

Functional Linkage of Watersheds and Streams (FLoWS): Network-based ArcGIS tools to analyze freshwater ecosystems

David M. Theobald, Ph.D.
John Norman
Erin Peterson
Silvio Ferraz
Natural Resource Ecology Lab
Colorado State University
Fort Collins, CO 80523-1499
davet@nrel.colostate.edu

Introduction

As part of an US Environmental Protection Agency-funded project (through the STAR program), we have been developing methods to assist EPA and other agencies to address science needs related to the Clean Water Act. A key difficulty of this type of environmental challenge is that traditional experimental designs (manipulated vs. controlled) cannot be conducted because the landscapes are so large and human activities so dominant. As a result, many studies have focused on identifying correlations of co-variates with measured response variables. Our goal is to develop landscape indicators that are useful for predicting aquatic responses, and we hope to assist in the general movement towards tenable hypotheses generated using understanding of ecological processes and are more direct measures of processes.

A critical challenge to develop improved landscape-scale indicators is a clearer representation of watersheds, their hierarchical relationships, and the nonlinearities of condition among different watersheds and along a stream segment (Fausch et al. 2002; Gergel et al. 2002; Allan 2004). Commonly metrics are computed using an entire watershed as the analytical unit, which generates “lumped” metrics such as the percentage urban or percentage agricultural land use, yet these estimates vary widely at a smaller spatial scale (e.g., Richards et al. 1996). Ignoring the spatial heterogeneity and scaling of watersheds has led to somewhat equivocal conclusions regarding general proportions of land use in a watershed as an overall indicator of biological condition.

To this end, we have developed geographic information systems (GIS) tools to assist scientists to quickly and easily generate indicators of aquatic response that capture functional relationships between watershed and streams.

Concepts

Typology of watershed-stream relationships

To examine the behavior of watershed and streams, there are a variety of ways that space has been represented and hydrologic features have been measured: *site*, *watershed*, *distance-base*, and *network*. Commonly, landscape data (from GIS, remotely-sensed) about covariates such as geology, dominant vegetation, elevation, etc. are needed to complement field-based data collected at a *site* or location (e.g., an EPA EMAP site). Occasionally covariate data for features that are nearby or that form the context of a site are needed, such as catchment area, population density, acres of agricultural land use, etc.

A second way to represent hydrographic features in a landscape is in terms of *watersheds* (or catchments). Co-variables are summed or averaged within watersheds (often called “lumped” models). These hydrologic units are used to compute landscape indicators, for example, average road density (Bolstad and Swank), dam density (Jones et al. 1997; Moyle and Randall 1998), connected impervious surface (Wang et al. 2001), or total number of dams within a watershed. These watersheds are often conceived of as overlapping, hierarchical areas defined from the “pour-point” or outlet up to the headwaters or watershed boundary. This conceptualization of watersheds follows directly from the River Continuum Concept (Vannote 1980)¹, where river systems are conceived of as a continuous gradient of physical conditions from headwaters to the mouth of a river. However, typically in practice these watersheds are tessellations of non-overlapping catchment areas, such as the commonly used Hydrologic Unit Codes (also called 8-digit HUCs), where only 55% of the 2,150 cataloguing units are true watersheds -- the rest are called “adjoint” watersheds or interior basins (Seaber et al. 1987). Moreover, flow direction is not represented between true watersheds and downstream adjoint watersheds.

A third way to represent hydrographic features is to explicitly examine the spatial relationships between sites (or locations), which can then be incorporated into a geostatistical model (e.g., Ganio et al. 2005). This is typically accomplished by including not only covariates at a site within a model, but also the measured responses from other nearby sites, conditioned on their distance apart. Typically the spatial relationships are measured by simply straight line distance (as the crow flies) between points (e.g., Olden et al. 2001). Increasingly, distance along the hydrological network (as the fish swims) is computed. Note that both of these are Euclidean (or assume a flat, 2D plane) distances, and hence “Euclidean distance”, though commonly used, is an ambiguous term.

A fourth way is to conceive and represent river systems and aquatic landscapes as a network (Benda et al. 2004). In this sense, relationships between sites can be represented through functional distance measures (e.g., *ver Hoef in press*). For many hydrological processes (not all!) downstream flow direction is an important ecological process, so that distance is not symmetric. Also, including important landscape attributes that modify the degree to which nearby locations are connected is important. This would include topographic considerations such as stream gradient and slope, as well as features that might impede the movement of a species or process such as waterfalls, dams, or certain vegetation types. Representing functional relationships can be done within a network, to recognize that physical conditions along a river are often controlled by the network geometry of the river system (Benda et al. 2004). “One consequence of this interplay [between pattern and process] is the form of functional connectivity found in a landscape. The landscape pattern-process linkage produces spatial dependencies in a variety of ecological phenomena, again mediated by organismal traits.... It is through the integration of these features of landscapes and of organisms that landscape ecology can offer new insights to freshwater ecologists, fostering a closer linking of spatial patterns with ecological processes” (Wiens 2002, pg. 511).

Reach Contributing Areas

Rather than conducting hydrologic analyses using overlapping, hierarchical watersheds, we employ a hydrologic framework composed of a complete, detailed tessellation of reach

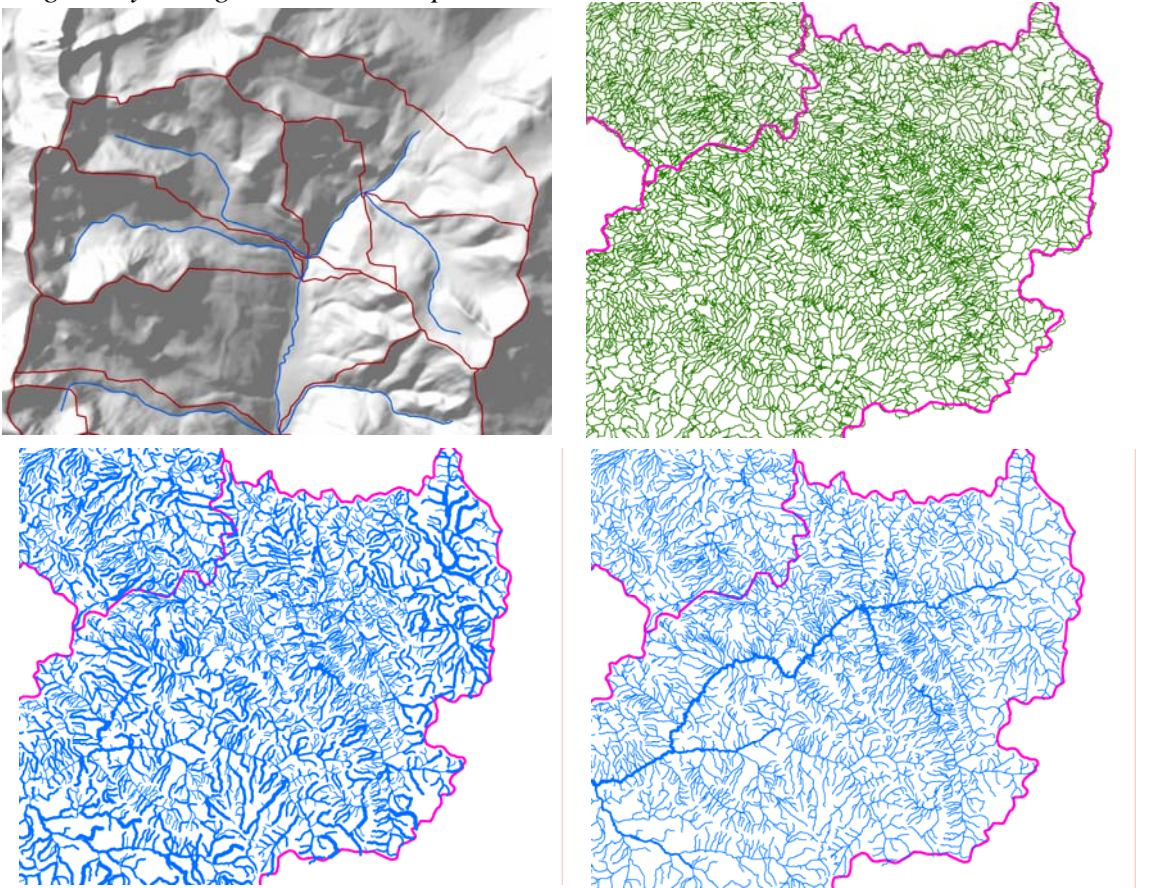
¹ See the following link for a graphic depiction of the River Continuum Concept:
<http://www.epa.gov/owow/watershed/wacademy/acad2000/ecology/s19.html>

catchment areas (RCAs). RCAs are non-overlapping, edge-matching polygons that are drawn so that their boundaries include nearby areas that would likely flow into a given reach. For example, we have generated RCAs for reaches defined in the USGS National Hydrography Dataset (at medium resolution, 1:100,000). In the NHD, a *reach* is defined usually as a significant segment of surface water that has similar hydrologic characteristic (<http://nhd.usgs.gov/chapter1/index.html>). A *transport reach* is delineated by lines that are oriented in the flow direction. Note that a *branched path reach* is generated to represent the 1D flow of water through a waterbody. Operationally in a GIS, a reach is typically represented as a polyline feature representing a unique head-to-confluence, confluence-to-confluence, confluence-to-mouth, or head-to-mouth segment in a river network.

Hydrologic network

There is a one-to-one relationship between reaches and RCAs. This allows functional relationships between watersheds and streams to be easily and explicitly modeled. In one sense, RCAs are simply more finely-resolved HUCs, where watershed conditions are aggregated based on an RCA unit. But the underlying hydrologic network that ties RCAs together is also explicitly represented in a topological relationship table. This allows measures that have been computed locally for each RCA to be accumulated downstream, such as catchment area (Figure 1). Also note that RCAs can be generated from linear networks that represent segments or portions of a stream reach that have uniform discharge, depth, area, and gradient, such as the eight geomorphic classes of Montgomery and Buffington (1997). This allows, for instance, meanders or pool/riffle sequences to be delineated within a confluence-to-confluence stream reach.

Figure 1. Top-left: A close-up view of some reach contributing areas (red lines) that follow ridge lines and are formed around reaches (blue lines). Top-right: RCAs in the upper Colorado basin. Bottom-left: the area of RCAs is joined over to the stream hydrology coverage so that thicker lines represent larger RCAs. Bottom-right: Catchment area (local) is accumulated downstream along the hydrologic network to represent cumulative watershed area.



Landscape network

A network is a data structure used to represent topological relationships between objects or features. Networks typically rely on graph theory, where a set of nodes (or locations) are related through edges (or linkages). A *landscape network* (Theobald 2005) is a simple data structure we have developed to represent geometric network that represent landscape features (including hydrologic networks), which stores the geometry of nodes and edges in addition to topological adjacencies. Note that edges are directed, so hydrologic flow can be represented.

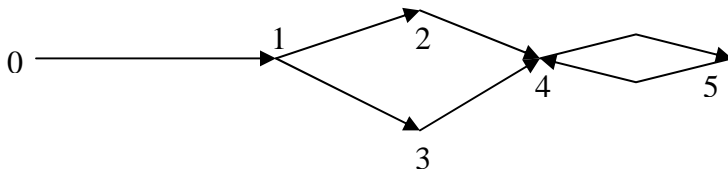
In ESRI's Geodatabase architecture at v8, this type of data structure was called a *geometric network* (Zeiler 1999), and edges were represented as a 1-dimensional polyline, which can be a simple straightline between two nodes or may be a complex "wiggly" line (with >2 vertices) like a stream. In a geometric network, the location where two or more edges intersect is represented by a node, which has a spatial location and associated attributes such as area (Zeiler 1999).

We began developing FLoWs around the geometric network, but we found it was cumbersome to automatically generate networks from our hydrology shapefiles and we did not make use of the supplied "solver methods for most of our analyses. As a result, we eventually opted to generate our own, open, simple network based on a ForwardStar data structure (Ahuja et

al. 1993). A landscape network's topology is defined by geometric coincidence of from/to nodes, and the polylines that connect the nodes can be represent geometry and can also cross (be non-planar). Note that we represent just simple edges (not complex edges of Geometric Network).

For example, the figure and table below represents a graph, where nodes are represented by numerical values. Note that the relationship table records are ordered by from feature.

Figure 2. The relationship table for a simple landscape network.



FID	FFeat	TFeat
0	0	1
1	1	2
2	1	3
3	2	4
4	3	4
5	4	5
6	5	4

FLoWS v1 Toolbox

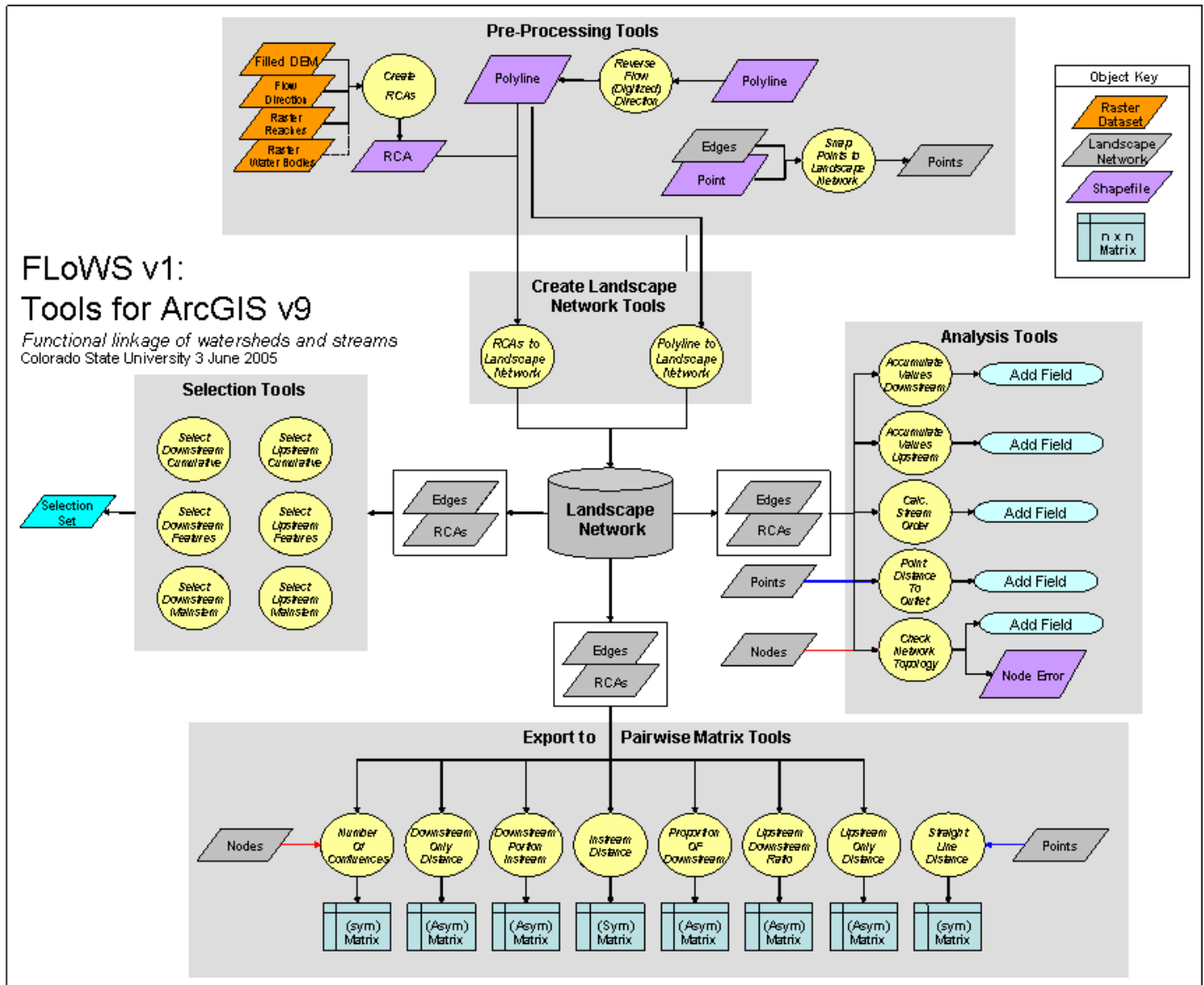
Software environment

The FLoWS v1 software was written as a Geoprocessing toolbox, written in Python (v2.1) using ArcGIS v9, Service pack 3. Nearly all tools in FLoWS require the basic ArcGIS desktop (no extensions). Only two tools (*Create RCAs* and *Fill DEM and Build Flow Direction Raster*) require the Spatial Analyst extension. If you are creating any landscape network (LN) -- which is a personal Geodatabase, an ArcINFO license is required. The other tools (query, selection, analysis, export, etc.) can function with just an ArcView license.

Structure of FLoWS v1

The FLoWS toolbox consists of five toolsets (Figure 3) that allow users to pre-process common GIS datasets, create LNs, conduct various selections and analyses on LNs, and export data in formats useful for statistical analysis. The *pre-processing* toolset contains miscellaneous tools that are useful to edit and convert raw datasets into appropriate inputs for other FLoWS tools. The *selection* tools allow interactive queries or selections on LNs within an ArcMap document. This allows users to create new selection set that represents upstream or downstream topological relationships to be summarized or used in further analysis. The *analysis* tools allow users to perform network-based analyses. These routines typically populate a user-defined field for a defined LN feature class. The *export* tools evaluate point to point relationships within a LN and create a comma delimited $n \times n$ matrix of distance values between pairs of locations.

Figure 3. Flowchart and structure of the FLoWs v1 toolbox for ArcGIS v9.



Pre-processing toolset

Create RCAs

This tool generates a polygon shapefile of Reach Catchment Areas (RCAs) for every unique polyline within an input hydrologic network. An RCA represents a sub-component (polygon) of a watershed that drains directly into a given stream segment. This script requires a filled DEM, a flow direction raster (output from *Fill DEM and BUILD Flow Direction Raster*), and rasterized hydrologic features (stream reaches and water bodies) to calculate RCAs. The output is a polygon shapefile that represents RCAs. These RCAs can be linked to the input reach features via the unique ID (field) used to generate the reach raster dataset. This linkage is used in the *RCAs to Landscape Network* tool to build the downstream relationships between RCAs and reaches within a given drainage network.

Fill DEM and Build Flow Direction Raster

This tool processes a DEM and based on a user-defined “fill z-limit” value (to fill in pits) to generate a filled DEM and flow direction raster. This is a preprocessing tool for the *Create RCAs* tool.

Reverse Flow (Digitized) Direction

This tool reverses the digitized direction of the input polyline features that represent a hydrologic network. This tool is useful if a hydrologic network has been digitized upstream (so that the starting point is lower elevation than the ending point) and it needs to be imported into an LN that builds feature to feature relationships assuming downstream flow.

Snap Points to Landscape Network Edges

This tool allows features represented by points (such as dams, stream gages, sample locations, point-source pollution, mines, etc) to be incorporated into the LN by associating each point to an edge via dynamic segmentation. Dynamic segmentation intersects the point with the closest edge segment, moves the point to that location and calculates a point’s distance ratio from the end of an edge. This allows for point to point distances to be calculated along a network without needing to permanently alter the hydrologic reaches by cutting them into two reaches at each point location.

Create landscape network toolset

Polyline to Landscape Network

This tool generates a LN based on geometric coincidence of the input polyline features. The output LN is a personal Geodatabase composed of four elements: an edge feature class, node feature class, relationships table, and a node relationships table. The edge feature class is the input polyline feature class with a relational field *rid*, which relates it to the relationships table. The node feature class is a point feature class of all the end points (or nodes) within the input polyline feature class. The field *pointid* joins it to the noderelationships table. The relationships table is a sorted table (as a ForwardStar structure) that represents flow or movement direction. It has two fields: *fromfeat* (from feature) and *tofeat* (to feature), which indicate feature to feature relationships within the edge feature class. The noderelationships table relates nodes to edges within the LN.

RCAs to Landscape Network

This tool generates an LN for RCAs based on geometric coincidence of the input polyline (hydrologic network) features. The output LN is a personal Geodatabase composed of four elements: an edge feature class, node feature class, relationships table, and a node relationships table. The RCA feature class is the input polyline feature class with a relational field *rid*, which relates it to the relationships table. The node feature class is a point feature class of all the end points (or nodes) within the input polyline feature class. The field *pointid* joins it to the noderelationships table. The relationships table is a sorted table (as a Forward Star structure) that represents flow or movement direction. It has two fields: *fromfeat* (from feature) and *tofeat* (to feature), which indicate feature to feature relationships within the edge feature class. The noderelationships table relates nodes to edges within the LN.

Selection toolset

Select Downstream Cumulative

This tool adds features to the selected set that are downstream from the selected features (as defined in ArcMap). The user needs to define a numeric field and a threshold value such that features will be included in the selection if downstream features have a cumulative value less than or equal to the threshold value.

Select Downstream Features

This tool adds features to the selected set that are downstream from the selected features (as defined in ArcMap). Like the *Select Downstream Mainstem* tool, this tool adds features that are directly downstream (along the mainstem), but also features that are upstream of added features. For example, all mainstem and tributary reaches below a dam can be identified (assuming the initial selected feature represents a reach with a dam on it).

Select Downstream Mainstem

This tool adds features to the selected set that are (strictly) downstream along the mainstem from the selected features (as defined in ArcMap).

Select Upstream Cumulative

This tool adds features to the selected set that are upstream from the selected features (as defined in ArcMap). The user needs to define a numeric field and a threshold value such that features will be included in the selection if upstream features have a cumulative value less than or equal to the threshold value.

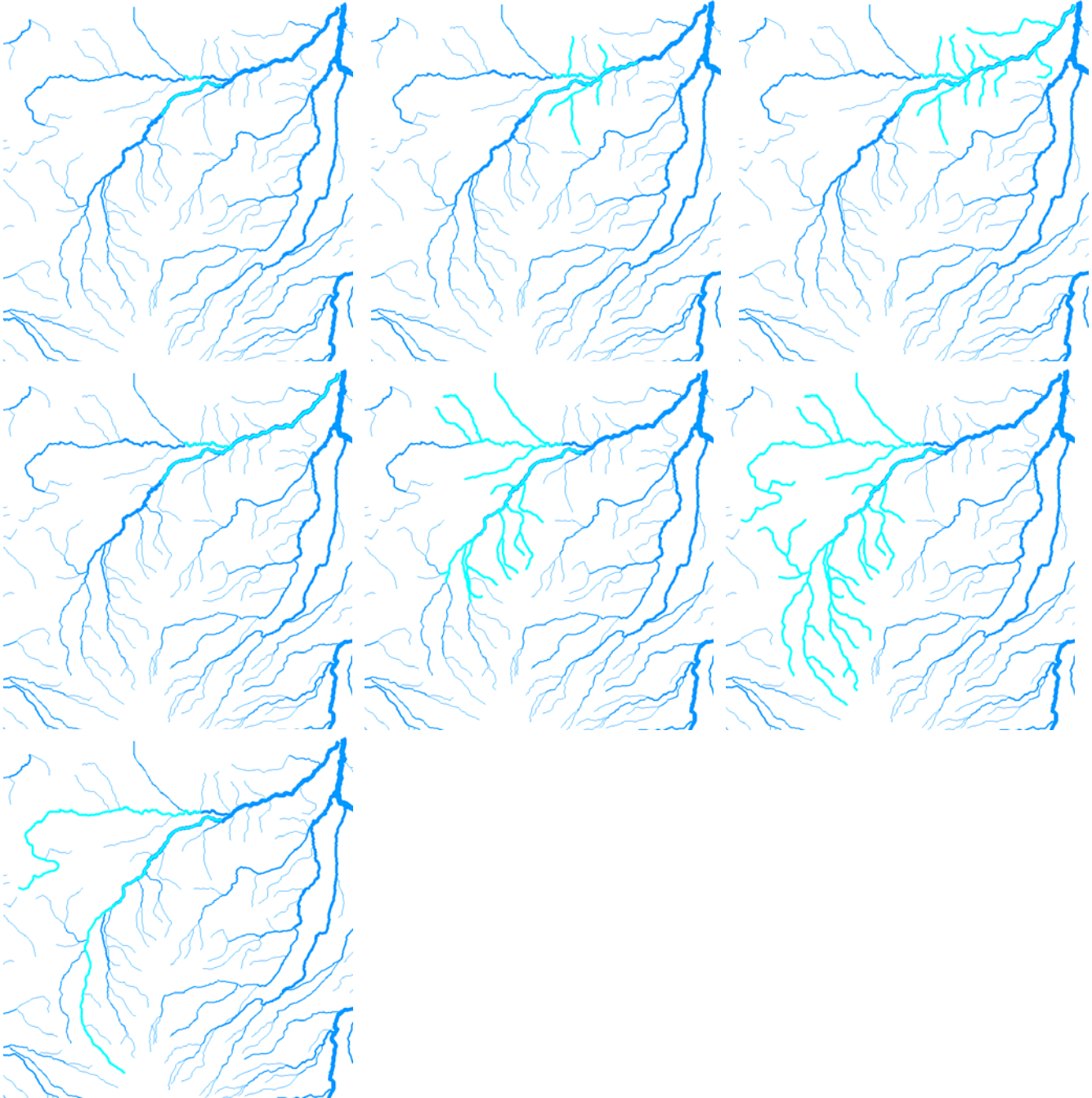
Select Upstream Features

This tool adds all features to the selected set that are upstream from the selected features (as defined in ArcMap).

Select Upstream Mainstem

This tool adds mainstem features to the selected set that are upstream from the selected features (as defined in ArcMap). The user needs to define a numeric field so that mainstem features are defined by finding the largest upstream accumulated value at each confluence upstream from the initial selection.

Figure 4. Selection of features using upstream and downstream flow connectivity: initial feature(s) that are selected (top left); downstream cumulative within 4 km (top center); downstream features (top right); downstream mainstem (middle left); upstream cumulative within 10 km (middle center); upstream features (middle right); upstream mainstem using cumulative length (lower left).



Analysis toolset

Accumulate Values Downstream

This tool accumulates values from a user-defined field downstream and populates the values of a new field for each feature with its downstream accumulated value.

Accumulate Values Upstream

This tool accumulates values from a user-defined field upstream and populates the values of a new field for each feature with its upstream accumulated value.

Calculate Downstream Distance From Points to Basin Outlet

This tool calculates the distance (along the mainstem) from each point in a drainage to its outlet and populates a user-defined field with the distance value. Points must be coincident (snapped) on a network line. The input point features must be a snapped and represented as a point feature class within the LN Geodatabase.

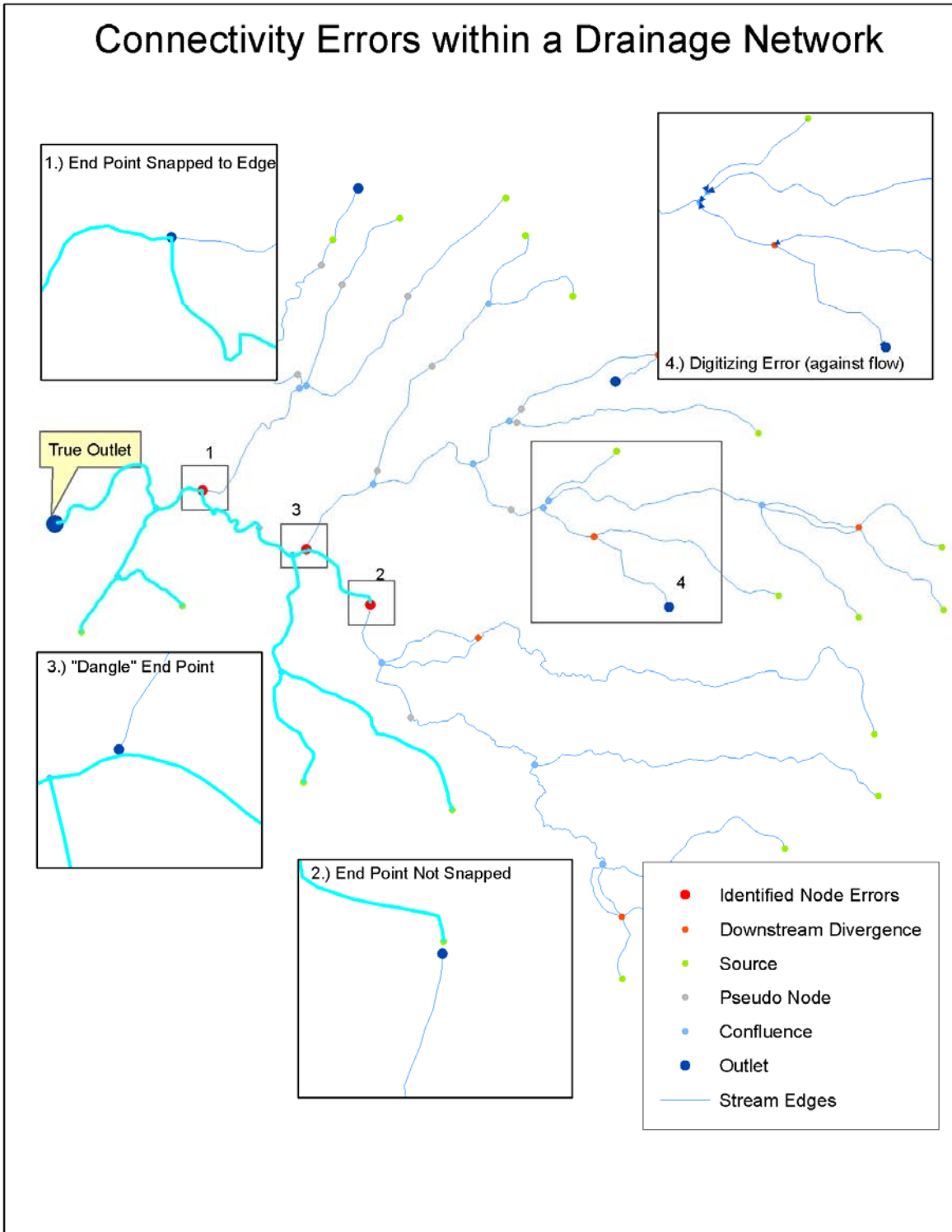
Calculate Stream Order (Strahler)

This tool calculates Strahler stream order for each reach within an LN feature class. It populates a user-defined field with a feature's associated stream order. Before running this tool, be sure to confirm the network topology by using the *Check Network Topology* tool which finds errors and adds a field that describes the node types used to determine stream order in braided channels (also be sure to visually inspecting the network by using various Selections).

Check Network Topology

This tool searches the node and edge feature classes for a LN for topological errors based on geometric coincidence (Figure 5). It attributes a user-defined field in the node feature class with node categories or types, which include: source, outlet, pseudo node, downstream divergence, converging stream, and confluence. *Source* nodes occur at the starting point of an edge if there is not an adjacent edge with its starting point coincident (i.e. does not have an upstream input). *Outlet* nodes occur at the ending point of an edge that does not have an adjacent edge with a starting point coincident (i.e. does not have a downstream output). *Pseudo* nodes are locations where a starting point of an edge is coincident with the two node of only one other edge. *Confluence* nodes occur where two or more edges (ending points) are coincident with exactly one other edge (starting point). *Downstream divergence* nodes represent locations where an edge's ending point is coincident with two or more edges' starting points. *Converging* nodes occur when two or more edges' ending points are coincident (with no starting point coincidence). Possible topological errors are computed and stored in a shapefile called *node_errors.shp* located in the same directory as the input LN. Topological node and edge errors include outlet nodes that are within a user-specified search tolerance distance of source nodes, and nodes that are snapped to an edge without a coincident node (i.e. an edge is connected to another edge at its middle). Note that it is also important to visually display node errors (especially to identify erroneous outlets at headwater streams if they are digitized in the wrong direction) and accumulated flow. Also, selections up and downstream often also identify possible topological errors. It must be stressed that the quality of an analysis rests on the topological correctness of the hydrologic network, which is not readily apparent (nor required) for most typical GIS analyses.

Figure 6. Common topological errors can be identified using the Check Network Topology tool. Connectivity errors may result in erroneous network calculations. The Landscape Network's relationship tables are generated by examining geometric coincidence of features, in particular their starting points (from nodes) and ending points (to nodes). Basic connectivity errors can fragment the topological connectivity of a drainage network, such as line features not being properly snapped to the starting or ending points of other features. These errors cause a network to be fragmented into false drainage basins. For example, the selected upstream features of the true outlet (shown below) do not include all line features in the drainage network. Also, features (lines) that are digitized in the opposite direction from the rest of the network often create falso outlets.



Export toolset

Downstream Only Distance (Asymmetric)

This tool creates an asymmetric matrix of downstream-only distances from all pairs of points in the input feature class based on a LN feature class (e.g., edges or RCAs). Distances are computed using values from the user-defined field and are output as a comma delimited text file. If a pair of points is not connected, their distance value is -1.

Downstream Portion of Instream Distance (Asymmetric)

This tool creates an asymmetric matrix that provides only the downstream portion of the instream distance between all pairs of points in the input feature class based on a LN feature class (e.g., edges or RCAs). Distances are computed using values from the user-defined field and are output as a comma delimited text file.

Instream Distance (Symmetric)

This tool creates a symmetric matrix of instream distances from all pairs of points in the input feature class based on a LN feature class (e.g., edges or RCAs). Distances are computed using values from the user-defined field and are output as a comma delimited text file.

Number of Confluences (Symmetric)

This tool creates a symmetric matrix that computes the number of confluences between all pairs of points (upstream and downstream) in the input feature class based on a LN feature class (e.g., edges or RCAs). Distances are computed using values from the user-defined field and are output as a comma delimited text file. A confluence is defined a junction where the ending points of two lines are coincident with the starting point of a single line.

Proportion of Downstream Only Distance (Asymmetric)

This tool creates an asymmetric matrix that provides the downstream proportion (or percent) of the total instream distance between all pairs of points in the input feature class based on a LN feature class (e.g., edges or RCAs). Distances are computed using values from the user-defined field and are output as a comma delimited text file.

Ratio of Upstream to Downstream (Asymmetric)

This tool creates an asymmetric matrix that provides the ratio of the upstream to downstream distance between all pairs of points in the input feature class based on a LN feature class (e.g., edges or RCAs). Distances are computed using values from the user-defined field and are output as a comma delimited text file. This tool is typically used to compute the ratio of an accumulated field (e.g., waterbasin area).

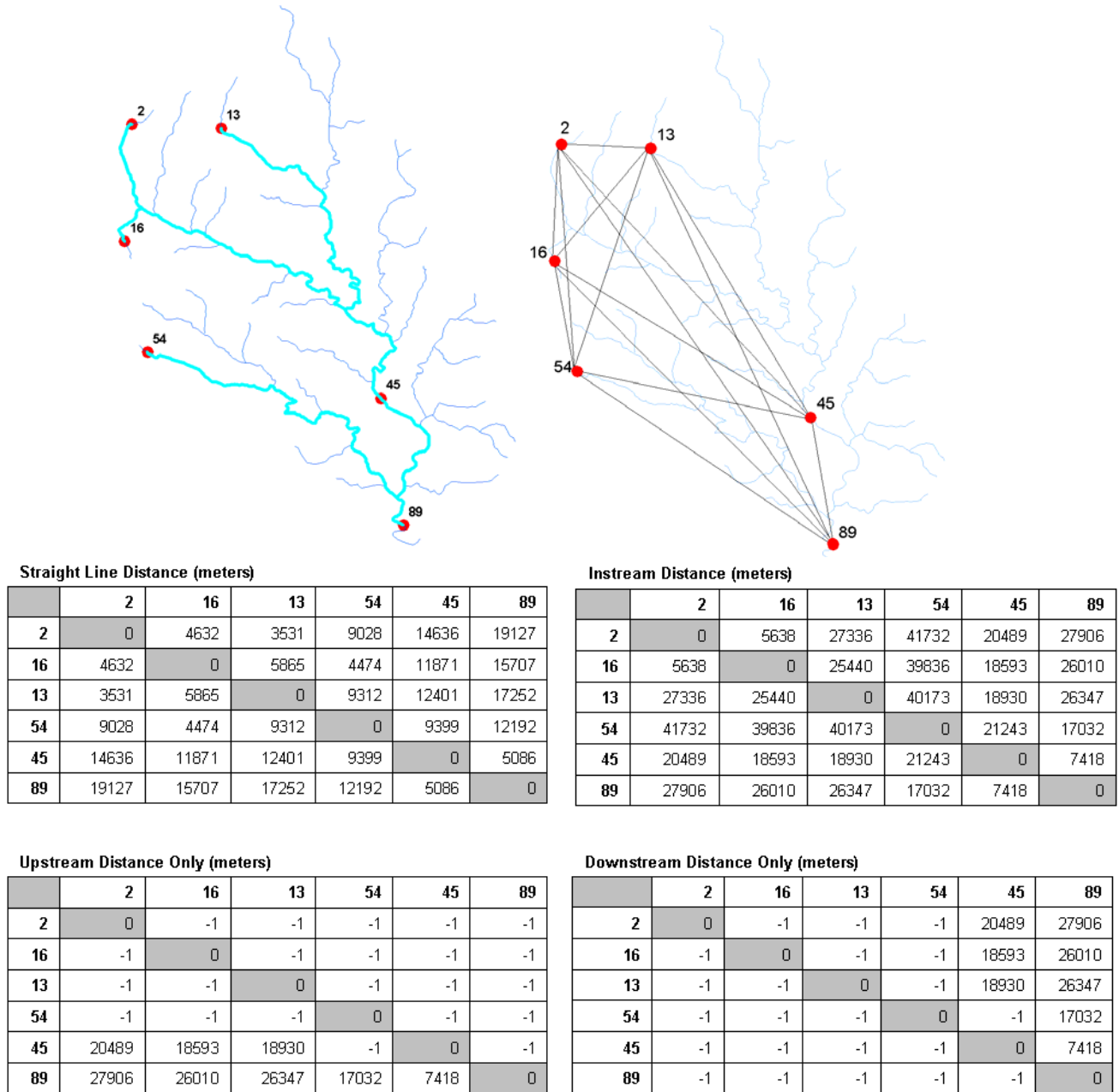
Straight Line Distance (Symmetric)

This tool creates a symmetric matrix that provides the straight-line distance (computed in map units) between all pairs of points in the input feature class based on a LN feature class (e.g., edges or RCAs).

Upstream Only Distance (Asymmetric)

This tool creates an asymmetric matrix of upstream-only distances from all pairs of points in the input feature class based on a LN feature class (e.g., edges or RCAs). Distances are computed using values from the user-defined field and are output as a comma delimited text file. If a pair of points is not connected, their distance value is -1.

Figure 6. An example of computing pair-wise distances and resulting matrices.



Literature cited

- Ahuja, R.K., T.L. Magnanti, and J.B. Orlin. 1993. *Network flows: Theory, algorithms, and applications*. Prentice-Hall: Upper Saddle River, New Jersey.
- Allan, D. 2004. Landscapes and riverscapes: The influence of land use on stream ecosystems. *Annual Review of Ecology, Evolution, and Systematics* 35: 257-284.

- Benda, L., N.L. Poff, D. Miller, T. Dunne, G. Reeves, G. Pess, and M. Pollock. 2004. The Network Dynamics Hypothesis: how channel networks structure riverine habitats. *BioScience* 54(5): 413-427.
- Bolstad, P.V. and W.T. Swank. 1997. Cumulative impacts of landuse on water quality in a Southern Appalachian watershed. *Journal of the American Water Resources Association* 33(3): 519-533.
- Fausch, K.D., C.E. Torgersen, C.V. Baxtger, and H.W. Li. 2002. Landscapes to riverscapes: bridging the gap between research and conservation of stream fishes. *BioScience* 52: 483-498.
- Ganio, L.M., C.E. Torgersen, and R.E. Gresswell. 2005. A geostatistical approach for describing spatial pattern in stream networks. *Frontiers in Ecology and Environment* 3(3): 138-144.
- Gergel, S.E., M.G. Turner, J.R. Miller, J.M. Melack, and E.H. Stanley. 2002. Landscape indicators of human impacts to riverine systems. *Aquatic Sciences* 64: 118-128.
- Jones, K.B., K.H. Ritters, J.D. Wickham, R.D. Tankersley, Jr., R.V. O'Neill, D.J. Chaloud, E.R. Smith, and A.C. Neale. *An Ecological Assessment of the United State Mid-Atlantic Region: A Landscape Atlas*. US Environmental Protection Agency, EPA/600/R-97/130.
- Montgomery, D.R. and J.M. Buffington. 1997. Channel-reach morphology in mountain drainage basins. *Geological Society of America Bulletin* 109: 596-611.
- Moyle, P.B. and P.J. Randall. 1998. Evaluating the biotic integrity of watersheds in the Sierra Nevada, California. *Conservation Biology*. 12:1318-1326.
- Olden, J., D.A. Jackson, and P.R. Peres-Neto. 2001. Spatial isolation and fish communities in drainage lakes. *Oecologia* 127: 572-585.
- Richards, C., L.B. Johnson, G.E. Host. 1996. Landscape-scale influences on stream habitats and biota. *Canadian Journal of Fishery and Aquatic Science* 53: 295-311.
- Seaber, P.R., F.P. Kapinos, and G.L. Knapp. 1987. *Hydrologic Unit Maps*. US Geological Survey Water-Supply Paper 2294.
- Vannote, R.L., G.W. Minshall, K.W. Cummins, J.R. Sedell, and C.E. Cushing. The river continuum concept. *Canadian Journal of Fisheries and Aquatic Science* 37: 130-137.
- Ver Hoef, J.M., E. Peterson, and D.M. Theobald. (*in press*). Spatial statistical models that use flow and stream distance. *Environmental and Ecological Statistics*.
- Wang, L., J. Lyons, P. Kanehl, and R. Bannerman. 2001. Impacts of urbanization on stream habitat and fish across multiple spatial scales. *Environmental Management* 28(2):255-266.
- Wiens, J. 2002. Riverine landscapes: taking landscape ecology into the water. *Freshwater Biology* 47: 501-515.
- Zeiler, M. 1999. *Modeling our world: The ESRI guide to geodatabase design*. Redlands, CA: ESRI Press.