# A Species Habitat Model Standard for the NatureServe Network

Version 1.0

December 6, 2021

NatureServe

NatureServe Network

This Standard was produced by the Species Habitat Model Standard Work Group; all members are listed in Appendix I.

Recommended Citation:
NatureServe. 2021. A species habitat model standard for the NatureServe Network. NatureServe: Arlington, VA.

Version Notes:

This document is a first release of a NatureServe Network Habitat Model Standard. We fully anticipate the publication of revised editions as procedures and processes for data management and exchange of habitat model products are tested and refined and as the science of habitat modeling evolves.

In the near-term, we anticipate:

- Updates to criteria for scoring of model confidence based on statistical performance, including for deductive models.
- Updated link to the Network Best Practices Wiki (still in development).
- Updates to data management framework, and potentially data format requirements, as processes for data exchange are built and tested and data management sub-team is further engaged.
- More clarity on how to communicate habitat model information when the same species has been modeled by multiple programs.

We highly recommend that programs interested in pursuing data exchange of model outputs contact NatureServe for additional guidance on data formatting and the exchange process prior to final preparation of models for exchange.

# Table of Contents

# EXECUTIVE SUMMARY

NatureServe and the Network of Natural Heritage Programs and Conservation Data Centers are the recognized leaders for providing decision-quality distribution and abundance data for at-risk species and ecosystems. Our focus has been on Network-collected occurrence data that identify key locations of demonstrable conservation value for those species and ecosystems. Now, with increasing pressures on biodiversity from land development, transportation, and energy extraction interests, there is a pressing need for decision-makers to extrapolate beyond documented locations, and to identify areas likely to be occupied by at-risk species. Although modeled information is not as certain as documented locations, it overcomes limitations on current survey information and can be applied in a precautionary manner to evaluate potential development impacts. Many Network programs have developed predictive models to identify likely habitat for the at-risk species within their jurisdictions, and several regional modeling efforts have been completed as well. As more models are developed, it has become clear that we need a straightforward and flexible standard for developing, compiling, and sharing species habitat models across the Network. Developing and adhering to such a standard will help position the Network as a leader in the provision of high-quality habitat models that are trusted by decision-makers and expand our Biodiversity Location Data offerings.

To create the standard, we convened a Species Habitat Model Standard Work Group (hereafter Work Group), comprised of individuals from multiple Network programs and NatureServe. It focused on three key objectives, led by Work Group sub-teams:

1) Develop a standard for Network species habitat models so that data on "likely habitat" can be provided as standardized Network-wide products. The standard covers:
    a. Data formats, including the structure of spatial data, key attribute fields and metadata requirements
    b. Data quality standards
2) Document best practices for developing species habitat models through an updatable wiki that provides additional guidance on the development of habitat models in a format that is flexible and adaptive as the science of modeling evolves.
3) Outline a data management framework to guide the development of IT systems supporting Network provision of habitat models.

This document focuses on (1) the Habitat Model Standard and (3) the Data Management Framework. Best practices documentation will be maintained as a separate online updateable Wiki. The Work Group was guided by the model development and assessment rubric outlined in Sofaer et al. (2019), the collective experience gained through various state and multi-jurisdictional modeling initiatives, and the previous efforts of Network habitat modeling resolutions and technical teams. Our goal was to create a flexible standard to support efficient sharing of information about likely habitat, while providing the flexibility for programs to adapt and improve modeling methods over time.

This standard places NatureServe and the Network in a better position to place the most current, complete, and consistent Biodiversity Location Data into the hands of researchers and decision-makers. It will leverage Network data to create products that enable conservation practitioners and the regulation community to better understand not only where species have been documented but where else they are likely to occur and where they are *unlikely* to occur. The standard, and particularly the Best Practices it references, will continue to evolve to meet the changing needs of the Network as we address pressing conservation challenges.

# 1. INTRODUCTION

## 1.1 The Need for a Species Habitat Modeling Standard

A core function of NatureServe's species programs is to facilitate the sharing of information on where at-risk species occur. Traditionally, this has been done through the provision of documented location data, most typically in the form of Element Occurrences, but also through observations. However, even with limitless resources, full documentation of all locations of a species is impossible, as we will never be able to survey everywhere. Predictive habitat models provide a means to address this limitation by providing information on habitat suitability, and thus likely species presence or absence, in areas that have not been comprehensively or recently surveyed. Conservation planning efforts frequently rely on models to evaluate the predicted importance of areas of interest to biodiversity, particularly at-risk species. They also help managers understand where at-risk species are *not* likely to occur. Models can help inform where additional field surveys should be conducted. When used to supplement information on documented occurrences, species habitat models provide essential information to guide the decision-making of conservation practitioners.

However, not all models are created equal and model products, without interpretation, can be challenging for decision-makers to correctly interpret and appropriately apply. A Species Habitat Model Standard is needed to ensure that model products are scientifically sound, easily interpretable, and consistent across the geographies in which we work. Development of the standard is an important milestone that will enable NatureServe to provide multi-jurisdictional model products, ensuring that we remain a trusted source for comprehensive biodiversity data.

Habitat modeling has been widely used across the NatureServe Network. Beginning with workshops on "Element Distribution Modeling" hosted by the Wyoming Natural Diversity Database in the early 2000s, members of the NatureServe Network have worked to develop scientifically defensible methods for developing species habitat models and to share those methods across the Network. In 2013, several Network programs came together to author the NatureServe Network Modeling Centers Initiative Charter, providing a vision for multi-jurisdictional habitat model products.

Since that time, NatureServe has supported Network habitat modeling efforts by hosting a bi-monthly webinar on species habitat modeling (at various points referred to as "element distribution modeling," "species distribution modeling," and "habitat suitability modeling"), hosting workshops, and convening habitat modeling technical and resolutions teams. Network programs have increasingly collaborated on regional modeling projects, while several states have pursued comprehensive efforts to generate models for most of the species they track. In 2018, NatureServe received funding to pursue the Map of Biodiversity Importance (MoBI) (Hamilton, Smyth, Young et al. in press), a major modeling initiative which significantly advanced collaborative modeling across the continental United States and resulted in the development of first- generation models for over 2,000 imperiled species.

The potential value of species habitat models for supporting conservation has led NatureServe to advocate for greater funding for their development, promote their use in the conservation community, and pursue inclusion of model products in the data we provide. However, standards are needed to guide their development, evaluation, distribution, and appropriate uses (Araujo et al. 2019). In 2019, NatureServe and the Network (represented by NYNHP) collaborated with USGS and USDA scientists on a publication outlining a framework to support the incorporation of models into decision-making processes (Sofaer et al. 2019). That framework, and other publications (Baker et al. 2020, Moskwik et al. 2019) increasingly make the case for greater use of models to guide conservation and move habitat modeling out of the academic sphere and into practical applications. This Species Habitat Model Standard is an important milestone to ensure that models shared by the NatureServe Network are sufficiently rigorous and understandable to end users. By following this standard, we will maintain our unique position in the conservation community as the source of high-quality conservation information products, building on our occurrence data and the species expertise of our Network to offer a wider variety of Biodiversity Location Data.

## 1.2. Goals and Objectives

NatureServe, in consultation with the Network Section Councils, formed a Species Habitat Model Standard Work Group. The goal of the work group was to establish a straightforward and flexible standard for the development and provision of modeled data representing areas of probable species habitat. These data are *not* a replacement for data on documented species occurrences (e.g., element occurrence and observation data); rather, in the absence of comprehensive surveys, these models provide complementary information on where species *may* occur.

The goal of the NatureServe Species Habitat Modeling Initiative is to produce rigorous, spatially-explicit modeled representations (i.e., maps) of likely habitat for each of several hundred at-risk plant and animal species. These models are intended to provide information on where species may occur beyond documented occurrences, and to guide decision-making among practitioners in land development, transportation planning, energy extraction, habitat management, and conservation. Habitat models cover large areas (e.g., states, ecoregions) at a relatively fine grain and are developed using readily available data on species locations together with data on environmental characteristics. The intent of the standard is to support production of these models with consistent processes and common end formats. Doing so will facilitate clear communication about how model products can be interpreted, facilitate rapid updates when new data become available, and facilitate edge-matching of maps produced for neighboring areas.

The work group focused on three main objectives:

1) Develop a standard for Network species habitat models so that data on “likely habitat” can be provided as standardized Network-wide products. The standard covers types of data needed to represent likely habitat, key attribute fields including information on model quality, and metadata requirements.
2) Document best practices for developing species habitat models through an updatable wiki that provides additional guidance on the development of habitat models in a format that is flexible and adaptive as the science of modeling evolves.
3) Outline a data management framework to guide the development of IT systems supporting Network provision of habitat models.

While modelers are expected to utilize modeling techniques appropriate for their situations and needs, which may include both deductive and inductive methods, this standard will help ensure compatible end products and information about methodologies that facilitate universal comparisons among models. This standard will leverage the power and expertise of the Network community by facilitating the compilation, comparison, and use of modeling products throughout the Network.

## 1.3 Terminology

The preferred terminology for the products covered by this standard, as determined by the working group, is **“species habitat models.”** A review of the ecological modeling literature reveals persistent inconsistencies in important terms. Models representing the relative likelihood of species occurrence are variously termed “range models,” “species distribution models,” “habitat suitability models,” “habitat distribution models,” and other similar phrases. We have chosen to use the term "species habitat model" for two primary reasons. First, our methods emphasize identification of the environmental conditions associated most consistently with a species' known locations, i.e., the "habitat" of the species (Figure 1). The actual species distribution may or may not include patches that the model has identified as habitat. Second, our model inputs do not, at least at this time, include the detailed demographic parameters that we feel should be required to imply anything about "suitability."

Operational definitions are as follows:

- *Range:* The generalized geographic area within which a species occurs, without consideration of specific habitat requirements.
- *Habitat*: The combination of resources and environmental conditions that promote occupancy, survival, and reproduction by individuals of a species within its range.
- *Distribution:* The set of locations occupied by individuals of a species within its range at a specific point in time.

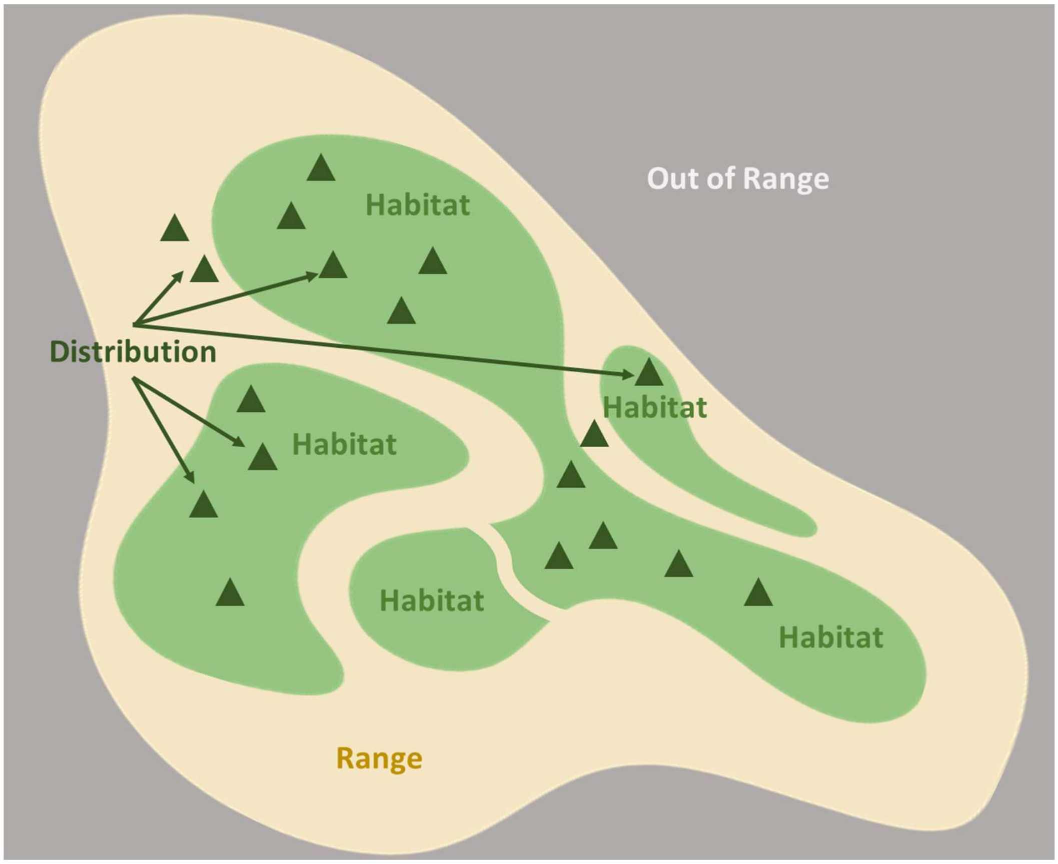


**Figure 1.** Conceptualization of habitat, distribution, and range. Green triangles represent the distribution of the species, while light green polygons represent habitat. Yellow denotes the species range.

Range maps are typically delineated by aggregating map units thought to be occupied by individuals. The map units comprising a range map can be political boundaries such as states, counties, or townships; ecological boundaries such as catchments, larger basins, or ecoregions; or regular units within a lattice, such as squares or hexagons. Because map units are usually coarse and simply-shaped, their utility is limited for resource managers making decisions at finer project scales (Hamilton, Smyth, Young et al. *in press*).

Both habitat and distribution are subsets of range, but they are not equivalent. Habitat is the spatial subset of the range in which individuals of a species are most likely to occur due to favorable environmental conditions. This is consistent with the USFWS regulatory definition of habitat as "the abiotic and biotic setting that currently or periodically contains the resources and conditions necessary to support one or more life processes of a species" (Endangered and Threatened Wildlife and Plants, 2020). Available suitable habitat is not always guaranteed to be occupied, and unsuitable non-habitat areas may sometimes be occupied as individuals move through. The distribution of individuals at any given time is driven by habitat suitability for the species as well as by the species' mobility, tolerance for crossing barriers, behavioral and physiological cues for dispersal of young or propagules, and other dynamic movement variables (Peterson et al. 2011, Soberón and Nakamura 2009, Soberón and Peterson 2005).

Despite the fact that the distribution of individuals sometimes includes unsuitable, non-habitat locations, it is reasonable to assume that individuals usually occur within environments that promote their occupancy, survival, and reproduction – i.e., within habitat. Therefore, a habitat model can be derived by sampling the known distribution of individuals and measuring a suite of environmental variables associated with their locations. A habitat model provides an estimate of the relative likelihood that an individual of the species would be found in a given location and is far more useful for land managers than a general range map.

This standard is specific to species habitat models, with an understanding that those models are often used as a proxy for distribution. Standards for range maps will be addressed elsewhere.

# 2. DEVELOPMENT OF THE STANDARD

## 2.1. Habitat Model Standard Work Group

The NatureServe Network has been engaged in developing species habitat models since the late 1990s. Several individual Network programs developed various solutions for standardized habitat model products, and network trainings, webinars, workshops, and documentation of best practices have been used to promote consistency in approaches across the Network. However, this document represents the first effort to truly standardize Network habitat model products. The Work Group, consisting of staff from NatureServe and Network programs (Appendix I), was established to review and build on those past efforts to create this standard.

## 2.2. Guiding Principles

The following principles were used to guide the development of the standard:

- Keep it simple: Make managing large numbers of species habitat models as simple as possible by focusing the standard on core data outputs that are needed for consistent model products.
- Ensure consistent multi-jurisdictional data: Develop standard definitions, methods for classifying habitat predictions, and rules for aggregating model outputs across boundaries that support clear communication of Network habitat model products to external audiences.
- Meet standards from the literature: Ensure that the standard is guided by recent publications on the development and delivery of habitat models products, including Sofaer et al. 2019.
- Provide options: Recognize that compatible data, especially from external sources, can be collected in varying ways and still be useful. Our requirements should provide flexibility to take advantage of these data (e.g., accommodate both inductive and deductive models).
- Promote assessment of model quality: Facilitate assessment and communication of the quality of and confidence in species habitat models.
- Facilitate the gathering and sharing of large amounts of data: Provide a data management framework to support the exchange and maintenance of species habitat models across the Network.

- Encourage the sharing of best practices for model development within the Network. While the standard is designed to be flexible and non-prescriptive, we address the desire for additional guidance on sound modeling approaches through creation of a living Species Habitat Modeling Best Practices Wiki. (*in development*)

Overarching our work is the concept that "imperfect information is better than no information." Models, by definition, will never be 100% accurate. However, by developing a scientifically robust standard for model products that includes means to communicate uncertainty, we believe biodiversity conservation ultimately will be well served.

## 2.3. Practical Considerations

The theory and techniques for modeling relationships between species and the environment have developed rapidly in recent decades and will continue to evolve. There is a robust body of literature on sophisticated methods, some of which are applicable only to small areas and require highly structured input data that is gathered at high cost over long time periods. In contrast, our initiative is explicitly practical and seeks to use long-proven methods and widely available technologies and datasets to produce many broadly applicable maps relatively quickly. Our work is not at the cutting edge of theory, and is not intended to advance modeling science – it is intended to produce useful information for natural resource professionals to incorporate into their day-to-day decisions; i.e., models achieving the "bronze standard" proposed by Aruajo et al. (2019). We seek to produce useful models for as many of North America's at-risk species as feasible. Given the taxonomic breath and geographic scope of this undertaking, and the need to frequently and rapidly update our models as new information becomes available, our methods may necessarily deviate from ideal approaches.

Perfectly structured field observations, collected expressly for modeling purposes and that cover state- or regional- scale areas completely, are essentially never available for any species. Thus, our work necessarily relies on a variable set of documented field observations, compiled from multiple sources, usually collected over long time spans and for a variety of purposes. Such observations typically record little beyond species presence with no information on reproduction, survivorship, body condition, or other variables that might lend deeper insight into the species-environment relationship. Such multi-sourced data are often biased spatially and temporally, and although pre-model filtering procedures can reduce biases, they do not eliminate them entirely. (Dorazio 2014, Fithian et al. 2014, Lahoz-Monfort et al. 2013, Kery 2011)

For predictor variables we are limited to environmental measures that are available digitally and that cover large areas, as our goal is to extrapolate final models in map form across the full area of interest. This constraint often excludes from our models important but unmapped predictor variables that may strongly influence a species' environmental use.

When possible, we use proven modeling algorithms (e.g., MAXENT, Random Forests, or ensembles thereof) to relate species observations to environmental predictors, as they are robust to multi-sourced data and do not require special technical platforms (Valavi et al. 2021). But for certain species, documented field observations are so scarce and/or spatially biased that they cannot produce an accurate or useful inductive model. In some cases, we recognize deductive modeling, in which expert opinion is used to define environments used by target species, as a legitimate alternative approach for defining habitat. Past modeling projects with scopes similar to ours have shown deductive modeling to be an important means for producing best available predictive occurrence maps, particularly for data-poor species, highly mobile generalist species, or species closely associated with a specific land cover type or feature that is well-mapped (Aycrigg et al. 2015, Boykin et al. 2008). This standard is thus intended to cover both inductive and deductive models.

## 2.4. Work Group Process

Plans to form a Species Habitat Model Standard were announced in Spring 2020 on U.S. and Canada Section Councils and during the regularly occurring Habitat Modeling in the Network webinar series. Members were recruited during those calls, and the Work Group first convened in June 2020. Sub-teams were established after that meeting, each with a Network and NatureServe lead and some cross-pollination between groups to keep efforts aligned. The three sub-teams focused on: Habitat Model Standards, Best Practices, and a Data Management Framework. The sub-teams met periodically over the following months via conference calls and extended online workshops. Periodic updates on progress made by the Work Group were reported to the Network on recurring Habitat Modeling in the Network webinars.

The draft Habitat Model Standard was completed in February 2021 and shared with the entire Work Group (i.e., members from all teams had an opportunity for review). An introduction to the standard was shared with U.S. and Canadian section councils in March 2021 and presented to the Network in more detail in a separate webinar. Following a period of network comment and revision, version 1.0 of the standard was finalized in December 2021.

The Habitat Model Data Standard is anticipated to be a dynamic standard. The Best Practices in particular are anticipated to change as the science of modeling evolves, and thus are to be provided in a living 'wiki' format. Periodic review of the currency of the Best Practices Wiki, including edits, additions, and links to additional Network resources, will be necessary to keep the Best Practices relevant. The Habitat Model Data Standard will also require updates, particularly as the Network gains more experience with the practicality of exchange and sharing of habitat model products. The Work Group will reconvene to make updates as needed.

# 3. THE STANDARD

This Species Habitat Model Data Standard includes a) the definition of a species habitat model and associated terminology, b) specifications for data formats (spatial data and metadata), and c) standards for model documentation required to assess model quality and confidence and determine appropriate uses. It identifies key aspects of the data that facilitate development of consistent Network products, while enabling communication of confidence based on model quality to ensure that the NatureServe Network upholds its reputation as the authoritative source for spatial biodiversity information.

The standard does *not* provide prescriptive criteria related to methods for model development. We encourage data providers to consult the NatureServe Network Species Habitat Modeling Best Practices Wiki for additional guidance on modeling approaches.

## 3.1. Definition of a Species Habitat Model

Here, we define a species habitat model as a predictive map of where environmental conditions favor the occurrence of a species within its range. Methods for developing habitat models may be inductive (using occurrence data, environmental covariates, and statistical methods to predict suitability) or deductive (using expert knowledge about species/environment relationships to map areas of likely occurrence based on environmental data). This definition of a species habitat model is intentionally broad, allowing models developed with different methods to conform to the standard.

## 3.2. Standard Formats for Species Habitat Models Products

This standard covers three types of data associated with species habitat models.

**Metadata** – Information about how the model was developed, model quality, and other key attributes of the model. *Required.*

**Categorical Prediction** – A categorical map product indicating the probability of habitat as distinct classes (i.e., high, medium, low, and not habitat) (Figure 2A). *Required.* Categorical products can be developed based on deductive models (i.e., simple expert-defined rules for defining habitat categories) or inductively derived from a continuous prediction.

**Continuous Prediction** – A continuous map product with values from 0 to 1 indicating relative likelihood that habitat occurs at a given location on the ground (Figure 2B). *Optional (only applies to certain model types).*

In addition, we will also collect spatial data on:

**Modeling extent** – A polygon representation or description of the boundaries within which the model was produced and justification for this extent. *Required.*

**Model confidence scores** will be assigned to each model to aid in interpretation of the data but are not addressed in this section because the assignment of confidence scores represents a post-processing step (see Section 3.3).

Spatial data may be provided either as raster data (the preferred format for inductive terrestrial models) or vector data (the format of many inductive aquatic models, as well as many deductive models). Specifications for each of the above are provided below. Adherence to exact data formats is not required, so long as key features can be cross-walked to standardized fields. However, programs may choose to adopt recommendations contained in this standard regarding data structure, file naming, and attribute field naming conventions to streamline data flow and sharing.

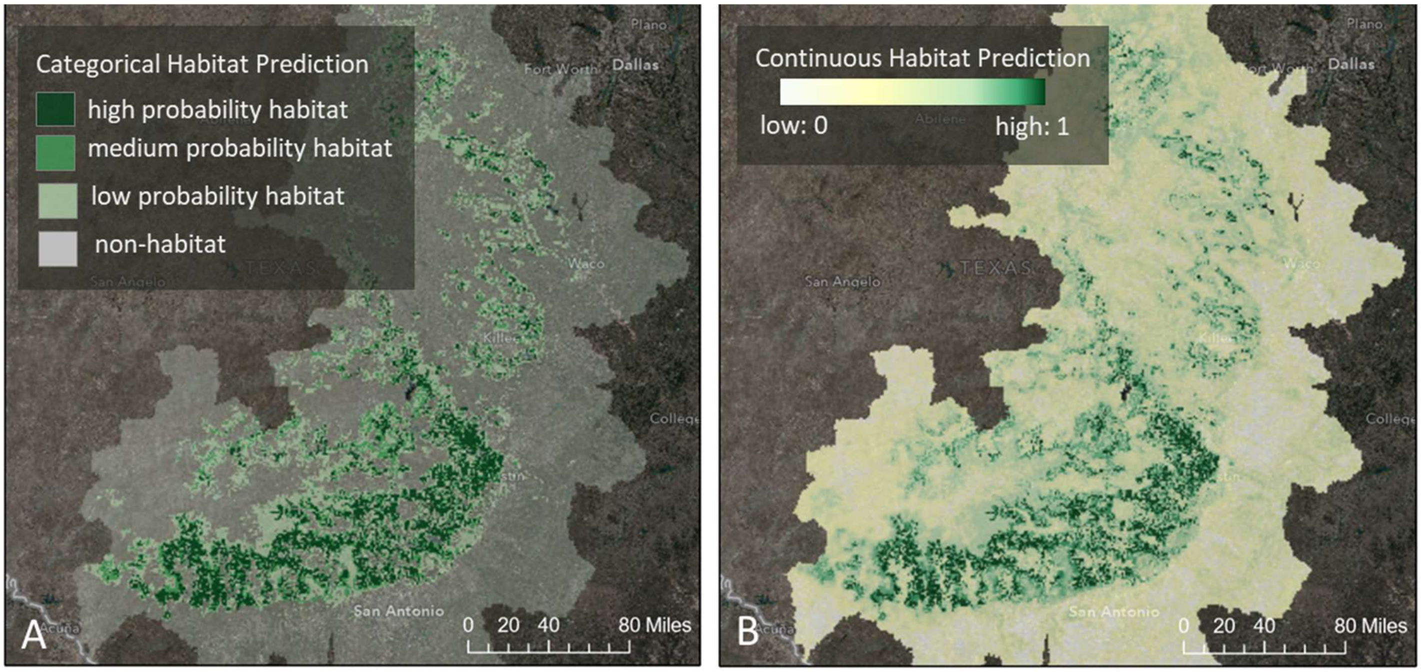


Figure 2. An example of standard species habitat model spatial data products, shown for the *Setophaga chrysoparia* (golden-cheeked warbler) including (A) a categorical prediction of species habitat, and (B) a continuous prediction of species habitat.

### 3.2.1. Metadata (required)

Metadata is required to assist with the cataloguing and evaluation of models. Key metadata information includes:

- **What** (Taxon/Type): the element represented by the distribution.
- **Where** (Location): the extent or footprint within which a species' habitat was modeled, provided in the form of a GIS spatial data type (e.g., shapefile), or optionally, by selecting subnational units from a list.
- **When** (Date): when the model was created, and the time period for which the model is relevant.
- **Who (**Modeling Entity): program and/or person who created the model.
- **How** (Type of Model): general type of model (e.g., inductive vs. deductive; algorithm used).
- **Model Quality**: an evaluation of model quality against scientific standards as described in Sofaer et al. 2019 and adapted for inductive and deductive models. Includes expert assessment of model performance. See Appendix II and Appendix III for details.

Table 1 summarizes the information that shall be provided for each model.

**Table 1**. Key Model Attributes

| Category | Field Label (Displayed) | Optional/Required | Definition | Format |
|---|---|---|---|---|
| WHAT | Element Unique ID | Required | NatureServe unique identifier, which captures both the ELEMENT_GLOBAL_OU_UID and ELEMENT_GLOBAL_SEQ_UID (the ELEMENT_GLOBAL_ID of the system associated with the ELEMENT_GLOBAL_OU_UID). | Text |
| WHAT | Scientific Name | Required | The full scientific name. Can be used for QC purposes. | Text |

| Category | Field Label (Displayed) | Optional/Required | Definition | Format |
|---|---|---|---|---|
| WHAT | Sensitive | Required | Indicator of whether the model outputs should be considered sensitive information | Yes/No |
| WHAT | Model Label | Required | Short model description (e.g., "MoBI Model for *Amystoma cingulatum*") | Text |
| WHAT | Version | Optional (if not provided, will be assigned by NatureServe) | Model version number or identifier used for tracking. | Text |
| WHAT | Location Use Class | Required | Descriptive label to indicate whether the habitat model is specific to a specific life stage or seasonal usage | Multiple Choice: see Biotics Location Use Class options |
| WHERE | Model Extent (footprint) | Required | The geographic area covered by the model. | Well- Known Text (WKT) representation of geometry |
| WHERE | Full Global Range Indicator | Required | Indicator of whether or not the model covers the full known range of the species | Yes/No |
| WHERE | Full National Range Indicator | Required | Indicator of whether or not the model covers the full known range of the species within the nation/nation(s) the model covers | Yes/No |
| WHERE | Full Range Comment | Optional | Comments on the national or global range coverage | Text |
| WHERE | Scale of Use | Optional | Minimum scale of use | Multiple Choice: Any, $1mi^2$ hexagon, $7mi^2$ hexagon (note: data sharing rules provided by Network programs will supersede values provided here) |
| WHEN | Creation Date | Required | Date when the model was produced | Date-time format yyyy-mm-dd |
| WHEN | Model Period | Required | The general period for which the model is intended relevant (current, historic, future) | Multiple Choice: current (default), historic, future |
| WHEN | Model Period – Start | Optional (required for future or historic models) | The time period for which the model is valid | Year |
| WHEN | Model Period – End | Optional (required for future or historic models) | The time period for which the model is valid | Year |
| WHO | Author Organization | Required | Name of organization producing model | Text |
| WHO | Authored For Organization | Optional | Name of organization for whom the model was generated | Text |
| WHO | Contact Name | Required | Name of person to contact for more information about the model | Text |
| WHO | Contact Email | Required | Email address of contact person | Text |
| WHY | Model Goals | Required | Information on why the model was developed and its intended use | Multiple choice (pick all that apply): guiding inventory, conservation planning, environmental screening, …., other (specify) |

| Category | Field Label (Displayed) | Optional/Required | Definition | Format |
|---|---|---|---|---|
| WHY | Model Goal Description | Optional | Additional short description of the project for which the model was developed, comments on confidence in the model for the intended uses listed above, and citation and web link for any associated reports or publications. | Text |
| HOW | Model Type | Required | Indicator of the general model type | Multiple choice: inductive, deductive, or combination |
| HOW | Model Method | Required | Indicator of specific modeling method(s) | Multiple choice: Random Forest, MaxEnt, Ensemble, other (with comments), etc. |
| HOW | Model Method Comments | Optional (required if "Other" is selected for Model Method) | Additional short explanation of the modeling method(s) used | Text |
| QUALITY | Presence Data Quality | Required (inductive and deductive models) | Indicator of the quality of presence data used for training the model | Multiple Choice: interpret with caution, acceptable, ideal |
| QUALITY | Presence Data Comments | Required (deductive models) or optional (inductive models) | Comments on the quality of presence data used for training (inductive models), or on the general state of knowledge of the species' habitat requirements (deductive models) | Text |
| QUALITY | Absence/ Background Data Quality | Required (inductive models) | Indicator of the quality of absence or background data used for training | Multiple Choice: interpret with caution, acceptable, ideal |
| QUALITY | Absence/ Background Data Comments | Optional | Comments on the quality of absence or background data used for training (e.g., the source of known absences or the method used to generate pseudoabsences.) | Text |
| QUALITY | Evaluation Data Quality | Required (inductive and deductive models) | Indicator of the quality of data used to test the model output | Multiple Choice: interpret with caution, acceptable, ideal |
| QUALITY | Evaluation Data Comments | Optional | Comments on the method (e.g., independent evaluation, cross-validation) and quality of data used to test the model output | Text |
| QUALITY | Predictor Relevance | Required (inductive and deductive) | Indicator of the relevance of predictor variables included in the model | Multiple Choice: interpret with caution, acceptable, ideal |
| QUALITY | Predictor Relevance Comments | Optional | Comments on the relevance of predictor variables included in the model, including methods used to select predictors. | Text |
| QUALITY | Spatio- Temporal Data Alignment | Required (inductive and deductive) | Indicator of the spatio-temporal alignment between training data points and the environmental predictor values assigned to them | Multiple Choice: interpret with caution, acceptable, ideal |
| QUALITY | Spatio-Temporal Data Alignment Comments | Optional | Comments on the spatio-temporal data alignment | Text |

| Category | Field Label (Displayed) | Optional/Required | Definition | Format |
|---|---|---|---|---|
| QUALITY | Algorithm Choice | Required (inductive and deductive) | Indicator of the suitability and statistical rigor of the algorithm(s) chosen for modeling | Multiple Choice: interpret with caution, acceptable, ideal |
| QUALITY | Algorithm Choice Comments | Optional | Comments on the algorithm choice | Text |
| QUALITY | Sensitivity Analysis | Required (inductive) | Indicator of the level of model sensitivity analysis | Multiple Choice: interpret with caution, acceptable, ideal |
| QUALITY | Sensitivity Analysis Comments | Optional | Comments on the sensitivity analysis | Text |
| QUALITY | Performance Evaluation | Required (inductive and deductive) | Indicator of the statistical rigor of the performance evaluation | Multiple Choice: interpret with caution, acceptable, ideal |
| QUALITY | Performance Statistic | Required (inductive and deductive) | Specific measure or measures used to evaluate performance | Multiple Choice: AUC, TSS, Overall Sensitivity & Specificity, Other, None |
| QUALITY | Performance Statistic Value | Required (unless Performance Statistic = none) | Numeric result (or results) of the performance evaluation | Number |
| QUALITY | Performance Evaluation Comments | Optional | Comments on model performance | Text |
| QUALITY | Model Review Status | Required (inductive and deductive) | Indicator of the level of model review | Multiple Choice: interpret with caution, acceptable, ideal |
| QUALITY | Model Review Status Comments | Required | Comments on model review process | Text |
| QUALITY | Model Review Outcome | Required (inductive and deductive) | | Multiple Choice: Good, Moderate, Poor, Unreviewed |
| QUALITY | Model Review Outcome Comments | Required | Additional information on who provided reviews and the comments received | Text |
| QUALITY | Iterative Comments | Optional | Comments on model iterations, including whether changes were made based on reviews received | Text |
| QUALITY | Map Products | Required (inductive and deductive) | Indicator of the quality of map products produced as model outputs | Multiple Choice: interpret with caution, acceptable, ideal |
| QUALITY | Map Products Comments | Optional | Comments on the map products | Text |
| QUALITY | Supplementary Metadata | Required (inductive and deductive) | Indicator of the quality of supplementary metadata supporting interpretation of the model and its outputs | Multiple Choice: interpret with caution, acceptable, ideal |
| QUALITY | Supplementary Metadata Comments | Optional | Comments about the quality of supplementary metadata | Text |
| QUALITY | Reproducibility | Required (inductive and deductive) | Indicator of the level of information provided so that others could reproduce the model | Multiple Choice: interpret with caution, acceptable, ideal |

| Category | Field Label (Displayed) | Optional/Required | Definition | Format |
|---|---|---|---|---|
| QUALITY | Reproducibility Comments | Optional | Comments on model reproducibility | Text |

For inductive models, the criteria for model quality ratings are based on Sofaer et al. (2019), with minor modifications. The criteria for deductive models were developed by the Working Group following a similar framework. These criteria are provided as Appendix II and Appendix III respectively. Criteria for scoring model review outcomes are provided in Appendix IV.

Supplementary metadata should be provided in the form of a PDF document. Recommendations for what this metadata should encompass are provided in the Guidelines for Accompanying Interpretive Materials section of the Species Habitat Modeling Best Practices Wiki, and include:

- the source, quantity, and date range of species occurrence data used for model training,
- environmental predictor data used and their sources, which variables contributed most to the model, and the relationship between those variables and the occurrence data,
- statistical thresholds for defining habitat categories and threshold selection methods,
- any statistical measures of model performance (e.g., AUC, kappa) beyond what is provided in the key model attributes, and
- details on reviews received and any modifications made based on those reviews.

See Appendix V for examples of PDF metadata.

### <u>3.2.2 Categorical Products (required)</u>

The categorical habitat model products indicate, based on inductive or deductive model outputs, how likely it is that an individual of a given species will occur at a given location using pre-defined categorical classes - i.e., high probability habitat, medium probability habitat, low probability habitat, non-habitat. Following the assumption that species will occur with greater frequency in habitats with the best combination of environmental variables for supporting survival and reproduction, the likelihood categories can be interpreted both as how suitable the habitat is to support the species and the likelihood that the species will be found in the that location.

#### 3.2.2.1 Definitions of Habitat Categories

NatureServe Network categorical species habitat model products should conform to the following categories:

- **High Probability:** Habitat is optimal for species occurrence. Environmental conditions at the location are nearly identical to conditions where the species is frequently observed or where the location is known to be high-quality species habitat. The probability that an individual of the species would be found in this location is high.
- **Medium Probability:** Environmental conditions at the location are similar to conditions where the species is regularly observed. The probability that an individual of the species would be found in this location is moderate.

- **Low Probability:** Environmental conditions at the location are similar to conditions where the species is rarely observed or where observations are known to be in poor quality species habitats. The probability that an individual of the species would be found in this location is low.
- **Non-Habitat:** Environmental conditions at the location are dissimilar to conditions where species observations have occurred. The habitat is considered unsuitable, and the species is not expected to be found in this location unless incidentally passing through.

The high and medium probability categories should be considered for management, planning, permitting, survey, and other decisions. The low probability category may be more contextual for these types of decisions.

Data conforming to the standard can represent one or more of those categories, though when possible, data submitters are encouraged to submit data that represents all four.

When inductive modeling techniques are used, these classes are typically defined by thresholding the probability surface, using a combination of statistically generated threshold values (e.g., minimum training presence, equal sensitivity plus specificity) and expert judgment. Because experience has taught that a single, inflexible approach for selecting thresholds to define categories does not produce good results for all species, this standard does not provide a prescriptive method for setting thresholds, leaving threshold selection up to the judgement of the model producer. Guidelines for threshold selection will be included in the NatureServe Network Habitat Modeling Best Practices Wiki. The reasoning behind the threshold selection should be provided in the metadata PDF.

Deductive methods often result in binary habitat maps – i.e., only two categories of data that can best be interpreted as habitat and non-habitat. In these cases, the modeler must document whether the habitat class is best interpreted as high or medium probability habitat, with the default being medium probability.

#### 3.2.2.2 Data Formats for Categorical Products

These data can be provided as any of several different spatial data formats (raster, polygon, polyline), but the projection must be adequately defined.

For raster data, a classified (categorical) raster should be provided that bins the prediction into the habitat classes described in Section 3.2.2.1 (i.e. high probability, medium probability, low probability, and non-habitat). The raster should have an associated attribute table containing the category assignments, with values within that field matched to one of the four standard data categories. Optionally, the original continuous prediction raster can also be provided (see Section 3.2.3).

For vector files, a field containing the category assignment should be identified, and values within that field matched to one of the four standard data categories.

While it may be feasible to match non-standard values to these standard attributes via field-mapping during the data upload, processes to do so have yet to be implemented. Recommendations for formatting of attribute fields are thus provided here. Application of these recommendations will streamline the exchange of habitat models.

**For rasters representing multiple habitat categories:**

- The raster value, and/or a separate field labeled "hab_value", shall correspond to unique category assignments
  - *preferred values:*

| Category | hab_value[1] | hab_cat[2] |
|---|---|---|
| High Probability | 1 | high probability |
| Medium Probability | 2 | medium probability |
| Low Probability | 3 | low probability |
| Non-Habitat | 4 | non-habitat |

[1] hab_value = habitat value (numeric code)

[2] hab_cat = habitat category (text description)

- A text field in the raster attribute table should be used to provide narrative text on the meaning of those values (optional, but recommended if the values do not conform to the recommendations above)
  - *preferred name for this field: hab_cat*
  - Non-standard entries in this field may be allowed if they can be matched to one of the four categories at the time of data upload.
    - *preferred field entries: high probability, medium probability, low probability, non-habitat*

**For vectors representing multiple habitat categories**

- The file attribute table shall contain a field providing category assignments
  - *preferred: hab_cat*
  - Non-standard entries in this field are allowed if they can be matched to one of the four categories at the time of data upload.
    - *preferred field entries: high probability, medium probability, low probability, non-habitat*

Alternately, processes may be developed whereby data could be provided in the form of a continuous raster or vector dataset with an attribute representing continuous values (see 3.2.3) alongside tabular data identifying the breaks at which to define the categories. This approach has the benefit of reducing data storage requirements and will be further explored as data management processes are implemented.

**For rasters and/or vectors representing a single habitat category**

In some cases, a model may represent just a single category of data (e.g., just "high probability habitat"). In these cases, a category field will not necessarily be required if the category represented by the data is provided at the time of data upload (pending full development of the model upload utility), though including "hab_cat" and "hab_value" attributes is still recommended as a best practice. Note however that if multiple categories of data are available for a particular area (e.g., high and medium probability habitat) those data should be merged into a single file and not provided individually.

### 3.2.3 Continuous Prediction (optional)

The continuous prediction of habitat is a map product with values from 0 to 1 indicating relative likelihood that habitat occurs at a given location on the ground. This is an optional product, and typically is only available when inductive modeling techniques are used. Because the probability surface will not exist for all habitat models produced by the Network, and because the values can be difficult to correctly interpret by the casual user, provision of this data is not a requirement. Recognizing that many educated users are highly interested in these data, and that the data can be responsibly applied to meet needs that categorical data cannot (for example, in weighted sampling designs), we encourage programs to provide these data when available.

Probability surfaces can be provided as any of several different spatial data formats if the projection is adequately defined and the probability value (0 to 1) is represented by the value attribute (rasters) or in a defined field within a polygon or polyline attribute table (*preferred name for this field: hab_prob).*

## 3.3 Assignment of Confidence Scores

A key need when sharing habitat models is to communicate information about confidence in the prediction, since models can vary widely in quality. The Working Group recognizes the need to uphold the high data quality standards for which NatureServe data are known, but also recognizes that models – by definition – will never be perfect representations of habitat.

All models reflect some degree of uncertainty. Nonetheless, models often provide the best available indication of where a species may potentially occur—essential information for the public, conservation practitioners, and all relevant stakeholders involved in land-use decision making. It is an obligation for all modelers to clearly communicate the nature and degree of uncertainty underlying the data and methods behind their models, to help data consumers understand model limitations and appropriate end uses.

To meet this need, we developed a rule set for assigning model confidence scores to individual models. Our goals with the confidence scoring are three-fold:

- to help data consumers understand appropriate uses for any particular model,
- to enable NatureServe to provide “less than perfect” models while upholding our scientific credibility, and
- to provide a means to document and communicate the nature in which certain models can be improved.

To ensure consistency, confidence scores will be assigned to individual models via a post-processing step, drawing on the metadata submitted with the model and model review scores.

### 3.3.1 Model Confidence Definitions and Appropriate End Uses

General definitions for model confidence are provided below, with specific rules for assigning confidence values outlined in the subsequent sections. Models will only be assigned a low, medium, or high confidence score if they meet minimum standards for inputs and methods.

**High Confidence:** The model was developed using acceptable or ideal inputs and methods and performs well based on both statistical validation measures and expert review. While the model results cannot guarantee species presence or absence, all measures suggest that the model provides a successful approximation of species habitat. It is appropriate to use the model products for a wide variety of end uses, from conservation planning to environmental screening.

**Medium Confidence:** The model meets or exceeds minimum criteria for inputs and methods and validation statistics and/or expert review indicate it provides a reasonable approximation of habitat. The model is appropriate for use in a variety of contexts but should not be the sole source of evidence for high-stakes applications.

**Low Confidence:** Statistical measures of performance, model review results, and/or the integrity of modeling methods call into question the accuracy of the habitat prediction. Low confidence models may successfully predict habitat, but without further information, we are unable to say so with certainty. These models are thus best used only for "low stakes" informational purposes such as exploring the relationship between the species and environmental variables, guiding additional inventory efforts, or broad-scale conservation planning applications such as inclusion in measures of biodiversity at a statewide scale.

Ultimately, users must decide on their own level of risk tolerance when determining whether a habitat model should be used in a particular context. However, Table 2 provides general guidelines on appropriate end uses for models with different confidence scores.

**Table 2**. Appropriate uses for models of differing confidence.

| | Low | Medium | High |
|---|---|---|---|
| Range determination | √ | √ | √ |
| Guiding field surveys | √ | √ | √ |
| Conservation planning - broad scale | √ | √ | √ |
| Conservation planning - fine scale | | √ | √ |
| Climate change vulnerability assessment | | √ | √ |
| Initial environmental screenings | | √ | √ |
| Evaluation of restoration opportunities | | √ | √ |
| Informing listing decisions | | √ | √ |
| Environmental review | | | √ |
| Informing critical habitat decisions | | | √ |
| Species translocations | | | √ |

### 3.3.2 Scoring of Model Confidence

Three factors will be considered when determining the confidence scores: (1) expert model review, (2) statistical performance, and (3) the integrity of model inputs and methods. Rules for assigning confidence for each of those factors as well as how the three factors get rolled up into an overall confidence score are outlined below.

#### 3.3.2.1 Expert Model Review

Model review scores provide an expert-derived measurement of model performance. While reviews can be subjective, feedback from scientists who know the species well is one of the clearest ways to determine whether predictive methods result in a habitat map that is consistent with what is already known about the species.

For more information on model review criteria see Appendix IV. A model review confidence score will be assigned based on (a) whether the model has been comprehensively reviewed by scientists with expertise covering the geographic distribution of the species and (b) the assessment by those experts of how well the model performs. Models that have been comprehensively reviewed and receive good marks from all reviewers (mean score ≥4 on a 5-point scale, with all review scores ≥3) will be considered "high" confidence. Models will be assigned a model review confidence score of "medium" if all review scores are moderate or above (mean score ≥3 on a 5-point scale, with all review scores ≥2), even if the model t has not been comprehensively reviewed across the full distribution of the species. Models will be assigned a model review confidence score of "low" if review scores do not meet the criteria for moderate performance (see

Appendix IV) or if the model has not been reviewed by a species expert or experts. Figure 3 illustrates these decision rules for assigning the model review confidence score.

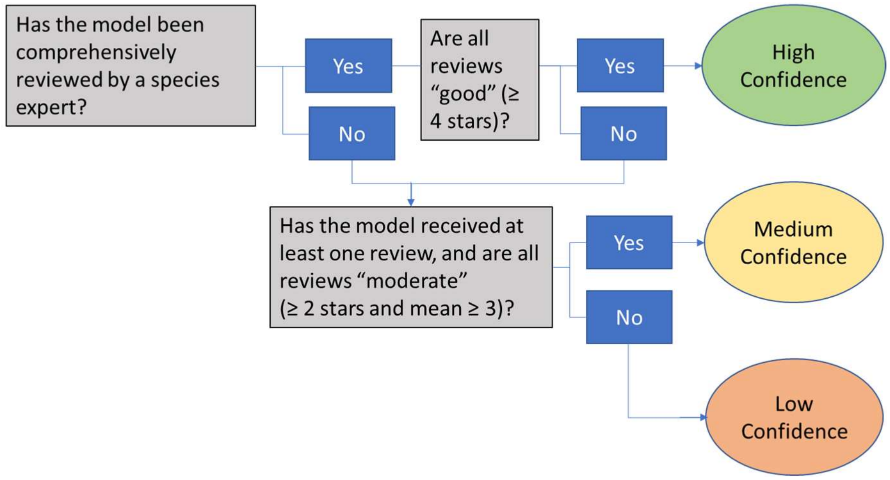


**Figure 3.** Decision tree for assigning model review confidence scores. See Appendix IV for more information on the assignment of "star" ratings.

#### 3.3.2.2 Statistical Performance

There are various statistical measures for assessing model performance, and they may vary with the modeling methods used, particularly between inductive and deductive models. Here we provide several options for statistical measures of model performance and rules for applying them to assign low, medium, or high statistical performance confidence scores. If no statistical measures of model performance are provided, a score of "low confidence" for this metric will be assigned. If modelers have used a different approach to measure model performance than those outlined here, they should contact NatureServe to ascertain whether this standard may be updated to include that measure.

**Inductive Models**

The performance of inductive models should be evaluated using one or more external validation procedures, details of which will be addressed in the Best Practices Wiki. Modelers should report threshold-independent measures such as the area under the curve (AUC) of the receiver-operating characteristic (ROC) plot and, if appropriate, threshold-based measures such as the True Skill Statistic (TSS). When threshold-based measures are used, the method for deriving the threshold and the habitat class to which it applies (high, medium, or low probability) should also be reported.

Performance measures are best used as relative methods to compare among modeling approaches or algorithms: if algorithm A validates with an AUC of 0.90 and algorithm B reports AUC = 0.81, then we might conclude that algorithm A does a better job at discriminating habitat vs. non- habitat within the context of the modeling environment applied. Performance measures are more difficult to interpret as stand-alone values. For example, using AUC alone as an indication of model performance may be misleading if the validation involves comparing large areas of non-habitat to occupied areas. With that said, we recognize that an AUC of 0.5 is more likely to be a 'poor' model and an AUC of 0.98 is more likely to be a 'good' model. Given this generality, we provide ballpark values for AUC and TSS that can be used in confidence assessment (Table 3). We anticipate that methods for scoring model confidence based on statistical performance will be further refined over time.

**Table 3**. General cutoffs for discriminating between low, medium, and high statistical performance scores using difference performance metrics.

| | Low | Medium | High |
|---|---|---|---|
| Area Under the Curve (AUC) | < 0.80 | 0.80 – 0.90 | > 0.90 |
| True Skills Statistic (TSS)[2] | < 0.50 | 0.50-0.80 | > 0.80 |
| *Additional metrics TBD* | | | |

*[1]Modelers should specific the habitat category on which the TSS is based. We recommend basing the confidence assessment off the high probability category.*

**Deductive Models**

Because deductive models are often created in situations where training and evaluation data are limited or employ methods that require using all documented occurrence locations, calculating statistical measures of model performance for deductive models is not always possible. In these cases, the statistical performance confidence rating will be recorded as "low," acknowledging the uncertainty associated with models of this type.

In some situations, independent data are available for use in model evaluation. When a model developer creating inductive models has enough independent samples of presence locations and non-detection locations (e.g. best estimate of absences), the statistical performance of categorical deductive model outputs can be evaluated in much the same way as inductive models, and the performance thresholds in Table 3 apply.

### 3.3.2.3 Integrity of Model Inputs and Methods

The final factor considered for the overall model confidence score is the integrity of model inputs and methods. Here, we apply the model evaluation rubric published in Sofaer et al. 2019 to determine whether the means by which a model was developed conforms to established scientific criteria - specifically the input data (species locality data and environmental predictor data) and aspects of the modeling process (algorithm and sensitivity). The confidence score for integrity of model inputs and methods is determined based on whether minimum criteria for different aspects of the modeling process are met (Table 2). The specific criteria for rating each category are outlined in Appendices II (inductive models) and III (deductive models).

**Table 4.** Minimum criteria for assigning confidence scores for model inputs and process.

| | | Minimum Criteria | | |
|---|---|---|---|---|
| | | Low Confidence | Medium Confidence | High Confidence |
| Species Data | Presence Data | Acceptable | Acceptable | Ideal |
| | Absence/Background Data | Interpret with Caution | Acceptable | Acceptable |
| Environmental Predictor Data | Ecological and Predictive Relevance | Interpret with Caution | Acceptable | Ideal |
| | Spatial and Temporal Alignment | Interpret with Caution | Acceptable | Acceptable |
| Modeling Process | Algorithm and Statistical Rigor | Interpret with Caution | Acceptable | Ideal |
| | Sensitivity | Interpret with Caution | Acceptable | Ideal |

If one or more elements of the modeling inputs and process score as “interpret with caution” the model will be considered as low confidence for this metric. The one exception to this is the species presence data inputs; models must have a minimum rating of “acceptable” for this factor to meet the minimum bar for NatureServe Network habitat model products.

If all elements of the modeling inputs and process score as “acceptable” the model can be considered medium confidence. To be considered high confidence, all elements of the modeling inputs and process must score as “ideal” with the exception of (1) absence data and (2) spatial and temporal alignment of the environmental predictor data, which may be “acceptable.” The bar for an ideal rating of those two factors, as set in Sofaer et al. 2019, is so high as to be nearly impossible to achieve for most modeling efforts outside the academic sphere.

### 3.3.3 Assigning Overall Confidence

The overall confidence score will be based on a combination of the (1) expert model review, (2) statistical performance, and (3) the integrity of model inputs and methods, based on the model quality ratings as defined above. In general, the overall confidence score is set to be equal to the lowest score of any of the three categories. That is, if a model scores as “low” for either model review, statistical performance, or integrity of modeling inputs and methods, the overall confidence score will be “low”. However, recognizing that none of these three measures is infallible, we have also integrated a system of checks and balances. If a model scores low in one category, but high in both the others, the overall confidence rating will be bumped up to “medium”. For example, a model that scores low for integrity of methods because of a lack of high-quality input data, but which performs well based on both expert review and validation statistics, would be considered of medium confidence. Models with overall high confidence must score high in all three categories. Models that score low in all three categories are considered sub-standard and will not be included in model products served by NatureServe.

**Table 5.** Assignment of Overall Confidence Score based on component scores.

| Statistical Performance | Integrity of Methods | Expert Review | Overall Confidence Score |
|---|---|---|---|
| High | High | High | High |
| | | Medium | Medium |
| | | Low | Medium |
| | Medium | High | Medium |
| | | Medium | Medium |
| | | Low | Low |
| | Low | High | Medium |
| | | Medium | Low |
| | | Low | Low |
| Medium | High | High | Medium |
| | | Medium | Medium |
| | | Low | Low |
| | Medium | High | Medium |
| | | Medium | Medium |
| | | Low | Low |
| | Low | High | Low |

| Statistical Performance | Integrity of Methods | Expert Review | Overall Confidence Score |
|---|---|---|---|
| | | Medium | Low |
| | | Low | Low |
| Low | High | High | Medium |
| | | Medium | Low |
| | | Low | Low |
| | Medium | High | Low |
| | | Medium | Low |
| | | Low | Low |
| | Low | High | Low |
| | | Medium | Low |
| | | Low | Deficient; Don't Include in Model Products |

# 4. DEVELOPMENT OF STANDARDIZED HABITAT MODEL PRODUCTS

The habitat model data standard will facilitate the development of consistent multi-jurisdictional model products, including habit model information provided via NatureServe Explorer and NatureServe Explorer Pro. To develop these products, processes will need to be put in place to (1) upscale data from multiple sources to a common scale, (2) handle potentially conflicting models covering the same geography, and (3) reflect the dynamic nature of model products, understanding that models can and should be refined over time with new data and/or improved methods. Here, we provide guiding principles for managing dynamic model products and for combining habitat model information from various sources. We acknowledge that continued engagement of a Network advisory group will be necessary to flesh out the specifics of how best these guiding principles should be applied.

## 4.1. Upscaling Models to a Common Spatial Framework

This standard is intended to cover models developed at disparate spatial scales, allowing model contributors to work at the spatial resolution that best meets their needs. Some audiences will be most interested in access to those models at the resolution at which they were produced. However, we also anticipate a need to upscale model products to a common spatial framework. Doing so will enable efficient provision of consistent multi-jurisdictional model products and facilitate data sharing.

The nested hexagon framework utilized by NatureServe Explorer provides an ideal means of summarizing habitat models (Figure 4). Using the four categories of data as previously defined (i.e., non-habitat, and low, medium, and high probability habitat), likely presence of species habitat can be summarized at multiple resolutions (i.e., 1 $mi^2$, 7$mi^2$, 49 $mi^2$, or 343 $mi^2$ hexagons).

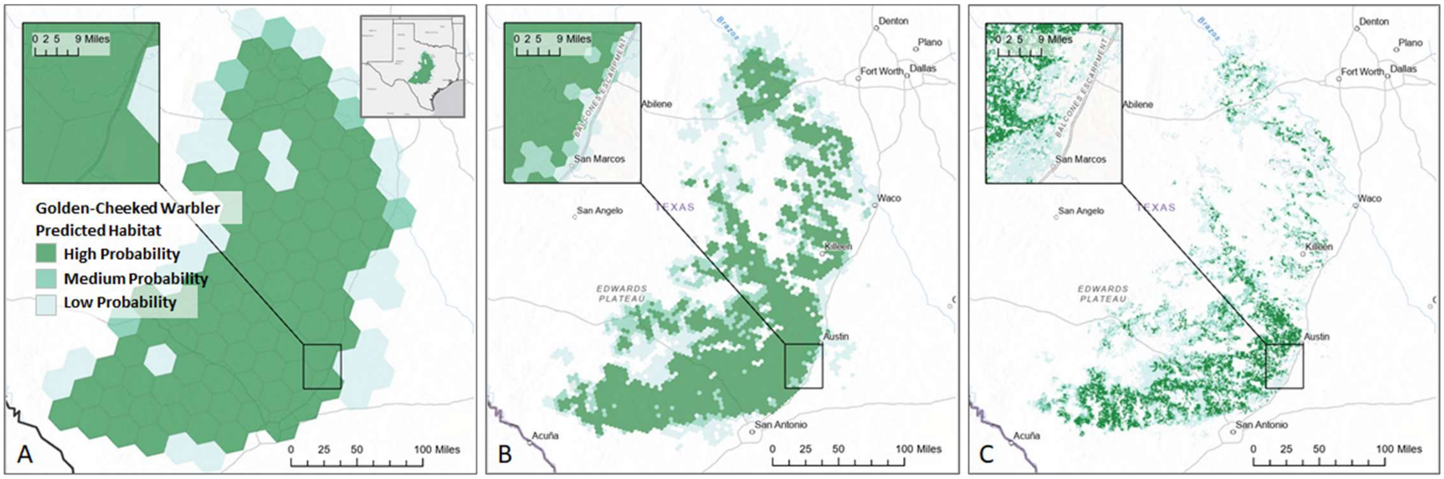


**Figure 4.** The habitat model for the golden-cheeked warbler (*Setophaga chrysoparia*) upscaled to 343 $mi^2$ hexagons (panel A), 7$mi^2$ hexagons (panel B), and at its native resolution as a 330$m^2$ raster (panel C).

Upscaling to hexagons also facilitates stitching together range-wide species habitat maps by combining models developed for different jurisdictions. Figure 5 illustrates a hypothetical example of how the hexagon framework can be used to provide multi-jurisdictional habitat model results. Panel A shows how habitat modeling outcomes might be communicated for a species that occurs in three states (Idaho, Wyoming, and Montana) but which has been modeled in only two (Wyoming and Montana). Panel B provides a more complicated hypothetical scenario, where the habitat categories provided by Wyoming and Montana differ in ways permissible under this standard (that is, only "high probability" habitat is mapped in Wyoming). The resultant map has edge mapping issues, but nonetheless represents a best available multi-jurisdictional Network habitat model product. Note that these maps show habitat model results using coarse scale hexagons (343 $mi^2$), but we anticipate interest will be highest for finer (e.g., 1 $mi^2$) polygons.

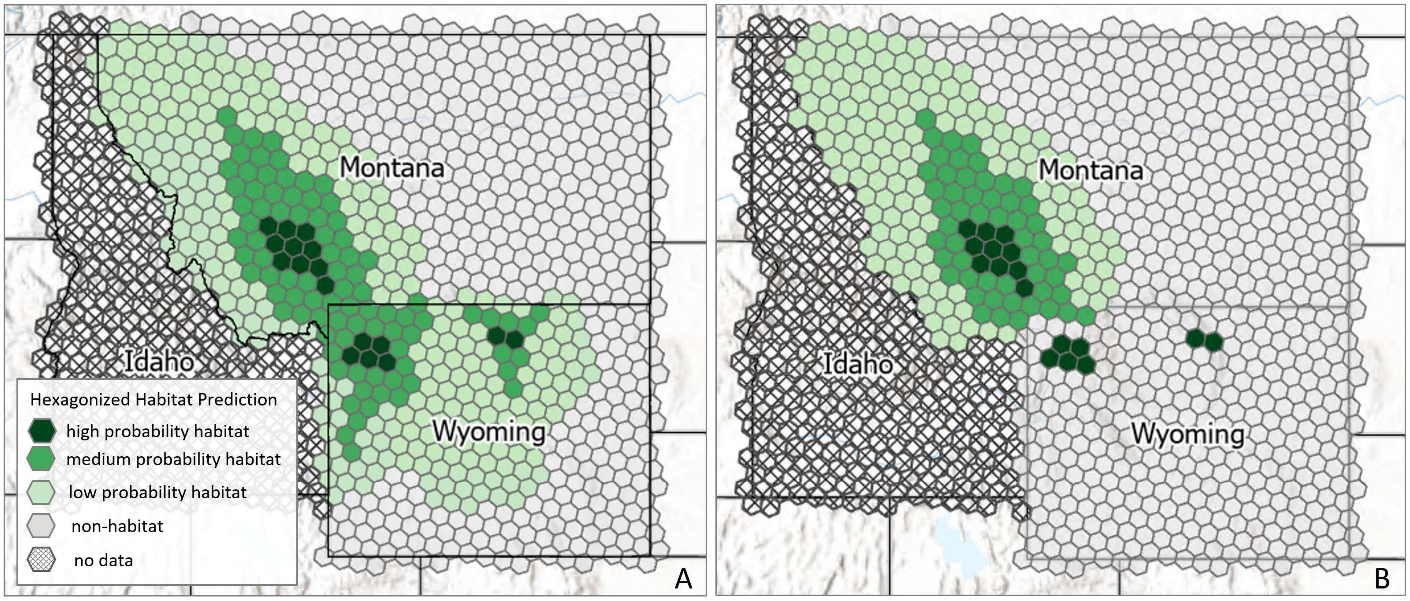


**Figure 5**. A hypothetical example of model results from multiple states scaled up to a common hexagon framework. In Panel A, habitat models available from both Montana and Wyoming, but not Idaho, using all four categories of habitat. Panel B shows how the results would look if only the "high probability" category was mapped in Wyoming.

Upscaling to hexagons may be accomplished in different ways. One recommended approach is to assign the maximum value occurring within a given polygon to that polygon – that is, if any "high probability" habitat occurs within the shape, the polygon will be assigned a value of "high probability". This is a relatively conservative (protective) approach that emphasizes identifying areas that may contain habitat and which will overpredict habitat, especially at coarse scales. Another means of summarizing predicted habitat at the hexagon scale is to use a majority rule, where the polygon is assigned a value based on whether a majority of the area is high or medium probability habitat. This approach works well for wide-ranging species, for which major overpredictions of habitat may occur if the maximum rule is followed. Applying species-specific minimum area or adjacency rules may also be appropriate. We anticipate that the maximum value approach will provide a good starting point for upscaling the habitat models, but that methods should be refined over time to be tailored based on species traits. Another option would be to allow data consumers to control the display by offering both maximum- and majority-rule summarizations. Whichever methods is applied should be documented and clearly communicated to model consumers.

## 4.2. Communicating Models from Multiple Sources

The common spatial framework provides a means to facilitate combination of models from disparate sources. This may at times result in edge-mapping issues (see Figure 5) but that alone is not sufficient reason to avoid serving multi-jurisdictional model products. The Network Habitat Model Standard Working Group recommends that when a range-wide model *is* available for a species, that model should have precedence, particularly when models are communicated at range-wide scales. This avoids edge-mapping issues and simplifies the communication of already complicated habitat model products. If multiple range-wide models exist, precedence should be given to the model with the higher confidence score. In the absence of a range-wide model, if multiple models exist with overlapping extents, precedence should also be given to the model with the higher confidence score.

Acknowledging though that range-wide models are not always the "best" models, particularly when used at local scales, the Working Group recommends that when both range-wide and sub-national models exist for a given species, information about the availability of sub-national models be communicated in some manner. Exactly how that is accomplished, whether through geographically specific reporting on model availability, display of different models at different zoom levels in dynamic mapping applications, preferential selection of models based on the confidence assessment, provision of ensemble models, or some combination of the above, is still to be determined, and will be a focus of continued engagement of a Network advisory group.

## 4.3. The Model Development Cycle

Models represent our best estimate of species habitat made with the information available at a given point in time. They thus can, and should, be refined as better information becomes available and modeling methods are advanced. We anticipate that models served by the NatureServe Network will be dynamic: models will be updated as methods advance and new information becomes available. As models are refined and improved, confidence scores will increase.

The dynamic model development cycle is illustrated in Figure 6. If a model fails to meet Network standard, revisions will be required before it is incorporated into NatureServe's Biodiversity Location Dataset. Models may also be revised to increase confidence scores, or to incorporate new species occurrence data, refined environmental predictors, or advanced methods. Habitat models should be retired and archived when a newer, higher confidence model for the same geography becomes available.

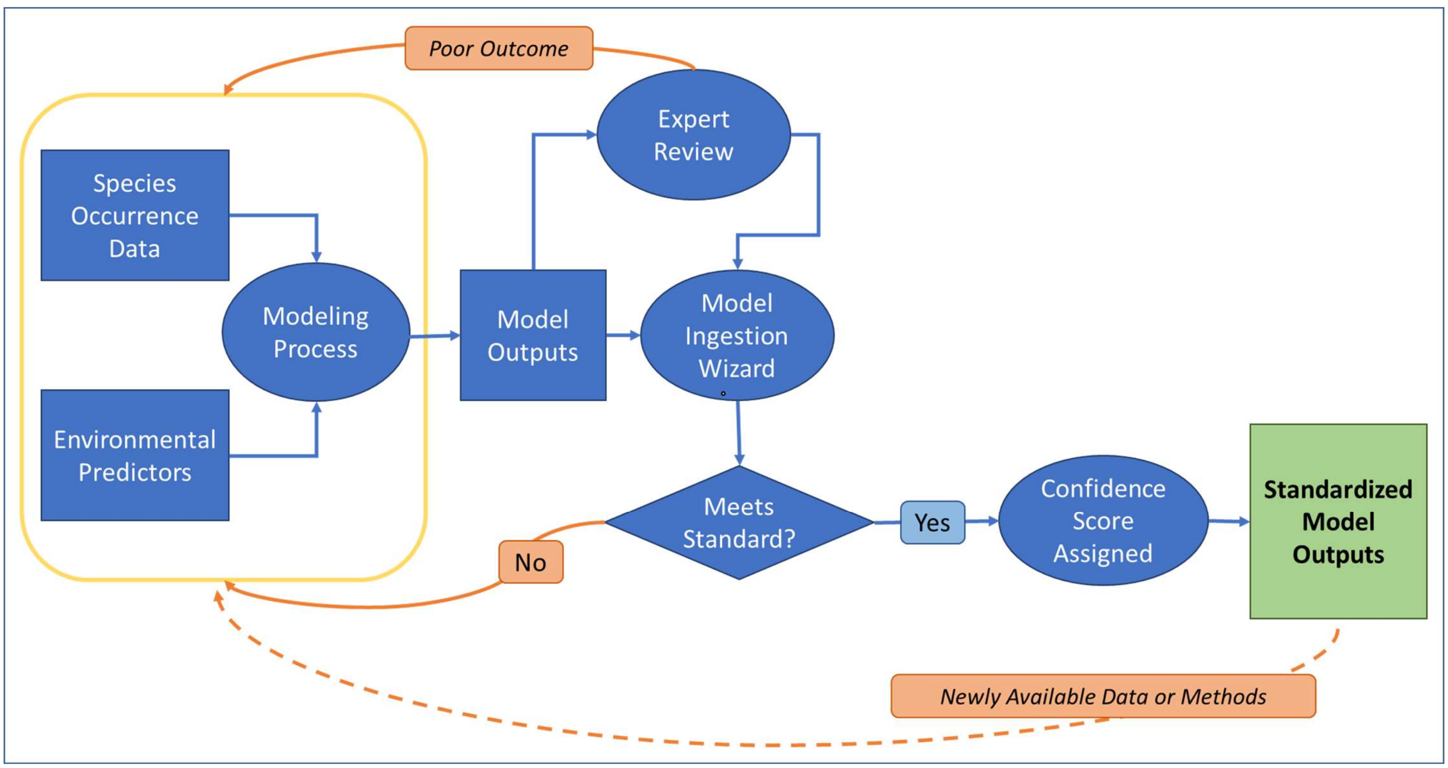


**Figure 6**. The model development cycle, from model development (yellow box) to creation of standardized model outputs. Orange lines indicate triggers for model refinement. The dotted orange line reflects the dynamic nature of all models; as better data and methods become available, refined models can be produced and model confidence increased. Models without expert review will, by definition, have lower confidence scores.

# 5. HABITAT MODEL DATA MANAGEMENT FRAMEWORK

## 5.1. A Vision for Collecting, Aggregating, and Managing Habitat Model Data

Implementing the Habitat Model Standard will facilitate the flow of habitat models throughout the NatureServe Network and support the creation of an expanded set of Biodiversity Location Data knowledge products. Adoption of the Habitat Model Standard by Network programs or partners will enable the exchange and aggregation of habitat models and support the creation and sharing of standardized habitat model products to support conservation applications.

Exchange of habitat models presents unique challenges. Habitat model data come in a variety of spatial data formats (i.e., raster and vector). File sizes can be large, especially when considered in aggregate. Metadata needs may differ by modeling approach. Raster data storage presents its own challenges, requiring Enterprise GIS solutions and image serving capabilities. A full technological solution for the exchange and maintenance of habitat models is beyond the scope of this standard. However, here we provide a framework for management and exchange of species habitat models to guide design and implementation of a full technical solution.

Key needs that the framework must address include:

- Tracking the availability and currency of models
- Linking models to species information stored in NatureServe central databases
- Ingesting data from diverse sources into standardized formats
- Distributed management of spatial data
- Central management of tabular data and metadata
- Seamless provision of those data into core data products

While we emphasize the exchange of models within the NatureServe Network, we intentionally leave the door open to integration of habitat models from a variety of sources.

A conceptual diagram for the flow of species habitat model information is provided in Figure 7. Models may be developed either in cooperation with NatureServe, using shared Microsoft Azure computing infrastructure, or produced elsewhere inside or outside the Network. The data ingestion wizard will provide a guided process for cross-walking key model attributes to standard fields and ensuring data are formatted correctly for exchange. Tabular data on key model attributes (Table 1) will be stored in a central database that will eventually link to Biotics and can be used to track model availability and versioning. This database should also incorporate information from the Model Review Tool, an interactive online means of gathering expert review of range-wide models. A model confidence rating, assigned at the time of exchange and/or following review, will also be stored in this database. Spatial data, including categorical data, continuous data (if available), and model footprints will be stored either in a central spatial data store (as in the case of multi-jurisdictional models produced using NatureServe's modeling infrastructure) or on local data hubs (as in the case of large libraries of models maintained by a Network program). These data will be used to produce standard Biodiversity Location Data products via NatureServe Explorer or custom Biodiversity Location Data products.

The solutions outlined in Figure 7 and described below offer a means to aggregate habitat models from a variety of sources, ensure that their taxonomic concepts conform with those in Biotics, and convert model results into standardized products provided as part of NatureServe's Biodiversity Location Data offerings. This approach will fill in holes in our current data offerings by including information on potential habitat alongside data on documented occurrence. This puts NatureServe and the Network in a better position to provide decision-makers with the most current, complete, and consistent biodiversity data.

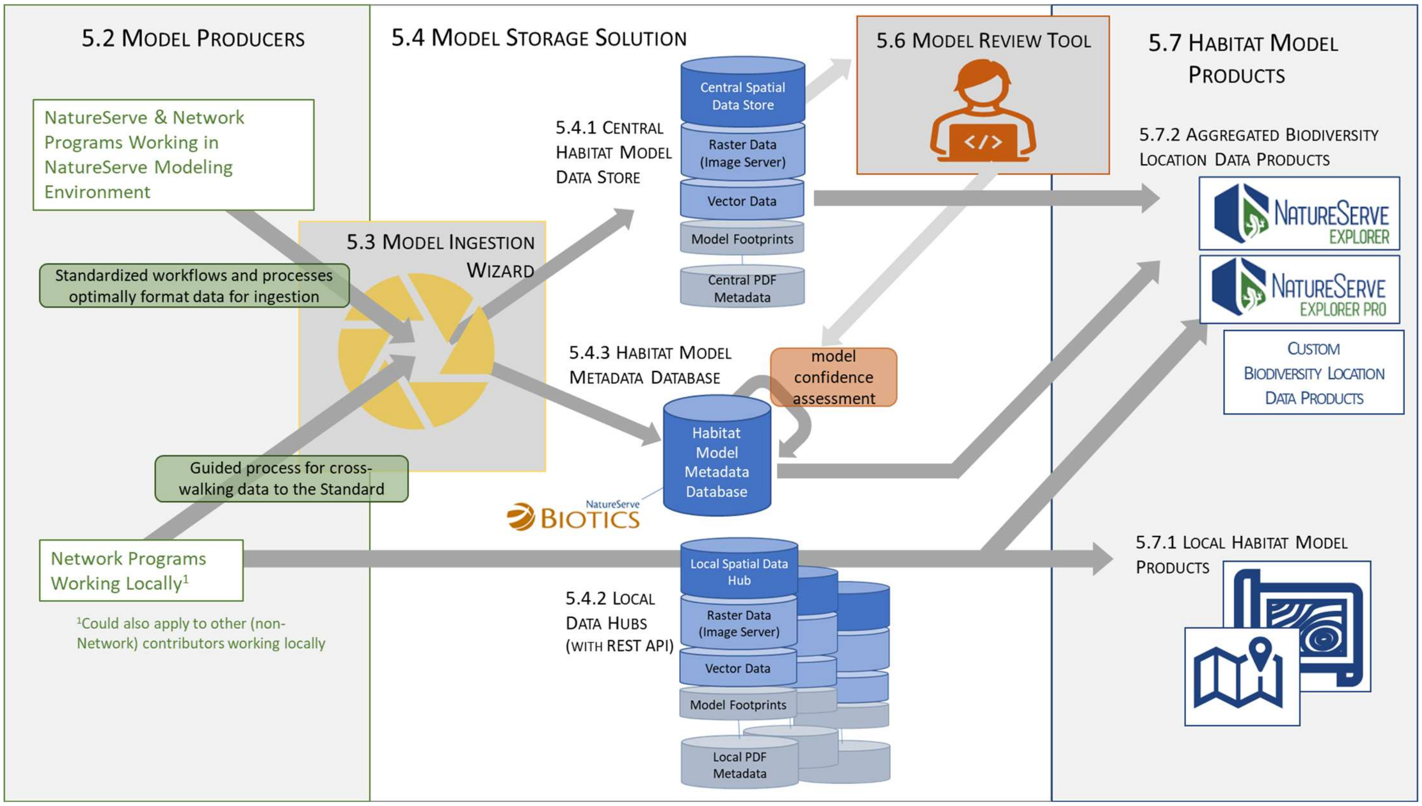


**Figure 7.** *Flow of habitat model data from Network programs and external sources to aggregated, multi-jurisdictional model products. More information on each numbered step is provided below.*

## 5.2 Model Producers

Models included in NatureServe's Biodiversity Location Data may come from a variety of sources. Many range-wide habitat models are being produced centrally at NatureServe, working in close collaboration with Network Programs. NatureServe's Microsoft Azure habitat modeling hub supports the co-development of these models by NatureServe and the Network and facilitates the production and storage of model products. Model development procedures and centralized model storage in this environment will be optimally designed to support the provision of models in a standardized format.

The format of models produced elsewhere in the Network is largely left to the discretion of programs engaged in modeling. However, to be integrated into NatureServe Biodiversity Location Data products, models must be able to be cross-walked to this standard. The data ingestion wizard will be designed to facilitate that process. Network habitat modeling best practices and scripted procedures shared via the Heritage Network Regional Habitat Model GitHub repository are available to help programs ensure that data they produce can be readily shared. The ingestion wizard will capture key model metadata, but spatial data could continue to be stored through a local data hub.

Models developed outside of the Network could potentially be incorporated through a similar process. This includes models from academic collaborators using Network data.

## 5.3. Model Ingestion Wizard

The model ingestion wizard will be used to register models in a centralized catalog (i.e. the Habitat Model Metadata Database) and enforce/assist in fitting models to the Network Habitat Model Standard. The process will include capturing pathways to distant services from which spatial data and other model artifacts can be retrieved, when appropriate, via REST APIs.

The production pathway for models produced by NatureServe and Network Programs using NatureServe's Azure modeling environment will ensure that model products (spatial data, pdf metadata, and a JSON file containing the information listed in Table 1) are optimally formatted for intake by the ingestion wizard and then upload them into central data stores.

For models produced elsewhere in the Network, model contributors will be guided through the cross-walking of their model products and metadata to standardized fields through an online form and/or customized tools designed to support batched uploads of multiple models. Similar pathways for model sharing from other non-Network sources could be developed as need arises.

## 5.4 Model Storage Solution

### 5.4.1 Central Habitat Model Store

The central habitat model storage solution will serve as a repository of spatial data associated with habitat models produced at NatureServe in collaboration with the Network. Continuous and categorical raster products will be managed as image services, where models for individual species are maintained as tiles. Vector products will be similarly managed, most likely in a geodatabase. Model footprints, and pdf metadata, provided at the time of model ingestion, can also be centrally managed.

### 5.4.2 Local Data Hubs

Local data hubs are spatial data stores managed by local programs. It is envisioned that these may be configured similarly to the Central Habitat Model Store, or otherwise designed to meet needs of local programs, so long at the configuration enables ingestion and resharing of model data through the model ingestion process by way of REST APIs. Map or image service API endpoints, referenced in the centrally-stored metadata provided through the model ingestion process, will facilitate central access to spatial data, and other model artifacts, stored locally. Alternately, model contributors could stand up API endpoints that themselves serve as model catalogs, rather than storing metadata information centrally, so long as those endpoints conform to specifications set by NatureServe.

Programs interested in developing local data hubs should work with NatureServe to determine the optimal design and implementation solutions for configuring hubs to meets local needs while ensuring efficient information sharing.

### 5.4.3 Habitat Model Metadata Database

The database will store all the information provided in Table 1, as well as the assigned model confidence scores and paths to the spatial data and PDF metadata for models stored both in the central habitat model database and on local spatial data hubs (possible exception: data provided via APIs from local data hubs could enable local, rather than central, storage of model metadata information - TBD). Note that the Habitat Model Metadata Database refers to the system for storing model attributes. PDF metadata about each model will be stored separately in the central or local data hubs.

The central habitat model database would optimally be a component of Biotics but could initially be configured as a separate database that links to Biotics through enforcement of taxonomic concepts that conform with those in Biotics.

## 5.5 The Model Review Tool

The Model Review Tool is an online application developed by NatureServe through which species habitat models can be reviewed by species experts inside or outside the Network within a secure online interface. The Model Review Tool facilitates broad Network collaboration in improving model quality by leveraging the extensive knowledge of Network botanists and zoologists. Feedback received via the review tool can be used to assign model confidence scores.

Currently, the Model Review Tool exists as a stand-alone application that calls data from an independent image service and stores data collected via the review process in ArcGIS Online. However, our vision is that in the long-term, the Model Review Tool could reference data stored in NatureServe's Central Habitat Model Store. Data collected via the review process, which is currently stored in ArcGIS Online, could be linked to the central model database and directly referenced in the assignment of model confidence scores. There is also the possibility of configuring the Model Review Tool, or a version there-of, to support expert review of habitat models stored on local data hubs.

Currently however, expert model review information, including that collected via the Model Review Tool, must be summarized by the model contributor and provided as model attributes as specified in Table 1.

## 5.6 Habitat Model Products

### 5.6.1 Local Model Products

Habitat models produced at the state, provincial, or other sub-national levels are commonly used locally in Environmental Review Tools, field guides, local conservation prioritizations, and to guide inventory. The guidelines provided in this standard support the production and application of local habitat models by providing a common language for Network habitat model products and recommendations for communication of confidence and appropriate end use. Network Programs may also find that embracing the standard will introduce efficiencies in model development and provision, as Network tools and resources, such as default configurations for Environmental Review Tools, the Model Review Tool, and habitat modeling code maintained in a Network github repository, are all likely to become aligned to the new standard.

### 5.6.2 Aggregated Biodiversity Location Data Products

The Species Habitat Model Standard also provides the foundation necessary for NatureServe to add habitat models to the multi-jurisdictional Biodiversity Location Data products we provide. Including habitat model products in NatureServe's standardized Biodiversity Location Data enables the NatureServe Network to provide valuable information on potential species habitat that complements the documented occurrence data (i.e., element occurrences and observations) for which the Network is well known. We envision that species habitat models will be served via NatureServe Explorer, NatureServe Explorer Pro, and provided in response to custom multi-jurisdictional data requests as representations of "likely habitat" alongside documented occurrence data. Model metadata, information on model confidence, and guidance on how models should be interpreted and applied will accompany these spatial data products. Providing these habitat models will increase the level of support the Network can provide to decision-makers seeking data to direct field inventories, guide land management, assess environmental risks, or pursue other uses as indicated on Table 2.

Sharing of habitat models provided by the Network, or developed using Network data, will be governed by Network data sharing permissions. Recognizing that habitat predictions are generally less sensitive than occurrence data, NatureServe's data permission form will allow data providers to set permissions for sharing of model products separately from those for sharing occurrence data. Enforcing stricter restrictions for species susceptible to persecution and harm will be possible.

Models provided via NatureServe Explorer/Explorer Pro, especially those provided as open data (i.e. data that is publicly available, with specified terms of use but no signed license agreement), may require upscaling the habitat predictions to a common spatial data framework (e.g. hexagons). Where multiple models are available for the same species, rules for selecting preferred models or creating ensemble models will need to be developed (see Section 4). Approaches for doing so, as well as derivative habitat model products (for example, products that combine information on predicted habitat with information on habitat condition) are not further addressed in this standard but will be addressed during further development of Explorer/Explorer-Pro, with opportunities for Network feedback.

# 6. CONCLUSIONS

This Species Habitat Model Standard focuses explicitly on spatial data and metadata requirements and means for assessing and communicating confidence in species habitat model products. It also provides a framework for data management of those models. By design, the standard is not prescriptive regarding how models are developed, leaving the specifics of inductive or deductive modeling methods up to the data provider. Instead, our goal is to provide criteria for ensuring that end products are consistent, interpretable, and of sufficient scientific integrity. For further guidance on methods for model development, see the NatureServe Network Species Habitat Modeling Best Practices Wiki (*under development)*.

Species habitat models are a valuable source of information for conservation decision-making and complement the data on documented locations of at-risk species for which the NatureServe Network has long been known. Given the maturation of habitat modeling as a science, the increasing availability of habitat model products from across the Network, and a call for greater use of standardized model products by land managers and other conservation decision-makers (Sofaer et al. 2019, Aruajo et al. 2019), NatureServe recognizes that it is uniquely positioned to channel this collective momentum into a cohesive resource for standardized habitat model products. This standard, by providing the means to collect and aggregate habitat models, will ensure the relevance of NatureServe Network data into the future and maximize our contributions to the conservation of imperiled species. The NatureServe Network has been producing species habitat models for almost two decades; with this standard we are now positioned to provide those data to conservation practitioners in a consistent and comprehensive manner. This standard will enable us to:

- Supplement the information on documented species occurrences for which the NatureServe Network has long been known, filling in information on the presence or absence of likely habitat in areas that have not been comprehensively surveyed.

- Make existing models more readily accessible, so that they are discoverable and available to guide inventory, contribute to conservation prioritizations, and otherwise support conservation efforts.
- Assist end users in interpreting and applying habitat models, which are often complex and difficult to interpret, by communicating consistent model products and information on model confidence and appropriate end uses.

# 7. REFERENCES


Araújo MB, New M. 2007. Ensemble forecasting of species distributions. Trends in Ecology and Evolution 22: 42–47.

Araujo M, Anderson R, Barbosa A, Beale C, Dormann C, Early R, Garcia R, Guisan A, Maiorana L, Naimi B, O'Hara R, Zimmerman N, Rahbek C. 2019. Standards for distribution models in biodiversity assessments. Science Advances 5:1-11.

Aycrigg J, Beauvais G, Gotthardt T, Huettmann F, Pyare S, Andersen M, Keinath D, Lonneker J, Spathelf M, and Walton K. 2015. Novel Approaches to Modeling and Mapping Terrestrial Vertebrate Occurrence in the Northwest and Alaska: An Evaluation, Northwest Science 89(4), 355-381.

Baker DJ, Maclean IMD, Goodall M, and Gaston KJ. 2020. Species distribution modelling is needed to support ecological impact assessments. Journal of Applied Ecology 58: 21-26.

Barbet-Massin M, Jiguet F, Albert CH, Thuiller W. 2012. Selecting pseudoabsences for species distribution models: How, where, and how many? Methods in Ecology and Evolution 3: 327–338.

Boykin KG, Drost C, and Wynne JJ. A gap analysis of terrestrial vertebrate species of the Colorado Plateau: Assessment from the Southwest Regional Gap Analysis Project. The Colorado Plateau III: Integrating research and resources management for effective conservation. University of Arizona Press, Tucson (2008): 77-107.

Dorazio RM. 2014. Imperfect detection and survey bias in presence-only data. Global Ecology and Biogeography 23: 1472-1484.

Dormann CF. 2007. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. Global Ecology and Biogeography 16: 129–138.

Dormann CF, et al. 2013. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. Ecography 36: 027–046.

Endangered and Threatened Wildlife and Plants; Regulations for Listing Endangered and Threatened Species and Designating Critical Habitat. 2020. Vol. 85, No. 242 Fed. Reg. Page 81411.

Fithian W, Elith J, Hastie T, and Keith DA. 2015. Bias correction in species distribution models: pooling survey and collection data for multiple species. Methods in Ecology and Evolution 6: 424-438.

Fourcade Y, Besnard AG, Secondi J. 2018. Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. Global Ecology and Biogeography 27: 245–256.

Graham CH, Elith J, Hijmans RJ, Guisan A, Peterson AT, Loiselle BA. 2008. The influence of spatial errors in species occurrence data used in distribution models. Journal of Applied Ecology 45: 239–247.

Guillera-Arroita G, Lahoz-Monfort JJ, Elith J, Gordon A, Kujala H, Lentini PE, McCarthy MA, Tingley R, Wintle BA. 2015. Is my species distribution model fit for purpose? Matching data and models to applications. Global Ecology and Biogeography 24: 276–292.

Guisan A, Zimmermann NE. 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135: 147–186.

Hamilton H, Smyth R, Young BY, Howard TG, Tracey C, Breyer S, Cameron DR, Chazal A, Conley AK, Frye C and Schloss C. Increasing taxonomic diversity and spatial resolution clarifies opportunities for protecting imperiled species in the US. Ecological Applications*, In press.*

Hallgren W, Santana F, Low-Choy S, Zhao Y, and Mackey B. 2019. Species distribution models can be highly sensitive to algorithm configuration. Ecological Modelling 408:108719.

Kéry M. 2011. Towards the modelling of true species distributions. Journal of Biogeography 38: 617-618.

Lahoz-Monfort JJ, Guillera-Arroita G, and Wintle BA. 2014. Imperfect detection impacts species distribution models. Global Ecology and Biogeography 23: 504-515.

Lozier JD, Aniello P, Hickerson MJ. 2009. Predicting the distribution of Sasquatch in western North America: Anything goes with ecological niche modelling. Journal of Biogeography 36: 1623–1627.

Liu C, Newell G, White M. 2016. On the selection of thresholds for predicting species occurrence with presence-only data. Ecology and Evolution 6: 337–348.

Liu CR, Berry PM, Dawson TP, Pearson RG. 2005. Selecting thresholds of occurrence in the prediction of species distributions. Ecography 28: 385–393.

Moskwik M, Mainali k, Pavelka M, Nicolaysen T, Juliusson L, Chhetri D, Shultz M. 2019. Standard Operating Procedure: USFWS Refined Range Maps for Threatened and Endangered Species v3.0. U.S. FWS Ecological Services.

Owens HL, Campbell LP, Dornak LL, Saupe EE, Barve N, Soberón J, Ingenloff K, Lira-Noriega A, Hensz CM, Myers CE, Peterson AT. 2013. Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. Ecological Modelling 263: 10–18.

Petitpierre B, Broennimann O, Kueffer C, Daehler C, Guisan A. 2017. Selecting predictors to maximize the transferability of species distribution models: Lessons from cross-continental plant invasions. Global Ecology and Biogeography 26: 275–287.

Peterson A, Soberón J,Pearson R,Anderson R, Martínez-Meyer E, Nakamura M, and Araújo M. 2011. Ecological Niches and Geographic Distributions. 10.1515/9781400840670.

Phillips SJ, Dudik M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S. 2009. Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. Ecological Applications 19: 181–197.

Qiao H, Soberón J, Peterson AT. 2015. No silver bullets in correlative ecological niche modelling: Insights from testing among many potential algorithms for niche estimation. Methods in Ecology and Evolution 6: 1126–1136.

Roberts DR, et al. 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40: 913–929.

Roubicek AJ, VanDerWal J, Beaumont LJ, Pitman AJ, Wilson P, Hughes L. 2010. Does the choice of climate baseline matter in ecological niche modelling? Ecological Modelling 221: 2280–2286.

Soberón J and Nakamura M. 2009. Niches and distributional areas: Concepts, methods, and assumptions. Proceedings of the National Academy of Sciences 106 (Supplement 2) 19644-19650.

Soberón J and Peterson A. 2005. Interpretation of Models of Fundamental Ecological Niches and Species' Distributional Areas. Biodiversity Informatics. 2. 10.17161/bi.v2i0.4.

Sofaer HR, Jarnevich CS, Pearse IS, Smyth RL, Auer S, Cook GL, Edwards TE, Guala GF, Howard TG, Morisette JT, and Hamilton H. 2019. Development and delivery of species distribution models to inform decision-making. BioScience 69:544–557.

Valavi R, Guillera-Arroita G, Lahoz-Monfort J J, and Elith J. 2021. Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. *Ecological Monographs* 00( 00):e01486.

# APPENDIX I - SPECIES HABITAT STANDARD WORK GROUP

**Table 1.** Members of the Observation Data Standard Work Group. Section column refers to either the NatureServe Section Councils or NatureServe organization, where CA = Canadian Section Council, NS = NatureServe, NSC = NatureServe Canada, US = U.S. Section Council.

| Section | Organization | Last Name | First Name | Position | Contact Information |
|---|---|---|---|---|---|
| US | Wyoming Natural Diversity Database | Beauvais | Gary | Director | beauvais@uwyo.edu |
| CAN | British Columbia Conservation Data Centre | Clare | Jacqueline | Data Management Lead | jacqueline.clare@gov.bc.ca |
| US | New York Natural Heritage Program | Conley | Amy | Spatial Ecologist | amy.conley@dec.ny.gov |
| US | NatureServe | Faber-Langendoen | Don | Senior Ecologist & Conservation Methods Coordinator | don_faber-langendoen@natureserve.org |
| US | Oklahoma Natural Heritage Inventory | Fagin | Todd | Conservation Data Analyst | tfagin@ou.edu |
| US | California Natural Diversity Database | Gogol-Prokurat | Melanie | Senior Environmental Scientist | melanie.gogol-prokurat@wildlife.ca.gov |
| US | Virginia Division of Natural Heritage | Hazler | Kirsten | Landscape Ecologist | kirsten.hazler@dcr.virginia.gov |
| US | Oregon Biodiversity Information Center | Henderson | Emilie | Modeling Lead | emilie.henderson@oregonstate.edu |
| US | New York Natural Heritage Program | Howard | Tim | Director of Science | tghoward@esf.edu |
| US | Alabama Natural Heritage Program | Lawson | Kaitlyn | GIS Analyst II | klawson@auburn.edu |
| US | South Carolina Heritage Trust | Lemeris | Joseph | GIS/Data Manager | lemerisJ@dnr.sc.gov |
| US | Montana Natural Heritage Program | Maxell | Bryce | Program Coordinator | bmaxell@mt.gov |
| US | NatureServe | McIntyre | Patrick | Senior Ecologist | patrick_mcintyre@natureserve.org |

| Section | Organization | Last Name | First Name | Position | Contact Information |
|---|---|---|---|---|---|
| US | Florida Natural Areas Inventory | Oetting | John | Chief of Conservation Planning Services | joetting@fnai.fsu.edu |
| US | NatureServe | Rapacciuolo | Giovanni | Director of Applied Science Programs | giovanni_rapacciuolo@natureserve.org |
| NS | NatureServe | Smyth* | Regan | Director of Spatial Analysis | regan_smyth@natureserve.org |
| US | Pennsylvania Natural Heritage Program | Tracey | Christopher | Conservation Planning Manager | Ctracey@paconserve.org |
| NS | NatureServe | Woo | Maggie | Senior Software Engineer | maggie_woo@natureserve.org |

*Regan Smyth was chair of the Work Group.

# APPENDIX II – CRITERIA FOR INDUCTIVE MODEL QUALITY ATTRIBUTES

For inductive models, the framework for scoring on model quality attributes is adapted from Sofaer et al. 2019 (Table 1). Similar criteria for scoring deductive models are provided in Appendix II. Note that for the purpose of collecting model metadata and scoring, we have relabeled Sofaer et al.'s "problematic" category as "interpret with caution". This reflects the idea that for some end uses with lower stakes, such as guiding inventory, practices that may be problematic in other contexts may still result in useful information.

In addition to the factors in Table II-1, this standard calls for a quality rating based on the outcomes of expert review of model outputs. Criteria for that are also included in Appendix IV.

**Table II-1.** Criteria for scoring model quality for inductive models, adapted from Sofaer et al. (2019). We expand on each criterion after the formatted table.

| | | Interpret with Caution | Acceptable | Ideal |
|---|---|---|---|---|
| Training and Validation Data | Presence data | Poor or unassessed quality of data (precision, accuracy, sampling bias, taxonomy) | Systematic spatial inaccuracies avoided or corrected. Spatial precision matches modeling scale. Samples span species' range; clustered samples handled appropriately. No taxonomic inconsistencies | High spatial accuracy and precision. Unbiased samples across species' range. No taxonomic inconsistencies. |
| | Absence/ background data | Background samples collected in statistically inappropriate manner. Background data cover inappropriate extent. | Background samples mimic presence sampling biases, sampling is weighted, and/or sensitivity analyses evaluate effects of different background datasets. | Design-based sampling of both presence and absence locations, or presence/absence data sets combined in statistically compatible manner. |
| | Evaluation data | Based on training data. | Based on spatial cross-validation of training data. | Based on independent data from separate sampling effort. |
| Environmental Predictors | Predictor relevance | Arbitrary set of predictors. Multicollinearity ignored. | Selection of predictors justified based on natural history. Multicollinearity handled appropriately. | Predictors represent factors known to govern distributional limits or are direct signals of species presence. Multicollinearity is eliminated or handled in an appropriate manner, |
| | Spatio-temporal data alignment | Spatio-temporal alignment of species data and predictors not considered or verified. | Predictor variables sufficiently approximate conditions at the time and place of the observations used to train the model | Predictor variables accurately reflect conditions at the time and place of the observations used to train the model. |

| | | | | |
|---|---|---|---|---|
| Modeling Process | Algorithm choice & Statistical Rigor | Models prone to overfitting; uncertain suitability for prediction; modeling assumptions not met | Algorithm suitable for prediction; modeling assumptions generally met | Algorithm(s) highly suitable for prediction; multiple algorithms evaluated; model assumptions met |
| | Sensitivity Analysis | Single algorithm used with default settings; no sensitivity analysis | Sensitivity analysis for algorithms, tuning parameters, and/or input data | Multiple algorithms, with evaluation of tuning parameters and input data sampling protocols; model agreement and uncertainties evaluated with ensemble techniques. |
| | Performance Evaluation | Single performance metric, and/or evaluation scores below generally accepted levels. | Multiple performance metrics; evaluation scores close to generally accepted levels; ecological plausibility considered. | Multiple performance metrics; evaluation scores at or above generally accepted levels; scores connected with intended use; ecological plausibility supported |
| | Model Review Status | No review, poor review, or unreliable review | Review by qualified regional and taxonomic experts; moderate review score; reviews inform model use and/or plans for updates | Comprehensive review by qualified regional and taxonomic experts; high review score or moderate score followed by update; transparency of review outcomes and influences on model |
| | Iterative | No. | Updated based on expert review and other performance assessments. Not updated based on new field observations. | Updated via targeted field sampling and incorporation of new field data into subsequent model iterations. |
| Model Products | Map products | Value ranges not interpreted For classified maps, single default threshold for all applications. Use of 0.5 as a threshold for poorly calibrated models. | Value ranges interpreted. For classified maps, thresholds based on test data but not necessarily linked to intended use. | Value ranges interpreted. For classified maps, thresholds based on intended use and model assessment, with exploration of sensitivity. Uncertainties mapped. |
| | Supplementary Metadata | No supplementary metadata or inadequate metadata to assess key decisions. Little or no description of predictor variables or methods. | Enough information to evaluate every row in this table. Where explanation is a goal, description of included variables and their importance. | Information to easily evaluate every row in this table. Where explanation is a goal, description of predictor variable importance and estimated relationships to response for focal variables. Engagement with user community to help define objectives, guide the development and interpret results. |
| | Reproducibility | Inputs not saved/published, settings from modeling GUI not saved or code not annotated and saved. | Inputs saved and made available (excepting locations of rare species), scripts, settings, and model results archived. | Inputs saved and made available (excepting locations of rare species). Scripts, settings, model results archived. Species expert and modeler identity known. |

## Training and validation data

**Presence data**

An assessment of presence data should consider taxonomic accuracy as well as locational precision and accuracy (Graham et al. 2008, Lozier et al. 2009).

**Interpret with Caution**

Precision of records is unrecorded and/or very low relative to the modeling scale (e.g., position represented only by county centroids, or too few significant digits for geographic coordinates). Accuracy of records is unverified or very poor (e.g., systematic spatial shifts due to incorrect specification of coordinate system or misregistration of spatial data). Spatial bias of samples is unacknowledged or unaddressed; records provide poor representation of environmental variability across the species' range. Taxonomic identification is unreliable, or records have higher-level taxonomic identification than intended for the model (e.g., genus or species instead of the targeted species or subspecies, respectively).

**Acceptable**

Most records have documented precision that is commensurate with the modeling scale; lower precision records are identified and handled with a sampling methodology that applies weights according to precision. Accuracy issues due to mis-specified coordinate systems or spatial misregistration have been avoided or corrected. Bias arising from spatial clustering of samples is handled appropriately (e.g. with spatial thinning or subsampling), and samples adequately capture environmental variability across the species' range. Taxonomic identification of samples is reliable and consistent with the taxonomic entity being modeled.

**Ideal**

The vast majority of records are verified accurate and precise, and any lower-precision records are handled appropriately. Records are from unbiased samples across the species' range. Taxonomic inconsistencies have been avoided or corrected.

**Absence/Background data**

An assessment of background input data should consider their spatial relationships to the input presence locations. (Barbet-Massin et al. 2012, Guillera-Arroita et al. 2015, Phillips et al. 2009)

**Interpret with Caution**

Background samples include locations that are coincident with, or very close to, presence locations. Background samples do not reflect sampling bias in presence locations. Background data cover an inappropriate extent, either extending far beyond the known species' range, or restricted to a subset of the known range.

**Acceptable**

Background samples have been filtered to avoid close proximity to presence locations. Background samples mimic sampling biases in presence data, background sampling is strategically weighted with the number of presence locations (Barbet-Massin et al. 2021), and/or sensitivity analyses are conducted to evaluate effects of different background datasets.

**Ideal**

Absence and presence data were collected via a statistically unbiased design-based protocol, or presence and absence datasets were combined in a statistically compatible manner. Sampling may require explicit modeling of detection biases.

**Evaluation data**

Model evaluation is an important part of model development, and evaluation results convey important information about the model to which all users of the model should have access. (Fourcade et al. 2018, Roberts et al. 2017)

**Interpret with Caution**

The training data are used to evaluate the success of the model.

**Acceptable**

Cross-validation (jackknife, leave-one-out, or other cross-validation) of the training data is used to assess the performance of the model. Cross-validation groups must be separated spatially, such that training and test data do not represent the same spatial occurrences.

**Ideal**

Model evaluation is based on independent data from a separate sampling effort. New data are collected, usually with a sampling design based on the original model, and results from this sampling effort are used to evaluate the model.

## Environmental Predictors

**Predictor relevance**

Taxonomic experts cannot be expected to know, *a-priori*, all ecological drivers of distribution for every species. However, if habitat preferences are generally known, relevant species-specific predictors should be included to the extent possible. Predictors with unknown ecological relevance may also be useful for modeling, but care needs to be taken when including them, especially with algorithms that don't handle co-varying predictors well. (Fourcade et al. 2018, Guisan and Zimmermann 2000, Petitpierre et al. 2017)

**Interpret with Caution**

An arbitrary set of predictors is used, with no tuning to remove correlated predictors or predictors of low statistical importance. Predictors representing environmental factors that are most important to the species are not included.

**Acceptable**

Predictors are selected based on known or suspected importance to the taxon, based on natural history characteristics. Multicollinearity is handled in an appropriate manner, and predictors without predictive power are eliminated.

**Ideal**

Predictors are selected based on demonstrated predictive power from previous research. Predictors represent mechanistic drivers of distributional limits or direct signals of species presence (e.g., remotely sensed indices). Multicollinearity is eliminated or handled in an appropriate manner,

**Spatio-temporal data alignment**

Landscapes and habitats change over space and time, and predictor variables vary in scale and accuracy. Very old observations may have occurred in habitats no longer present, and excellent quality observations are of no use if they occur in places not covered by the predictor variables. (Roubicek et al. 2010)

**Interpret with Caution**

The values of predictor variables do not accurately reflect conditions at the time and/or place of the observations used to train the model, or the spatio-temporal alignment of species data and predictors was not considered. Time-varying predictor variables reflect conditions at a single point in time that does not necessarily match the time at which species observations were made. Spatial data may be shifted or misaligned, and/or of low resolution relative to training data.

**Acceptable**

The values of predictor variables sufficiently approximate conditions at the time and place of the observations used to train the model. If applicable, time-varying predictor variables have been developed for a range of time periods, at a moderate temporal scale, so that they can be approximately matched to observation data collected at different times. Spatial data are aligned correctly and of moderate resolution relative to training data.

**Ideal**

The values of predictor variables accurately reflect conditions at the time and place of the observations used to train the model. If applicable, time-varying predictor variables have been developed for a range of time periods, at a fine temporal scale, so that they can be matched to observation data collected at different times. Spatial data are aligned correctly and of sufficiently high resolution relative to training data.

## Modeling Process

### Algorithm choice (& Statistical Rigor)

Model structures and algorithms vary considerably in the kinds of relationships they can model, the types of assumptions they can accommodate, and the types of user applications for which they are most suited (Qiao et al. 2015). Different statistical approaches carry different assumptions about the input data and methodologies. (Dormann 2007, Dormann et al. 2013)

**Interpret with Caution**

Models prone to overfitting are used for extrapolation to new areas or new times. Goals of prediction versus explanation are confounded. Input data do not meet fundamental assumptions of the modeling approach, or assumptions are not recognized or evaluated.

**Acceptable**

Selection of an algorithm is aligned with the objectives for prediction. Model assumptions recognized and considered; input data do not deviate strongly from the assumptions of the modeling approach.

**Ideal**

Multiple algorithms are evaluated, and selection of one or more algorithms is aligned with the objectives for prediction. Model assumptions formally evaluated; input data meet the assumptions of the modeling approach.

**Sensitivity analysis**

Model outputs may change considerably with only small alterations to inputs or settings, or model outputs may be robust to changes in initial states and settings. (Araújo and New 2007), Hallgren et al. 2019)

**Interpret with Caution**

A single algorithm is used with default settings. No sensitivity analysis is conducted.

**Acceptable**

Multiple algorithms, tuning parameter settings, and/or input data sampling protocols are evaluated, and adjustments made to optimize results.

**Ideal**

The modeling process includes multiple algorithms, adjustment of tuning parameters, and optimization of input data sampling protocols and all settings are reported in metadata. Algorithms are tuned for best results. Model agreement and uncertainties are evaluated with ensemble techniques and the level of uncertainty and model sensitivity is reported. High levels of uncertainty and sensitivity might cause a reduction to "Acceptable."

**Performance**

The performance of the model can be evaluated through a variety of approaches and metrics. (Jarnevich et al. 2015)

**Interpret with Caution**

Evaluation of performance is based on a single metric, and/or evaluation scores are below generally accepted levels.

**Acceptable**

Evaluation of performance is based on multiple metrics, evaluation scores are close to generally accepted levels, and ecological plausibility is considered.

**Ideal**

Evaluation of performance is based on multiple metrics, and evaluation scores are at or above generally accepted levels. Scores connected with implications for intended use are considered. Ecological plausibility is described and supported with data or references.

*Note that model providers are also be asked to submit additional details on performance measures and outcomes (see Table 1).*

**Model Review Process**

Part of the modeling process can include a review phase where taxon experts evaluate the model and output (map) products. (Guisan et al. 2013)

**Interpret with Caution**

The model was released without review, was reviewed by individuals having potential conflicts of interest or little expertise, or received a low review score (e.g., average of 1-2 on a 5-point scale) with no attempt to rectify issues identified by reviewers.

**Acceptable**

The model was reviewed across at least some of the species' range by 1-2 regional and taxonomic experts, (with no conflicts of interest), and received at least a moderate review score (e.g., average of at least 3 on a 5-point scale). Reviewer comments inform recommendations for model use and/or plans for model updates.

**Ideal**

The model was reviewed across most of the species' range by 3 or more regional and taxonomic experts (with no conflicts of interest). The model received a high review score (e.g., 4-5 on a 5-point scale), or received a moderate review score and was then updated in iterative fashion based on reviewer comments and/or additional field data. Review outcomes and their influence on model updates are transparent.

*Note that model providers are also be asked to submit details on model review* outcomes *(see Table 1)*. *For criteria for scoring Model Review Outcom*es *see Appendix IV.*

**Iterative**

The modeling process can be iterative, in that it is re-run after feedback on review, with new information, or in response to other alterations of inputs or settings. Modelers should provide information on whether the model was refined in an iterative fashion (a) based on iterative development by modeler(s) (e.g., via variable selection), (b) to address comments received during the review process, or (c) with newly available data. This information will not directly be used in the confidence evaluation but is of value for communicating about model status.

**Interpret with Caution**

No iteration.

**Acceptable**

Updated based on expert review and other performance assessments.

**Ideal**

Updated via targeted field sampling and incorporation of new field data into subsequent model iterations.

## Model Products

**Map Products**

The type of outputs can vary based on user need; in addition, the amount and type of information released with a model influences its use and interpretability. (Guillera-Arroita et al. 2015, Liu et al. 2016, Liu et al. 2005, Owens et al. 2013)

**Interpret with Caution**

A binary, classified, or continuous map is produced without a clear description to help users interpret ranges of values. If a thresholded map is produced, a single default threshold is used for all applications. A default threshold of 0.5 is used for poorly calibrated models.

**Acceptable**

A continuous map is accompanied by a clear description to help users interpret ranges of values. Thresholds for classified maps are set based on test data (e.g., sensitivity equals specificity), although not necessarily linked to intended uses.

**Ideal**

A continuous map is accompanied by a clear description to help users interpret ranges of values, and/or is used as basis for derived products (e.g. sampling design). Thresholds for classified maps are selected based on intended use and model assessment, with exploration of sensitivity. A map representing level of uncertainty is included.

### Supplementary Information (Interpretation Guidance)

Supplementary metadata is necessary to ensure that model products can be correctly understood and interpreted.

**Interpret with Caution**

Insufficient information is provided for the user to assess key decisions. There is little or no description of methods, and predictor variables are not listed or are poorly described.

**Acceptable**

Enough information is provided to evaluate every row in this table. Methods are adequately explained, and predictor variables itemized and briefly described. Where explanation is a goal, more detailed descriptions of the most important variables may be provided.

**Ideal**

Detailed information is provided, making it easy to evaluate every row in this table. Methods are thoroughly explained, and predictor variables itemized and described. Where explanation is a goal, focal variables and their estimated effects on the likelihood of species occurrence are described in more detail. Modelers actively engage with user community to help define model objectives, guide the development, and interpret results.

### Reproducibility

Model results should be transparent and reproduceable.

**Interpret with Caution**

Modeling inputs were not saved/published. If a modeling GUI was used, specified settings for the model run were not documented. If modeling scripts were used, the code not adequately annotated or not saved.

**Acceptable**

Modeling inputs were saved and made available to users (excepting locations of rare species). Scripts, settings, and model results were archived.

**Ideal**

Modeling inputs were saved and made available to users (excepting locations of rare species). Scripts, settings, model results were archived. Species experts and modelers are identified.

**APPENDIX III – CRITERIAL FOR DEDUCTIVE MODEL QUALITY ATTRIBUTES**

In some cases, predictive habitat models are developed by mapping specific land cover types or habitat features known to represent habitat for a species. These deductive models are based on expert knowledge rather than statistical analysis. Deductive modeling can produce useful estimations of habitat for many species, including those that are data-poor and thus ill-suited to inductive models, highly mobile generalist species, or species tied closely to a specific land cover type or feature that is well-mapped. Deductive models may produce a binary output (habitat, non-habitat), or a categorical output (low, med, high suitability) based on an expert opinion habitat suitability ranking. For deductive models, the framework for scoring on model quality attributes was developed by the Habitat Model Standard Work Group and was modeled after the framework for inductive models adapted from Sofaer et al. 2019 (Table 1). Criteria for the scoring of each quality factor are provided below.

In addition to these criteria, this standard calls for a quality rating based on the outcomes of expert review of model outputs. Criteria for that are also included in Appendix IV.

**Training and validation data**

**Presence data**

Deductive models may be appropriate when there are too few species occurrence locations to develop a statistical model, or when the existing species occurrences are not a representative sample of the habitats used by the species across its range. Metadata comments should describe why the available presence data are best suited for a deductive model.

**Interpret with Caution**

There are no available occurrence locations for the species, or existing occurrence locations are old and/or spatially imprecise. Deductive model is based primarily on written accounts of the habitats and locations used by the species.

**Acceptable**

Sufficient presence data exists to determine the range/extent of potential habitat for the species, although presence locations may not be representative of species habitat use. A combination of presence locations and written accounts of species habitat use allow mapping of potential habitat across the species range.

**Ideal**

Consistently accurate records of species occurrence across the range of habitat types or features used by the species are available to determine both the type and extent of potential habitat.

**Evaluation data**

Model evaluation is an important part of model development and evaluation results convey important information about the model that all users of the model should have access to. (Fourcade et al. 2018, Roberts et al. 2017) Statistical model validation can be performed on deductive models when sufficient data are available, though often the ability of modelers to do so will be limited if sample sizes are small and/or deductive techniques require use of all available presence data.

**Interpret with Caution**

No model evaluation was conducted.

**Acceptable**

Qualitative model evaluation was done by overlaying species data such as species range or occurrence locations on the deductive model and addressing any inconsistencies.

**Ideal**

A statistical evaluation of the deductive model results was conducted, or a qualitative evaluation was conducted based on independent data from separate sampling effort. For the latter, additional field work was conducted and results from this sampling efforts were used to evaluate the model.

## Environmental Predictors

### Predictor relevance

For deductive models, *a-priori* knowledge of species habitat preferences is required. This may be based on expert knowledge or written descriptions of species habitat use, or on scientific studies of species habitat requirements.

**Interpret with Caution**

Species habitat requirements are poorly known and are based on limited written descriptions from only one or two sources. -OR- Species habitat requirements are adequately known but spatial data representing those environmental variables are of poor quality or coarse resolution.

**Acceptable**

Species habitat requirements are based on written descriptions and expert knowledge from multiple sources and spatial data mapping approximating those environmental factors are of high quality (precise and accurate).

**Ideal**

Species habitat requirements are based on scientific studies of species habitat use and spatial data mapping those specific environmental factors are of high quality (precise and accurate).

### Spatio-temporal data alignment

The landscape and habitats change over space and time. Very old observations may have occurred in habitats no longer present, and excellent quality observations are of no use if they occur in places not covered by the predictor variables (Roubicek et al. 2010). Deductive models are generally produced by selecting habitat types or features for a species based on written descriptions and expert knowledge. If species occurrence points are used to supplement this information or for model evaluation, the alignment of the species occurrence points with predictor variables may be impacted by spatio-temporal data alignment.

**Interpret with Caution**

Species occurrence points were used to identify suitable habitat types or for model evaluation, and the values of predictor variables do not accurately reflect conditions at the time and/or place of the observations or the temporal and/or spatial alignment of species data and predictors was not considered.

**Acceptable**

Species occurrence points were used to identify suitable habitat types or for model evaluation, and the values of predictor variables sufficiently approximate conditions at the time and place of the observations. OR Species occurrence points were not used in model development or evaluation, and the predictor variable accurately reflects the time period to which the model applies (e.g., the predictor sufficiently reflects present-day condition of the landscape if the model is intended to reflect present-day condition).

**Ideal**

Species occurrence points were used to identify suitable habitat types or for model evaluation, and the values of predictor variables accurately reflect conditions at the time and place of the observations used to develop or evaluate the model.

## Modeling Process

### Algorithm choice

Algorithms vary considerably in the kinds of relationships they can model, the types of assumptions they can accommodate, and the types of user needs they are most suited for. (Qiao et al. 2015) Choice of a deductive model may be based on 1) inadequate number of occurrence data points or non-representative distribution of occurrence data points, 2) tight relationship of species distribution with a well-mapped predictor variable, or 3) selection of a deductive model over an inductive model during expert review.

**Interpret with Caution**

Use of deductive model is not justified by a limited availability of occurrence points or by a tight relationship with a well-mapped predictor variable, and other model algorithms have not been explored.

**Acceptable**

Selection of algorithm aligned with objectives and choice of a deductive model is justified by a limited availability of occurrence points or by a tight relationship with a well-mapped predictor variable.

**Ideal**

Multiple algorithms evaluated and deductive model was selected by expert review.

### Model Review and Updating

Part of the modeling process can include a review phase where taxon experts evaluate the model and output (map) products. (Guisan et al. 2013)

**Interpret with Caution**

The model was released without review, was reviewed by individuals having potential conflicts of interest or little expertise, or received a low review score (e.g., average of 1-2 on a 5-point scale) with no attempt to rectify issues identified by reviewers.

**Acceptable**

The model was reviewed across at least some of the species' range by 1-2 regional and taxonomic experts, (with no conflicts of interest), and received at least a moderate review score (e.g., average of at least 3 on a 5-point scale). Reviewer comments inform recommendations for model use and/or plans for model updates.

**Ideal**

The model was reviewed across most of the species' range by 3 or more regional and taxonomic experts (with no conflicts of interest). The model received a high review score (e.g., 4-5 on a 5-point scale), or received a moderate review score and was then updated in iterative fashion based on reviewer comments and/or additional field data. Review outcomes and their influence on model updates are transparent.

*Note that model providers are also be asked to submit details on model review* outcomes *(see Table 1)*. *For criteria for scoring Model Review Outcom*es *see Appendix IV.*

**Iterative**

The modeling process can be iterative, in that it is re-run after feedback on review, with new information, or in response to other alterations of inputs or settings. Modelers should provide information on whether the model was refined in an iterative fashion (a) based on iterative development by modeler(s) (e.g., via changes to model parameters), (b) to address comments received during the review process, or (c) with newly available data. This information will not directly be used in the confidence evaluation but is of value for communicating about model status.

**Interpret with Caution**

No iteration.

**Acceptable**

Updated based on expert review and other performance assessments. Not updated based on new field observations.

**Ideal**

Updated via targeted field sampling and incorporation of new field data into subsequent model iterations.

**Performance**

The performance of the model can be evaluated through a variety of approaches and metrics, though, as noted in the evaluation data section, performance metrics for deductive models may not always be able to be calculated due to a paucity of evaluation data.

**Interpret with Caution**

An evaluation of performance has not been performed and/or evaluation scores are below generally accepted levels.

**Acceptable**

Evaluation of performance is based on multiple metrics (e.g., sensitivity, specificity, kappa), evaluation scores are close to generally accepted levels, and ecological plausibility is considered.

**Ideal**

Evaluation of performance is based on multiple metrics, and evaluation scores are at or above generally accepted levels. Scores connected with implications for intended use are considered. Ecological plausibility is described and supported with data or references.

*Note that model providers are also be asked to submit additional details on performance measures and outcomes (see Table 1).*

## Model Products

### Map Products

The type of outputs can vary based on user need but also the amount and type of information released with a model influences its use and interpretability. (Guillera-Arroita et al. 2015, Liu et al. 2016, Liu et al. 2005, Owens et al. 2013)

**Interpret with Caution**

Binary or categorical map produced without clear description to interpret the classes.

**Acceptable**

Binary or categorical map produced with a clear description to interpret the classes.

**Ideal**

Categorical predictions mapped reflecting different levels of habitat suitability and including areas of uncertainty.

### Supplementary Information (Interpretation Guidance)

The type and amount of information released with a model can influence interpretation and use.

**Interpret with Caution**

Insufficient information is provided for the user to assess key decisions. There is little or no description of methods, and predictor variables are not listed or are poorly described.

**Acceptable**

Enough information is provided to evaluate every model quality attribute. Methods are adequately explained, and any predictor variables itemized and briefly described. Where explanation is a goal, more detailed descriptions of environmental variables and their importance may be provided.

**Ideal**

Detailed information is provided, making it easy to evaluate every model quality attribute. Methods are thoroughly explained, and predictor variables itemized and described. Where explanation is a goal, focal variables and their estimated effects on the likelihood of species occurrence are described in more detail. Modelers actively engage with user community to help define model objectives, guide the development, and interpret results.

### Reproducibility

Model results should be transparent and reproduceable.

**Interpret with Caution**

Modeling inputs were not saved/published. If a modeling GUI was used, specified settings for the model run were not documented. If modeling scripts were used, the code not adequately annotated or not saved.

**Acceptable**

Modeling inputs were saved and made available to users (excepting locations of rare species). Scripts, settings, and model results were archived.

**Ideal**

Modeling inputs were saved and made available to users (excepting locations of rare species). Scripts, settings, model results were archived. Species experts and modelers are identified.

## Appendix IV – CRITERIA FOR MODEL REVIEW OUTCOMES

Model review outcomes are an additional important measure of model quality. For ease of communication, model review outcomes will be summarized into four categories: good, moderate, poor, or unreviewed. This represents a simplification of the more nuanced five-point scale available to model reviewers using NatureServe's online Model Review Tool (Table IV-1).

**Table IV-1.** Guidelines for the rating of model performance using the 5-point scale in NatureServe's Model Review Tool.

| Star Rating | Brief Meaning |
|---|---|
| ★ | **Poor** representation of habitat. Habitat for this species is unlikely to be successfully mapped using standard habitat modeling approaches given unique species traits. |
| ★★ | **Significant concerns**. A large proportion of the map shows habitat in areas where the species is unlikely to occur or does not predict habitat where the species is known. Revisions are needed before the model is used for any formal application or decision, with the possible exception of guiding inventory. |
| ★★★ | **Some concerns** about model performance in specific areas. The model would benefit from additional refinement, but the general pattern of mapped habitat is consistent with expert expectations. |
| ★★★★ | Model **generally good.** Potential for further improvement through additional iteration but provides a good approximation of likely habitat. |
| ★★★★★ | Modeled habitat is a **very good** representation of likely habitat. Further iteration is unlikely to result in significant improvements. |

Model review scores can be indicated by those submitting models, or, preferably (for documentation and consistency purposes), obtained via NatureServe's Model Review Tool. For unreviewed or incompletely reviewed models, NatureServe may pursue acquisition of model reviews using the online Review Tool so as to improve confidence ratings. Information about review status and outcomes would then be updated in central databases.

Criteria for scoring model review determinations is as follows:

### Model Review Outcomes

#### Good

The mean model review score is high (4-5 star on five-point system) and the model has been assessed across entire species range. Reviews by species experts indicate that while it may be possible to further improve the model with additional iteration, the model serves as a good approximation of habitat and is appropriate for use for a variety of applications.

#### Moderate

The mean model review score is high, but the model has not comprehensively assessed across the entire range of the species OR the mean model review score is moderate (2.5-3.9 on five-point scale) and the species has been assessed across majority of range. Reviewers have some concerns about how the model performs in certain areas and the model would benefit from additional refinement, but the general pattern of mapped habitat is consistent with expert expectations.

**Poor**

The mean model review school is low (average of 1-2 on five-point scale). The predicted habitat map has significant problems. May be suitable for use for certain applications (e.g., guiding inventory) but should not be used as a general representation of areas of likely/unlikely habitat.

**Unreviewed**

Has not been reviewed by species experts