Species Distribution Modeling Ensemble Tool (SDMet)
or
Ensemble Tool for Species Distribution Modeling (eSDM)
or
Ensemble Tool for species distribution models (ETsdm aka ET)

# Table of Contents

# Manual
# Species distribution modeling ensemble tool (SDMet)
# (beta version, Oct 2017)

## 1. Introduction

### 1.1. Why create ensemble models?

As the field of species distribution modeling (SDM) for marine species in their dynamic environment has advanced, the resulting spatial models have greatly enhanced the ability of marine resource managers to assess and mitigate potential impacts to protected species at the appropriate spatial scales. However, different data sets or different analytical approaches often yield different modeled results, creating uncertainty and challenges for management decision-making. As an example, there are currently multiple spatial and habitat-based models of blue whale (*Balaenoptera musculus*) density, distribution, and biologically important areas off the U.S. West Coast (Forney et al. 2012, Redfern et al. 2013, Irvine et al. 2015, Calambokidis et al. 2015, Becker et al. 2016, Hazen et al., 2016). These models can directly inform management decision-making, e.g. for reducing ship-strike risk but different models may suggest different actions because of strengths, biases, and limitations for underlying data set and model.

When original data sources are available to support an integrated analysis, i.e. in a Bayesian hierarchical framework, this can provide a robust, probabilistic assessment of the combined information on spatial distributions of protected species. However, this is generally a time-consuming and (at times) analytically challenging approach, and highly disparate data sources may be too complex to allow an effective joint analysis given computational capabilities. In other cases, the different investigators may only be willing to share the output from their SDMs, and not the original survey, telemetry, or other data sources.

For these reasons, a method of combining model outputs (i.e., predicted species distributions) is needed, allowing end-users and managers to explore and evaluate 'ensemble predictions' that take into account each model's strengths and weaknesses and highlight areas of greater or lesser uncertainty about species occurrence. Predictions derived from an ensemble of models are generally more robust than single-model predictions. Integration and comparisons of predictions across models provides critical information on uncertainty, and provides a foundation for improving modeling methodologies by highlighting key assumptions and limitations of each model and/or data set. Combining multiple models can be complex when models are developed at different spatial scales, with different geographic projections, and using differing prediction units (e.g. animal density, probability of occurrence, presence/absence).

The package SDMet has been designed as a user-friendly spatial tool with a web-based interface that allows end-users to import the SDM model layers (i.e. as GIS layers or raster data) and create and explore ensemble predictions to inform management and explore spatial

uncertainties. The manual below describes how to use this tool, referencing sample data sets that are provided as part of the software package.

### 1.2. Structure of Manual

[Do this last]

### 1.3. Supported import file types

#### Excel .csv

- [describe format]

#### Raster files

[describe format]

#### Shapefile files

[describe format]

#### File geodatabase feature class files

[describe format]

### 1.4. Supported export file types

#### Excel .csv

[describe format]

#### Shapefile

[describe format]

#### KML files

[describe format]

### 1.5. Supported image file formats for download

#### JPG

[describe format]

#### PDF

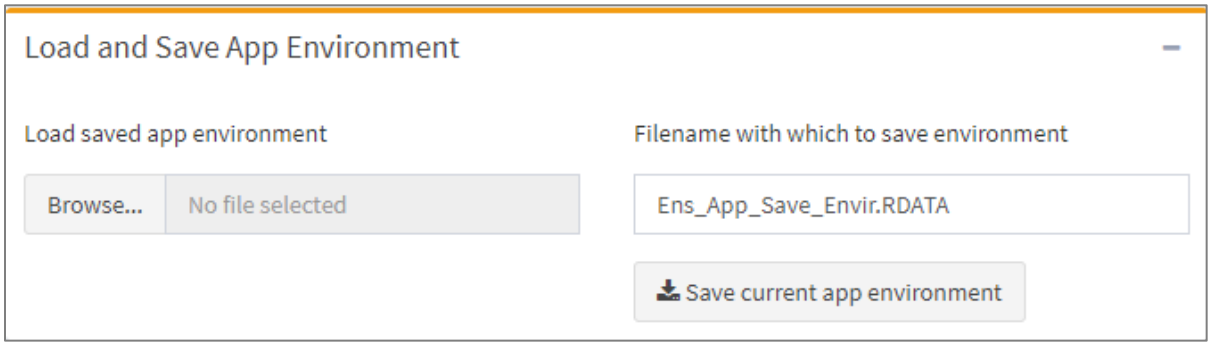
[describe format]

#### PNG

[describe format]

## 2. Load Model Predictions Tab

## 2.1. Load and Save App Environment




*Load saved app environment*

- Browse to and load an .RDATA file created using tool in order to restore saved environment.

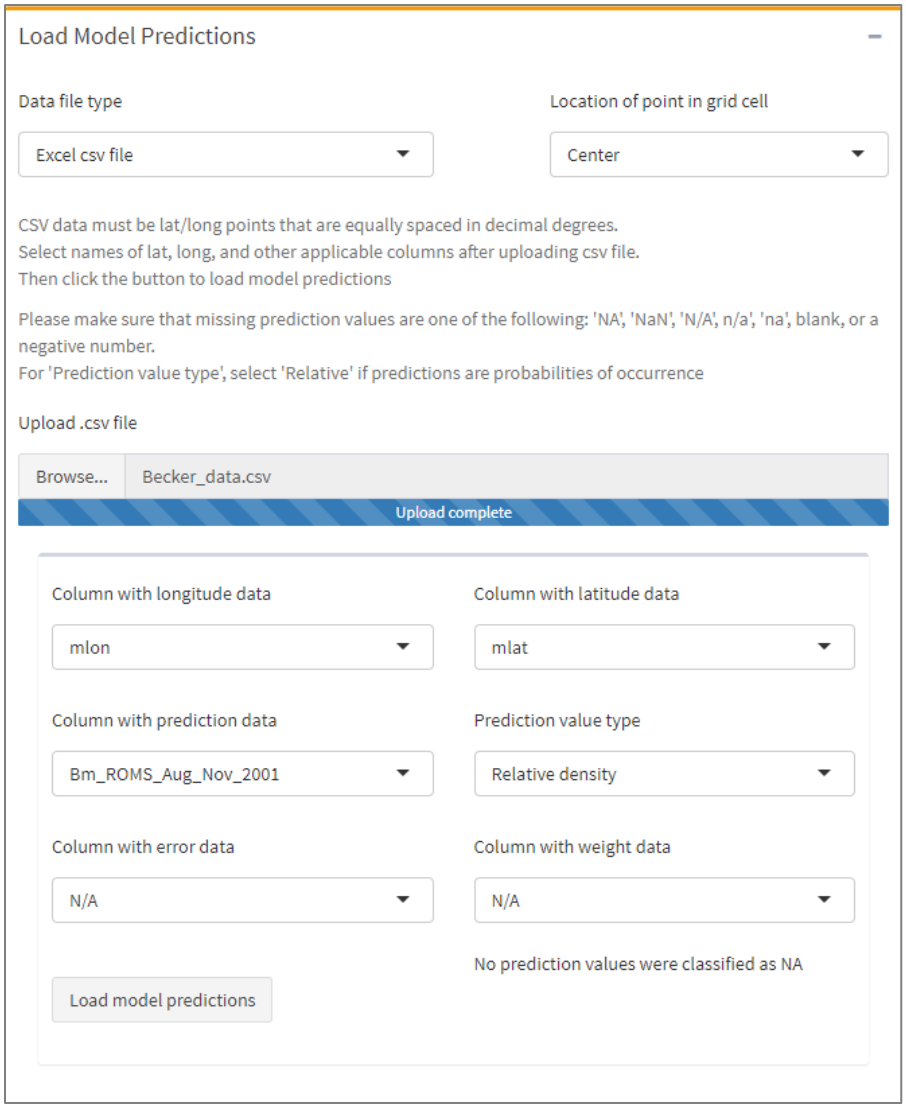
*Filename with which to save environment*

- Name of .RDATA file that will be downloaded if you click *Save current app environment*.

*Save current app environment*

- Click to download an .RDATA file that stores data loaded in the current session of the tool. This file can be loaded using *Load saved app* environment, thereby allowing the user to mostly pick up where they left off in their last session (selections made in the tool are not currently saved). Aspects of the app that are saved include but are not limited to: original model predictions, study area polygon, land polygon, overlaid model predictions, ensemble predictions, validation data, and calculated evaluation metrics.

## 2.2. Load Model Predictions




*Data file type*: "Excel csv file"

*Location of point in grid cell*:

- Specifies what part of each prediction grid cell the longitude and latitude coordinates in the loaded .csv file represent. The options are "Center", "Top right", "Top left", "Bottom left", and "Bottom right".

*Upload Excel csv file (.csv)*

- Browse to and select the .csv file that contains the model prediction data you want to load into the app. The longitude and latitude points must be geographic coordinates that are equally spaced in decimal degrees. This file must have headers, and an error message will appear if you select a file that is not a .csv file.

*Column with longitude data*

- Select the name of the column with the longitude data.

*Column with latitude data*

- Select the name of the column with the latitude data.

*Column with prediction data*

- Select the name of the column with the prediction data.

*Prediction value type*

- Select "Absolute density" if all necessary correction factors have been applied to the model predictions for the applicable study area, and thus provide a true prediction of density. Select "Relative density" if the model prediction have **not** had all necessary correction factors applied for the applicable study area, and thus only accurately predict the density relative to the other model predictions. If you select "Absolute density", then the abundance will be displayed in the *Loaded Model Predictions table* and you will be able to select "None" as a rescaling option in the 'Create Ensemble Predictions' tab.

*Column with weight data*

- Select the name of the column with the weight data. All weight values must be between zero and one. If you do not have weight data for these model predictions, then select "N/A". Weight data can be used as pixel-level spatial weights in a weighted ensemble in the 'Create Ensemble Predictions' tab. Thus, these weights could be some type of error value or values with zero for all points south of some latitude, among other options.

*NA prediction values message*

- A message detailing how many of the provided prediction values will be classified as NA. A prediction value will be classified as NA if the provided data is one of the following: 'NA', 'NaN', 'N/A', 'n/a', 'na', blank, or a negative number.

*Load model predictions*

- Click to load model predictions and other specified data from the uploaded .csv file into the app.

*Data file type:* "GIS raster"

*Band number of prediction data*

- The band number of the prediction data within the loaded raster .tif file.

*Upload raster file (.tif)*

- Browse to and load the TIFF file that has the extension '.tif. The raster can be in any projection, but the raster coordinates must be between the equivalent of -180 and 180 decimal degrees. An error message will appear if you select a non-.tif file, or if the selected raster does not have any data at the provided band number.

*Prediction value type*

- See *Prediction value type*

*NA prediction values message*

- See *NA prediction values message*

*Load model predictions*

- Click to load model predictions and other specified data from the uploaded .tif file into the app.

*Data file type:* "GIS shapefile"

*Upload GIS shapefile files*

- Browse to and select all files of the desired GIS shapefile. Although they will have different file extensions, these files will all have the same file name (this will be the name of the shapefile in ArcCatalog). An error message will appear if not all of the files of the desired shapefile are selected, or if extraneous files are selected.

*Column with prediction data*

- Select the name of the column with the prediction data.

*Prediction value type*

- See *Prediction value type*

*Column with error data*

- See **Error! Reference source not found.**

*Column with weight data*

See *Column with weight data*

*NA prediction values message*

- See *NA prediction values message*

*Load model predictions*

- Click to load model predictions and other specified data from the uploaded shapefile into the app.

*Data file type:* "GIS file geodatabase (.gdb) file"

*Full path to file geodatabase*

- Enter the full path up to and including the file geodatabase that contains the file geodatabase feature class you wish to load. In the file path, the file geodatabase will act as a folder and will have a '.gdb' extension. Do not put any additional text, such as a '/', after the .gdb extension. You can copy and paste the file path from the top bar of ArcCatalog after navigating to the desired file geodatabase. On a Windows machine, you can also copy and paste the file path from the top bar of the Windows Explorer.

*Name of file geodatabase feature class*

- Enter the name of the file geodatabase feature class you wish to load. You can find this name in ArcCatalog, where the file type should be 'File Geodatabase Feature Class'.

*Upload file geodatabase feature class*

- Click to upload the specified file from the specified path. An error message will appear if the app does not find a file with the specified name at the specified path.

*Column with prediction data*

- Select the name of the column with the prediction data.

*Prediction value type*

- See *Prediction value type*

*Column with error data*

- See **Error! Reference source not found.**

*Column with weight data*
See *Column with weight data*

-

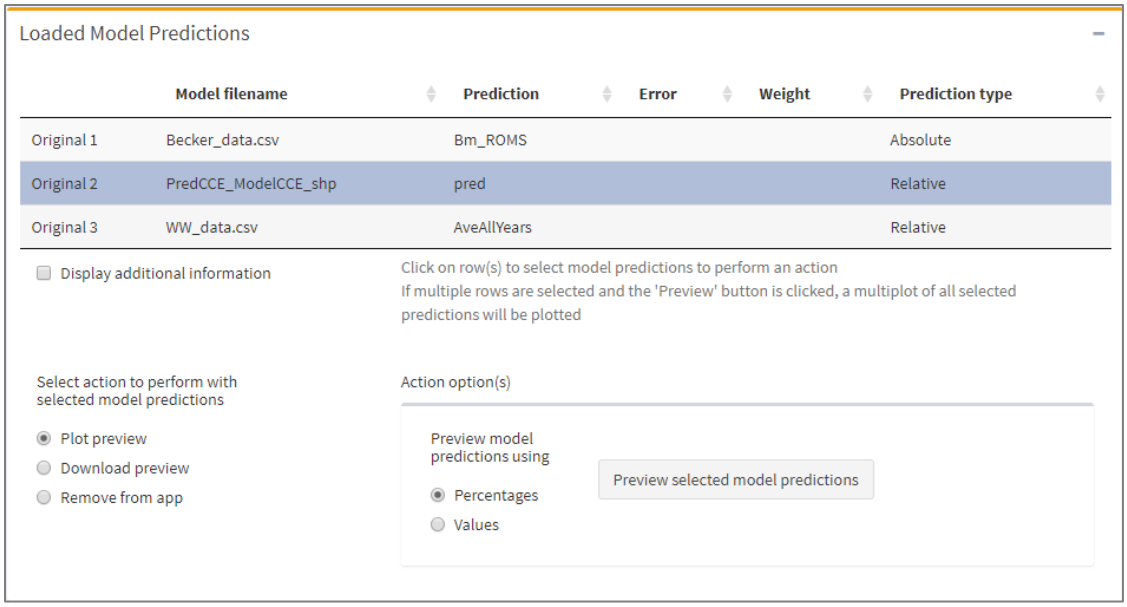
*NA prediction values message*

- See *NA prediction values message*

*Load model predictions*

- Click to load model predictions and other specified data from the uploaded file geodatabase feature class into the app.

## 2.3. Loaded Model Predictions




*Loaded Model Predictions table (Load tab)*

- Select or deselect a set of model predictions by clicking on the row of that set of model predictions in the table. You may select multiple sets of model predictions at one time. If selected, a row is highlighted grey-blue.
- Description: The first table (displayed above) is displayed when *Display additional information* is unchecked and reports information specified when the model predictions were being loaded in. This information consists of the name of the file that was loaded, the names of columns from which prediction, error, and weight data were loaded, and the specified prediction type. The 'Error' and 'Weight' columns are blank if "N/A" was specified for those columns. The second table is displayed when *Display additional information* is checked, and it reports additional information about the loaded model predictions. This information consists of the resolution of the predictions (see below for more details), the number of cells in the grid of model predictions, the count of the non-NA predictions, the predicted abundance (if the predictions are absolute densities), and the range of the model predictions.
    - 'Resolution' column: When model predictions are loaded into the app, the app attempts to determine the resolution at which the predictions were made, meaning the distance between the centroids of adjacent grid cells. The app only attempts to calculate the resolution in the native projection of the loaded model predictions, so if model predictions were generated on an equal area grid but were loaded into the app via a shapefile in WGS 84 geographic coordinates, then the app will not be able to calculate the resolution. This functionality is still in its infancy, so please check the reported resolution if you know the resolution of the loaded predictions.

*Action to perform with selected model predictions*

- Select the action you wish to perform with the set(s) of model predictions selected in the *Loaded Model Predictions table*. The options in the *Action option(s)* box will change depending on the selected action.

*Action option(s)* for “Plot preview” (see image above)

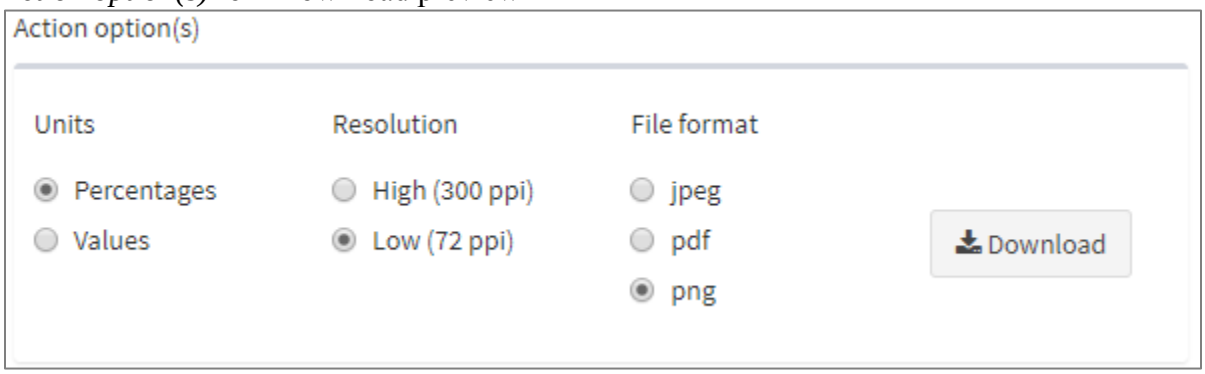
*Preview model predictions using*

- Select “Percentages” to have the color of predictions be based on the prediction value relative to the other prediction values in the set of model predictions. For instance, the color scheme will delineate, among other ranges, the top 2% of prediction values, the prediction values in the top 2% to 5%, and the prediction values in the top 5% to 10% of all of the prediction values. Select “Values” to have the color of predictions be based on the numerical values of the predictions

*Preview selected model predictions*

- Click to generate the preview in the “Preview” box. This can be a preview of a single set of predictions or a multiplot of multiple sets of predictions.

*Action option(s)* for “Download preview”




*Units*

- See here

*Resolution*

- The resolution in which the preview will be downloaded. It is recommended to use the “High” resolution for multiplot previews.

*File format*

- The file format in which the preview will be downloaded. The current options are JPEG, PDF, and PNG file types.

*Filename*

- The filename of the specified plot if it is downloaded. The filename is reset to the default if different predictions are chosen to be plotted or if any of the options in the *Action option(s)* box are changed.

*Download*

- Click to download preview of selected set(s) of model predictions with the parameters and filename specified in the *Action option(s)* box.

*Action option(s)* for “Remove from app”


*Remove selected model predictions*

- Click to remove the selected set(s) of model predictions from the app.

## 3. Overlay Model Predictions Tab

### 3.1. Load Study Area Polygon

Study area polygon description:

- A study area polygon should be used if you have model predictions that cover a broad area, but you only want to create ensemble predictions in a specific study area (region) within the broader area. For instance, if you have model predictions that span the US west coast but only want ensemble predictions in the Southern California Bight, then you could upload a study area polygon that covers the Bight. If the study area boundary cuts through a prediction cell, then only the portion of the prediction cell that is within the boundary is kept during the overlay.

*Use a study area polygon in the overlay process*

- Check to upload a study area polygon, and uncheck to remove a loaded study area polygon. Note that a message is displayed when a study area polygon is loaded.

*File type:* "Excel csv file"

*Upload Excel csv file (.csv)*

- Browse to and load the file with the .csv extension that contains the desired polygon(s). The file must have headers, the first column must contain the longitude values, and the second column must contain the latitude values. The longitudes and latitudes must be in geographic coordinates in the range [-180, 180]. Multiple polygons may be demarcated using blank cells or cells with 'NA' entries. If the provided points do not form a closed polygon, then the last point is connected to the first point. Please be aware that this could create an invalid polygon.

*File type:* "GIS shapefile"

*Upload GIS shapefile files*

- See *Upload GIS shapefile files*

*File type:* "GIS file geodatabase (.gdb) file"

*Upload GIS shapefile files*

- See *Full path to file geodatabase*

*Upload GIS shapefile files*

- See **Error! Reference source not found.**

*Upload GIS shapefile files*
See *Upload file geodatabase feature class*

### 3.2. Load Land Polygon

Land polygon description:

- A land polygon should be used if you have model predictions for a strictly marine species. If a land polygon is loaded, then during the overlay process all land area specified by the land polygon will be erased from the base grid and thus all overlaid models as well.

*Use a land polygon in the overlay process*

- Check this box to upload a land polygon, and uncheck this box to remove a loaded land polygon. Note that a message is displayed when a land polygon is loaded.

*Land polygon source:* "Use provided"

- Use one of the land polygons provided with the app. The provided land polygons were downloaded from the Global Self-consistent, Hierarchical, High-resolution Geography (GSHHG) Database. See the GSHHG website for more information about these land polygons.

*Resolution of land polygon*

- The resolution options are "full", "high", "intermediate", "low", and "crude". Each resolution option is an approximately 80% reduction in size and quality from the previous resolution option.

*Load provided land polygon*

- Load the GSHHG land polygon with the specified resolution

*Land polygon source:* "Load personal"

*File type:* "Excel csv file"

*Upload Excel csv file (.csv)*

- Browse to and load the file with the .csv extension that contains the desired polygon(s). The file must have headers, the first column must contain the longitude values, and the second column must contain the latitude values. The longitudes and latitudes must be in geographic coordinates in the range [-180, 180]. Multiple polygons may be demarcated using blank cells or cells with 'NA' entries. If the provided points do not form a closed polygon, then the last point is connected to the first point. Please be aware that this could create an invalid polygon.

*File type:* "GIS shapefile"

*Upload GIS shapefile files*

- See *Upload GIS shapefile files*

*File type:* "GIS file geodatabase (.gdb) file"

*Upload GIS shapefile files*

- See *Full path to file geodatabase*

*Upload GIS shapefile files*

- See **Error! Reference source not found.**

*Upload GIS shapefile files*

See *Upload file geodatabase feature class*

### 3.3. Loaded Model Predictions

*Loaded Model Predictions table (Overlay tab)*

- Select one row in order to use the grid of the corresponding set of model predictions as the base grid in the overlay process (see Appendix *tbd* for a description of the overlay process). See *Loaded Model Predictions table (Load tab)*
- for descriptions of the data contained in each column and the functionality of the *Display additional information* checkbox.

### 3.4. Overlay Model Predictions

Overlay process description:

- See Appendix *tbd* for a more detailed description of the *Overlay options* and overlay process.

*Overlay options: projection*

*Perform overlay in lat/long WGS 84 geographic coordinates*

- The overlay of model predictions onto the selected base grid will be performed with the base grid and all of the model predictions in WGS 84 geographic coordinates.

*Use the selected model predictions' projection during the overlay process*

- This option appears if *Perform overlay in lat/long WGS 84 geographic coordinates* is unchecked. The overlay process will be performed in the projection of whichever set of model predictions is selected.

*Overlay options: percent overlap*

- The slider bar specifies the percent that the original model prediction(s) must overlap a base grid cell for that cell to have a non-NA overlaid prediction value. A slider bar value of "0" means that cell will have a non-NA overlaid prediction value if there is any overlap with any original model prediction.

*Overlay all predictions onto specified base grid*

- Click this button to overlay all sets of model predictions onto the specified base grid.

### 3.5. Preview of Base Grid

*Preview*

- Click to generate a preview of the base grid, including loaded study area and land polygons. The preview will not be generated if *Use a study area polygon in the overlay process* is checked and no study area polygon is loaded, or if *Use a land polygon in the overlay process* is checked and no land polygon is loaded. Note that if the grid of the model predictions selected to be the base grid was made at a high resolution, then the base grid in the preview may appear to be completely black because of the small grid cell size.

### 3.6. Preview of Overlaid Model Predictions

*Overlaid model predictions to preview*

- Select the set(s) of overlaid model predictions to preview. The sets of overlaid model predictions are listed as "Overlaid 1", "Overlaid 2", etc. These numbers correspond to the "Original 1", "Original 2", etc., entries in the far left column of the *Loaded Model Predictions table (Overlay tab)*, meaning that the "Overlaid 1" set of model predictions is the "Original 1" model predictions overlaid onto the base grid, etc. You may select multiple sets of predictions in order to do a multiplot.

*Preview*

- Click to generate a preview of the selected set(s) of overlaid model predictions. This can be a preview of a single set of overlaid model predictions or a multiplot of multiple sets of overlaid model predictions. This feature is meant to allow you to be able to quickly preview the overlaid models, possibly at a coarser resolution than the actual overlaid predictions, and see how they differ from the original model predictions. If you want high quality images of the overlaid model predictions, please use the 'High Quality Maps' tab.

## 4. Create Ensemble Predictions Tab

### 4.1. Overlaid Model Predictions

### 4.2. Create Ensemble Predictions

### 4.3. Created Ensemble Predictions

## 5. Evaluation Metrics Tab

### 5.1. Select Predictions to Evaluate

*Original, Overlaid, and Ensemble Tables*

- Click on the rows of the original, overlaid, and/or ensemble model predictions for which you wish to calculate evaluation metric(s).

## 5.2. Load Validation Data

*Validation data file type*

"Excel .csv file"

*Upload Excel .csv file*

See *Upload Excel csv file (.csv)*

- 

"GIS shapefile"

*Upload GIS shapefile files*

- See *Upload GIS shapefile files*

"GIS file geodatabase feature class"

*Full path to file geodatabase*

- See *Full path to file geodatabase*

*Name of file geodatabase feature class*

- See …*Full path to file geodatabase*

*Upload file geodatabase feature class*

- See …

*Validation data type:*

"Counts (numerical)"

- Validation data is numerical, meaning it consists of actual counts of organisms.

"Presence or absence"

- Validation data is presence/absence, meaning that it consists of codes that specify whether a point is a presence point or an absence point.

*Select, in this order, the longitude, latitude, and data columns for …*

- When you click on the input box, the headers of all of the columns in the uploaded file will be displayed. Select, in order, the headers that correspond to the longitude column, the latitude column, and the column that contains either the count data or the presence and absence data. You can delete selected column names from the input box by selecting them and clicking your computer's delete button.

*Select the validation data column for the uploaded GIS data*

## 5.3. Calculate Metrics

## 5.4. Metrics

## 5.5. Metric Descriptions and References

# 6. High Quality Maps Tab

## 6.1. Select Predictions to Plot

## 6.2. Plot Parameters

### 6.2.1. Map range

### 6.2.2. Title and axis labels

### 6.2.3. Tick marks and tick labels

### 6.2.4. Color scheme of predictions

# 7. Export Predictions Tab

## 7.1. Select Predictions to Export

## 7.2. Export Predictions

# 8. Manual Tab

- Contains this manual

# 9. Submit Feedback Tab

- Use this form to: 1) describe functionality you would like to see in future releases, 2) report errors in the tool and 3) comment on any other facet of the tool. In all cases, please provide as much detail as possible. Thank you!

# 10. Appendix

## 10.1. Appendix 1: The overlay process

The overlay process is the backbone of this tool, as it allows you to create ensembles of species distribution model (SDM) predictions when the SDMs were made using grids with different positions and/or spatial resolutions.

The first step of the overlay process is converting all of the loaded sets of model predictions, which are displayed in the *Loaded Model Predictions Table (Overlay tab)*, to the projection or geographic coordinates specified in *Overlay options: projection*. Next, the base grid, which is selected in the *Loaded Model Predictions Table (Overlay tab)*, is created and processed. If a study area polygon is loaded, then the base grid is clipped to the specified study area. If a land polygon is loaded, then the land area specified by this polygon is erased from the base grid.

After the base grid is created, then the other SDM predictions are overlaid onto the it. To do this, the app performs the following process for each base grid polygon. For this description, the SDM predictions that are being overlaid will be referred to as “the original model”. First, the app determines which polygons from the original model intersect with the base grid polygon, and the percentage of each of these polygons that overlaps with the current base grid polygon. Next, the app calculates the predicted abundance for each of the overlapping polygons from the original model.