Responsible Parties	Ecological Services Program Employees Ecological Services Program Supervisors Species Range Project Team

Authors: Matthew Moskwik, Kumar Mainali, Mark Pavelka, Tara Nicolaysen, Lara Juliusson, Dilip Chhetri, Michaelle Shultz

Peer reviewed by USFWS staff, Science Applications staff, USGS staff

Approver: Martha BalisLarsen

STANDARD OPERATING PROCEDURE

USFWS REFINED RANGE MAPS FOR THREATENED AND ENDANGERED SPECIES

1.0   PURPOSE

The U.S. Fish and Wildlife Service (Service) is considered the definitive source for species listed as threatened or endangered under the Federal Endangered Species Act (ESA). As such, the Service will provide the best representation of where each listed species is currently found (see section 4.0 for definition of current range). For species that are wide ranging and have sufficient documented information about locations and environmental covariates that affect their life histories, the Service will utilize species distribution models (SDMs) to create current range maps. For narrow range species or those with insufficient data, alternative models will be used.

STANDARD OPERATING PROCEDURE

The following procedures define the parameters for creating refined range maps for ESA-listed threatened and endangered species. To maintain transparency and accountability, all data (i.e., occurrence data and gridded products of covariates) and code will be made publicly available. Final maps of current ranges will be available through the Environmental Conservation Online System (ECOS) for public viewing and download. This SOP is not a static document and will need to be altered as the models are developed and refined.

2.0   SCOPE

This SOP establishes the standards by which current range maps for listed species will be developed and assessed as part of the Refined Range Mapping project for the Service. It also lays out guidelines for how range maps that do not explicitly follow these guidelines will be evaluated and accepted as the official current range. These range maps will be uploaded to ECOS, but not used as Area of Influence (AOI) ranges for section 7 consultations. Therefore, this SOP does not address any standards for updating AOI ranges. These standards have been created by reviewing peer-reviewed published literature and using two recent reviews on the topic (Araujo et al. 2019, Sofaer et al. 2019).

There are four general methods that will be applied, depending on the quality and amount of occurrence and environmental data available and the geographic size of the range. This SOP details these methods, as well as the processes required for using species distribution models (SDM) such as:

- Obtaining/cleaning occurrence data
- Selecting environmental covariates
- Building a geographic region for range depiction
- Using that range for selecting pseudoabsence/background points where required
- Constructing and assessing models
- Building an ensemble model for a species

3.0   RESPONSIBILITIES

Creating a model, and ultimately a refined range map for any species will require close collaboration between regional leads, field biologists/species leads, GIS specialists, and Headquarters (HQ) Species Range project team. Additionally, the final map will be approved for upload to ECOS by the species lead regional office Assistant Regional Director (ARD) for Ecological Services. Roles and responsibilities for individuals or teams are as follows:

STANDARD OPERATING PROCEDURE

Regional Leads

Since the overall process involves multiple individuals and teams, regional leads are instrumental in coordination of the overall range refinement process. Overall, regional leads will work in concert with the HQ Species Range Team to coordinate species occurrence data acquisition, data validation, and collaboration with field biologists and species leads. Once the current range map is refined and ready for approval, the regional lead will seek ARD approval. Once approved, the regional lead will inform the Species Range Team to post the map in ECOS.

Field Biologists/Species Leads

Field biologists/species leads will be engaged throughout the range map refinement process as they are the subject matter experts that can help guide the modelers with expert advice and information on species being modeled. The Service has designated a lead office for each species, as well as a lead biologist from the lead office. The lead biologist will make final decisions during range development. Many species also have field biologists from offices with jurisdiction over the species across its range. Species leads and field biologists will:

- Make available internally collected occurrence and absence data of species
- Validate and provide quality assurance/quality control (QA/QC) of species occurrence data from external sources (e.g., State/Heritage and other Federal collaborators)
- Provide guidance on potential environmental covariates critical to species
- Participate in model and map review/selection process
- Assist, as needed, in obtaining necessary approvals from the lead region's ARD

Species Range Project Team

Species Range Project Team is an interdisciplinary team with experience spanning ecology, species distribution modeling, statistics, data science, spatial science, programming, computing, and infrastructure development. As such, the team will lead and take on most of the tasks involved in modeling and map creation, including pre- and post-processing datasets and resultant maps.

In addition, the team will provide hands-on support to regional teams/individuals (with necessary R and SDM training) if they wish to model/map their own species for which they have datasets.

Finally, the team will assist the lead biologists, and be responsible for coordinating and facilitating communication with biologists for species that span multiple field office jurisdictions.

Assistant Regional Director (ARD)

Assistant Regional Director, with the lead region responsibility for the species, has the final authority to approve and sign off on a range map before it gets published in ECOS.

STANDARD OPERATING PROCEDURE

4.0   PROCEDURE

Overview

The definition of current range that has been used for the development of procedures in this document is: 'the general geographic area where we know or suspect that a species currently occurs. Current range can include areas that have suitable habitat that have not been surveyed for presence of the species, but are within the range of the species and are considered by experts likely to be occupied during all or part of the annual cycle or life cycle of the species. However, it is not limited to locations for which there are data on species known occurrence. Mapped current range can reflect species occurrence where other supporting evidence indicates that the species has a moderate to high probability of being present. Current range includes seasonally used areas as well as areas used year-round. The time frame covered by the term "current" varies by species and needs to be updated as appropriate.' This definition can also be found here:
https://my.usgs.gov/confluence/display/FEAH/How+to+Enter+Species+Current+Range+and+Area+of+Influence+Data#HowtoEnterSpeciesCurrentRangeandAreaofInfluenceData-3.1WhatIsaSpeciesCurrentRange?
As a starting point, this SOP considers 'current' to be between 1972 and 2019. Therefore, data (i.e., occurrence and environmental data) that occurs during this time period can be used for the construction of current ranges. Please see **Obtaining and Cleaning Species Occurrence Data for Models** for more details.

The Species Range Project Team will consult with species biologists to determine the appropriate method for estimating a species current range. For narrow range species, delimiting polygons around occurrence data or a statistical model using latitude and longitude may be sufficient (see section entitled **Delineating Ranges for Narrow Range Species** for more details). For widespread species with sufficient occurrence and environmental data, a species distribution modeling (SDM) approach may produce a better estimate of the current range. The minimum set of species distribution models that will be used to create an ensemble include generalized linear models (GLM), generalized additive models (GAM), random forest (RF), boosted regression trees (BRT), and maximum entropy (MAXENT), presuming appropriate assumptions, e.g., data density are met for each. Other model forms may be added to the ensemble assuming they meet the conditions described in this SOP.

In any modeling effort, there can be a tradeoff between under- and over-prediction. Overprediction or a commission error occurs when a model incorrectly predicts a species occurrence. Conversely, underprediction or an omission error occurs when a model incorrectly fails to predict an occurrence of a species. The procedure below will attempt to limit both, but for some species, settings can be chosen to minimize underprediction (i.e., omission error), given the Service's mandate to protect threatened and endangered species wherever found.

<div style="background-color: orange; text-align: center;">

STANDARD OPERATING PROCEDURE

</div>

Depending on the taxonomic group being modeled, additional considerations may need to be made. For migratory species, especially birds, if sufficient occurrence data are available, seasonal (i.e., winter, migration, and summer) ranges will be created separately. Similarly, for anadromous species, separate models will be done for marine and freshwater environments. Finally, aquatic species present additional challenges; therefore, a section entitled **Aquatic Species Modeling** is included near the bottom of this document to address these challenges.

There may also exist models or ensembles from within or outside the Service that depict the current range of a species and do not follow the SOP. A model or ensemble can be used as the current range of the species, if the following conditions are met: (1) undergoes an external review process, similar to the one for publication in a peer-reviewed scientific journal; (2) receives approval from species biologists within the Service; (3) the output encompasses the complete current range of the species (i.e., regional models will not be accepted unless they represent some form of seasonal life history characteristic); and (4) all data, metadata and documentation are available conforming to open data act (https://www.congress.gov/bill/115th-congress/house-bill/1770/).

Obtaining and Cleaning Species Occurrence Data for Models

1. Acquire occurrence data from publicly available databases, including: GBIF (www.gbif.org), eBird (www.ebird.org), BISON (bison.usgs.gov), iNaturalist (www.inaturalist.org), Berkeley Ecoinformatics (ecoengine.berkeley.edu), VertNet (vertnet.org), iDigBio (www.idigbio.org), OBIS (obis.org), as well as other appropriate taxon specific databases. These include vouchered museum records, as well as records from verified citizen science databases, such as iNaturalist and eBird.
   a. All threatened and endangered species on iNaturalist have their geographic coordinates obscured by default. Obscured data from iNaturalist will not be used for models. Thus, it is necessary to create a project page in iNaturalist and request that users send their unobscured coordinates to the project. Only research grade (i.e., verified by a photo or sound and another user) observations will be used for models.
   b. When possible, additional data from within the Service, state natural heritage programs, and partner agencies and organizations should also be included. To be included in models, the data must be considered 'research' quality, where either biologists with significant knowledge of the species have made the original identification or it's been verified by an independent expert.
   c. To meet Open Data Governance, all occurrence records used will need to be made publicly available at the end of the process, although spatial coordinate data may be withheld or fuzzed to protect species localities. Therefore, proprietary or closed-source data cannot be accepted unless provider(s) agree to the open data act.
2. Data will be acquired using the valid taxonomic name, as well as any species synonyms according to the Integrated Taxonomic Information System (ITIS) database https://www.itis.gov/. In some cases, additional data for a species may be available, but listed

under another valid and related species. This can occur when one species is split into two or more species, based on newly published taxonomic guidance. In many cases, data for the previous species will not have been transferred to the newly recognized species. Therefore, it may be important to research each species to determine previous taxonomic names and search for those to see whether additional data can be obtained and included. All taxonomic entries that have been used to obtain occurrence data will be documented.

    a. Including data from previous taxonomic names and synonyms could lead to the inclusion of erroneous records. Additionally, vagrancy, when an individual occurs well outside its accepted range, can happen in some taxonomic groups. To address these issues, the records will first be obtained for the valid taxonomic name and depending on species characteristics, an alpha convex hull, minimum convex polygon, or point-buffer created around the occurrence data. The selected geometry will be combined with the current ECOS range and a default 10km buffer applied to create a mask. The distance of this buffer may be modified depending on the traits of the individual species. This mask will then be used to find/assign erroneous and vagrancy records, which will be removed.

    b. The approach above will not work for listed Distinct Population Segments. In this case, the ECOS range or another defined geographic region suggested by species biologists will be used to subset the occurrence data.

3. One should remove records that do not have an observation/collection year, because it is unknown when the observation occurred. For species that will require multiple seasonal ranges (e.g., migratory species) a full date will be necessary.

4. After this is complete, historical records will be removed prior to any SDM construction. Historical includes anything prior to 1972. This cutoff year was chosen because the available climate data (i.e., Oregon State University's PRISM dataset http://www.prism.oregonstate.edu/) ranges from 1981-2010. Occurrence data from 2019 is 9 years beyond this time frame; therefore, to maintain consistency, occurrence data 9 years previous to this time frame should also be included. Historical records are removed in order to model the species current range, as well as keep them temporally consistent with environmental covariates/layers, such as landcover and/or climate data, which can change over time.

    a. A different cutoff date may need to be chosen in consultation with species biologists.

5. Many occurrence records from public databases are incorrectly georeferenced because of observer error or efforts to georeference records with locality descriptions at a later date. Therefore, any records that contain geographic coordinates with suspicious values or those with less than two decimals for both latitude and longitude should be removed. Suspicious values include the following: 0, 0.25, 0.33 (with additional digits of 3), 0.5, 0.66 (with additional digits of 6 or 7), 0.67, 0.75, or 1. Coordinates that match the centroid of counties will also be removed. These are removed because they are likely not georeferenced correctly, if at all.

6. Remove duplicate records (i.e., records with the same coordinates). This process will yield only one record for a specific geographic location. Duplicate records would contain the same environmental information during the modeling process, thereby leading to pseudoreplication.

    a. Many of the same records are included in more than one database, and the number of decimals for the same record differs amongst the databases (i.e., rounding occurs in

some databases). Therefore, it will be necessary to identify the database with the fewest decimals and round the data from the others to match before duplicates are removed. Note: For most species, four decimal places appears to be the fewest number used across all public databases, so all records will be rounded to the fourth decimal place before duplication removal occurs.

7. In some cases, occurrence data may take the form of polygons instead of points, as is the case with data from many natural heritage programs. When the area of a polygon is less than or equal to one square kilometer, an occurrence point will be created for the medial axis of the polygon. For larger polygons, additional information, such as field notes, may be used to assign an occurrence point(s) or the polygon may be removed.

8. The issue of uncertainty of georeferenced occurrence points will be addressed by having species biologists vet and approve the data before modeling can occur. Before vetting the data, biologists will be informed of the model output resolution so an appropriate accuracy level can be used to remove records. Manual removal of occurrence data was chosen because many occurrence data do not have uncertainties; therefore, removing occurrence data that contains uncertainty information could create an inconsistent dataset.

9. After the review process by species biologists, there may exist multiple valid occurrence points within the same grid cell. This would lead to pseudoreplication in models. Therefore, prior to modeling, all grid cells will be limited to a maximum of one occurrence per cell.

10. In order to proceed with a standard species distribution model, there needs to be at least 30 grid cells with occurrence records (Guisan et al. 2017). If a species has less than 30, but at least 10, an alternative modeling approach may be considered as described in Breiner et al. (2015). This approach builds a suite of model iterations with one predictor for each model iteration that are averaged to create an ensemble model.

Selection of Environmental Covariates/Data

1. For each species or subspecies, a standardized list of covariates will be made available, but selection of appropriate covariates for models will be left to species biologists. Species biologists can also suggest additional environmental covariates not included in the list, based on Service documents or expert knowledge. If possible, suggested covariates will either be found or created and included in models. The list of currently available covariates can be found here: ecos.fws.gov/docs/SR_SOP/Predictor_Covariates_Summary.xlsx

2. Species biologists will select no more than one environmental covariate for each set of 10 occurrence points (Harrell et al. 1996). For example, if a species has 40 occurrence points, four covariates may be selected.
   a. Correlations between covariates will be assessed by computing pairwise correlation coefficients. Any pairwise correlation that is higher than 0.70 (Dormann et al. 2012) will be addressed by removing one of the predictors (e.g., usually the covariate of lesser importance).

3. Environmental covariates will initially be created at 1km resolution. If after consultation with species biologists, a higher resolution is needed; either 100m or 30m will be used instead,

contingent upon available environmental covariates at the chosen resolution and the accuracy of occurrence data.

    a. When large changes in grid size are needed, resampling will need to take place in stages to limit the introduction of mathematical artifacts. For example, when downscaling 800m data to 30m, it will need to be downscaled successively from 800m to 600m to 400m to 300m to 100m and then down to 30m. The number and size of the steps will vary depending on the dataset and change in resolution; however, all steps will be documented in the metadata for the final layers.

Creating the Background Region and Selecting Pseudoabsences/Background Points

1. The geographic background, or the region for selecting pseudoabsences/background points, will be created by building a minimum convex polygon around all vetted occurrence data (Rodda et al. 2011).

    a. For some species, a different background may need to be selected. Ideally, the background region should include the current range of the species, as well as the region that the species can disperse to over 'relevant' periods of time (Barve et al. 2011). In these cases, the region will be delineated in consultation with species biologists.

    b. When peer-reviewed research demonstrates that a species occurs in different environmental spaces throughout its range, more than one background will need to be created and the occurrence and pseudoabsence/background data subsetted by those regions (Mainali et al. 2015). Species distribution models will then be run for each region individually and the resulting outputs are combined. This approach presumes that sufficient data densities exist for regional models.

2. For all pseudoabsence/background point selection, the set seed function should be set in R, so the pseudoabsence/background geographic locations can be replicated if needed.

3. Pseudoabsences or background points will be drawn spatially using two separate methods. First, they will be randomly drawn from the background. Second, pseudoabsence/background points will be drawn with the same bias/intensity of the occurrence data. This will be achieved by using a target-group background approach (Phillips et al. 2009), which uses occurrence data from closely related species occurring in the same geographic space for pseudoabsences/background points.

    a. Additionally, pseudoabsence/background points will not be selected from within grid cells that contain occurrence data.

    b. If requested, species biologists will also be provided an opportunity to vet pseudoabsence/background data.

4. The number of pseudoabsence points that will be drawn from the background will be equal to the number of grid cells with occurrence points for RF and BRT. For GLM, GAM, and MAXENT, at least 10x the number of grid cells with occurrence points will be drawn. This approach has been shown to maximize sensitivity or to minimize the possibility of an omission error in the model output (Barbet-Massin et al. 2012).

STANDARD OPERATING PROCEDURE

   a. If true absence data are available for the species within the background region, these data will be used in place of or supplemented with pseudoabsences, if necessary, to meet the quota for regression models.

5. Occurrence data will be combined with pseudoabsence/background data. This resulting dataset will be intersected with the selected environmental layers and run through five different modeling algorithms.

Building Individual and Ensemble Models

1. Five algorithms will be run for every species including: generalized linear models (GLM), generalized additive model (GAM), random forest (RF), boosted regression tree (BRT), and maximum entropy (MAXENT). These models were chosen because they have been shown to perform well in a number of peer-reviewed studies (Drake et al. 2006; Elith et al. 2006; Elith and Graham 2009). Some initial specific algorithm settings are listed below.
   a. For GLM, a backward stepwise regression will be performed that uses AIC for covariate removal. Interaction and quadratic terms may also be included.
   b. For GAM, the mgcv package in R will automatically choose the 'best' degree of smoothing based on internal cross-validation.
   c. For RF, the number candidate covariates taken randomly at each node should be equal to the square root of the number of covariates (Breiman 2001). Additionally, at least 1,000 individual trees will be built.
   d. For BRT, at least 1,000 trees will be built and the learning rate will be set no higher than 0.01 (Elith et al. 2008). Additionally, tree complexity will be set to 2-5 nodes, the bag fraction will be set to 0.5, and 10-fold cross-validation will be conducted (Elith et al. 2008).
   e. It may be necessary with MAXENT models to alter the regularization parameter to prevent over- or under-fitting of the data.

   Based on the characteristics of different datasets, we will need to fine-tune the parameters for different models to reach the best performance. The parameters for the best models will be recorded and stored with the model for future reference.

2. K-fold cross-validation will be done to evaluate models. The appropriate number of folds will be chosen based on occurrence data density.
3. Evaluation metrics, including kappa, true skill statistic (TSS), area under the curve (AUC), specificity, and sensitivity will be used to validate and evaluate individual models.
4. Each model will be projected onto the union of the background region and the ECOS current range polygon buffered by 10km. In cases when a species may occur outside this combined region, it may be necessary to buffer by a different amount (e.g., 100km) or use an ecoregion. Other projection regions may be considered under consultation with species biologists.
5. Two types of ensemble models will be created. The first by averaging the individual model's suitability values, yielding an average suitability surface. Then a variety of thresholds will be applied to these mean suitability surfaces. Second, each individual model will be thresholded

using kappa and then stacked to create a model concordance surface. Then several thresholds (i.e., number of models that predict presence) will be applied.

6. All output maps, including threshold and ensemble models, will be sent to species biologists for review. After receiving review, the maps may be updated by the inclusion of, for example, new occurrence data, environmental covariates, or changing the threshold.

7. If output maps are not able to be corrected by altering the models, then species biologists may add locations where the species is present, but the model does not predict it, to the output maps for inclusion into ECOS. This will likely be the case with critical habitat. If the model does not predict a species occurrence within critical habitat, that location will be appended to the output. Similarly, if the model inaccurately predicts the species within block cleared areas, those predictions will be removed. There may be additional adjustments required for the final map for some species, such as removing predictions in urban areas. Any change that occurs to model outputs will be documented in the associated metadata.

Aquatic Species Modeling

1. The model resolution unit for aquatic species will be the line segment catchment polygons associated with the National Hydrography Dataset (NHD)-plus dataset from the United States Geological Survey. Using these catchments, instead of a square grid, better aligns with the distributions of aquatic species, which are restricted to watersheds.

2. Similar to the square grid for terrestrial species, the environmental predictors will be summarized for each catchment. This means that land cover data will be summarized to the catchment, as well as stream conditions (i.e., velocity, volume). Additionally, occurrences will be summarized, pseudoabsence/background points drawn, and model predictions made into catchment polygons.

3. As with terrestrial species, the background region will consist of creating a minimum convex polygon around occurrence points. However, the resulting polygon will be overlaid on the catchment polygons, and those that intersect will be extracted. These extracted polygons will be the background for modeling.

Delineating Ranges for Narrow Range Species

1. If a species has a very small range, it may be more appropriate to delineate the range by using a statistical model with latitude and longitude as covariates, by delineating a polygon(s), or by using another modeling method.

2. Polygons will be delineated based upon Service documents, taxonomic websites (e.g., amphibiaweb.org), range descriptions on NatureServe Explorer, and peer-reviewed literature. All sources that were used to delineate ranges will need to be recorded and cited.

3. For deductive models, covariates and associated thresholds will be chosen based on recommendations by species biologists and examining occurrence data.

STANDARD OPERATING PROCEDURE

5.0 FILE LOGISTICS/ATTRIBUTES

MetaData for Species Occurrence Data and Final Current Range Map

The following data standards shall be met for each geospatial file:

File Format

ESRI shapefiles shall serve as the common storage format for all data. A shapefile is actually a collection of files:

.shp – (the shapefile) feature data
.dbf – Attribute table
.shx – Shape index file
.prj – Projection file
.xml – Metadata file (see below for additional information)

Each shapefile must contain features and attributes for only one species.

File Naming Convention

All shapefiles must comply with the following shapefile naming convention. This naming convention applies to the prefix of all components of the shapefile, including the metadata record.

<SPCODE>_<VIPCODE>_<EXT>
SPCODE: 4 place alphanumeric species code
VIPCODE: 3 place alphanumeric species code
EXT: this indicates whether the shapefile is the current range <CURR> or occurrence data <OCC> for the species.

*The SPCODE and VIPCODE for any listed entity can be searched for at https://ecos.fws.gov/ecp0/reports/ad-hoc-species-report?status=E&status=T&status=EmE&status=EmT&status=SAE&status=SAT&fleadreg=on&fstatus=on&finvpop=on&fvipcode=on&fspcode=on&header=All+Listed+Species or by going to https://ecos.fws.gov/tess/speciesModule/PopulationSearchForm.do. Once you find the desired listed entity in the search results, click on its View/Edit button. A page that includes the SPCODE and VIPCODE will be displayed.

Example:

Shapefile prefix for the Marbled murrelet where:
SPCODE = B08C

STANDARD OPERATING PROCEDURE

VIPCODE = V02
Shapefile prefix = B08C_V02_CURR

Coordinate Format

All coordinates shall be in NAD 83 CONUS Albers (EPSG:5070).

Attributes for Occurrence Data Shapefiles

All x/y coordinate records used in the range delineation should have at least the following attributes:

Name: EOO_Source_Name
Description: This is the name of the source from which the x/y coordinate was added to this database. It may be a report title, a database name, the name of a museum collection, etc.
Type: Text, up to 254 characters

Name: EOO_Source_Date
Description: This is the date the x/y coordinate was copied/extracted from the source. For hard copy data use the date of digitization, for data from existing databases use the date of copy/extraction.
Type: Date (mm/dd/yyyy)

Name: EOO_Date
Description: This is the date of the observation. If the record represents a date range, use the most recent date. If the date is fuzzy (e.g., May 1980) enter it at the start of the period (e.g., May 1, 1980); or if your date is accurate only to year (e.g., 1980) enter it as January 1, 1980.
Type: Date (mm/dd/yyyy)

Name: EOO_Source_Details
Description: This is an optional field to record any additional desirable information about the source. This field will not contain any personally identifiable information. (e.g., author of the report, collection number, etc.)
Type: Text, up to 254 characters

Attributes for Current Range Shapefiles

Name: Range_Status
Description: This indicates whether the polygon was "added" or "removed" from the final model output. For polygons that were created directly from model output, they should be designated as "model".
Type: Text, up to 254 characters

Name: Range_Contact
Description: The individual requesting the change (i.e., addition or deletion) should be recorded.

STANDARD OPERATING PROCEDURE

Type: Text, up to 254 characters

Name: Range_Reason
Description: If the polygon was created or removed, a reason should be
    provided, along with any supporting documentation.
Type: Text, up to 254 characters

Shapefile Metadata

All submitted shapefiles are required to have FGDC compliant metadata records. The metadata records can be created using the ESRI ArcGIS metadata tools or using other FGDC-compliant metadata creation tools. Please refer to the FGDC Metadata website at http://www.fgdc.gov/metadata/ for detailed information on metadata standards, minimal metadata record requirements, and metadata document creation tools.

The naming convention for the metadata files is the same as for other shapefile files. See the "File Naming Convention" section of this document for a description of the required file name prefixes.

Metadata Format

All metadata records must be submitted in FGDC compliant XML format. The ESRI ArcGIS metadata tools will create compliant XML formatted metadata records by default. If you choose to use another metadata creation tool, ensure that the tool creates FGDC-compliant XML formatting.

Metadata Content

Use ArcGIS or another dedicated tool for metadata editing and add the following information to the shapefile's metadata:

Fieldname: Item Description – Summary:
Value to enter: "Current range for" or "Occurrence data for" <species common name>
    "(<species scientific name>)."

Fieldname: Item Description – Description:
Value to enter: This will be a brief description of the contents of the file.

Fieldname: Contacts – Name:
Value to enter: <name of the lead Regional Office for the species>

Fieldname: Contacts – Organization:
Value to enter: "U.S. Fish and Wildlife Service"

Fieldname: Contacts – Phone:
Value to enter: <main phone number for the lead Regional Office for the species>

STANDARD OPERATING PROCEDURE

Final Current Range Map in ECOS

It will not be possible to upload the final current range map to ECOS until we have made significant code changes to the system. Therefore, through coordination with the species biologist, the Species Range Team (SRT) will ensure that the final current range map is posted to ECOS and made publically available. Please send an email to the SRT (srpteam@fws.gov) and indicate that the final map can be posted to ECOS. If the final current range map was produced outside of the SRT, the regional species lead must verify that one of the four conditions for not following the SOP (outlined on page 5 of this SOP) was met, and provide all of the appropriate shape files and metadata, as well as documenting the source of the data before it will be posted in ECOS.

Storage of data points

Ecological Services is working to provide a long-term storage solution.

6.0   LITERATURE CITED

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.

Barbet-Massin M, Jiguet F, Albert C, and Wilfried T (2012) Selecting pseudo-absences for species distribution models: how, where and how many? Methods in Ecology and Evolution 3:327-338.

Barve N, Barve V, Jimenez-Valverde A, Lira-Noriega A, Maher S, Peterson A, Soberon J, Villalobos F (2011) The crucial role of the accessible area in ecological niche modeling and species distribution modeling. Ecological Modelling 222:1810-1819.

Breiman L (2001) Random forests. Machine Learning 45:5-32.

Breiner F, Guisan A, Bergamini A, Nobis M (2015) Overcoming limitations of modelling rare species by using ensembles of small models. Methods in Ecology and Evolution 6:1210-1218.

Dormann C, Elith J, Bacher S, Buchmann C, Carl G, Carré G, Marquéz J, Gruber B, Lafourcade B, Leitão P, Munkemüller T, McClean C, Osborne P Reineking B, Schroder B, Skidmore A, Zurell D, Lautenbach S (2012) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36:27-46.

Elith J, Graham C (2009) Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. Ecography 32:66-77.

STANDARD OPERATING PROCEDURE

Elith J, Graham C, Anderson R, Dudík M, Ferrier S, Guisan A, Hijmans R, Huettmann F, Leathwick J, Lehmann A, Li J, Lohmann L, Loiselle B, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton J, Peterson A, Phillips S, Richardson K, Scachetti-Pereira R, Schapire R, Soberón J, Williams S, Wisz M, and Zimmermann N. (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29:129-151.

Elith J, Leathwick J, Hastie T (2008) A working guide to boosted regression trees. Journal of Animal Ecology 77:802-813.

Guisan A, Thuiller W, Zimmermann N (2017) Habitat suitability and distribution models: with applications in R. Cambridge, United Kingdom: Cambridge University Press.

Harrell F, Lee K, Mark D (1996) Tutorial in biostatistics multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in Medicine 15:361-387.

Mainali K, Warren D, Dhileepan K, McConnachie A, Strathie L, Hassan G, Karki D, Shrestha B, Parmesan C (2015) Projecting future expansion of invasive species: comparing and improving methodologies for species distribution modeling. Global Change Biology 21:4464-4480.

Phillips S, Dudik M, Elith J, Graham C, Lehmann A, Leathwick J, Simon F (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data Ecological Applications 19:181-197.

Rodda G, Jarnevich C, Reed R (2011) Challenges in identifying sites climatically matched to the native ranges of animal invaders. Plos One 6:1-18.

Sofaer H, Jarnevich C, Pearse I, Symth R, Auer S, Cook G, Edwards T, Guala G, Howard T, Morisette J, Hamilton H (2019) Development and delivery of species distribution models to inform decision-making. BioScience 69:544-557.

7.0 CONTACTS

Michaelle Shultz
Chief, Branch of Data Integration
Michaelle_Shultz@fws.gov 703-358-1869

Matthew Moskwik
Management Analyst, Species Range Delineation Project
matthew_moskwik@fws.gov 703-358-1889