Calculating Strahler Stream Order from NHD Data

Barry J Kronenfeld, Eastern Illinois University

1. Introduction

Arther Strahler developed a conceptual ordering (ranking) of streams in the 1960s that is useful for measuring and visualizing branching networks such as stream systems, and is still discussed in many textbooks today. Strahler's system can theoretically be applied to a typical, topologically consistent branching network assuming that flow is in one direction and there are no loops. Practical application of the stream ordering system, however, is more complicated. Hydrology tools in ArcMap allow for calculation of Strahler stream order from raster data but not vector data. Further, the National Hydrology Dataset (NHD), which provides high quality stream data for the United States, includes polygonal features (e.g. lakes) and network loops that present both practical and conceptual obstacles to determining stream order.

This exercise works through the process of converting NHD stream data into simple hierarchical trees without loops, and calculating Strahler stream order and related measures on the resulting data. The conversion process involves some degree of generalization, as does the Strahler stream ordering systems itself. The advantage is that the resulting model can be used to characterize flow qualities of individual streams as well as the overall form and efficiency of the stream network.

2. Obtain Tools

This exercise uses a set of Python scripts written by the author for a graduate-level GIS modeling class at Eastern Illinois University. A toolbox containing these scriptss can be downloaded at http://ux1.eiu.edu/~bjkronenfeld/projects/coding.html. Download and unzip the zipped folder located in the section "GIS Modeling Scripts" to your working folder. The exercise also utilizes many ArcMap tools. Tools will be referenced by toolbox/toolset name; the scripts contained in the downloaded zip folder are accessible through ArcCatalog in a toolbox called *GIS_Modeling*.

3. Obtain Data

Pre-staged data for NHD subregions can be obtained from http://nhd.usgs.gov/data.html (go to *pre-staged subregions*) or directly from ftp://nhdftp.usgs.gov/DataSets/Staged/SubRegions/FileGDB/MediumResolution/. These data correspond to 4-digit hydrological regions. NHD data at other levels can be obtained from the National Map.

Examples in this walk-through will use data for subregion 0512 in eastern Illinois, central Indiana and a small part of western Ohio. I recommend you try following along using a different subregion – pick one at random, or you can view NHD 4-digit subregion outlines on the National Map.

4. Prepare Map

- Open a new map document in ArcMap and save to a suitable working folder.
- Download the NHD dataset for your region and unzip (extract) it to your working folder.
- Add the *Hydro Net* geometric network layer from the geodatabase into your map.
- Add other data as appropriate for geographic context. Elevation data in particular will be useful to confirm flow direction.
- Create a new geodatabase called *working* to store your working data and results.

5. Remove Loops from NHD Flowlines

NHD data is built as a geometric network, and can be used as input for tools in the *Geometric Network* toolset in the *Data Management* toolbox. One such tool, *Trace Geometric Network*, can be used to identify loops (circuits) in the network.

- Note: NHD datasets contains flow direction data for each network edge. This generally means that there are no *flow loops* – that is, if you follow the direction of flow, you will never visit the same location twice. However, the network contains many instances of flow diverging and then reconverging, resulting in *diagram loops* – connected sets of edges that close to form a polygon. These loops must be eliminated and replaced by acyclic linear features (i.e. centerlines) to determine Strahler stream orders.

- In some cases, there is more than one way to replace a loop with lines, which may affect Strahler stream order assignment. The strategy we will take is to:
  1. Identify large loops and resolve them manually. It is best to use elevation data to visually resolve ambiguities.
  2. For the more numerous small loops, which often represent lakes and other areal water features, use the *medial axis* of the polygon created by each loop to represent the likely flow path of the original stream network.

5.1. Identify Loops

- Loops can be detected by the *Trace Geometric Network* tool is located in *Data Management Tools* >> *Geometric Network*. Run this tool and specify:
  - Input Geometric Network: the *HYDRO_NET* geometric network in the NHD geodatabase
  - Output Group Layer Name: Network_Loops
  - Flags: HYDRO_NET_Junctions
  - Trace Task Type: FIND_LOOPS
- This tool may take a while to run - it took 28 minutes on my admittedly antiquated 2009 Compaq Presario CQ56 (with an AMD V140 2.3 GHz processor and 4GB RAM).
- When the tool is finished, a new group layer is added to the map. The NHDFlowline dataset in this layer contains all flow lines, but with loops (loops) selected. **Do not clear/alter this selection yet…**

5.2. Create Polygons representing Loops

To determine medial axis centerlines it is necessary to create polygons for each loop. With the loops still selected from step 4.1, run the *Feature to Polygon* tool (*Data Management* >> *Features*) on the NHDFlowline data created by the *Trace Geometric Network* tool. Save your results as a temporary feature class and then run the *Dissolve* tool to dissolve adjacent polygons.

- To dissolve adjacent polygons, uncheck the box for *Create multipart features* in the *Dissolve* tool.
- Name your output layer "*Loop_Polygons*".
- Delete the temporary feature class.

You can now clear the selection on the NHDFlowline data layer.

5.3. Handle Large Loops

If there are any large loops, it is best to break them up manually to ensure that the acyclic (loop-free) network is a good representation of the actual flow. Otherwise, the automated procedure outlined below will create an acyclic network, but it may not be representative of the true network. This is especially true when loops span a divide – that is, when they contain streams that flow in opposite directions outward from the loop (an example can be found in the very first pre-staged NHD subregion for northern Maine). To handle these cases:

- Using manual selection, separate the loop polygons you will edit manually and polygons you will handle using automated tools. Place the loops for manual edting in a separate data layer.
- Locate and display high quality elevation data or maps to determine the actual flow directions, then manually edit the Flowline data to break up the large loop(s) that you designated above for manual editing.

5.4. Create Feature Class containing Flow Lines Exterior to Loops

- This step and further steps below assume that the *Loop_Polygons* feature class only contains loops that have not been handled through manual editing.
- Use the *Select by Location* function to select NHDFlowline features that are within the *Loop_Polygons* feature class. Verify that segments both within the polygons and on the polygon boundaries are selected (but segments outside the polygon boundaries are *not* selected).

- Invert the selection to get flow lines exterior to loop polygons (open the attribute table and click the *Switch Selection* button).
- Export the selection to your working geodatabase as a new feature class called "*Exterior_Lines*".

At this point, you can remove all NHD data from your map except for the *Loop_Polygons* and *Exterior_Lines* feature classes you created.

5.5. Generate approximate medial axes in loop polygon.

- Use the tool located in the GIS Modeling toolbox.
- To create more accurate medial axes, use a smaller densify distance. I recommend using ~100 meters, or about half the average distance between successive stream intersections on your flow line data. Note that small distances will result in longer processing time and may cause the script to fail due to limited memory.
- Refer to your resulting feature class as *MA_approx*

5.6. Create points of intersection between Exterior Lines and Loop Polygons

- Use the *Intersect* tool to intersect *Exterior_Lines* and *Loop_Polygons*. Make sure the result is a point feature class, and save it as a temporary layer.
- The result will be a multipoint feature class. To use in later tools, run the *Multipart to Singlepart* tool (*Data Management >> Features*) to convert to a single part feature class called *Edge_Points*.

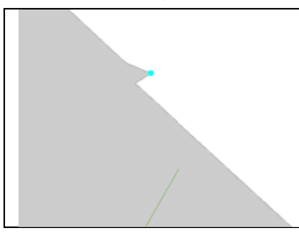
5.7. Extend medial axes to points

5.7a) Automated Extension

- Use the Connect Edge Points to Interior Lines tool in the GIS_Modeling toolbox. Name the result "*MA_Extensions*".

5.7b) Manually Extension

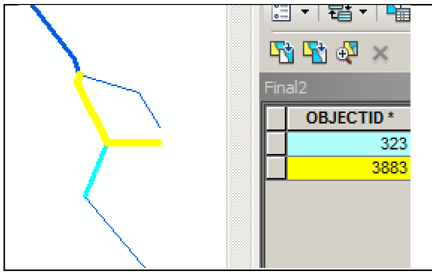
At the time of this writing, the Connect Edge Points to Interior Lines tool does not create medial axis extensions for every polygon – in my test it misses four of the smallest polygons. Here is an example:



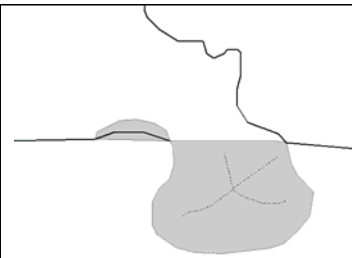
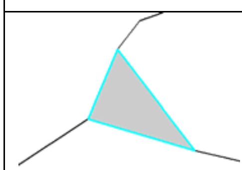
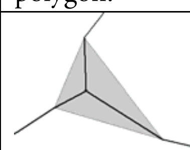
Situation in which medial axis cannot be algorithmically extended to the marked edge point unless a very small densify distance is used. Polygon is gray, and edge point is turquoise. Note that a straight path from the edge point to the medial axis would cross outside of the polygon, which is not allowed. In this case, it is probably quicker to manually create the extension.

The number of missed extensions will decrease in proportion to accuracy of the medial axis approximation, and thus in inverse proportion to the densify distance. To find locations where the medial axis has not been extended to edge points:

- Use the *Select by Location* function to find all features in the *Edge_Points* layer that are adjacent to a feature in *MA_Extensions*.
- Switch the selection to find features in *Edge_Points* that don't have a medial axis extension connect to them.
- Manually add features to *MA_Extensions* to connect each such point to a vertex on the medial axis.
- ➥ **Important:** Connect each point to a vertex of the medial axis. Do not connect to a location between vertices. Why is this necessary? Otherwise, you will end up with a situation similar to that below (but note the example is not quite right, because the line does meet at a vertex, which is ok):

	In this diagram, line 323 (turquoise) is a stream that flows into line 3883 (yellow). However, because the intersection is in the middle of line 3883, it is not detected by the Strahler stream ordering tool. **Since capturing this screenshot, the tools have been modified to capture intersections at vertices. However, intersections in between vertices are still not captured.*

- If a polygon does not have a medial axis, create a medial axis extension feature right through the polygon connecting the exterior lines on either side.

 **Before editing:** The upper left *Loop_Polygon* does not have a medial axis.	 **After editing:** A *MA_Extension* feature has been created in the upper left *Loop_Polygon*. Note that extensions still need to be created in the lower right polygon.
 **Before editing:** *Loop_Polygon* with no medial axis is selected.	 **After editing:** Three *Exterior_Line* features have been extended to connect through the *Loop_Polygon*.

5.8. Weed your medial axes

- Use the *Merge* tool to merge *MA_approx* and *MA_Extensions*. Call the result *MA_Merge*.
- Use the *Unsplit_Line* tool (*Data Management >> Features*) to unsplit *MA_Merge*. Call the result *MA_Unsplit*.
- Use the *Strahler Stream Ordering* tool in the *GIS_Modeling* toolbox to calculate "stream orders" for the *MA_Unsplit* feature class.
- Use *Select by Attributes* to select features with stream order of 1.
- Use the *Select by Location* to ***remove*** features from the selection if they touch (intersect) features in *Edge_Points*.
- Invert your selection. Zoom in to confirm - you should be able to see that selection contains the lines required to connect all *Edge_Points*, but omits most superfluous lines that are not necessary for connectivity.
- Export your selection to a new feature class called *MA_Final*.

5.9. Merge the Different Lines Together

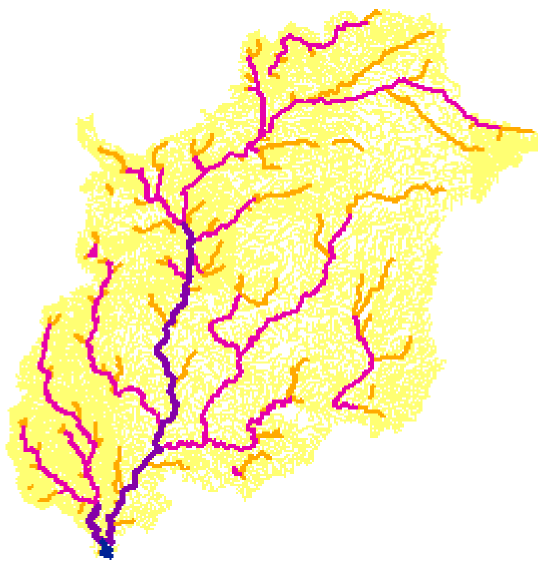
- Use the *Merge* tool to merge together *MA_Final* and *Exterior_Lines*.
- If you've done everything correctly, you now have a topologically connected, acyclic network that can be used to calculate Strahler stream orders.

6. Calculate Strahler Stream Order

- Select the furthest downstream segment of each separate stream system.
- Run the *Strahler Stream Ordering* tool.

- Apply symbology as desired

Here are the results for NHD subregion 0512:



If you look carefully you may be able to see a couple of missing links which need to be connected. This is because the example was constructed prior to working out all of the steps necessary to ensure topological connectivity. You can see, however, that even when there are broken connections, the tool is able to capture much (if not all) of the structure of the streamnetwork.

7. Further Analysis & Visualization

7.1. Streamsheds

- Use the *GIS_Modeling* >> *Strahler Streamsheds* tool to distinguish sub-stream systems at a given Strahler stream order.
- Tabulate the total length of streams in each Strahler order. The relationship between stream order and total stream length can be used to characterize the efficiency of the stream network.