Rank Calculator Prep Toolbox

The Rank Calculator Prep Toolbox for ArcGIS Pro provides custom tools designed to increase the efficiency of preparing input species datasets and obtaining spatial statistics needed for use with NatureServe's Conservation Rank Calculator. The following guide includes general instructions for setting up and using the tools, as well as limitations to consider when deciding whether these tools are appropriate for your use.

Toolbox name: Rank Calculator Prep.pyt
Toolbox access: https://github.com/PNHP/DataManagement/tree/master/RankCalculator
Version 1.0 release date: 11/21/2023
Author: Molly Moore, Pennsylvania Natural Heritage Program

1) When to use these Tools

If tasked with calculating the conservation status of an element (or multiple elements) as a part of NatureServe methodology, you will likely be using NatureServe's Conservation Rank Calculator Tool. As a part of the Rank Calculator Tool, you will need to compile spatial data and provide spatial statistics about element range, occupancy, and occurrence numbers. If you are adept with ArcGIS Pro, compiling the spatial data, grouping observations into occurrences, and calculating the necessary spatial statistics requires relatively simple GIS operations. However, manually performing these GIS operations can be time intensive, especially if you are running the Rank Calculator Tool for several species. The Rank Calculator Prep Toolbox is intended to make the process of compiling spatial data and calculating spatial statistics within ArcGIS Pro more efficient. Please note that the use of these tools does not supersede QA/QC of your data sources and it is recommended that you establish a workflow that allows for consistent geospatial processing and comprehensive quality checks.

Best practices include:

1. Ensure all input datasets are appropriately projected into a consistent metric coordinate system.
2. Ensure all input datasets are of the same geometry.
3. Ensure records (especially if downloaded from community science databases) are within the elements reasonable range to avoid gross errors in range extent.
4. Run the 'Check Geometry' tool on input datasets and repair when necessary to avoid errors.
5. Ensure there are no special characters in the species identifier field.
6. Ensure there are no NULL values in the species identifier field and no NULL values in the separation distance field.

2) Accessing and setting up the Rank Calculator Prep Toolbox

1. If you are familiar with GitHub, you can fork or clone the PNHP/DataManagement repository to your local machine. Use the ArcGIS Pro Catalog to open the folder where you forked or cloned the repository. The Rank Calculator Stats Toolbox will be in the Rank Calculator folder and ready to use. If you are not familiar with GitHub, use the following steps to set up the toolbox.

2. Open a web browser and go to: https://github.com/PNHP/DataManagement/blob/master/RankCalculator/Rank%20Calculator%20Prep.pyt
3. In the upper-right corner of the code block, click the "Download raw file" button 
4. The toolbox will now be in your Downloads folder. If you wish, copy and paste the toolbox to another location on your machine.
5. Open ArcGIS Pro and in your Catalog, right-click the 'Toolboxes' folder and select 'Add Toolbox'.
6. Navigate to the location where you saved the toolbox, click the toolbox, and click "Okay".
7. Your toolbox is now ready to use just like any ArcGIS Pro toolbox!

3) Occurrence Grouping Tools

The occurrence grouping tools can be used to compile one or more input datasets into a single merged dataset and group species/element observations into occurrence groups based on separation distance.

3.1 Occurrence Grouping – Terrestrial and Lentic Species Tool

This tool merges input datasets and groups observations into occurrences using straight-line separation distance. Occurrence groups are identified by an Occurrence ID field in the output dataset. The data and parameters needed to run the tool include:

1. One or more feature classes of element observations (projected into a consistent metric coordinate system). These datasets must include:
   a. A field that uniquely identifies the species/element, and is consistently named among all input datasets.
   b. A field that designates the species/element separation distance in kilometers.
2. The geodatabase location and name of the output feature class.

3.2 Occurrence Grouping – Riverine Species Tool

This tool merges input datasets and groups observations into occurrences using separation distance along a stream network. Occurrence groups are identified by an Occurrence ID field in the output dataset. In addition to the above data and parameters needed for the Terrestrial and Lentic Species Tool, the following are needed to run the Riverine Species tool:

1. Network dataset built from a flowline feature class.
2. Snap distance in meters, which is the distance to the flowline beyond which observations will not be grouped.

LIMITATIONS: The accuracy of occurrence grouping along a river network is heavily dependent on the input spatial data. Occurrence grouping using a river network often performs poorly for observations located in large rivers, reservoirs, wetlands, headwater streams, and other waterbodies that are not well represented by linear features. For species that primarily occur in these kinds of aquatic habitats, it is recommended that you use the Terrestrial and Lentic Species Tool or consider alternatives for occurrence grouping.

3.3 Post Occurrence Grouping QC

After running the occurrence grouping tools, it is good practice to check the results and resolve any issues prior to running the Rank Calculator Stats Tool.

1. General separation distance QC –manually inspect all occurrence groups to ensure groupings look correct. To do this, select all records that have the same Occurrence ID, zoom to the records, and evaluate accuracy of grouping.
2. Separation barriers QC – Manually inspect all occurrence groups against aerial imagery to decide whether there are any separation barriers that need to be considered.
3. Null values for Occurrence ID after running the Riverine Tool – If a record contains a null value for the Occurrence ID after running the Occurrence Grouping – Riverine Species Tool, it means that the record was further away from a flowline than the snap distance set by the user in the parameters. Zoom to that record and decide whether it should be included in an occurrence group and update the Occurrence ID manually. **NOTE**: It is best to resolve null observation IDs before running the Rank Calculator Stats Tools. Records with null Occurrence IDs will NOT be counted toward the occurrence numbers, but will be counted toward the total observation numbers, and the range extent and area of occupancy calculations.

4) Rank Calculator Stats Tool

The Rank Calculator Stats Tool can be used to calculate the range extent, occupancy area, number of occurrences, and number of observations for one or multiple species/elements.

4.1 Running the Rank Calculator Stats Tools

The data and parameters needed to run these tools include:

1. An input feature class that includes all observations for one or multiple species/elements. This dataset must include:
   a. A field that uniquely identifies the species/elements.
   b. Optionally, a field that identifies occurrence groups. If no field is chosen that identifies occurrence groups, the occurrence number will equal the observation number.
2. Range extent type selection. This determines the method by which the range extent will be calculated. For more information about choosing the appropriate range extent type for your data, see NatureServe Conservation Status Assessments: Factors for Evaluating Species and Ecosystem Risk.
3. If “Occupied HUC Watershed” is chosen for the Range Extent Type, a HUC Watershed layer is required.
4. An optional feature layer of a subnational boundary that is used to clip the range extent. If left blank, the range extent may extend beyond the subnational boundary and this area may be calculated in the range extent.
5. Grid size area selection. This determines the size of the grid squares used for calculating the area of occupancy. For more information about choosing the appropriate size of grid squares for

your data, see NatureServe Conservation Status Assessments: Factors for Evaluating Species and Ecosystem Risk.

6. The geodatabase location and name of the output table containing the rank calculator stats for all species.
7. An optional location and name for an output feature class with range extent polygons that were used for range extent area calculations. If left blank, this will not be saved.

4.2 Rank Calculator Stats Tools Output

The output table from the Rank Calculator Stats Tools will include the following four fields:

1. range_extent_km2 – If "Convex Hull" is chosen as the Range Extent Type, range extent is calculated as the area of the convex hull polygon generated around all occurrences in square kilometers. If "Occupied HUC Watershed" is chosen as the Range Extent Type, range extent is calculated as the sum of areas of occupied HUC watersheds in square kilometers. If a subnational boundary is used, the range extent is calculated as the area after it is clipped to the subnational boundary layer.
2. occupancy_area_km2 – Occupancy area is calculated as the sum of the area of occupied grid cells in square kilometers, using a random grid surface generated by the tool. The size of the grid cells is chosen by the user.
3. occurrence_num – The occurrence number is calculated as the count of unique values in the Occurrence ID field. If this optional parameter is not used, the occurrence number will be equal to the observation number. If this optional parameter is used, records with null Occurrence IDs will NOT be counted toward the occurrence numbers, but will be counted toward the total observation numbers.
4. obs_num –The observation number is calculated as the count of observations, or total records, for the species.