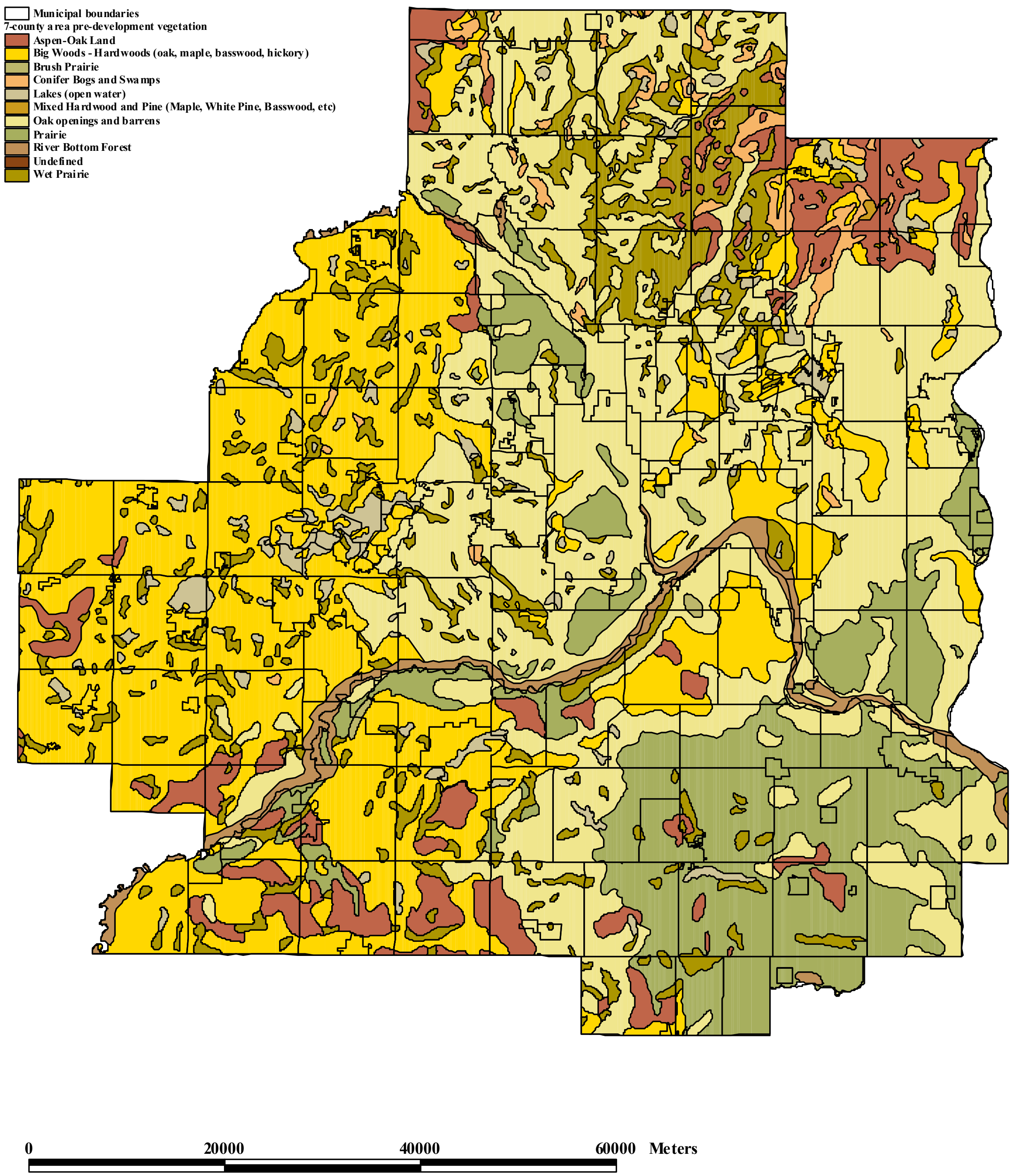
June 2005

Figure 5. Aspect Data Used in Modeling Kittentail



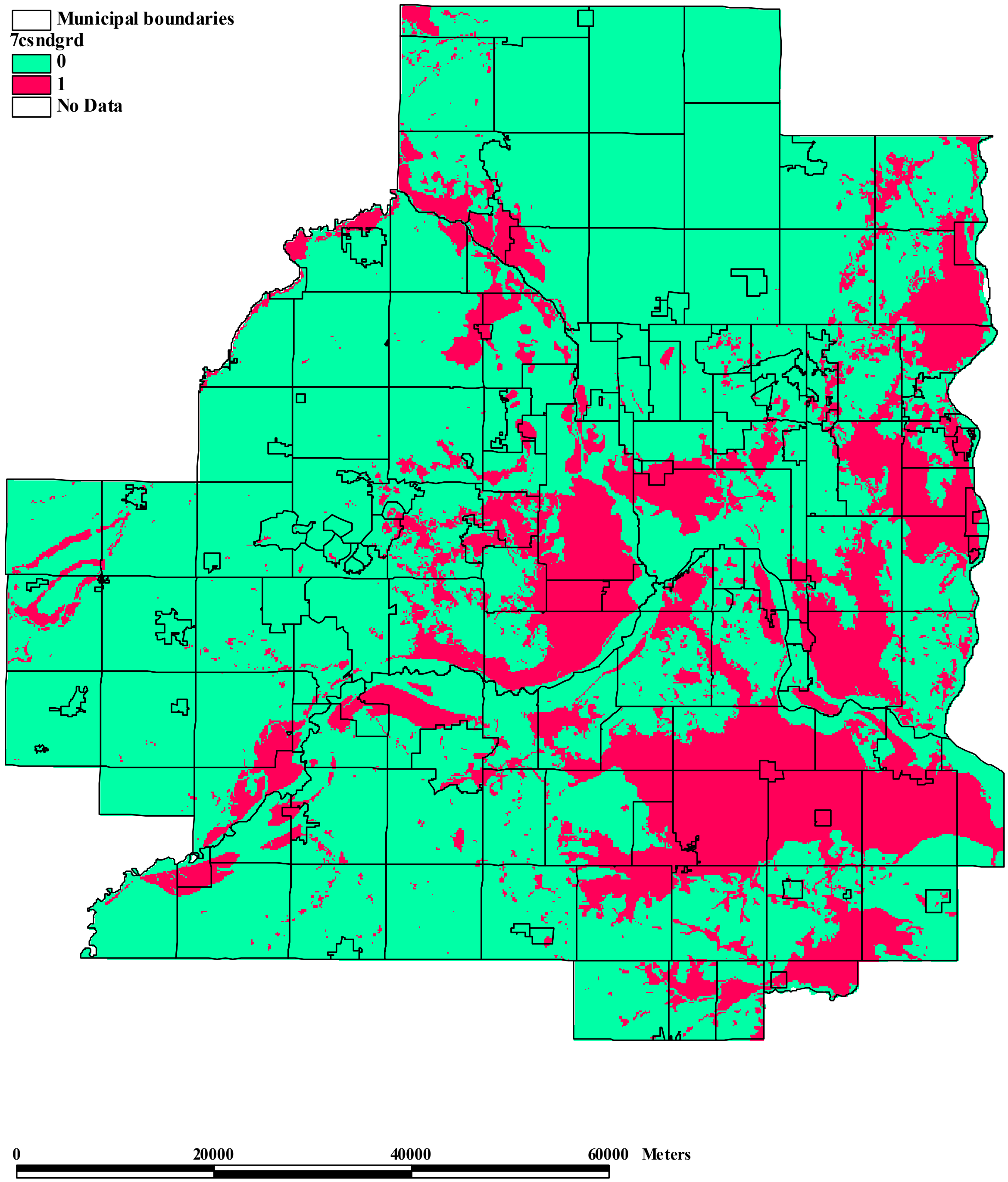
June 2005

Figure 6. Pre-Development Vegetation Used in Modeling Kittentail

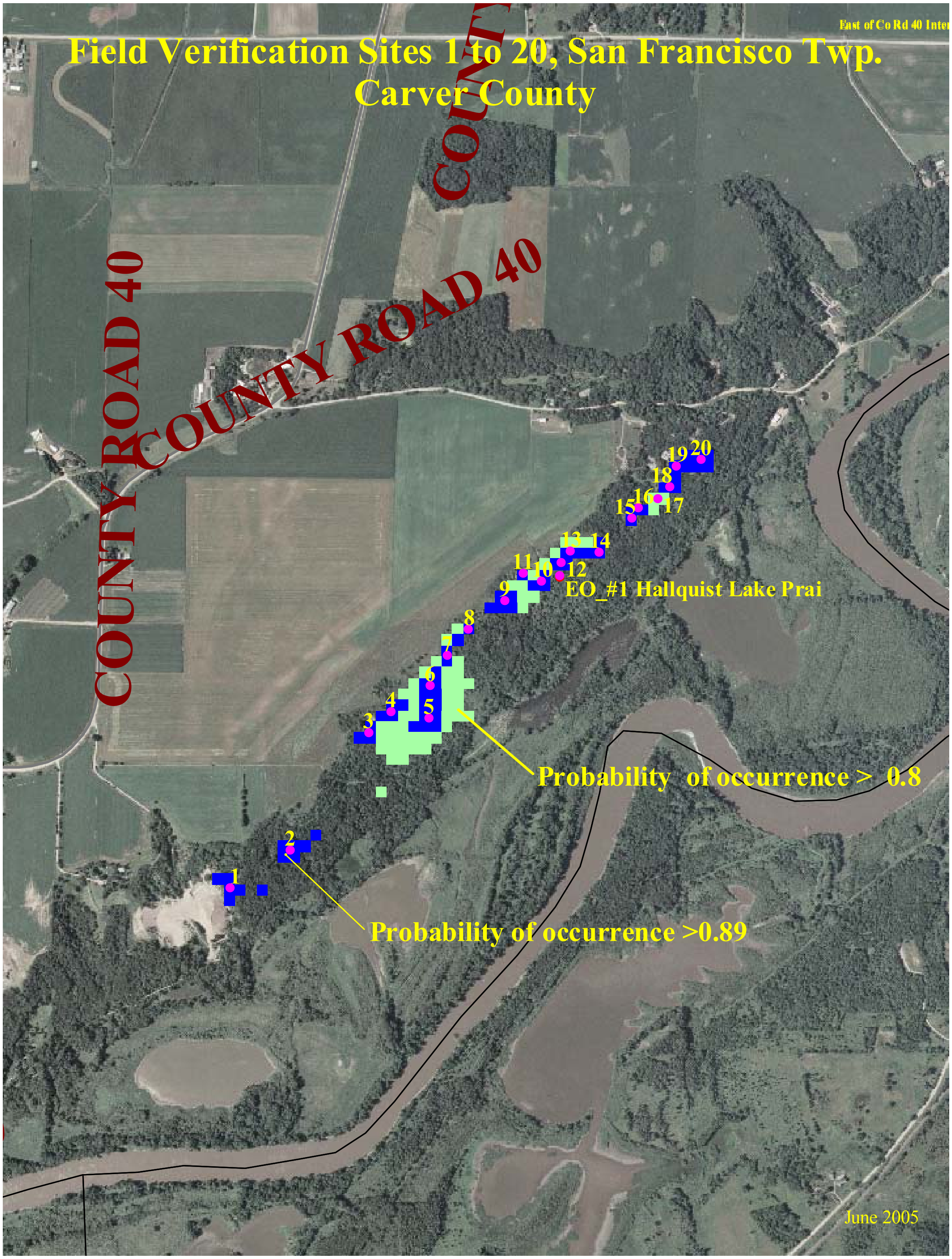


June 2005

Figure 7. Gridded Sand and Gravel Data Used in Modeling Kittentail



June 2005



Field Verification Sites 1 to 20, San Francisco Twp.
Carver County
East of Co Rd 40 Inter
COUNTY ROAD 40
COUNTY ROAD 40
EO_#1 Hallquist Lake Prai
Probability of occurrence > 0.8
Probability of occurrence >0.89
June 2005

Figure 8.

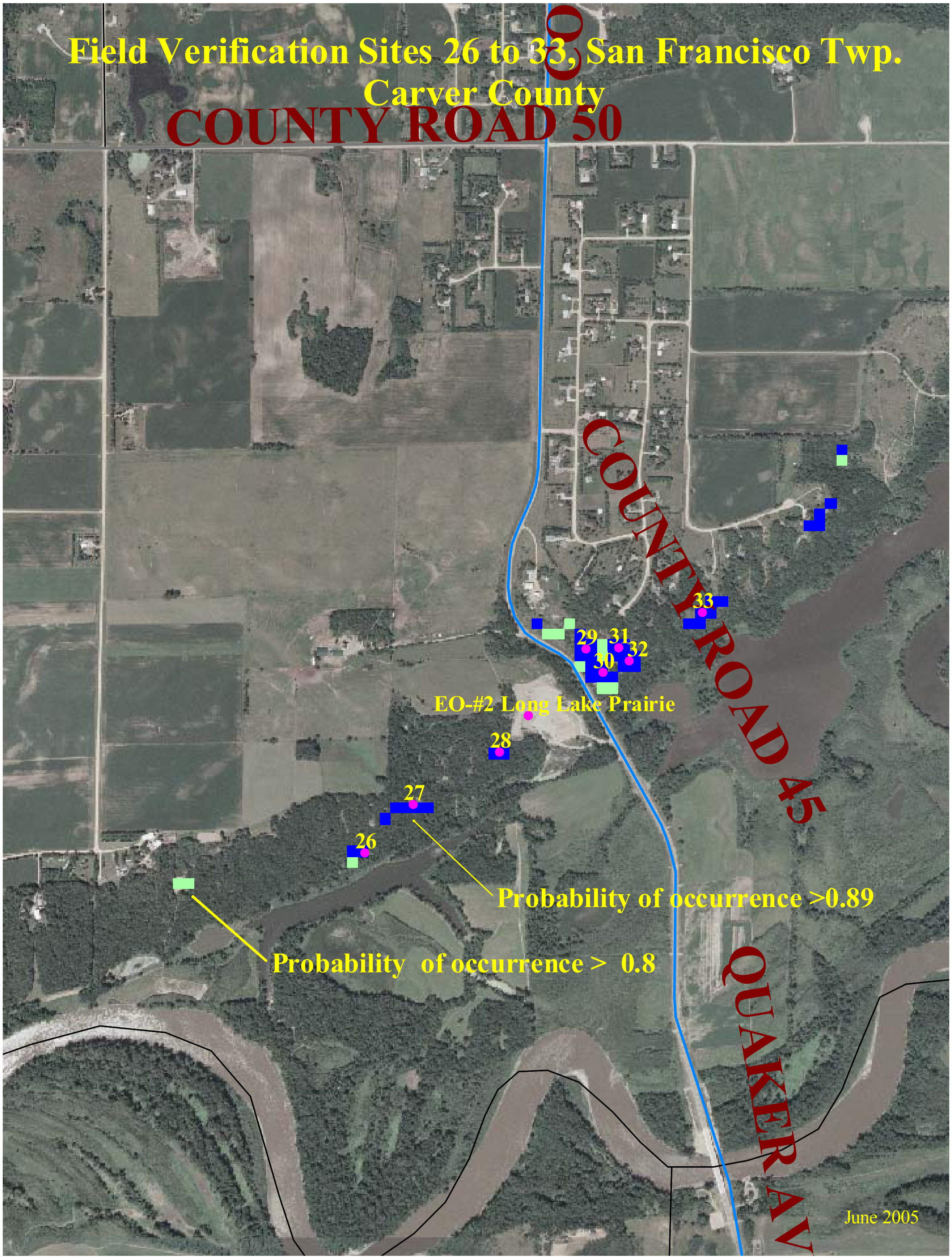

Field Verification Sites 26 to 33, San Francisco Twp.
Carver County
COUNTY ROAD 50
COUNTY ROAD 45
QUAKER AV
33
29
31
32
30
EO-#2 Long Lake Prairie
28
27
26
Probability of occurrence >0.89
Probability of occurrence > 0.8
June 2005

Figure 9.

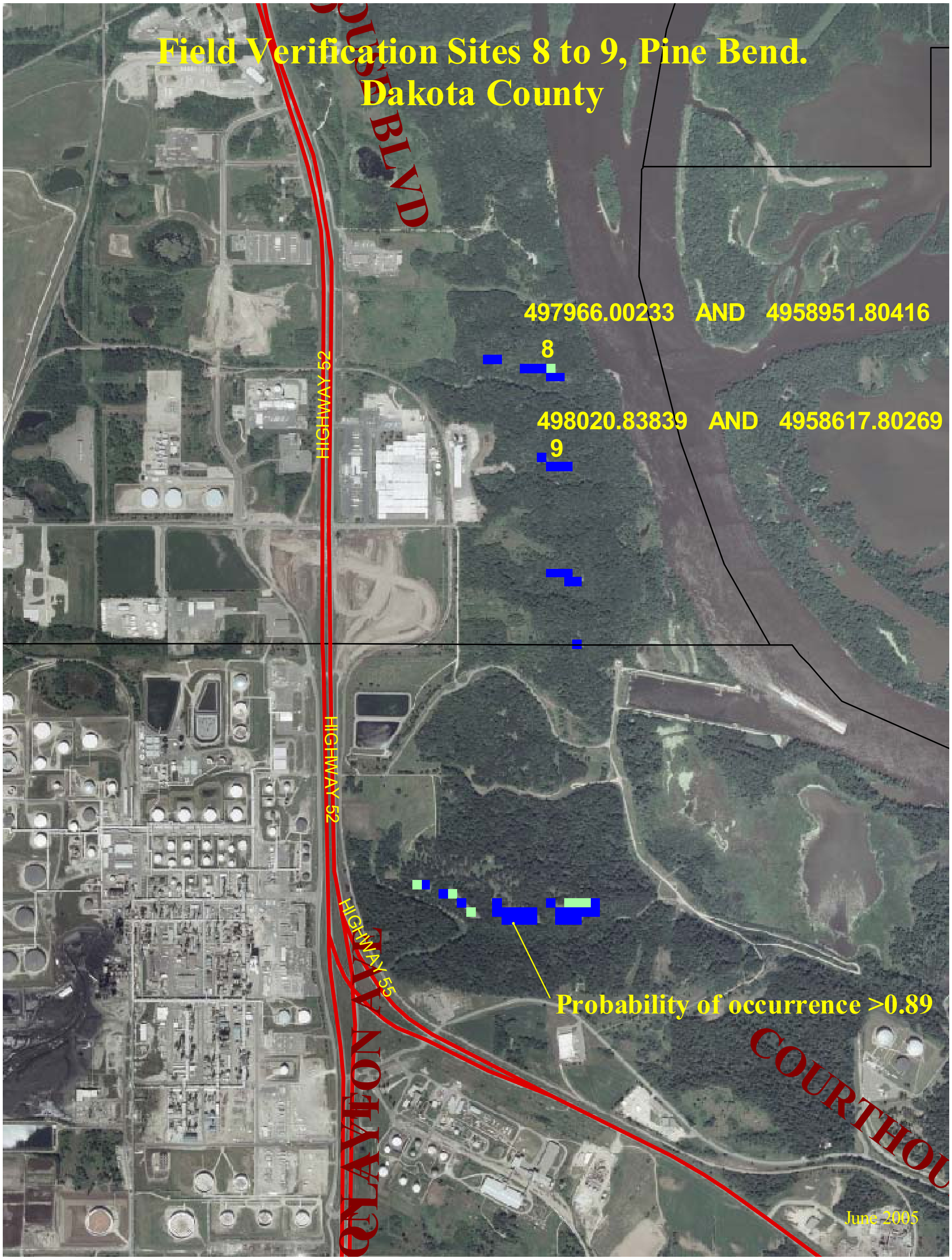

Field Verification Sites 8 to 9, Pine Bend.
Dakota County
497966.00233 AND 4958951.80416
8
498020.83839 AND 4958617.80269
9
HIGHWAY 52
HIGHWAY 52
HIGHWAY 55
Probability of occurrence >0.89
June 2005

Figure 10

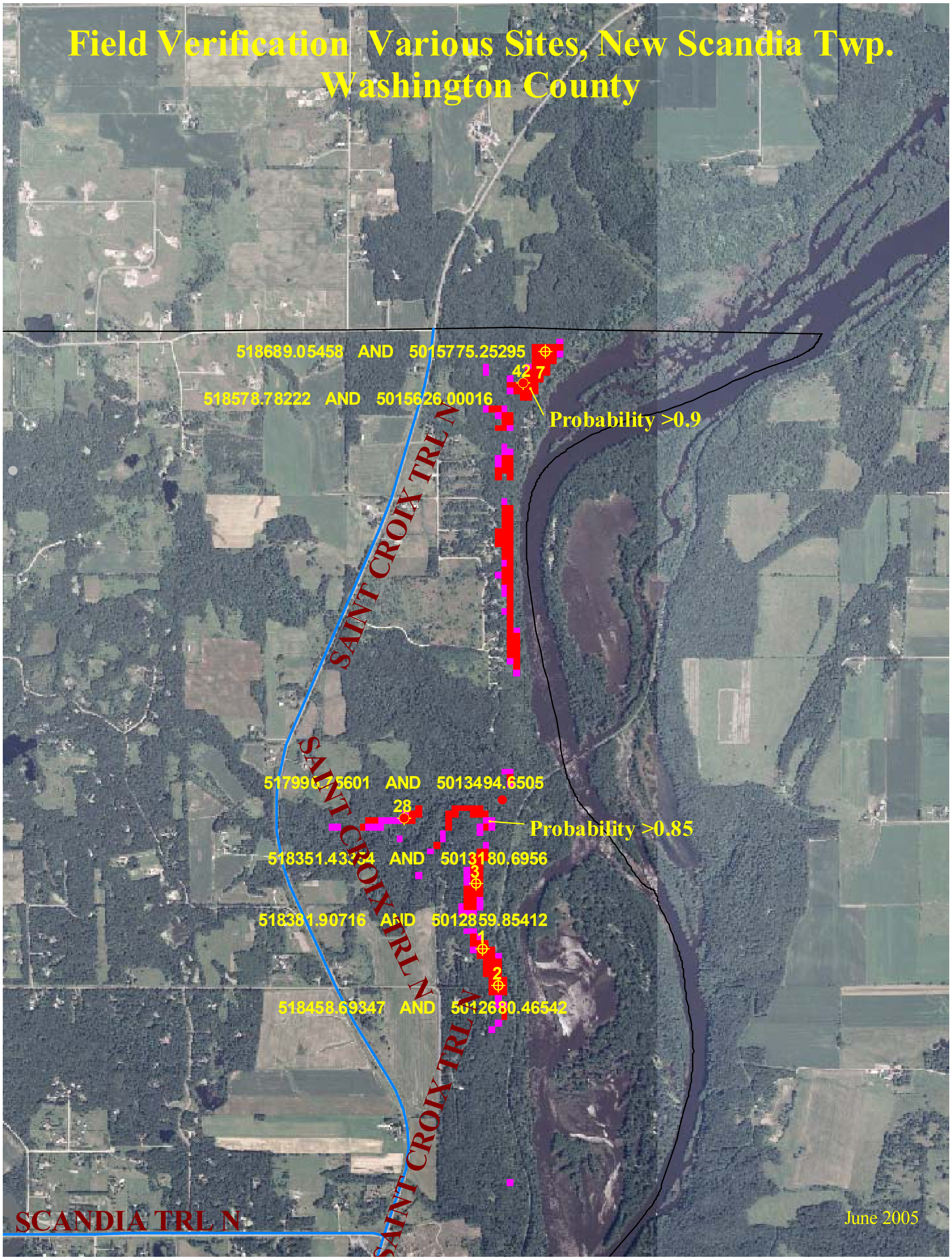
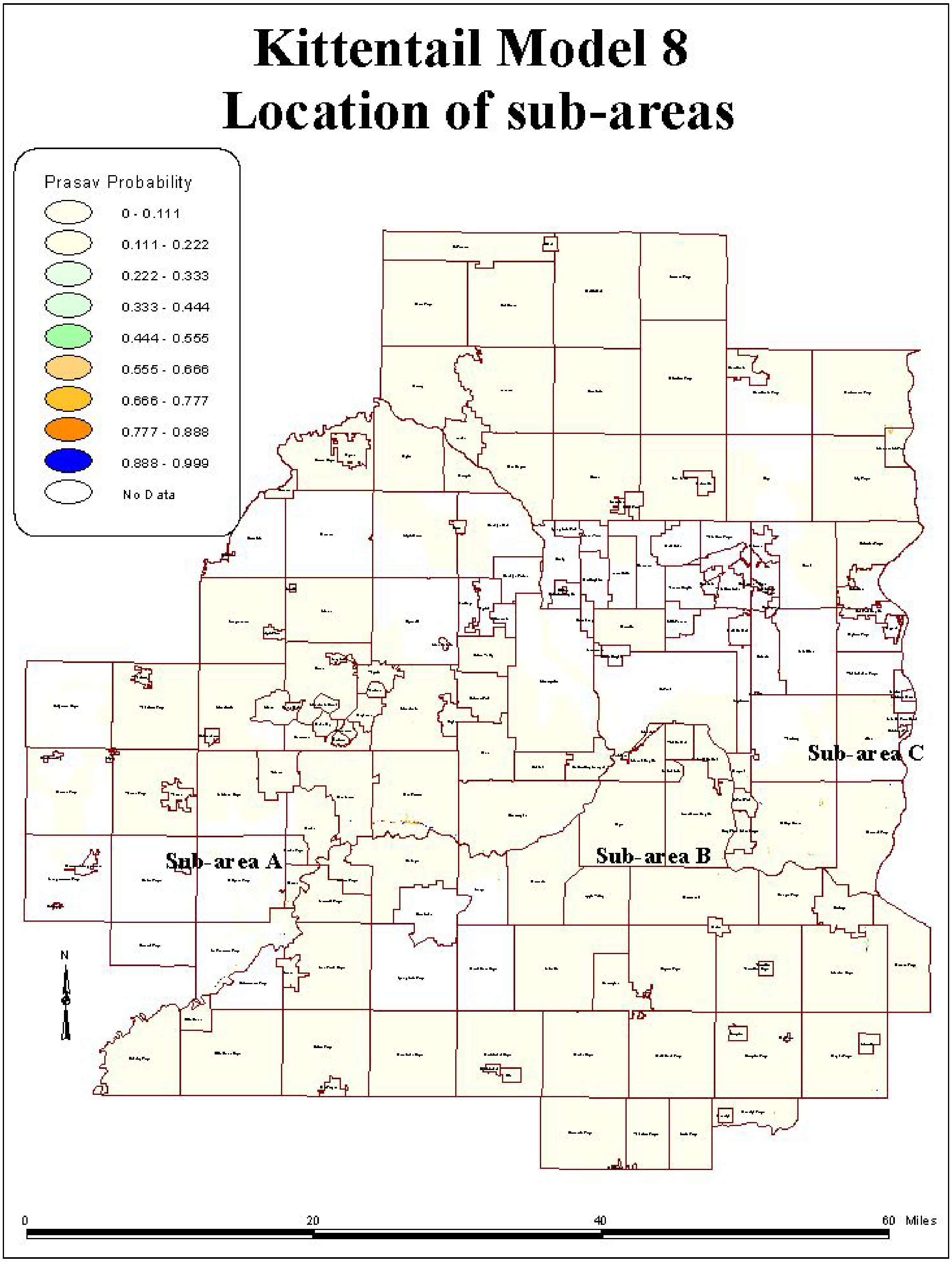

Field Verification Various Sites, New Scandia Twp.
Washington County
518689.05458 AND 5015775.25295
42 7
518578.78222 AND 5015626.00016
Probability >0.9
SAINT CROIX TRL N
517990.75601 AND 5013494.6505
28
Probability >0.85
518351.43354 AND 5013180.6956
3
518381.90716 AND 5012859.85412
1
2
518458.69347 AND 5012680.46542
SAINT CROIX TRL N
SAINT CROIX TRL N
SCANDIA TRL N
June 2005

Figure 11



Kittentail Model 8
Location of sub-areas
Prasav Probability
0 - 0.111
0.111 - 0.222
0.222 - 0.333
0.333 - 0.444
0.444 - 0.555
0.555 - 0.666
0.666 - 0.777
0.777 - 0.888
0.888 - 0.999
No Data
Sub-area A
Sub-area B
Sub-area C
N
0
20
40
60 Miles

Figure 12

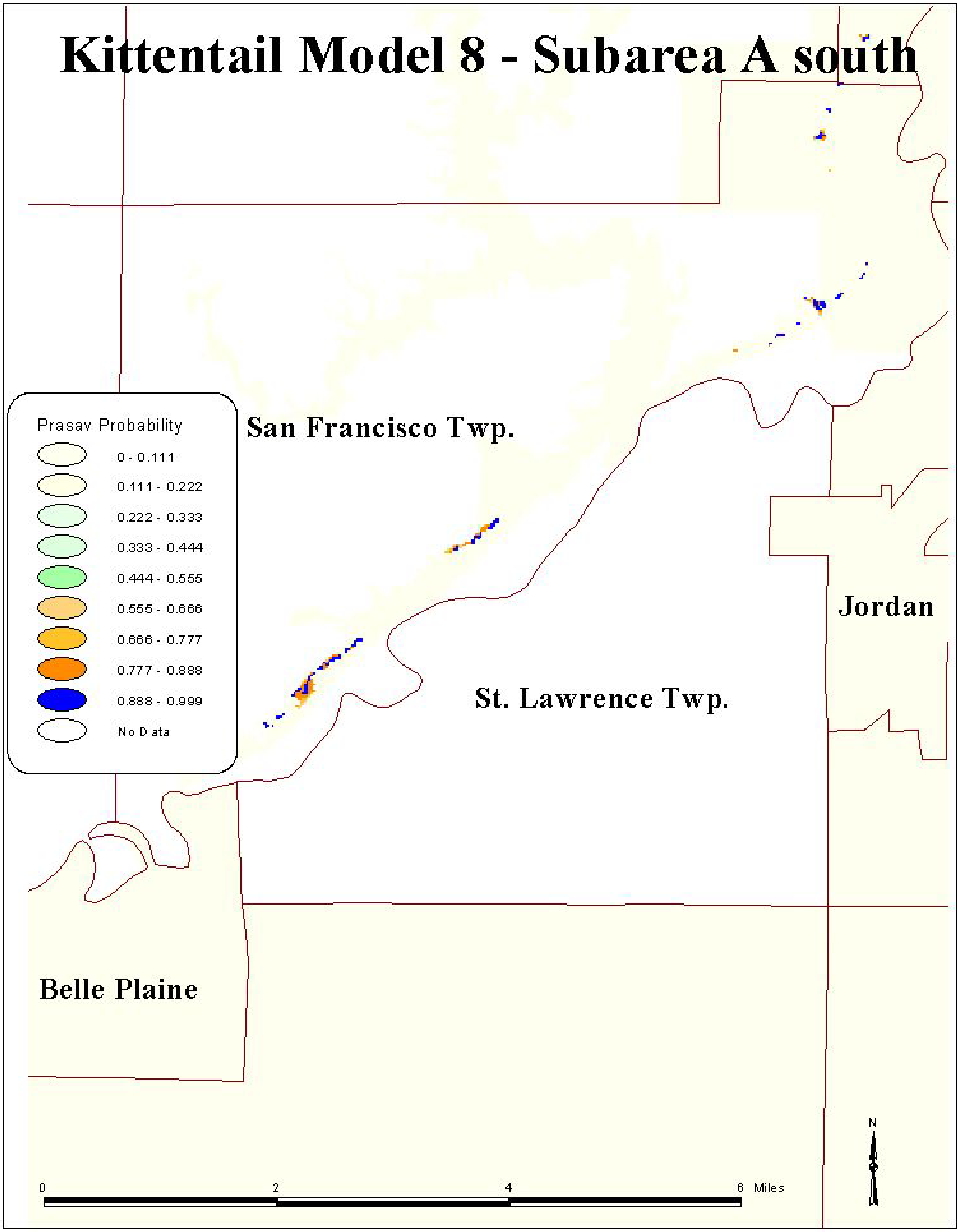

Kittentail Model 8 - Subarea A south
Prasav Probability
0 - 0.111
0.111 - 0.222
0.222 - 0.333
0.333 - 0.444
0.444 - 0.555
0.555 - 0.666
0.666 - 0.777
0.777 - 0.888
0.888 - 0.999
No Data
San Francisco Twp.
St. Lawrence Twp.
Jordan
Belle Plaine
0
2
4
6 Miles
N

Figure 13

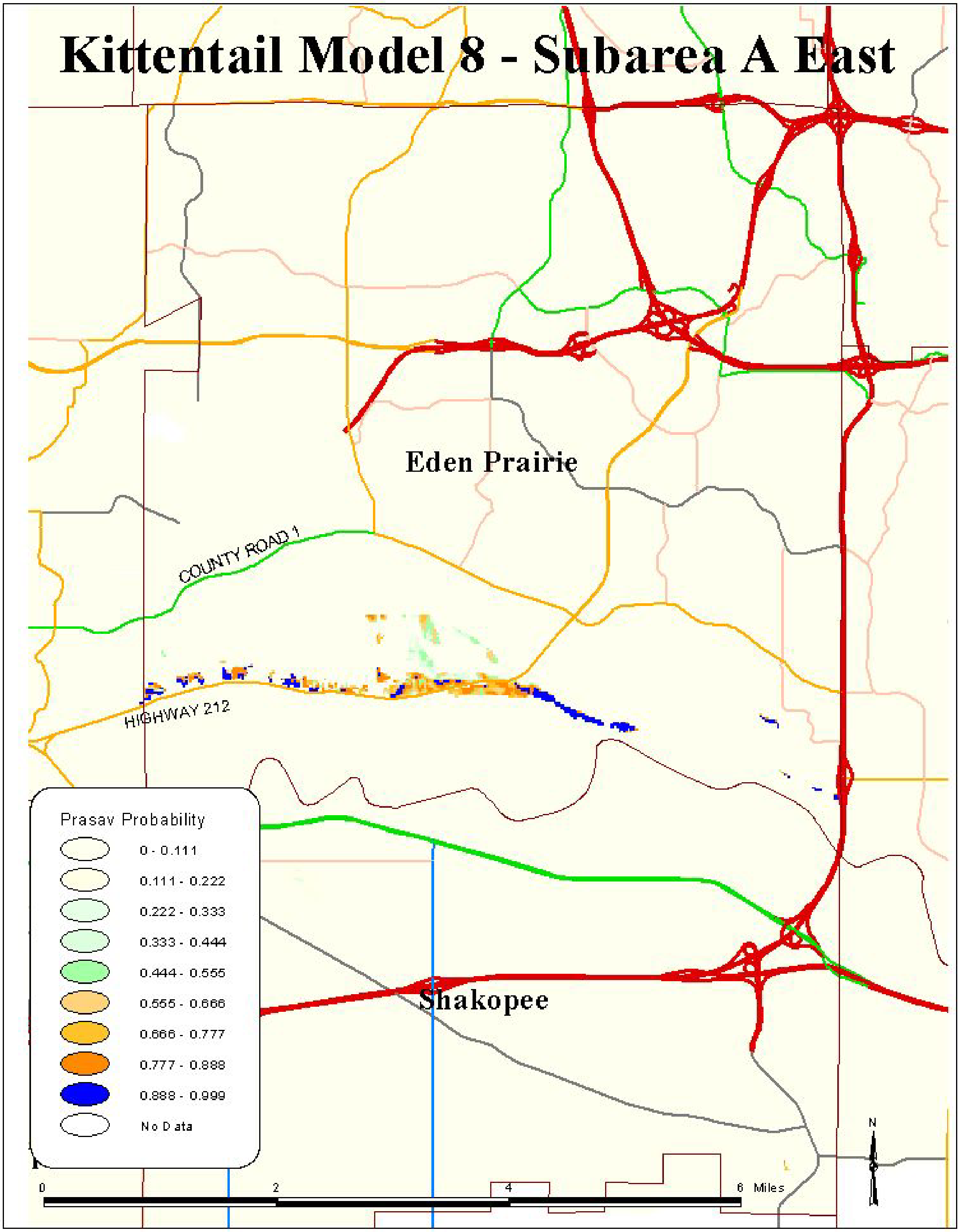
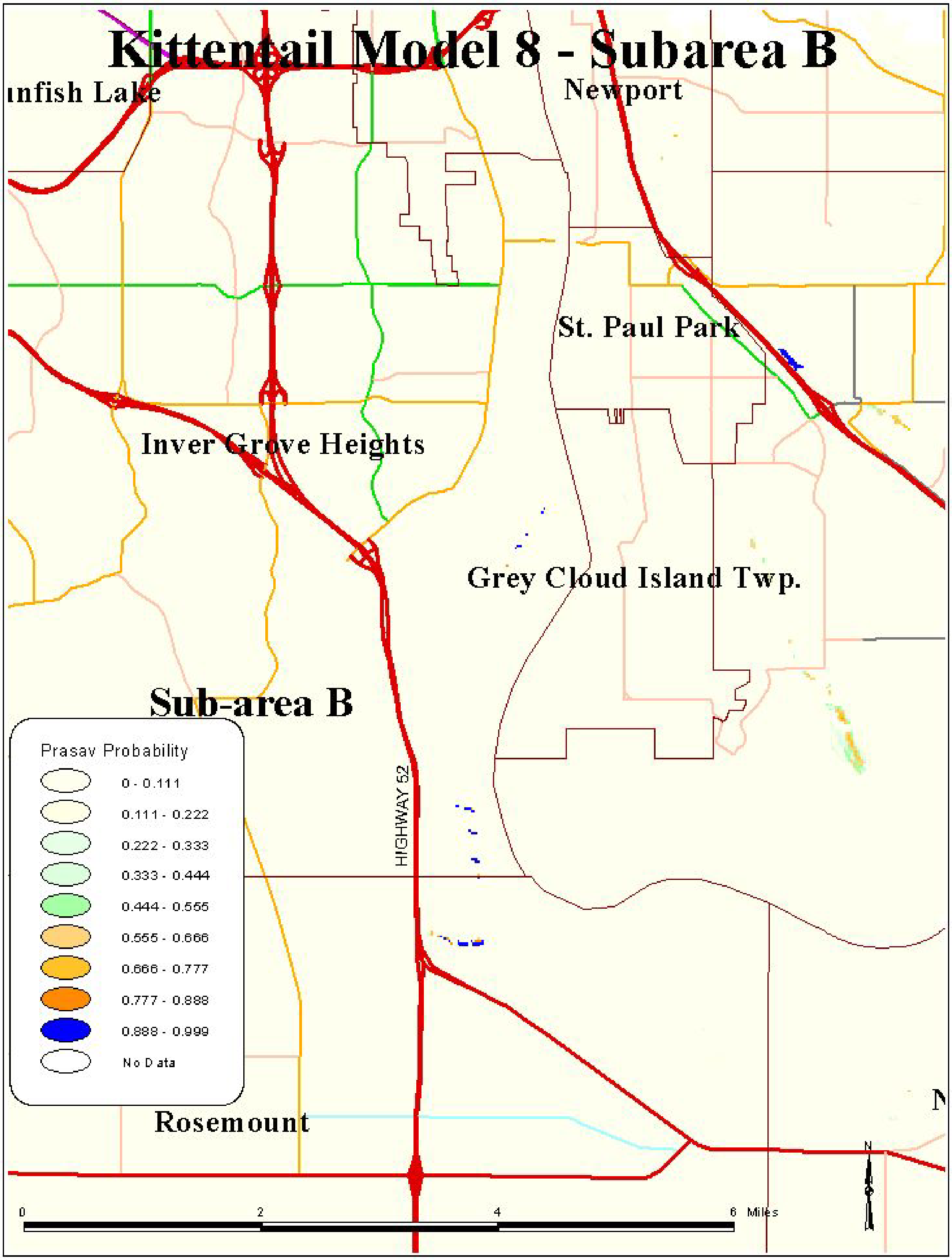

Kittentail Model 8 - Subarea A East
Eden Prairie
COUNTY ROAD 1
HIGHWAY 212
Shakopee
Prasav Probability
0 - 0.111
0.111 - 0.222
0.222 - 0.333
0.333 - 0.444
0.444 - 0.555
0.555 - 0.666
0.666 - 0.777
0.777 - 0.888
0.888 - 0.999
No Data
0
2
4
6 Miles
N

Figure 14



Kittentail Model 8 - Subarea B
ınfish Lake
Newport
St. Paul Park
Inver Grove Heights
Grey Cloud Island Twp.
Sub-area B
Prasav Probability
0 - 0.111
0.111 - 0.222
0.222 - 0.333
0.333 - 0.444
0.444 - 0.555
0.555 - 0.666
0.666 - 0.777
0.777 - 0.888
0.888 - 0.999
No Data
HIGHWAY 52
Rosemount
0
2
4
6 Miles
N

Figure 15

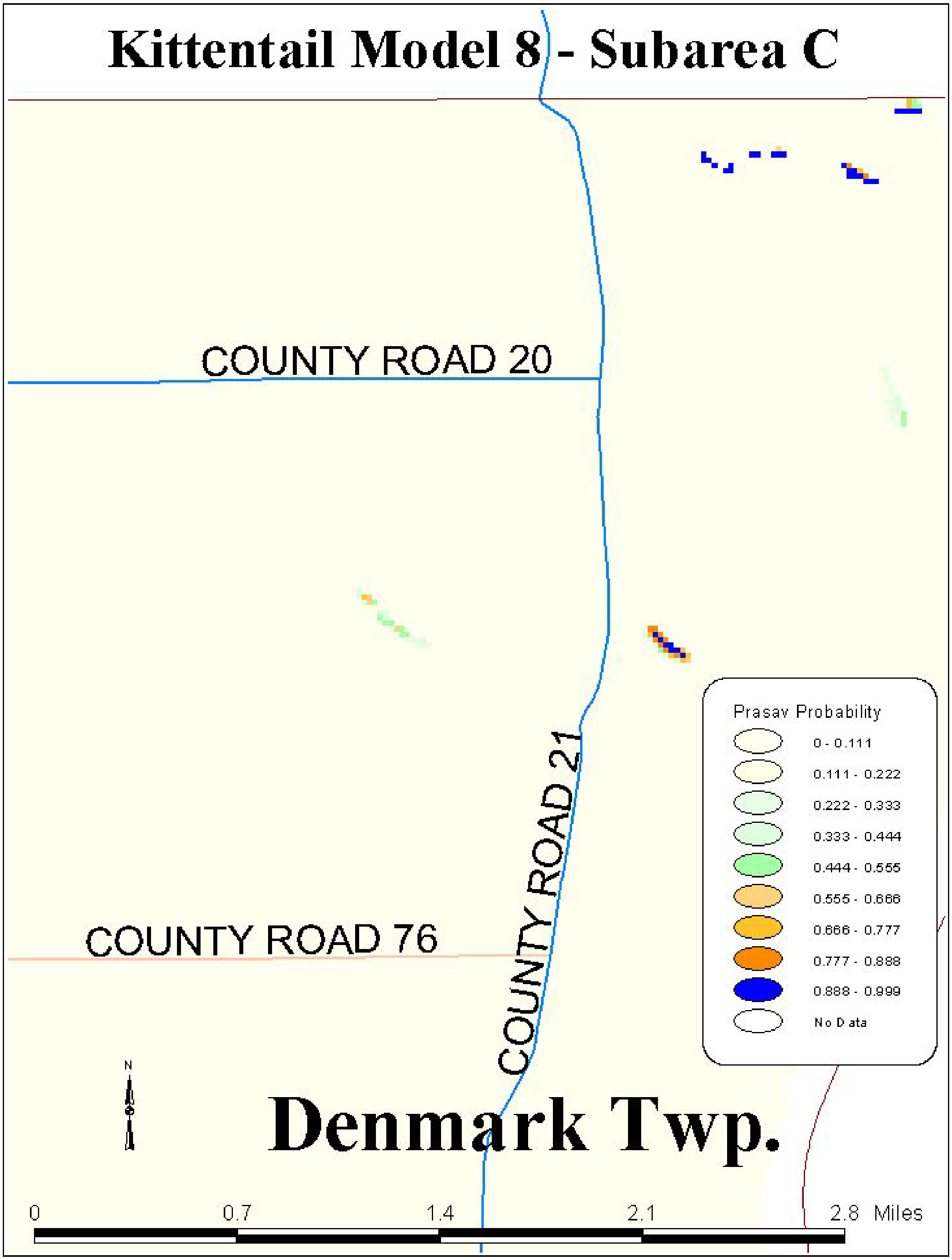

Kittentail Model 8 - Subarea C
COUNTY ROAD 20
COUNTY ROAD 21
COUNTY ROAD 76
Prasav Probability
0 - 0.111
0.111 - 0.222
0.222 - 0.333
0.333 - 0.444
0.444 - 0.555
0.555 - 0.666
0.666 - 0.777
0.777 - 0.888
0.888 - 0.999
No Data
N
Denmark Twp.
0
0.7
1.4
2.1
2.8 Miles

Figure 16

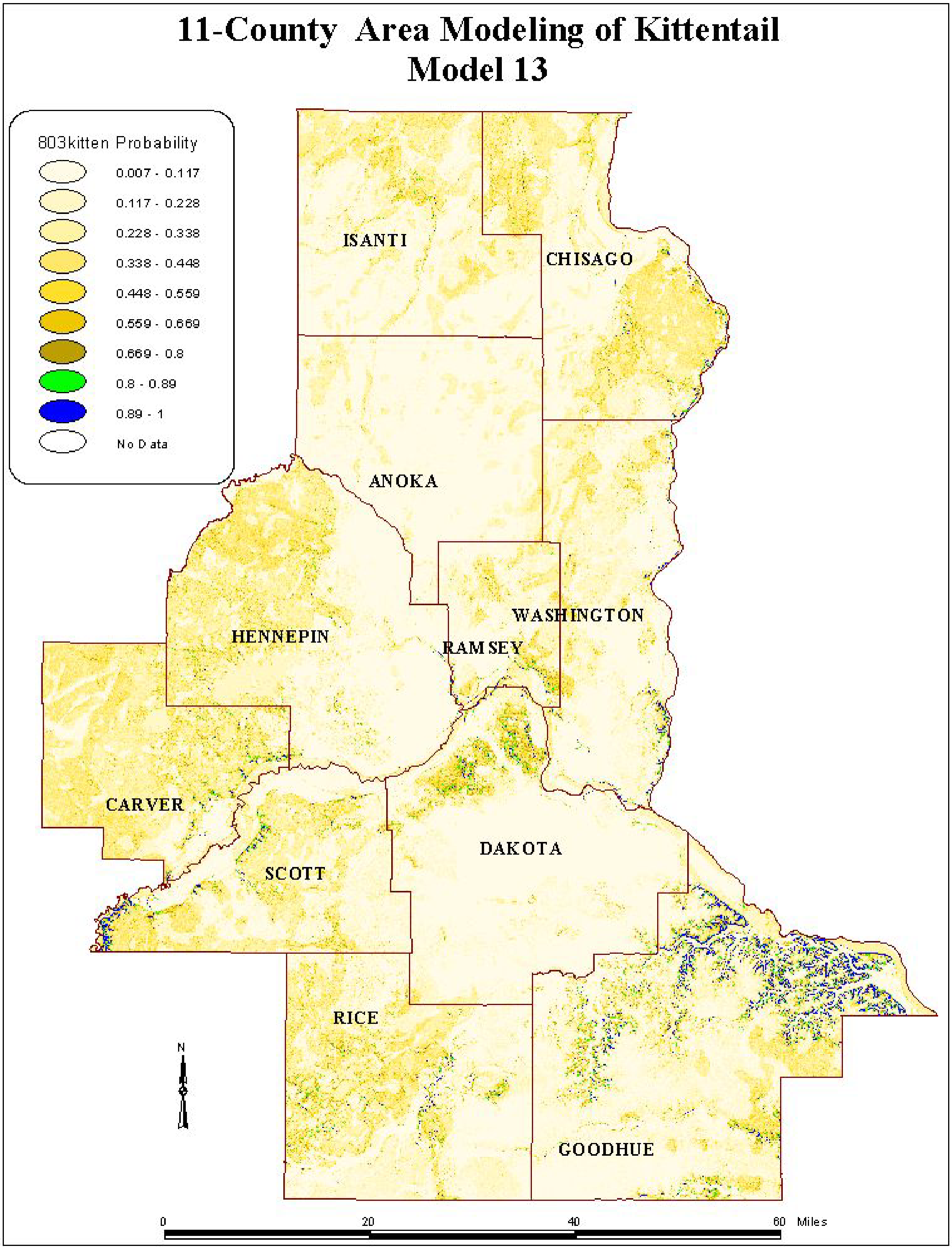

11-County Area Modeling of Kittentail
Model 13
803kitten Probability
0.007 - 0.117
0.117 - 0.228
0.228 - 0.338
0.338 - 0.448
0.448 - 0.559
0.559 - 0.669
0.669 - 0.8
0.8 - 0.89
0.89 - 1
No Data
ISANTI
CHISAGO
ANOKA
WASHINGTON
HENNEPIN
RAMSEY
CARVER
DAKOTA
SCOTT
RICE
GOODHUE
N
0
20
40
60 Miles

Figure 17

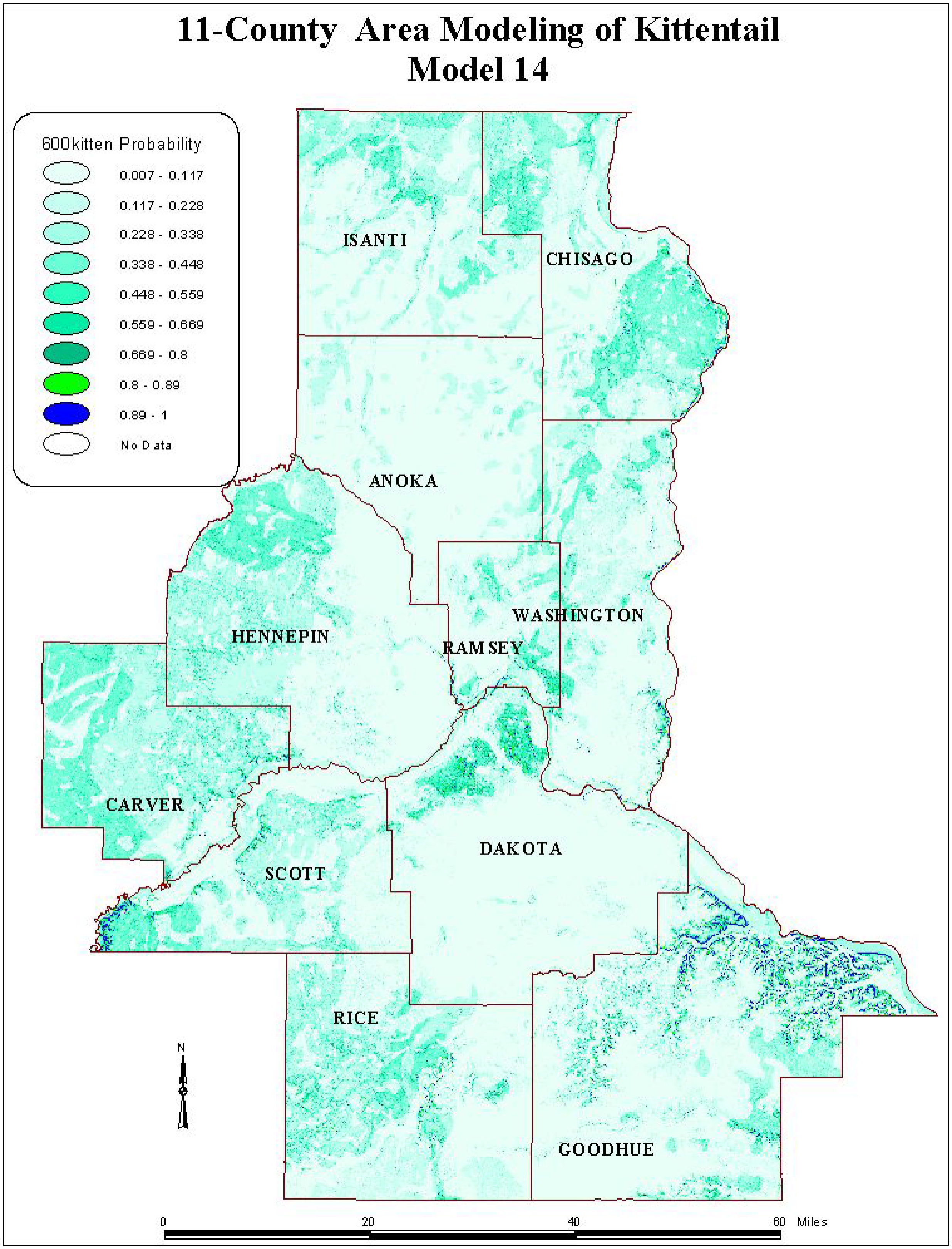

11-County Area Modeling of Kittentail
Model 14
600kitten Probability
0.007 - 0.117
0.117 - 0.228
0.228 - 0.338
0.338 - 0.448
0.448 - 0.559
0.559 - 0.669
0.669 - 0.8
0.8 - 0.89
0.89 - 1
No Data
ISANTI
CHISAGO
ANOKA
WASHINGTON
HENNEPIN
RAMSEY
CARVER
DAKOTA
SCOTT
RICE
GOODHUE
N
0
20
40
60
Miles

Figure 18