1:24K Hydrography Attribution Data
Science Services, Fisheries & Aquatic Research
March 13, 2013
Matt Diebel, Diane Menuz, and Aaron Ruesch

Background

This project attributed channel, riparian, and watershed level data for streams in the 24K hydrogeodatabase (24KGDB) with a variety of geologic, land cover, and other base data. The immediate goal of this project was to provide fine-resolution stream attribute data to be used to model stream flows and fish distributions. However, there are several general outputs of the project that can be useful for many projects. First, most features in the 24kGDB are attributed with data at six scales- the stream channel, traces from the stream channel to other features (such as lakes), in a 60-m riparian buffer around features, traces along all upstream buffers, in the watershed surrounding features and traces along all upstream watersheds. Second, the stream, riparian and watershed layers are available so that additional data can be attributed to them. For example, updated land cover data could be attributed to the watershed rasters and this information could be added to the overall attribution. Third, data on the topology between features has been generated. This makes it possible for new traces to be done and to identify the general to-from connectivity of all features. Last, tools have been developed that, with some minimal customization, will allow for changes in the 24KGDB to be incorporated into updates. These tools could also potentially be used to undertake a similar project across another large region. Further documentation of the tools can be found in the Tutorial for the 1:12K Hydrography Creation Toolset.

The purpose of this document is to provide users with mid-level detail of the process that produced these files and more information about feature attributes when further explanation beyond the metadata associated with them is needed. Users who intend to use the attributes already associated with features may want to only read the sections of this document pertaining to attributes they are interested in. Users who want to create new attributes for the database may want to read the Tutorial and gain some familiarity with this document. In particular, these users should become familiar with the "STATUS" attribute in the BaseAttributes_24K table to understand which new attributes are valid. Users who want to use the tools developed by this project to undertake a similar effort and/or to update data from this project should read both this document and the tutorial and potentially also look at the scripts themselves to determine if/how they may need to be modified to meet specific project needs.

Feature Selection

All stream and lake data was obtained from a copy of the 24KGDB from 08/24/2012. The following features were included in the dataset: 1) Features with hydrocode 7061, 7062, or 7071 and >= 5 acres in area ("lakes") 2) landlocked primary flowlines connected to landlocked lakes 3) primary, non-landlocked flowlines and 4) flowlines through lakes. All selected flowlines from now on will be called "streams" though some do flow through lakes or other waterbodies. Waterbody features with Hydrotype 710 (Unspecified open water) were attributed as Hydrocode 7061, 7062, or 7071 in the 24kGDB and thus are included as lakes for this project though they may not actually be lakes.

The features mentioned above were considered part of reaches, the basic unit that was attributed. For lakes, the lake polygon and all stream features flowing under them were assigned a single REACHID equivalent to the HYDROID of the lake polygon. All other streams were considered as separate features, so that each HYDROID is associated with a unique REACHID. To more accurately represent all of the dimensions of streams, we spatially associated other waterbodies with the streams flowing under them. For example, a stream along the Wisconsin River ends up being much wider than a single pixel because of the wider area of the underlying Wisconsin River polygon. This is discussed in more detail in the section on reach rasterization.

In a few cases, we needed to add or modify lines in the 24KGDB and thus had to assign our own HYDROIDs to these features. All HYDROIDs that we added start with 29 to distinguish them from HYDROIDs already present in the data. New features were sequentially assigned values starting with 290000001. Features from the 24KGDB that had to be split into two features were assigned new values that incorporated the old feature HYDROID. If the original feature HYDROID was 200042977, the two new features will be assigned the HYDROIDs 291042977 and 292042977.

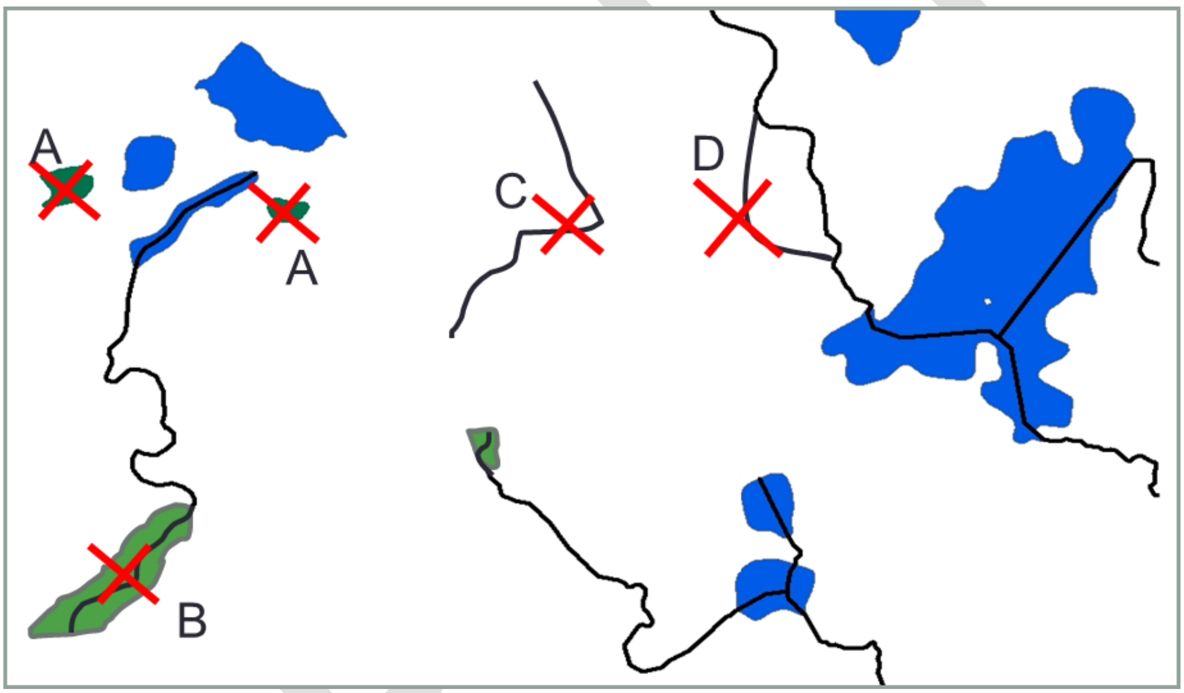


Figure 1. **Overview of features included and excluded from study.** Blue indicate lakes, green indicate non-lake waterbodies, and black indicate flowlines. Flowlines continuing out of the image are considered not landlocked while the remaining lines are landlocked. All features in image were included in the analysis except the following: *A* because they are non-lake waterbodies that are not attached to stream lines, *B* because it is a non-lake waterbody attached to an isolated lake, though the flowline under the non-lake waterbody IS included. These features were excluded on accident, but there are only approximately 300 features like this that were excluded. *C* because it is an isolated streamline without any attached lakes and *D* because it is a secondary feature (braided channel).

Feature Location

The goal of this project is to attributes features within the state of Wisconsin. Some features in Wisconsin have watersheds that include areas outside of the state. Examples of this include the Nemadji drainage in Minnesota and the entire Mississippi River. We included necessary out-of-state features whenever possible in order to have complete watershed coverage. We also included all out-of-state features that are within HUC12s that cross the Wisconsin border. This latter inclusion allowed for more accurate watershed delineations. Because we delineated sheds within HUC12 boundaries, it was important to have all possible features included in delineations so that no feature was assigned an overly large watershed. The major exceptions to the out-of-state inclusions were for four major boundary rivers: the St. Louis, St. Croix (along the state border, not the inland portion), Mississippi, and Menominee. We did not include out of state features that flow into these rivers because 1) we are not as interested in attributing these large rivers and 2) there would be a substantial number of additional out-of-state features that would need to be obtained in order to attribute these rivers. The database only has base attribution and channel-level attribution for these features.

Data Cleaning

FLoWS

In order to select features of interest and ensure correct to-from topology for upstream traces, we performed considerable data cleaning steps. We ran the tool "Check Network Topology" in FLoWS v9.3. This program assigns node statuses to each node in a stream network, including source, outlet, psuedonode, confluence, diverging and converging nodes. All nodes listed as diverging (places were flowlines split into two different directions) and converging (places were two flowlines come together but do not flow into a third line) were considered errors. Some of the fixes associated with these issues included incorrect landlocked status and/or primary vs. secondary designation in the 24kGDB and lines that needed to be flipped. In addition, outlets that were within the state were checked for possible errors if they were associated with non-landlocked features. Isolated features were allowed to have converging nodes; however, a file of single line "dangles" was created with lines that extended from the converging node in order to have correct topology in the landscape network. The dangles were assigned an identical REACHID to one of the converging line features; generally, they were assigned to a lake feature if possible; otherwise , the underlying DEM served as a guide for determining which REACHID to associate the dangle with.

Check landlocked status

In additional to using FLoWS, we checked whether the attribute "landlocked" was correct in two additional ways. First, we visually inspected all landlocked features that were 3rd order or higher to determine whether these features should actually be connected. We used underlying DEM, imagery, and database of wetland polygons to help make these determinations. These determinations were passed on to William Ceelen and Ann Schachte at WDNR and only those corrections that they verified were changed for our analysis. Second, we made an ArcGIS tool that flags any landlocked features that are within a user-specified distance of non-landlocked features. We ran this tool using a tolerance of 0, meaning that features were only flagged if they actually directly touched a non-landlocked feature. These features were then visually inspected to determine whether they, or their adjoining features, needed to be updated.

Check waterbody snapping

We checked whether flowlines were correctly snapped to the edges of the lakes they were associated with. This is necessary so that flowlines under lakes are not incorrectly attributed with their own REACHID distinct REACHID separate from the lakes REACHID and so that flowlines not under lakes are not incorrectly attributed as part of lakes. Changes were only made in the data if the snap tolerance was large enough to affect output.

Identify lakes that need lines

In order to have to-from relationship between features, it was important that lake features that are adjacent to streamlines have lines that go through the lake and join the network. We developed a tool to check for lakes that should, but do not, have internal flowlines.

Reach Rasterization

Lake and stream features were rasterized and snapped to the 10-m DEM along with non-lake waterbodies. Non-lake waterbody pixels were assigned the REACHID of the nearest lake or stream feature using the Cost Allocation tool in ArcGIS. This created wider features where rivers become larger or pass through small (<5 acre) lakes.

Lake Michigan, Lake Superior, and Lake Winnebago each have HUC12 boundaries that encompass the entirety of the lakes. We considered these features to be seeds that were able to receive overland flow directly across HUC12 boundaries rather than only through stream lines that flowed into them. However, the 24KGDB boundary for these lakes is not exactly aligned with the USGS HUC12 boundaries. For this reason, we did the following: For Lake Michigan and Lake Superior, we used the HUC12 boundary to create a rasterized line feature, and then expanded the feature by one. The line was given the same HYDROID as Lake Superior or Lake Michigan, as appropriate. We mosaicked the HUC12 boundary raster with the remaining lake and stream reach rasters, prioritizing the reaches so that stream and lake pixels could cross the Great Lakes boundaries. For Lake Winnebago, we combined the HYDRO24 lake feature with the HUC12 lake feature to create a larger seed for the lake. In both cases, some (non-lake) 24KGDB features end up being partially or entirely underneath the HUCs for the three lakes. These features were generally removed from the analysis; we are treating the HUC12 boundaries as the true borders. Data for features that are partially under HUC12 boundaries should be treated with caution- data will only reflect attributes for the area outside of the HUC12 boundary and thus may not be comprehensive for all parts of the feature.

Riparian buffers

Riparian buffers were created around the completed rasterized reaches at a distance of 60 m. We first removed the Great Lakes and Lake Winnebago boundary seeds and then used a cost allocation to create the riparian buffers (the expand function would not work with more than 1000 unique seeds)

Watershed Delineation

We delineated watersheds for each water feature in the state that filtered through all selection procedures above. This was a several step process.

The first step was "conditioning" the DEM. Conditioning a DEM means slightly altering a raw DEM (including the removal of small errors) to ensure continuous watersheds that have the fewest possible irregularities. Our conditioning steps followed protocols of the AGREE method embedded in ArcHydro tools—filling sinks (small internally draining depressions), "burning" in vector flowlines (lowering elevations under streams), building "walls" (increasing elevations at pre-defined watershed boundaries), and "breaching walls" by lowering elevations where streams crossed watershed boundaries. We added two new improvements to the AGREE method. First, we built "columns" at stream confluences (increased elevation at single pixel location) to prevent watershed boundaries from crossing confluences (see Figure). Second, in addition to burning streams into the DEM, we also burnt in isolated lakes, which amplifies the depth of the depression. To ensure that the DEM does not re-fill the depression during the "Fill" routine, we only partially filled the DEM to half the depth of the burn. This ensures all minor depressions were filled without filling isolated-lake depressions.

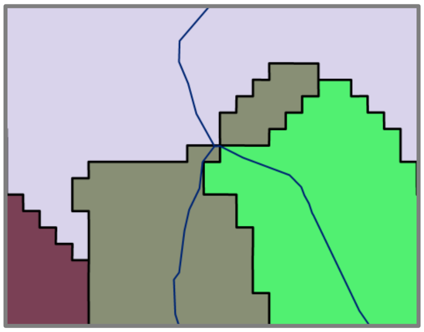


Figure 2. Example of a watershed delineation that crosses a confluence.

The second step was the delineating of watersheds themselves using ArcGIS hydrology tools that are packaged in the Spatial Analyst extension. We used the "Flow direction," "Flow accumulation," and "Watersheds" tools on the conditioned DEM to output raster watershed area. The raster watersheds were vectorized (without simplification) and additional irregularities were removed. These irregularities included very small watershed delineations (less than 1000 m^2), "zippers" (see Figure 3), watersheds disconnected through zippers, and watersheds of tributaries of the Great Lakes that did not cross the Great Lakes boundary.

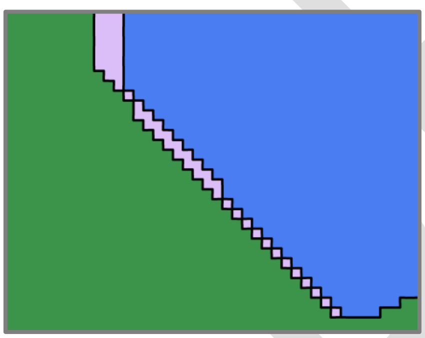


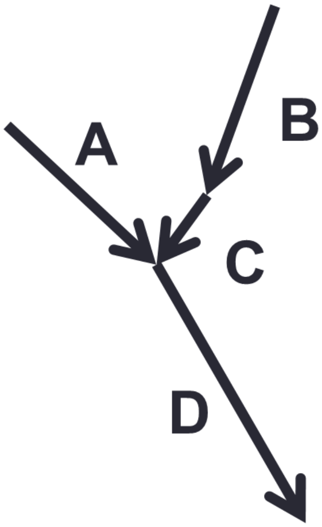
Figure 3. Example of a "zipper" and a watershed connected by a zipper.

Fix watershed expands

Some watersheds had small gaps in them after the watershed delineation tool ran. These gaps were identified by 1) merging all sheds delineated within a HUC12 together and then 2) determining whether the merged sheds were identical to the original HUC12 (total of about 35 had gaps, including Horicon Marsh). Features with gaps went through an additional Cost Allocation phase so that all land in the study area is associated with a watershed. A cost allocation option was not written into the original tool because it seemed prone to crashing the script.

Watershed Topology Attribution

Topology, or the connectivity of features in space, must be known to summarize connectivity attributes on a stream network. To create watershed topology, we conflated stream network (1-dimensional) topology to the tessalation of watersheds (2-dimensional). Stream network topology was created using the FLoWS ArcToolbox described in the above "Data Cleaning" section. FLoWS topology is a simple data structure consisting of polylines associated with a relationship table. Each segment in the polyline network has an ID that is listed only once in an ID column in the relationship table. The relationship table has an additional column that lists the ID of the downstream segment (Figure 4).



ID	TO_ID
A	D
B	C
C	D
D	--

Figure 2. Data structure of stream network topology.

To conflate topology to watersheds, each watershed was given the same ID as the stream it flows into. Then, the relationship table for the stream network topology can simply be re-used. However, two complicating factors exist: there is not a one-to-one relationship of stream segments to watersheds and watersheds are attributed to stream segments that are isolated from the network. To correct the problem of having more stream segments than watersheds, we simply skipped those segments without a watershed, and continued searching downstream until a segment with a watershed was found and denoted that segment's ID as the downstream segment. To correct the problem of isolated features was more challenging.

To ensure complete topology, including watersheds that are isolated from the network, we inferred connectivity based on terrain analysis. Most isolated features are isolated because of some topographic anomaly, such as past glaciation causing pocketed terrain, or highly permeable soils that allow groundwater to continuously feed a water feature (otherwise known as a "seep" lake). For each isolated feature, we clipped out pixels of a flow-accumulation grid that was created using a filled DEM and selected the pixel with the highest flow accumulation. The ID of the adjacent watershed that intersected this pixel was considered the downstream ID. In simpler terms, this location describes the place where water would theoretically "spill" if inundated with water (Figure 5). When topology was created using this terrain-based topology, occasionally topological networks resulted in circular connectivity.

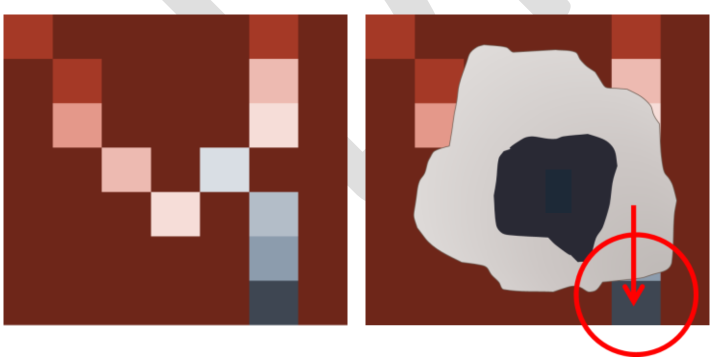


Figure 3. Schematic of terrain-based topology assignment. Colors on the grid from red-to-blue represent a grid of flow accumulation. The grid cell with the highest flow accumulation adjacent to the watershed (light blue polygon) of a lake (dark blue polygon) informs the topological relationship in flow. The adjacent watershed intersecting the dark blue grid cell would be assigned as the downstream ID.

A little over 200 features exhibited circular topology. This means that Shed A flowed into Shed B based on flowline direction, but Shed B flowed into Shed A based on flow accumulation. We decided to resolve these issues case by case because of the vast number of potential causes for the issue. Here are the general rules we followed:

1) Features that were supposed to flow into Lake Michigan or Lake Superior based on flowlines were assigned a “to” shed that was the HYDROID of the Great Lake they flowed into.
2) Features that headed out of state based on flowlines to catchments that were not delineated were assigned a “to” shed of 88888.

The remaining circular topologies generally resulted from two scenarios- incorrect flowline directionality or direction of flow within isolated basin not the same as direction that flow would go if water were to leave the basin. When the underlying DEM indicated that flowlines were heading uphill, these lines were usually flipped and watershed topology changed to reflect the flipped lines (see Figure 6 for examples). Generally a flow accumulation was run over the area in question in order to determine exactly where flow should go though in some cases only the DEM was used because the result was pretty clear. In some cases, flowlines were not flipped though the watershed topology was reversed as if the lines had been flipped. This happened when 1) local knowledge indicated that flowlines actually did move against the DEM (see watershed 200024270) or 2) evidence that the flowlines were incorrect was slight (e.g. difference in DEM less than 1 m).

We applied one additional rule in changing topologies. Watersheds for isolated features were not allowed to cross HUC12 boundaries. When the flow accumulation indicated that they should cross the HUC12 boundary, we instead routed the topology to the lowest resistance path within the HUC12. Non-contributing HUCs were an exception to this rule. Since these HUCs are not generally breached at all, we decided that in extremely high water, breaching may occur in more than one location. See HUC12 070300050604 as an example of this. Features without circular topologies that crossed HUC12 boundaries were not fixed.

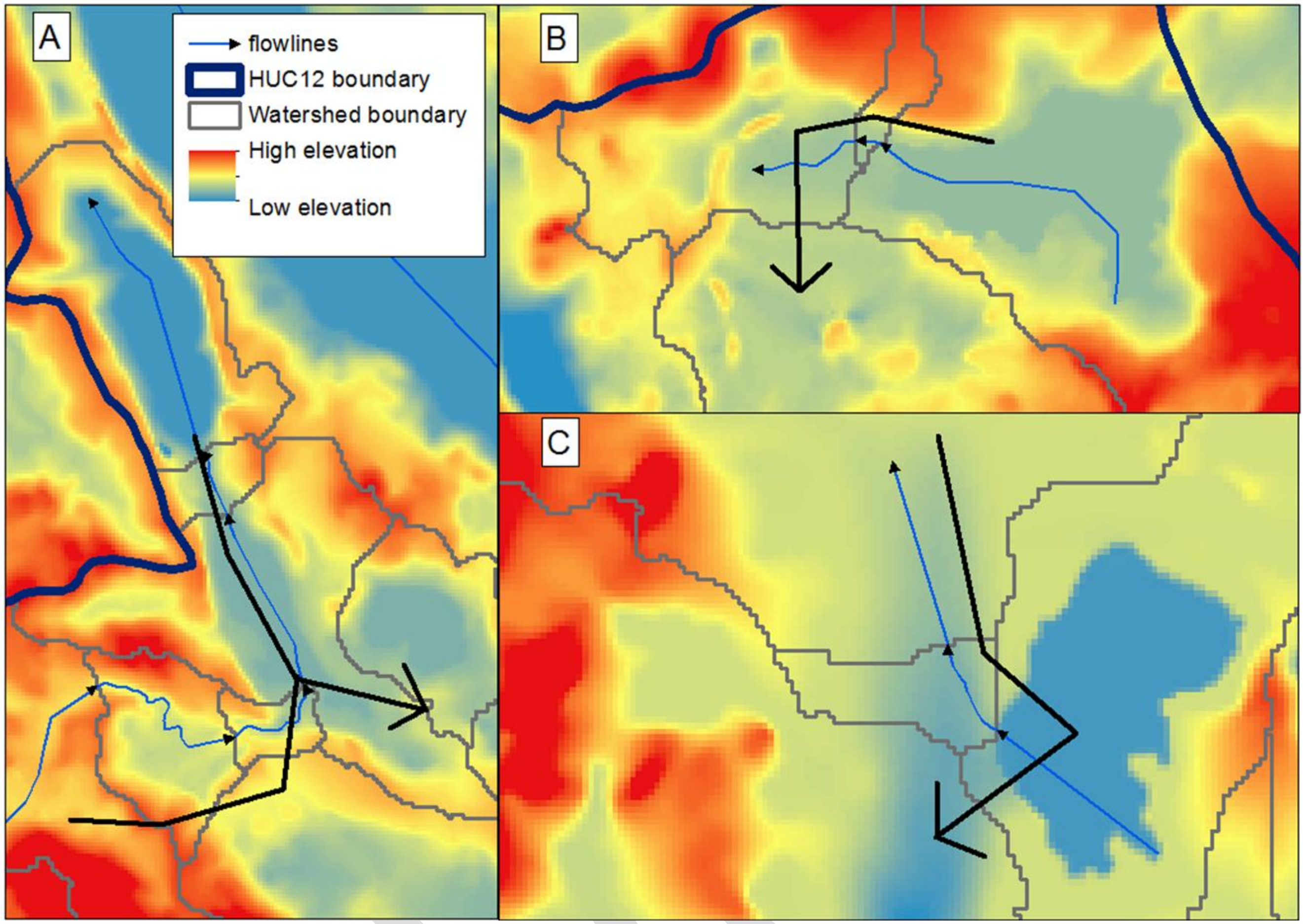


Figure 4. Areas with watersheds that had to be rerouted to correct circular topology. Black arrows indicate the rerouted watershed flow. In Figure A and B, permanently flipping flowlines was not recommended while in C it was to better represent the underlying DEM. In Figure A, the flowlines accurately represent the probable direction of flow within the isolated basin while the watershed topology accurately represents where overflow would be rerouted to. In B, the difference in the start to end of the flowlines is less than 1 m, so the database directionality is kept. In C, the DEM suggests that both the watershed and flowline topology should be flipped.

Feature Attribution

We used data sources obtained from the EPA/Gap Star project and/or obtained new rasters for more complete coverage of our study area and/or to obtain more up-to-date versions. All shapefiles were converted to rasters and all rasters were reprojected to NAD_1983_HARN_Transverse_Mercator and converted to a 10-meter grid snapped to the DEM across the study area. Data that we obtained from the EPA/Gap project was already reclassified using GAP/EPA Star Categories and generally matched at state borders if data came from multiple sources. Units and sources of data can be found in the metadata and metadata_Sources tables in the same database as the attribute values. Below is a limited discussion of additional changes made to input source data. More information about how reclassification was done can be found in metadataAttributeReclassification.xslx.

Soil permeability

EPA/Gap Star project converted 'Perm' Field to Integer (Multiplied by 100) and Deleted Polygons with -10 values (So Lakes became 'No Data')

Adjusted soil permeability

To adjust soil permeability values by impermeable surfaces in urban areas, we first created rasters of the percent impervious area associated with each urban class in 1992, 2001, and 2006 land cover data, with all other land cover pixels assigned an impermeability value of 0. We then calculated adjusted soil permeability with the equation (soil permeability)*((100- % impervious)/100). Values used to determine % impervious for each urban class can be found in metadataAttributeReclassification.xslx.

Darcy

We used Darcy calculations from GAP/EPA Star work. These calculations used a 30-m DEM and K values based on estimates for surficial geology categories. Details can be found in http://www.michigandnr.com/PUBLICATIONS/PDFS/ifr/ifrlibra/research/reports/2064rr.pdf and values of K used for each surficial geology category can be found in metadataAttributeReclassification.xslx.

Presettlement Land Cover

We obtained presettlement land cover data from Minnesota to complement data already obtained by the GAP/EPA Star project. Data was reclassified using GAP/EPA Star categories and mosaicked to the existing presettlement land cover raster.

2006 Land Cover and Modeled Land Cover, all years

This data has not been reclassified from the original values to EPA/Gap project values. Modeled land cover was obtained in December 2012 from Bryan Pijanowski and Jarrod Doucette at Purdue University. Future projections were based on NLCD 2001V2.

High Capacity Wells

Tabular data on high capacity wells in the state of Wisconsin were obtained from Robert Smail, Water Supply Specialist in the Bureau of Drinking Water and Ground Water at the Wisconsin Department of Natural Resources on December 4, 2012 and spatial data on well locations were obtained from the Wisconsin Department of Natural Resources SDE on or around November 20, 2012. Spatial data were spatially associated with underlying catchments and joined to the tabular data using the HICAP_WELL_NO field. We used three attributes from the tabular data to calculate the amount of pumping in each catchment in the year of interest, *y*. The fields Active Year (year well first in use) and

ABD Year (year well abandoned) were used to determine which wells were active in year *y.* Wells were considered active if Active Year ≤ *y* and ABD Year ≥ *y*. The field Est Total Annual WDRL (kGal) was then summed for all active wells within a given year for each catchment. Est Total Annual WDRL was calculated by Robert Smail using the following methods: "Reported withdrawals in 2011 were used to determine the percentage of daily capacity used each month per well. These usage coefficients were then averaged by water use category and an estimated monthly withdrawal was recalculated for each well. For those wells with missing water use codes, coefficients were calculated for property types and a monthly estimate was given for each. Estimated withdrawals for all sources were totaled and compared against reported withdrawals for 2011. The difference between the reported (213 bGal) and estimated withdrawals (212 bGal) was only .0038%."

Curve Number

We calculated curve numbers statewide by combining land-cover information (National Land Cover Dataset and WISCLAND) with soil "hydrologic group" information (SSURGO soils database). We created a lookup table (based on USDA Technical Report 55) that associated combinations of hydrologic soil groups and land cover classes with a specific curve number describing runoff potential. The lookup table can be found in metadataAttributeReclassification.xslx.

Artificial Drainage

We created a layer that estimates if a land area is artificially-drained (e.g., tile-drainage). This map is created from land-cover information (National Land Cover Dataset and WISCLAND) and soil-drainage-class information (SSURGO soils database). The estimate assumes that if a soil-drainage-class (ability of soil to drain water without human intervention) is considered poorly drained and it coincides with a cultivated-crop land-cover class, that area has likely been artificially drained to create a field suitable for agriculture. The Mapunits of the SSURGO database that were selected as "poorly drained" were those under the [drclassdcd] column attributed as "Poorly drained" or "Very poorly drained." Where no data existed, the same values of the STATSGO database column [drainagec] were used as a replacement. The areas considered to be cultivated crops were WISCLAND values 111, 112, 113, 118, and 110 (Herbaceous/Field Crops, Row Crops, Corn, Other Row Crops, and Agriculture) and NLCD value 82 (Cultivated Crops).

Sinks

We created watershed summaries of "sinks," which are internally draining topographic depressions that can be identified using a DEM. Sinks are defined according to a threshold which represents the theoretical depth of rain (or fill) necessary for the sink to overflow. We defined sinks according to two depth thresholds, 1-meter and 5-meter.

Gradient

We calculated stream gradient by dividing the elevation difference of a stream reach by its length. Initial elevations were derived from an un-conditioned 10 m DEM (National Elevation Dataset). To account for potential misalignment of stream reaches and topography, we assigned the lowest elevation pixel in the reach watershed as the minimum reach elevation. We then assigned maximum elevation as the minimum elevation of the upstream reach. If there was no upstream reach (i.e., a headwater), the maximum elevation was assigned the elevation of the upstream node of the reach. In some cases, this approach resulted in negative stream gradients. We corrected negative stream gradients using an interpolation approach. The interpolation algorithm is relatively complex. More details can be found in Appendix A.

Dams

All dam records were individually reviewed to eliminate duplicate or inappropriate entries and to identify dams that had been removed or lost despite being listed as active, intact structures. Records in which dams were not assigned to a size class but had sufficient information on height and/or impoundment size were assigned to the appropriate size class (small or large) based on the National Inventory of Dams criteria: large dams are those with a structural height of over 6 feet (1.83 m) and impounding 50 acre-feet (61,681 m3) or more, or having a structural height of 25 feet (7.62 m) or more and impounding more than 15 acre-feet (18,504 m3) [USACE, 2005, http://crunch.tec.army.mil/nid/webpages/nid.cfm]. Dams with a status code of "unbuilt" or "levee", or with similar status information in the comment field were removed from our analysis; dams classified as "planned" (i.e., construction permits approved by WDNR) were assumed to have been built and were included (following advice from the Wisconsin Office of Dam Safety; M. Galloway, WDNR Office of Dam Safety, Madison, WI, pers. comm.). Several dams were described as being located on non-navigable waterways or were classified as "nowat"- not situated on a waterway (e.g., stormwater retention ponds). Because our focus was on the effects of dams on drainage systems, we sought to eliminate structures that were not situated on stream or river channels. The presence and location of structures that were >2 m from a stream line on the WDNR 1:24k hydrography layer were assessed with aerial photographs. Structures that could not be located were removed from the analysis. All remaining dams were then snapped to the nearest stream line for subsequent spatial analyses. Records lacking latitude/longitude data were also excluded. It is likely that many of the removed records represent dams that were or are situated in river reaches, but taking this conservative approach was preferable to making assumptions about reach location for incorrectly sited or un-sited structures. A small number of records included multiple dams in sequence; these records were split so that each record represented an individual dam, and the location of each structure was determined from aerial photographs. We also added dams that had apparently not been authorized (i.e., were not in the database) but whose existence had been confirmed through field visits. In all cases, these structures were dams that were recently removed. Collectively, these actions resulted in the removal of 646 of the original 5158 records, and addition of 14 new records (n = 4526 records analyzed). 3698 of these dams are classified as active.

Of these 3698 dams, 140 of them were located on stream segments not present in the final selected flowlines used for the attribution project. Stream segments have between 0 and 7 dams located on them. The attribute DAMSIDE for features with more than one dam associated with them was determined as follows: features were split at dam locations, the segment length for the segments associated with the start and end nodes was found, and the side with the smaller segment length was determined to be the side that the dam was located on. In other words, the side of the stream segment (upstream or downstream) with the least amount of open water before hitting a dam was considered the DAMSIDE.

Cross tabulate data values

Pixel counts were converted to percentages in each category. These percentages do not take into account areas with missing data, so all values should add up to 100% within a category. For each category, there is also a column with the percent of data missing for that feature in that category.

Connectivity traces

The stream distance between each feature and the nearest feature of interest (e.g. nearest lake or nearest large watershed) was traced along the stream pathway. Features that connect with a dam before reaching the feature of interest are given a null value for the connectivity trace unless they are able to reach a different feature of interest in a different direction (i.e. there is a lake upstream though feature is blocked from closer lake downstream by a dam). All dammed streams (see ***Dams***) were attributed as associated with either the up- or down- stream end of streams based on which stream end had the longest distance after being split by the dam. Streams with multiple dams were still attributed as being dammed on only one side, even if there were dams near both ends. If a stream has a dam on the downstream node and a dam 10 m from the upstream node, the segment would be attributed as having upstream connectivity, because it has *more* connectivity on the upstream side, though overall connectivity is low. Only Wisconsin dam data was included in the analysis, so some data may be incorrect for some features if they trace to out-of-state lakes and/or if they are out of state features. The distance in a connectivity trace for a feature runs from the middle of the feature to the edge of the nearest target feature (e.g. large lake or large watershed), or 0 if the feature itself is a target feature (e.g the feature has a large watershed).

Distance to the nearest lake

Distance from a lake with multiple lines under it to the nearest lake of a given size are attributed with 0 if they are in the correct size class or attributed as the mean value of the distances calculated for each of the centroids of the flowlines under given lake.

Distance to watershed of specified size

Some flowline features were too small to have individual watersheds delineated for them. If these features were located within a shed that is one of the target sizes, for example, 100 km², they should be attributed as being 0-m from the nearest 100 km² shed. However, these features are instead attributed with a small distance equal to the distance from the midpoint of the flowline feature to the upstream node of the first flowline feature that has a watershed delineated for it. These distances will generally be half of the length of the feature, but could be longer if there are several contiguous flowlines without watersheds. These values should almost always be less than 40-m and usually less than 10-m.

Stream Temperature

An artificial neural network flow model was used to predict daily mean stream temperature for the Jun-Aug periods in 1990-2008.[1] Daily temperatures were summarized for each year as:

- Mean Jun-Aug
- Mean July
- Maximum daily

The mean maximum daily temperature across the 1990-2008 period is included in the attribute database. Note that 2001 land cover was used for the entire period and predictions were not made for

[1] Stewart, J., Mitro, M., Roehl, E., and Risley, J. (2006). "Numerically optimized empirical modeling of highly dynamic, spatially expansive, and behaviorally heterogeneous hydrologic systems—Part 2." Proceedings of the 7th International Conference on Hydroinformatics, Nice, France, 1-8.

reaches with more than 6% missing upstream accumulated land cover data or with missing cumulative riparian and watershed Darcy values.

Stream Flow

Empirical stream flow models were developed for 23 flow metrics, including mean Jun-Sep flow; 5, 10, 25, 50, 75, 90, and 95% annual exceedance flows, and 10, 50, and 90% exceedance flows for three seasonal (spring, summer, fall) and two monthly (April, August) periods. The models are mixed effects regressions with log(discharge) as the response, several covariates (e.g., watershed area, precipitation, geology, land cover) as fixed effects, and USGS gage ID as a random effect on the intercept. Flow predictions were made for USGS water years 1984-2010, and are reported as the mean of these annual predictions.

Discharge data were downloaded from the USGS NWIS database for all Wisconsin stream gages with at least 1 year of daily discharge data between 1981 and 2011. Sites where impoundments made up more than 10% of the watershed area were excluded. At each site, exceedance flows for each time period were calculated for each year where discharge was recorded on at least 90% of the days in that period.

The following watershed characteristics were used as predictors in all models. Land cover variables for each site-year combination were derived from the closest in time of WiscLAND 1992, and NLCD 2001 and 2006. Effective precipitation was summarized over seven time frames: the period of discharge data and the following time periods prior to the period of discharge data: 1-3, 4-6, 7-9, 10-12, 13-24, and 25-48 months. All variables were transformed (square root or log) to approximate normality and then converted to z-scores by dividing by subtracting the mean and dividing by the standard deviation.

Variable Name	Units	Description
shedL	km^2	Total upstream watershed area
adjPermS	in/hr	Soil permeability from STATSGO soils multiplied by the proportion of pervious land cover
slopeL	%	Average land slope from 10 m NED
ag	%	Sum of pasture/hay and cultivated crops
drainedS	%	Agricultural drainage estimated as the intersection of agricultural land cover and poorly or very poorly drained SSURGO soils
waterL	%	Open water
wetland	%	Sum of woody wetlands and emergent herbaceous wetlands
pmean, plag1, plag4, etc.	mm/day	Average daily precipitation plus snowmelt from DayMet
pet	mm/day	Potential evapotranspiration derived from daily solar radiation and air temperature using Hargreaves method

Models were fit using the lmer function in the lme4 package in R. In addition to the fixed effects of the watershed characteristics described above, the models include a random effect of gage ID to account for correlated errors among multiple observations (years) at the same gage. Exploratory model selection indicated minor differences among models in which variables were significant predictors, so a common model structure was used for all response variables. The model formula is:

logQ~(1|site)+shedL+plag1+plag4+plag7+plag10+plag13+plag25+pmean*(pet+adjPermS+slopeL+ag+drainedS)+pet*(waterL+wetland)+adjPermS*slopeL

To evaluate model fit, the mean predicted flow metric was compared to the mean observed metric over the period of record at each gage. Gages with fewer than 10 years of record were omitted from this evaluation because the objective was to estimate the accuracy of long-term mean predictions rather than predictions for individual years.

In log space, the models all have very good fit, with R^2 values from 0.93-0.98. The root mean square error of predictions ranges from 35-84% among models. In general, higher flows (e.g., 10% exceedance) were predicted more accurately as percentages of their observed values than were lower flows. Similarly, flows in larger streams were predicted more accurately than flows in smaller streams. For example, the standard error of a summer mean flow prediction is 55% for a small stream with a predicted flow of 5 cfs, but only 21% for a river with a predicted flow of 500 cfs.

Pre-settlement stream flows were simulated by setting agricultural, drained, and impervious surface variables to zero and estimating wetland percent from "Original Vegetation Cover of Wisconsin" (Finley 1976).

Spatial Files Metadata

Flowlines

Polyline file of all streamlines included in the study, including lines that serve as individual seeds, lines that are part of lake seeds, and lines used in connectivity traces but not included as seeds. Most features were derived from the 24KGDB; these features generally have data in the HYDROCODE, HYDROID, HYDROTYPE, ROW_NAME, and RIVER_SYS_WBIC fields taken directly from the 24KGDB. The accuracy of these values was not verified by this project and is dependent on the accuracy of values in the 24KGDB. For a more in-depth description of these fields, see Wisconsin DNR 24K Hydrography Data Capture and Feature-Coding Decision Rules (http://dnr.wi.gov/maps/gis/documents/24khyd_decision_rules.pdf).

Field Name	Description
ROW_NAME	***Register of Waterbodies (ROW) name:*** The name that ranks highest on the ROW name source hierarchy: Geographic Names Information System 1 Wisconsin Geographic Names Council 2 Quad Map 3 WI Lake Book 4 WI Lake Map 5 County Surface Water Book 6 Master Waterbody File 7 WDNR 24K Hydro 8 Local Name 9

Unknown 10

RIVER_SYS_WBIC — ***River System WBIC***

7 digit unique number assigned to a river system. WBIC = Waterbody ID Code

HYDROID — ***Unique ID assigned to each individual hydro feature***

HYDROCODE — ***Code to indicate flow and duration***

100 - Primary Flow Over Land Perennial
101 - Secondary Flow Over Land Perennial
110 - Primary Flow Over Land Intermittent
111 - Secondary Flow Over Land Intermittent
120 - Primary Flow Over Land Fluctuating
121 - Secondary Flow Over Land Fluctuating
200 - Primary Flow In Water Perennial
201 - Secondary Flow In Water Perennial
210 - Primary Flow In Water Intermittent
211 - Secondary Flow In Water Intermittent
220 - Primary Flow In Water Fluctuating
221 - Secondary Flow In Water Fluctuating

HYDROTYPE — ***Type of water feature***

502 - Cranberry Bog
503 - Stream/River Centerline
504 - Wetland Gap
505 - Ditch/Canal
506 - Stream Extension
507 - Flow Potential
508 - Stream/River, single-line

seedtype — ***Feature type as considered by the attribution project***

Dangle- feature added to correct topology but not attributed
Lake- flowline associated with a feature designated as a lake
Great Lakes- flowline under the Great Lakes
Isolated Stream- landlocked stream (non-lake) flowline
Network Stream- Non-landlocked stream (non-lake) flowline

REACHID — ***Unique ID assigned to each seed (unit of attribution)***

ID is same as HYDROID for stream features and same as waterbody HYDROID for lake features; some features in the Great Lakes do not have REACHIDs

AGGID — ***Unique ID assigned to lakes and confluence-bounded streams***

ID is same as REACHID for lake features; new IDs were created for stream reaches

TRACEID — ***Unique ID for each feature generated by attribution project***

The TRACEID corresponds with values in the relationship table and can be used to determine to-from connectivity between features and to run traces along the flowlines

Waterbodies

Polygon file of all waterbodies included in the study, including lakes assigned individual REACHIDs and other waterbodies that served to inform the dimensions of stream features. Most features were derived from the 24KGDB; these features generally have data in the WATERBODY_ROW_NAME, HYDROID, HYDROCODE, HYDROTYPE, and WATERBODY_WBIC fields taken directly from the 24KGDB. The accuracy of these values was not verified by this project and is dependent on the accuracy of values in the 24KGDB. For a more in-depth description of these fields, see Wisconsin DNR 24K Hydrography Data Capture and Feature-Coding Decision Rules (http://dnr.wi.gov/maps/gis/documents/24khyd_decision_rules.pdf).

Field	Description
WATERBODY_ROW_NAME	***Waterbody Register of Waterbodies (ROW) name***
	The name for the areal water feature that ranks highest on the ROW name source hierachy.
HYDROID	***Unique ID assigned to each individual hydro feature***
HYDROCODE	***Code to indicate flow and duration***
	6011- Ditch/canal, perennial
	6021 - Stream/River, perennial
	6103 - Cranberry Bog, fluctuating
	7011 - Backwater, perennial
	7012 - Backwater, intermittent
	7021 - Fish Hatchery, perennial
	7031 - Flooded Excavation, perennial
	7043 - Innundated Area, fluctuating
	7051 - Industrial Waste Pond, perennial
	7061 - Lake/Pond, perennial
	7062 - Lake/Pond, intermittent
	7071 - Reservoir/Flowage, perennial
	7081 - Sewage Disposal Pond, perennial
	7091 - Tailings Pond, perennial
HYDROTYPE	601 - Ditch/canal
	602 - Stream/river
	610 - Cranberry bog
	701 - Backwater
	702 - Fish Hatchery
	703 - Flooded Excavation
	704 - Innundation Area
	705 - Industrial Waste Pond
	706 - Lake/Pond
	707 - Reservoir/Flowage
	708 - Sewage Disposal Pond
	709 - Tailings Pond
	710 - Unspecified Open Water

WATERBODY_WBIC	***Unique 7-digit waterbody ID code (WBIC) assigned to the areal water feature***
REACHID	***Unique ID assigned to each seed (unit of attribution)***
	Features that were not included as seeds do not have REACHIDs
LAKE	***Indicator of whether feature was considered a lake (*****≥5 acres)**
	0- not a lake
	1- lake

Nodes

Spatial data of all flowline nodes extracted from a landscape network. The field **node_cat** indicates the node category of each feature. *Confluence* indicates that a node is at the point where two flowlines meet and flow into a third flowline. *Source* indicates nodes at the headwaters of the network and *outlet* indicates nodes at the termination of the network. A designation of *Pseudo node* indicates that one feature flows into another feature in the absence of a confluence at that point. The field **pointid** are unique identifiers for each point.

Node Relationships

Table providing information on the relationship between node features and flowline features, extracted from a landscape network. The field **TRACEID** contains unique flowlines identifers that are also found in the **TRACEID** field in the flowlines file. The fields **fromnode** contains the unique pointid (from the feature class ***nodes***) at the upstream end of a flowline and the field **tonode** contains the unique pointid at the downstream end of the flowline.

Reaches

Raster file of all seeds used in analysis, including lakes, streams, and border lines for Lake Michigan and Lake Superior so that these features can receive direct overland flow. Values correspond with feature REACHIDs. This raster can be used to create new channel-level attributes for analysis or to visualize the dimensions of the seeds used in analysis.

Relationships

Table of to-from relationships between all adjoining flowline features (derived from landscape network relationship tables). Values in the **FROM_TRACEID** and **TO_TRACEID** fields correspond with values in the **TRACEID** field of the flowlines file. Features with a particular **TRACEID** in the **FROM_TRACEID** field connect downstream directly with the matching feature in the **TO_TRACEID** field. Terminal features with no further downstream connections are not listed in the **FROM_TRACEID** column of this table.

Riparian Buffers

Raster file of 60-m riparian buffers created for all seeds in the study area. Values in the **Value** and **CATCHID** columns correspond with feature REACHIDs. Values in the **TOCATCHID** column correspond with the riparian buffer that is directly joining a feature on the downstream side. Features with 999999 in the **TOCATCHID** do not have a downstream feature into which they flow. These features may be the end of an internal drainage network or they may flow out of state or into one of the Great Lakes. This raster can be used to create new riparian buffer-level attributes for analysis. This raster can also be used to run trace attribution in order to obtain total cumulative upstream riparian data for features.

Watersheds

Polygon file of all watersheds delineated for features. The **CATCHID** field contains the REACHID value for each feature. Three non-contributing HUC12s (that do not have any internal seeds) have **CATCHIDs** of 100000001, 100000002, and 100000003. The **TOCATCHID** field indicates the **CATCHID** of the watershed that each feature flows into, either through direct flowline connections or through overland flow. **TOCATCHID** values of 999999 indicate that catchments flow into one of the Great Lakes or flow out of state to features we did not include in analysis. This polygon file can be used to create new watershed-level attributes for analysis and to run trace attribution to obtain total cumulative upstream watershed data for features.

Appendix A: Correcting Stream Gradient

In some cases, stream gradients that we calculated using an overlay of a DEM were negative (i.e., the minimum elevation was higher than the maximum). We developed a novel algorithm for correcting these cases. The algorithm is relatively complex, but is founded on a simple rule: if a stream gradient is found to be negative, anchor the downstream feature (the “low anchor”) as well as the upstream feature which has a higher minimum elevation (the “high anchor”), and linearly interpolate elevations for all reaches between. The high anchor refers to the next upstream reach that has a minimum elevation greater than the minimum elevation of the low anchor. A more thorough explanation of the process is described in Figure A1.

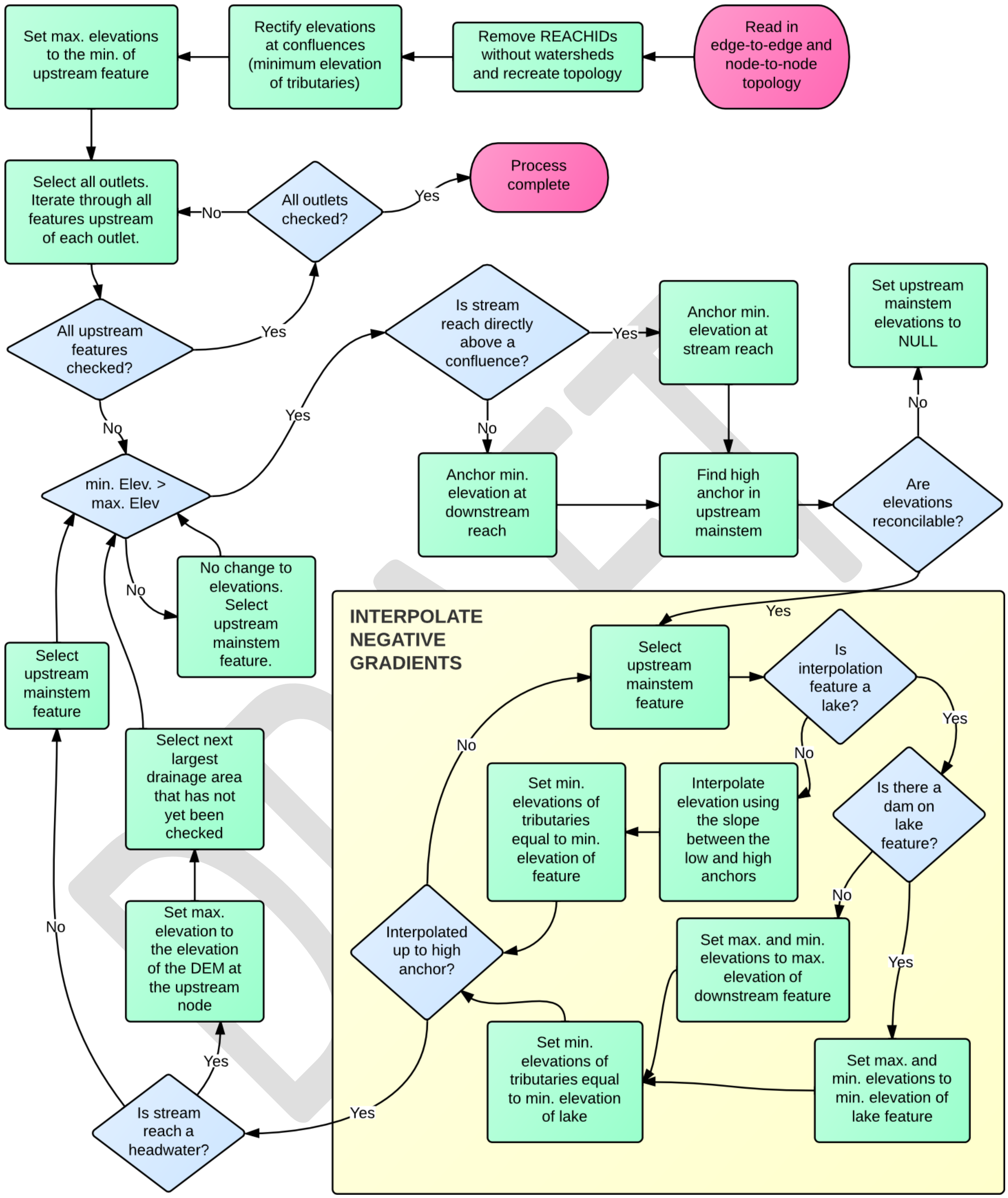


Figure A1. Flowchart describing the algorithm used to correct negative stream gradients. The term "upstream mainstem" refers to the next upstream reach with the greatest drainage area. The term "reconcilable" refers to cases where a negative gradient could be corrected using a high anchor with a minimum elevation higher than the minimum elevation of the low anchor. Gradients were "irreconcilable" if a high anchor could not be found before reaching a headwater.