STARS: An ArcGIS toolset used to calculate the spatial information needed to fit spatial statistical models to stream network data

Erin E. Peterson
CSIRO, Division of Mathematics,
Informatics and Statistics

Jay M. Ver Hoef
NOAA, Alaska Fisheries
Science Center

Abstract

Goals: to describe the **STARS** ArcGIS geoprocessing toolset, which is used to calculate the spatial information needed to fit spatial statistical models to stream network data using the **SSN** package. This document may be used as an introduction to new users. The methods used to calculate the spatial information and format the final `.ssn` object are also explicitly described so that users may create their own `.ssn` object using other non-proprietary software.

Keywords: GIS, spatial statistical modeling, streams, SSN.

1. Introduction

Spatial autocorrelation is an intrinsic characteristic in freshwater stream environments where nested watersheds (i.e., the entire land area that drains to a single location within the stream), dendritic network structure, and the directional flow of water may produce spatial relationships that are not explained by Euclidean distance. Yet, many common autocovariance functions used in spatial models are statistically invalid when Euclidean distance is replaced with hydrologic, or within network, distance (Ver Hoef, Peterson, and Theobald 2006). This issue made it necessary to develop new spatial statistical methodologies for stream networks, which permit valid covariances to be generated based on a variety of hydrologic, or within-network, relationships (Ver Hoef and Peterson 2010).

Fitting spatial statistical models to stream network data is challenging because it requires multidisciplinary skills in aquatic ecology, geographic information science, and spatial statistics. In addition, specialized geographic information system (GIS) tools are needed to generate the spatial information needed to fit spatial models to stream network data. Two ArcGIS version 9.3 (ESRI 2009) geoprocessing toolboxes have been provided to help users generate these spatial data: the Functional Linkage of Waterbasins and Streams (**FLoWS**) toolbox (Theobald, Norman, Peterson, Ferraz, Wade, and Sherburne 2006) and the Spatial Tools for the Analysis of River Systems (**STARS**) toolbox. The **FLoWS** toolbox is a set of graph theoretic-based analysis tools that functionally link aquatic and terrestrial components of the landscape based on hydrologic processes (Theobald, Norman, Peterson, Ferraz, Wade, and Sherburne 2006). These tools provide an efficient framework for navigating throughout the network, which makes it possible to calculate a variety of attributes related to network

distance, flow direction, and terrestrial contributing areas.

The **STARS** toolbox was developed to take advantage of the **FLoWS**-based framework and provides tools to generate and format the feature geometry, attribute data, and topological relationships of GIS datasets so that they may be used to fit spatial statistical models to streams data. The **STARS** tools create a new directory to store this information, which we refer to as a `.ssn` object. Once the data has been reformatted and exported as a `.ssn` object, it can be imported and analyzed in R statistical software (R Development Core Team 2010) using the **SSN** package (Ver Hoef, Peterson, Clifford, and Shah in review). This document provides an overview and introduction to the **STARS** toolset and the methods used to create the `.ssn` object.

2. Mathematical Framework and Notation

We represent each stream segment as a line (i.e., edge) bounded by nodes, with multiple stream segments forming a dendritic stream network (Figure 1a). Water flows downstream and so the lines have direction. All of the directed lines within a network drain to a single most-downstream location, referred to as the stream outlet (Figure 1a). Any location on a stream network can be connected by a continuous line to the stream outlet, and hence distance from the outlet is simply the length of that line. We define this as "upstream distance".

There are an finite number of stream segments within a network and we index them arbitrarily using a reach identifier (rid), $j = 1, 2, \ldots, n$ (Figure 1b). Many locations will have the same upstream distance in a dendritic stream network. Therefore, we denote each location as x^j, which is the upstream distance on the jth stream segment, in order to uniquely define each location and keep track of upstream distance. It is also convenient to arbitrarily assign indices to points in the network, which we denote as x_i^j for the ith point, dropping the superscript when knowledge of stream segment is unnecessary. The most downstream location on the jth segment is denoted as ℓ^j, and the most upstream location as u^j. As an example, $\ell^0 = \ell^3 = u^4 = x_{11}$ (Figure 1b). The whole set of stream segment indices is denoted as I. The index set of stream segments upstream of segment j, including j is denoted U_j, while the index set of stream segments upstream of segment j, excluding j is denoted as U_j^*. Likewise, the index set of stream segments downstream of segment j, including j is denoted as D_j, while the index set of stream segments downstream of segment j, excluding j is denoted as D_j^*. As an illustration, $U_4^* = \{0, 1, 3\}$ and $D_4 = \{4, 5\}$ in Figure 1b.

3. The GIS-based spatial modeling framework

A landscape Network (LSN) is a directional graph used to represent spatial context and relationships with additional geographic information (Theobald, Norman, Peterson, Ferraz, Wade, and Sherburne 2006). It is created using the **FLoWS** toolset and stored as an Environmental Systems Research Institute (ESRI) ArcGIS personal geodatabase (ESRI 2009). The personal geodatabase is stored as an Access database, which may contain point, polyline (a.k.a. line), and polygon feature classes. Each feature class is made up of similar spatial features and an Access table (i.e., attribute table), which contains attribute data for every spatial feature.

The nodes feature class represents topologic breaks in the stream network such as confluences,
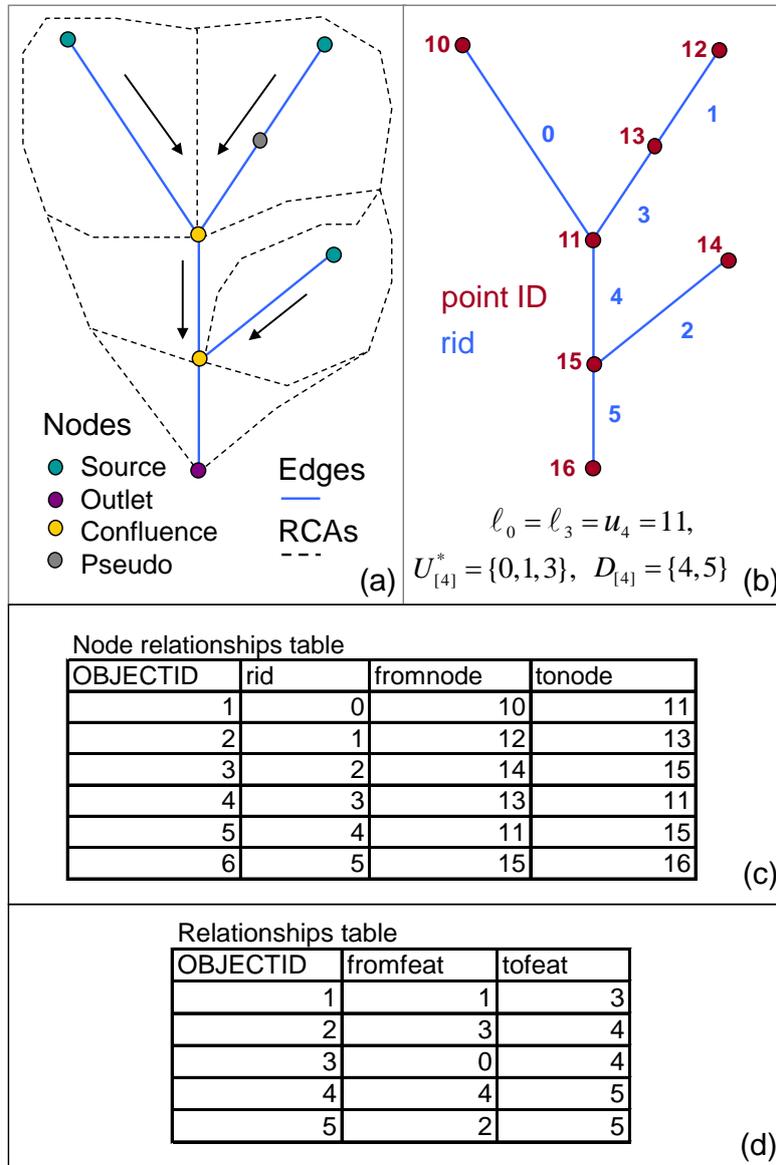
Figure 1: A stream network is made up of a group of directed edges that drain to a single stream outlet (a). Source and outlet nodes occur at the end points of the network, while confluence and pseudo nodes mark locations where edges intersect one another (a). Reach contributing areas (RCAs) form a tessellation of non-overlapping polygons (a). There is a one-to-one relationship between edges and RCAS, which represent the aerial extent that would theoretically contribute overland flow to a given edge. An identifier is allocated to each edge (rid) and node (pointID) in the landscape network (b). The geographic coincidence of nodes and edges is recorded in the noderelationships table (c), while the spatial relationship of edges to one another is stored in the relationships table (d). Both the noderelationships (c) and relationships (d) tables retain information concerning flow direction in the network.

stream sources, or stream outlet points, while edges represent flow paths (i.e., line segments) from node to node (Figure 1a). The stream reach contributing area (RCA) corresponds to the aerial extent that would theoretically contribute overland flow to a given edge in the absence of other hydrologic processes such as infiltration or evaporation. RCAs have a one-to-one relationship with edges, and form a non-overlapping, contiguous tessellation of polygons (Figure 1a). The **FLoWS** toolset also allows RCAs to be incorporated into the LSN, but they do not need to be explicitly included for spatial statistical modeling. Instead, landscape characteristics such as land use, elevation, or road density, can be summarized as areas, means, or counts for each RCA, RCA_j, and then recorded in the edges attribute table.

The LSN is based on a ForwardStar data structure (Ahuja, Magnanti, and Orlin 1993), which provides a way to store, search, and calculate new spatial information based on the geographic coincidence of features. These spatial relationships are stored within three tables, which may be related to one another through the node and edge identifiers (`pointID` and `rid`, respectively). The `nodexy` table contains four attributes: the `OBJECTID` (internal GIS identifier), `pointID`, and the x-, y-coordinates for each node. The `noderelationships` table links the nodes to their associated edges based on the `rid`, j, and the `pointID` (Figures 1b,c), while the `relationships` table represents the downstream flow paths from edge to edge, which are based on the digitized direction of the line segments. Topological errors are common in GIS data and must be corrected before the LSN is constructed to ensure that the `noderelationships` (Figures 1b,c) and `relationships` (Figures 1b,d) tables accurately represent direction and connectivity within the network. Together, these tables provide a wealth of information about the topology of the network, including the spatial location of each node, the most upstream and downstream location on each edge (`fromfeat` $= u^j$ and `tofeat` $= \ell^j$, respectively), the direction of the edges, and the spatial relationship between edge features.

One challenge of working with GIS data is that survey sites collected within a stream usually do not intersect the edge, or segment, where they were collected. This is a common phenomenon that can result for a variety of reasons. Coordinates collected using global positioning system (GPS) units typically have uncertainty associated with them, which translates to locational uncertainty in the site data. Streams have area (i.e., both length and width), but are often represented by lines in a GIS. Consequently, survey sites located on the banks of a stream may not fall directly on a line segment. There are also various degrees of mapping errors and generalizations in digital streams datasets, such as the absence of small tributaries and the homogenization of form. The magnitude of these errors is dependent upon the spatial resolution of the dataset. In addition, the physical location of streams is temporally dynamic, with some segments meandering from their mapped position over time. Regardless of the error source, the survey sites must fall exactly on an edge so that they may be incorporated into the LSN framework.

The **FLoWS** tools are used to modify site locations and incorporate site data into the LSN (Theobald, Norman, Peterson, Ferraz, Wade, and Sherburne 2006). This includes both observed site locations where measurements were collected, as well as locations where predictions will be generated. Each site, s_i, is associated with an edge using dynamic segmentation. The process involves intersecting the site with the closest edge segment, physically moving the site to the new location, x_i^j, and calculating the distance ratio, r_i, from the most-downstream location on the edge to the site location:

$$r_i = \frac{d(\ell^j, x_i^j)}{L_j} \tag{1}$$

where $d(\ell^j, x_i^j)$ is the distance travelled along edge j (i.e., hydrologic distance) between locations ℓ^j and x_i^j and L_j is the total length of the jth edge, $L_j = d(\ell^j, u^j)$.

When the tool has finished successfully, a new point feature class is created in the LSN containing the modified site data. The sites attribute table contains the original attribute data, as well as, a number of new fields; the most relevant of which are the `ratio` (field type = double), `rid` (field type = long), `NEAR_X` and `NEAR_Y` (field type = double) fields. The `rid` field indicates which edge the site has been moved to, while `ratio` provides the exact location along the edge. These two pieces of information allow attributes to be calculated for site-specific locations. In addition, the `NEAR_X` and `NEAR_Y` fields provide the x-, y-coordinates for the modified site locations, which are used to calculate the Euclidean distance matrix in the **SSN** package.

An LSN must contain six datasets before it can be used to calculate the spatial information needed to fit the spatial statistical models described by Ver Hoef and Peterson (2010): three feature classes representing the edges, nodes, and survey sites, as well as, three Access tables containing information about the spatial relationship between features (Figure 2). An additional feature class representing prediction locations may also be included. This information is used to navigate up and downstream between features, allowing new attributes to be generated based on spatial context and the relationship between features. As such, the LSN provides a powerful and flexible GIS-based hydrologic modeling framework for the **STARS** toolset to operate on. For a more detailed description of the LSN, please see Theobald, Norman, Peterson, Ferraz, Wade, and Sherburne (2006). For step-by-step instructions on how to calculate an LSN appropriate for spatial statistical modeling in stream networks, please see Peterson (2011).
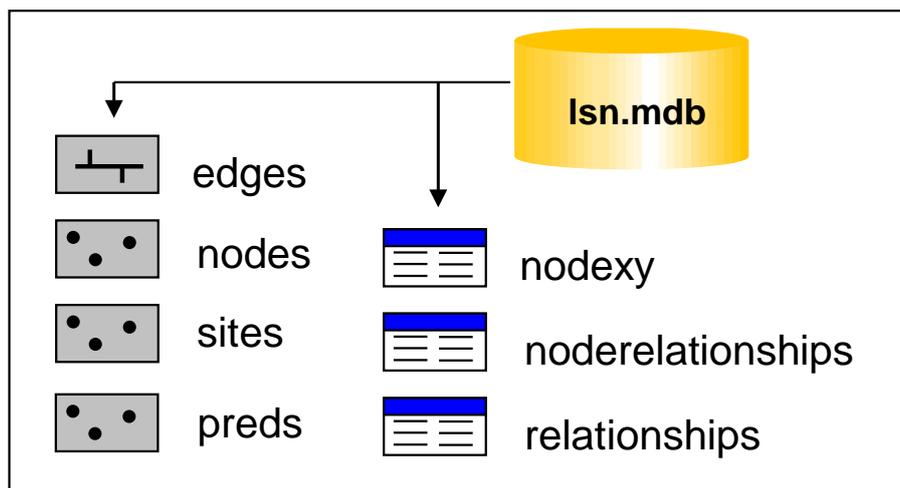

Figure 2: A landscape network (LSN) must contain six datasets before it can be used to calculate the data needed to fit the spatial statistical models: three feature classes: edges, nodes, and sites, as well as three Access tables: nodexy, noderelationships, and relationships. An additional feature class representing prediction locations (preds) may also be included.

4. STARS Geoprocessing Toolset

The **STARS** toolset is used to generate and format the spatial data needed to fit spatial statistical models to streams data in the **SSN** package. These data include information about hydrologic distances (with flow-direction preserved), the spatial additive function used to calculate the spatial weights, and the covariates for all observed and prediction locations in the stream network. In addition, the topological structure of the network is exported to a format that can be efficiently stored, accessed, and analyzed in the **SSN** package. The **STARS** toolset was developed for ArcGIS version 9.3.1 (ESRI 2009), with an ArcInfo license, and the **Spatial Analyst** extension. It was developed for a Windows operating system, with scripts written in Python version 2.5 (Hammond and Robinson 2000). A Python editor for Windows, PythonWin (Hammond and Robinson 2000), must also be installed to successfully run the toolset.

The **STARS** toolset contains eight tools that are used to transform an LSN into a `.ssn` object, which is the data structure the **SSN** package has been designed to utilize. The tools are grouped into three general categories: Pre-processing, Calculate, and Export.

4.1. Pre-processing

Identify Complex Confluences

It is *extremely* critical that the edges are topologically correct to ensure that hydrologic distances and spatial relationships are calculated properly using the **STARS** toolset. The process of topologically correcting the data can easily be the most time-consuming aspect of the modeling process, especially for large datasets. There are a number of useful tools provided in the **FLoWS** toolset to help identify and remove errors (Theobald, Norman, Peterson, Ferraz, Wade, and Sherburne 2006) and so the majority of the pre-processing occurs before the final LSN is generated. However, LSNs used to create `.ssn` objects have two unique topological restrictions, which are not considered a true topological error in a GIS or within a LSN. First, converging stream nodes are not allowed. These nodes occur at the downstream node of two edges that converge (Figures 3a and b), but do not flow into another downstream edge. This commonly occurs at the boundaries of the streams dataset (Figure 3a) or may be the result of topological errors within the stream network (Figure 3b). Converging nodes are identified using the **FLoWS** `Check Network Topology` tool and must be manually removed when the LSN is being generated. The second restriction is that only two edges may converge and flow into a single downstream edge at a confluence. The `Identify Complex Confluences` tool has been included in the **STARS** toolset to help users to identify LSN nodes that violate this condition (Figure 3c). The tool produces a text file that contains the `pointID` for every node with more than 2 upstream edges. Once these errors have been identified, they must be manually edited and a new LSN generated before a `.ssn` object can be created. Note, only the topological relationships shown in Figure 1a are permitted in an LSN used to generate a spatial statistical model using the **SSN** package.

4.2. Calculate

Six tools are provided to calculate the spatial data necessary for spatial statistical modeling: `Watershed Attributes`, `Segment PI`, `Additive Function - Edges`, `Additive Function -`
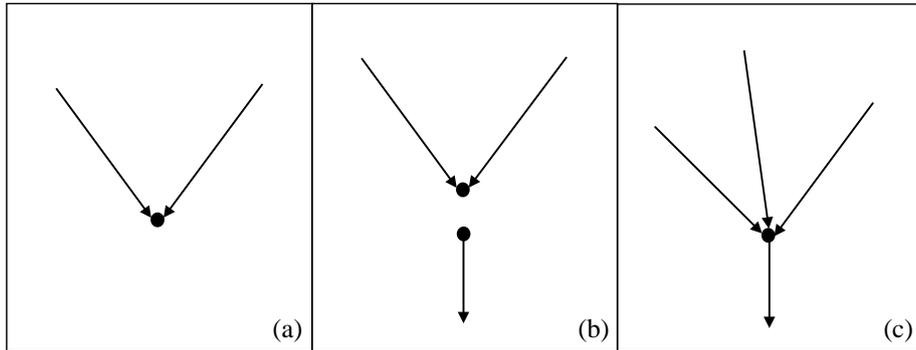
Figure 3: Landscape Networks (LSNs) used to create `.ssn` objects have two unique topological restrictions; converging nodes (a, b) and complex confluences (c) are not allowed. Converging nodes occur at the downstream node of two edges that converge (a, b), but do not flow into another downstream edge. Complex nodes occur anytime more than two edges converge at, or flow into, a node (c).

`Sites`, `Upstream Distance - Edges`, and `Upstream Distance - Sites`.

Watershed Attributes

The RCA is used to characterize landscape conditions found near each edge (Figure 1a), but in most cases it is only a subcomponent of the watershed; the exception being source edges where the watershed and the RCA are equivalent (Figure 1a). Instead, the watershed is commonly used as an analytical unit to summarize characteristics about the landscape that have the potential to affect in-stream conditions (Johnson and Host 2010). The **FLoWS** toolset provides the capability to sum edge attributes downstream (`Accumulate Values Downstream` tool). When an RCA attribute is summed, a watershed attribute is generated for the most downstream location on each edge. Yet, survey sites may fall anywhere along an edge. The **STARS** `Calculate Watershed Attributes` tool was developed to estimate site-specific watershed attributes:

$$W_i = (1 - r_i)RCA_i + \sum_{k \in U_i^*} RCA_k \qquad (2)$$

where W_i is the watershed attribute for each survey site i, RCA_i is an attribute summarized over the RCA where the ith site resides, and r_i is as defined in Equation 1. A new field is added to the sites attribute table that contains the W_i for each site. Note, W_i is simply an estimate of the true watershed attribute because RCA_i does not contain spatial information about attribute variability at the sub-RCA level. As an alternative, other more spatially-explicit methods could be used to calculate watershed covariates for sites, such as grid-based approaches (Peterson, Sheldon, Darnell, Bunn, and Harch 2011; Tarboton, Schreuders, Watson, and Baker 2009) or deterministic models (Rathburn and Wohl 2001), but automated tools for those approaches have not been included in the **STARS** toolset.

Segment PI

Calculating the spatial weights needed to fit a spatial statistical model to streams data is a

three step process: 1) calculating the segment proportional influence (PI), 2) calculating the additive function values, and 3) calculating the spatial weights (Peterson and Ver Hoef 2010). Steps 1 and 2 are performed using the **STARS** toolset, while step 3 is undertaken in R using the **SSN** package.

The PI for each edge, ω_j, is defined as the relative influence that the jth edge, or segment, has on the edge directly downstream. In the following example, ω_j is based on watershed area, but other simple measures, such as Shreve's stream order (Shreve 1966), could also be used (see Cressie, Frey, Harch, and Smith (2006) for an example). To begin, watershed area is calculated for the most downstream location of each edge in the network: $W^j = \sum_{k \in U_i} RCA_k$, using the **FLoWS** toolset. When two edges, denoted j and j', join at a node, the proportional influence (PI), ω_j, for the jth edge that flows into the node is then:

$$\omega_j = \frac{W^j}{W^j + W^{j'}}. \tag{3}$$

Note that, the ω_j values for edges directly upstream from a single confluence always sum to 1 because they are proportions. The `Calculate Segment PI` tool adds a new field to the `edges` attribute table (field type = double) that contains ω_j.

Calculate Additive Function - Edges and Sites

Two tools have been provided in the **STARS** toolset, which are used to calculate the additive function value (AFV) for every edge and site in the LSN: `Calculate Additive Function – Edges` and `Additive Function – Sites`. Separate tools are provided so that the AFV can be calculated for multiple sets of sites without having to recalculate the values for the edges.

The AFV for the jth edge, AFV_j, is equal to the product of the segment PIs found in the path downstream from the jth edge to the stream outlet:

$$AFV_j = \prod_{k \in D_j} \omega_k. \tag{4}$$

Given that ω_j is a proportion, the AFV_j will always range between 0 and 1, with the AFV_j for the most downstream edge in the network equal to 1. The AFV for the ith site, AFV_i, is simply equal to the AFV_j of the edge it lies on. As a result, multiple sites located on a single edge will always have the same AFV value. The `Calculate Additive Function – Edges` and `Additive Function – Sites` tools create a new field in both the edges and sites attribute table representing the AFV (field type = double). For additional details, please see Peterson and Ver Hoef (2010) and Peterson (2011).

Upstream Distance - Edges and Sites

There are two tools provided in the **STARS** toolset to calculate the upstream distance for each of the edges and sites: `Calculate Upstream Distance – Edges` and `Upstream Distance – Sites`. The `Calculate Upstream Distance – Edges` tool is used to calculate:

$$upDist_j = \sum_{k \in D_j} L_k \tag{5}$$

where $upDist_j$ is the upstream distance from the stream outlet to the upper end of the jth edge (u^j) and L_j is the total length of the jth edge. The `Upstream Distance – Sites` tool

is used to compute the upstream distance from the outlet to the ith site, $upDist_i$; i.e., the x part of x_i^j, and is given by:

$$upDist_i = r_i L_i + \sum_{k \in D_j^*} L_k, \qquad (6)$$

where L_i is the total length of the jth edge of x_i^j and recall that D_j^* is the set of all segments downstream of x_i^j, excluding the jth segment. The $upDist_j$ and $upDist_i$ values are recorded in the edges and sites attribute tables, respectively, within a new field named `upDist` (field type = double). The information stored in the two upstream distance attributes provides part of the spatial information needed to calculate pair-wise hydrologic distances between locations in the **SSN** package.

4.3. Export

Create SSN Object

The purpose of the `Create SSN Object` tool is to reformat the feature geometry, attribute data, and topological relationships of each spatial dataset contained in the LSN into a `.ssn` object so that it can be imported into R using functions provided in the **SSN** package. As we mentioned previously, a new directory is created to store this information, with the naming convention `lsn-name.ssn` (i.e., `lsn.ssn`); we refer to this as the `.ssn` object (Figure 4).
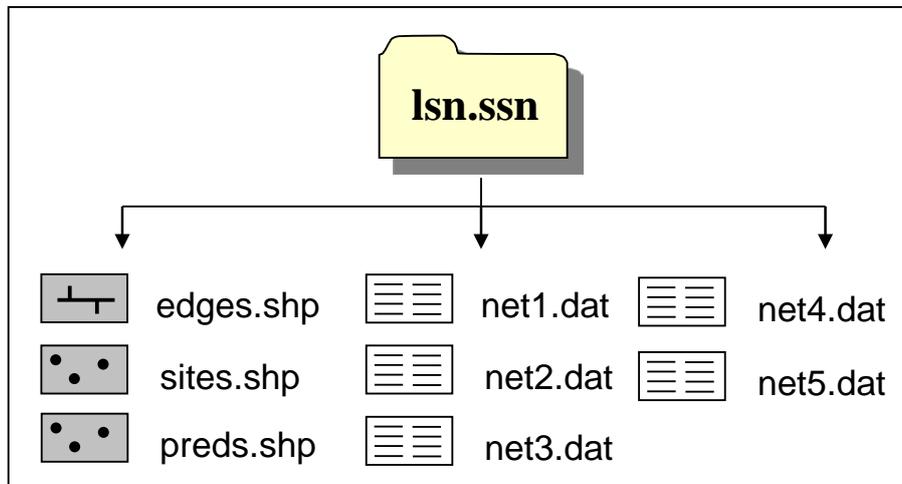

Figure 4: The **SSN** object contains the spatial, attribute, and topological information of the landscape network (LSN). It will always contain two shapefiles: edges and sites, as well as multiple text files containing the edge binary identifiers for each stream network in the LSN. Multiple shapefiles representing the prediction locations may also be included.

Shapefiles can be easily imported into R using the maptools package (Bivand, Pebesma, and Gomez-Rubio 2008), with all of the associated shape geometry and attributes. Consequently, the spatial datasets (edges and sites) are converted from ESRI (2009) feature class to shapefile format and stored in the `.ssn` object directory. However, shapefiles in standard form cannot be used to represent the topological relationships of the LSN. Our solution was to generate

network and binary identifiers (IDs), which provide an efficient way to assess connectivity and spatial relationships between features on the network.

The process of assigning binary IDs is relatively straightforward. First, the outlet edge (i.e., the most downstream edge in the network) is identified and assigned a binary ID equal to 1 (Figure 5). The information stored in the relationships table (Figure 1d) is used to identify edges that are directly upstream from the outlet edge. Binary IDs are assigned to the upstream edge(s) by arbitrarily appending a 0 or 1 to the downstream binary ID. For example, binary IDs 10 and 11 are directly upstream from binary ID 1 in Figure 5. This process of moving upstream and assigning binary IDs continues until every edge in the stream network has been assigned a binary ID.
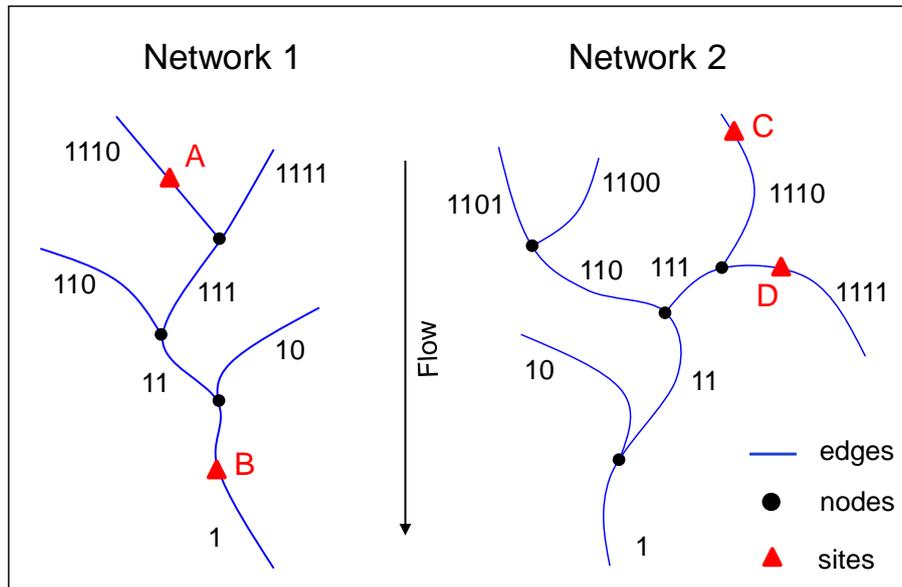

Figure 5: Binary identifiers (IDs) are assigned to each edge in the landscape network.

The binary IDs are useful because they contain information about whether locations have a flow-connected or flow-unconnected relationship. Two locations are considered flow-connected when water flows from an upstream location to a downstream location. In contrast, two locations are flow-unconnected when they share a common outlet downstream, but water does not flow between them. In Figure 5, the binary ID for the edge where site B resides is completely nested within the binary ID for the edge where site A lies ("1" is nested within "1110"), which indicates that the two locations are flow-connected. In contrast, the binary IDs for the edges where sites C and D reside are not nested ("1110" is not nested within "1111") because the two locations are flow-unconnected. The binary IDs for flow-unconnected locations also contain information about the closest common downstream location. As an example, the most common downstream confluence between sites C and D is u^{111} (Figure 5) and so "111" is where the two binary IDs diverge.

It is common for LSNs to contain multiple stream networks, with unique stream outlets. As an example, consider a streams dataset with two individual stream networks in the edges feature class (Figure 5). In this case, two edges from different networks will have the same

binary ID. Thus, a network identifier (`netID`, field type = long) is also assigned to the edges, sites, and prediction sites (if included) attribute tables to differentiate between locations. In addition, a location ID (`locID`, field type = long) and a unique point ID (`pid`, field type = long) are assigned to the sites and prediction sites attribute tables in order to distinguish between repeated measurements at a single location.

The `rid` and `binary` ID for each edge are stored in a comma delimited text file (Figure 6), with a separate file for each network. The naming convention for these files corresponds to the network ID (i.e., `net1.dat`, `net2.dat`, etc.). Once the binary and network IDs have been calculated, the edges, sites, and prediction sites feature classes are converted to shapefiles and exported to the `.ssn` object directory. When the `Create SSN Object` tool is complete, the `.ssn` object will have the file structure shown in Figure 4 and will contain the information necessary to fit a spatial statistical model using the **SSN** package.

```
"rid", "binaryID"
418,  1
409,  10
410,  11
407,  100
408,  101
176,  110
405,  111
347,  1010
```

Figure 6: Binary identifiers (IDs) and their associated reach identifier (`rid`) are stored in a comma delimited text file format within the `.ssn` object.

5. Accessing the Tools

The **FLoWS** and **STARS** custom ArcGIS toolsets, as well as the **SSN** package for R statistical software, are freely available through the "SSN and STARS" website hosted by the U.S. Forest Service, Boise Aquatic Science Lab http://www.fs.fed.us/rm/boise/AWAE/projects/SpatialStreamNetworks.shtml. A tutorial for **STARS** (Peterson 2011) is also available from the website; it is written for users who have relatively little experience with GIS and provides detailed instructions about how to generate the spatial information needed to fit spatial statistical models for stream networks (Ver Hoef and Peterson 2010). An example dataset is available for download and is referred to throughout the tutorial.

6. Future Developments

The **STARS** toolset currently contains all of the functionality needed to create the `.ssn` object, which can then be imported into the **SSN** package and used to fit a spatial statistical model to stream data. The LSN is currently stored as a personal geodatabase, which provides a self-contained data structure for multiple feature types (e.g., edges, nodes, survey sites). However,

the personal geodatabase is not a supported data format in ArcGIS versions released after 9.3. Therefore, the next major development will be to select a new data format for the LSN. The ArcGIS file geodatabase provides a straight-forward alternative; it is similar in structure to the personal geodatabase and an application programming interface (API) has been released for 32 and 64-bit Windows and Linux OS, which should provide better non-ArcObjects based access to the data. A second alternative is to recreate the **FLoWS** and **STARS** functionality in an open-source GIS software, such as GRASS (GRASS Development Team 2008), that can be directly linked with R using the **spgrass6** package (Bivand *et al.* 2008). Although the second alternative requires more labor to implement, it is attractive for a number of reasons. First, the ability to link an active GIS session directly with R reduces the chances of user error when files are transferred between software packages. Second, a direct linkage would allow some of the current functionality, such as calculating covariates for observed and prediction sites, to be more accessible in R. This would allow users who are unfamiliar with GIS software to calculate new covariates without having to move between software programs. Third, ArcGIS is proprietary software, while R is open-source with a vast community of contributors and users. Implementing a modeling framework based on open-source GIS and statistical software would not only be powerful, but also philosophically appropriate. In the meantime, we have tried to provide enough detail about the structure of the `.ssn` object so that users can recreate it using other, non-proprietary GIS software packages if necessary.

7. Acknowledgments

References

Ahuja R, Magnanti T, Orlin J (1993). *Network Flows: Theory, Algorithms, and Applications.* Prentice-Hall, New Jersey.

Bivand RS, Pebesma EJ, Gomez-Rubio V (2008). *Applied spatial data analysis with R.* Springer, NY. URL `http://www.asdar-book.org/`.

Cressie N, Frey J, Harch B, Smith M (2006). "Spatial Prediction on a River Network." *Journal of Agricultural, Biological, and Environmental Statistics*, **11**(2), 127–150.

ESRI (2009). "ArcGIS Desktop: Release 9.3.1." *Technical report*, Environmental Systems Research Institute, Redlands, CA.

GRASS Development Team (2008). "Geographic Resources Analysis Support System (GRASS) Software." URL http://grass.osgeo.org.

Hammond M, Robinson A (2000). *Python Programming on Win32: Help for Windows Programmers*. OâĂŹReily Media, Inc, Sebastopol.

Johnson L, Host G (2010). "Recent developments in landscape approaches for the study of aquatic ecosystems." *Journal of the North American Benthological Society*, **29**(1), 41–66.

Peterson EE (2011). "STARS: Spatial Tools for the Analysis of River Systems - A tutorial." Commonwealth Scientific Industrial Research Organisation (CSIRO), Technical Report EP111313. March 28, 2011. 42 p. URL http://www.fs.fed.us/rm/boise/AWAE/projects/SSN_STARS/software_data.html#doc.

Peterson EE, Sheldon F, Darnell R, Bunn S, Harch B (2011). "A comparison of spatially explicit landscape representation methods and their relationship to stream condition." *Freshwater Biology*, **56**(3), 590–610.

Peterson EE, Ver Hoef JM (2010). "A mixed-model moving-average approach to geostatistical modeling in stream networks." *Ecology*, **93**(3), 644–651.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rathburn SL, Wohl E (2001). "One-dimensional sediment transport modeling of pool recovery along a mountain channel after a reservoir sediment release." *Regulated Rivers: Research and Management*, **17**, 251–273.

Shreve RL (1966). "Statistical law of stream numbers." *Journal of Geology*, **74**, 17–37.

Tarboton DG, Schreuders KAT, Watson DW, Baker ME (2009). "Generalized terrain-based flow analysis of digital elevation models." In RS Anderssen, RD Braddock, LTH Newham (eds.), *18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation*, pp. 2000–2006. Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation. URL http://www.mssanz.org.au/modsim09/F4/tarboton_F4.pdf.

Theobald D, Norman J, Peterson EE, Ferraz S, Wade A, Sherburne MR (2006). *Functional Linkage of Waterbasins and Streams (FLoWs) v1 User's Guide: ArcGIS Tools to Analyze Freshwater Ecosystems*. Natural Resource Ecology Lab, Colorado State University, Fort Collins, Colorado.

Ver Hoef JM, Peterson E (2010). "A Moving Average Approach for Spatial Statistical Models of Stream Networks (with discussion)." *Journal of the American Statistical Association*, **105**, 6–18. doi:10.1198/jasa.2009.ap08248; Rejoinder pgs. 22 - 24.

Ver Hoef JM, Peterson E, Clifford D, Shah R (in review). "SSN: An R Package for Spatial Statistical Modeling on Stream Networks." *submitted to Journal of Statistical Software*.

Ver Hoef JM, Peterson EE, Theobald D (2006). "Spatial Statistical Models That Use Flow and Stream Distance." *Environmental and Ecological Statistics*, **13**(1), 449–464.

Affiliation:

Erin E. Peterson
Division of Mathematics, Informatics and Statistics
Commonwealth Scientific and Industrial Research Organisation (CSIRO)
PO Box 2583, Brisbane, QLD 4001
E-mail: Erin.Peterson@csiro.au
URL: http://www.csiro.au/org/CMIS.html

Jay M. Ver Hoef
NOAA National Marine Mammal Laboratory
NMFS Alaska Fisheries Science Center
International Arctic Research Center, Room 351
University of Alaska Fairbanks, Fairbanks, AK 99775-7345
E-mail: jay.verhoef@noaa.gov
URL: http://sites.google.com/site/jayverhoef