Esri Green Infrastructure Data Description
Updated: March 15, 2017

Summary:
This map package contains intact habitat cores (minimally disturbed natural areas at least 100 acres in size and greater than 200 meters wide) and the source data used to derive them for one state.

The package is a replacement to the original package and in addition to all the datasets in the initial version it includes:

- New data layers:
  - Intact habitat cores by betweenness
  - Fragments (natural areas that did not meet the size criteria)
  - Least cost paths connecting cores layer
  - Cost surface used to find the least cost paths connecting the cores
- New core attributes:
  - Two new Betweenness Centrality attributes for each core. The core geometry remains the same.
- New data structure
  - All source data in Data.gdb
  - Derived data in DerivedGIData.gdb
  - Raster datasets clipped to a 10km buffer of the state border

Purpose:
Esri's Green Infrastructure Initiative is a collection of authoritative geospatial resources, newly generated data, online applications, and downloadable models with the aim of empowering local organizations engaged in GI work, all while initiating a national vision of GI planning.

This map package provides data related to the newly generated national database of intact habitat cores and the source data used to generate these data. Its goal is to support green infrastructure planning efforts at national, regional and local scales and was constructed using the best available national data – typically at a resolution of 30 meters. These data should be supplemented and/or replaced with more current or higher resolution data when available, especially for local analysis and mapping.

Please see visit the Esri Green Infrastructure website for more information about these data, access to the models utilized to generate them, and additional online applications for green infrastructure planning.

Description:
Following a methodology outlined by the Green Infrastructure Center Inc., Esri staff created a national intact habitat cores database for the lower 48 United States which were generated from the 2011 National Land Cover Data. Cores were derived from all "natural" land cover classes and excluded all "developed" and "agricultural" classes including crop, hay and pasture lands. The resulting cores were tested for size and width requirements (at least 100 acres in size and greater than 200 meters) and then converted into unique polygons. This process resulted in generation of over 550,000 intact habitat cores.

These polygons were then overlaid with a diverse assortment of physiographic, biologic and hydrographic layers to populate each core with attributes (53 in total) related to the landscape characteristics found within. These data were also compiled to compute a "core quality index", or score related to the perceived ecological value of each core, to provide users with additional insight related to the importance of each core when compared to all others.

Source data:
Number of Endemic Species (Mammals, Fish, Reptiles, Amphibians, Trees) (Jenkins, Clinton N., et. al, (April 21, 2015) *US protected lands mismatch biodiversity priorities*, PNAS vol.112, no. 16, www.pnas.org/cgi/doi/10.1073/pnas.1418034112)

Biodiversity Priority Index Areas: Endemic species, small home range size and low protection status. (Jenkins, Clinton N., et. al, (April 21, 2015) *US protected lands mismatch biodiversity priorities*, PNAS vol.112, no. 16, www.pnas.org/cgi/doi/10.1073/pnas.1418034112)

Unique Ecological Systems (based upon work by Aycrigg, Jocelyn L, et. al. (2013) Representation of Ecological Systems within the Protected Areas Network of the Continental United States. PLos One 8(1):e54689). New data constructed by Esri staff, using TNC Ecological Regions as summary areas.

Ecologically relevant landforms (Theobald DM, Harrison-Atlas D, Monahan WB, Albano CM (2015) *Ecologically-Relevant Maps of Landforms and Physiographic Diversity for Climate Adaptation Planning.* PLoS ONE 10(12): e0143619. doi:10.1371/journal.pone.0143619 ,http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0143619

Local Landforms (produced 3/2016) by Deniz Karagulle and Charlie Frye, Esri, 30 m* resolution. "Improved Hammond's Landform Classification and Method for Global 250-m Elevation Data" by Karagulle, Deniz; Frye, Charlie; Sayre, Roger; Breyer, Sean; Aniello, Peter; Vaughan, Randy; Wright, Dawn, has been successfully submitted online and is presently being given consideration for publication in Transactions in GIS.
*we scaled the neighborhood windows from the 250-meter method described in the paper, and then applied that to 30-meter data in the U.S.

Ecological Land Units (Sayre, R., J. Dangermond, C. Frye, R. Vaughan, P. Aniello, S. Breyer, D. Cribbs, D. Hopkins, R. Nauman, W. Derrenbacher, D. Wright, C. Brown, C. Convis, J. Smith, L. Benson, D. Paco VanSistine, H. Warner, J. Cress, J. Danielson, S. Hamann, T. Cecere, A. Reddy, D. Burton, A. Grosse, D. True, M. Metzger, J. Hartmann, N. Moosdorf, H. Dürr, M. Paganini, P. DeFourny, O. Arino, S. Maynard, M. Anderson, and P. Comer. 2014. A New Map of Global Ecological Land Units — An Ecophysiographic Stratification Approach. Washington, DC: Association of American Geographers. 46 pages) http://www.aag.org/cs/global_ecosystems. 2015 updated data accessed 3/2016, 250 m resolution.

National Elevation Dataset, USGS, 30 m resolution, http://viewer.nationalmap.gov/launch/

NWI – National Wetlands Inventory "Classification of Wetlands and Deepwater Habitats of the United States". U.S. Department of the Interior, Fish and Wildlife Service, Washington, DC. FWS/OBS-79/31, U.S. Fish and Wildlife Service, Division of Habitat and Resource Conservation (prepared 10/2015)

NLCD 2011 – National LandCover Database 2011 http://www.mrlc.gov/nlcd2011.php (downloaded 1/2016) Homer, C.G., et. al. 2015,Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. *Photogrammetric Engineering and Remote Sensing*, v. 81, no. 5, p. 345-354

NHDPlusV2 - https://www.epa.gov/waterdata/nhdplus-national-hydrography-dataset-plus Received from Charlie Frye, Esri 3/2016. Produced by the EPA with support from the USGS.

gSSURGO - Soil Survey Staff. Gridded Soil Survey Geographic Database for the Conterminous United States. Natural Resources Conservation Service, United States Department of Agriculture. Available online at http://gdg.sc.egov.usda.gov/. Accessed 3/2016, 30 m resolution

SSurgo - Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey. Available online at http://websoilsurvey.nrcs.usda.gov/. Accessed 3/2016

GAP Level 3 Ecological System Boundaries (downloaded 4/ 2016)
http://gapanalysis.usgs.gov/gaplandcover/data/download/

NOAA CCAP Coastal Change Analysis Program Regional Land Cover and Change
downloaded by state (3/2016) from: **https://coast.noaa.gov/ccapftp/#/**
Description: https://coast.noaa.gov/dataregistry/search/collection/info/ccapregional
30 m resolution, 2010 edition of data

NHD USGS National Hydrography Dataset http://nhd.usgs.gov/data.html

TNC Terrestrial Ecoregions http://maps.tnc.org/gis_data.html#TNClands (downloaded 3/2016)

2015 LCC Network Areas https://www.sciencebase.gov/catalog/item/55b943ade4b09a3b01b65d78

Derived Data:

Intact Habitat Cores - Cores were derived from the 2011 National Land Cover dataset. The "natural" land cover classes were included and all "developed" and "agricultural" classes including crop, hay and pasture lands were excluded. The resulting cores were tested for size and width requirements (at least 100 acres in size and greater than 200 meters) and then converted into unique polygons.

Fragments – Fragments are those natural areas that for size or shape reasons cannot be classified as a core. While smaller than a core, they are still very important areas for connectivity as well as habitat restoration.

Cost surface - A cost surface was generated using a raster overlay process to create a composite comprised of several landscape variables. Characteristics within each variable were scored based on their perceived impact on species movement whereas reduced movement is reflected as high cost. Landscape variables were categorized into three themed classes based on their expected influence on the cost surface. The first class, impedance, represents the expected cost of species movement as it relates to land cover. The second class, bonuses, represents reductions in cost resulting from being within an existing core, fragment or proximal to surface water; these conditions are assumed to enhance movement. The third class, penalties, represents increases in cost resulting from steeply sloping terrain and road infrastructure; these conditions are assumed to discourage movement.

Least Cost Path - The newly developed national connectivity dataset depicts a network of least cost path (LCP) connections between neighboring cores. This LCP network was used to calculate a Betweenness Centrality (BC) attribute for each core. Esri generated this comprehensive network of LCPs using the Cost Connectivity tool which was introduced in ArcGIS 10.4 and ArcGISPro in 1.3. The inputs and outputs of this operator are described in the Cost Connectivity Help file. http://pro.arcgis.com/en/pro-app/tool-reference/spatial-analyst/cost-connectivity.htm

Betweenness Centrality Measure – Betweenness Centrality (BC) provides a metric depicting each core's importance in the network. BC represents the number of paths that flow through a given habitat core normalized by the total number of shortest paths between all pairs of nodes, except those paths connecting immediate neighbors. High BC values reflect greater use of that core in traversing the network, thus elevating the core's importance in facilitating connectivity when compared to cores of lower value.

The NetworkX python library was used to compute BC based on the optional neighboring connections output from Cost Connectivity. The networkx.betweenness_centrality_source function was implemented based largely on the Parallel Betweenness advanced example provided on the NetworkX site. Click here for additional information about NetworkX.

Evaluation:

The creation of a national core quality index is a very ambitious objective, given the extreme variability in ecosystem conditions across the United States. The additional attributes were intended to provide flexibility in accommodating regional or local environmental differences across the U.S.

Scripts for constructing local cores and scoring them using the Green Infrastructure Center's methodology are available on http://www.esri.com/about-esri/greeninfrastructure

Two general approaches were used in the developing core quality index values. The first (default) follows the guidance of the Green Infrastructure Center's scoring approach developed for the southeastern US where size of the core is the primary determinant of quality. The second; Bio-Weights puts more emphasis on bio-diversity and uniqueness of ecosystem type and de-emphasizes slightly the importance of core size. This is to compensate for the very large intact core habitat areas in the west and southwest which also have comparatively low biodiversity values.

Detailed Description of the Core Quality Index Methodology:
The creation of a national core quality index was a challenging undertaking given the extreme heterogeneity of ecosystem conditions across the United States. As a result, 9 separate index scores were generated for each core, each placing varying weights on landscape characteristics of regional or local significance. This was done to account for variation across the U.S. and to provide users with additional flexibility in accommodating regional or local environmental priorities.

Scripts for constructing local cores and scoring them using the Green Infrastructure Center's methodology are available on http://www.esri.com/about-esri/greeninfrastructure

Two general approaches were used in developing the core quality index values. The first, Default Weights, uses core size as the primary determinant of quality, following the guidance of the Green Infrastructure Center's scoring approach developed for the southeastern US. The second, Bio-Weights, puts more emphasis on characteristics associated with bio-diversity and uniqueness of ecosystem types and de-emphasizes slightly the importance of core size. This alternative was developed to compensate for the very large intact core habitat areas in the west and southwest which also have comparatively low biodiversity values.

Scoring values:

Default Weights

0.4, # Acres
0.1, # Thickness
0.05, # Topographic Diversity (Standard Deviation)
0.1, # Biodiversity Priority Index (Species Richness in GIC original version)
0.05, # Percentage Wetland Cover
0.03, # Ecological Land Unit – Shannon-Weaver Index (Soil Variety in GIC original version)
0.02, # Compactness Ratio (Area relative to the area of a circle with the same perimeter length)
0.1, # Stream Density (Linear Feet/Acre)
0.05, # Ecological System Redundancy (Rare/Threatened/Endangered Species Abundance (Number of occurrences) in GIC original version)
0.1, # Endemic Species Max (Rare/Threatened/Endangered Species Abundance (Number of unique species in a core) in GIC original version)

Bio-Weights

0.2, # Acres
0.1, # Thickness
0.05, # Topographic Diversity (Standard Deviation)
0.25, # Biodiversity Priority Index (Species Richness in GIC original version)
0.05, # Percentage Wetland Cover
0.03, # Ecological Land Unit – Shannon-Weaver Index (Soil Variety in GIC original version)
0.02, # Compactness Ratio (Area Relative To The Area Of A Circle With The Same Perimeter Length)
0.1, # Stream Density (Linear Feet/Acre)
0.1, # Ecological System Redundancy (Rare/Threatened/Endangered Species Abundance (Number of occurrences) in GIC original version)
0.1, # Endemic Species Max (Rare/Threatened/Endangered Species Diversity (Number of unique species in a core) in GIC original version)

Key datasets for Green Infrastructure planning include wildlife diversity, rarity and abundance information. These data are frequently extremely localized in their coverage and often quite sensitive with regards to rare and endangered species information. As a result they are not typically available for national, landscape or regional analysis.

As surrogates for the variables "*Rare/Threatened/Endangered Species Abundance*" and "*Rare/Threatened/Endangered Species Diversity*" in the Green Infrastructure Center's methodology, we used data from the BiodiversityMapping.org site. Specifically their **Priority Areas for USA Biodiversity Conservation** data layer and a composite file summarizing the number and distribution of their endemic **mammal, reptile, amphibian, freshwater fish, and trees** files in the US.

The research that produced these data is described in detail in the scientific paper, "U.S. protected lands mismatch biodiversity priorities," by Clinton N. Jenkins, Kyle S. Van Houtan, Stuart L. Pimm, and Joseph O. Sexton. Published April 6, 2015, in *Proceedings of the National Academy of Sciences*. www.pnas.org/cgi/doi/10.1073/pnas.1418034112

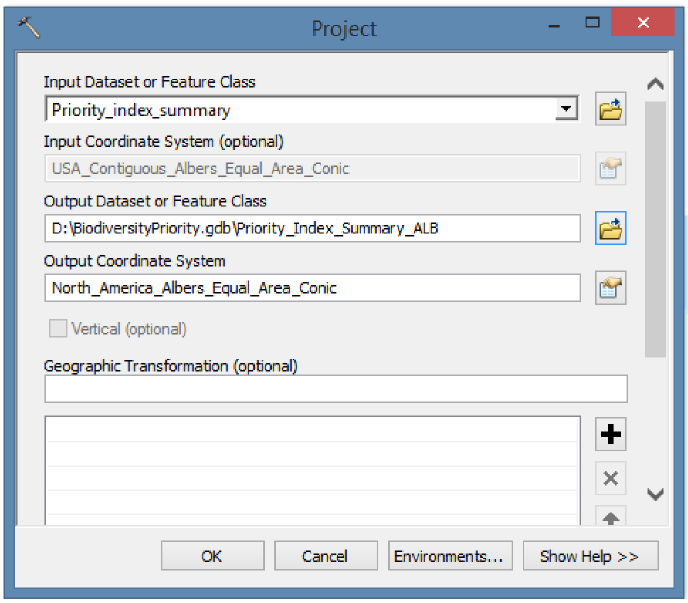
In our intact core habitat dataset these data are reflected in the fields **BiodiversityPriorityIndex** and **EndemicSpeciesMax** respectively.

We do not have permission to redistribute these files directly, but they are freely available on the BiodiversityMapping.org website and are not large. Click the 'Download GIS data' button and download the USA maps to download a map package.

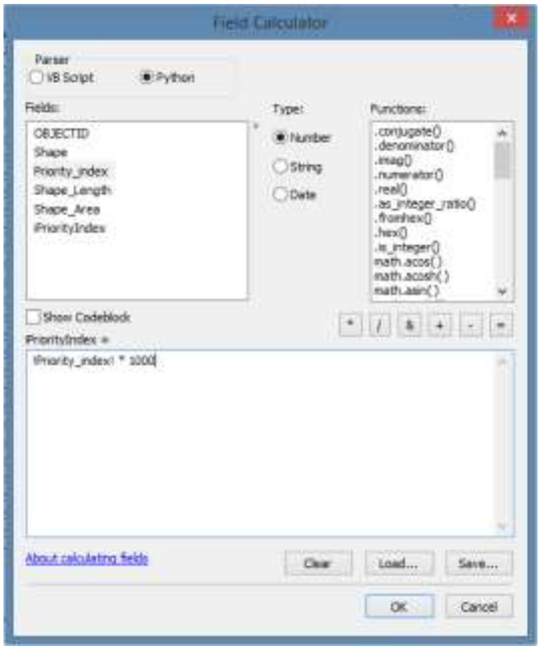
These data can be used directly for visual reference; however, some preprocessing is required to use the datasets in the model to create intact habitat cores. The workflow assumes intermediate knowledge of ArcGIS and is described below.

Priority Areas for USA Biodiversity Conservation Workflow

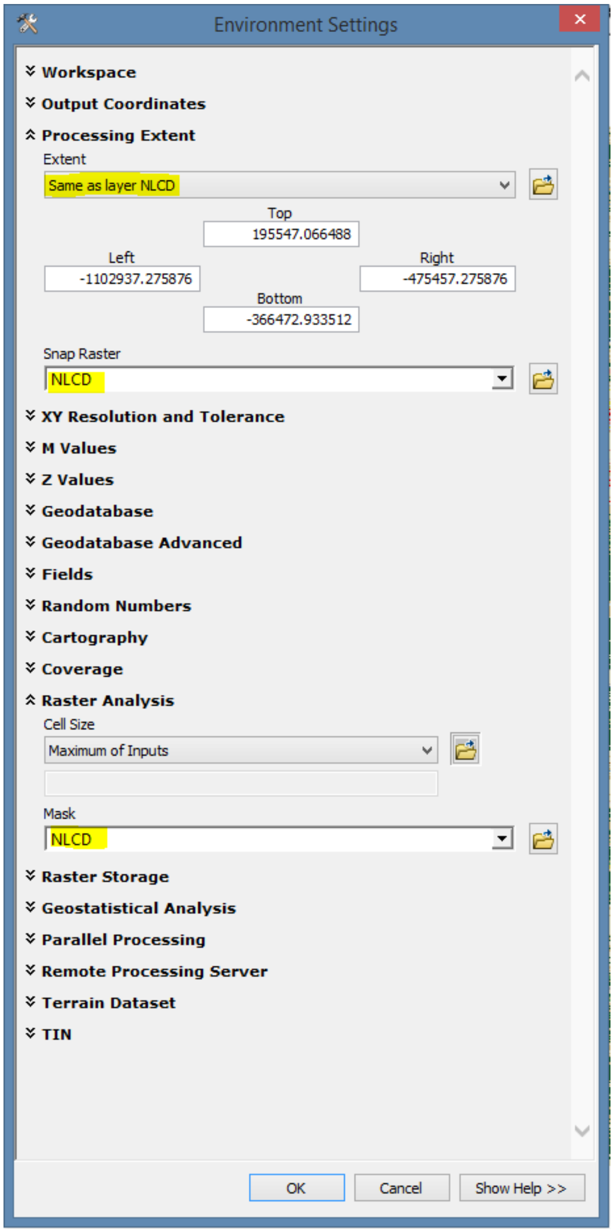
- Project to common spatial reference. The data are distributed as a polygon layer (website_layers_simple_names.gdb\Priority_index_summary) with a different spatial reference than the national intact cores project used. Project to:
    - North America Albers Equal Area Conic
        - WKID = 102008



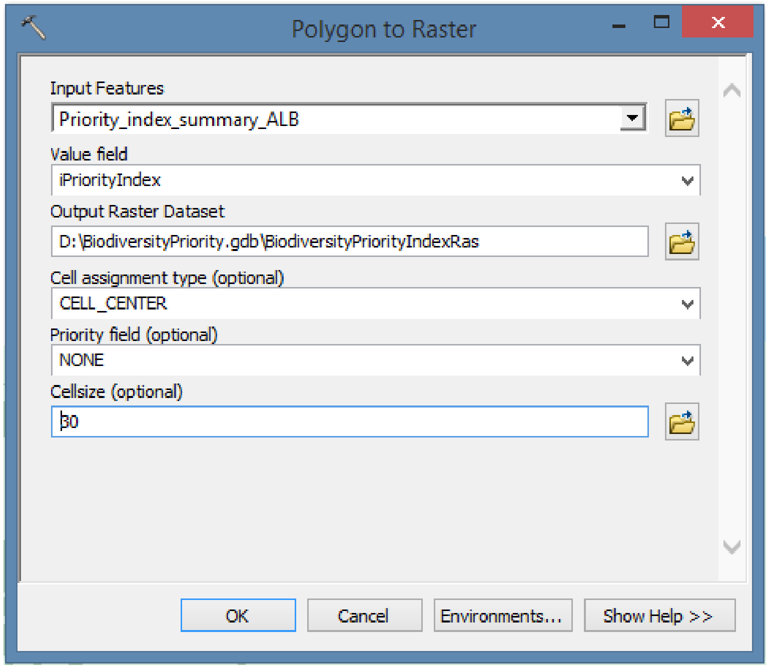
- Convert to a raster dataset and reclassify. The model calculates the intact core habitat biodiversity priority statistic field from a raster dataset.
    - The values in the dataset range from 0 to over 5000 with a mean value of about 2.4. Add an integer field and apply a multiplication factor the convert to a raster dataset.
    - AddField
        - Add a new field, iPriorityIndex of type **LONG** to the polygon attribute table
    - CalculateField
        - Calculate iPriorityIndex = PriorityIndex * 1000



- The model requires a raster data input. Convert to raster using the newly added field, iPriorityIndex, as the value item using the **Polygon To Raster** tool. Open the tool dialog and click the 'Environments' button. Set environment values: Extent, Snap Raster and Mask. Set these all to the NLCD layer included with the state data zip file or map package.

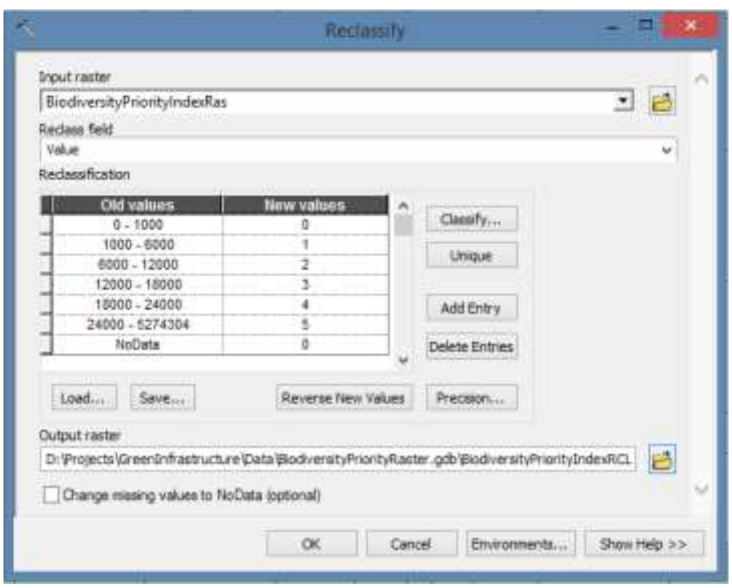


- Press OK to apply the environment settings.



- Press OK to create the new raster dataset.

- Reclassify the values in the raster dataset to range from 0 to 5 using the Spatial Analyst Reclassify tool:



The dialog above shows the reclassification that was used for the data across the nation. Note that the data ranges you observe in the “Old values” column may vary from these given the spatial extent of your data.

- Press OK to reclassify the data.

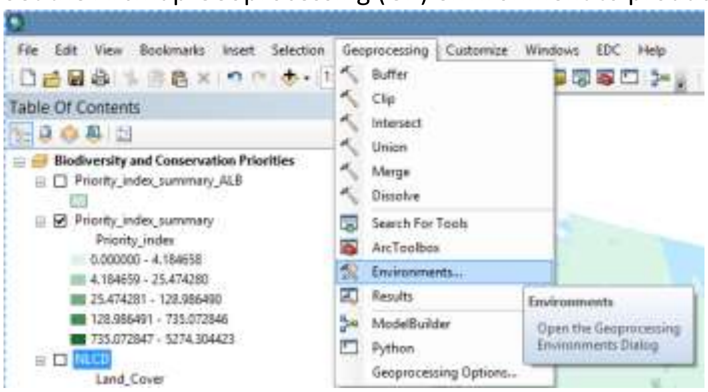
The output raster dataset is now ready for use in the GIC Core Tool -> Biodiversity Priority Index Raster input field.

Endemic Species Distribution Workflow

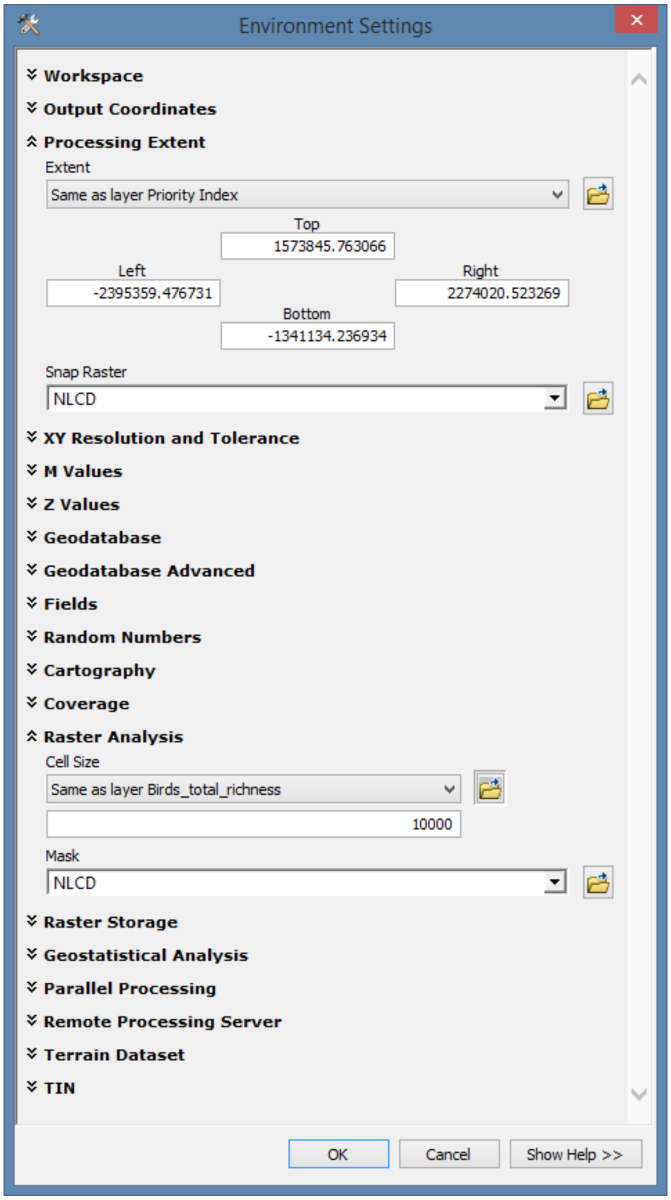
The endemic species distribution layer is the sum of six endemic species layers representing; mammals, birds, amphibians, reptiles, fish and trees.

The native format of these data was both raster and vector layers. To ensure that the file formats were consistent for all data, the two polygon layers were converted to match the 10km resolution of the other four raster datasets.

- Set the ArcMap Geoprocessing (GP) environment to produce consistent overlay results.

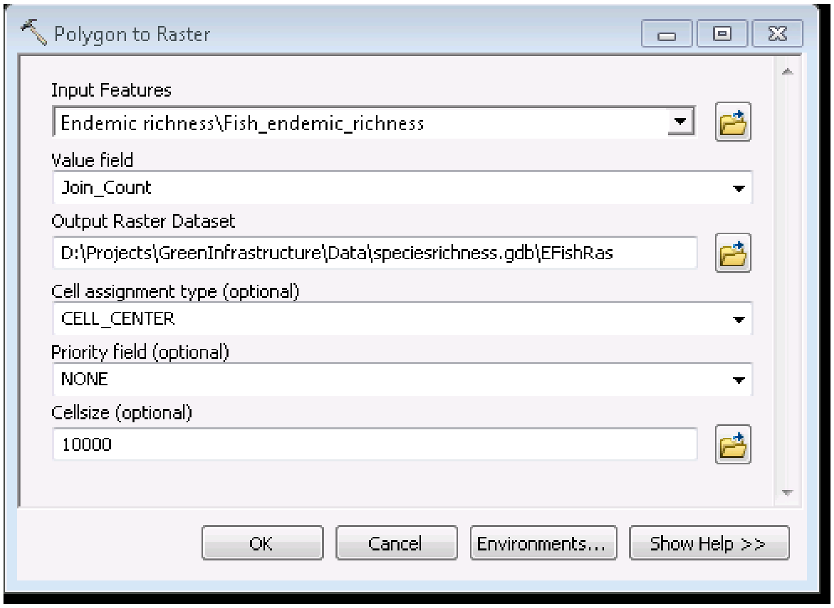


- Set SnapRaster and Cell Size to an existing endemic species raster dataset. Set the Extent and Mask values to the Area of Interest (AOI) Layer which is frequently the NLCD layer but could be your study area or county boundary.

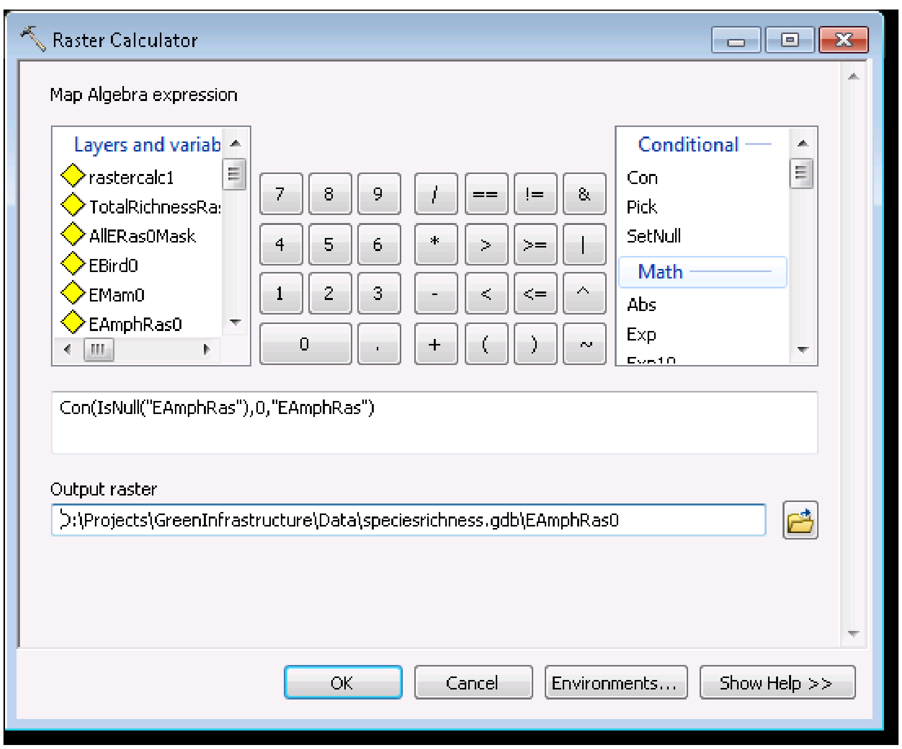


Having set the GP environment, we then converted the polygon layers (Amphibians_endemic_richness and Fish_endemic_richness) to raster datasets.

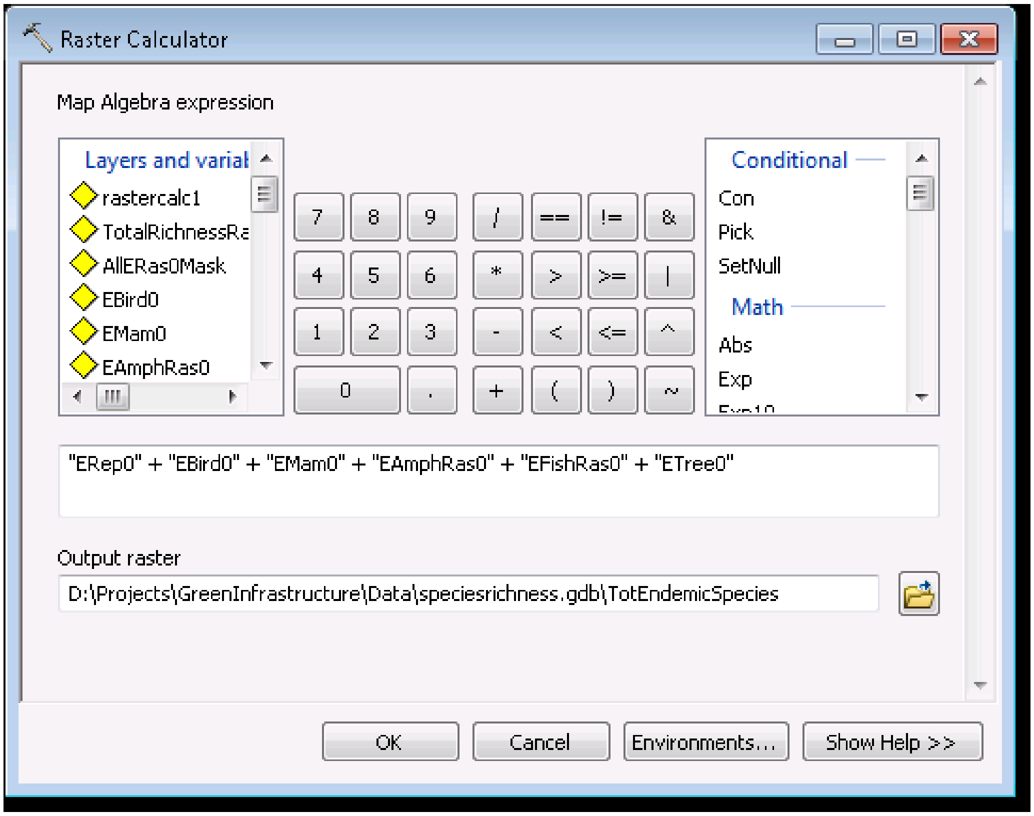
- Using the PolgonToRaster GP tool, convert fish and amphibian polygon layers to 10 km resolution raster datasets.



- This completes the normalization of all inputs to 10km raster datasets. Before summing these data, cells represented as NoData in each layer must be replaced with a value of "0" within the AOI. Failing to complete this step will result in an incomplete output. The Spatial Analyst Map Algebra dialog was used to replace NoData with "0" as follows:

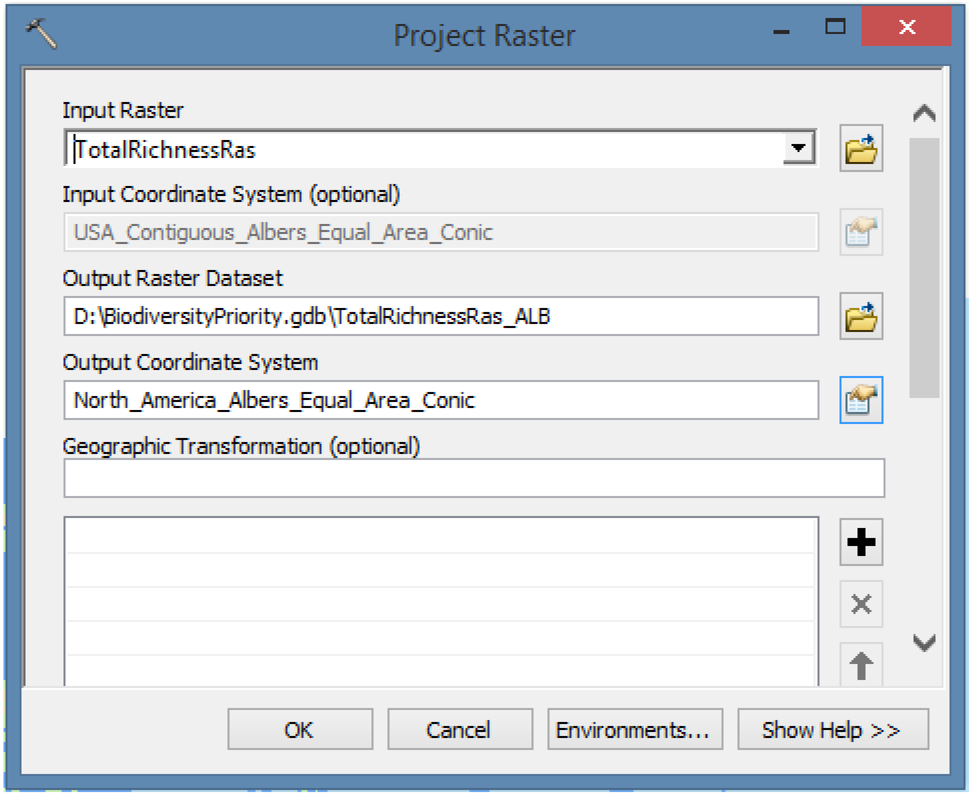


- Completing this process ensures that all layers are "padded" with "0" values to the edge of the AOI (having previously applied the GP mask and extent). Map Algebra was then utilized to sum the endemic species layers.



Using the expression above, the output raster cells reflect the sum of endemic species from the six input layers. Some areas, particularly near International borders, will have a value of zero by virtue of sensitivities of the original data and the analysis extent. This is due to the fact that as habitat ranges cross borders, they are no longer exclusive to the U.S., thus they are not included. Additionally, some ranges extend the full analysis extent including the border, resulting in the likelihood of NoData values for those species beyond the border

- Project to common spatial reference. The data are distributed with a different spatial reference than the national intact cores project used. Use the Project Raster tool:



The result, TotEndemicSpecies_ALB raster is now ready to be used in the GIC Core Tool-> Endemic Species Max Raster input field.