ALASKA GAP ANALYSIS PROJECT

External Review Process

Background

The Alaska Gap Analysis Project is predicting habitat for 430 vertebrate species and subspecies that reside, breed, or use habitat in the state of Alaska for a substantial portion of the their life history. Gap Analysis uses the predicted distributions of animal species habitat to evaluate their conservation status relative to existing land management (Scott et al. 1993). However, the maps of species distributions may also be used to answer a wide variety of management, planning and research questions relating to individual species or groups of species. In addition to the maps, great utility may be found in the literature and occurrence data that is assembled to produce the species distribution models.

The premise of our endeavor is that we are modeling to identify areas of the landscape that contain physical and biotic features that likely will or do support occurrence of specific animal taxa. That modeling is based on a set of associations (wildlife habitat relationships or environmental and climatic) developed for each taxa relative to a set of landscape features that are compiled at the statewide scale. Namely, we are modeling potentiality for occurrence of suitable habitat features or environmental niches for each animal taxa; we are not preparing predictions of absolute occurrence of any individual taxa on any given day.

The list of species to model was determined by identifying decision rules for taxon inclusion. In preparation for modeling, we compiled over 1.6 million occurrence records from 650 unique data sources, developed watershed-scale range maps for each target species, and populated a habitat-associations database that cross-walks species habitat descriptions from the literature and expert input to ecological systems. A species' distribution, at 60 meter resolution, was created using a model to predict areas suitable for occupation within its range. We used a combination of deductive and inductive modeling techniques to produce our final models.

Deductive models were derived using a suite of spatial variables including habitat types, elevation, hydrological characteristics, and distance to/from forest edge. Inductive models were derived using known points of occurrence and their intersection with a suite of environmental parameters. Final distribution maps are intersections of these two independently derived models, delimited by range limits of the target species, and evaluated for classification success. To create the most accurate models possible we are engaging taxa experts to provide a review of the watershed derived range maps and species distribution models.

Objectives

An important factor for model implementation is understanding the objectives of the modeling effort and the assumptions of the models. The objective of the species distribution models are to:

1. Provide maps that predict the distribution of terrestrial vertebrate species throughout their range in Alaska to support analysis of conservation status; and
2. Develop a database of geographic ranges, wildlife habitat relationships, and predicted distribution of each vertebrate species for the long-term utility of GAP and its cooperators (Csuti and Crist 2000).

Along with these objectives are several assumptions associated with GAP vertebrate habitat models (Csuti and Crist 1998):

1. Species are assumed to occur within a distribution model representing potential habitat but are not predicted to occur at any particular point within that model.
2. Species are assumed to be present within their predicted distribution model, but no assumptions are made about the abundance of the species within their distribution.
3. Species are assumed to be present during some portion of their life history, but not necessarily during the entire year. This is especially the case for breeding birds, who are only present in Alaska during the pre-breeding, breeding, and post-breeding season (spring through early fall). Therefore, for many avian taxa, we only produced distribution models for the breeding season.

We encountered many challenges while creating both range and distribution maps. Thus, we are soliciting external review from knowledgeable individuals on the modeled terrestrial vertebrates across the state. The purpose of the model review is both to inform the process with which models are developed and potentially revised, and to provide user’s confidence that species models are accurate and useable within the scale and context they are intended. This document describes the expert review process within AKGAP.

Review Approach

The AKGAP habitat models have three model components that we would like reviewed. These are:

1. Range extent (Range)
2. Wildlife habitat relationships (Report)
3. Distribution models (3 model types = 1 deductive, 1 inductive, 1 combined)

Range Extent

Review of individual species range maps is to ensure that the range extent accurately depicts the known range of the species. The review should include an evaluation of: 1) extent, and 2) seasonal coding (this section is most relevant to migratory taxa, e.g., permanent resident, summer breeder).

Some considerations include:

1. Does the range extent, as depicted by hydrologic units, reflect the known range of species?
2. Are the hydrologic units correctly coded?

Wildlife Habitat Relationships

Review of this section will either substantiate or refute the habitat relationships used to produce the deductive models. This process should include: 1) review of each relationship used in the model, 2) concurrence with the relationships; and 3) review of references to ensure that important citations are not missed.

Questions to focus on this part of the review include:

1. Are the habitat relationships (within the limits of available information) correctly identified?
2. Are there additional relationships not identified, which should be included? Knowledge regarding the limitations of the habitat relationships are also requested.

Distribution Models

Review of predicted distribution maps is a subjective review based on expert knowledge. The review of this tier should focus on the following questions:

1. Does the depiction look plausible?
2. Does the depiction identify too much habitat?
3. Does the depiction not identify enough habitat?
4. Does the predicted distribution appear to be spatially correct?

Rules for model modification

Modifications must be identified based on a reference with associated source code or documented as personal communication (Table 1).

Table 1. Categories of reference information.

Category	Description
1	Information is based on substantive direct investigation and published (printed or electronic) in an outlet subject to peer evaluation.
2	Information is based on direct investigation or general review and is available in any of a variety of general publications that are serial or ad hoc documents of a technical nature subject to uncertain degree of professional review.
3	Information is derived directly or indirectly from individuals with demonstrated limited or broad expert credence; formats include but are not limited to word-of-mouth accounts, field journals, specimen record tags or forms, labeled photographs, etc.
4*	Information is inferred from associations applicable to similar taxa.

* Though not a preferred reference, because of taxonomy changes this option may be applicable.

Literature Cited

Csuti, B. and P. Crist. 1998. Methods for Assessing Accuracy of Animal Distribution Maps, Gap Analysis Program, University of Idaho, Moscow, Idaho. http://www.gap.uidaho.edu/ Date Accessed: 02 July 2003.

Material for Expert Reviewers

Expert Review packet: All documents and materials for the AKGAP expert review are accessible via the internet at: http://sealab.uas.alaska.edu/page.php?npID=13.

Reviewer Items

- **External Review and Process** – this document**.**
- **External Review Form** - questionnaire we are asking each reviewer to fill out. One questionnaire per species.
    - For reviewers that want to submit reviews as hard copies, please print and fill in the External Review Form and mail all hard copy documents to:
      Tracey Gotthardt
      Attn: AKGAP Expert Review
      707 A Street, Suite 103
      Anchorage, AK 99501

      Or email to: tagotthardt@uaa.alaska.edu
- **Landfire Overview** – Document describing the LANDFIRE Ecological Systems, including a map, map legend and searchable table.
- **Ancillary Data documentation** – overview of development of several ancillary data layers used for modeling.
- **Table of Accuracy Statistics for Models**

Documents to be reviewed: Pdfs or jpegs of these documents can be obtained from the website http://sealab.uas.alaska.edu/page.php?npID=13. Search for your species under the headings **Mammals, Birds**, or **Amphibians**. Species are arranged in phylogenetic order. Documents to be reviewed include:

- Seasonal Range by HUC (jpg)
- Predicted Distribution Model – deductive, inductive, combination (jpg)
- Wildlife Habitat Relationships report (pdf)

Additional information:

- **Ancillary Data** – maps depicting each of the ancillary data layers used to develop distribution models.

Modeling Methods

Range Mapping

We define range as the total areal extent occupied by a given taxon. Range maps are usually characterized by large all-encompassing polygons with very little interdigitation of occupied and unoccupied space (Aycrigg and Beauvais 2007).

The first map product developed for each target taxon was a range map, using 8-digit hydrologic units (Hydrologic units) as map units, following methods employed by other recent regional GAP range mapping efforts (Boykin et al. 2007). 8-digit HUCs (HUC8s) were used to identify species ranges. The intent of the HUC8 was to constrain the habitat model. In some cases the HUC-depicted range can extend well beyond the known range of the species.

We acquired initial polygon range maps for individual taxa from NatureServe (http://www.natureserve.org/getData/animalData.jsp) and from the Alaska Natural Heritage Program. We then tessellated each polygon range map into its constituent HUC8s (Figure 1).

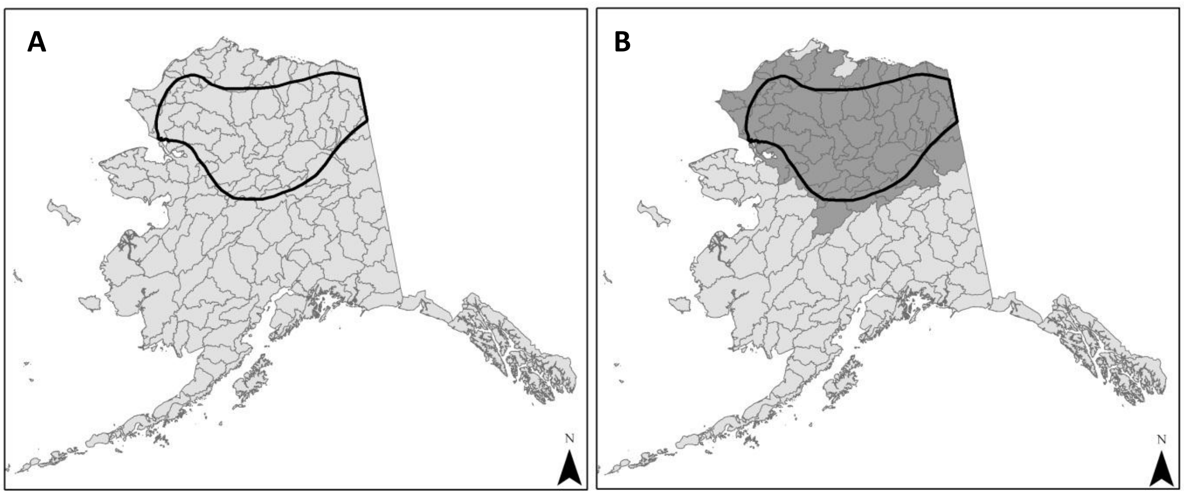


Figure 1. Area outlined in black indicates the original (polygon) range map for the Alaska Marmot overlaid on 8-digit HUCs (A). Dark gray area indicates all HUC8s that intersected or were included within the original polygon range map (B), and is considered the final HUC8 range map.

We then assigned initial values for two attributes to each HUC8:

> **Season:** Possible values were Summer, Spring/Fall, Winter, Yearround. Especially for migratory taxa, the value of the Season attribute was assigned with the specific modeling season (equate Breeding with Summer), and modeling season date, in mind.
>
> Seasons were defined as follows: Winter (December - February); Fall/Spring (March - May and August - November); Summer (June or July); Yearround (all months).

Occurrence: Possible values were Known, Suspected, Historical, or Accidental. "Known" equated to the presence of documented occurrences of the target taxon, or confident expert prediction of occurrence, within a given HUC8. Less confident conclusions were grounds for selecting the "Suspected" modifier. "Historical" indicated the last known record of occurrence for a given HUC8 predated 1910. "Accidental" was only selected when infrequent or irregular records were available for a given HUC8.

Distribution Models

A species' distribution, at 60 meter resolution, was created using a model to predict areas suitable for occupation within its range. We used a combination of **deductive** and **inductive** modeling techniques to produce our final models.

Deductive Models

Deductive distribution modeling followed the traditional, land cover-based procedures of previous Gap Analyses. Deductive models were derived using a suite of spatial variables including habitat types, elevation, hydrological characteristics, and distance to/from forest edge. This process can be described as designating land cover types from a given classification system as either suitable or unsuitable for occupation by a given taxa (Beauvais et al. 2012).

We used the LANDFIRE Existing Vegetation Type (EVT) map as our statewide land cover map. The EVT layer represents the species composition currently present at a given site. Vegetation map units are primarily derived from NatureServe's Ecological Systems classification, which is a nationally consistent set of mid-scale ecological units. Additional units are derived from NLCD, National Vegetation Classification Standard (NVCS) Alliances, and LANDFIRE specific types.

We developed a database of wildlife habitat relationships to help delineate habitats that were considered suitable for occupation by a given taxa. Habitat descriptions were extracted from the NatureServe Explorer database, the Alaska Natural Heritage Programs (AKNHP) Biotics database, and through exhaustive literature review. The descriptive habitat associations from the literature were then cross-walked to Ecological Systems and other associated ancillary variables by AKGAP species modeling team, with substantial assistance from vegetation ecologists at AKNHP (Figures 2 and 3).

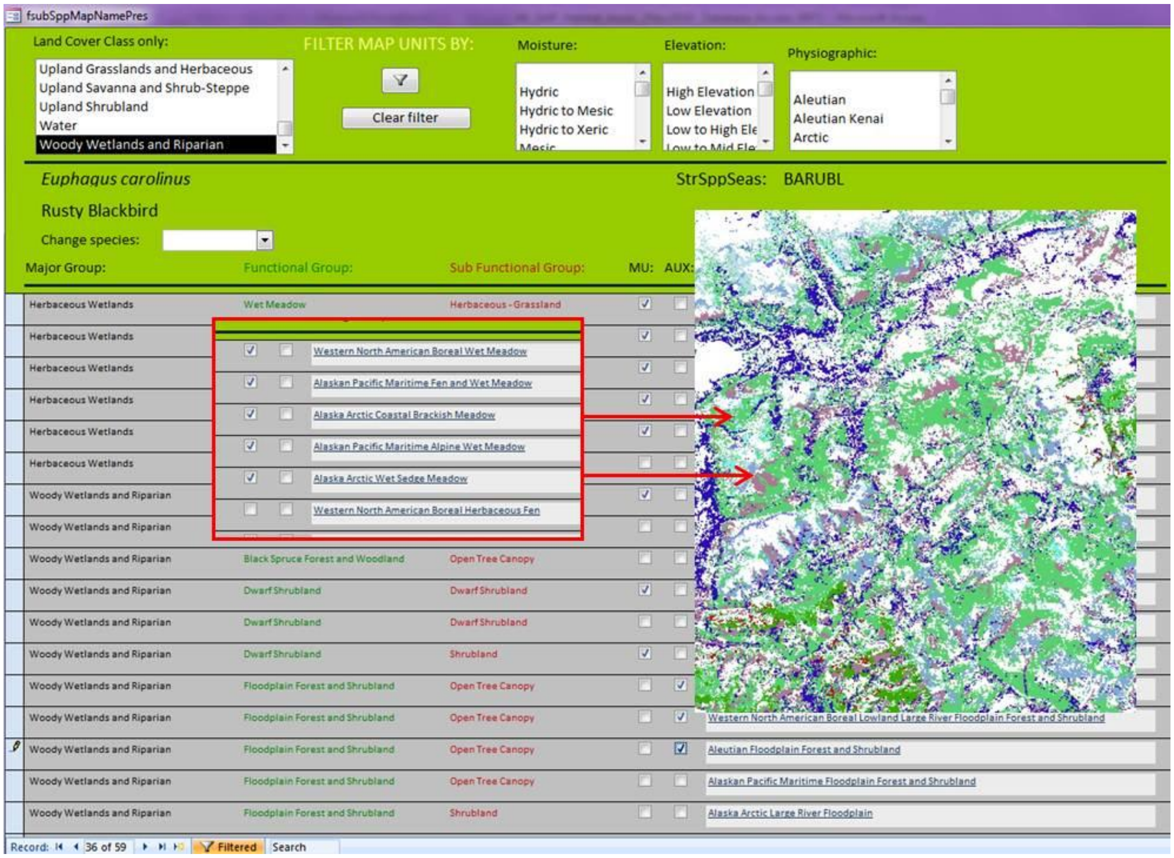


Figure 2. Example of the habitat associations crosswalk to ecological systems. There were 142 ecological systems defined for Alaska. Map inset is of habitat types selected for.

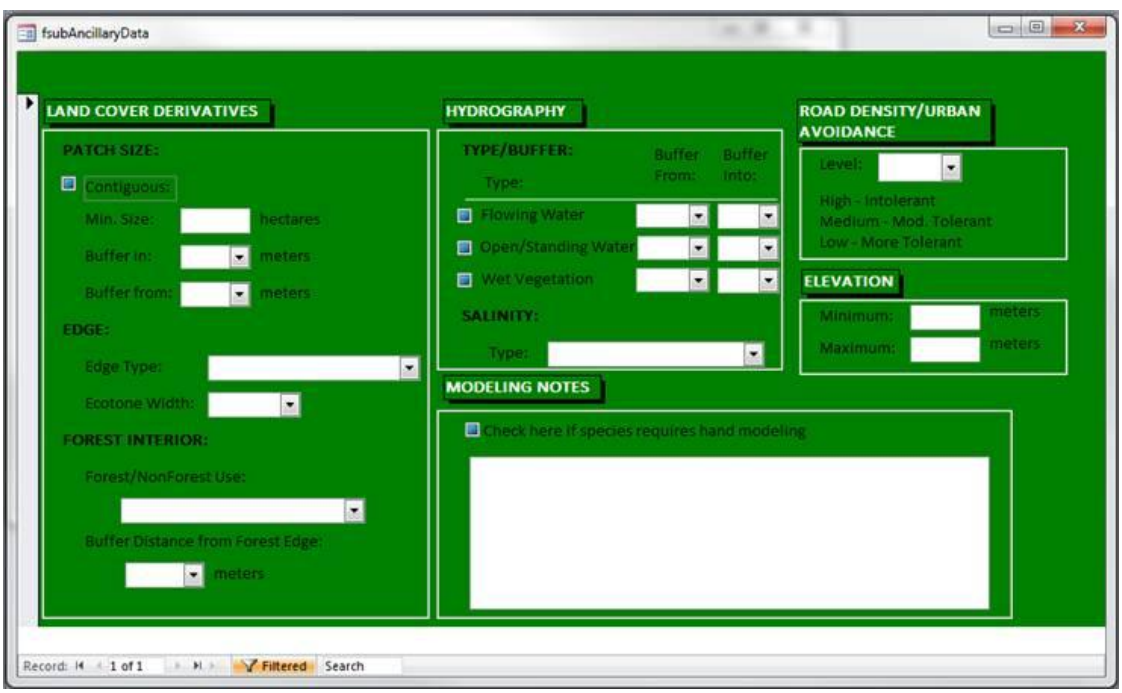


Figure 3. Additional ancillary data selections for further model refinement.

Models were developed to incorporate habitat utilization across the taxon's entire range in Alaska. We found that many wide ranging taxa utilized habitat differently across their range or elevational limitations were different due to latitudinal differences over the study area. In an attempt to capture regional variation in habitat utilization, Ecological Systems were filtered and then selected by physiographic region (including: Aleutian, Artic, Boreal, Sub-boreal, North American Pacific Maritime and Temperate Pacific) and elevation (high, medium, low).

Deductive models were then derived using a combination of the Ecological Systems that best described land cover types suitable for occupation by each target taxon, plus any additional categorical variables (e.g. distance to edge, elevation) selected by the species modeling team. The HUC8 range maps were then used to delineate the final modeled extent.

Inductive Models

Inductive models were derived using known points of occurrence and their intersection with a suite of environmental parameters. Inductive modeling included data processing and filtering, ancillary data layer development or refinement, and the application of the Maximum Entropy algorithm (MaxEnt version 3.3.1) to produce models.

Occurrence Data Collection and Processing

Occurrence data were acquired from over 650 unique data sources, resulting in a dataset of approximately 1.6 million records for 430 species. Records were summarized in a common format and attributed with 30 common fields. Positional accuracy (if not provided) was estimated based on the record's mapping protocol using standards established by the Natural Heritage Network (http://www.natureserve.org/prodServices/standardsMethods.jsp). All records were stored in a geodatabase that was queried as needed for analysis and modeling.

For migratory species, all occurrences outside the designated modeling season were removed from the dataset. For avian species, the primary season of interest was the breeding season, in which case, all non-breeding season occurrences were eliminated. Breeding season was broadly defined as such: for breeding waterfowl, May through August, for all other breeding birds, June, July and August. We then eliminated duplicate records. Next, we eliminated remaining records with mapping precisions >2000 m. Finally, we eliminated any remaining records of observation made before 1990. We selected 1990 as an arbitrary cutoff for two reasons: 1) 87% of the occurrence data were collected between 1990 and 2010, and 2) we felt that over the past 20 years, environmental conditions have remained reasonably stable across the study area.

Preliminary models were run using all occurrence data that met the above criteria. These preliminary datasets were then reviewed to identify species with highly autocorrelated data, which can sometime bias environmental niche models (Jimenez-Valverde and Lobo 2006, Johnson and Gillingham 2008). We thinned dense clusters of occurrences resulting from oversampling by applying a stratified sampling method using 12-digit HUCs to spatially separate occurrences. At least two, and up to ten occurrences were randomly selected from each HUC to be included in the modeling procedure. The number of occurrences used depended on the number of overall occurrence data points available for, and the results of further iterations of modeling.

After initial review, preliminary models for species that had poor model results were re-run using alternative data selection procedures. The first alternative data selection method removed or reduced the year restriction and included data from years prior to 1990, as long as they met the other filtering restrictions. This method was only used if the prior models for the species did not meet internal review criteria. The initial filtering restrictions resulted in several species that simply lacked adequate occurrence data to run a model. For these species, we reduced both the accuracy and date restrictions, in an attempt to produce a large enough sample to run a model, cognizant of the fact that by reducing accuracy restrictions we were potentially reducing the accuracy of the modeled output . Taxa with <10 final modeling records were excluded from the inductive modeling process. The distributions of such taxa were modeled entirely through the deductive process.

Environmental Data Collection and Processing

We selected 20 environmental predictor variables to use in all of our inductive distribution modeling. Environmental predictor variables were comprised of climatic data, elevation, geology, soils, and distance to specific landscape features (e.g. distance to coast). Environmental predictor layers were projected in the Alaska Albers Equal Conical projection and resampled to 60 m cell size, such that there projection, extent, cell size and alignment were consistent. These processes were performed in ArcGis 10.0.

Refer to the report titled "Ancillary Data Documentation" under the **Reviewer Items** tab at: http://sealab.uas.alaska.edu/page.php?npID=17 for more detailed description in individual predictive layers and their development, their range of parameters, and usage in modeling species predicted distributions.

Model Generation and Validation

For each target taxon, we used the Maxent algorithm, version 3.3.1 (http://www.cs.princeton.edu/~schapire/maxent/), to produce our inductive species distribution models. All models were produced using the same 20 environmental variables. Thirty percent of the occurrence data were held back to test the model. We used area under the curve (AUC) statistics derived from receiver operating characteristics analyses, which is automatically calculated by Maxent, to estimate performance. Models with an AUC of .75 and higher were considered acceptable, while models with AUCs lower than .75 were rejected.

Model Display

Model outputs include an ASCII file which was converted to a continuous raster grid for import into ArcGIS. Each cell in the raster contains a probability value that represents the probability of occurrence for that particular species, ranging from 0.0 to 1.0. For these models, a binary threshold was applied that divided the continuous output into two categories: predicted absence (0) and predicted presence (1). We overlaid the occurrence data used to produce the model with the modeled output to determine the raster value for each cell. We then calculated the mean raster cell value (and sd), and applied this as our threshold. The final modeled output was then clipped to the species known and suspected range within the state – thus, limiting predictions to areas of the state that are believed to be part of the species range.

Final Distribution Models

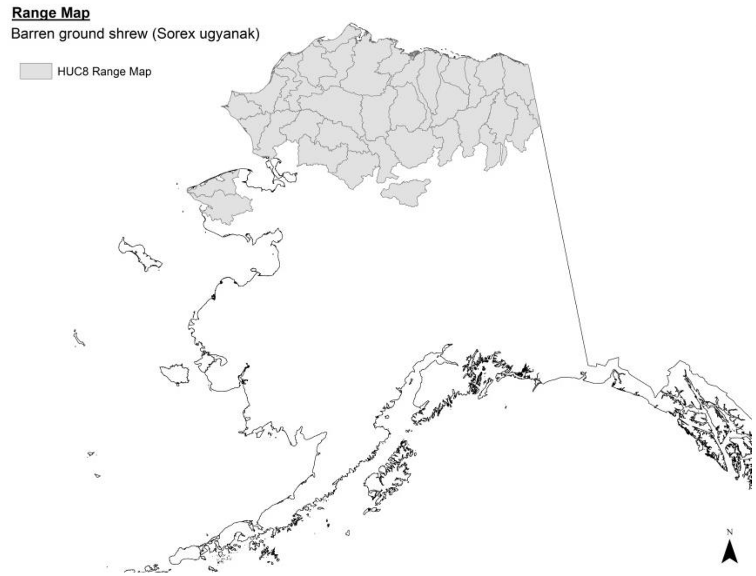
Final distribution models are intersections of the independently derived deductive and inductive models, delimited by range limits of the target species, and evaluated for classification success (Figure 4).

For those resident taxa for which only a deductive distribution model was generated (because they had <10 post-filtering occurrence records), the final deductive distribution model was designated as the final project distribution model. The mapped expression of that model within the boundaries of each taxon's final range was used as the final distribution map.

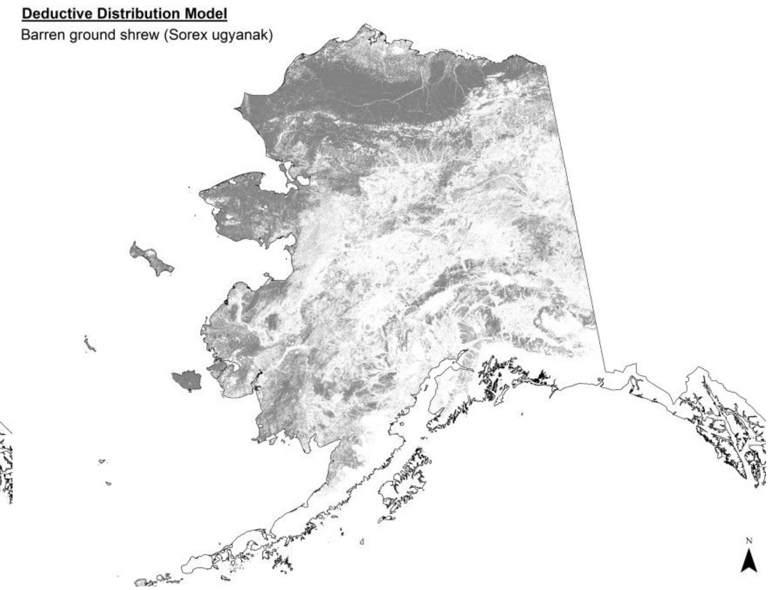
For those taxa for which we produced both a deductive and in inductive model, we intersected the maps of both models and clipped the result to the taxon's final range boundaries. We then visually inspected the clipped result to assess whether it predicted presence throughout most of the taxon's range (in which case the intersection map was accepted as the taxon's final distribution map) or left large portions of the taxon's range with no predicted presence (in which case the intersection map was rejected, and the mapped expression of the taxon's final deductive model was chosen as its final distribution map).

Figure 4. Examples of intersecting an inductive distribution model with a deductive distribution model, within range boundaries, to form a final model of predicted distribution for a given taxon. Map series is for Barren ground shrew (*Sorex uganak*): **1** is the HUC8 range map; **2** is the deductive (i.e., landcover-based) distribution model; **3** is the inductive distribution model; and **4** is the intersection of the deductive and inductive models within the range boundaries.

1


Range Map
Barren ground shrew (Sorex ugyanak)
HUC8 Range Map
N

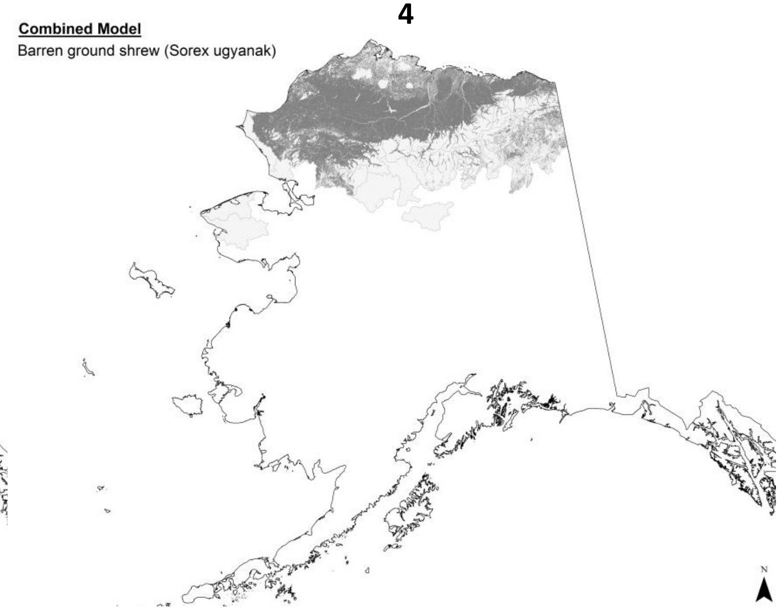
2


Deductive Distribution Model
Barren ground shrew (Sorex ugyanak)
N

3


Inductive Distribution Model
Barren ground shrew (Sorex ugyanak)
Predicted Present
N

4


Combined Model
Barren ground shrew (Sorex ugyanak)
N

Literature Cited

Aycrigg, J. and G.P. Beauvais. 2007. Novel approaches to mapping vertebrate occurrence for the Northwest Gap Analysis Project. Gap Analysis Bulletin 15:27-33.

Boykin, K. G., B. C. Thompson, R. A. Deitner, D. Schrupp, D. Bradford, L. O'Brien, C. Drost, S. Propeck-Gray, W. Rieth, K. Thomas, W. Kepner, J. Lowry, C. Cross, B. Jones, T. Hamer, C. Mettenbrink, K.J. Oakes, J. Prior-Magee, K. Schulz, J. J. Wynne, C. King, J. Puttere, S. Schrader, and Z. Schwenke. 2007. Predicted animal habitat distributions and species richness. Chapter 3 in J.S. Prior-Magee t al. (eds). Southwest Regional Gap Analysis Final Report. U.S. Geological Survey, Gap Analysis Program. Moscow, Idaho.

Beauvais, G., M. Anderson and D. Keinath. 2012. Modeling range, distribution, and habitat importance for terrestrial vertebrates in the Northwest ReGAP region. Wyoming Natural Diversity Database, University of Wyoming, Laramie, Wyoming.

Jimenez-Valverde, A. and J. M. Lobo. 2006. The ghost of unbalanced species distribution data in geographical model predictions. Diversity and Distributions 12:521-524.

Johnson, C. J. and M. P. Gillingham. 2008. Sensitivity of species-distribution models to error, bias, and model design: an application to resource selection functions for woodland caribou. Ecological Modeling 213:143-155.