Indiscriminate Cuckoo Bumble Bee (*Bombus insularis*)
Predicted Suitable Habitat Modeling

Distribution Status: Resident Year Round
State Rank: SNR
Global Rank: G4G5

Modeling Overview

Created by: Braden Burkholder
Creation Date: September 22, 2016
Evaluator: Braden Burkholder
Evaluation Date: October 14, 2016

Inductive Model Goal: To predict the distribution and relative suitability of general year-round habitat at large spatial scales across the entire state of Montana.
Inductive Model Performance: This is a preliminary model and is not intended to be the final version. The model was only examined briefly for obvious deficiencies. Data and model output need to be reviewed in greater detail prior to broad use of this model.

Suggested Citation: Montana Natural Heritage Program. 2016. Indiscriminate Cuckoo Bumble Bee (*Bombus insularis*) predicted suitable habitat models created on September 22, 2016. Montana Natural Heritage Program, Helena, MT. 12 pp.

Montana Field Guide Species Account: http://fieldguide.mt.gov/speciesDetail.aspx?elcode=IIHYM24340

Inductive Modeling

Model Limitations and Suggested Uses

This model is based on statewide biotic and abiotic layers originally mapped at a variety of spatial scales and standardized to 90×90 meter raster pixels. Furthermore, the spatial accuracy of the training and testing data are varied (typically 20-400 meters) and may result in additional statistical noise in the model. As a result, model outputs may not be appropriate for use on smaller areas or at fine spatial scales. Model outputs should not typically be used for planning efforts on land areas smaller than one quarter of a public land survey system (PLSS) section (<64 hectares) and model outputs for some species may only be appropriate for broader regional level planning efforts. Model outputs should not be used in place of on-the-ground surveys for species, and wildlife and land management agency biologists should be consulted about the value of using model output to guide habitat management decisions for regional planning efforts or local projects. See Suggested Contacts for State and Federal Natural Resource Agencies attached to this document.

Inductive Model Methods

Modeling Process

Presence-only data were obtained from Montana Natural Heritage Program Databases, which serve as a clearinghouse for animal and plant observation data in Montana. These data were then filtered to ensure spatial and temporal accuracy and to reduce spatial auto-correlation (summarized in Table 1). The spatial extent of this model was limited to the known geographic range of the species, by season when applicable, in order to accurately assess potentially available habitat.

We then used these data and 19 statewide biotic and abiotic layers (Table 2) to construct the model using a maximum entropy algorithm employed in the modeling program Maxent (Phillips et al. 2006, Ecological Modeling 190:231-259). Entropy maximization modeling functions by first calculating constraints and then applying the constraints to estimate a predicted distribution. The mean, variance, etc. of the environmental variables at the training data locations are used to estimate the constraint distributions. Maxent requires that the final predicted distribution fulfills these constraints. Maxent avoids overfitting of models to the training data by “regularizing” or relaxing the constraints so that modeled distributions only have to be close to, rather than exactly equal to, the constraint distributions (Elith et al. 2011, Diversity and Distributions 17:43-57).

Maxent fits a model by first assuming the predicted distribution is perfectly uniform in geographic space and moves away from this distribution only to the extent that it is forced to by the constraints. Constrained by training data, Maxent successively modifies the coefficients for each environmental variable via random walk, accepting the modified coefficient if it increases the gain. Gain is a measure of the closeness of the model concentration around the presence samples that is similar to goodness of fit in generalized linear models. The random walk of coefficients continues until either the increase in the gain falls below a set threshold or a set maximum number of iterations are performed. The gain value at the end of a model run indicates the likelihood of suitability of the presence samples relative to the likelihood for random background points. The overall gain associated with individual environmental variables can be used as a measure of the relative importance of each variable (Merow et al. 2013, Ecography 36:1058-1069).

We employed a k-folds cross validation methodology, in this case using ten folds for model training and validation (Elith et al. 2011). Each fold consists of 90% of the data designated for training and 10% of the data reserved for testing. Each record is used for training nine times and testing once. Ten models are estimated and averaged to produce the final model presented here.

Model Outputs and Evaluation

The initial model output is a spatial dataset of continuous logistic values that ranges from 0-1 with lower values representing areas predicted to be less suitable habitat and higher values representing areas predicted to be more suitable habitat (Figure 2). The standard deviation in the model output across the averaged models is also calculated (Figure 3). If enough observations were available to train and evaluate the models, the continuous output is reclassified into suitability classes - unsuitable, low suitability, moderate suitability, and high suitability. Thresholds for defining suitability classes are presented and described below (Table 4).

We evaluated the output of the Maxent model with two metrics, an absolute validation index (AVI) (Hirzel et al. 2006, Ecological Modelling 199:142-152) and deviance (Phillips and Dudik 2008, Ecography 31: 161-175). These metrics are described below in the results (Table 5). Area under the curve (AUC) values are also displayed for reference, but are not used for evaluation (Lobo et al. 2008, Global Ecology and Biogeography 17:145-151). Additionally, standard deviation in logistic output of the ten individual models is plotted as a map to examine spatial variance of model output. Finally, a deviance value was calculated for each test data observation as a measure of how well model output matched the location of test observations. In theory, everywhere a test observation was located, the logistic value should have been 1.0. The deviance value for each test observation is calculated as -2 times the natural log of the associated logistic output value.

Table 1: Model Data Selection Criteria and Summary

Location Data Source	Montana Natural Heritage Program Databases
Total Number of Records	191
Location Data Selection Rule 1	Records with <= 1600 meters of locational uncertainty
Number of Locations Meeting Selection Rule 1	110
Location Data Selection Rule 2	No overlap in locations within 800 meters in order to avoid spatial autocorrelation
Observation Records used in Model (Locations Meeting Selection Rules 1 & 2)	87
Season Modeled	Entire state, Year-round
Number of Model Background Locations	60,000

Table 2: Environmental Layer Information

Layer	Identifier	Original Scale	Description
Land Cover	catesys	30m	Categorical. Landcover classes (25) from the 2016 Montana Spatial Data Infrastructure Land Cover Framework; Level 2 classes used with a few minor changes including removal of linear and point features: Alpine Grassland and Shrubland, Alpine Sparse and Barren, Conifer-dominated Forest and Woodland (mesic-wet), Conifer-dominated Forest and Woodland (xeric-mesic), Deciduous dominated forest and woodland, Mixed deciduous/coniferous forest and woodland, Lowland/Prairie Grassland, Montane Grassland, Agriculture, Introduced Vegetation/Pasture/Hay, Developed, Mining and Resource Extraction, Wetland or Marsh, Floodplain and Riparian, Open Water, Wet meadow, Harvested Forest, Insect-Killed Forest, Introduced Vegetation, Recently burned, Deciduous Shrubland, Sagebrush Steppe or Desert Scrub, Sagebrush or Saltbush Shrubland, Bluff/Badland/Dune, Cliff/Canyon/Talus http://geoinfo.msl.mt.gov/msdi/land_use_land_cover
Geology	catgeol	vector	Categorical. Basic rock classes (5) as defined by USGS (plus water for large water bodies): Sedimentary, Unconsolidated, Metamorphic, Plutonic, and Volcanic. https://mrdata.usgs.gov/geology/state/state.php?state=MT
Soil Order	catsoilord	Vector	Categorical. Major soil orders (7) as defined by USDA based on STATSGO2 general statewide soil maps, along with non-soil (Rock, Water) classifications: Entisols, Inceptisols, Aridisols, Mollisols, Alfisols, Andisols, and Vertisols. http://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx
Soil Regime	catsoiltemp	vector	Categorical. Soil Moisture and Temperature regimes (11) classification pairs as defined by USDA (plus water): Cryic/Udic, Cryic/Udic Ustic, Cryic/Typic Ustic, Cryic/Aridic Ustic, Cryic/Typic Xeric, Frigid/Aquic, Frigid/Udic, Frigid/Typic Ustic, Frigid/Aridic Ustic, Frigid/Typic Xeric, Mesic/Ustic Aridic. http://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx
Elevation	contelev	≈10m	Continuous. Elevation in meters above mean sea level. https://www.sciencebase.gov/catalog/item/4f70aa9fe4b058caae3f8de5
Aspect (East-West)	contewasp	≈10m	Continuous. Aspect of slopes, ranging from 1 (east) to -1 (west). https://www.sciencebase.gov/catalog/item/4f70aa9fe4b058caae3f8de5
Aspect (North-South)	contnsasp	≈10m	Continuous. Aspect of slopes, ranging from 1 (north) to -1 (south). https://www.sciencebase.gov/catalog/item/4f70aa9fe4b058caae3f8de5
Slope	contslope	≈10m	Continuous. Percent slope (x100) of landscape. https://www.sciencebase.gov/catalog/item/4f70aa9fe4b058caae3f8de5
Ruggedness	contvrm	≈10m	Continuous. Vector ruggedness measure (0 to 1). https://www.sciencebase.gov/catalog/item/4f70aa9fe4b058caae3f8de5
Summer Solar Radiation	contsumrad	≈10m	Continuous. Solar radiation (WH/m^2) for the day of the summer solstice. https://www.sciencebase.gov/catalog/item/4f70aa9fe4b058caae3f8de5
Winter Solar Radiation	contwinrad	≈10m	Continuous. Solar radiation (WH/m^2) for the day of the winter solstice. https://www.sciencebase.gov/catalog/item/4f70aa9fe4b058caae3f8de5
Annual NDVI	contndvi	900m	Continuous. Normalized Difference Vegetation as a measure of yearly mean greenness from the MODIS Terra satellite. ftp://mco.cfc.umt.edu/ndvi/terra/yearly_normals/
Annual Precipitation	contprecip	≈800m	Continuous. Average annual precipitation (mm) for 1981-2010. http://www.prism.oregonstate.edu/normals/
Percent Winter Precipitation	contwinpcp	≈800m	Continuous. Average percent (0 to 1) of the total annual precipitation that occurs during winter (Nov-Apr) for 1981-2010. http://www.prism.oregonstate.edu/normals/
Max Summer Temp	conttmax	800m	Continuous. Average maximum temperature (°C) in July for 1981-2010. ftp://mco.cfc.umt.edu/tmax/monthly_normals/
Min Winter Temp	conttmin	800m	Continuous. Average minimum temperature (°C) in January for 1981-2010. ftp://mco.cfc.umt.edu/tmin/monthly_normals/
Degree Days	contddays	800m	Continuous. Average annual total of degree days (°F) above 32°F for 1981-2010. http://services.cfc.umt.edu/arcgis/rest/services/Atlas/Temperature_CropDegreeDays32F/ImageServer
Distance to Stream	contstrmed	vector	Continuous. Distance to major streams in meters, based on major streams identified in TIGER files or USGS topographic maps (Stream_Lake_1993 dataset). http://ftp.geoinfo.msl.mt.gov/Data/Spatial/NonMSDI/Shapefiles/
Distance to Forest Cover	contfrsted	30m	Continuous. Distance to any forest land cover type in meters. http://geoinfo.msl.mt.gov/msdi/land_use_land_cover

Inductive Model Results

Table 3: Environmental Layer Contributions to Model Fit

Layer ID	Percent Contribution[a]	Layer ID	Percent Contribution[a]
catsoiltemp	32.1%	contwinpcp	1.0%
catesys	25.0%	contfrsted	0.9%
contstrmed	15.2%	contelev	0.5%
catgeol	7.5%	contsriwin	0.5%
contslope	4.0%	conttmin	0.3%
catsoilord	4.0%	contsrisum	0.3%
contnsasp	3.8%	contprecip	0.2%
contndvi	1.9%	contvrm	0.1%
contewasp	1.4%	conttmax	0.0%
contddays	1.2%		

[a] Relative contributions of the layers to the model based on changes in fit (gain) during iterations of the training algorithm.

Table 4: Habitat Suitability Thresholds

Measure	Value
Low Logistic Threshold[a]	0.073
Moderate Logistic Threshold[b]	0.324
Optimal Logistic Threshold[c]	0.746
Area of entire modeled range (percent of Montana)	380,529.0 km^2 (100.0%)
Total area of predicted suitable habitat within modeled range	236,810.8 km^2
Area of predicted low suitability habitat within modeled range	158,762.4 km^2
Area of moderate suitability habitat within modeled range	73,070.1 km^2
Area of predicted optimal habitat within modeled range	4,978.3 km^2

[a] The logistic threshold between unsuitable and low suitability as determined by Maxent which balances training data omission error with predicted area.
[b] The logistic threshold value where the percentage of test observations above the threshold is what would be expected if the observations were randomly distributed across logistic value classes (Hirzel et al. 2006). This is equivalent to a null model. When sample sizes are small, it may be undetermined.
[c] The logistic threshold where the percentage of test observations above the threshold is 10 times higher than would be expected if the observations were randomly distributed across logistic value classes. When sample sizes are small, it may be undetermined.

Table 5: Evaluation Metrics

Metric	Value
Low AVI[a]	94.2%
Moderate AVI[a]	69.0%
Optimal AVI[a]	18.4%
Average Testing Deviance ($\bar{x} \pm sd$)[b]	1.971 ± 1.756
Training AUC[c]	0.890
Test AUC[d]	0.823

[a] Absolute Validation Index: The proportion of test locations that fall above the low, moderate, or optimal logistic threshold.
[b] A measure of how well model output matched the location of test observations. In theory, everywhere a test location was located, the logistic value should have been 1.0. The deviance value for each test location is calculated as -2 times the natural log of the associated logistic output value. For example, the equivalent deviance values for the low, moderate and optimal logistic thresholds of this model would be 5.221, 2.257 and 0.585, respectively. Deviances for individual test locations are plotted in Figure 6.
[c] The area under a curve obtained by plotting the true positive rate against 1 minus the false positive rate for model training observations (averaged over 10 folds). Values range from 0 to 1 with a random or null model performing at a value of 0.5.
[d] The same metric described in c, but calculated for test observations.

Figure 1. Jackknife assessment of contribution by individual environmental layers to training gain.

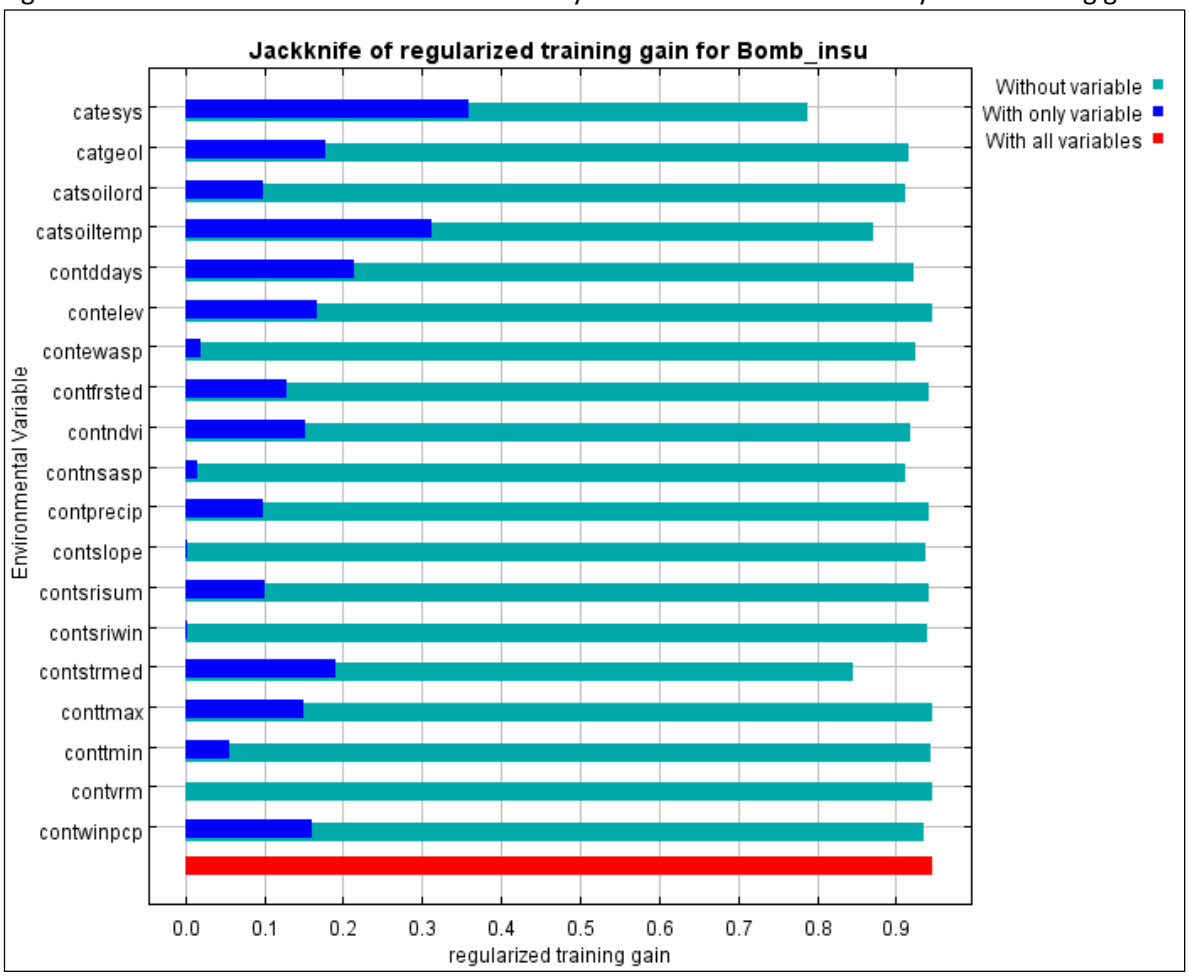
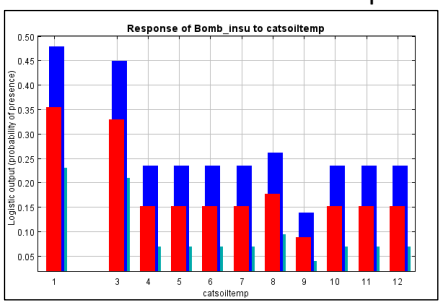
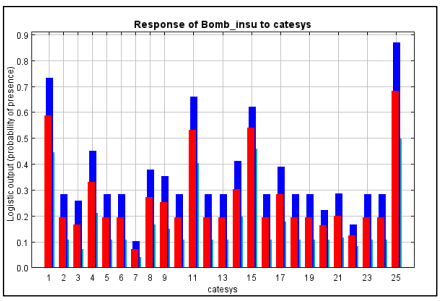
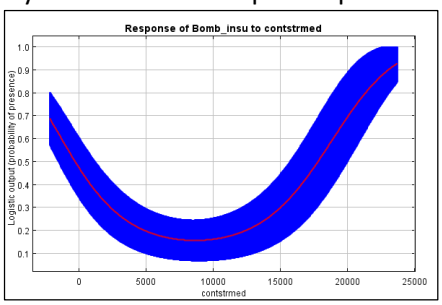


Figure 2. Response curves for the top three contributing environmental layers, mean value in red, +/- one standard deviation in blue. Response curves for additional environmental layers are available upon request.







Inductive Model Map Outputs

Figure 3. Continuous habitat suitability model output (logistic scale).

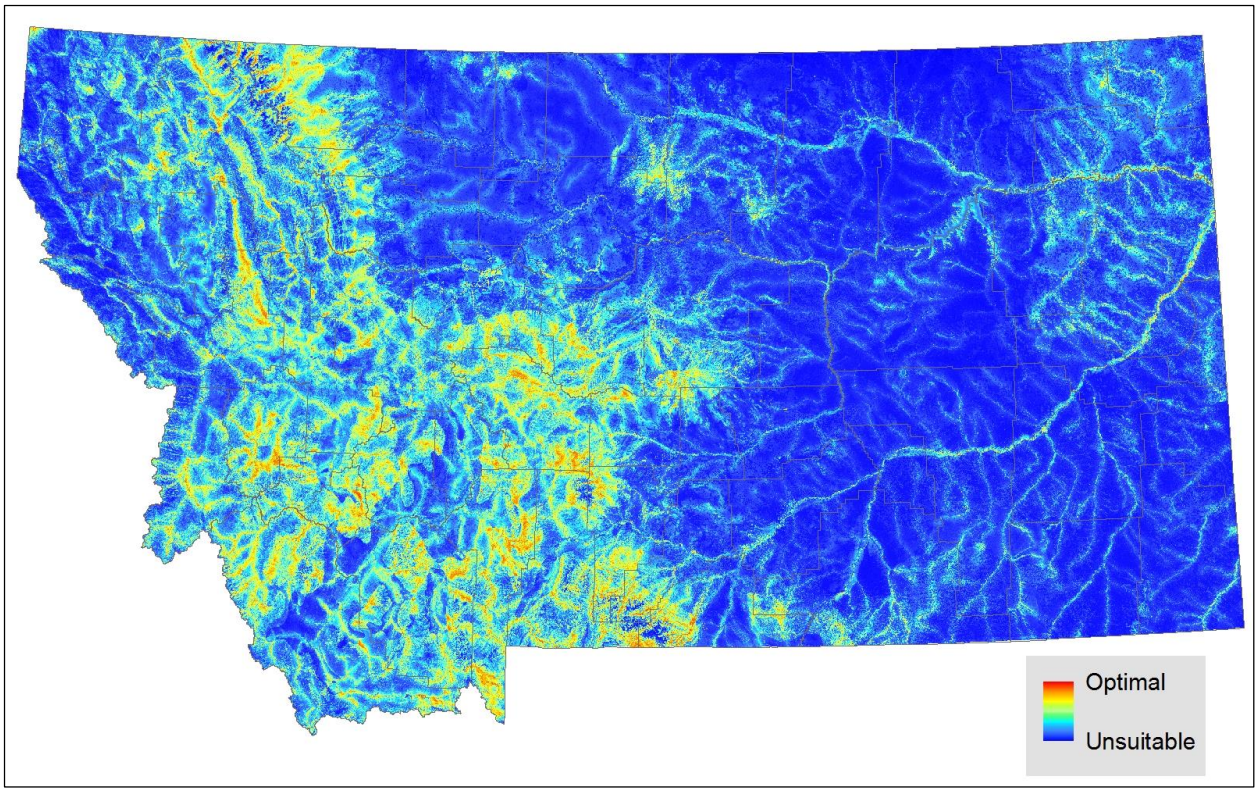


Figure 4. Standard deviation in the model output across the averaged models.

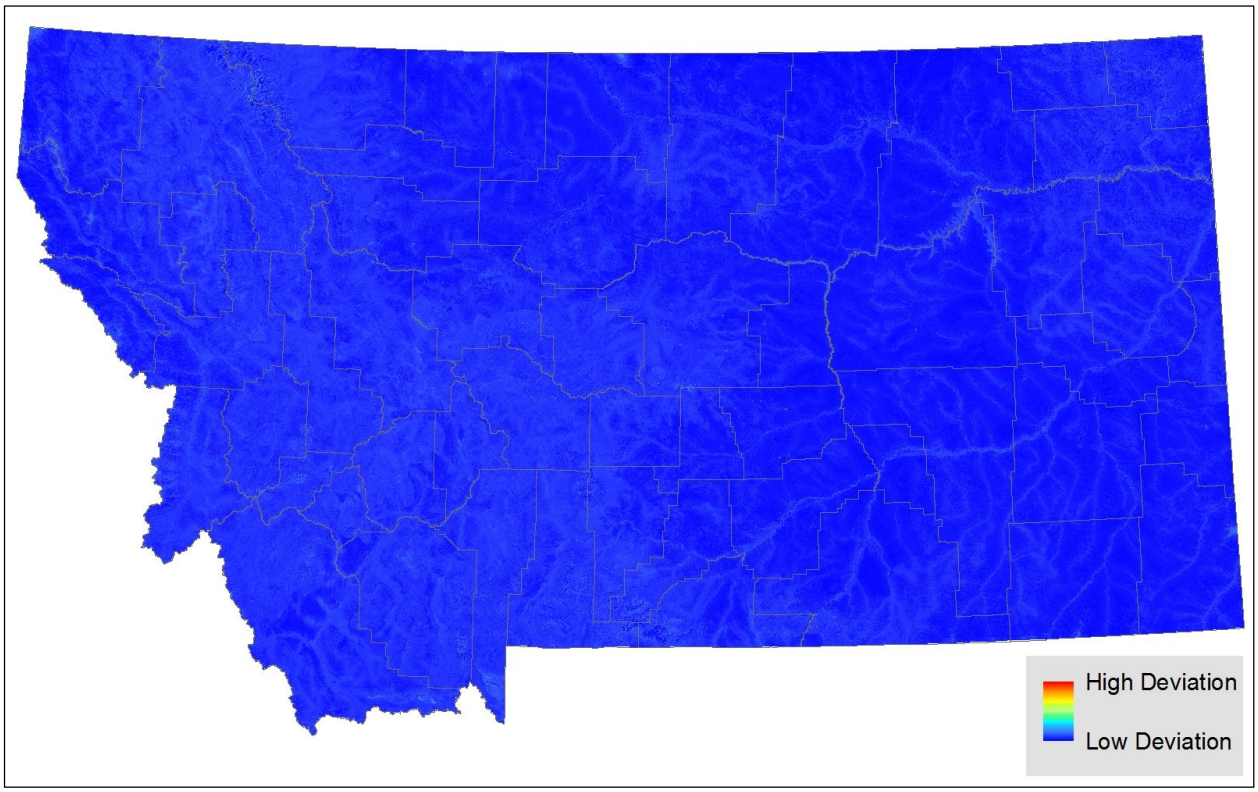


Figure 5. Continuous habitat suitability model output with the 87 observations used for modeling.

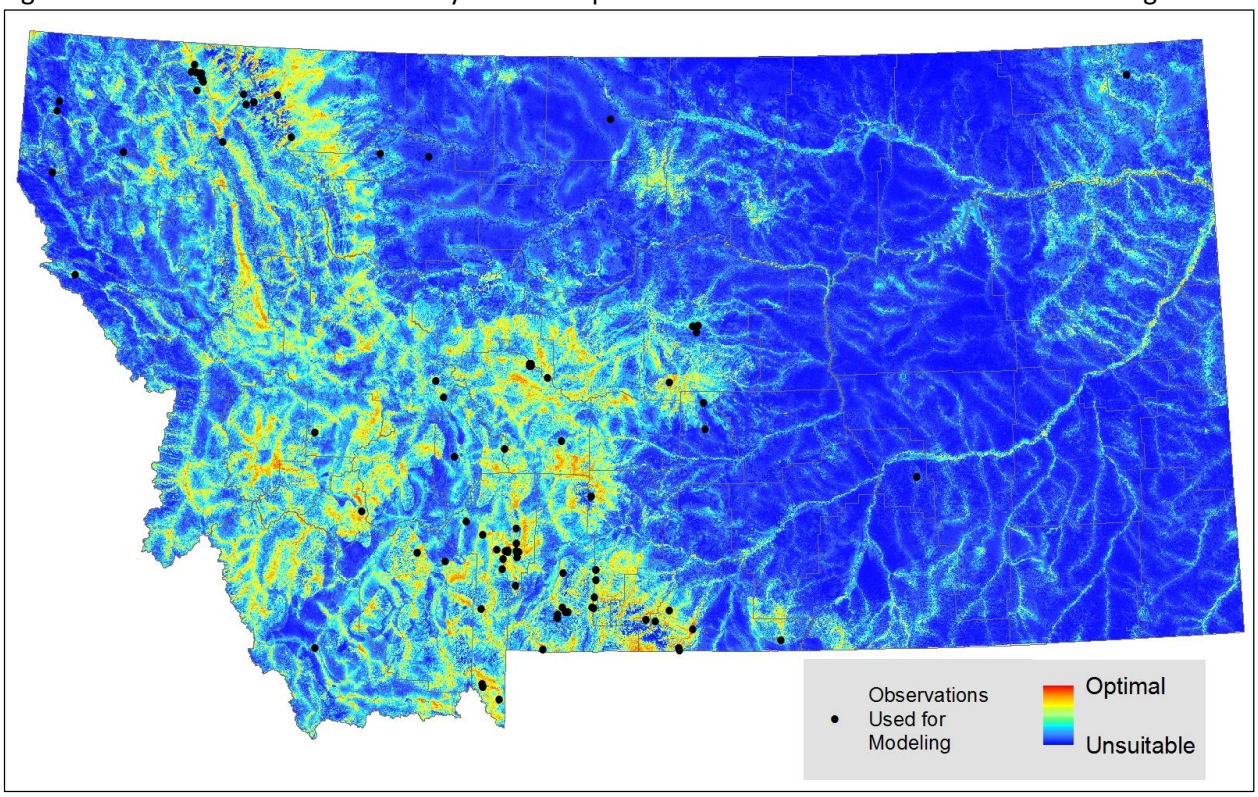


Figure 6. Continuous habitat suitability model output with relative deviance for each observation. Symbol size corresponds to the difference between 1.0 and the optimal, moderate, and low habitat suitability threshold.

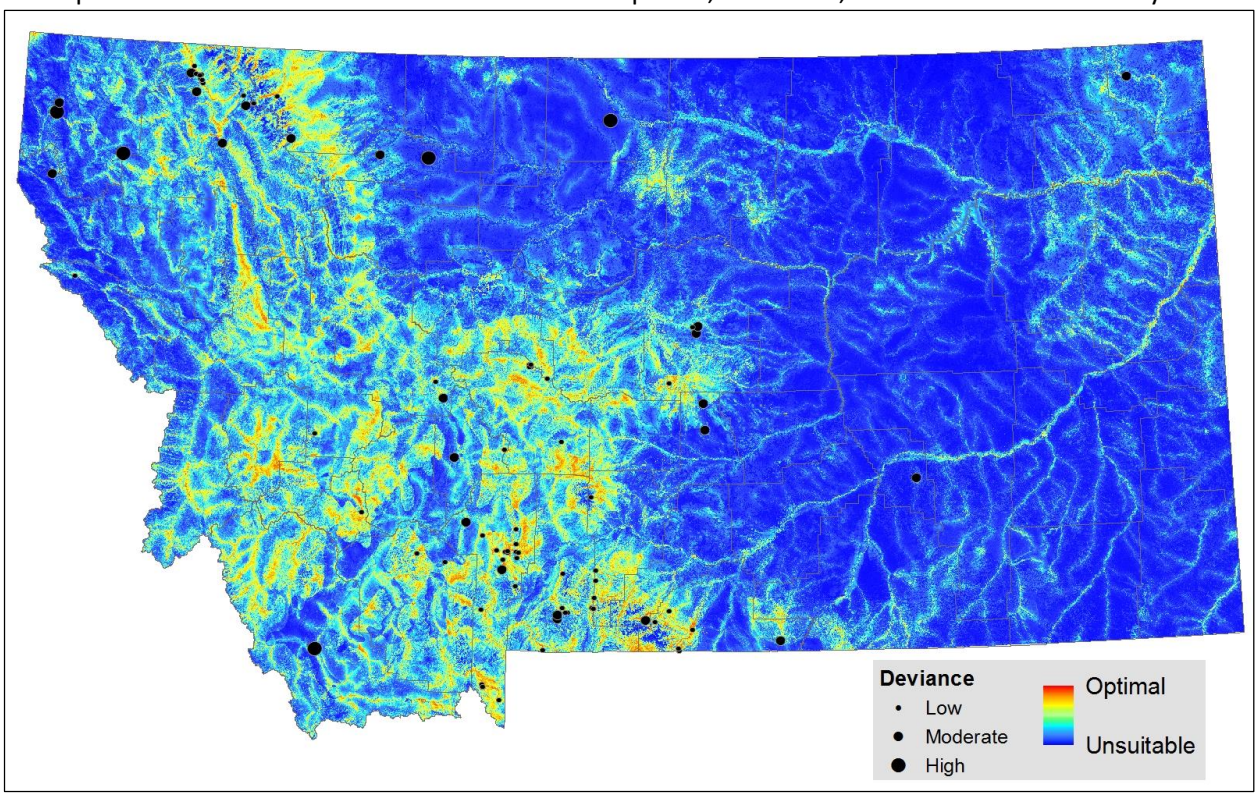


Figure 7. Continuous habitat suitability model output with all 191 observations (black) and survey locations that could have detected the species (gray).

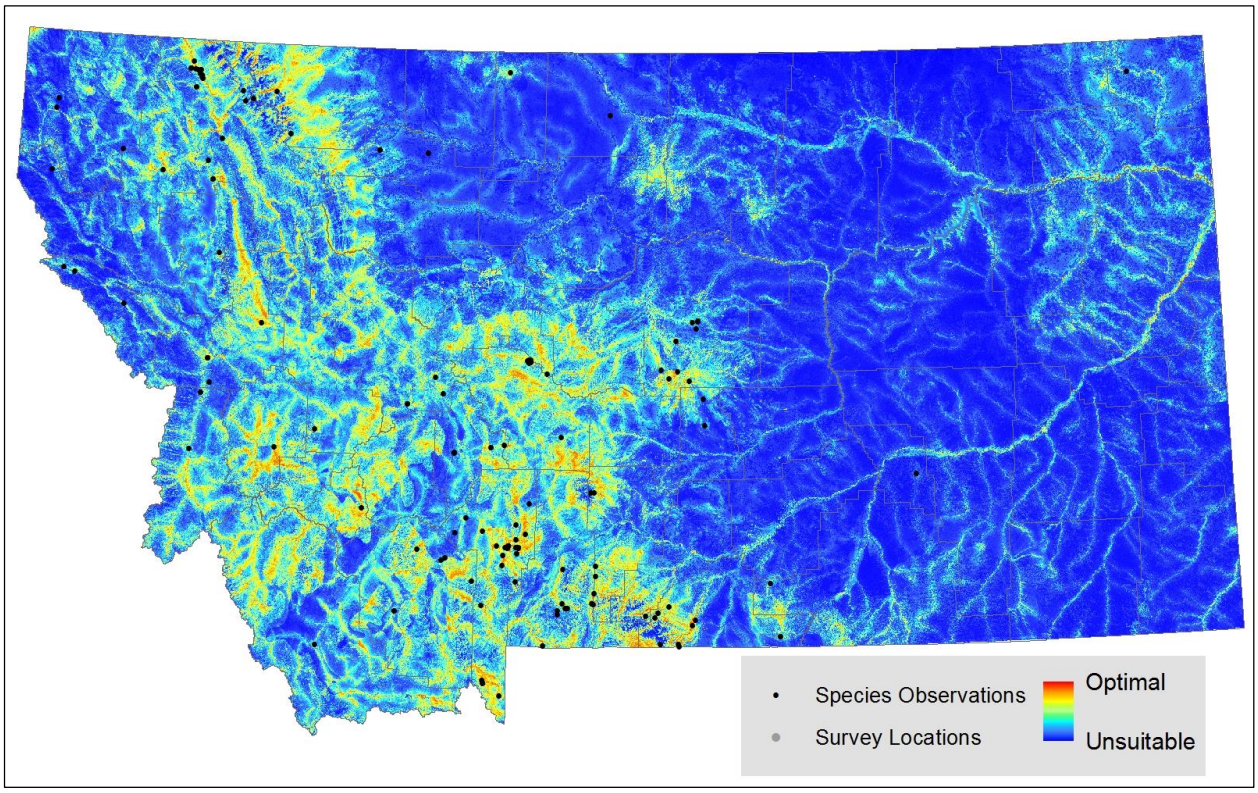


Figure 8. Model output classified into habitat suitability classes.

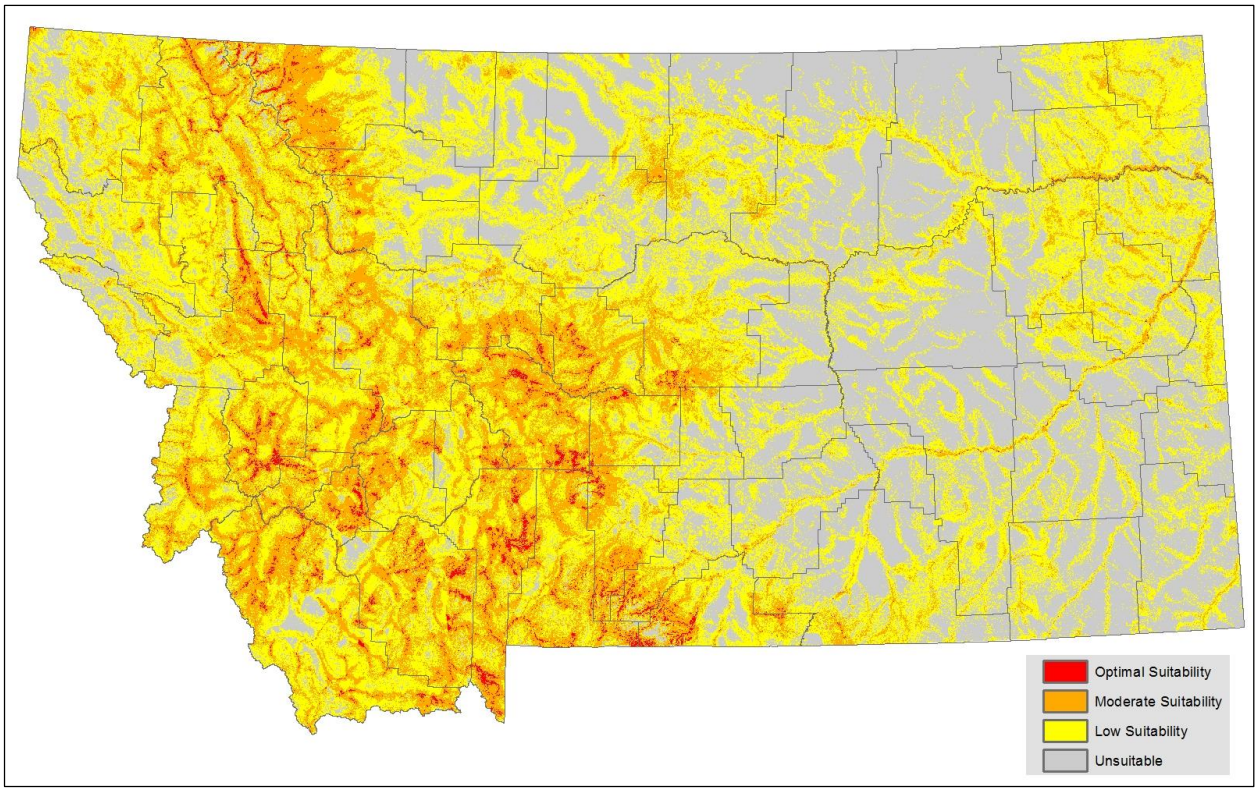
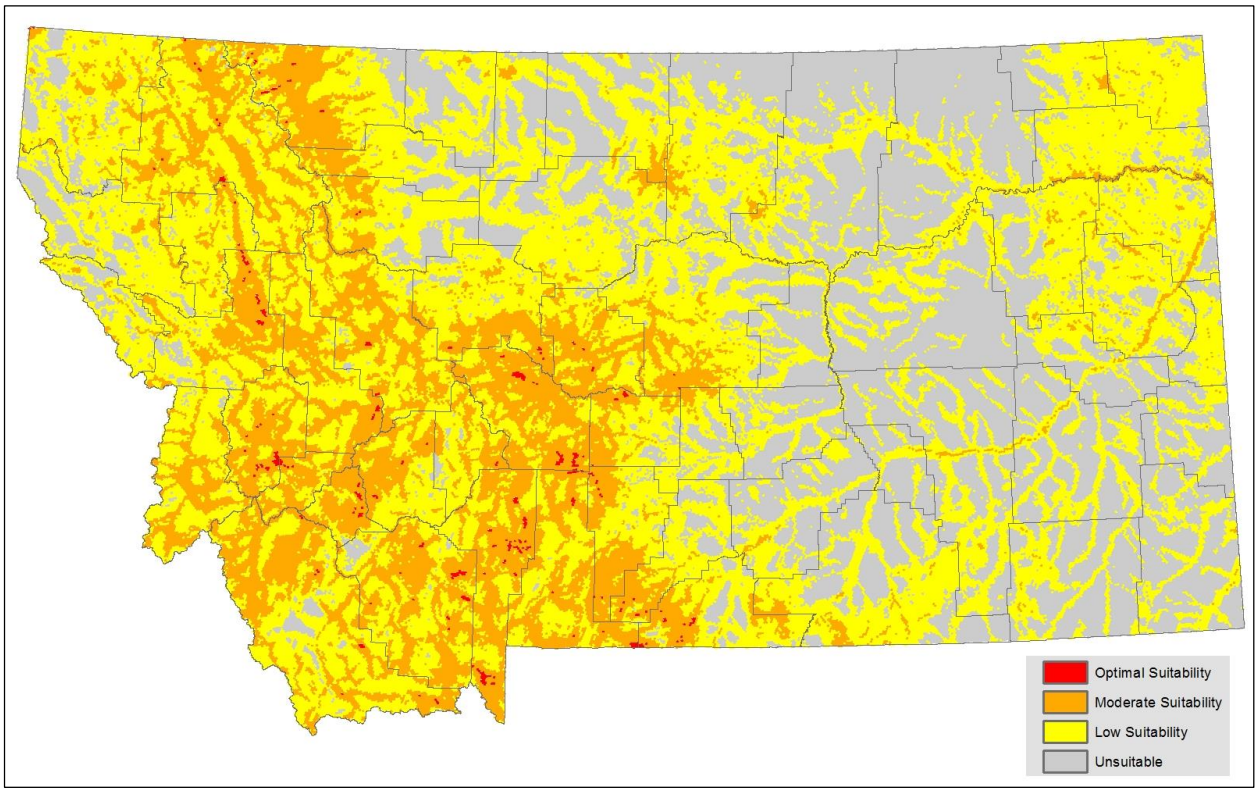


Figure 9. Model output classified into habitat suitability classes and aggregated into hexagons at a scale of 259 hectares per hexagon. This is the finest scale suggested for management decisions and survey planning.

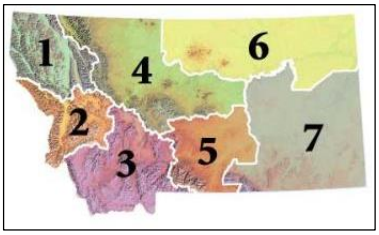


Suggested Contacts for State and Federal Natural Resource Agencies

As required by Montana statute (MCA 90-15), the Montana Natural Heritage Program works with state, federal, tribal, nongovernmental organizations, and private partners to ensure that the latest animal and plant distribution and status information is incorporated into our databases so that it can be used to inform a variety of planning processes and management decisions. In addition to the information you receive from us, we encourage you to contact state and federal resource management agencies in the area where your project is located. They may have additional data or management guidelines relevant to your efforts. In particular, we encourage you to contact the Montana Department of Fish, Wildlife, and Parks for the latest data and management information regarding hunted and high profile management species and to use the U.S. Fish and Wildlife Service's Information Planning and Conservation (IPAC) website http://ecos.fws.gov/ipac/ regarding U.S. Endangered Species Act listed Threatened, Endangered, or Candidate species.

For your convenience, we have compiled a list of relevant agency contacts and links below:

Montana Fish, Wildlife, and Parks

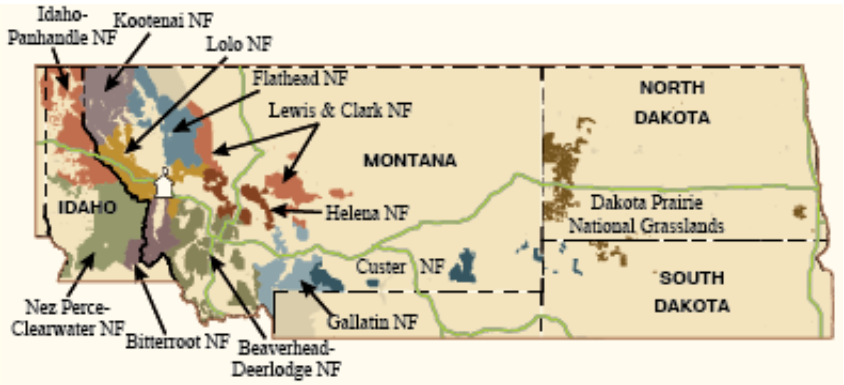
Fish Species	Zachary Shattuck zshattuck@mt.gov (406) 444-1231 or Lee Nelson leenelson@mt.gov (406) 444-2447
American Bison Black-footed Ferret Black-tailed Prairie Dog Bald Eagle Golden Eagle Common Loon Least Tern Piping Plover Whooping Crane	Lauri Hanauska-Brown LHanauska-Brown@mt.gov (406) 444-5209
Grizzly Bear Greater Sage Grouse Trumpeter Swan Big Game Upland Game Birds Furbearers	John Vore jvore@mt.gov (406) 444-5209
Managed Terrestrial Game and Nongame Animal Data	Adam Messer – MFWP Data Analyst (406) 444-0095, amesser@mt.gov
Fisheries Data	Bill Daigle – MFWP Fish Data Manager (406) 444-3737, bdaigle@mt.gov
Wildlife and Fisheries Scientific Collector's Permits	http://fwp.mt.gov/doingBusiness/licenses/scientificWildlife/default.html Merissa Hayes for Wildlife (406) 444-7321 merhayes@mt.gov Beth Giddings for Fisheries (406) 444-7319 begiddings@mt.gov
Fish and Wildlife Recommendations for Subdivision Development Regional Contacts  	Renee Lemon RLemon@mt.gov (406) 444-3738) and see: http://fwp.mt.gov/fishAndWildlife/livingWithWildlife/buildingWithWildlife/subdivisionRecommendations/ Region 1 (Kalispell) (406) 752-5501 Region 2 (Missoula) (406) 542-5500 Region 3 (Bozeman) (406) 994-4042 Region 4 (Great Falls) (406) 454-5840 Region 5 (Billings) (406) 247-2940 Region 6 (Glasgow) (406) 228-3700 Region 7 (Miles City) (406) 234-0900

Bureau of Land Management

Montana Field Office Contacts:

Billings:	(406) 896-5013
Butte:	(406) 533-7600
Dillon:	(406) 683-8000
Glasgow:	(406) 228-3750
Havre:	(406) 262-2820
Lewistown:	(406) 538-1900
Malta:	(406) 654-5100
Miles City:	(406) 233-2800
Missoula:	(406) 329-3914

United States Forest Service



Regional Office – Missoula, Montana Contacts

Wildlife Program Leader:	Tammy Fletcher	tammyfletcher@fs.fed.us	(406) 329-3588
Wildlife Ecologist:	Cara Staab	cstaab@fs.fed.us	(406) 329-3677
Fish Program Leader:	Scott Spaulding	scottspaulding@fs.fed.us	(406) 329-3287
Fish Ecologist:	Cameron Thomas	cathomas@fs.fed.us	(406) 329-3087
TES Program:	Kristi Swisher	kswisher@fs.fed.us	(406) 329-3558
Interagency Grizzly Bear Coordinator:	Scott Jackson	sjackson03@fs.fed.us	(406) 329-3664
Regional Botanist:	Steve Shelly	sshelly@fs.fed.us	(406) 329-3041

U.S. Fish and Wildlife Service:

Information Planning and Conservation (IPAC) website: http://ecos.fws.gov/ipac/

Montana Ecological Services Field Office: http://www.fws.gov/montanafieldoffice/ (406) 449-5225