Spring 2010

Development of a Statewide Model of Culex tarsalis Habitat Suitability Using GIS

Jason Larsen
Carroll College, Helena, MT

Follow this and additional works at: https://scholars.carroll.edu/lifesci_theses

Part of the Biodiversity Commons, Biology Commons, Ecology and Evolutionary Biology Commons, and the Entomology Commons

Recommended Citation
Larsen, Jason, "Development of a Statewide Model of Culex tarsalis Habitat Suitability Using GIS" (2010). Life and Environmental Sciences Undergraduate Theses. 111.
https://scholars.carroll.edu/lifesci_theses/111

This Thesis is brought to you for free and open access by the Life and Environmental Sciences at Carroll Scholars. It has been accepted for inclusion in Life and Environmental Sciences Undergraduate Theses by an authorized administrator of Carroll Scholars. For more information, please contact tkratz@carroll.edu.
Development of a Statewide Model of *Culex tarsalis* Habitat Suitability Using GIS

Submitted in partial fulfillment of the requirements for graduation with honors from the Environmental Studies Program at Carroll College, Helena, MT

Jason Charles James Larsen
Carroll College, Helena, MT
April 13, 2010
This thesis for honors recognition has been approved for the Environmental Studies Program by:

Dr. Grant Hokit, Director,
Department of Natural Sciences

Dr. Murphy Fox, Director
Environmental Studies Program

Dave Marshall, Associate Professor
Department of Computer Science
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>4</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>5</td>
</tr>
<tr>
<td>FIGURES AND APPENDICES</td>
<td>6</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>7</td>
</tr>
<tr>
<td>METHODS</td>
<td>9</td>
</tr>
<tr>
<td>RESULTS</td>
<td>15</td>
</tr>
<tr>
<td>CONCLUSIONS</td>
<td>16</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>20</td>
</tr>
</tbody>
</table>
Acknowledgements

I would like to thank Grant Hokit for his continued guidance, Cathy Maynard and Linda Vance for their help in obtaining environmental data layers, MAGIP for a scholarship received, the Wiegand Undergraduate Research Laboratory for providing the research facilities, and the Montana INBRE foundation and the Margaret A. Cargill Foundation for their financial support of the project.
Abstract  My study combines the use of the maximum entropy approach to species habitat prediction with the need for a West Nile Virus (WNV) infection risk model for the state of Montana to produce a statewide null model of *Culex tarsalis* habitat to be used as the basis for a future infection risk model. My study assumes that *C. tarsalis* is the primary vector for WNV in Montana, and that MAXENT software can reliably generate a habitat suitability model for *C. tarsalis* throughout Montana using presence only data. I successfully generated a statistically sound, statewide model of *C. tarsalis* habitat using MAXENT software, *C. tarsalis* presence data, and readily available environmental datasets. Presence points were determined by verification in the lab via microscopic determination of the target species from mosquito samples collected over a four year period. Samples were acquired by me and my undergraduate colleagues, as well as by graduate students from MSU and a number of cooperators around the state, and the sample site locations were verified either through direct GPS measurement or a combination of GIS (Geographic Information System Software) and aerial photos. The final product is a map of the state showing potential *C. tarsalis* habitat in fourteen increments of increasing suitability (indicated by color), as well as a set of charts produced by the software demonstrating the statistical methods used in determining the various levels of suitability predicted by the model. In the process of creating the final model, a series of preliminary models were generated, and unnecessary environmental layers eliminated to ensure that the final model uses as few variables as possible to produce statistically significant results. Included along with the final model results, in appendices to the report are examples of statistical data and maps generated in preliminary model attempts that demonstrate the reasoning behind refinements made at various stages of the model’s evolution.
FIGURES AND APPENDICES

APPENDIX A.................................................................................................................... i - ii
Datasets and Their Sources................................................................. i
Correlated Landcover Types................................................................. ii

APPENDIX B ........................................................................................................ iii - vi
Figures 1A- 1B................................................................. iii
Figures 2A- 2B................................................................. iii
Figure 3................................................................. iv
Figure 4A ................................................................. iv
Figure 4B ................................................................. v
Figure 5A ................................................................. v
Figure 5B ................................................................. vi

APPENDIX C ....................................................................................................... vii - xi
Figure 6................................................................. vii
Figure 7................................................................. viii
Figures 8A – 8C................................................................. ix
Figures 9A – 9B................................................................. x
Figure 10................................................................. xi
**Introduction:** West Nile Virus (WNV) is a virus that is transmitted to birds and mammals, including humans, via a mosquito vector (Hayes Et al, 2005). WNV is divided into two genetic variants, one of which is can result in encephalitis (sometimes resulting in death) in humans and other mammals, particularly horses. WNV originated in Uganda in the 1930s and has since spread to many portions of the world, including the United States. During the 1990s, there was an increase in the frequency of outbreaks, including severe outbreaks in humans correlating with outbreaks in birds and horses (Peterson, 2001).

WNV arrived in the Western Hemisphere in 1999, first infecting the East Coast of the United States including New York, New Jersey and Connecticut (Peterson, 2001). The virus rapidly spread across the continental United States reaching western states and the pacific coast in just a few short years. The rapid spread of the disease and record number of cases experienced throughout various regions of the U.S. and Canada, including the western United States, indicate a need for more research into spatiotemporal patterns associated with the spread of the disease in order to properly manage the threat it poses (Artsob Et al, 2009).

West Nile virus was first detected in Montana in 2002 (CDC). Since then, a number of state and local agencies and university research centers have worked to document and study the annual cycle of the disease; coordinating with federal agencies in tracking, reporting and educating the public about the spread of West Nile Virus. This effort to understand the spread of the disease and to alert the public to geographic areas of highest risk within the state continues to be a priority for West Nile researchers.
*Culex tarsalis* is the primary vector for West Nile Virus in the Western United States. *C. tarsalis* has increased its distribution due to irrigation in semi-arid regions (Pahk, 2003). Thus the null model for this study is a habitat suitability model for the mosquito species *C. tarsalis*.

Similar habitat suitability models have been generated for *C. tarsalis* and related species for other geographic areas, with the goal of modeling WNV risk. One such model was generated for the province of British Columbia (Tachiiri Et al, 2006). Though similar in concept, it did not use the same specific modeling method (maximum entropy) or software used for my study.

The MAXENT approach for modeling species habitat suitability starts with a geographic region represented as a grid with finite resolution. Inputs include a set of environmental variables representing each cell of the grid and a set of occurrence points that fall within the boundaries of the region. The software then uses these inputs, and an algorithm to evaluate each variable at each occurrence location and then estimate the likelihood of species occurrence for all cells in the grid based on a distribution of maximum entropy (Phillips Et al, 2004). MAXENT software has been found to be effective and reliable for predicting species occurrence when only presence data is available (Phillips Et al, 2008).

Habitat suitability for various other species have been modeled for the state of Montana, using MAXENT, with good results (Maxell, B. 2009).

**Hypothesis:** The Maximum entropy method and presence only data, in conjunction with pseudo-absences, can be used to successfully generate a statewide predictive model of suitable *C. tarsalis* habitat for Montana using readily available environmental data. A
model generated in this method, and tested against a sub-sample of presence points not used in the training of the model will be statistically valid with AUC (area under the curve) values of at least 0.95 for both model training and model testing.

Methods: Habitat suitability and statistical analysis were generated using GIS data and MAXENT software. The need to use maximum entropy modeling arises from limited number of presence points available for *C. Tarsalis* relative to the large geographical area of the state, and lack of absence points. The software generates a set of absence data randomly from all non presence points – called pseudo-absences. This method of modeling achieves a high level of statistical accuracy when a limited number of presence points and no absence points are available. (Elith Et al. 2006)

My project involved trapping mosquitoes using CDC (Center for Disease Control) light traps at various locations of suspected *C. tarsalis* habitat throughout the state. With the help of undergraduate colleagues, mosquitoes were then sorted, *C. tarsalis* removed from the general sample and quantified. The geographic coordinates of the trap site locations were recorded using GPS readings taken when the trap sites are placed (when placed by members of the project) or by verification in GIS software using georeferenced aerial photography (when 3rd party cooperators aid in trapping).

In addition to species presence data, a number of continuous statewide raster datasets were acquired for different environmental variables. These variables were input into MAXENT along with the presence data and several test models were generated. Utilizing a commonly accepted convention, seventy five percent of presence points were used to generate the model, and 25% were used to verify it. By comparing the statistical outputs generated by the software, the number and types of input variables were narrowed and a final model produced using the most statistically relevant variables.
Environmental data acquisition: Statewide GIS datasets for various environmental variables were downloaded, including: elevation, soil, land cover, land use, average precipitation, relative effective annual precipitation, average maximum temperature, average minimum temperature, surface water locations, humidity, wind speed data, and wind power data (wind speed contributed the least relatively, but the performance of wind power data in the model relied on its being coupled with wind power). Slope data was derived from the elevation dataset using the surface evaluation tools in ArcGIS (Spatial Analyst).

For a more detailed description of datasets used and their sources, see appendix A.

Relative Effective Annual Precipitation (REAP) data was compiled by the Montana Natural Resources Conservation Service (NRCS) and provided by the Natural Resource Information System (NRIS) of the Montana State Library. Initial attempts at using REAP data from data downloads failed as mosaicing of the many individual edge-matched datasets proved too difficult. A continuous statewide REAP raster dataset and updated (RE-GAP) landcover dataset, provided by Cathy Maynard and Linda Vance of the NRCS, did not require mosaicing and worked well in the model.

Data & Data Layer Control: During preliminary model runs and data preparation, it was found that some data layers were problematic and/or did not contribute much statistically to the model. For these reasons, not all data layers were used in generating the final model (see Appendix B for examples of preliminary models and statistics leading to the removal of specific points and data layers). The final model relied on only the following datasets: elevation, landcover, REAP, slope, wind power and wind speed. Elevation, slope, REAP, and the wind information datasets are continuous. The re-gap landcover data is categorical.

Producing a statewide raster layer from various vector layers for surface water proved difficult. Data at the resolution required to include small ponds that would be likely mosquito breeding grounds, was not available statewide. During the initial stages, a
composite layer containing both large and small scale vector datasets was produced, and then rasterized. The intent was to create a distance-to-water raster. However, preliminary runs of the model with this dataset were overfit (likely because so many of the sample points were chosen near waters edge) and showed potential habitat extending both outward from the water’s edge as well as inward from the water’s edge to include the open water areas of large water bodies. As such, the distance to water raster was dropped from the model. (See Appendix B figures 2A & 2B – showing an example of exaggerated potential habitat over water).

Average precipitation was not statistically useful in initial model runs, and tended to over-predict suitability in areas of steep slope where mosquito breeding is unlikely to be occurring. Average precipitation data was thrown out early in the preliminary modeling phase during modeling of a pilot area within the state. Montana REAP (Relative Effective Annual Precipitation) data was used in place of average precipitation data for all preliminary attempts at statewide modeling and in the final model. REAP data, which is a measure of effective precipitation rather than actual precipitation, and takes into account slope and other variables in its calculation, more accurately reflects the effect of precipitation in rugged terrain, especially mountains and mountain valleys which may receive relatively little direct precipitation but may have much larger amounts available do to overland runoff and aquifer flow. Use of the REAP data corrected the problems of over prediction on steep slopes that occurred in preliminary model runs, but could be introducing some redundancy into the model, as slope information is a part of the REAP data calculation.

Although studies show a high correlation between the transmission of WNV by C. tarsalis and temperature (Reisen Et al, 2006), and mosquitoes have certain minimum temperature requirements for development and activity, average temperature data didn’t prove as useful to the model as was initially expected. The effect of temperature on the
model, however may require further investigation in the future, other temperature related inclusions such as degree-day information may prove beneficial. Average minimum and maximum temperature datasets were ultimately dropped because they contributed little to the model statistically. The tables in figure 3 of Appendix B show an analysis of variable contribution when max and min temperature and humidity are included in a variation of the model otherwise similar to the final model. Although the contribution of minimum temperature is not completely irrelevant in this example, its relevance seems to be highly dependent on its being coupled with maximum temperature and both data layers were removed from the model.

Soil data, which also held a lot of promise at the onset of the project, ultimately had to be eliminated, at least for this stage of the modeling. Neither of the two soils datasets available at the time, the STATSGO2 (State Soil Geographic Database) and SSURGO (Soil Survey Geographic Database) soils datasets, could be incorporated into the final model.

Use of the older, STATSGO data was not advisable because of the coarseness of its spatial resolution, the age of the data, and because the data is too generalized. The newer, higher resolution and more correct SSURGO data (which is of sufficient scale to capture small geographical areas such as those that might represent mosquito habitats) was not available on a continuous statewide basis during the early phases of modeling. Though only a few small areas were without information, it would have made running the model in MAXENT problematic (the MAXENT software requires continuous and complete data set in order to process). In order to produce preliminary models using the available dataset, areas of missing data had to be filled in with the value “no data”, which allowed the model to run but left holes in the model map produced (see figure 4A of appendix B). Additionally, the SSURGO data is a very complex dataset and will require additional processing in order to be relevant to the needs of this model.
Though statewide model runs using the incomplete SSURGO data were successful (with areas of “no data”), because of the nature of the data, the model relied too heavily on the soils data layer. The problem lies in the correlation between the presence points and the soil types in which they occurred. Areas of potential habitat were limited to only those very specific soil types in which actual presence points existed, while ignoring numerous similar (same series, but different phase) soil types which are likely to occur in potential mosquito habitat. (see Figure 4B of appendix B).

SSURGO soils data still holds future promise of increasing the accuracy of the model. However, it is likely going to require more than just the complete statewide dataset (now available). The SSURGO data is highly complex with individual soil polygons generated according to soils classification at a high level of soil taxonomy, namely various phases of individual soil series. In addition, the categorical nature of the data, combined with the large number of categories makes it difficult to work with for this type of analysis. In order to be used in the model, one of two things will likely need to be done with the data. Either similar soil types will need to be aggregated to produce a more reasonable number (fewer) of discrete data types (essentially a reclassification at a lower level of soil taxonomy); or, a dataset of a specific soil property could be derived from the data. For example, a derived dataset for either soil particle size or water retention properties could prove useful.

*Environmental Data processing:* The acquired and derived raster datasets were processed using ArcGIS (versions 9.2-9.3), to prepare them for analysis in MAXENT. In order to meet MAXENT’s formatting requirements, environmental datasets were re-sampled, re-formatted, re-classified, extracted (by mask), and re-projected as necessary to produce complete, coincidental statewide raster datasets with the same resolution (90 meter) for all variables.
Sample Data Acquisition: GIS point data of successful *C. tarsalis* trap locations were used as the sample input in MAXENT. These presence point locations were provided by statewide cooperators, MSU Bozeman, and the research team from the Wiegand Undergraduate Research Center at Carroll College.

GIS point locations provided by Carroll and MSU research teams were taken using a handheld GPS receiver at the site of the mosquito trap locations. This method of acquiring locations was deemed adequate because the approximately 15 meter accuracy of such units is well within the 90 meter resolution used for all data layers of the model.

Some cooperator trap site locations were determined by direct GPS measurement, others were verified with cooperator assistance using aerial photography. Photos printed from Google Earth, or in ArcGIS with the Digital Orthophoto Quarter Quadrangle (DOQQ) 2005 statewide dataset as a backdrop were marked up with exact trap site locations by the cooperators and those markups compared to the 2005 DOQQ photography in ArcGIS to acquire the coordinate locations. Easily identifiable landmarks such as road intersections, water bodies and building corners assured that coordinates are well within the 90 meter resolution.

Trap site location data was reviewed for all locations with positive *C. tarsalis* identifications. In order to generate the most statistically significant data for species specific habitat suitability, only the data points for locations with relatively high numbers of *C. tarsalis* were used as sample inputs in the model. In some trap site locations not used in the model, *C. tarsalis* presence was as low as 2 in 500 individuals. Avoiding points for locations dominated by other mosquito species should help prevent overestimating habitat suitability in marginal areas and concentrate model predictions to areas where the species readily occurs.

For some trap locations, primarily cooperator locations and sites monitored by the MSU researchers, data has been collected historically since 2006. In other locations, primarily
those traps set by the Carroll research team, 2009 was the first year of trapping, and the only year for which data was available. Attempts were made to include as many of the locations for which two or more years of data were available. The final model was generated with a total of 40 presence points, 30 (75%) used for modeling and 10 (25%) for testing. In some instances, trap locations of relatively high *C. tarsalis* presence had to be omitted from the model because of conflicts with the landcover dataset. This was not surprising as several trap locations were within 90 meters of water. In one instance a point occurred in a location designated as “water” in the landcover data layer. For two other points, though they fell just outside of areas designated “water,” they were close enough to cause preliminary models to over-fit to waterways. Removing these points reduced the over-fit to waterways, and had little other effect on the model, other than a slight reduction in the number of useful points (these eliminations being prior to the final point counts of 30 and 10 for modeling and testing, respectively) (See appendix B – figures 1A &1B – which show an area of modest habitat suitability covering the interior of Canyon Ferry Lake).

**Results:** Elevation was negatively associated with habitat distribution decreasing as elevation increased and with a very low likelihood in high elevations habitat suitability is almost non-existent above 2000 meters.

Slope was also negatively associated with habitat. The highest habitat suitability occurs in areas with little or no slope. Predictions are the highest in areas with approximately five degrees or less.

Wind speed is nearly inversely proportionate to habitat suitability, with the highest levels of suitability occurring between 2 and 4 m/s (meters per second), and almost no suitability in areas with wind speeds above 10 m/s.
Wind power had a drastic, negative effect on suitability above levels of approximately 400 W/m² (watts per square meter), with the majority of suitable habitats occurring between 0 and 400 W/m².

REAP also showed an inversely proportionate relationship with highest suitability occurring around 30 centimeters and dropping to virtually no suitability in areas with REAP values above 100 centimeters. This relationship reinforces the need to incorporate soils data (and possibly irrigation data) in future generations of the model, as mosquito habitats tend to be damp, yet occur in areas of relatively low rainfall/effective rainfall. It may be that hydric and/or poorly drained soils account for this phenomenon.

The re-gap landcover data, being categorical correlated to fifteen specific landcover types found at or near the location of presence points. The landcover types correlating to the model are listed in a chart in Appendix A.

**Conclusions:** REAP is derived, in part, from slope – which may cause some problems with the model due to redundancy when REAP and slope data are both used.

Wind speed and wind power datasets were originally at a much lower resolution (400 meters) than the 90 meter resolution used for modeling, and so are overly generalized values, having been re-sampled to a higher resolution.

Another variable which was not used, even in preliminary models, may have been useful to the model. Aspect, the exclusion of which was an oversight, is easily derived from elevation data. The relevance of aspect should be explored in future years during refinement of the model.

The location of sample points near waterways have a tendency to overfit the model to waterways, even after certain problematic points have been removed - though not nearly as much as in initial runs of the model.
The objective of my study was to generate a statewide spatial risk model of West Nile Virus (WNV) infection risk by using the distribution of C. Tarsalis as a null model. The result of my modeling efforts is a statewide map showing variable habitat suitability along with a set of statistical outputs which qualify the accuracy of the model.

Mosquitoes require aquatic habitats for the larval stage of their life-cycle. They thrive where water is shallow and less likely to contain fish and other predatory species in abundance. (Juliano, 2009)

Given what is known about mosquito ecology, it is not surprising that the model shows high likelihood of suitability in areas that are relatively flat, near (or covering) water bodies and likely to be wet enough for the mosquito larval stage.

In fact, initial models showed areas of extremely high suitability directly over nearly all waterways, including deep water such as large lakes and major rivers. This over-fitting of the model to deep open water was corrected for in subsequent model generations.

A number of preliminary models were run in MAXENT using varied combinations of environmental variables. Initially all variables for which data was acquired were used. The only exception being the SSURGO soils data, which would not run in MAXENT because of the data layer not being complete for the entire state at the time the modeling was being conducted. The data is now available, but due to time restrictions, was not included in this analysis. (In future project years, attempted incorporation of SSURGO soils data into the model is advised).

The first models contained a number of data layers which did not contribute significantly to the model, determined by the jackknife statistics generated by the MAXENT program. With each succession of the model, one or more of the least significant data layers was removed, and the model was reprocessed. Additionally, AUC rates for both testing and training continued to rise from 0.9 until the final rates of 0.979 for training, and 0.966 for
testing were achieved in the final model. Although some variations of the model had a suspect “perfect” fit of 1.0 for training, the paired testing AUC rates for those fits were less than acceptable and suggested that those model variations were unreliable. One such example are the AUC/ROC numbers generated along with the model variation including the SSURGO soils data, in which the soils dataset contributed over 84% (see figure 5A in appendix B). Omission on model training for this model was less than random predicted for omission (figure 5B).

A total of six environmental datasets remained in the final model, contributing to the highest AUC rates: elevation, slope, wind power, wind speed, landcover (categorical Re-GAP data), and REAP (Relative Effective Annual Precipitation). Models generated using only occurrence data often show bias, particularly spatially auto correlated biases. However, studies performed to evaluate the performance of presence only distribution models have shown the maximum entropy method (MAXENT) to be particularly good at minimizing bias and producing accurate models (Phillips Et al, 2006). Attempts were made to keep biases to a minimum. A subset of the presence points were used to test the model, and the statistical data generated by the preliminary model runs was carefully evaluated. Statistical analysis of the final model, with AUC values for training and testing surpassing the conventionally accepted 95% threshold, was generally indicative of an accurate model.

There is a possibility of some model bias due to the sampling bias, with samples being taken from areas of suspected species occurrence. However, these sampling biases are based on known aspects of mosquito habitats. In future years, additional sample sites will add to the number of presence points, with traps set in areas less likely to capture the subject species, and eventually absence data will be included in a model that doesn’t rely only on pseudo-absences.
The final model generation produced a statewide map of potential habitat for *C. tarsalis*, with fourteen colors representing levels of increasing suitability from 0.0 (lowest) to 1.0 (highest). Included on the map is a legend showing the value represented by each color. The locations of each point used by the software in generating the model are included on the map, with color differentiation for training and testing. Assuming WNV is only found were we find *C. tarsalis*, this model can serve as a null model for WNV infection risk for Montana.

Accompanying the map are the omission/commission, Receiver Operator Characteristic (ROC), and jackknife statistics, as well as two sets of response curves. One set of response curves shows the response of each variable as it contributes to the success of the model independent of all other variables, and the other shows the contribution of each variable in combination with all other variables.

For the final model represented in map form, and all associated statistical data generated by MAXENT – see Appendix C.
References:


CDC http://www.cdc.gov/ncidod/dvbid/westnile/Mapsactivity/surv&control02Maps.htm


Hayes EB, Komar N, Nasci RS, Montgomery SP, O'Leary DR, Campbell GL. Epidemiology and transmission dynamics of West Nile virus disease. Emerging Infectious Disease. Volume 11, No. 8, August 2005


Maxell, Bryce. “Predicted Distribution and Landscape-Level Habitat Suitability of Amphibians and Reptiles in Montana” Wildlife Biology Program, College of Forestry and Conservation, University of Montana, Missoula


Reisen, William K.; Fang, Ying; Martinez, Vincent M. “Effects of Temperature on the Transmission of West Nile Virus by *Culex Tarsalis* (Diptera Culicidae)” Journal of Medical Entomology, Volume 43, Number 2, March 2006, pp. 309-317(9)

Appendix A

Dataset and Source Information:

Average minimum temperature, average maximum temperature, and humidity data came from the “Daily Surface Weather Data and Climatological Summaries” (DAYMET) database.

- Humidity datasets were downloaded for the entire contiguous continental U.S. from the University of Montana, Numerical Terradynamic Simulation Group’s website at: http://www.daymet.org/. The dataset was then clipped to the Montana boundary.
- Average minimum and maximum temperature datasets, for the state of Montana only, were downloaded from the NRIS GIS data list website at: http://nris.mt.gov/gis/gisdatalib/gisDataList.aspx. These temperature datasets are also available from the DAYMET website for the contiguous continental U.S.

Relative Effective Annual Precipitation (REAP) data was compiled by the Montana Natural Resources Conservation Service (NRCS) and the Natural Resource Information System (NRIS) of the Montana State Library. Initial attempts at using REAP data were from data downloads available from the NRIS REAP data website (http://nris.mt.gov/nrsc/reap/datapage.asp).

A continuous statewide REAP raster dataset, as well as an updated (RE-GAP) landcover dataset, were provided by Cathy Maynard and Linda Vance of the NRCS.
Wind power and Wind speed data was compiled by the Montana Department of Environmental Quality (DEQ) and downloaded from the NRIS GIS Data List website (http://nris.mt.gov/gis/gisdatalib/gisDataList.aspx).

STATSGO soils data (an NRCS product), and the digital elevation data (produced by the USGS) were also downloaded from the NRIS GIS Data List website.

Correlated Landcover Types:

<table>
<thead>
<tr>
<th>Value</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>Open Water</td>
</tr>
<tr>
<td>21</td>
<td>Developed, Open Space</td>
</tr>
<tr>
<td>22</td>
<td>Developed, Low Intensity</td>
</tr>
<tr>
<td>81</td>
<td>Pasture/Hay</td>
</tr>
<tr>
<td>82</td>
<td>Cultivated Crops</td>
</tr>
<tr>
<td>4232</td>
<td>Northern Rocky Mountain Dry-Mesic Montane Mixed Conifer Forest</td>
</tr>
<tr>
<td>4234</td>
<td>Northern Rocky Mountain Mesic Montane Mixed Conifer Forest</td>
</tr>
<tr>
<td>5312</td>
<td>Northern Rocky Mountain Montane-Foothill Deciduous Shrubland</td>
</tr>
<tr>
<td>5455</td>
<td>Inter-Mountain Basins Montane Sagebrush Steppe</td>
</tr>
<tr>
<td>7112</td>
<td>Northern Rocky Mountain Lower Montane, Foothill, and Valley Grassland</td>
</tr>
<tr>
<td>8403</td>
<td>Introduced Upland Vegetation - Annual and Biennial Forbland</td>
</tr>
<tr>
<td>9103</td>
<td>Inter-Mountain Basins Greasewood Flat</td>
</tr>
<tr>
<td>9155</td>
<td>Northern Rocky Mountain Lower Montane Riparian Woodland and Shrubland</td>
</tr>
<tr>
<td>9159</td>
<td>Northwestern Great Plains Floodplain</td>
</tr>
<tr>
<td>9256</td>
<td>Western Great Plains Saline Depression Wetland</td>
</tr>
</tbody>
</table>
Figures 1A & 1B: The area circled in red is Canyon Ferry Lake, the model is predicting moderate habitat suitability directly over deep open water. The white square represents a presence point used in the model and located over land near the water’s edge. This point however, is not one of the points that caused the exaggerated habitat suitability over land. The graphic on the right represents the same area as it is in the final model – generated without the problematic points.

Figures 2A & 2B – This set of figures compares the exaggerated habitat potential over water due to the inclusion of a surface-to-water raster (on left) and the same area as it is in the final model which does not include this layer. The layer clearly increased predicted habitat suitability over the open-water of Flathead Lake.
Figure 3 - Table showing variable contributions analyses. Figure 3 shows humidity and temperature data contributing little to the success of the model.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Percent contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>slope</td>
<td>35.4</td>
</tr>
<tr>
<td>windpower</td>
<td>24.3</td>
</tr>
<tr>
<td>landcover_spc</td>
<td>13.8</td>
</tr>
<tr>
<td>elevation</td>
<td>10.4</td>
</tr>
<tr>
<td>min_raster</td>
<td>10</td>
</tr>
<tr>
<td>reap_90m</td>
<td>5.5</td>
</tr>
<tr>
<td>windspeed</td>
<td>0.5</td>
</tr>
<tr>
<td>humidity</td>
<td>0</td>
</tr>
<tr>
<td>max_raster</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 4A - shows a model run using the SSURGO soils data as available at the time, black areas indicate portions of the database for which there was no data at the time (data is now available). Note also how little potential habitat is predicted due to overfitting and the specificity of the SSURGO soils data.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Percent contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>surgo_raster</td>
<td>84.3</td>
</tr>
<tr>
<td>landcover_spc</td>
<td>12</td>
</tr>
<tr>
<td>elevation</td>
<td>1.6</td>
</tr>
<tr>
<td>windspeed</td>
<td>1.2</td>
</tr>
<tr>
<td>windpower</td>
<td>0.4</td>
</tr>
<tr>
<td>reap_90m</td>
<td>0.2</td>
</tr>
<tr>
<td>slope</td>
<td>0.2</td>
</tr>
<tr>
<td>min_raster</td>
<td>0.1</td>
</tr>
<tr>
<td>max_raster</td>
<td>0</td>
</tr>
<tr>
<td>humidity</td>
<td>0</td>
</tr>
</tbody>
</table>

**Figure 4B** – shows extreme reliance on SSURGO raster data, leaving the model too reliant on specific soil polygons and a single dataset.

**Figure 5A** – Test and Training curves for AUC of an unusually high training fit when SSURGO soils data included in model.
Figure 5B – Omission on training curve is inconsistent with a reliable statistic, and lower than random predicted omission.
Figure 6: Response curves for all variables when each variable is used independently in producing the model.
**Figure 7:** Response curves for all variables when each variable is used in conjunction with the other variables in producing the model.
Figures 8A – 8C: Jackknife analysis of a) training gain; b) AUC; and c) test gain.
Figures 9A & 9B: Figure 8A shows an omission curve for both training and test samples compared to a random predicted omission. Figure 8B shows the AUC curve for training and testing, with a random prediction of AUC = 0.5 included for comparison.
Figure 10: Map of C. lasiurus habitat suitability throughout Montana, generated using MAXENT software. White squares indicate sample locations used for training, purple squares indicate sample locations used for testing.