Memorandum   February 23, 2011

To: File

From: Peter Leinenbach

Subject: Automated watershed digitization using NHDPlus datasets and the ArcHydro extension.

Procedures described in this memorandum can be used to quickly develop a "first cut" estimates of upstream contributing areas (i.e., watersheds) for biological sampling sites in the Puget Sound basin. All of the input datasets used with this method are readably downloadable from the NHDPlus website (http://www.horizon-systems.com/nhdplus/data.php). One important thing to point out is that the NHDPlus Grid data associated with this data source is at a relatively course resolution (30m), however these datasets have been processed in order to minimize hydrological errors when calculating watershed boundaries. Therefore, many of the shortfalls associated with using 30m DEM data to calculate watershed areas are eliminated. Also, NHDPlus processed datasets are available for the entire United States which helps ensure that a consistent watershed area calculation can be developed for various projects over the years.

Method Verification

Verification for this tool (method) was accomplished by comparing watershed boundaries calculated using the tool against watershed boundaries that had previously calculated with high resolution DEM data (10m), and were verified with topomaps, aerial images, and nhd line work. These 89 sites were located throughout the Puget Sound Basin. It is assumed that the watershed boundaries calculated using the 10m DEM were "accurate" as a result of the using high resolution data and the extensive post-processing evaluations for each site.

There were two factors evaluated for the method verification.

The first factor evaluated was whether the pour point initiation location was accurately determined from the lat/long information associated for each of the sites. In other words, it is important to check if the site location is situated exactly on the 30M DEM derived stream line. (As will be described further on in this memo, several automated tools are used to ensure that this will occur. Despite these efforts, it is still very important to verify the results.) It is important to note that this effort associated with calculating the 10m DEM derived watersheds was accomplished through a labor intensive manually effort. There are two possible errors associated with using this new automated methodology to estimate the pour point: 1) Snapping to the "wrong" 30m DEM derived stream line (i.e., The lat/long for a test site is located next to several 30m DEM derived stream lines and it picks the closest one, which is not the "target" stream.); 2) Not snapping to any stream line (i.e., The lat/long for a test site is located too far from any 30m DEM derived stream line and therefore does not snap to anything.). The error associated with the first issue results is a watershed boundary which is not located on the correct stream; it is very difficult to determine this error without visual inspection of each result using topomaps, aerial images, and nhd line work. The error associated with the second issue is much easier to determine; the watershed boundary will be very small and strangely shaped.

The second factor evaluated during model verification it the "shape" of the calculated watershed boundary, as compared to the previously determined 10m derived watershed areas. Observed differences will be a result of the DEM datasets used in the analysis (i.e., 30M NHDPlus DEM vs. 10m NED). There is no user actions that can rectify any observed discrepancy, and slight differences will be expected. Previously analysis associated with this topic during the EMAP-West effort in the 1990's indicated that slight differences associated with calculated watershed boundaries do not significantly affect analytical tests. Accordingly, verification for this parameter will be more of a "qualitative" check.

Only two (2) of the 89 watersheds had a problem with the calculated pour point location. One of the problem watershed boundaries was associated with Purdy Creek; a very small creek located in the very flat floodplain of the lower Skokomish River (see image below). The red dot is the lat/long location of the sampling location and the red line is the associated watershed calculated using the 10m DEM. The Orange dot is the "snap" location for the sampling point (Recall that the point was snapped to ensure that it will be located on top of the 30m hydrologic stream line (Indicated by the blue line).). The resulting watershed (Orange line) is incorrectly associated with the Skokomish River. The snapping tool snapped to the closest line, and the closest line was not the Purdy Creek. A similar problem was observed for Lower Vance Creek, which is also located in a very low gradient floodplain area. Accordingly, it is very important to validate all estimated watershed boundaries to ensure that they are located on the correct stream (Especially for very low gradient areas.).
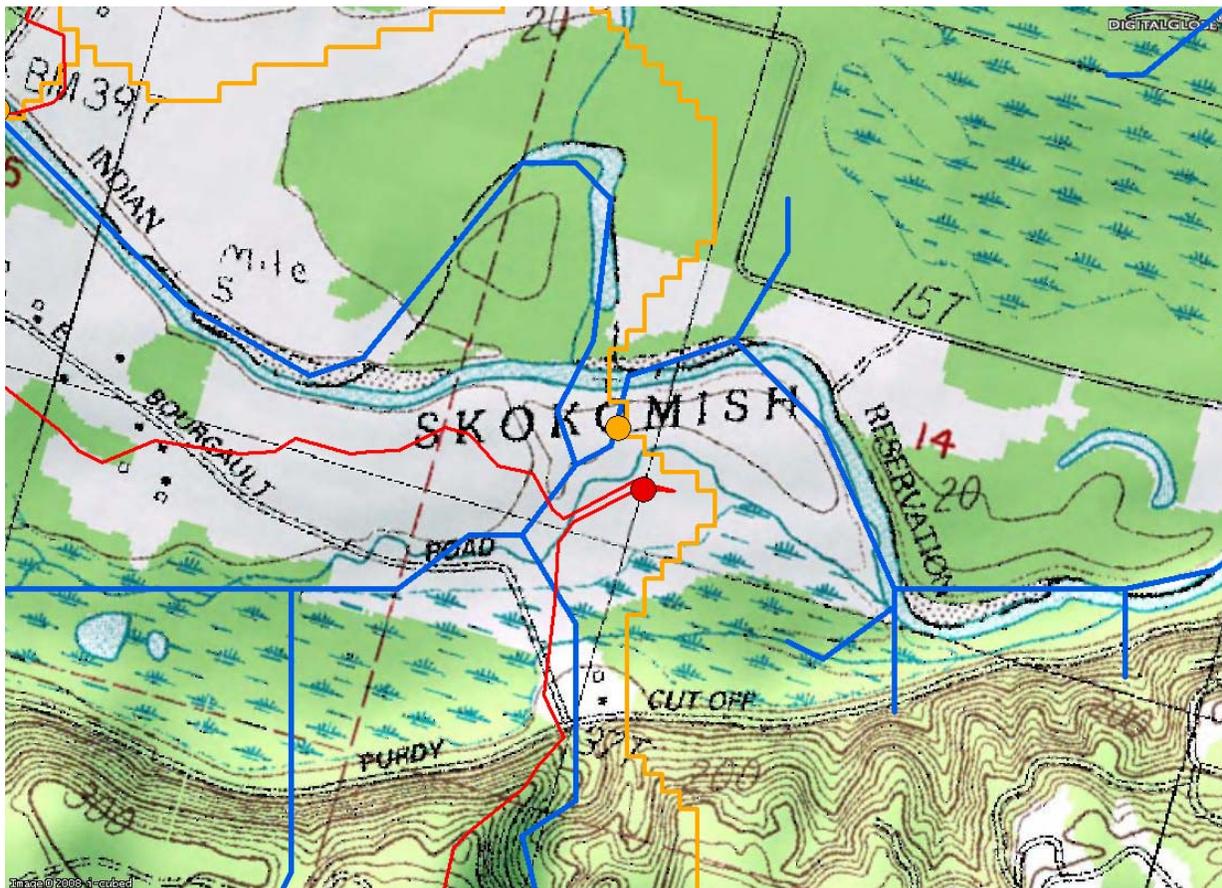
The image below illustrates a "typical" example of watershed boundaries calculated using the two methods. Once again, the orange line is the watershed area calculated using this new methodology, and the red line is the watershed calculated using the 10m DEMs. The shapes of the two watershed boundaries closely follow each other. There are two differences between the two methods. First, the orange line has a more "jagged" edge as a result of using 30m data to calculate the boundary. This condition does not affect subsequent landscape sampling results (i.e., the landscape data is often at 30m resolution also). The second effect is small boundary differences. The most obvious example is located near the mouth. Both watershed boundaries are accurate, based on their respective input datasets, and both seem very reasonable using a "qualitative" filter. Regardless of these small differences, the ultimate consequence of these small differences are minor.
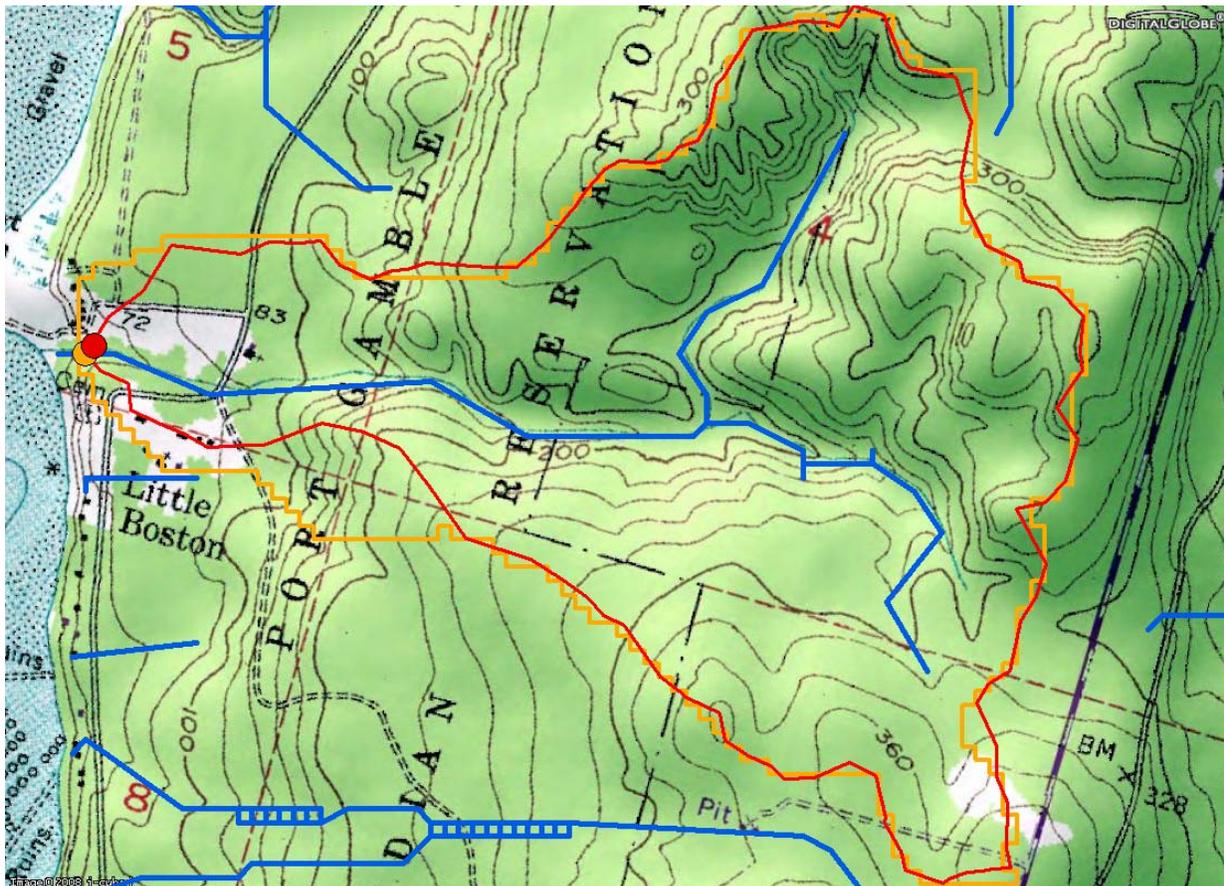
Methods

There are several specific tasks associated with the automotive process of developing watershed boundaries using the NHDPlus dataset and ArcHydro tools.

Task One – Data Preparation

The first step is to download data associated with the Puget Sound Basin from the NHD website - http://www.horizon-systems.com/nhdplus/data.php. The next step is to unzip the downloaded files onto your harddrive (ensure that the paths are preserved when unzipping). The download files should have the structure illustrated in the image below. Make sure there are no spaces in the folder names.

Task Two – Setup Attributes and Spatial Indexes and build Pyramids with the NHD Datasets

Follow instructions associated with the **Tutorial One** on the NHDPlus Tutorial webpage - http://www.horizon-systems.com/nhdplus/tutorials.php. Although it might not be necessary to implement this step, it will increase the processing efficiency associated with the tools presented in this memorandum.

Task Three – Using ArcHydro to create datasets used for watershed delineation.

Follow instructions associated with the **Tutorial Seven** on the NHDPlus Tutorial webpage - http://www.horizon-systems.com/nhdplus/tutorials.php. This process utilizes the NHD datasets you setup in the previous steps. It is important to **not** follow one of the directions in this tutorial – There is a bug in the program which makes it not work as directed. Specifically, do not "save" the mxd before you add data into the mxd project under the section called "Using the preprocessed Data for Watershed Delineation". However, it will work if you add the data and then save the file.

Task Four – Automotive "snapping" of sampling locations to the DEM derived stream layer

Obtain a "point" shapefile of the sampling locations. The file needs to be in a Geographic Projection (There is a bug in the ArcMap tool "Near" tool presented further on in this memo, and it was found to work correctly on geographic projected data.).

Next, using the "Reclassify" tool in the Spatial Analyst extension, reclassify the NHDPlus flow accumulation (FAC) grid so that 0-100 is "NoData", and all other values are classified as "1". Make sure that the "Change Missing values to NoData" is checked. Also make sure there are not spaces in any part of the "Output raster" name.
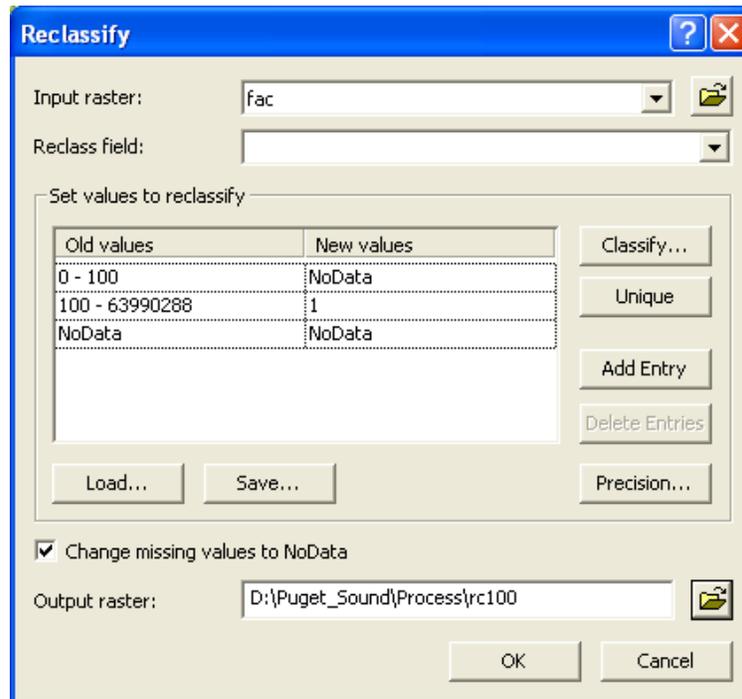
The next step is to convert this reclassified raster into a line feature using the "Conversion - Raster to Feature" tool in "Spatial Analyst". Once again, project the created polyline feature to the same Geographic Projection as the point feature described above.
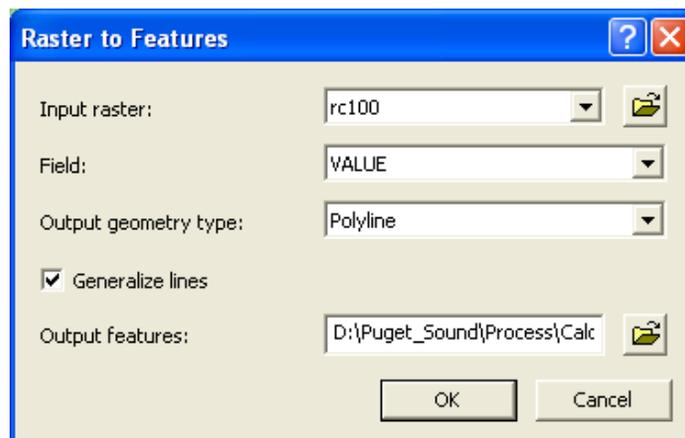
The next step is to use the "Near" tool in the ArcMap Toolbox to determine the latitude and longitude of the nearest calculated stream line to the point. The Search Radius should be set to 0.005 decimal degrees (approximately 0.35 mile). (For the 89 test sites in the Puget Sound a distance of 0.001 worked for almost all of the sites, however it was necessary to increase it to 0.005 for a couple of them. Make sure that the "Location" button is checked; it will add "Near_X" and "Near_Y" columns in the point cover shape file, and represent the longitude and latitude to the nearest line segment from the point (i.e., sampling location).
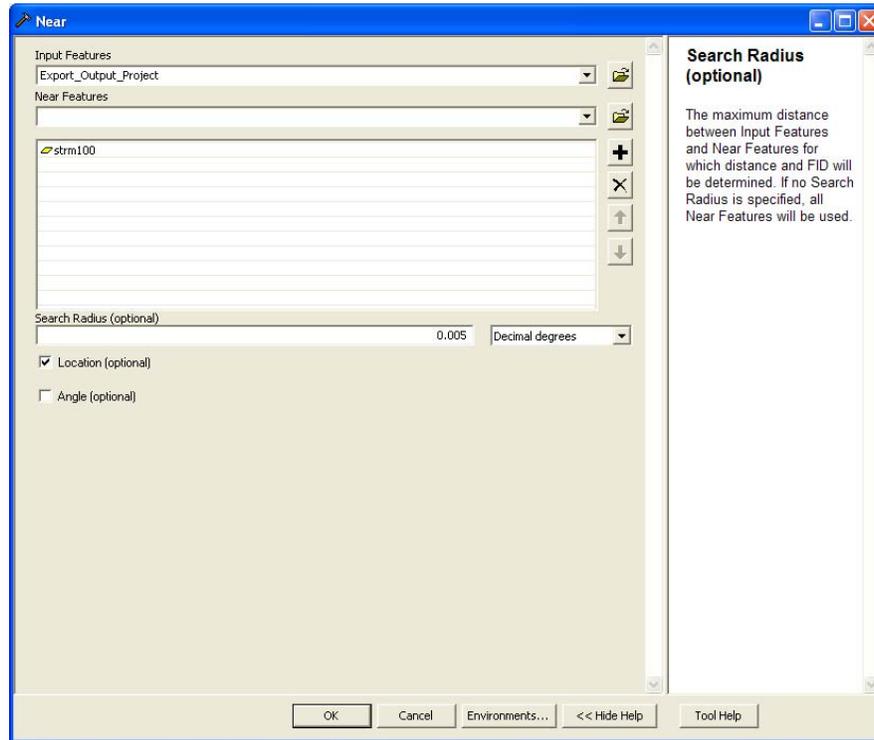

FID	Shape	NAME	New_Name	Code_Order	NEAR_DIST	NEAR_X	NEAR_Y
67	Point	BERTRAN	Bertrand Creek	1	0.000097	-122.508289	48.963191
68	Point	DEEPCR	Deep Creek	2	0.000123	-122.547013	48.845039
66	Point	WHATCOM	Whatcom Creek	3	0.000015	-122.427864	48.753695
62	Point	CHUCKAN	Chuckanut Creek	4	0.00005	-122.484793	48.70211
64	Point	OTOOLEC	Otoole Creek	5	0.000159	-121.919164	48.510679
69	Point	PRESSEN	Pressetin Creek	6	0.000283	-121.850652	48.516888
65	Point	PILCHUC5	Pilchuck Creek 1	7	0.000268	-122.216625	48.215306

Record: 0   Show: All Selected   Records (0 out of 79 Selected)   Options ▾

The next step is to export the attribute table associated with this point shapefile to a "dbf" file by using the "Export" tool in the "Options" button located at the bottom of the Attribute Table.
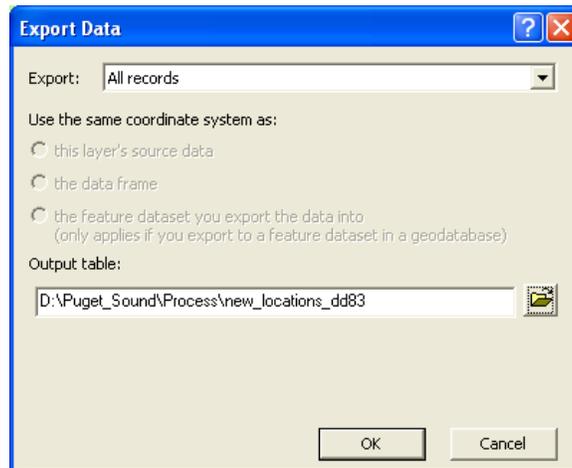
Then add this exported table into the ArcMap project. The next step is to "Display XY Data" for this added attribute table. Set the projection, and assign the X and Y coordinates to "Near_X", and "Near_Y". Save the result of this process as a new shapefile. Reproject this saved file back into the projection associated FAC grid.
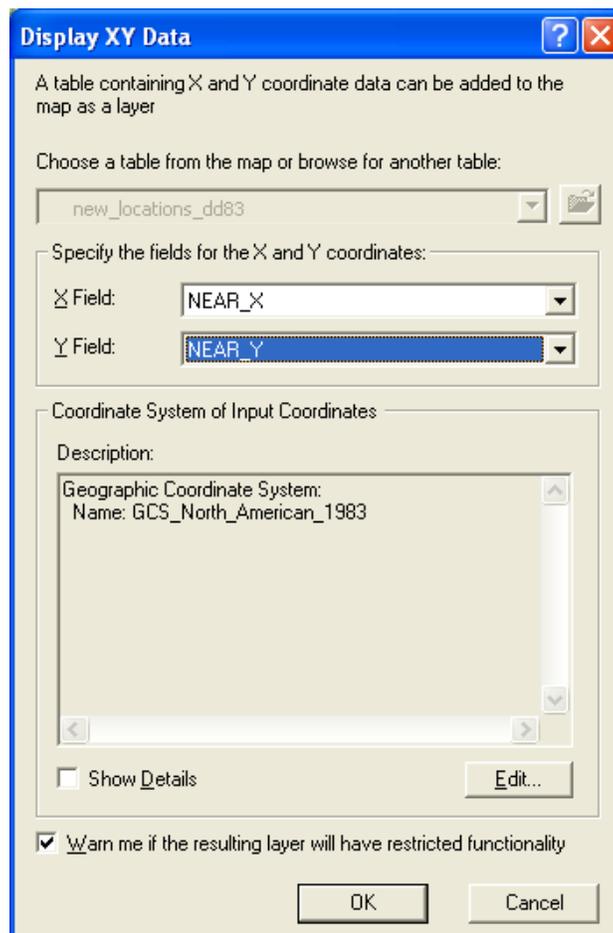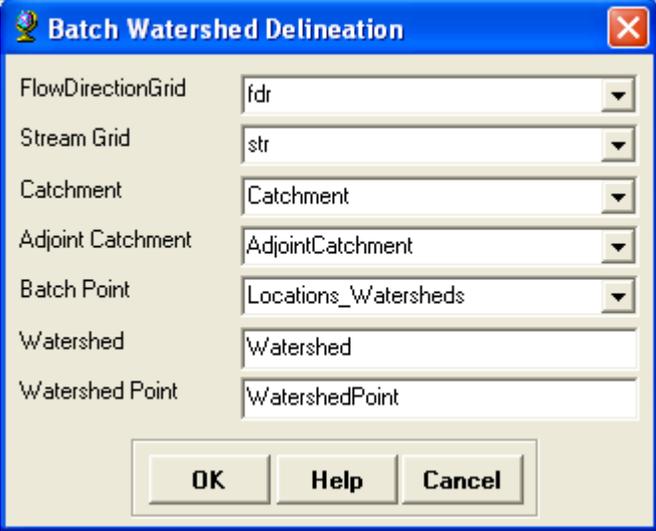
The resulting point coverage will be located right on top of the stream line shapefile which was created from the hydrologic stream layer (i.e., calculated from the FAC grid). The red dot on the image below illustrates the "original" location of the sampling point; the yellow dot is the "moved" location. The green blocks in this image represent the defined stream GRID developed from the FAC grid. The steps implemented up to this point for Task 4 were intended to "move" the sampling point on top of the stream Grid dataset so that an accurate watershed area can be calculated from the NHDPlus datasets.
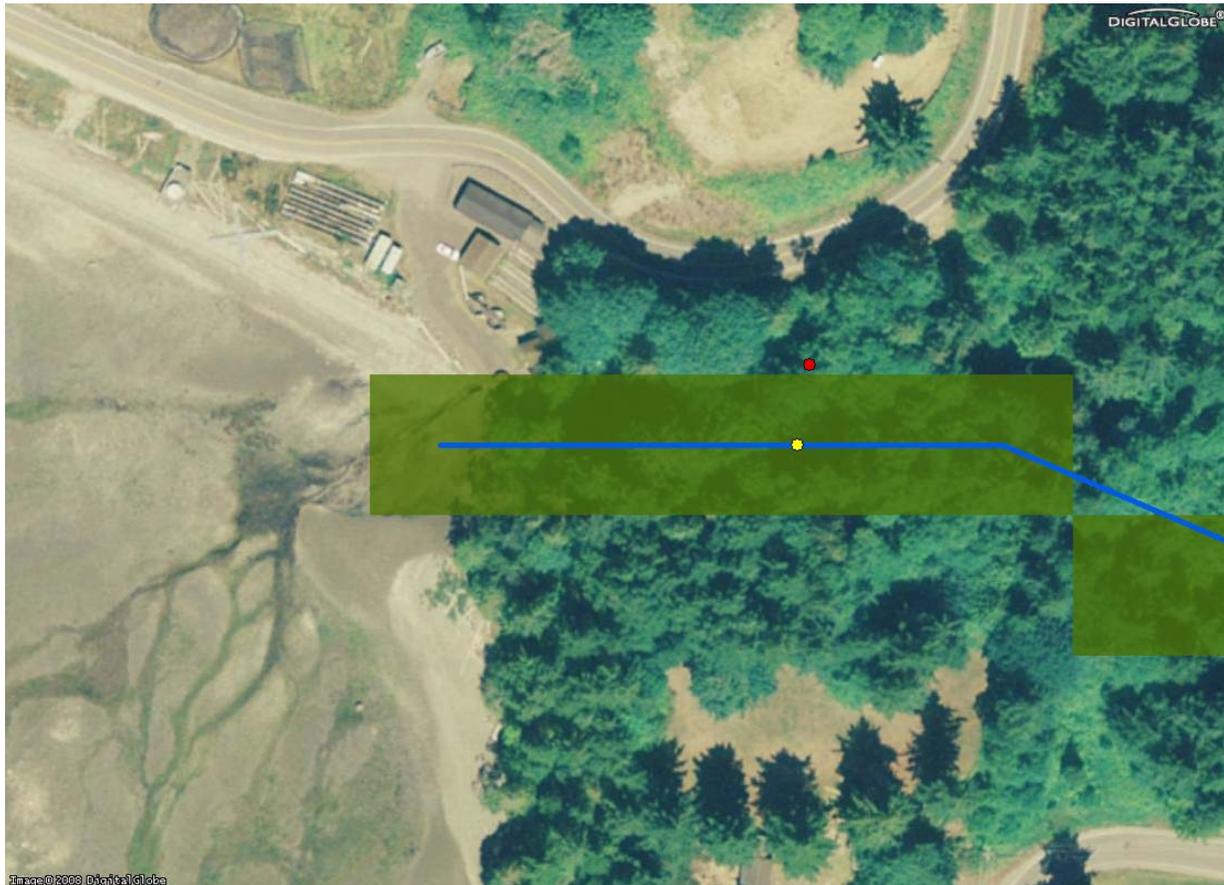
Task 5 – Calculate Watershed Areas

The next step is to calculate the watershed area using the "Batch Watershed Delineation" tool located in the ArcHydro extension. Most input datasets used with this tool were developed during Step 3, however the "Batch Point" dataset was developed during Task 4. It is necessary to add several columns in the "Batch Point" dataset attribute table: 1) "Name", 2) "Batchdone", 3) "Snapon", and 4) "Descript" (The "Batch Watershed Delineation" tool calls these columns of information during its processing.) "Name" and "Descript" columns are text format and the "Batchdone" and "Snapon" are in short integer format. The "Name" column is used as the unique identifier by the tool. The "Batchdone" indicates if a watershed needs to be calculated for the point; "0" indicates a watershed boundary is needed and will be calculated, "1" indicates to "skip" – The tool automatically puts a "1" in this column when it finishes calculating a watershed area, or a "-1" when it fails to accomplish this task. The "Snapon" should be set to "0" in order to turn this function off – It did not work properly during tool testing ("Snapping" is accomplished using other steps outlined in this memo.).

Task 6 – Check the results

It is important to check for errors in the calculated watersheds. The quickest way to check for errors is to sort watershed area column in the resulting watershed's shapefile attribute table: The most obvious error is that the calculated watershed area is very small as a result of not having the "pour point" located on the Grid stream layer (which was derived from the FAC Grid). An example of this situation is indicated by the "yellow" dot in the image below – This dot represents the result of "snapping" the "original" location (indicated by a "red" dot in the image) to the line (Recall Task 4). Although the point (i.e., the yellow dot) is now located on the line, it is possible that this location is not directly over the FAC stream Grid (Indicated by the green blocks on this image.). Because this point is not located on top of the Grid data, the watershed boundary will be small and not represent the actual watershed area. These small incorrect watershed areas need to be recalculated using the methods described in the next step.

Task 7 – Correct problem watershed area pour point location

The first step is to identify watersheds with have been incorrectly derived through an inspection of the size of the watershed. The quickest way to accomplish this task is to sort the "Shape_Area" column in the resulting **watershed shapefile** attribute table. As you can seen in the image below, there are several watershed which have very small watershed areas – The values represent square meters, and therefore a "Shape_Area" of 900 represent just one 30 meter pixel. You will need to visually investigate the watersheds with larger areas represented in this column. Often there is a large break in size which clearly shows you "problem" watersheds; For example, there is a large break between 13,500 and 1,756,799. Next, open the **pour point shapefile** and highlight the pour points associated with these incorrectly derived watersheds and save these highlighted pour points as a new point shapefile.

Shape *	OID *	Shape_Length	Shape_Area	HydroID	Name
Polygon	24	120	900	239	Potlatch Creek - site 1
Polygon	8	120	900	207	SEABECK
Polygon	21	120	900	233	Big Anderson Creek
Polygon	18	120	900	227	Swift Creek - site 2
Polygon	10	120	900	211	Steele Creek - South Fork
Polygon	20	180	1800	231	Vance Creek - site 2
Polygon	16	240	1800	223	Potlatch Creek - site 2
Polygon	5	180	1800	201	DUCKABU
Polygon	17	360	2700	225	Skokomish River - Site 2
Polygon	23	420	3600	237	Enetai Creek - site 1
Polygon	25	480	3600	241	Purdy Creek - site 2
Polygon	4	300	3600	199	Dickerson Creek
Polygon	7	360	4500	205	Little Boston
Polygon	6	720	13500	203	Little Anderson Creek
Polygon	12	8280	1756799.999999	215	Enetai Creek - site 2
Polygon	15	10800	2807100	221	Olalla Creek
Polygon	13	11100	3182399.999999	217	Harding
Polygon	2	17580	4073399.999999	195	Boyce Creek
Polygon	22	17580	7781400.000002	235	DEWATTO
Polygon	9	25500	8564400.000001	209	Stavis Creek
Polygon	14	30780	15303600.000001	219	LITTLEA
Polygon	11	26880	16187400.000002	213	Burley Creek
Polygon	3	31800	24082200.000001	197	Chico Creek

Record: 0   Show: All Selected   Records (0 out of 25 Selected)

The next step is to use the "Snap to Pour Point" tool in the ArcMap toolbox on the newly created pour point shapefile which contains the sites with incorrectly derived watershed areas. Make sure that you designate a **unique** numeric Identifier in the "pour point field" (This is the value that will be attributed to the resulting grid produced by this tool.). The snap distance should be set to 30 meters; Setting the value at 30 will ensure that the tool will search in the adjacent cells to find to lowest point (which is likely to be on the stream Grid because of Task 4 has already placed these locations near the stream Grid). The image below illustrates the result of this analysis; the orange highlighted cell is the low point next to the sampling location (i.e., yellow dot).
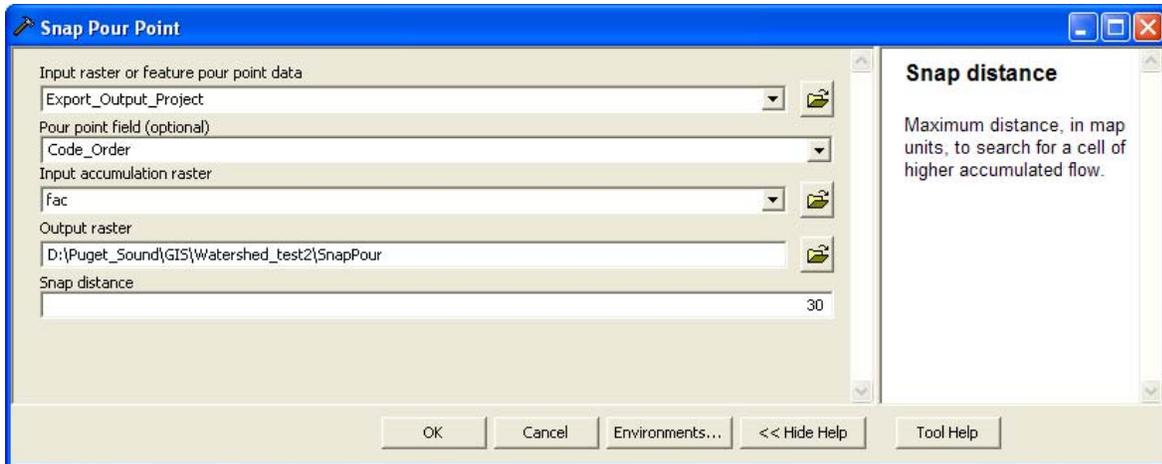
The next step is to convert the grid results produced by the "Snap to Pour Point" tool to a Point shapefile by using the "Convert – Raster to Features" tool in Spatial Analyst. The results are indicated by a green dot in the image below.
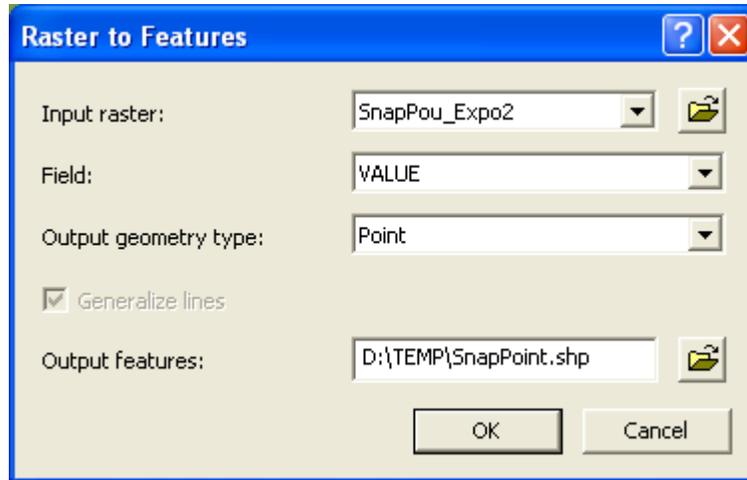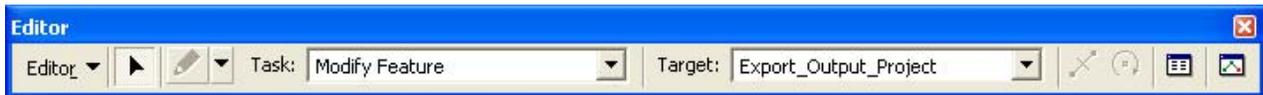
Once again, calculate the watershed area using the "Batch Watershed Delineation" tool located in the ArcHydro extension for these newly created pour points. Recall that several columns of information are required to be added to the attribute table associated with this point shapefile (see Task 5). The red line represents the watershed boundary that is associated with the green point.

Task 8 – Final Check to correct problem watershed area pour point location

It is important to check the results associated with the "Snap to pour point" tool described in Task 7 (i.e., check for "small" areas associated with these watersheds). Although most problem watershed areas will be eliminated using the method in Task 7, there will still be a couple of watersheds that are still not calculated correctly because the pour point still might not be located on the stream grid.

It will be necessary to manually edit the locations of these continuingly problematic pour points using the edit tool in ArcMap. Specifically, use the "Modify Feature" "Task", and make sure the "Target" is set to your point shapefile. Move the feature using the "▶" button.

Once again, run the watershed calculations using the "Batch Watershed Delineation" tool and check if the watershed boundaries are calculated correctly (It is possible that the point is placed just slightly in the incorrect place – It happens sometimes). If there is a problem, the pour point needs to be moved a little bit and tool needs to be run again.

Task 9 - Visually check if the pour point is located on the correct stream.

All previous steps utilized automated processes which dramatically quicken the speed which watershed boundaries can be calculated. Several steps were included in the methodology to ensure that the watershed boundary started on the "stream line". As with any automated watershed calculation, it is important to have a final check to see if the watershed area is derived on the correct stream. As mentioned in the beginning of this memo, there may be situations which the watershed boundary will be incorrectly derived because the pour point, although "snapped" to nearest stream line, is not located on the correct stream (Recall the image attached below illustrates this type of error). Accordingly, it is important to check the results using topomaps, aerial photos, and NHD stream line work in order to check for this error. Any problem sites need to be recalculated using the more labor intensive "manual" method described in the Task 8.
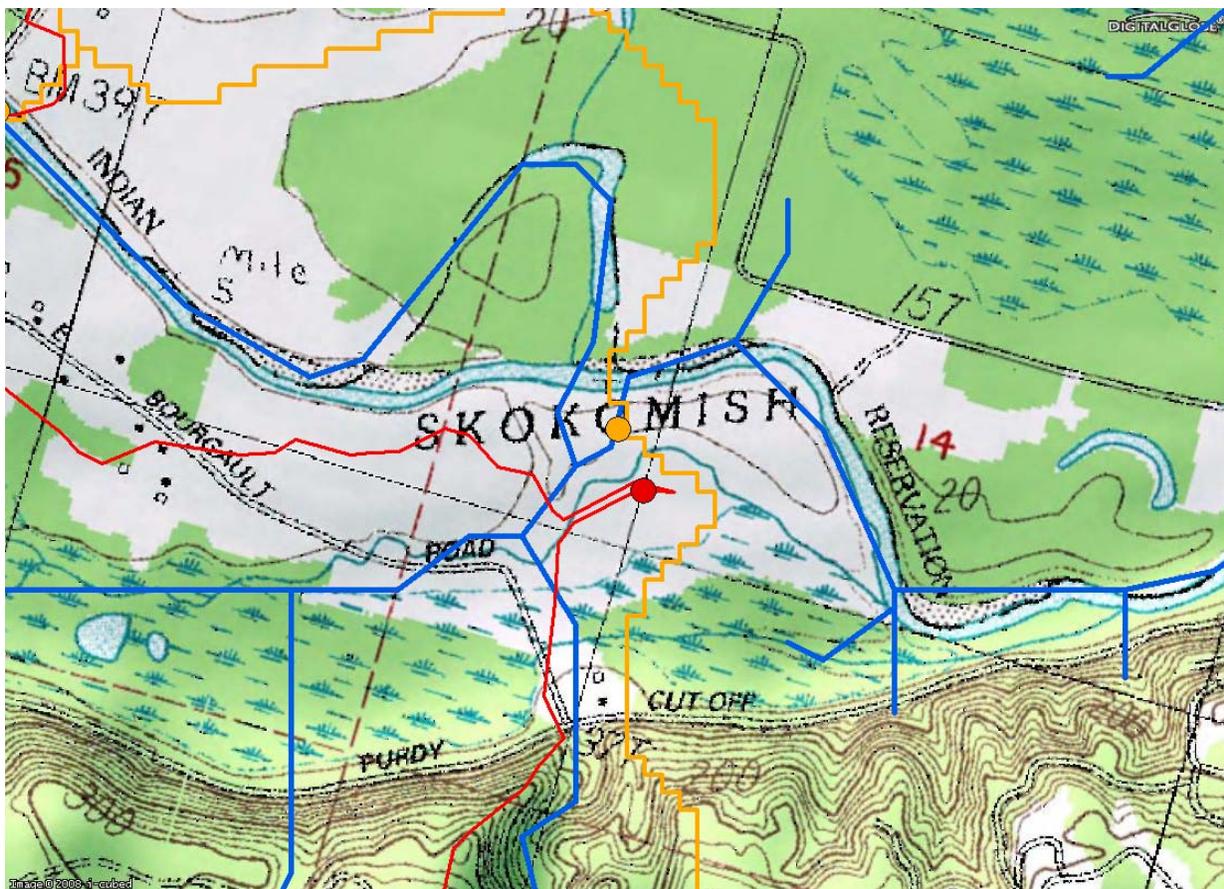
Example Results

The image below illustrates 1011 watershed boundaries calculated using this methodology for biological monitoring sites located within the Puget Sound region. It took around two days of effort to accomplish this task. It would have taken weeks of work using manual methods. It was necessary to use the manual method for only 10 sites. Without using the automotive processes, it would have been required to use the "labor intensive" manual method on all 1011 sites.