GIS Notes: GIS Methods for Spatial Statistical Modeling on Stream Networks

May 7, 2014

Software Requirements

1. ArcGIS v10.1 or 10.2
2. Advanced license and the Spatial Analyst Extension
3. STARS custom ArcGIS toolset version 2.0.0
4. Python version 2.7
5. PythonWin for version 2.7 (pywin32-218.win32-py2.7.exe)

Getting Started

1. Unzip the lab_example.zip file that we provided and save it in a new folder named C:\spmodels (C:\spmodels\datasets)

2. Create a C:\spmodels\work folder where you will save new datasets created in this lab. Then create a second directory within work and name it C:\spmodels\work\lsn.

3. Check for the C:\temp folder. Some of the STARS tools write temporary files to this folder and will not work without it. Open Windows Explorer and create a C:\temp folder if you do not already have one.

4. Add ArcGIS Toolboxes

    1. Open ArcGIS.

    2. Click on the ArcToolbox button, , located at the top of the screen.

    3. In ArcToolbox, right click on the words 'ArcToolbox' and select Add Toolbox from the dropdown menu. Navigate to the STARS toolbox and click Open.

    4. Right click on ArcToolbox, scroll down to Save Settings and click on To Default.

    5. Change Environment Settings. Go to Menu, click on Geoprocessing, and select Environments. Expand M Values and set Output has M Values to Disabled. Repeat these steps for Z Values and click OK.

    6. Overwrite outputs by default. In the menu, select Geoprocessing> Geoprocessing Options and check the box next to Overwrite the outputs of geoprocessing operations.

Example Dataset

The data has been pre-processed so that it all has the same projection. In addition, all of the rasters have a 25m spatial resolution and the raster cells line up perfectly. The datasets directory contains all of the data used in this example and includes:

- sites.shp: Observed site locations and attribute data
- preds.shp: Prediction sites locations and attribute data
- streams.shp: Stream network and attribute data
- demfill: Filled DEM
- edgegrid: Raster of streams derived from edges. Cell values representing edges contain the reachid value and all other cells contain NoData values
- wbgrid: Waterbodies converted to raster format.
- grazing: Raster of grazing land use: 1 = grazed land and NoData = everything else
- rca.shp: RCAs in vector format. The GRIDCODE can be used to link RCAs to edges. GRIDCODE = reachid.
- rcagrid: RCAs in raster format. The Value attribute can be used to link the RCAs to the edges. Value = reachid.

Adding Data to the View

1. Load the streams, sites, and prediction sites into ArcGIS

   Click on the Add Data tool  and navigate to the C:\spmodels\datasets folder. Select the streams.shp, sites.shp, and preds.shp files. Click Add. Note that, you may need to open ArcCatalog and Connect to Folder C:\spmodels before you can add the data.

2. Switch from the ArcToolbox tab to the Table of Contents (TOC) tab.

3. Turn layers on and off in the map.

   In the TOC, check and uncheck the box next to a layer name to make it visible or invisible in the view. The TOC can be viewed in multiple ways depending on which option is selected at the top. One useful option is the List By Drawing Order, , which lists Layers in the order in which they are displayed in the data view. Thus, the layers at the top of the TOC are displayed ‘on top of’ layers farther down in the list. Drag and drop layers to change the order in which they are drawn. Another useful option is List by Source, , which groups Layers depending on the source directory they reside in.

4. Examine the sites attribute data.

   Right click on sites in the TOC, scroll down, and select Open Attribute Table. You can scroll up and down to view the records and left and right to see the fields. Close the sites.shp attribute table.

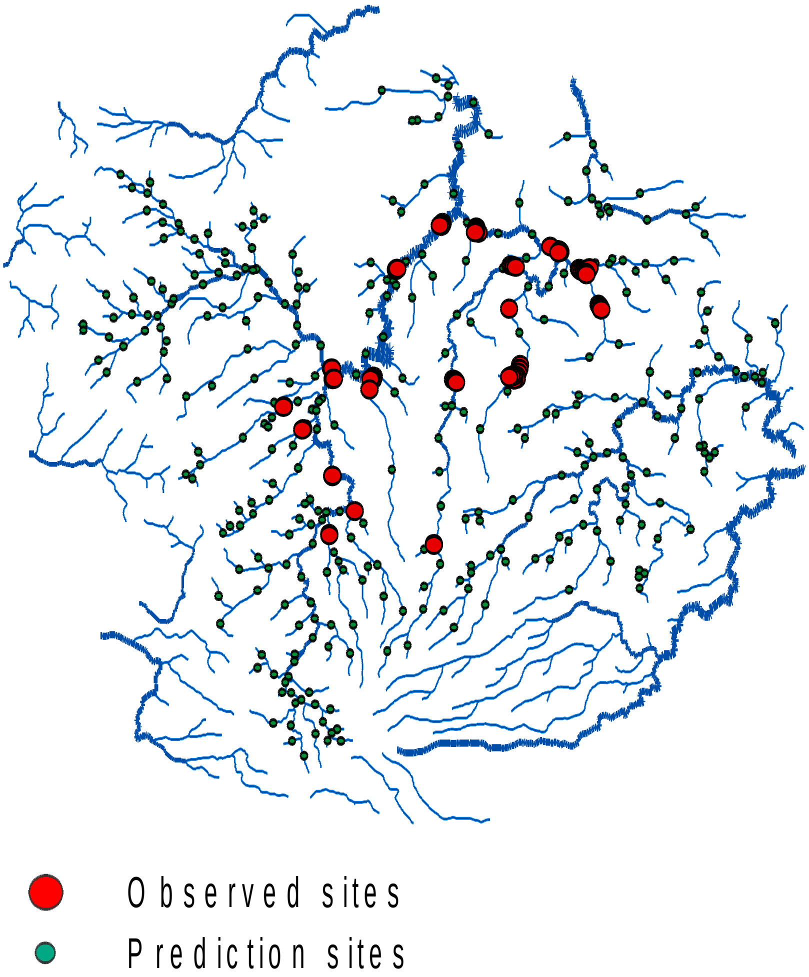


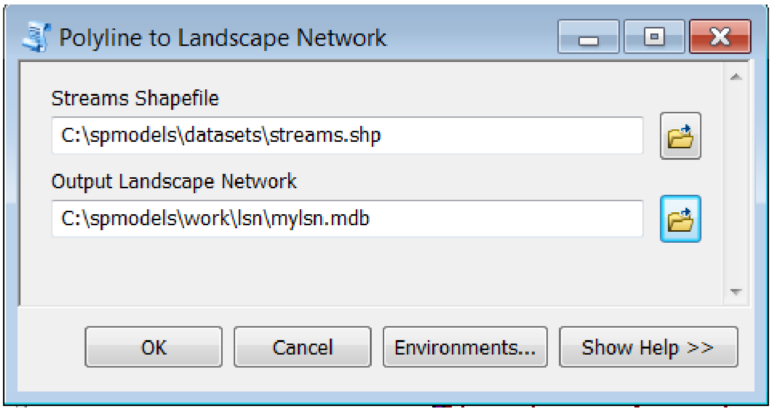
Figure 1. The observed sites are locations where data were collected, while prediction sites represent unobserved locations where predictions will be made.

Build a Landscape Network (LSN)

1. Click on the ArcToolbox tab to switch from the TOC to ArcToolbox.

2. Go to ArcToolbox>STARS>Pre-processing and double click on Polyline to Landscape Network to open the Polyline to Landscape Network window.

3. Set the parameters and click OK. Be sure to include the file extension .mdb in the Output Landscape Network.



If everything works correctly, the message FINISHED Polyline to Landscape Network Script will be shown in green. The tool produces a personal geodatabase with five components

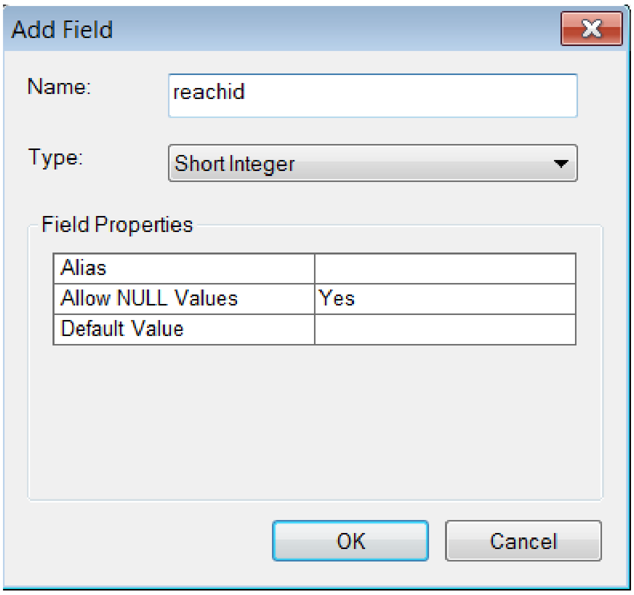
- nodes (point feature class)
- edges (polyline feature class)
- relationships (table)
- noderelationships (table)
- nodexy (table)

The program *failed to produce a LSN* if any of these components is missing.

4. Add the new LSN to the view
   1. Click the Add Data tool, and navigate to the C:\spmodels\work\lsn folder.
   2. Double click on mylsn.mdb.
   3. Select all five components and click Add.

5. Add a Short Integer field named reachid to the edges attribute table. This attribute will be used to relate the RCAs to the edges later in the lab.
   1. Open the edges attribute table and click on the Table Options button, , found at the top left corner of the Attribute Table. Select Add Field to open the Add Field window.

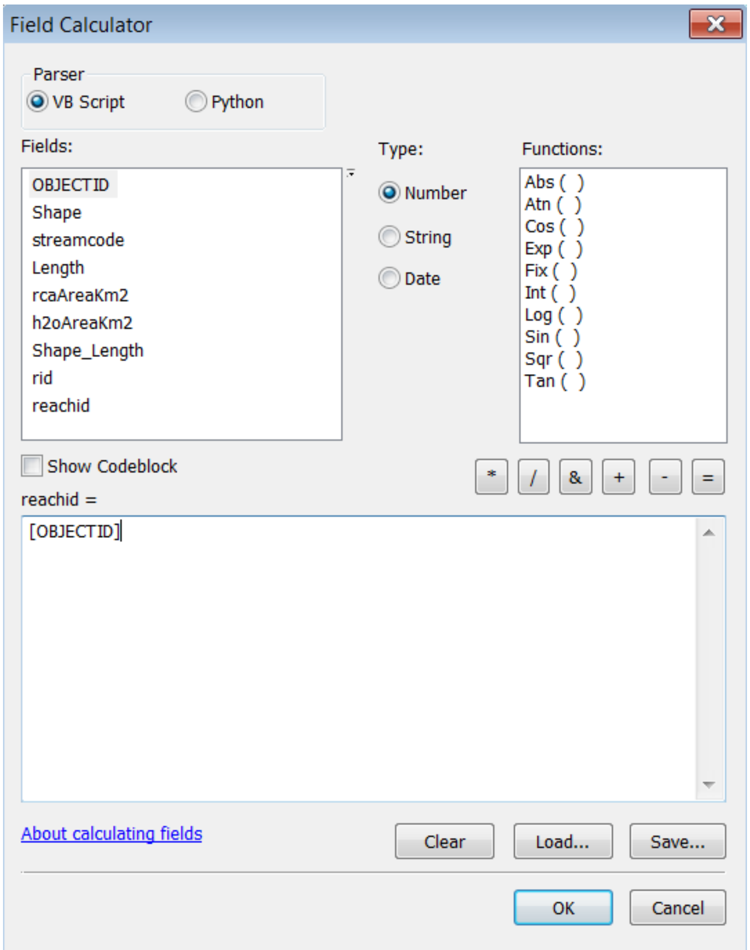
6. Set the parameter values and click OK.



7. Calculate the reachid field
    1. Scroll to the right until you see the reachid field. It will be filled with NULL values.
    2. Right click on the reachid field name and select Field Calculator to open the Field Calculator window.

8. Set the reachid = [OBJECTID] and click OK. Close the edges attribute table.

    Hint: You can build expressions by clicking on field names, functions, or buttons in the Field Calculator. This reduces the chance of errors in the expression.

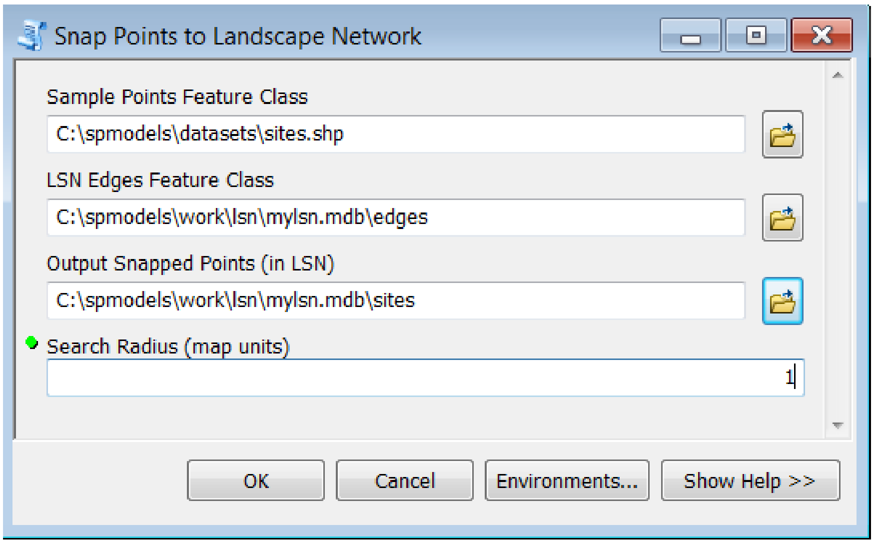


9. Remove streams.shp from the view. In the TOC, right click on streams and select Remove.

Incorporating Sites into the LSN

1. Go to ArcToolbox>STARS>Pre-processing and double click on Snap Points to Landscape Network to open the Snap Points to Landscape Network window.

    1. Set the parameter values. In this example, set the Search Radius = 1 because the sites.shp have already been manually edited and snapped to the appropriate reach. Click OK. The tool may take a few minutes to run.



A new sites feature class will be written to the LSN geodatabase (mylsn.mdb) if the survey locations are successfully incorporated. The sites feature class attribute table will contain some new fields, two of which are the ratio and rid fields. The rid field indicates which stream reach the site has been snapped to and the ratio field contains the percent distance along the stream reach. Together, these values are used to identify a site's location within the LSN. This is extremely useful because it enables hydrologic distances, spatial weights, and watershed attributes to be calculated for the sites.

If the tool finishes successfully, you should see a green message saying Finished Snap Points to Landscape Network.

2. Repeat steps 1 and 2 using preds.shp as the Sample Points Feature Class.

3. Remove sites.shp and preds.shp from the view.

4. Add the sites and preds feature classes to the view. They are located in the LSN geodatabase (C:\spmodels\work\lsn\mylsn.mdb).

Note: You will receive a warning saying that the sites and preds feature classes are missing a projection. Click OK.

Creating Reach Catchment Areas (RCAs)

1. Add input data into ArcGIS: demfill, edgegrid, and wbgrid datasets from the C:\spmodels\datasets folder

2. Check to see whether the Spatial Analyst toolbar is activated in ArcToolbox. If it is grayed out, you will need to enable it. On the main menu, go to Customize>Extensions. This will open the Extensions window. Make sure that the box next to the Spatial Analyst is checked.

3. Set the Geoprocessing environment

    1. Click on the Geoprocessing menu, scroll down and select Environments.

    2. Under Workspace, set the Current Workspace and Scratch Workspace to C:\spmodels\work.

    3. Under Processing Extent, set Extent to Same as layer demfill.

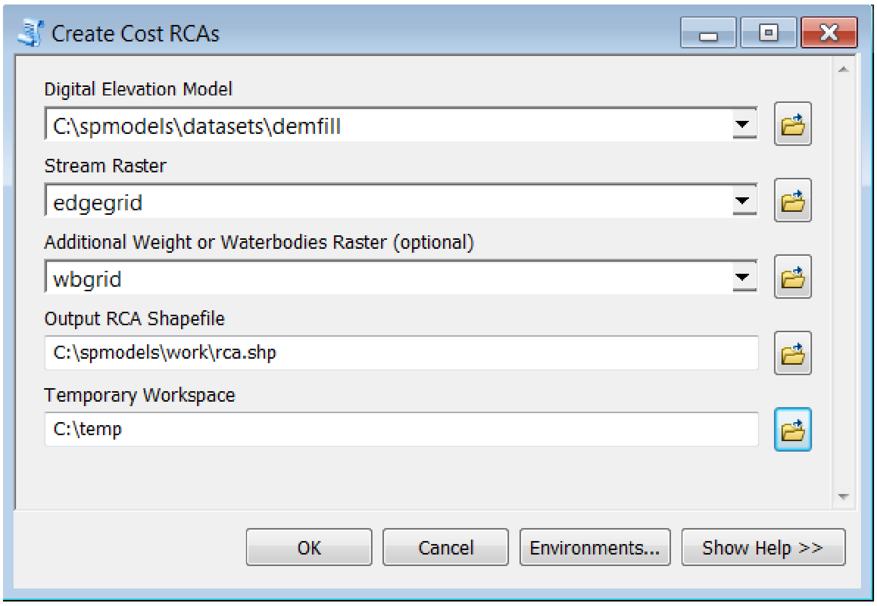
    4. Under Raster Analysis, set the cell size to Same as layer demfill. This will automatically set the cell size. No Mask is used.

    5. Click OK.

4. Create Cost RCAs

    1. Go to ArcToolbox>STARS>Pre-processing and double click on Create RCAs. This will open the Create Cost RCAs window.

    2. Set the parameters and click OK. The .shp extension in the Output RCA Shapefile name will automatically be included.



The program may take a few minutes to run.

The Create Cost RCAs tool will create a shapefile of RCAs. The GRIDCODE attribute in rca.shp is equal to the reachid value found in the edges attribute table. This attribute enables each RCA to be directly linked to a single edge.

5. If it wasn’t automatically added, add the rca.shp to ArcGIS and examine the tessellation of RCAs.

It is not uncommon to find that a relatively small number of stream reaches have not been assigned an RCA. This may occur if the length of the stream reach is short in relation to the spatial resolution of the DEM. For example, if a stream reach is 10 meters in length and the DEM has a 25 meter spatial resolution. If this happens, the RCA is essentially too small to delineate. When a group of stream reaches is part of a waterbody, such as a lake or reservoir, some reaches may not be assigned RCAs. Instead, there may only be two or three RCAs for an entire waterbody. Errors, such as multi-part features, can also occur when a large waterbody covers multiple reaches.

6. Remove rca.shp and the demfill, edgegrid, and wbgrid rasters from the view.

Calculating RCA Attributes

The RCAs must be converted to raster format before the RCA attributes are calculated. This raster, rcagrid, was provided in the example dataset.

1. Add the rcagrid and grazing rasters to the view. They can be found in the C:\spmodels\datasets folder.

   The rcagrid contains a Value attribute that is equivalent to the reachid found in the edges feature class. Each cell contains the reachid that the individual cell is associated with.

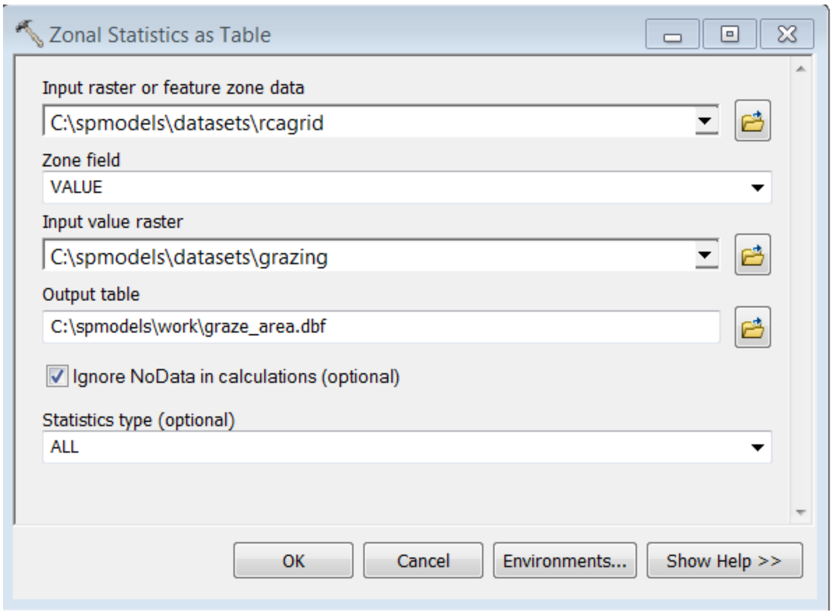
   Cells in the grazing grid contain either a 1 (grazed land) or NoData (other) value.

2. Calculate the RCA attribute for grazed area

   1. Open the Spatial Analyst Tools, scroll down to Zonal, and open the Zonal Statistics as Table tool.

   2. Set the argument values. In this example, the landscape value raster, grazing, represents grazed areas.

   3. Click OK.



This tool produces a dbf table with grazed land statistics, such as mean, count, or area for each RCA. The Value attribute in the table is equivalent to the reachid, which enables the RCA attributes to be associated with a single reach.

***Note that**, numerous users have reported issues with the Zonal Statistics As Table tool in ArcGIS 10.1, but not version 10.2. Please see the STARS tutorial page 29 for potential solutions.

3. Add the new RCA attribute to the edges attribute table.

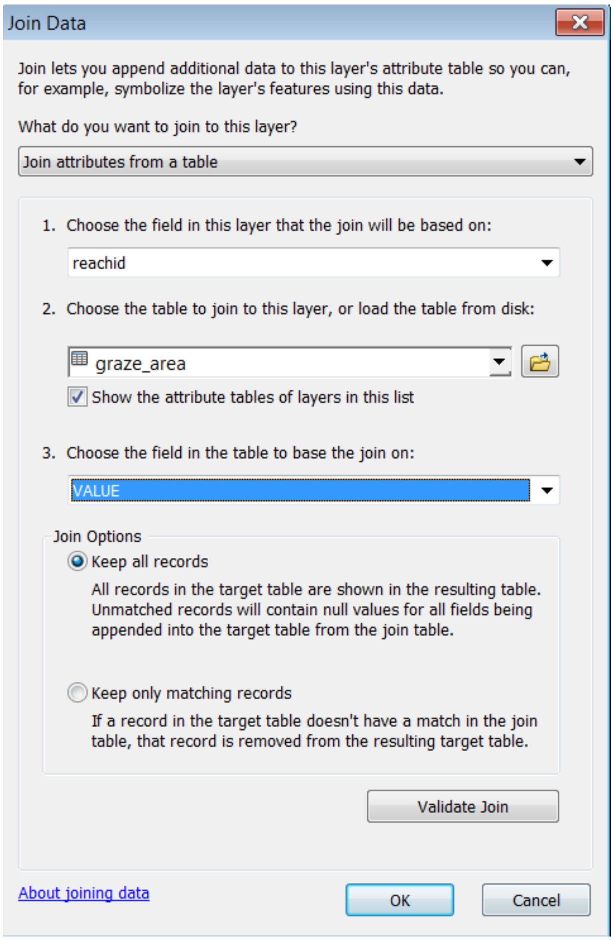
    1. Add a field of type DOUBLE to the edges attribute table that represents the grazed land use in the RCA

        Name = rcaGrazeKm2
        Type = Double

    2. Join the dbf table to the edges attribute table

        1. In the Table of Contents, right click on edges, scroll down, and select Joins and Relates>Join… This will open the Join Data window.

        2. Set the arguments (below) and click OK.



If a window opens asking whether you would like to index the field, click Yes.

3. Open the edges attribute table and scroll to the right until you see the empty RCA landscape value attribute. In this example, it will be rcaGrazeKm2.

4. Calculate the field. In this example, the area is also being converted from square meters to square kilometers.

   edges.rcaGrazeKm2 = [graze_area.Area] * 0.000001

   When the edges attribute table and the zonal statistics table are joined, RCAs that do not contain grazed areas will have NULL statistics values. As a result, an error box may appear when the field is calculated that says "The calculated value is invalid for the row with OBJECTID = …" This is normal – click OK.

   *Note: In this example, we are interested in the area of grazed land in each RCA. In other cases, we might be more interested in the mean or the count statistic.

5. Remove the Join: In the TOC, right click on edges, scroll down, select Joins and Relates>Remove Join(s)> Remove All Joins

6. Set NULL rcaGrazeKm2 values = 0
    1. In the edges attribute table, open the Select by Attributes window.

    2. Select NULL values

        [rcaGrazeKm2]is null

    3. Use the Field Calculator to set the rcaGrazeKm2 field = 0 for the selected records. Note, only the selected records will be calculated. Clear the selection and close the edges attribute table.

    4. Remove the rcagrid and grazing rasters, as well as the graze_area.dbf from the TOC.

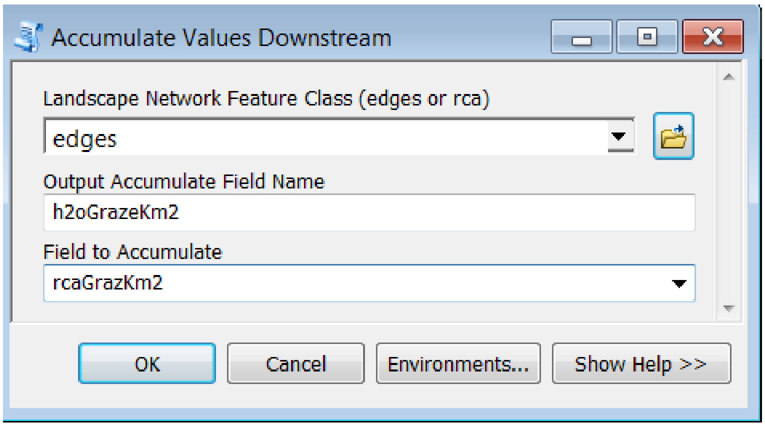
    It is necessary to set the NULL values = 0 for areas and counts. However, this step should be skipped if means are being calculated for continuous values, such as elevation, because zero may be a meaningful value. Please see the STARS tutorial page 30 for additional details.

Accumulating Watershed Attributes

In this example, the rcaGrazeKm2 attribute will be accumulated to calculate the total grazed area in the watershed.

1. Go to STARS>Calculate and double click on Accumulate Values Downstream to open the Accumulate Values Downstream window.

2. Set the parameter values and click OK.

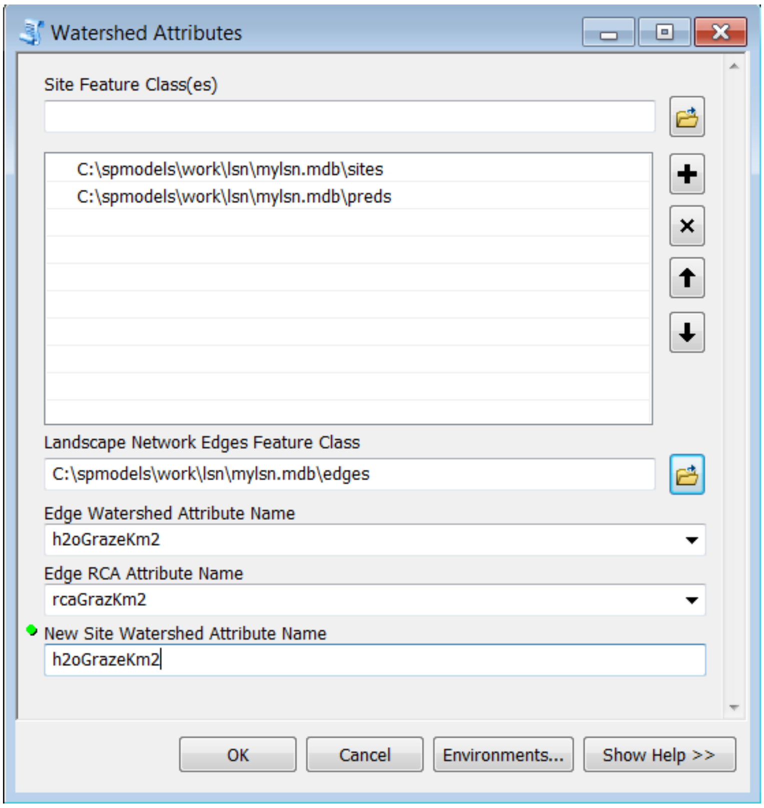


If the tool runs without errors, a green message will appear: Finished Accumulate Values Downstream Script. It adds a new field to the edges attribute table that represents the watershed attribute for the downstream node of each edge. The values shown in the new watershed attribute area field will be in square projection units. In this case, they are in square kilometers because the projection is Albers and we converted square meters to square kilometers.

Watershed Attributes

1. STARS>Calculate and double click on Watershed Attributes to open the Watershed Attributes window.

2. Set the parameter values and click OK.
    1. Watershed attributes may be calculated for multiple point feature classes simultaneously (i.e. observed and prediction sites).
    2. The Edge Watershed Attribute Name must be an accumulated field.
    3. The Edge RCA Attribute Name should be the RCA field that was accumulated to produce the Edge Watershed Attribute Name field. For example, rcaGrazeKm2 was accumulated to produce h2oGrazeKm2.

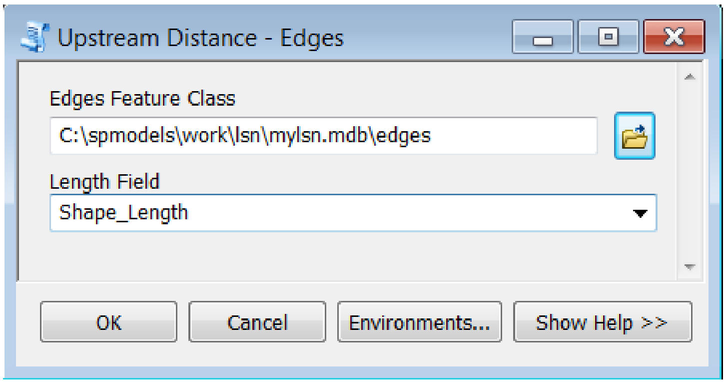


If the program finishes without errors, a green message will appear: Program finished successfully. A new field will be added to the sites attribute table that contains the watershed attribute for each site.

Calculate Upstream Distance – Edges & Sites

1. Double click on STARS>Calculate>Upstream Distance – Edges to open the Upstream Distance – Edges window.

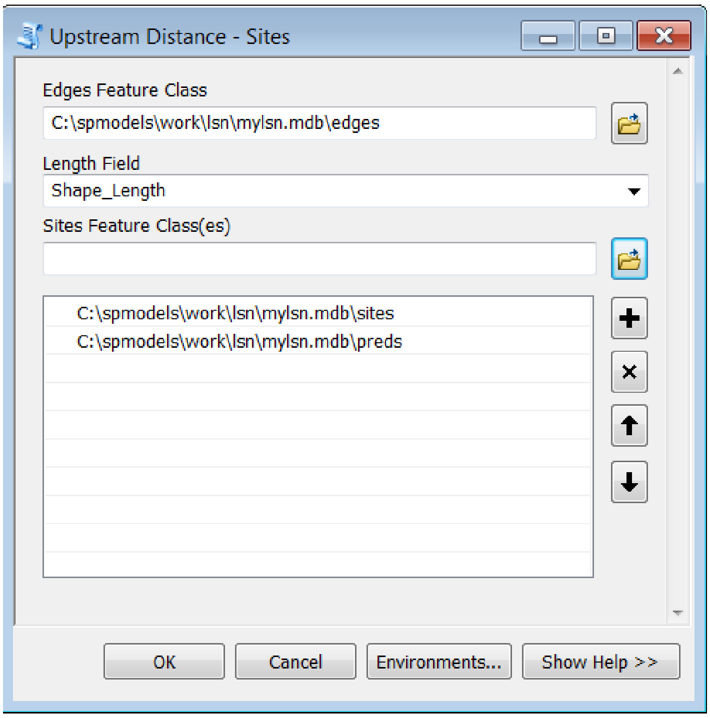
2. Set the parameters



You should see a green message Program finished successfully. In addition, a new field should appear in the edges attribute table named upDist.

3. Double click on STARS>Calculate>Upstream Distance – Sites to open the Upstream Distance – Sites window.

4. Set the parameters and click OK.



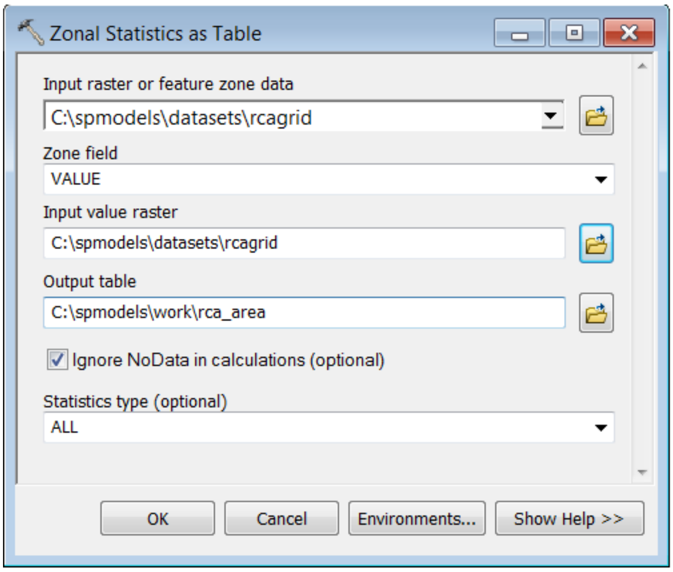
If no errors occurred, a green message, Program finished successfully, will appear.

***Note**: The upDist attribute must be present in the edges attribute table before the Upstream Distance – Sites tool can run. This tool adds an upDist attribute to all of the sites feature classes that have been selected.

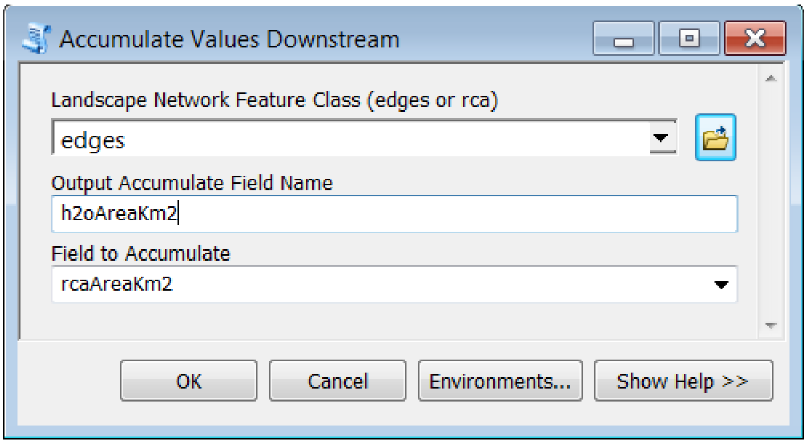
Calculate Segment PI

In this example, we will use watershed area as the basis for our spatial-weighting scheme. However, other attributes could be used as long as they are pre-processed using the Segment PI, and two Additive Function Values tools.

If you are using the NHDPlus or another attributed streams dataset, watershed area for the downstream node of each line segment may already be included as an attribute. If not, the instructions given in the Calculating RCA Attributes and Calculating Watershed Attributes sections can also be used to derive watershed area. For example, the Zonal Statistics as Table tool can be used to calculate the area of each RCA:



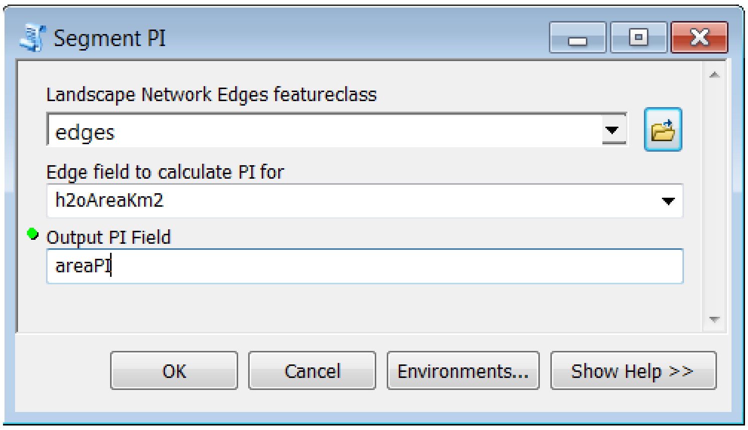
The rcaAreaKm2 is calculated in the same way that the rcaGrazeKm2 was calculated in the Calculating RCA Attributes section. The new rcaAreaKm2 attribute is then accumulated downstream to produce the watershed area for the downstream node of each line segment (h2oAreaKm2):



Once the h2oAreaKm2 has been calculated, it can be used to calculate the Segment Proportional Influence (PI).

1. Double click on the STARS>Calculate>Segment PI

2. Set the parameters and click OK.



A green message will appear, Program finished successfully, if it runs without errors. This tool adds a new field to the edges attribute table that contains the segment PI values. In this example, it is named areaPI.

The segment PI values are proportions and so values should range between 0 and 1. If any values are greater than 1, then the segment PI values are invalid and should be recalculated.

3. Check the areaPI field to ensure that values range between 0 and 1.

    1. Open the edges attribute table, right click on the areaPI field, scroll down and select statistics. This will open the Statistics of edges window, which contains summary statistics for the areaPI field.

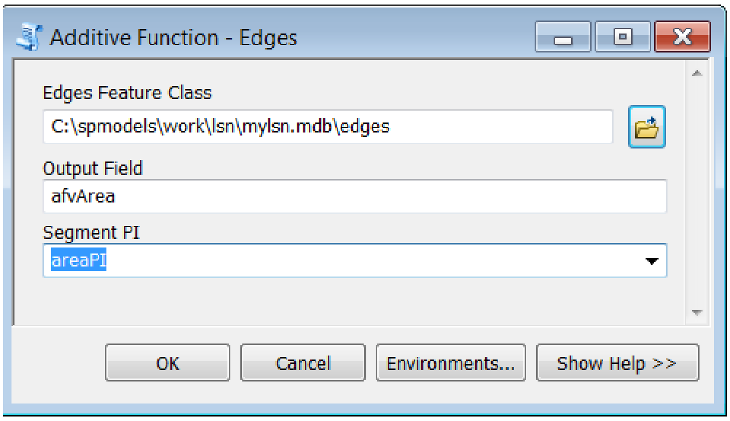
    2. Examine the minimum and maximum values to ensure that they range between 0 and

Calculate Additive Function – Edges & Sites

1. Calculate Additive Function – Edges

    1. Go to STARS>Calculate>Additive Function – Edges to open the Calculate Additive Function - Edges window.

    2. Set the parameters and click OK.

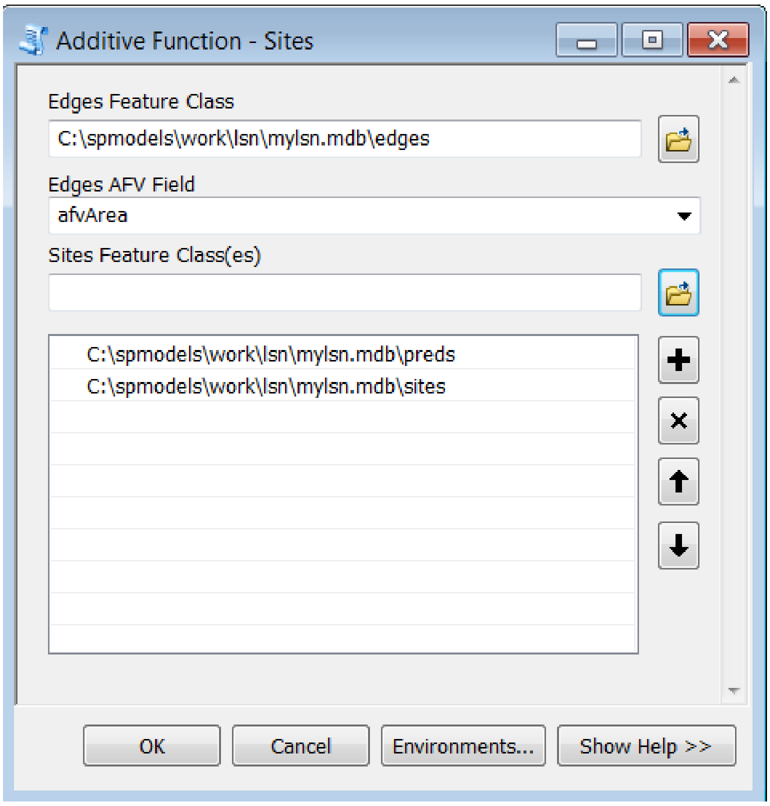


If the script runs without errors a green message will appear: Finished Additive Function - Edges.

2. Calculate Additive Function – Sites

    1. Go to STARS>Calculate>Additive Function – Sites to open the Calculate Additive Function - Sites window.

    2. Set the arguments and click OK.

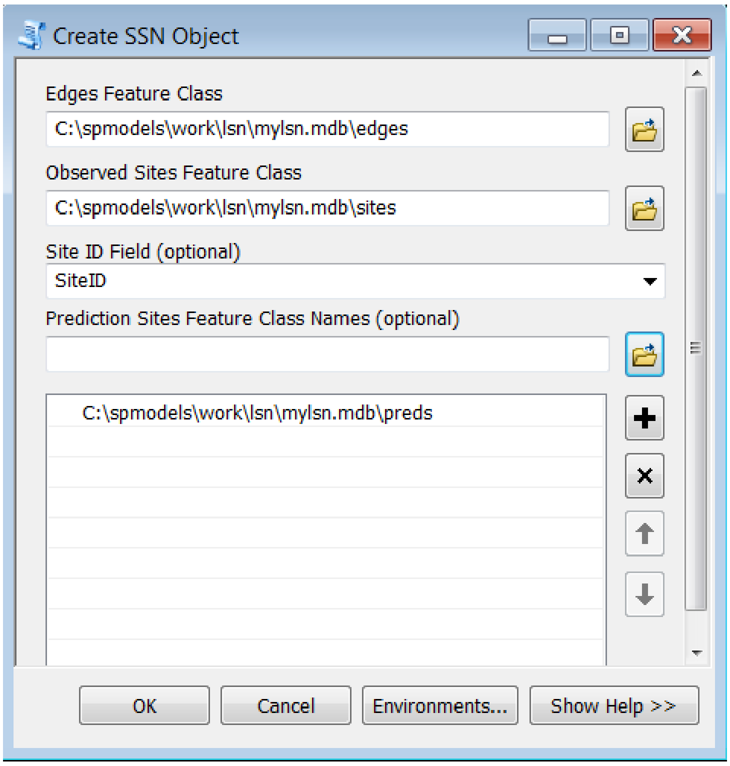


When the tool finishes running, a green message should appear: Finished Additive Function Script. The Additive Function tools create new fields in both the sites and edges attribute tables representing the additive function values (AFV). The AFV is a product of proportions (segment PI values) and so the AFV should always range between 0 and 1. Check to ensure that this is the case.

Create SSN Object

1. Double click on STARS>Export>Create SSN Object to open the Create SSN Object window.

2. Set the parameters and click OK.



The Create SSN Object tool may take a while to run depending on the number of edges and sites in the LSN. When the tool has finished successfully, a green message, Successfully Finished Create SSN Object Script, will appear. In addition, a .ssn object will have been created in the same directory as the LSN used to create it.

Examine the .ssn directory

1. Navigate to C:/spmodels/work/lsn

   This directory should contain the mylsn.mdb and a new directory called mylsn.ssn. This is the .ssn directory that will be imported into R and used to create a SpatialStreamNetwork object.

2. Open the mylsn.ssn directory. It should contain:
   3 shapefiles: edges, sites, and preds
   16 .dat files: netID1.dat to netID16.dat

Checking the .ssn object

Numerous attributes are calculated for the edges, sites, and prediction sites before it can be exported as a .ssn object. If a mistake happens during the pre-processing steps, it can often go unnoticed until after the .ssn object has been imported into R and an error occurs in one of the modeling or prediction functions. At this point, it is often difficult to diagnose the cause of the error. We have created a document with instructions on **Checking the .ssn Object** to help users ensure that their object is formatted properly. The document has been provided with the short course materials (Check_ssn.pdf) or alternatively, can be downloaded from the SSN & STARS website:
http://www.fs.fed.us/rm/boise/AWAE/projects/SSN_STARS/software_data.html .

If the .ssn object does not meet all of the criteria in the document, an error occurred. Please see the Troubleshooting section of the tutorial for more information.