Ensemble Tool for Species Distribution Modeling

eSDM

Samuel Woodman

October 2017

Table of Contents

User Manual
Ensemble Tool for Species Distribution Modeling (eSDM)
(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 eSDM 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.

1.2. Structure of Manual

The manual below describes how to use this tool, referencing sample data sets that are provided as part of the software package. Key steps in the ensemble process include: Loading model prediction (Section 2), Providing study area boundary/coastline files, and selecting a model as the base grid for overlaying all models (Section 3); Selecting methods of creating ensemble predictions and creating those ensembles (Section 4); Calculating performance metrics for models (Section 5); Creating high-quality maps (Section 6), and Exporting model predictions (Section 7). Support and assistance is provided via a PDF manual (Section 8), comments/feedback tool (Section 9), and Appendices with details for some of the internal operations and calculations

1.3. Supported import file types: SDM prediction data

SDM prediction data requirements for all file types

The prediction data loaded into the app must be either relative or absolute density predictions. Note that in this app, probability of occurrence is considered a relative density. Abundance predictions are not supported at this time and thus must be converted to densities before loading them into the app. Predictions that are one of 'NA', 'NaN', 'N/A', 'n/a', 'na', blank, or a negative number will be classified as NA predictions and not used in any of the analyses. (*SMW note: this may be slightly misleading because cells with NA predictions will still be used as base grid cells if that model is chosen?) For GIS file types, the predictions can be in any projection, but the longitude coordinates must be between the equivalent of -180 and 180 decimal degrees.

Excel .csv

SDM prediction data can be loaded from an Excel .csv file that has headers and data that meets the following requirements. The data must have longitude and latitude coordinates for each prediction, and these coordinates must be in WGS 84 geographic coordinates (decimal degrees). Longitude values may be in the range [-180, 180] or [0, 360]. The coordinates must be on a regular grid and thus equally spaced, but they can be in any order in the file and represent the center or any of the four corners of the grid cells.

GIS raster files

SDM prediction data can be loaded from a raster TIFF (.tif) file.

GIS shapefile files

Shapefiles are stored as multiple files on local file systems, and thus you must browse to and select all of the files from the shapefile you wish to load into the app. There may be as few as four files to load, or more than ten. The extensions of these files will likely be some subset of the following: ‘.shp’, ‘.shx’, ‘.dbf’, ‘.prj’, ‘.sbn’, ‘.sbx’, ‘.shp.xml’, ‘.fbn’, ‘.fbx’, ‘.ain’, ‘.aih’, ‘.ixs’, ‘.mxs’, ‘.atx’, or ‘.cpg’.

GIS file geodatabase feature class files

To load SDM prediction data from a file geodatabase feature class, you must enter the absolute file path of the file geodatabase, which will end with '.gdb', as well as the name of the feature class object within the file geodatabase. If possible, use ArcCatalog to get both the file path of the file geodatabase and the name of the feature class object. At this time, data cannot be read from personal geodatabases or from file geodatabase raster datasets.

1.4. Supported export file types: SDM prediction data

Any set of SDM prediction loaded into the app or created within the app (overlaid and ensemble prediction) may be exported in the following formats.

Excel .csv

To export SDM predictions to an Excel.csv file, the app calculates the centroid for each polygon that contains a prediction. Thus, the .csv file consists of columns with the longitude and latitudes of these centroids, in the specified projection specified, the predictions, and (if applicable) the weight values. Note that depending on the projection selected and if the prediction polygons have had any land area clipped from them, you may not be able to reload the saved predictions into the app because the centroids will not be equally spaced.

GIS Shapefile

SDM predictions exported as shapefiles will be exported as polygons with their respective prediction and weight values. These polygons will exactly match the polygons in the app. Predictions cannot be exported to either a GIS file geodatabase or a GIS personal geodatabase.

KML files

Not quite sure how to phrase this description yet-to be discussed with TJ

1.5. Supported export file types: image files

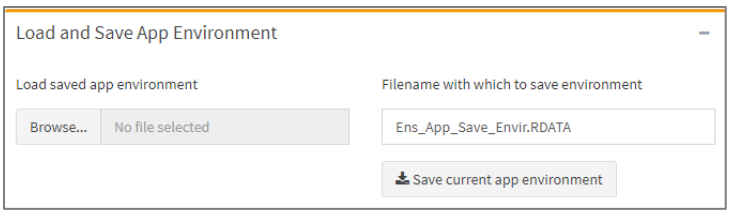
You may download eSDM files as JPG, PDF, or PNG file types at low (72 ppi) or high (300 ppi) resolutions.

2. Load Model Predictions Tab

2.1. Load and Save App Environment

Load eSDM working environment

- If you saved the working environment from a previous eSDM

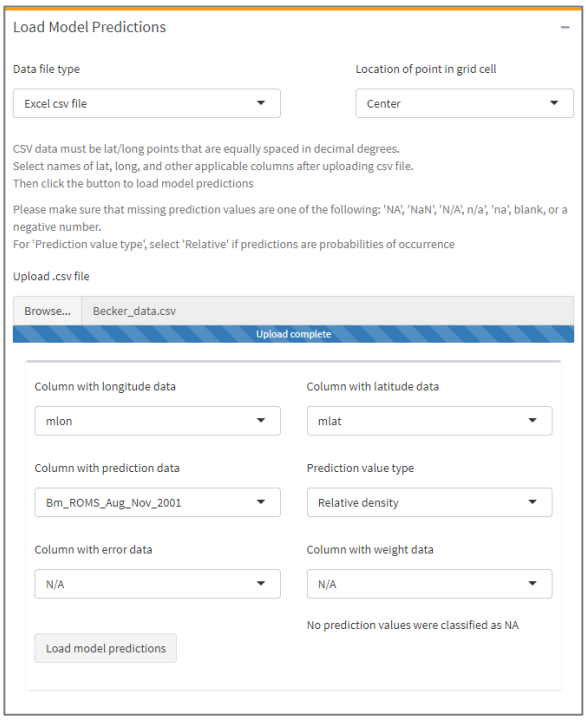


session, you can browse to the corresponding .RDATA file and load it to restore the working environment of the previous session.

Save eSDM working environment

- You may save the current eSDM working environment to an .RDATA file so that during a later eSDM session you can load the file and resume your work where you left off. To save the working environment of the current session, use *Filename with which to save current eSDM working environment* to provide a file name, including the '.RDATA' extension, for the .RDATA file and click *Save current app environment* to download the file to your computer. You can load this file later using *Load saved eSDM session.* Aspects of the eSDM working environment 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



Description: Load SDM predictions into the eSDM so that you can use them in one or more of the subsequent tabs. The *Data file type* pull down menu allows you to specify the type of input file, and adjusts the page to request relevant additional details, as described below.

2.2.1: "Excel csv file"

Description: Select this option if the SDM predictions you wish to load are in an Excel .csv file with columns consisting of longitude, latitude, and model prediction values.

Location of point in grid cell

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

Upload Excel csv file (.csv extension)

- 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, which must be in WGS 84 geographic coordinates (decimal degrees).

Column with latitude data

- Select the name of the column with the latitude data, which must be in WGS 84 geographic coordinates (decimal degrees).

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 predictions 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. Also select "Relative density" if the model predictions are probability of occurrence predictions or habitat suitability indices. 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, inclusive. 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.

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 Excel .csv file into the app.

2.2.2: "GIS raster (GeoTIFF)"

Description: Select this option if the SDM predictions you wish to load are a band of a raster GeoTIFF file. eSDM does not currently support loading model predictions from other raster formats.

Band number of prediction data

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

Upload raster GeoTIFF file (.tif extension)

- 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.

2.2.3: "GIS shapefile"

Description: Select this option if the SDM predictions you wish to load are in a GIS shapefile with at least one associated data column for the predictions values.

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 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.

2.2.4: "GIS file geodatabase (.gdb) file"

Description: Select this option if the SDM predictions you wish to load are in a file geodatabase feature class with at least one associated data column for the predictions values. eSDm does not currently support loading a raster dataset from a file geodatabase or loading a model from a personal geodatabase.

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 will 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 given name in the specified file geodatabase.

Column with prediction data

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

Prediction value type

- See *Prediction value type*

Column with weight data

- See *Column with weight data*

NA prediction values message

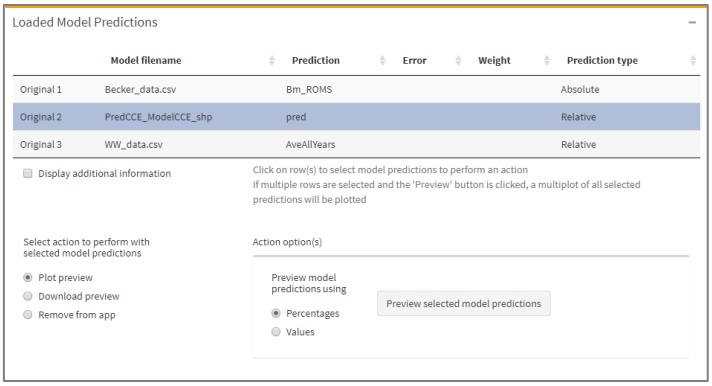
- See *NA prediction values message*

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

Description: This window contains a table with summary information about each loaded model, and can



be used to select one or more models to create or download a plot preview and remove model predictions from the app. You may select model predictions by clicking on the corresponding row in the table; this will highlight the model row in gray-blue. You may select multiple model predictions at one time if desired, but you can only select row(s) when *Display additional information* is unchecked.

2.3.1: *Select loaded model predictions with which to perform an action*

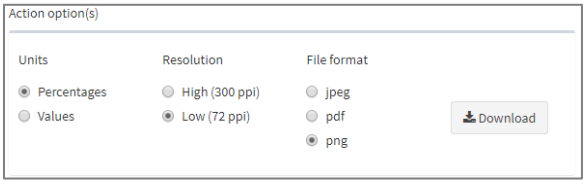
- Select or deselect a set of model predictions by clicking on the row of that set of model predictions in the table. You may select multiple sets of model predictions, and you can perform action(s) with these set(s) of model predictions below the table. If selected, a row is highlighted grey-blue. The table has two sets of information that you may toggle using the *Display additional information* check box. The table illustrated above is shown when the box is unchecked, and displays the name of the file that was loaded, the names of columns from which prediction and weight data were loaded, and the specified prediction type. The 'Weight' column is blank if "N/A" was specified for that column. If *Display additional information* is checked, the table displays the resolution of the model 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 eSDM, the app attempts to determine the resolution at which the predictions were made, meaning the distance between the centroids of adjacent grid cells. eSDM only attempts to calculate the resolution in the native coordinate system of the loaded model predictions; thus if model predictions were generated on an equal area grid but were loaded into eSDM via a shapefile in some form of lat/long coordinates, then eSDM would not be able to calculate the resolution. This functionality is still in a developmental stage, so please check the reported resolution if you know the resolution of your predictions.

2.3.2: *Action to perform with selected model predictions*

Description: Select the action you wish to perform with the set(s) of model predictions selected in the table. These actions include, for one or more sets of model predictions: showing a preview, downloading a preview, or removing them from the app. The options shown in the *Action option(s)* box depend on the selected action, and are described below.

2.3.2.1: "Plot preview"

Units



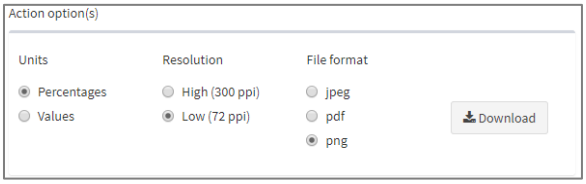
- Select "Percentages" to have the colors of predictions depend on the prediction value relative to the other prediction values in the set of model predictions. For instance, the color scheme will delineate, among other ranges, the top 2% of prediction values, the prediction values in the top 2% to 5%, and the

prediction values in the top 5% to 10% of all of the prediction values. Select "Values" to have the colors of predictions depend 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 preview of multiple sets of predictions.

2.3.2.2: "Download preview"



Units

- See *Units*

Resolution

- The resolution of the downloaded image. It is recommended to use the "High" resolution for multiplot previews.

Image file format

- The file format in which to download the preview of the selected model predictions. The current options are JPEG, PDF, and PNG file types.

Filename

- The filename of the preview when it is downloaded. The filename is reset to the default if different model predictions are chosen to be plotted, or if any of the other inputs 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.

2.3.2.3: "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

Description: A study area polygon restricts the region in which the overlay will be performed to the area covered by the study area polygon. 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 this feature if you have model predictions that cover a broad area, but you only want to create ensemble predictions in a specific study area 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. Use the checkbox to indicate whether or not you want to use a study area polygon, and use the *File type* selection to specify how you wish to load the study area polygon. Additional options shown depend on the specified *File type*.

Use a study area polygon as the boundary for the base grid 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 in blue text when a study area polygon is loaded.

3.1.1: "Excel csv file"

Upload Excel csv file (.csv)

- Browse to and load the file with the .csv extension that contains the desired polygon. 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], and provided in decimal format (e.g. 37.345). 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.

3.1.2: "GIS shapefile"

Upload GIS shapefile files

- See *Upload GIS shapefile files*

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

Full path to file geodatabase

- See *Full path to file geodatabase*

Name of file geodatabase feature class
See *Name of file geodatabase feature class*
Name of file geodatabase feature class
Name of file geodatabase feature class
Name of file geodatabase feature class
Name of file geodatabase feature class
Name of file geodatabase feature class

Upload file geodatabase feature class
See *Upload file geodatabase feature class*

3.2. Load Land Polygon

Description: 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. Therefore, all overlaid models will have the land erased as well. Use the checkbox to indicate whether or not you want to use a land area polygon, and use the *Land polygon source* to specify if you want to use a provided land polygon or load your own. If you are loading your own, then use the *File*

type selection to specify how you wish to load the study area polygon. Additional options shown depend on the *Land polygon source* and *File type* selections.

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.

3.2.1: "Use provided"

Description: If you do not have a land polygon of your own but would like to use one to remove land from the base grid, you may use one of the provided land polygons 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, from highest to lowest, "full", "high", "intermediate", "low", and "crude". Each resolution option is an approximately 80% reduction in size and quality from the previous option.

Load provided land polygon

- Load the GSHHG land polygon with the specified resolution

3.2.2: "Load personal"

Description: Use the *File type* selection to specify how you wish to load the study area polygon. Additional options shown depend on the *File type* selections.

3.2.2.1: "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], and provided in decimal format (e.g. 37.345). 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.

3.2.2.2: "GIS shapefile"

Upload GIS shapefile files

- See *Upload GIS shapefile files*

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

Full path to file geodatabase

- See *Full path to file geodatabase*

Name of file geodatabase feature class
See *Name of file geodatabase feature class*
Name of file geodatabase feature class
Name of file geodatabase feature class
Name of file geodatabase feature class
Name of file geodatabase feature class
Name of file geodatabase feature class

Upload file geodatabase feature class
See *Upload file geodatabase feature class*

3.3. Loaded Model Predictions

Description: See *Description*. The only difference is that in this table you can only select one row because the base grid must come from a single set of model predictions.

Select loaded model predictions to use as the base grid

- See *Select loaded model predictions with which to perform an action*. The only difference is that in this table you can only select one row because the base grid must come from a single set of model predictions.

3.4. Overlay Model Predictions

Description: Specify options used in the overlay process, and then overlay all original model predictions onto the base grid to create overlaid model predictions that you may use to create ensemble predictions. If all of your model predictions were made using the same base grid, then use *Overlay version to perform* to select the same-grid overlay version. This is a simplified version of the standard overlay process designed to overlay the model predictions quicker if their grids are the same. See *Appendix 1*: The overlay process for a more detailed description of the standard overlay proess, as well as how the *Overlay option* inputs are used in the standard overlay process.

3.4.1: "Perform standard overlay

Description: Overlay models with different resolutions and/or coordinate systems onto a single base grid.

Overlay options: projection

Perform overlay in WGS 84 geographic coordinates

- If this box is checks, the base grid and all loaded model predictions will be in WGS 84 geographic coordinates during the overlay process. Thus, the following assumptions will be made for all area and overlap calculations: 1) 'Equatorial axis of ellipsoid' = 6378137 and 2) 'Inverse flattening of ellipsoid' = 1/298.257223563. See this article for more details about assumptions that must be made when calculating the area using WGS 84 geographic coordinates.

Use the coordinate system of the selected model predictions during the overlay

process

- If *Perform overlay in WGS 84 geographic coordinates* is unchecked, then you must specify the coordinate system in which you wish the overlay process to be performed. Currently you may specify this coordinate system by selecting the set of loaded model predictions that was imported with the desired coordinate system.

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 the specified base grid

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

3.4.2: "Perform same-grid overlay"

Description: Overlay models that have the same grid, and thus same resolution and coordinate system. If the model predictions do not meet these requirements, the app will display error messages and will not allow you to perform a same-grid overlay. The app will display warning messages if the loaded model predictions could be overlaid using the same-grid overlay but certain flags are raised.

Perform same-grid overlay

- Overlay all loaded model predictions using the same-grid overlay process.

3.5. Preview of Base Grid

Description: Click *Preview* 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

Description: Click *Preview* to generate a preview of the set(s) of overlaid model predictions selected in *Overlaid model prediction to preview*. 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 was designed 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 wish high quality images of the overlaid model predictions, please use the 'High Quality Maps' tab.

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 table in the *Loaded Model Predictions* section, meaning that the "Overlaid 1" set of model predictions is the "Original 1" model predictions overlaid onto the base grid, and so on. You may select and thus preview multiple sets of model predictions at one time.

4. Create Ensemble Predictions Tab

4.1. Overlaid Model Predictions

Description: Specify the overlaid models that you wish to use when creating ensemble model predictions.

Select overlaid model predictions to ensemble

- If the *Create ensemble using a subset of the overlaid models* check box is unchecked, then all overlaid models will be used in the ensemble. If the box is checks, then you can click on rows of the table to select or deselect the overlaid models you wish to use when creating ensemble model predictions.

4.2. Create Ensemble Predictions – Ensembling method

Description: For all ensembling methods, the ensemble is created by, for each grid cell, performing an average of all the predictions in that grid cell. This average may be an unweighted or weighted average, depending on the *Ensembling method* selection. The displayed options in this section depend on the *Ensembling method* selection, and these options are described below. If any of the overlaid models predictions are NA for a particular cell, then that NA is ignored for that cell and the other predictions are added together.

4.2.1: "Unweighted"
Description: The predictions will be averaged together via an unweighted average (simple mean).

4.2.2: "Weighted"
Description: The predictions will be averaged together via a weighted average. You can specify the weighting method via your *Weighted ensembling method* selection. The options displayed and the process of getting the different weights are described below.

4.2.2.1: "Manual entry"
Description: Weight each set of model predictions by a single value.

Ensemble weights

- The values entered into the text box will be used as weights in the weighted average of the predictions. The first value will be applied to the first set of overlaid model predictions specified in *Select overlaid model*

predictions to ensemble, the second value will be applied to the second set of specified overlaid model predictions, and so on. These values must be numbers greater than or equal to zero that are separated by a single comma and a single space.

4.2.2.2: "Evaluation metric"
Description: Weight each set of model predictions by an evaluation metric value. To use this weighting method, go to the 'Evaluation Metrics' tab and, for all of the overlaid models you wish to use in the ensemble, calculate the metric you wish to use as a weight.

Metric to use for weights

- This section will display the metric(s) that you have calculated for all of the overlaid models you have selected to use in the ensemble. Select the metric that you wish to use as weights for the sets of model predictions. The table displayed shows the metric calculated for each of the overlaid models selected to be in the ensemble, as well as their weights relative to the largest metric.

4.2.2.3: "Pixel-level spatial weights"
Description: Weight each individual model prediction by its own value. 'Pixel-level spatial weights' refer to the weight data specified via the *Column with weight* data entry when the prediction data was loaded into the app in the 'Load Model Predictions' tab. These weights are applied to the prediction in their same pixel. At least one model must have this weight data to use this weighted ensembling method, and the table displayed shows which set(s) of model prediction have pixel-level spatial weights. The set(s) of model predictions that do not have pixel-level spatial weights will be included in the ensemble, but will not be weighted.

4.2.2.4: "Polygon(s) with spatial weigh ts"
Description: Weight model predictions within the area defined by a loaded polygon. You may load polygons, hereafter referred to as 'weight polygons', and specify the weight that will be applied to model predictions that have a specified percentage of their area within that weight polygon. Each weight polygon can be applied to one or more set(s) of overlaid model predictions. Some of the options shown depend on the *File type* selection (sections 4.2.2.4.1 – 4.2.2.4.4). The assigned weight polygons are summarized in a table below the load and assign options. You may also preview the weight polygons for a single set of overlaid model predictions by specifying the overlaid predictions to preview and clicking *Preview* in the 'Polygon(s) with weights method (cont)' box in the right part of the window.

Overlaid predictions to which to apply polygon weights

- Select the set(s) of overlaid predictions to which to apply the loaded weight polygon when you click *Assign loaded weight polygon to selected predictions*.

Weight for ___ polygon(s)

- The weight that will be applied to the model predictions within the weight polygon.

Percentage of overlap…

- Specify the percentage of a model prediction grid cell that must be overlapped by the weight polygon for that weight polygon to be applied to that model prediction.

Assign loaded weight polygon to selected predictions

- Click this button to assign the loaded weight polygon to the selected set(s) of overlaid model predictions. These changes will be reflected in the summary table.

Remove selected weight polygons

- Click to remove the weight polygons selected in *Select loaded weight polygon(s) to remove* from the app. These changes will be reflected in the summary table.

4.2.2.4.1: "Excel .csv"

Upload Excel .csv file (.csv extension)

See *Upload Excel csv file (.csv extension)*

4.2.2.4.1: "GIS raster GeoTIFF"

Upload raster GeoTIFF file (.tif extension)

See *Upload raster GeoTIFF file (.tif extension)*

4.2.2.4.3: "GIS shapefile"

Upload GIS shapefile files

- See *Upload GIS shapefile files*

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

Full path to file geodatabase

- See *Full path to file geodatabase*

Name of file geodatabase feature class

See *Name of file geodatabase feature class*

Name of file geodatabase feature class

Name of file geodatabase feature class

Name of file geodatabase feature class

Name of file geodatabase feature class

Name of file geodatabase feature class

Upload file geodatabase feature class

See *Upload file geodatabase feature class*

4.3. Creating Ensemble Predictions – Rescaling method

Description: If any of the loaded model predictions have a prediction type of 'Relative abundance' rather than 'Absolute abundance', then you will have to select a rescaling method. This means that the model predictions will be rescaled in using one of the provided methods before they are ensembled together via a weighted or unweighted average.

4.3.1: "None"

Description: Overlaid predictions will not be rescaled before being ensembled together. This option is only available if all prediction types are 'Absolute density'.

4.3.2: "Abundance"

Description: Overlaid predictions will be rescaled so that the predicted abundance for each set of overlaid model predictions is the value entered in *Abundance to which to rescale predictions*. This value must be greater than zero.

4.3.3: "Normalization"

Description: Overlaid predictions will be rescaled so that each set of overlaid predictions has a range of [0, 1]. Normalization uses the following formula, where X is the overlaid predictions: $X_{new} = (X - X_{min}) / (X_{max} - X_{min})$

4.3.4: "Standardization"

Description: Overlaid predictions will be rescaled so that each set of overlaid predictions has a mean (μ) of 0 and a standard deviation (σ) of 1. Standardization uses the following formula, where X is the overlaid predictions: $X_{new} = (X - \mu) / \sigma$

4.3.5: "Sum to 1"

Description: Overlaid predictions will be rescaled so that they sum to one for each model.

4.4. Creating Ensemble Predictions – Create

Description: Click *Create ensemble* to create an ensemble using your specified ensembling and rescaling method inputs.

4.5. Created Ensemble Predictions

Description: This window contains a table with summary information about each set of created ensemble predictions, and can be used to select one or more ensembles to create or download a plot preview, remove from the app, or calculate the predicted abundance of the ensemble. You may select ensemble predictions by clicking on the corresponding row in the table; this will highlight the model row in gray-blue. You may select multiple model predictions at one time if desired.

This section has the same actions to perform and action option(s) as those described under *Select loaded model predictions with which to perform an action*, except you also have the

option to calculate the predicted abundance for the selected ensembles. However, you only can calculate the predicted abundance if the rescaling method was "None" or "Abundance".

5. Evaluation Metrics Tab

5.1. Select Predictions to Evaluate

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

5.2. Load Validation Data

Description: Load validation data into the eSDM that you wish to use to evaluate original, overlaid, or ensemble model predictions. You can specify the type of input file via the *Validation data file type* menu, and you can specify the type of validation data via the *Validation data type* menu. The app adjusts the page to request relevant additional details depending on the file type and data type selected. A blue 'Validation data loaded' message will be displayed if validation data is loaded in the app.

5.2.1: "Excel .csv file"

Upload Excel .csv file

See *Upload Excel csv file (.csv extension)*

-

Select, in this order, the longitude, latitude, and validation data column for the uploaded .csv file

- When you click on the input box, the app will display the headers of all of the columns in the uploaded file. Select, in order, the headers that correspond to the longitude column, the latitude column, and the column that contains the validation data. You can delete selected column names from the input box by selecting them and clicking your computer's delete button. If *Validation data type* is "Counts (numerical)", then the data column must consist of numbers.

5.2.2: "GIS shapefile"

Upload GIS shapefile files

- See *Upload GIS shapefile files*

Select the validation data column for the uploaded GIS file

- Select the name of the data column in the shapefile that contains the validation data. If *Validation data type* is "Counts (numerical)" then the data column must consist of numbers.

5.2.3: "GIS file geodatabase (.gdb) file"

Full path to file geodatabase

- See *Full path to file geodatabase*

Name of file geodatabase feature class

See *Name of file geodatabase feature class*
Name of file geodatabase feature class
Name of file geodatabase feature class
Name of file geodatabase feature class
Name of file geodatabase feature class
Name of file geodatabase feature class

Upload file geodatabase feature class
See *Upload file geodatabase feature class*

Select the validation data column for the uploaded GIS file

- Select the name of the data column in the shapefile that contains the validation data. If *Validation data type* is "Counts (numerical)" then the data column must consist of numbers.

"Presence or absence" selected
Description: You must specify the code(s) that specify presence observations using the *Select presence code(s)* input, and the code(s) that specify absence observations using the *Select absence code(s)* input. You must classify all of the codes as either presence or absence codes, and you cannot classify a code as both a presence and an absence code.

Load specified validation data into app

- Load the validation data into the eSDM with the specified inputs.

5.3. Calculate Metrics

Description: Get information about the loaded validation data, select the metrics you wish to calculate, and then calculate the metrics. The *Validation data info* table displays information about the data specified as the validation data column in the 'Load Validation Data' section.

Metric(s) to calculate

- Check the respective boxes to calculate area under the curve (AUC), True skill statistic (TSS), and/or root mean squared error (RMSE). You may only calculate RMSE if the loaded validation data is "Count (numerical)". See Fielding and Bell 1997 for more information about AUC, Allouche *et al.* 2006 for more information about TSS, and this page for more information about RMSE.

Calculate metrics

- Click this button to calculate the selected metrics for the specified set(s) of model predictions using the loaded validation data.

5.4. Metrics

Description: This section displays the calculated metrics for the specified models after you click *Calculate metrics*. You may also download the metric values.

Download metrics

- Click to download an Excel .csv file that contains the metric value(s) and model information for the models for which the metrics were calculated. Because ensemble predictions have different information than original and overlaid predictions, if evaluation metrics have been calculated for both ensemble predictions and at least one of original and overlaid predictions, then some column headers will be formatted as 'Original+Overlaid info name/Ensemble info name'.

5.5. Metric Descriptions and References

Description: This section provides additional information about the metrics you can calculate.

6. High Quality Maps Tab

6.1. Select Predictions to Map

Description: Click on the row of the original, overlaid, and/or ensemble model predictions to select the one you wish to map. Currently, you can only create a map of one set of model predictions at a time.

6.2. Map Control

6.2.1. Generate map in-app

Generate map

- Click this button to generate a map in the box above with the parameters specified in the 'Map Parameters' sections. Plotting a large set of model predictions may take several minutes.

6.2.2. Download map

Resolution

- The resolution of the downloaded image. It is recommended to use the "High" resolution for multiplot previews.

Image file format

- The file format in which to download the preview of the selected model predictions. The current options are JPEG, PDF, and PNG file types.

Filename

- The filename of the preview when it is downloaded. The filename is reset to the default if different model predictions are chosen to be plotted, or if *Resolution* or *Image file format* are changed.

Download map

- Click this button to generate a map in the box above with the parameters specified in the 'Map Parameters' sections. Downloading a large set of model predictions may take several minutes.

6.3. Map Parameters – Section 1

6.3.1. Map coordinate system and range

Description: Control the coordinate system of the generated or downloaded map, as well as the range of the map. Note that if the specified coordinate system is degree-based, then map range values must be entered as decimal degrees with the range [-180, 180] for the longitude values and [-90, 90] for the latitude values. If the specified coordinate system is not degree-based, then the map range values must be entered as the unit of the specified coordinate system.

Generate map in WGS 84 geographic coordinates (decimal degrees)

- If checked, the map will be generated in WGS geographic coordinates (decimal degrees). If unchecked, you may specify the coordinate system in which to generate the map with *Filename of original model predictions with desired projection for map*

Filename of original model predictions with desired coordinate system for map

- The map will be generated in the native coordinate system of the specified original model prediction.

Longitude minimum

- The left-most limit of the map. This value have the same units as the specified coordinate system, and in the range [-180, 180].

Longitude maximum

- The right-most limit of the map. This value have the same units as the specified coordinate system, and in the range [-180, 180].

Latitude minimum

- The bottom-most limit of the map. This value have the same units as the specified coordinate system, and be in the range [-90, 90].

Latitude maximum

- The top-most limit of the map. This value have the same units as the specified coordinate system, and be in the range [-90, 90].

6.3.2. Title and axis labels

Map title

- The title of the map, which will be displayed at the top of the image. Currently this title may be only one line. Leave this box blank if you do not wish to have a title.

X-axis label

- The label of the x-axis, which will be displayed at the bottom of the map.

Leave this box blank if you do not wish to have an x-axis label.

Y-axis label

- The label of the y-axis, which will be displayed on the left of the map. Leave this box blank if you do not wish to have an x-axis label. Currently this label cannot be displayed with horizontal text.

Title size

- The size of the title, relative to one. For instance, a size entry of 1.1 makes the title slightly bigger than the default size, which a size entry of 0.4 makes the title much smaller than the default size.

Axis label size

- The size of the axis labels, relative to one. For instance, a size entry of 1.1 makes the labels slightly bigger than the default size, which a size entry of 0.4 makes the labels much smaller than the default size.

6.3.3. Tick marks and tick labels

Description: Control whether tick marks will be displayed on the map, and if so where they will be. There are no minor ticks, and tick marks and tick labels will only be displayed on the bottom and left sides of the map. If tick marks are displayed, then you also control whether or not tick labels are generated at the specified ticks. If ticks are not displayed, then tick labels cannot be displayed either.

Include tick marks in the map

- Check the box to include tick marks in the map. The other options for tick marks and tick labels are only displayed if this box is checked.

Tick location options

- Specify whether you want the tick locations to be the default tick locations determined by R ("Use default tick locations") or if you want to specify the tick locations yourself ("Enter tick locations manually").

Longitude tick locations

- Only displayed if "Enter tick locations manually" is selected. Enter the longitude values at which you want tick marks. These values must have the same unit as the coordinate system specified in the 'Map coordinate system and range' box, and must be numbers separated by ", " (e.g. "-135, -130, -115"). If the unit of the specified coordinate system is degrees, then the values must be decimal degrees and be in the range [-180, 180].

Longitude tick locations

- Only displayed if "Enter tick locations manually" is selected. Enter the latitude values at which you want tick marks. These values must have the same unit as the coordinate system specified in the 'Map coordinate system

and range' box, and must be numbers separated by ", " (e.g. "25, 30, 40"). If the specified coordinate system's unit is degrees, then the values must be decimal degrees and be in the range [-90, 90].

Tick length

- The length of the ticks, relative to one. For instance, a size entry of 1.1 makes the ticks slightly longer than the default size, which a size entry of 0.4 makes the ticks much smaller than the default size.

Include tick labels in the map

- Check the box to include tick labels in the map. These tick labels will be displayed wherever there are tick marks.

Tick label size

- The size of the tick labels, relative to one. For instance, a size entry of 1.1 makes the tick labels slightly bigger than the default size, which a size entry of 0.4 makes the tick labels much smaller than the default size.

6.4. Map Parameters – Section 2

6.4.1. Color scheme of predictions

Description: Specify the units of the color scheme of the predictions in the map, as well as the color palette and, if applicable, the number of colors 'bins' into which to divide the predictions. The *Color scheme preview* is a preview of the specified color palette with the specified number of colors on the right side of the box. The lowest number in the preview corresponds to the first color bin (highest 2% or largest value), and so on.

Prediction display option

- Specify how you want to color-code the predictions. 1) If you select "Color-code predictions by relative percentage", then the predictions will be split into 10 bins: highest 2%, 2% - 5%, 5% - 10%, 10% - 15%, 15% - 20%, 20% - 25%, 25% - 30%, 30% - 35%, 35% - 40%, and lowest 60%. 2) If you select "Color-code predictions by numerical value", then the predictions will be split into *Number of colors* bins of equal size.

Color palette

- Select the color palette that you wish to use to color the model predictions. The palette names are formatted as "'R package the palette came from': 'Name of color palette'". The 'Default: blue to white to red' color palette was made by hand and did not come from an R package. Some of the palettes have requirements for the number of colors used with them; thus, if "Color-code predictions by relative percentage" is selected, then you cannot select some of the palettes because the "Color-code predictions by relative percentage" selection require that there be ten colors.

Number of colors

- Enter the number of colors that you wish to use, and thus the number of bins into which the predictions will be split. This input option is only available when "Color-code predictions by numerical value" and specific palettes are selected, because there are ten relative percentage bins and some of the palettes require a specific number of bins. Also, some of the palettes have a min and max number of colors; if those palettes are selected, then there is a '(Min: #; Max: #)' text display after the *Number of colors* label.

6.4.2. Background color and legend

Click to select background color

- When you click the input rectangle, a pop-up window will appear. Click in this window to select the color you want the background of the map to be. Background area is non-prediction and non-land area.

Reset background color to white

- Click this button to set the color specified in *Click to select background color* to white.

Include legend with map

- Check this box to plot a legend to one side of the map

Legend position

- Specify the side of the map on which the legend is displayed. The options are "Right", "Bottom", "Left", and "Top".

Legend labels: number of decimals

- Only displayed if color code predictions by numerical value" is selected. Specify the number of decimals to display in the legend labels.

6.4.3. Additional polygons

Description:

Include additional polygons in the map

- Check in order to include the study area and/or land polygon in the map

Include selected polygons in the map

- Check in order to include the study area and/or land polygon in the map

Include additional polygons in map

- Check in order to include the study area and/or land polygon in the map

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.