Transactions in GIS, 2005, 9(2): 247–263

Technical Note

GIS Tools for Freshwater Biodiversity Conservation Planning

Thomas W. FitzHugh
The Nature Conservancy
Olympia, Washington

Abstract

The Nature Conservancy (TNC) has developed a set of customized GIS tools to help ensure that freshwater biodiversity is adequately incorporated into its regional conservation planning efforts. The tools generate GIS data and attributes that support a regional-scale classification of freshwater systems and analysis of the potential integrity of these same systems. They have been implemented by TNC in more than 30 ecoregions across the U.S., and work is expanding to include other parts of the Americas. The tools are now in their second iteration. While the first version of the tools worked only with the RF3 hydrography dataset available in the U.S., the second version is more flexible and will function with a variety of datasets. Future improvements in the tools could include an improved interface and improved algorithms to reduce data pre-processing needs. The GIS tools, documentation, and data from completed ecoregions are now publicly available on the Internet.

1 Introduction

Regional conservation planning for freshwater biodiversity requires mapped information on conservation targets and threats at a relatively fine scale. Because specialized attributes must be developed while efficiently processing large amounts of data, the planning process is greatly facilitated by the use of customized GIS tools. The Nature Conservancy (TNC) has developed a set of GIS tools that generate data and attributes that can then be used to conduct a regional-scale ecological classification of freshwater systems and to analyze the potential ecological integrity of these same systems. These tools were developed to aid in TNC's ecoregional assessment process, the purpose of which is to identify a set of conservation areas that best represents the native species, communities, and ecosystems of an ecoregion and the underlying ecological processes

Address for correspondence: Thomas W. FitzHugh, The Nature Conservancy, 120 East Union Avenue, Olympia, WA 98501, USA. E-mail: tfitzhugh@tnc.org

9600 Garsington Road, Oxford OX4 2DQ, UK and
350 Main Street, Malden, MA 02148, USA.

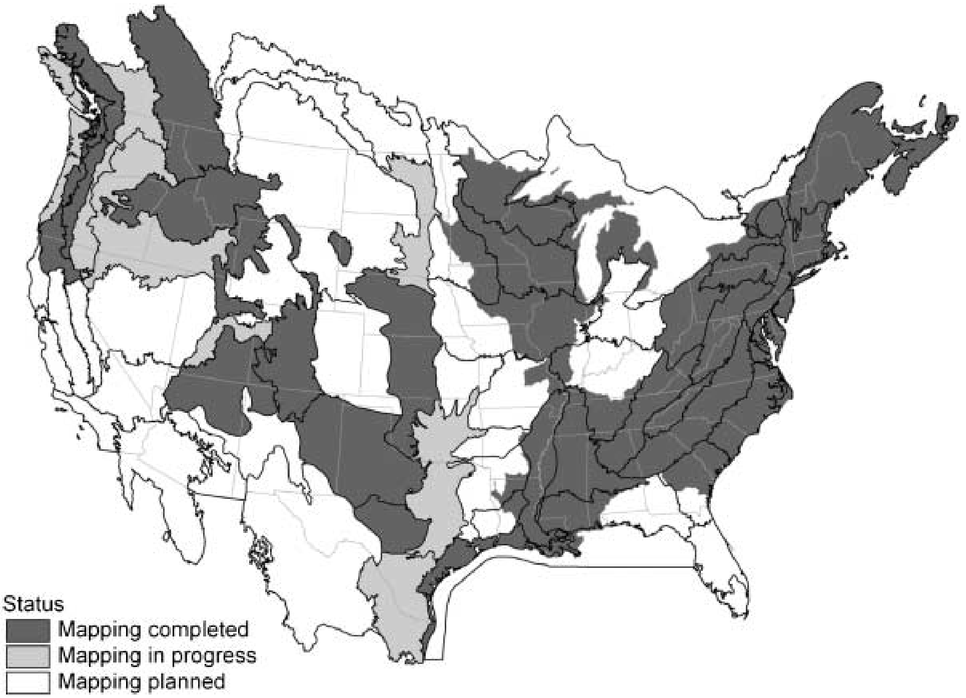


Figure 1 Status of freshwater ecological classification mapping in Nature Conservancy ecoregions in the U.S. lower 48 states

that sustain them (Groves et al. 2002). The tools have been implemented by TNC across the U.S. (Figure 1), and work is expanding to include other parts of the Americas.

The GIS tools described in this paper are used for two primary purposes in ecoregional assessment. First, they generate stream and lake attributes and watershed data which are then used to conduct a comprehensive classification of freshwater systems in each ecoregion, based on abiotic factors that are derived from readily available GIS data. Because developing and implementing a freshwater classification based on biotic data is beyond the resources of most regional conservation planning efforts, and often requires data that are unavailable, a classification based on abiotic factors is an essential tool for ecoregional assessment (Higgins 2003). Second, the tools are used to generate GIS indicators of the potential quality of these same streams and lakes, using information on human impacts. While such GIS-based quality indicators are no substitute for field-collected data on ecosystem health, they are nevertheless useful as an initial filter to evaluate the suitability of different areas for conservation (Higgins 2003).

This paper will describe the outputs of the tools, the algorithms used, input data requirements, and some of the limitations of the tools. In order to generate stream and lake attributes that depend on upstream characteristics, the vector GIS data of stream and lake hydrography used by these tools must satisfy certain requirements. This issue will be discussed in some detail, because generation of fine-scale hydrographic GIS data is still in process in many parts of the world, and it is instructive to reflect on which data structures are most appropriate for conducting this sort of analysis.

2 GIS-Generated Data Products

2.1 Freshwater Ecological Classification

The freshwater ecological classification used by TNC to support ecoregional assessments is described in Higgins (2003) and Higgins et al. (2005). The classification is based on abiotic factors such as stream and lake size, local connectivity to other hydrologic features, elevation, gradient, and upstream geology. The classification is hierarchical in nature, and the GIS tools are used to facilitate the classification of the two finest-scale levels in the hierarchy, macrohabitats and aquatic ecological systems (AESs). Macrohabitats are individual river valley segments and small- to medium-sized lakes or lake basins that are relatively homogeneous with respect to the abiotic factors that shape freshwater system structure and functions and influence the distribution of biota. AESs are larger stream and lake networks that represent areas with distinct geomorphological patterns tied together by similar environmental processes (e.g. hydrologic, nutrient, and temperature regimes), which are reflected in the patterns of macrohabitats they contain (Higgins et al. 2005).

The primary purpose of the classification is to identify coarse-filter conservation targets. Targets refer to those elements of biodiversity that are a focus for conservation planning or action. The main idea behind the coarse-filter approach is that by conserving representative examples of coarse-filter targets such as ecosystems or communities, the majority of species in the ecoregion will also be conserved (Groves et al. 2002). The AES is the coarse-filter target that TNC uses in the freshwater component of ecoregional assessment. Macrohabitats are rarely used as conservation targets themselves, because of the large numbers of macrohabitats in each ecoregion (in the range of 10,000–100,000) and the numerous possible unique macrohabitat types. AESs provide a much more tractable planning unit, which is coarse-scale enough to be useful as a conservation target to a planner working across an entire ecoregion, yet is also detailed enough to make meaningful distinctions between freshwater system types.

The GIS tools are used to automatically generate the stream and lake attributes that are used to classify macrohabitats. Figure 2 shows a final macrohabitat classification product for a small section of stream network. Lakes can also be classified into macrohabitats in a similar fashion, using variables such as size, shoreline complexity, upstream or underlying geology, and position in the flow network. While the GIS tools generate the raw classification attributes, creating the final macrohabitat product shown in Figure 2 requires that these attributes be translated into ecologically significant classes. For example, raw stream gradient values calculated for each stream arc are used to create a series of discrete classes (e.g. low, medium, or high gradient). The process for identifying these classes is beyond the scope of this paper, but it is based on review of relevant literature, consultation with experts, and other research (Higgins et al. 2005).

Methods for classifying AESs have evolved over time, but TNC's current methodology is to represent them using a series of nested watersheds. The GIS tools are used to create these watersheds, which are classified into different system types according to the variety of macrohabitats that they contain. AESs are typically classified at three to five scales (see Figure 3 for an example). While existing watersheds could be used for this purpose, this would restrict the scale of analysis to that of available watershed data. In order to distinguish between different types of AESs, watershed sizes need to be tailored to the characteristics of the freshwater systems being classified. The actual process of classifying AESs is fairly involved and also beyond the scope of this paper, but it is described in more detail in Higgins et al. (2005). AES classification over large

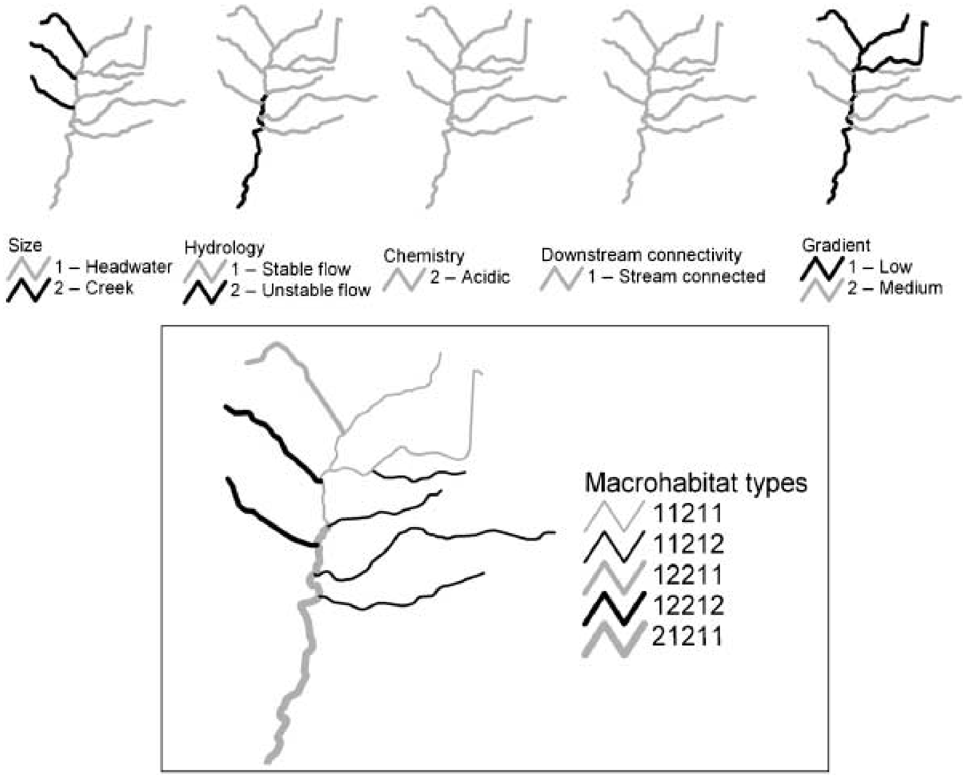


Figure 2 Stream macrohabitat classification. Five variables (stream size, hydrology, chemistry, downstream connectivity, and gradient) are used to classify stream reaches into five different macrohabitat types. The hydrologic regime and chemistry variables are not usually calculated directly, but are modeled based on upstream geology. Each unique combination of variables is a unique macrohabitat type

areas is greatly facilitated by use of a statistical clustering algorithm, in order to identify watersheds in each size class which have similar patterns of macrohabitats.

2.2 *Quality Analysis*

The GIS tools are also used to generate indicators of the type and amount of human presence in the watershed of each stream and lake macrohabitat and AES. Some typical quality indicators are upstream land-cover distribution (e.g. area of agriculture, urban development, and impervious surface) and density of roads, dams, and point sources of pollution. The scientific basis for using these indicators comes from studies that have found predictive relationships between such indicators and freshwater biotic integrity (Steedman 1988, Scheuler 1994, May et al. 1997, Wang et al. 1997, Moyle and Randall 1998, Arya 1999, Stormwater Manager's Resource Center 2001, Wang and Lyons, 2003). Such indicators have also been found to be predictive of in-stream sediment and nutrient loadings (Jones et al. 2001), which can in turn impact biotic integrity. While experts can be found who are familiar with many of the streams in an ecoregion, such knowledge is never comprehensive. These GIS-based quality indicators can be used to

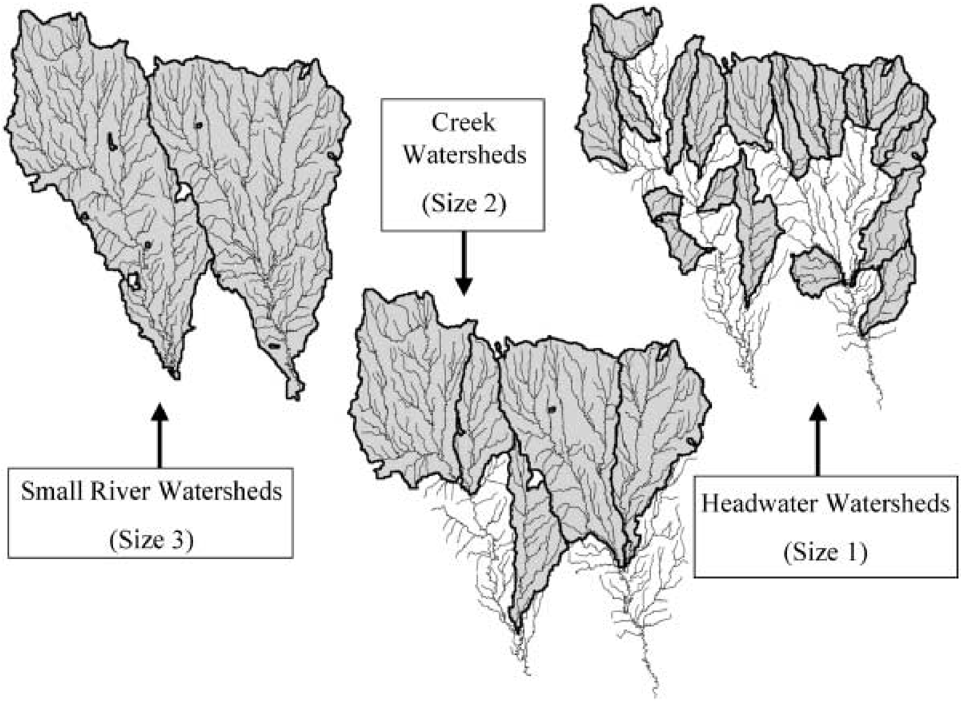


Figure 3 Nested watersheds used to represent aquatic ecological systems. Shows watersheds created at three scales (small river, creek, headwater) for the same drainage

fill in gaps in expert knowledge, so that at least some information exists about the potential quality of every stream.

An important caveat is that while indicator-integrity relationships have been empirically established by research at specific sites, the knowledge and data rarely exist to create a comprehensive model of biotic integrity using spatial data for a large basin. The only indicator that has been consistently validated in numerous locations is percent impervious area in the watershed (Stormwater Manager's Resource Center 2001). For other indicators to be viewed as anything more than preliminary assessments, they must be screened and validated through comparison with stream sampling data or expert opinion. The use of quality indicators in conservation planning is part of an ongoing process of learning and experimentation in TNC. In addition to preliminary assessments of biotic integrity, quality indicators are also useful for analyzing the types of threats to freshwater systems that may exist across entire ecoregions, and the socio-economic characteristics of watersheds that may impact the cost of doing conservation work and the types of conservation strategies employed.

3 GIS Tools, Processing Methods, and Input Data Requirements

3.1 *Overview of the Tools*

The GIS tools consist of a suite of algorithms written in three programming languages (Arc Macro Language (AML), Avenue, and Visual Basic) that operate in three different software packages (Arc/Info (ESRI Inc. 1998a), ArcView (ESRI Inc. 1998b), and Microsoft Access (Microsoft Inc. 2002)). AML and Avenue are used for tools which require spatial operations, while Visual Basic is used for tools that involve only database operations. For database operations, Access provides much faster processing than Arc/Info or ArcView. The different tools are applied in sequence to produce the desired output products. The tools were compiled and documented by the author, but many of the algorithms are based on previous work by other individuals and organizations (see the Acknowledgements for a list of contributors to the tools). At a minimum, the tools are designed to be used with a vector hydrographic layer and a raster Digital Elevation Model (DEM). Full implementation of the tools requires other GIS data layers such as geology, land-cover, roads, dams, and point sources of pollution.

There have been two versions of the tools. The first version of the tools worked with the Reach File 3-Alpha (RF3) hydrography (EPA Office of Wetlands, Oceans, and Watershed, Office of Water 1994), a 1:100,000-scale dataset of streams, lakes and wetlands in the lower 48 U.S. states. The second version of the tools, (completed in January 2003) works with the National Hydrography Dataset (EPA and USGS 2000) another 1:100,000-scale dataset that is the successor to the RF3. It will also function with any other hydrographic datasets that satisfy certain input data requirements. The tools were upgraded to take advantage of the improved accuracy of the NHD and also to make them usable with a variety of datasets. While many of the tools in the first version were dependent on stream connectivity attributes that were unique to the RF3, the second set of tools uses arc-node topology to keep track of connectivity in the flow network. This makes the tools less dependent on the attributes of a particular dataset, and also makes it possible to combine together different datasets in situations where watersheds cross political boundaries (e.g. the U.S.-Canadian border).

Table 1 lists the outputs generated by these tools, with brief summaries of the processing methods and input data requirements. Tools are subdivided into outputs

Table 1 Summary of attributes generated by the GIS tools, with processing methods and input data requirements.

Local Analysis

Outputs	*Processing Methods*	*Input Data Requirements*
Stream gradient and elevation	Derive elevation of from and to nodes from DEM. Elevation is the average elevation of the two nodes. Gradient is the ratio of the change in elevation along the arc's length to the length of the stream arc.	Stream arcs must be pointed downstream.
Stream upstream and downstream connectivity classes (whether connected to a stream, lake, or coastline; or unconnected)	Use arc-node topology to identify the types of features that are upstream and downstream of each stream arc.	Stream arcs must be pointed downstream, and must be connected to lake arcs at nodes.

Table 1 *Continued*

Lake elevation and underlying geology	Overlay lake polygons on DEM and geology coverages. Generate elevation of the polygon label point and the most common geology class within the perimeter of the lake.	Lakes must exist as polygons.
Lake number of surface connections	Use arc/node and polygon/arc topology to count up the number of stream arcs connected to each lake polygon.	Lakes must exist as polygons, and lake arcs must be connected to stream arcs at nodes.
Upstream Analysis		
Outputs	*Processing Methods*	*Input Data Requirements*
Stream and lake orders and arbolate sums	Use flow network attributes to perform calculations.	Flow network must be prepared, with appropriate attributes (seqno, level, srchflag).[1]
Stream and lake watershed area (total and by class) and upstream quality indicators	Create reach catchments, and overlay on relevant themes (geology, land-cover, roads, dams, point sources). Use flow network attributes to perform calculations.	Flow network must be prepared, with appropriate attributes (seqno, level, srchflag).[1]
Nested watersheds used to represent AESs	Create reach catchments. Use flow network and size attributes to perform calculations.	Flow network must be prepared, with appropriate attributes (seqno, level, srchflag).[1]
Watershed (AES) quality indicators	Use flow network attributes to derive downstream reach of each watershed, and transfer quality indicators from flow network.	Flow network must be prepared, with appropriate attributes (seqno, level, srchflag).[1]

---

[1] Proper preparation of the flow network depends on different attributes of the data for each version of the tools. For the first version of the tools, RF3 attributes defining connectivity in the flow network must be accurate. For the second version of the tools, (a) stream arcs must be connected at nodes, (b) centerlines must exist through any lake or large river polygons, (c) stream and centerline arcs must be pointed downstream, and (d) main paths must be delineated through flow divergences. See the text for further discussion of these requirements.

which are generated by analyzing GIS data in or directly next to each stream arc or lake polygon (local analysis) and outputs that have to do with the upstream characteristics of the stream or lake (upstream analysis). The remainder of this section will further describe the tools, although upstream analysis will be discussed in far more detail because the processing involved is more sophisticated.

The input data requirements presume that hydrographic data is properly attributed to identify the different types of hydrographic features (lakes, wetlands, rivers, coastlines). Closure of polygons and arc connectivity issues are mentioned in Table 1 because in the author's experience, these cannot always be taken for granted in hydrographic GIS data. Many of the tools require that stream arcs be pointed downstream, so that the from- and to- nodes can be properly identified. Most of the hydrographic datasets used with these tools already have arcs that are properly pointed downstream, but the second version of the tools also has a tool for checking and possibly correcting arc directionality. This tool will be discussed later in more detail.

3.2 Local analysis

Attributes based on local analysis are generated through standard GIS query and overlay operations, as described in Table 1. Other lake attributes that are important for ecological classification are size and shoreline complexity, but these are not described because they can be easily generated directly from the GIS data without any special tools.

Stream gradient and elevation calculations require accurate DEM elevations and accurate spatial co-registration between DEM and stream data. These two conditions may be self-evident, but they are mentioned here because when these tools have been used with readily available GIS data in the U.S., these sorts of errors sometimes cause short streams in flat areas to receive negative gradients, even when pointed in the right direction. Fortunately, such situations are not very frequent, and these streams can still be easily classified into the lowest gradient class, so this error is unlikely to have a significant impact on the final classification product.

3.3 Upstream analysis

Attributes that are based on upstream analysis are defined by the characteristics of the watershed draining into the stream (or lake) in question. Many of these attributes are estimates of stream size or the amount of flow entering a lake, such as Strahler order, Shreve order, arbolate sum (the total length of all streams upstream), and watershed area. Upstream analysis is also used to generate watershed characteristics such as the distribution of geology or land-cover classes and the density of roads, dams, and point sources. The last output product from upstream analysis is the set of nested watersheds that are used to classify AESs. Since all upstream analysis depends on a standard flow network data structure and upstream accumulation algorithm, these will be described next, after which issues specific to each particular tool will be discussed.

3.3.1 Flow network data structure

The flow network is defined here as the set of arcs that are used to represent hydrographic features that flow into each other. These arcs typically represent streams and also wide rivers and lakes. The method of representing wide rivers and lakes varies depending on the dataset. They are generally represented either by arcs along their shorelines or by centerline arcs that have been delineated through wide river and lake features. The former method is used in the RF3 and the latter in the NHD.

Prior to performing upstream analysis, the data must be pre-processed to add some standard attributes to all arcs in the flow network. These attributes are required by the upstream accumulation algorithm. This flow network data structure and the algorithms used to generate and analyze it were developed by the Research Triangle Institute (Bondelid 2002 pers. comm.). The attributes are:

1. *Stream level*. The stream level identifies the main paths through the flow network. A level of 1 is assigned to the flow path that extends from the terminus of the drainage network all the way upstream through the main path of flow to a headwater reach where flow begins. A level of 2 is assigned to all reaches that terminate at this level 1 path (that is, all tributaries to the path) and to all reaches that trace the main path of the flow along each tributary back to its headwater reach. A level of 3 is assigned to all tributary paths of the level 2 path, using the same criteria. Levels of increasing value are assigned using these criteria to all reaches in the flow network. The stream level is used as the index in an array that keeps track of values being accumulated down the flow network.
2. *Sequence number*. This controls the order in which reaches are processed, and is assigned using a depth-first search algorithm starting at the bottom of the flow network. When doing upstream analysis, reaches are processed in descending sequence number order, which provides the GIS tools with an efficient way to traverse the flow network.
3. *Start reach flag*. This is a flag which distinguishes start reaches (where flow begins) from other reaches.

See Figure 4 for further description of these added attributes.

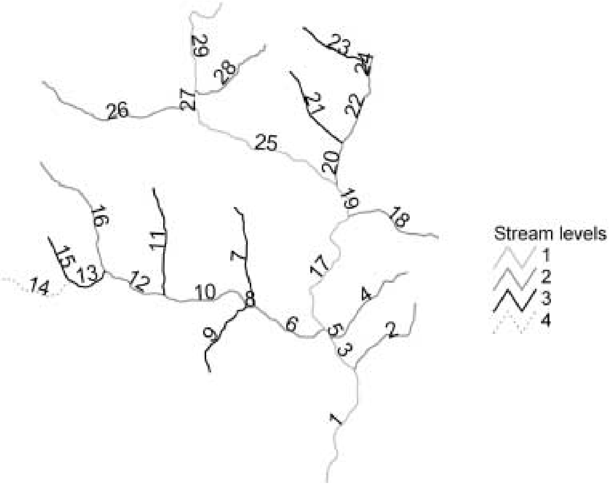


Figure 4 Flow network attributes. Different stream levels are shown in different shades of gray, and each reach's sequence number is shown next to the reach. Start reaches are those reaches that have no reaches upstream, such as sequence numbers 14, 15, and 16, but not 12 and 13

These additional attributes allow the upstream accumulation algorithm to traverse the flow network in a manner that is much more efficient than if the original dataset attributes were used. Records can be sorted and processed in descending sequence number order, so that it is never necessary to do time-consuming searches through the entire table. Using this method provides great advantages in terms of processing time for very large datasets (with more than 100,000 arcs in the flow network).

3.3.2 Building the flow network

The GIS tools use the following procedure to assign flow network attributes. The pre-processing tool starts at a terminal arc (at the bottom of the flow network), and then proceeds upstream (using a depth-first search algorithm) and assigns the three flow network attributes to each upstream arc. When all arcs upstream of the first terminal arc have been visited and processed, the algorithm goes to the next terminal arc, and repeats the process. The two versions of the GIS tools used slightly different methods to implement this pre-processing procedure, because of the different datasets used. The first version used the flow network attributes of the RF3 data to traverse the flow network. RF3 has attributes that identify, for each arc in the flow network, the arcs that are upstream and the next arc that is downstream. The second version of the tools uses arc-node topology and arc directionality (arcs must be pointed downstream) to identify these same flow relationships.

3.3.2.1 Flow divergences

Flow divergences (locations where an arc in the flow network connects to two other arcs in the downstream direction) can cause problems for this standard data structure and the algorithms that utilize it. Flow divergences in hydrographic data can vary from simple cases where an arc splits into two arcs which then immediately recombine, to complex braided streams and rivers, to large loops in the flow network (Figure 5). The

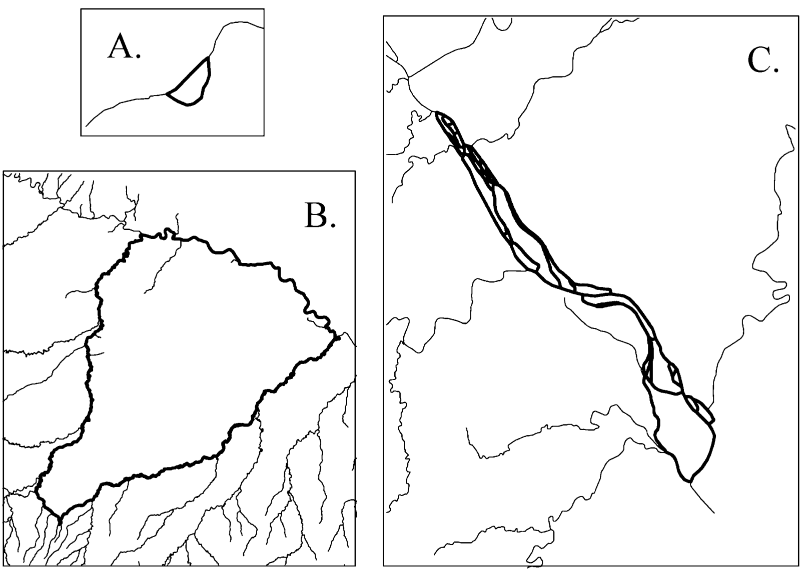


Figure 5 Flow divergences: (A) shows a simple divergence; (B) shows a large loop in the flow network; and (C) shows a braided river

problem introduced by these divergences is that the stream level concept only makes sense if tributaries do not reconnect to each other further upstream, and if this rule is violated, the upstream accumulation algorithm does not work properly. Different methods for dealing with flow divergences are used by the two versions of the tools.

RF3 data has an attribute that identifies arcs that enter flow divergences, and this attribute was used by the first version of the pre-processing tool to assign levels and sequence numbers so that the upstream accumulation algorithm would work properly. The method works well for simple divergences, but results for more complex divergences are sometimes unpredictable. Another feature of RF3 that is designed to mitigate problems with divergences is the exclusion of complex braided streams from the flow network. Instead, these arcs are encompassed by a “braided stream envelope”, consisting of a line of arcs on either side of the complex braid that can be used to simplify the connectivity along braided sections of the river.

The NHD (along with most other hydrographic data) does not have specialized attributes to identify flow divergences. So the second version of tools identifies flow divergences using arc-node topology and arc directionality. Divergences are incorporated into the flow network by identifying a main path through each divergence, and then deleting all secondary paths from the dataset. Using a main path to traverse flow divergences is the same approach used in the British Columbia Watershed Atlas (British Columbia Ministry of Environment, Lands, and Parks, Fisheries Branch 1996), a 1:50,000-scale hydrographic dataset of British Columbia. Figure 6 shows some examples of flow divergences with main and secondary paths identified. Deleting arcs is not ideal, but at the regional scale removing these few arcs is unlikely to have much impact on the resulting analysis. Indeed, for complex braided streams, flow divergences are sometimes fairly ephemeral, so representing them in the hydrographic dataset may not really even be appropriate.

Identification of main paths can be a time-consuming process, so a special tool is used to automatically identify main paths. For each divergence, it identifies the shortest

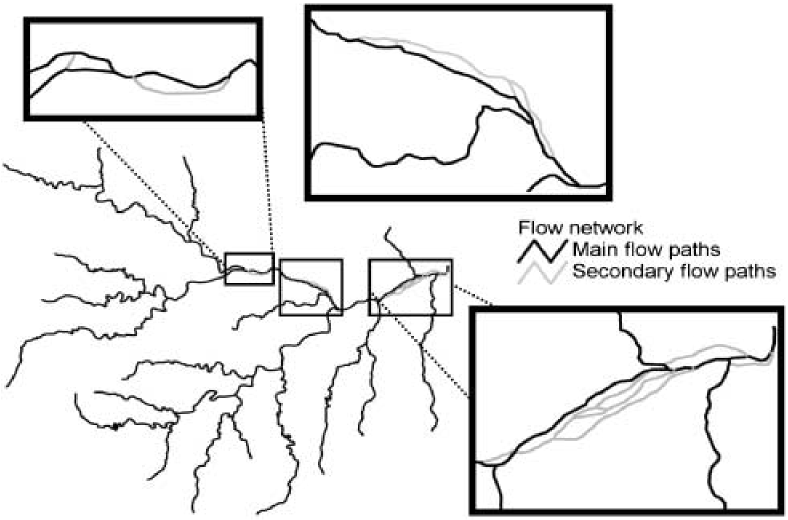


Figure 6 Section of a stream network with main paths delineated through flow divergences

path that connects all of the points that enter or exit the divergence, and this is used as a main path. The tool works well for simple divergences and many braided streams, but the most complex divergences and large loops in the flow network must be manually checked and often fixed. Arc directionality is also checked at this stage of the processing, because deletion of secondary paths usually requires that directionality be corrected for a few arcs. These algorithms work similarly to the pre-processing tool described earlier, starting at a terminal arc, and then proceeding up the flow network to check directionality and identify flow divergences.

3.3.2.2 Other pre-processing issues

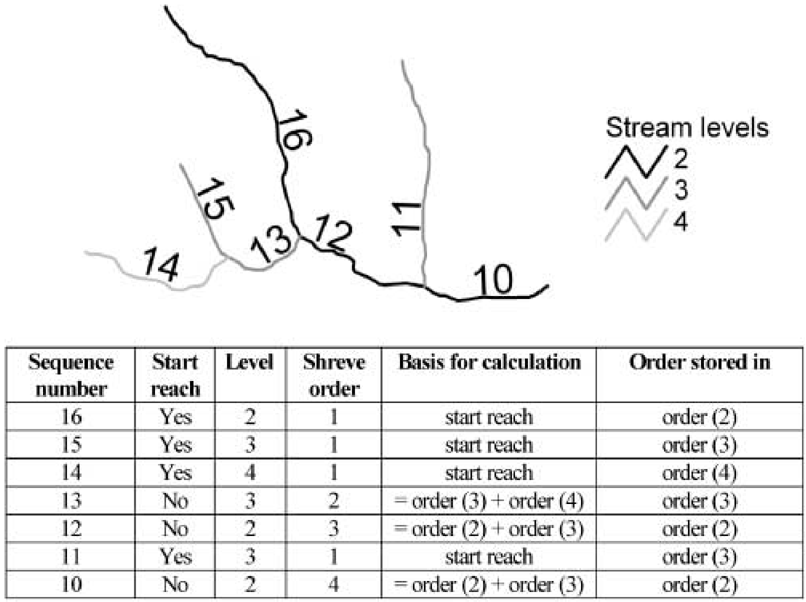
Another situation which can cause problems for this flow network data structure and the algorithms that utilize it is the existence of junctions in the flow network where more than two arcs flow into the next arc downstream. Four-way connections such as this occur in the NHD, and five-way connections (four streams flowing into one) have also been found, although they are very rare. The flow network data structure described above only works for three-way connections. In order to deal with this problem, additional records can be added to the data table of arcs to turn each 4-way connection into a pair of 3-way connections. This can be done by adding a "dummy" record which has no length and altering the from- and to-node numbers of the connecting records. A similar approach can be used for five-way connections, although this has not been extensively tested, since these situations are so rare.

Building the flow network is usually the most time-consuming part of applying the GIS tools. Some amount of manual editing to the hydrographic data is always necessary, for either version of the tools. The RF3 fields that identify the arcs that are upstream of each arc in the flow network contain a few errors, and these must be corrected for the tools to work properly. The process of identifying main paths through divergences in the NHD or other datasets inevitably requires some manual editing to deal with the most complex divergences.

3.3.3 Upstream accumulation algorithm

Figure 7 shows how the flow network attributes are used to facilitate the accumulation of values down the flow network. This example shows how Shreve stream orders are calculated for a small section of the flow network. The Shreve stream order of a stream is the number of first-order streams (start reaches) that are upstream. The algorithm proceeds in descending sequence number order through the table of stream reaches, i.e. from the top to the bottom of the table in Figure 7. The first reach (sequence number 16) receives an order of 1, because it is a start reach. This order is also stored in an array indexed to the level of this reach (2). The second and third reaches (sequence numbers 15 and 14) also receive orders of 1 (because they are start reaches), and these orders are stored in arrays indexed to the levels of these reaches (3 and 4, respectively). The next reach (sequence number 13) is not a start reach, so it needs to receive an order of 2 (the number of start reaches upstream). This is calculated by summing the orders stored in the arrays indexed to the level of the current reach (3) and the next higher level (4). Similar calculations are performed for all of the reaches in Figure 7 (see the table included with the figure).

The flow network attributes ensure that this calculation will be accurate for every reach in the flow network. Processing proceeds through the entire flow network in this manner, and similar algorithms can be used to calculate Strahler order, arbolate sum,



Sequence number	Start reach	Level	Shreve order	Basis for calculation	Order stored in
16	Yes	2	1	start reach	order (2)
15	Yes	3	1	start reach	order (3)
14	Yes	4	1	start reach	order (4)
13	No	3	2	= order (3) + order (4)	order (3)
12	No	2	3	= order (2) + order (3)	order (2)
11	Yes	3	1	start reach	order (3)
10	No	2	4	= order (2) + order (3)	order (2)

Figure 7 Example of using flow network attributes to calculate Shreve stream orders. See text for a description of this example

watershed areas, and quality indicators. These attributes can be generated for lakes as well as streams, assuming that the shoreline or centerline arc at the downstream end of the lake can be automatically identified.

3.3.4 Attribute generation

3.3.4.1 Stream and lake orders and arbolate sums

Shreve orders are calculated using the algorithm just described. Strahler orders are calculated similarly, except that the order number is only incremented when the two reaches upstream have the identical order. Arbolate sums (total upstream stream length) are calculated using the length of each reach, and summing it up as the algorithm proceeds down the flow network.

3.3.4.2 Stream and lake watershed area (total and by class) and upstream quality indicators

Calculations having to do with watershed area require an additional item of information, which is the area draining into each reach. This is generated by running an AML which generates a polygon for each arc in the flow network that delineates the area that flows directly into that arc. This is called a reach catchment (see Figure 8 for additional details). This is accomplished using a DEM and the Arc/Info GRID commands for defining flow direction and flow paths from a raster DEM. Reach catchments can be generated for each arc or lake polygon in the flow network.

This tool works quite well, except that in flat areas the reach catchments have a much higher level of error than in areas with more relief. While improved algorithms

Figure 8 Reach catchments delineated for each stream reach

have been developed which would likely minimize these problems in flat areas (e.g. Mackay and Band 1998), such methods have not been used because TNC is currently constrained to using algorithms in readily available software packages such as Arc/Info. Fortunately, the impacts of these errors in flat areas should be greatly lessened in TNC's regional-scale planning work because for most reaches, attributes based on upstream analysis are based on the contents of multiple reach catchments, i.e. the local reach catchment plus the reach catchments of all reaches upstream. The larger the upstream watershed, the more likely it is that any attribute errors that occur because of errors in the location of individual reach catchment boundaries will be averaged out in the final attribute calculations.

Once these reach catchments are generated, each arc in the flow network can be assigned an area, and these areas can be accumulated in the same manner as stream orders and arbolate sums, in order to generate the area of the watershed of each stream or lake. The reach catchments can also be overlaid on other GIS layers, such as geology and land-cover, to generate the area of each geology and land-cover class in each reach catchment. These areas can be accumulated in the same way as the total watershed area. Lastly, similar overlays can be performed with data such as roads, dams and point sources, and sums of the length or number of these features calculated. Densities of these features relative to watershed area can then be easily calculated.

3.3.4.3 Nested watersheds to use for creating AESs

Nested watersheds are created by proceeding through the table of reaches in descending sequence number order, keeping track of the accumulated watershed area for each arc. Watersheds are generated in discrete size classes, for example, 4–10, 10–100, 100–1000 km^2. For each size class, the tool identifies the sets of reach catchments that can be merged together in order to form watersheds that fall within the desired size classes, by assigning all of the reach catchments in each prospective watershed a unique identification number. The reach catchment polygons can then be dissolved together using these identification numbers to form the nested watersheds. A second tool identifies the downstream reach of each watershed, so that the reach level quality indicators generated by the other tools can be assigned to these watersheds. In this manner quality indicators can be generated for each AES.

4 Conclusions

The GIS tools have been instrumental in allowing TNC to incorporate freshwater biodiversity into its ecoregional assessment process, because they can efficiently produce the necessary output products using widely available GIS datasets. The GIS tools have been used in over 30 of TNC's ecoregions in the U.S., and have also been used to process hydrographic data in Canada and Mexico. Many of the classification products that have been produced are the first ever developed for their regions, and represent a major step forward in defining patterns of freshwater biodiversity in these areas (Higgins et al. 2005). A recent example of a project for which the tools were used is Weitzell et al. (2003), who identified freshwater conservation priorities in the Upper Mississippi River Basin. The tools were used to generate attributes for approximately 150,000 stream and lake reaches, in an area equivalent to 6% of the entire area of the lower 48 United States.

Though the tools have greatly enhanced TNC's ability to analyze patterns of freshwater biodiversity and threats, there are still improvements that could be made. Because they function in three different software packages, the tools are still more time-consuming to use than is ideal, are only usable by experienced GIS users, and require a significant amount of staff time for training and support. Recoding all of the tools in the new ArcGIS software package (ESRI Inc. 2001) would provide a more integrated and user-friendly set of tools. No such conversion is planned as of now, but making this conversion would be facilitated by the fact that much of the code is written in Visual Basic, the language now used in ArcGIS.

Another area where significant improvements could be made would be to reduce the amount of pre-processing and manual fixes that have to made in hydrographic data prior to using the tools for upstream analysis. This is generally the most time-consuming part of applying the tools. One way to improve this situation would be to improve the algorithm for identifying main paths through flow divergences, particularly with regard to how it solves the more complex divergences. Ideally, data management agencies would also create hydrographic datasets that take into account the input data requirements of tools such as the ones discussed in this paper. Assignment of main paths and secondary paths to each flow divergence (as was done in the BC Watershed Atlas mentioned earlier) is one method of preparing data so that it works more smoothly with the tools.

An important issue in applying the tools is the potential impact on final data products of errors resulting from inaccuracies in input data or less than optimal processing algorithms. Examples of the most obvious errors observed have already been mentioned, but their impact on final data products is typically not severe, because these errors are either relatively infrequent or because their impact is greatest at very fine scales that are not that important for the regional-scale analyses being conducted. But it is important to understand that these errors may have significant impacts on the characteristics of a few individual reaches or reach catchments, and hence some outputs of the tools may not be appropriate for finer-scale analyses.

To conclude, in addition to supporting ecoregional assessment, the data produced by these tools and the conservation planning products that the tools contribute to are also useful for other purposes. The fine-scale macrohabitat data and reach-level quality indicators can be used as an initial characterization of the freshwater systems within conservation areas and the potential threats to those systems. These data are also useful for conducting further research. Some research questions that these data are relevant to are the relationship between the distribution of aquatic organisms and habitat characteristics, and the use of watershed characteristics as predictors of biotic integrity. Finally, conservation priorities identified by TNC and accompanying data are being published so that other conservation and environmental agencies and organizations can make use of the information.

Tool and Data Availability

These GIS tools can be downloaded at http://www.freshwaters.org/info/large/documents.shtm#gis. Detailed documentation and a tutorial dataset are also available. Data generated by the tools is available for large parts of the USA at http://gis.tnc.org/community/projects/FWI/index.php

Acknowledgements

Comments from two anonymous reviewers were very helpful in improving this paper. Production and design of the GIS tools would not have been possible without the support and advice of the staff of Nature Conservancy's Freshwater Initiative program, including Jonathan Higgins, Mary Lammert, and Mark Bryer. Some of these tools were wholly created by the author, but many of them were based on other tools already in existence. The following individuals and organizations developed or contributed algorithms and/or programming code that are incorporated into the GIS tools: Tim Bondelid and Chris Spoerri at Research Triangle Institute, under contract with The Nature Conservancy; Chuck Horne at Jordan, Jones, and Goulding; Mark Caldwell, Missouri Department of Conservation; Environmental Systems Research Institute, under contract with The Nature Conservancy; Jim Barber, U.S. Forest Service; and Bart Butterfield, Idaho Fish and Game.

References

Arya S 1999 Explaining biotic integrity and habitat across multiple scales: An empirical analysis of landscape, land use, and land cover variables in an Ohio ecoregion. WWW document, http://www.ucgis.org/oregon/papers/arya.htm

British Columbia Ministry of Environment, Lands, and Parks, Fisheries Branch 1996 *An Introduction to the British Columbia Watershed Atlas, Revision 2.1*. North Vancouver, BC, Spatial Vision Consulting Ltd

EPA Office of Wetlands, Oceans, and Watershed, Office of Water 1994 *EPA Reach File Version 3.0 Alpha Release (RF3-Alpha) Technical Reference*. Washington, DC, U.S. Environmental Protection Agency

EPA and USGS 2000 The National Hydrography Dataset, Concepts and Contents. WWW document, http://nhd.usgs.gov/chapter1/index.html

ESRI Inc. 1998a *Getting started with ARC/INFO*. Redlands, CA, Environmental Systems Research Institute

ESRI Inc. 1998b *Using Arcview GIS*. Redlands, CA, Environmental Systems Research Institute

ESRI Inc. 2001 *Getting started with ArcGIS*. Redlands, CA, Environmental Systems Research Institute

Groves C G, Jensen D B, Valutis L L, Redford K R, Shaffer M L, Scott J M, Baumgartner J V, Higgins J V, Beck M W, and Anderson M G 2002 Planning for biodiversity conservation: Putting conservation science into practice. *Bioscience* 52: 499–512

Higgins J V 2003 Maintaining the ebbs and flows of the landscape: Conservation planning for freshwater ecosystems. In Groves C G (ed) *Drafting a Conservation Blueprint: A Practitioner's Guide to Planning for Biodiversity*. Washington, DC, Island Press: 291–318

Higgins J V, Bryer M T, Khoury M L, and FitzHugh T W 2005 A freshwater classification approach for biodiversity conservation planning. *Conservation Biology* 19: in press

Jones K B, Neale A C, Nash M S, Van Remortel R D, Wickham J D, Riitters K H, and O'Neill R V 2001 Predicting nutrient and sediment loadings to streams from landscape metrics: A multiple watershed study from the United States Mid-Atlantic Region. *Landscape Ecology* 16: 301–12

Mackay D S and Band L E 1998 Extraction and representation of nested catchment areas from digital elevation models in lake-dominated topography. *Water Resources Research* 34: 897–901

May C R, Horner C R, Karr J, Mar B, and Welch E 1997 Effects of urbanization on small streams in the Puget Sound Lowland Ecoregion. *Watershed Protection Techniques* 2: 483–94

Microsoft Inc. 2002 Microsoft Access. WWW Document, http://www.microsoft.com/office/access/default.asp

Moyle P B and Randall P J 1998 Evaluating the biotic integrity of watersheds in the Sierra Nevada, California. *Conservation Biology* 12: 1318–26

Schueler T 1994 The importance of imperviousness. *Watershed Protection Techniques* 1: 100–11

Steedman R J 1988 Modification and assessment of an index of biotic integrity to quantify stream quality in Southern Ontario. *Canadian Journal of Fisheries and Aquatic Sciences* 45: 492–501

Stormwater Manager's Resource Center 2001 Impervious Cover Model. WWW Document, http://www.stormwater.net

Wang L and Lyons J, 2003. Fish and benthic macroinvertebrate assemblages as indicators of stream degradation in urbanizing watersheds. In Simon T (ed) *Biological Response Signatures: Indicator Patterns Using Aquatic Communities*. Boca Raton, FL, CRC Press: 227–49

Wang L, Lyons J, and Gatti R 1997 Influences of watershed land use on habitat quality and biotic integrity in Wisconsin streams. *Fisheries* 22(6): 6–12

Weitzell R E, Khoury M L, Gagnon P, Schreurs B, Grossman D, and Higgins J 2003 *Conservation Priorities for Freshwater Biodiversity in the Upper Mississippi River Basin*. Arlington, VA, NatureServe and The Nature Conservancy