Delineating Reach Contributing Areas Using QGIS

Erin E Peterson
April 7, 2024

Contents

1. Before you begin

What are RCAs?

Reach contributing areas (RCAs) form a detailed tessellation of non-overlapping, edge-matching inter-basins. The RCA boundaries are based on topography and include nearby areas that would theoretically contribute overland flow, were it to occur to a given reach (Theobald et al. 2005). Delineating RCAs is an optional step, which will not be required in many cases. The two main reasons to undertake this extra data processing are to calculate:

1. Variables such as watershed area or stream order for each line segment, used to generate the spatial weights in the tail-up covariance model.

2. Additional variables representing characteristics RCA or watershed characteristics that can be used as covariates (i.e. predictors) in statistical models.

Do I need to calculate RCAs?

Oftentimes, national-scale streams datasets include RCAs, but they are referred to using different terminology. For example, they are referred catchments in the US NHDPlus V2 (Horizon Systems Corporation 2007) and NHDPlus High Resolution (Moore et al. 2019) datasets and subcatchments in the Australian Hydrological Geospatial Fabric (Geofabric) Surface Network dataset (Bureau of Meteorology 2012). These datasets are also attributed, which means that RCA attributes and watershed attributes describing land use, area, and stream order have been precomputed for each line segment and made available in separate relational databases (Stein et al. 2012; Wieczorek and LaMotte 2010). These attributes can be joined to the edges and if necessary, summed downstream in R using the SSNbler package function `accum_edges`.

If your streams dataset has an associated set of RCAs or it is attributed, you may not need to delineate RCAs using the methods outlined in this tutorial. If the streams are attributed, the variables needed to satisfy (1) and (2) above are likely already available. If there are additional variables you would like to calculate at the RCA scale, the existing RCAs associated with the stream's dataset can be used to summarise those for each edge segment.

In addition, variables other than watershed area can be used to generate spatial weights for the tail-up model that do not require RCAs to be generated. The simplest option is to weight every set of edges equally at confluences (e.g. all edge segments receive the same variable value). Other additive variables can also be used as long as they are additive, meaning that the values sum downstream. One common example is Shreve's stream order (Shreve 1967), which can be calculated in QGIS without generating RCAs. Note that Strahler's stream order (Strahler 1957) is unsuitable because it is not additive. Ultimately, it is up to each user to decide whether alternative variables, such as stream order, will adequately represent influence in a branching stream network, given the response (i.e. dependent variable). See Frieden et al. (2014) for examples and a general discussion about how to select variables used to compute the spatial weights.

2. Data and software requirements

Example data

This tutorial comes with an example dataset, which can be found in the zip archive **create_rcas.zip**. It contains all the data required to delineate RCAs (Table 1). This includes an error-free, topologically corrected set of edges that have been used to build a Landscape Network (LSN) using the SSNbler function `lines_to_lsn`, in addition to a digital elevation model (DEM) and waterbodies dataset. The create_rcas.zip archive also contains three sub-folders: 1) lsn, which is the LSN where the edges are stored; 2) work, an empty folder to store outputs created in this tutorial; and 3) final_outputs, which contains a reference version of the final outputs created in this tutorial.

Table 1. Data requirements for delineating RCAs.

Data	Format	File Type	Description
Edges	Vector	GeoPackage	Topologically corrected edges
DEM	Raster	GeoTIFF	Digital elevation model (m)
Waterbodies	Vector	GeoPackage	Lakes, reservoirs, estuaries, etc.

Software requirements

Three open-source software packages must be installed to successfully complete this tutorial (Table 2). QGIS version 3.34-5 (QGIS Development Team 2024) was used to develop this tutorial, which was the current stable long-term release (LTR) at the time. It has also been tested with QGIS 3.36.1, where there are slight differences in the tool inputs and outputs. The QGIS WhiteboxTools plugin provides many of the hydrologic processing tools required to generate RCAs, which depends on the WhiteboxTools Open Core (Lindsay 2014). All three software packages must be installed users.

Table 2. Software requirements

Software	URL for Download
QGIS Version ≥ 3.34 Long Term Release (LTR)[1]	https://www.qgis.org/en/site/forusers/download.html
WhiteboxTools QGIS Plugin	Video instructions[2]: https://www.youtube.com/watch?v=xJXDBsNbcTg&t=3s
WhiteboxTools Open Core	https://www.whiteboxgeo.com/geospatial-software/

[1] Be sure to download the QGIS Long Term Release (LTR), not the newest Development version shown in the big green button. Instead, click on the text underneath the button that says, “Looking for the most stable version? Get QGIS <version number> LTR”.

[2] To access QGIS Preferences on Windows, select Settings in the QGIS menu, scroll down and select Options. Follow all the additional instructions in the video.

We assume that the reader has a working knowledge of QGIS and is familiar with vector and raster data and processing tools. For those that need additional help, the QGIS documentation is a good place to start (https://docs.qgis.org/3.34/en/docs/index.html):

- Chapter 5: Getting Started
- Chapter 12: General Tools
- Chapter 16: Working with Vector Data
- Chapter 17: Working with Raster Data vector data.

ChatGPT is also a quick and easy way to get suggestions when you get stuck (https://chat.openai.com/). And of course, online discussion forums are a great resource. Remember, it’s rare to have a question that hasn’t been asked in some form before!

3. Getting started

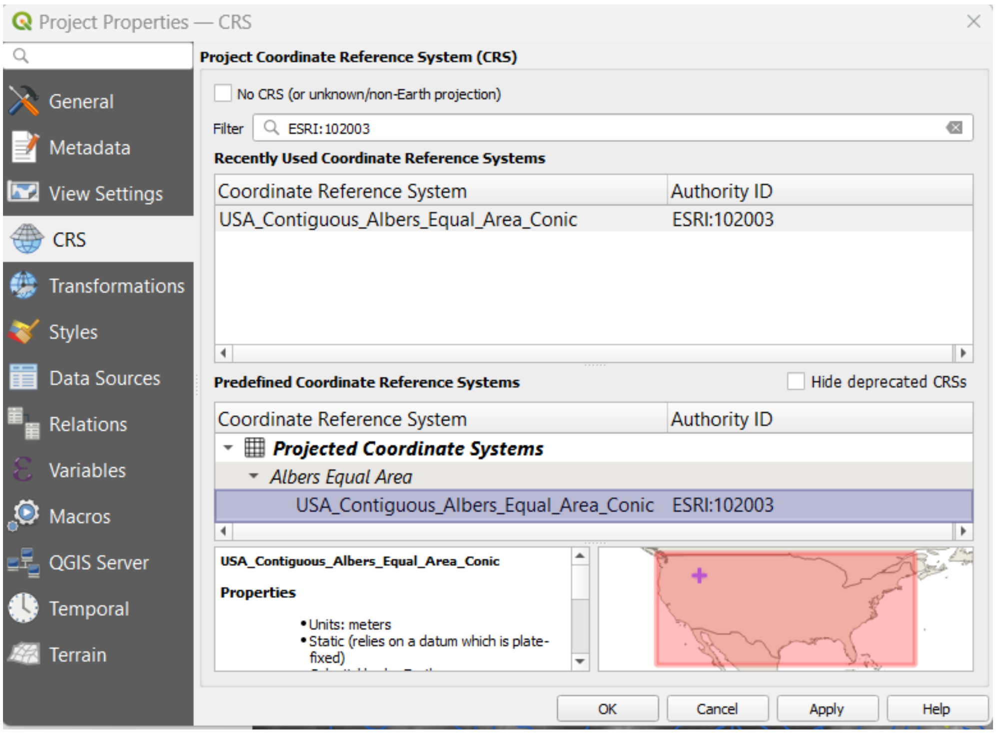
To begin,

1. Open QGIS and load edges.gpkg, dem.tif, and waterbodies.gpkg**,** which are found in the lsn directory.

2. Set the Project projection.

   a. In the main menu, click on Project, scroll down, and click on Properties to open the Project Properties window.

   b. QGIS will automatically set the projection to the projection of the first dataset loaded into the project. If you do need to set the projection, click on CRS in the left panel, search for ESRI:102003, and select USA_Contiguous_Albers_Equal_Area_Conic, as shown below. Click OK.



3. Save the Project. In the main menu, click on Project, scroll down, and click on Save As. Name and save the QGIS project in a suitable location. Remember to periodically save the Project throughout this tutorial if you would like to save your work.

4. Change the symbology of the layers to a colour scheme that makes it easy to distinguish between features.

   c. Double click on edges in the Layers Panel to open the Layer Properties window.

d. In the left panel, select Symbology. Change the colour, line width, etc. as desired and click OK. Repeat this process for all the layers. The layers should look something like this, depending on the colour schemes you selected.

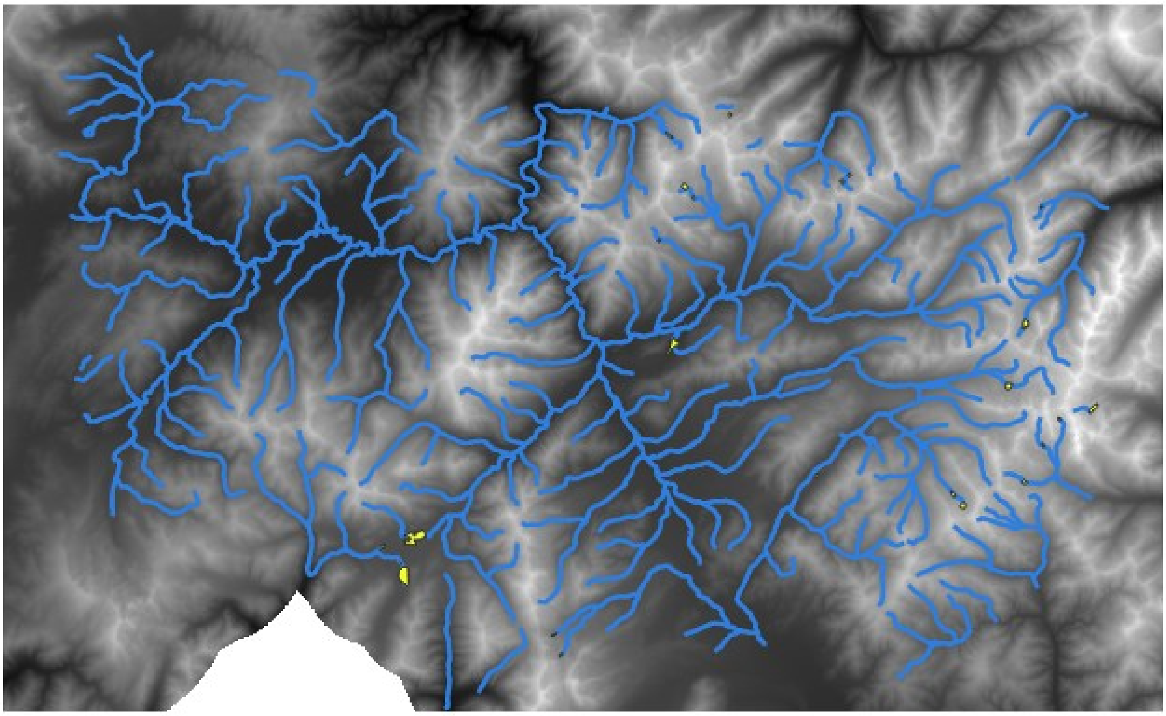


Figure 1. The example dataset includes edges shown as blue lines and waterbodies as yellow polygons. Elevation is displayed using a range of dark grey (lower elevations) to white (higher elevations).

Examine the edges near the borders of the DEM. Notice there are edge segments flowing away from ridgelines towards the boundaries. These edges are outside the true study area extent, which is the MiddleFork04 dataset used in the SSNbler and SSN2 vignettes. Extra edges along the periphery are included because RCAs delineated along the DEM boundaries may fail terminate at ridgelines, making them erroneously large. Additional edges surrounding the study region help ensure the RCAs within the actual study area are delineated properly.

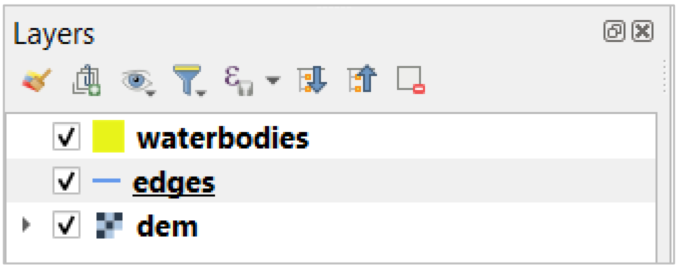
4. Format streams and waterbodies

Format edges

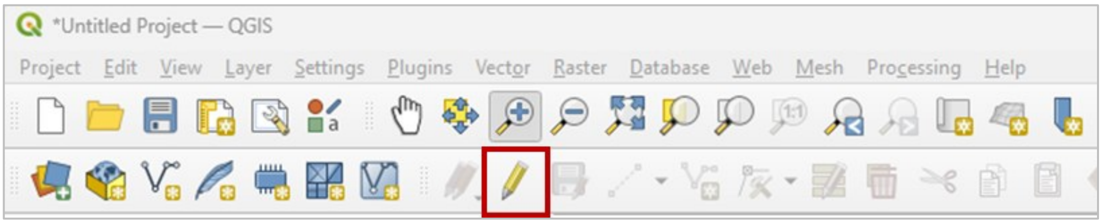
1. Add a new identifier column to edges

   An edges attribute should be created to relate the RCAs to the edges. It is better not to use the fid or rid fields because they may change if the LSN is regenerated rebuilt. Instead, create a new column for this purpose that will remain unchanged regardless of data transformations, conversions, and other manipulations.

a. Select the edges in the Layers Panel by left clicking on it.

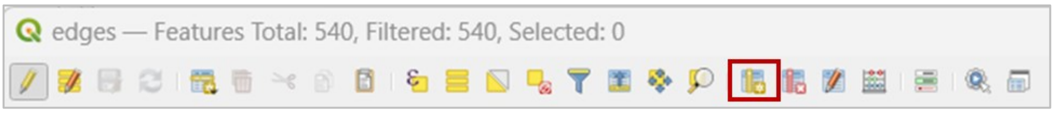


b. Turn on the layer editor by clicking the Toggle Editing button in the Digitizing toolbar (button indicated below in red square). If the Digitizing tool bar is not available, right click anywhere in the Menu, scroll down to Toolbars, and check the box next to the Digitizing toolbar to add it. The edges are now in editing mode. Note that the Toggle Editing button can also be accessed inside the attribute table menu.

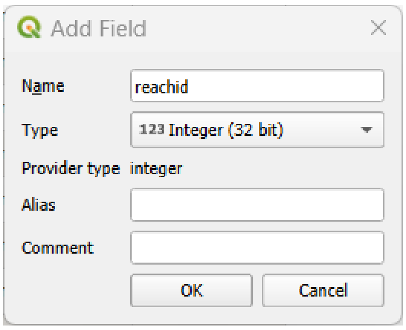


c. Open the edges attribute table by clicking on the Open Attribute Table button, , in the Attributes toolbar. If this toolbar is not visible, right click anywhere in the Toolbar and check the box next to Attributes Toolbar to add it.

d. Click on the New field button (red square below) to open the Add Field window.



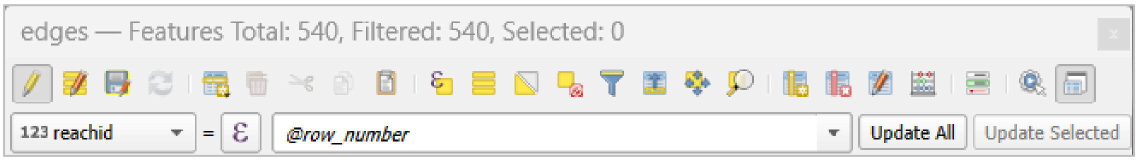
e. Add the field reachid using the arguments below and click OK.



f. Calculate reachid

i. If the message in the top of the attribute table indicates that any of the edges are selected, click on the Deselect features button, , in the attribute table menu to clear the selection.

ii. Set the arguments as shown below and click the Update All button. We assign reachid based on row number because this ensures that reachid ≥ 0.

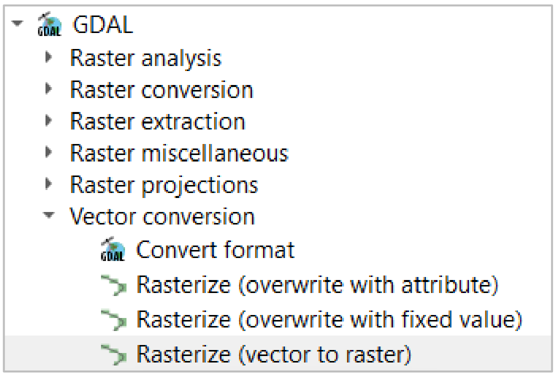


g. Click on the Save edits button, , Toggle editing mode off, and close the edges attribute table.

2. Convert edges from vector to raster

a. Click on the Toolbox button, , to open the Processing Toolbox.

b. Click on the arrow next to the GDAL toolbox and then again at Vector conversion to expand the list of tools. Double click on Rasterize (vector to raster) to open the tool.



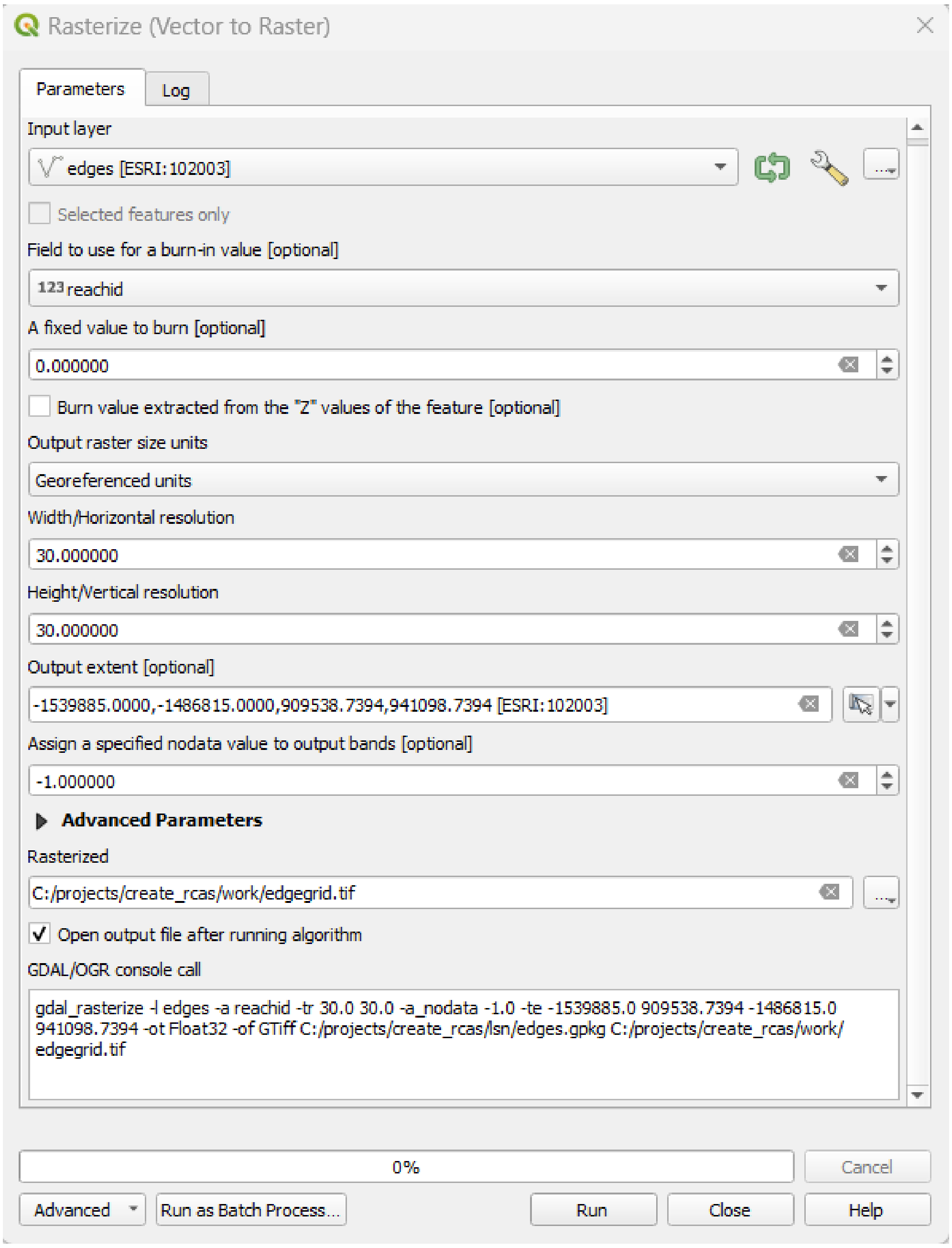
c. Set the tool arguments as shown below. Some things to note:

i. Field to use for a burn-in value [optional]. This is the edges attribute field value that will be stored in each raster cell. Selecting reachid will produce a raster layer with cell values equal to the associated edge reachid. Non-edge values are set to no data.

ii. To set Output extent [optional], click on the dropdown menu on the right-hand side, select Calculate from Layer, and then dem. This will populate the Output extent.

iii. Width/Horizontal resolution and Hight/Vertical resolution. This is the spatial resolution of the example DEM, but this will differ by dataset. Make sure to set this spatial resolution to match your DEM when working with your own data. You can find this in dem>Properties>Information. Look for Pixel Size under Information from provider.

iv. Click Run and then Close the tool when finished. The new raster edgegrid should be added to the Layers Panel.



d. Examine the edgegrid values. Use the Zoom In tool, , found in the main menu to take a closer look at the edgegrid. The edgegrid values should range from 3 to 539.

e. Use the Identify features tool, , to query values in edgegrid and edges. To query both datasets at the same time, select one of them in the Layers Panel, hold down the Ctrl key, and select the second one. Now query an edgegrid that intersects an edge (most should). Notice that the edgegrid raster cell values match the edges reachid. If they do not, a mistake has been made. Go back and rerun the Rasterize tool.

f. Visually assess the cell size and alignment for the edgegrid and dem. It is important that these two rasters have the same spatial resolution and that the raster cells are perfectly aligned.

    i. Make sure the dem and edgegrid are visible in the Layer Panel.

    ii. Zoom into the edgegrid and visually inspect the cell alignment of the two raster layers. Note that you may have to change the colour scheme of the DEM to make the cell boundaries more obvious. In the example below, the edgegrid is shown in black and the DEM values range from blue to yellow.

    Do not proceed until the cells are perfectly aligned. If differences do occur, it is likely due to:

    - Errors/typos in the input arguments. Pay particular attention to the Output raster size units, resolution inputs, and extent. Check carefully and run the tool again.
    - Extent differences. Make sure the extent of the DEM is larger than that of the edges.
    - Projection mismatch: The example data have the same CRS. If you are working with your own data, ensure that all inputs (edges, waterbodies, and the DEM) are in the same projection.

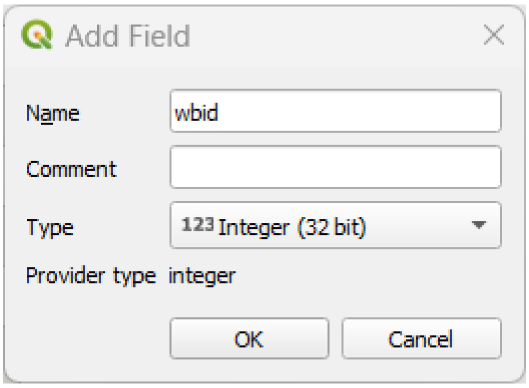
Define edge-waterbody relationships

1. Identify edges that intersect waterbodies

    a. Add a new identifier field to the waterbodies dataset

i. Open the waterbodies attribute table and Toggle editing mode on.

ii. Add a new field as shown below and click OK.



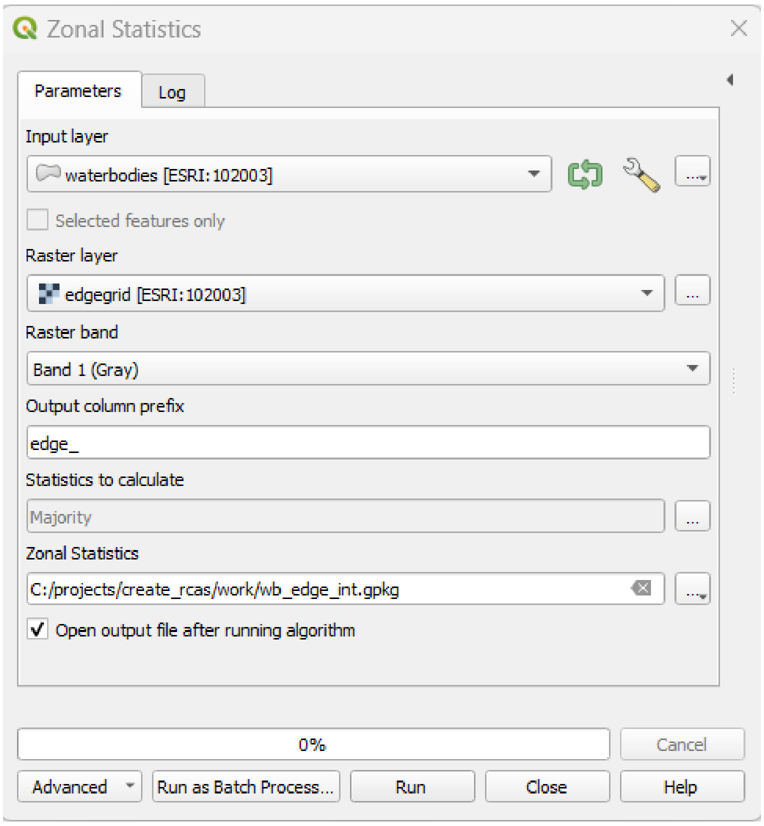
iii. Set wbid = fid and click Update All. Note that this would need to be set to @row_number if the fid column contained negative values or fid = 0.

iv. Click Save edits, Toggle editing mode off, and close the attribute table.

b. In the Processing Toolbox search for Zonal Statistics. Double click to open the tool.

c. Enter the arguments as shown below. When the Statistics to calculate argument is set to Majority, a new field is added to the output waterbodies layer named edge_majority. This will contain the reachid for the edge with the largest number of cells intersecting the waterbody feature.

*Note that in QGIS 3.36.1, there is a separate option in the 'save-output-as-file' menu for GeoPackages. When the file is saved as a GeoPackage, you will be prompted to choose a layer name.

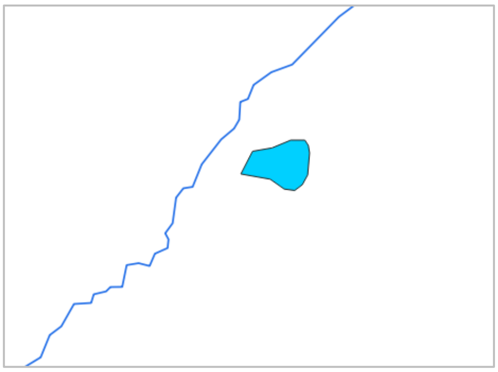


d. Click Run and close the tool when it completes successfully.

If the tool fails to run and returns an error message that the waterbodies contain invalid geometries, use the Vector Geometry Check Validity tool to identify where the topological errors are occurring and correct these, as necessary. The example dataset should not contain invalid geometries.

e. Look for NULL edge_majority values. Open the new wb_edge_int attribute table and click on the edge_majority field heading to sort the data in ascending order.

Notice that all but two of the waterbodies have been assigned an edge_majority (i.e. reachid) value. Manually select one of the two waterbodies with a NULL edge_majority value and click on the Zoom map to selected row tool, , to zoom in. Notice that the waterbody does not intersect an edge (see below), which explains why the edge_majority field is NULL. Repeat this process for each of the waterbodies where edge_majority = NULL. These waterbodies must be assigned an edge_majority value.

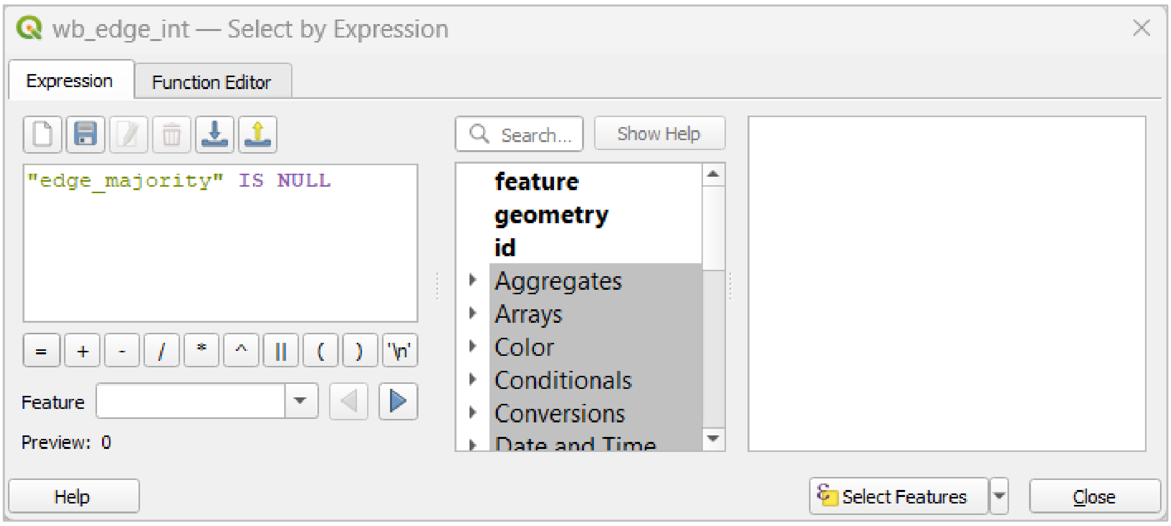


2. Replace edge_majority = NULL

   If the number of missing edge_majority values is small, as is the case here, values for the missing waterbodies can be manually added to the attribute table when Toggle Editing is turned on. However, this is impractical when many waterbodies do not intersect streams. The following steps should be taken in this case.

   a. With the wb_edge_int attribute table open, click on the Select Features with Expression button, , to open the tool.
   b. Enter the expression shown below, click Select Features, and then Close. Two features should be selected.

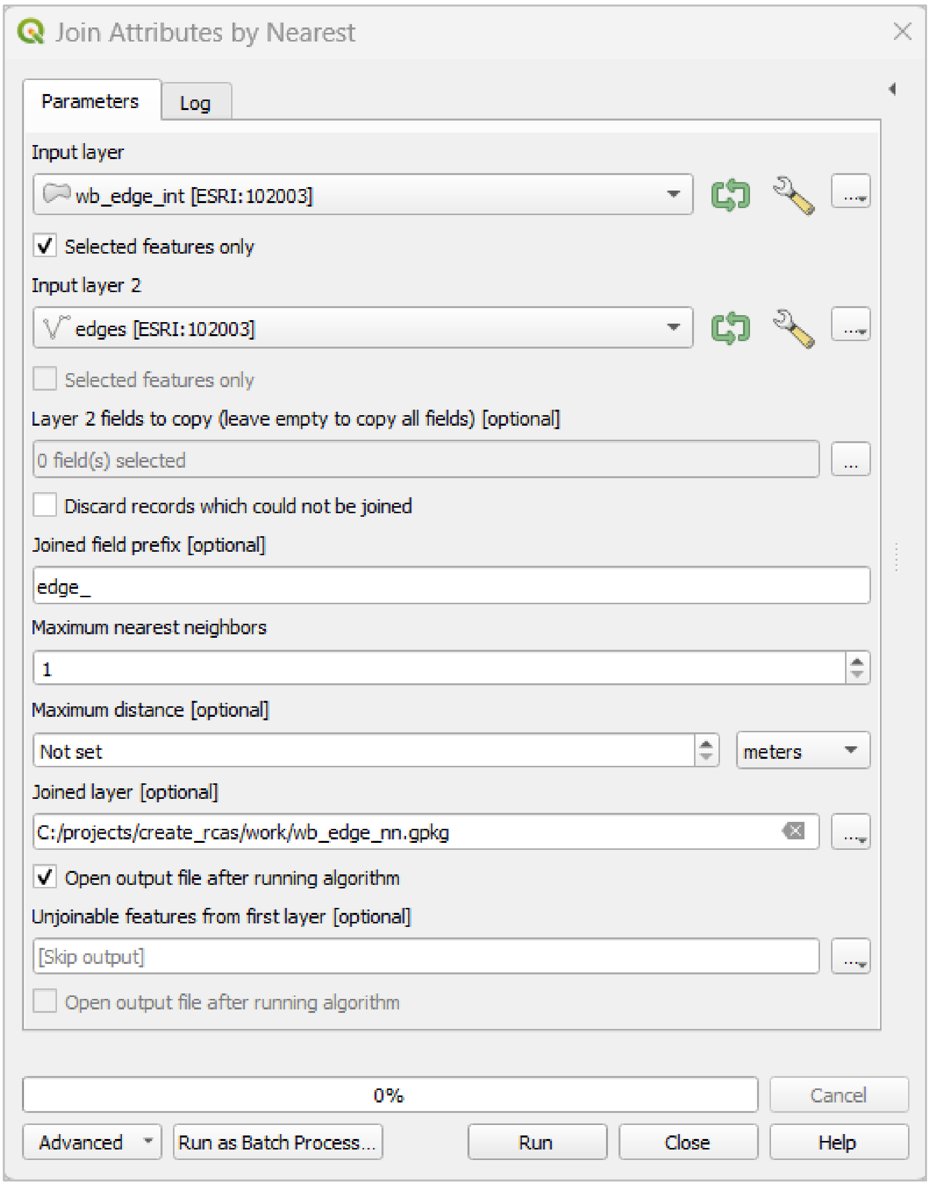
      *Note that in QGIS 3.36.1, edge_majority does not need to be in quote.



   c. Close the wb_edge_int attribute table, but do not clear the selection.

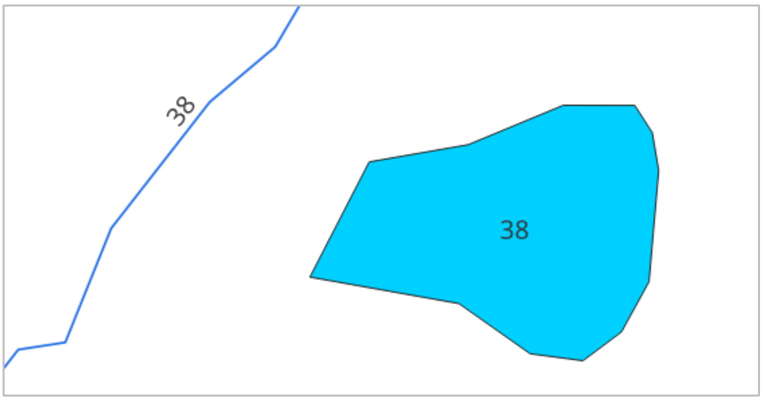
   d. In the Processing Toolbox, search for Join attributes by nearest. Double click on the result to open the tool.

   e. Input the arguments as shown below. Be sure to check the box next to Selected features only under Input layer 1. Click the Run button and then Close once the tool has successfully completed.



The output polygon layer wb_edge_nn should be added to the Layers Panel. Open the attribute table and examine the contents. The layer should only contain two records for the features selected in wb_edge_int. Notice that the table contains the original attributes from wb_edge_int, as well as the attributes for the nearest edge feature.

Select one of the records and click on  to zoom to the selected waterbody. Use the Identify tool to confirm that the closest edge reachid attribute is equal to the edge_reachid attribute in wb_edge_nn, as shown below.



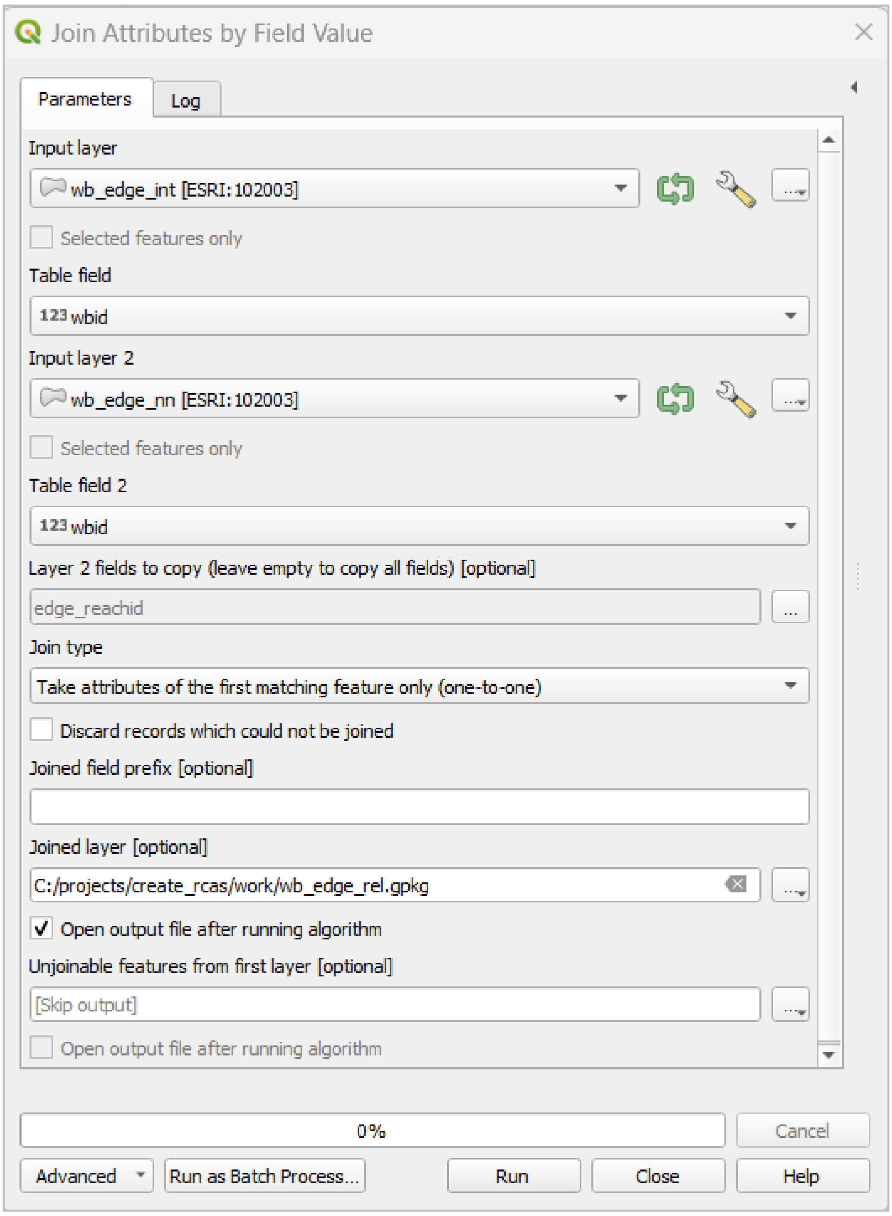
f. Populate remaining edge_majority values in wb_edge_int

   i. Click on the Deselect Selection from all Layers button, , in the Selection Toolbar to clear all the active selections.

   ii. In the Processing Toolbox, search for Join attributes by field value. Double click on the result to open the tool.

   iii. Enter the arguments shown below, click Run.

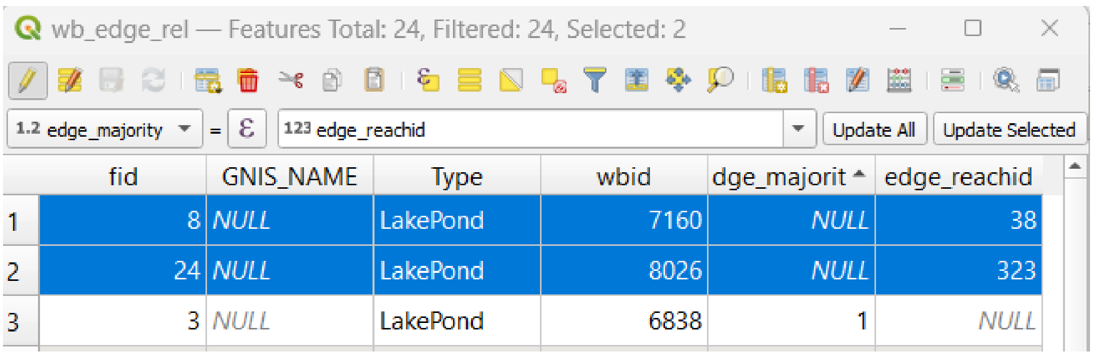
   You will receive a red warning message in the log output stating that 22 feature(s) from input layer could not be matched. Do not be concerned – this is what we expect because wb_edge_nn only has two records. Click Close.



iv. Open the wb_edge_rel table and Toggle editing mode on.

v. Select the two records where edge_majority = NULL, as we did previously.

vi. Set edge_majority = edge_reachid. Click on the <u>Update Selected button</u> NOT Update All.



vii. Clear the selection and check that all records contain an edge_majority value.

viii. Click the Delete Field button, , to open the Delete Fields tool. Select the edge_reachid field and click OK.

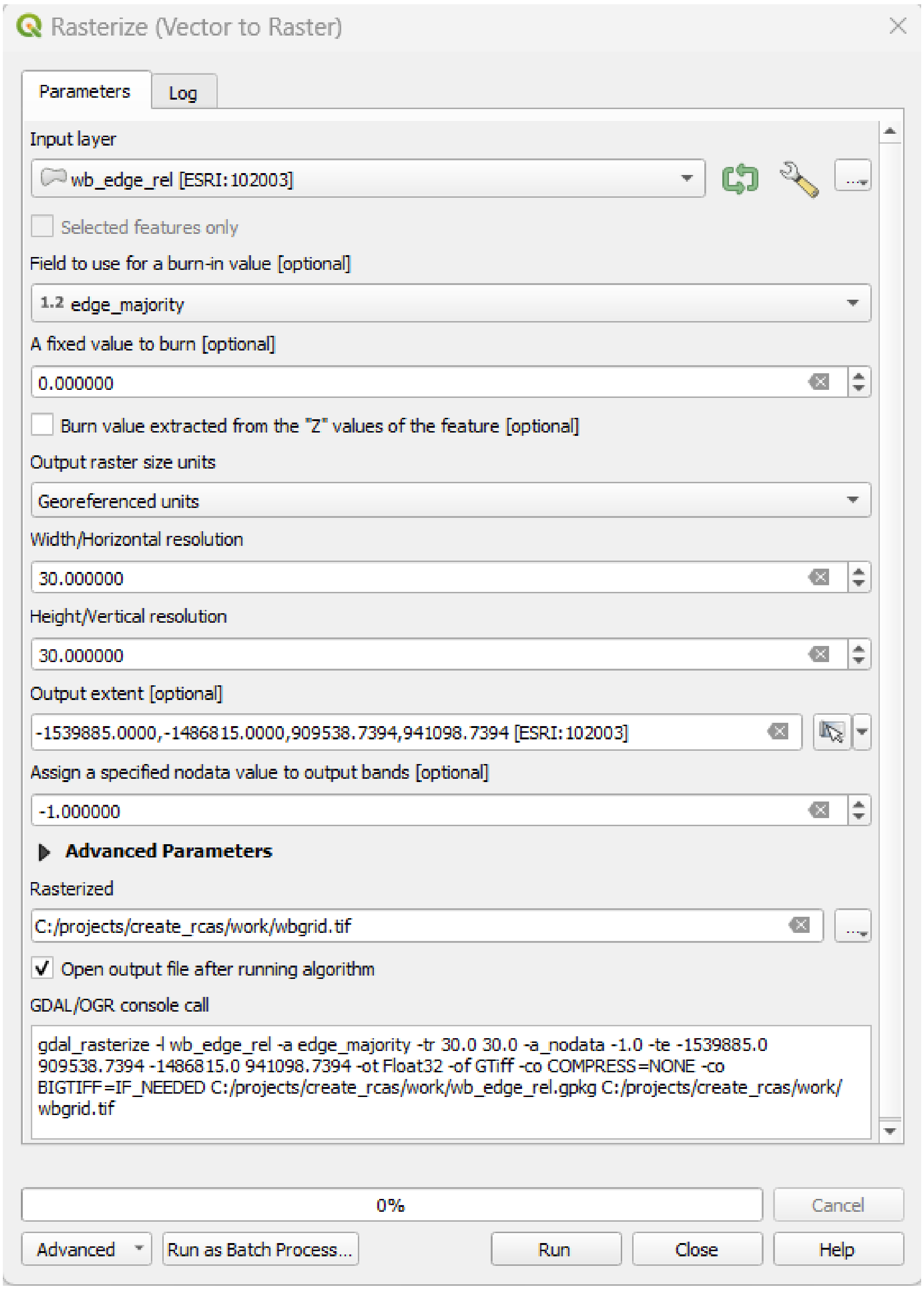
ix. Click Save edits and Toggle editing mode off. Close the wb_edge_rel table.

x. Remove wb_edge_int and wb_edge_nn from the Layers Panel. Select all three layers, right click on one of them, scroll down and select Remove Layer. Click OK to confirm in the dialog box.

3. Convert the wb_edge_rel vector layer to raster

    a. In the Processing toolbox search for and open the Rasterize (vector to raster) tool.

    b. Enter the arguments shown below. The Output extent should be calculated from the dem layer. Click Run and then Close.



c. Check the new wbgrid to ensure that cells have the same spatial resolution as the DEM and that the cells are perfectly aligned before continuing. It should have values ranging from 1 to 323.

d. Remove wb_edge_rel from the Layers Panel.

4. Merge edgegrid and wbgrid

   a. In the Processing Toolbox, search and open the GDAL tool Merge.

   b. Input the arguments as described and shown below.

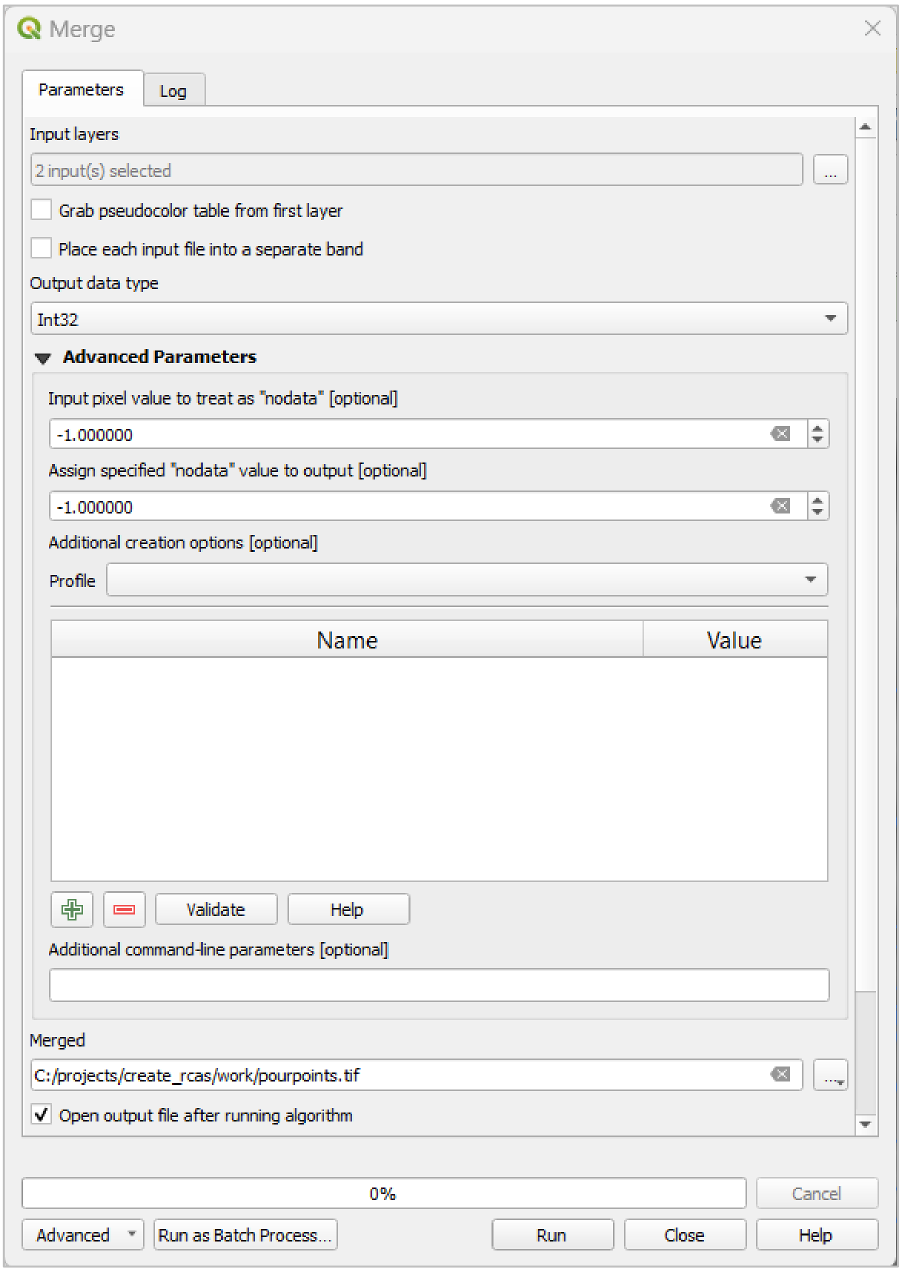
i. Input layers: Click on the more options button (three dots) to the right and check the boxes next to edgegrid and wbgrid. Click OK.

ii. Click on the arrow next to Advanced Parameters to expand the section.

iii. Enter arguments as shown below, click Run, and then Close when the tool successfully finishes.

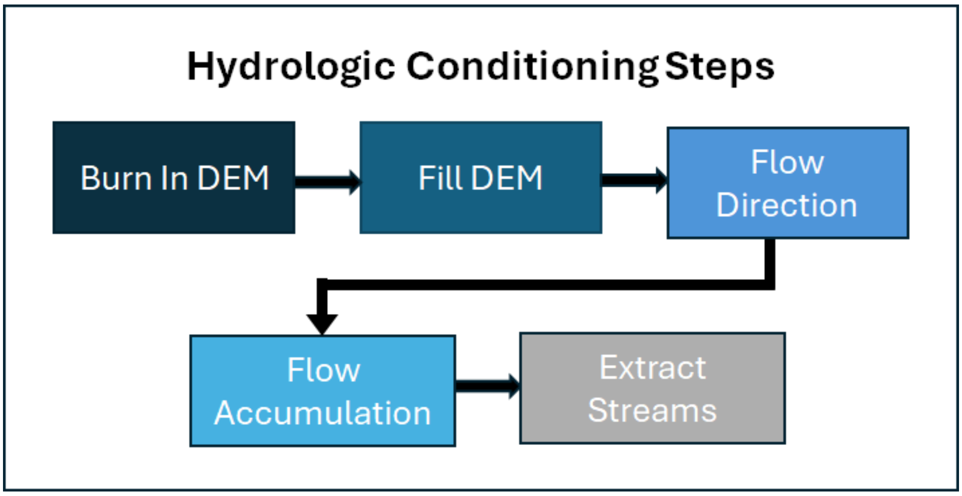
iv. Examine the new pourpoints raster layer. It should contain both the edges and waterbodies in raster format, with values equal to the associated reachid. Raster cells in other areas should contain 'no data' values. Note that in QGIS 3.36.1, the other cells may contain 0 instead of no data – this is ok.

v. Remove edgegrid and wbgrid from Layers Panel.



5. Creating a hydrologically conditioned DEM

Hydrologic conditioning of the DEM ensures that the flow paths and drainage patterns represent topography and surface flows as accurately as possible. In this section, we implement five key hydrologic conditioning steps using the WhiteboxTools plugin for QGIS. For more information about the individual tools and arguments, please see the WhiteboxTools Tool Reference found in the online User Manual (https://www.whiteboxgeo.com/manual/wbt_book/available_tools).



It is worth noting that the instructions provided here are an oversimplification of the hydrologic conditioning process for most study areas. In reality, it is often an iterative process of identifying and filling depressions (i.e. sinks), identifying anthropogenic structures (i.e. bridges, culverts, etc.) and breaching them, burning in pour points, calculating flow direction, and then examining the flow accumulation and streams derived from it to ensure the hydrologic conditioning produces the desired outputs. This is an active area of research (e.g. Ai-Ling et al. 2023) and there are many online resources and tools describing alternative hydrological conditioning approaches. However, the instructions provided here offer a foundation to build and improve on.

Burn in pour points

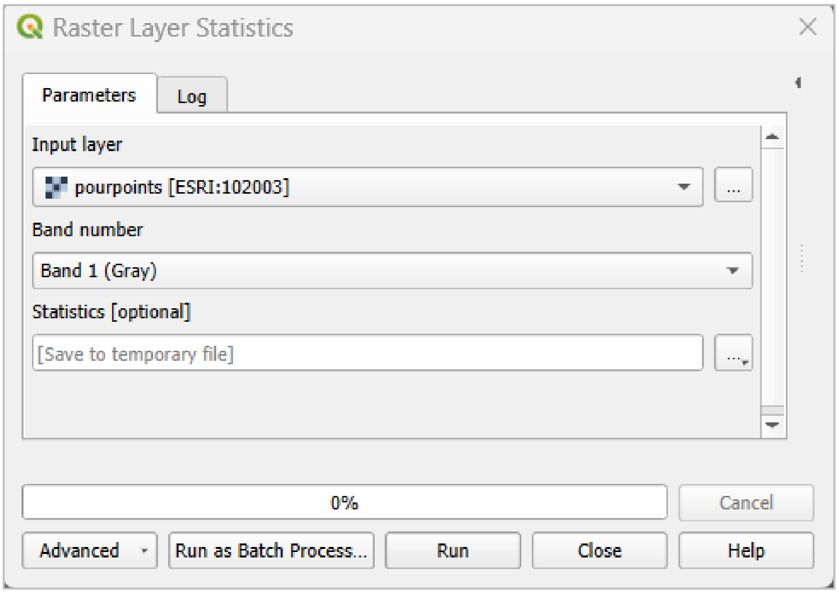
Burning the pour points (edges and waterbodies) into the DEM is a way to incorporate stream network information into the elevation data, creating a hydrologically correct surface. A constant positive value is assigned to pour point cells and then subtracted from the DEM. This artificially lowers the elevation along drainage lines and waterways, forcing flow along these paths; ensuring that flow paths accurately represent the drainage patterns of edges and waterbodies when RCAs are delineated. This step is important and particularly essential if your streams (i.e. edges) were not derived from a DEM or were derived from a different DEM than the one used to delineate RCAs in subsequent steps.

Although the process of burning in pour points is simple, the decision about the depth to burn in requires careful consideration of data accuracy and resolution, as well as the topographic variability of terrain characteristics. In this example we use a value of 10m, but it may be necessary to trial different burn in values to see which one works best.

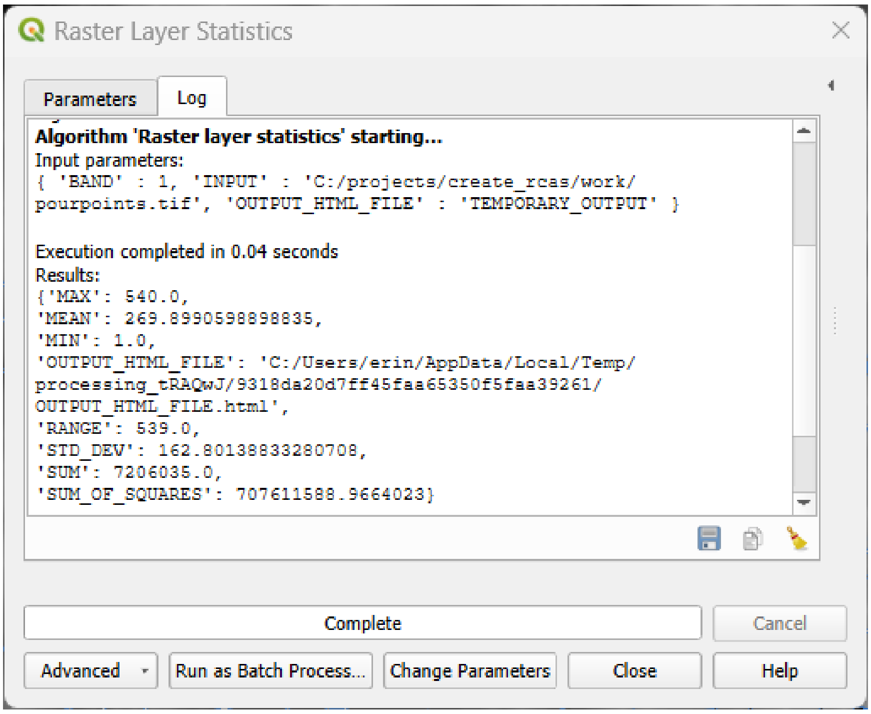
1. Create the burn in raster

a. In the Processing Toolbox, search for Raster Layer Statistics and open the tool.

b. Input the following arguments and click Run. Do not close the tool.



c. Statistics for the pourpoints raster are printed as text in the Log tab as shown below. Take note that the minimum and maximum values are 1 and 540, respectively. If your pourpoints raster contained 0's instead of no data values (QGIS version 3.36.1), the range will be 0-540. Click Close.



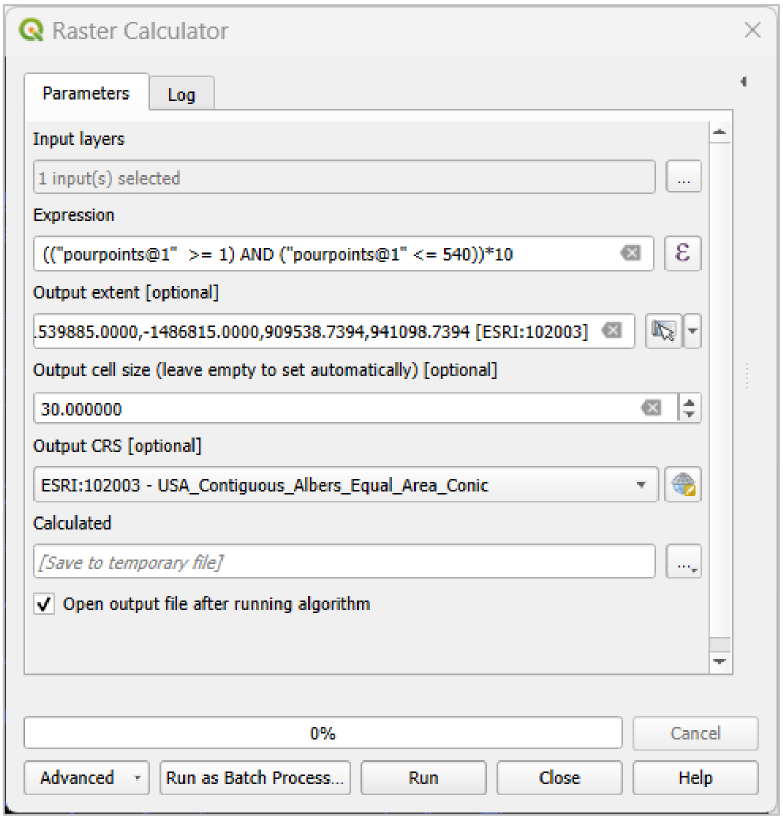
2. Open the Raster Calculator tool in the Raster analysis toolset

a. Input layers: Click on the more options button (...) and select pourpoints. Click OK.

b. Enter the other arguments shown below.

c. Output extent: Select Calculate from Layer > dem

d. Click Run and then Close.



Examine the temporary layer, Calculated. It should be a raster layer with values at pour points equal to 10 and no data everywhere else (raster will have value of 10 and 0 in QGIS 3.36.1). Note that the burn in value is always in the same units as the elevation data. In this case, it is in metres, but that may differ by DEM.

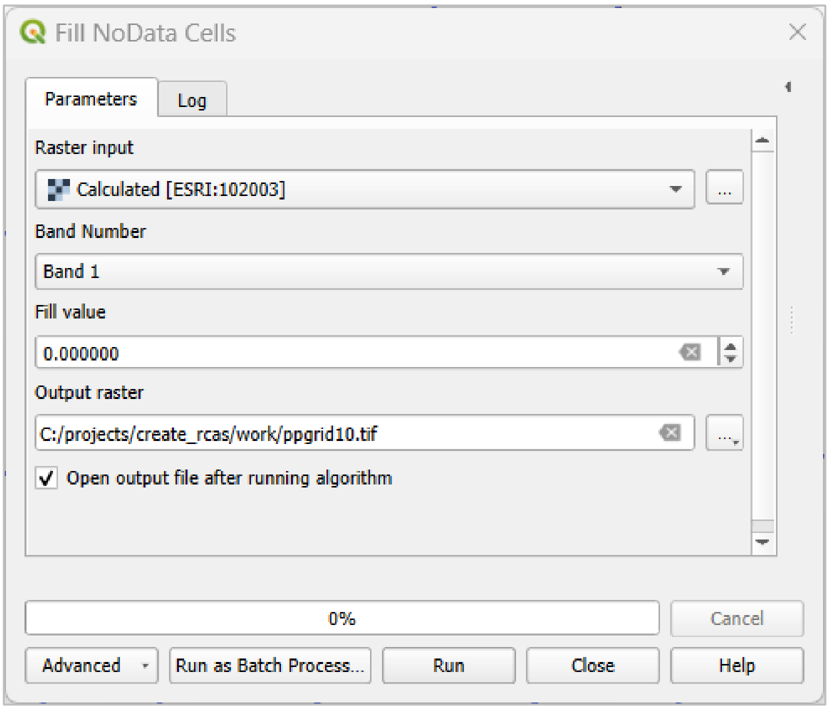
e. Replace no data values with 0

   i. In the Processing toolbox, search for and open the Fill NoData cells tool.

   ii. Enter the arguments shown below, click Run and then Close when the tool completes successfully.

   iii. Remove the temporary layer, Calculated, from the Layers Panel.

* Note that this step is not necessary if your temporary raster layer already contains values of 10 and 0 (QGIS 3.36.1) instead of 10 and no data. In that case, simply save the temporary raster layer as ppgrid10.tif and move to the next step.

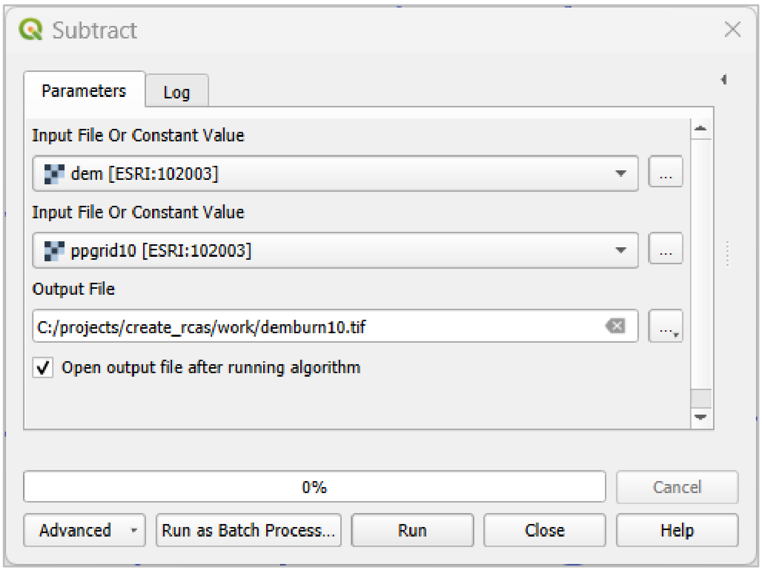


The result, ppgrid10, is a raster layer of the pour points, with raster cells representing edges and waterbodies allocated a value of 10 and all other cells a value of 0. The ppgrid10 will be used to ‘burn’ these pour points into the DEM.

3. Burn pour points into the DEM

   a. In the Processing Toolbox, open the WhiteboxTools Subtract tool.

   b. Enter the arguments below, click Run, and then Close.



   c. Use the Identify tool to examine elevation in demburn10 and dem.

      i. Zoom into an area with pourpoint values equal to 10.

      ii. Click on the Identify tool

iii. In the Layers panel, select dem, hold down the Ctrl key and select demburn10. Both layers should be selected.

iv. Use the Identify tool to compare elevation in the two DEMs

    1. Click on a pourpoint cell with a value of 10. Notice that the elevation in the Identify Results window is 10m lower in demburn10 than dem. Click on  to expand new results by default.

    2. Click on a cell with a pour point value equal to 0. The elevation values should be the same. If so, the pourpoints have been successfully burned into the DEM.

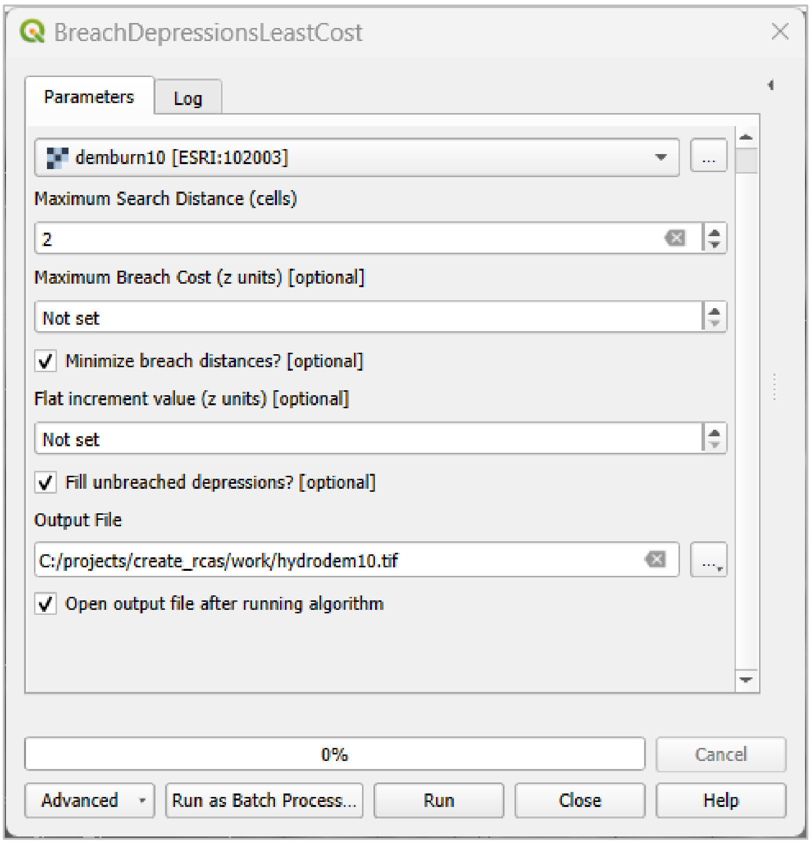
    3. Remove ppgrid10 from the Layers Panel.

Fill DEM

Sinks and peaks in the DEM may be errors caused by artifacts in the data, data resolution, or rounding of elevations. These sinks must be filled to ensure proper delineation of RCAs. The WhiteboxTools Hydrological Analysis toolset includes several tool options for filling depressions including FillDepressions, FillDepressionsPlanchonAndDarboux, FillDepressionsWangAndLui, FillSingleCellPits, as well as BreachDepressions, BreachDepressionsLeastCost, BreachSingleCellPits, and StochasticDepressionAnalysis. It is worth taking the time to read about the options in the WhiteboxTools User manual before deciding which approach best suits the characteristics of your dataset and topography of the study area. We use the BreachDepressionsLeastCost tool in this example.

1. In the Processing Toolbox, search for BreachDepressionsLeastCost and open the tool.

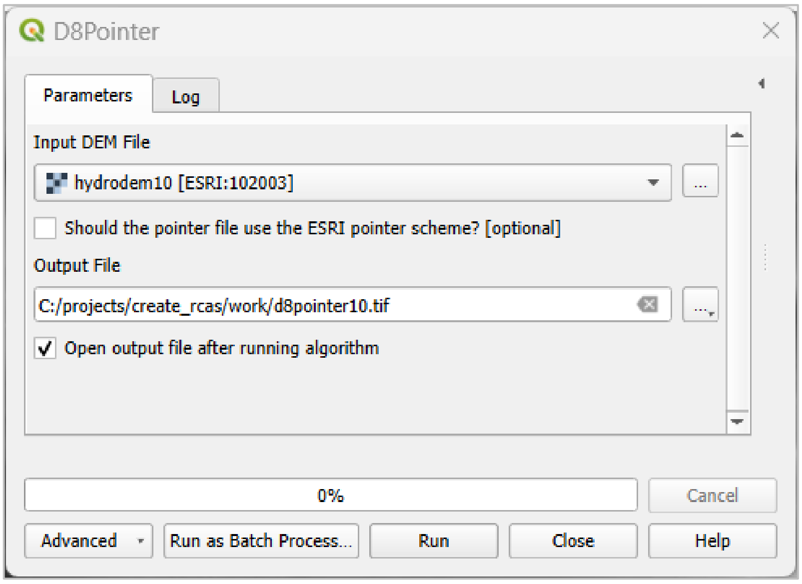
2. Enter the arguments shown below. Note that the Flat increment value is not set because we are using the default settings. Click Run and then Close when the program finishes successfully. The raster hydrodem10 is added to the Layers Panel.



D8 Flow Direction

A D8 rather than multidirectional flow direction algorithm must be used here because it is required by the tool used for RCA delineation.

1. Open the WhiteboxTools D8Pointer tool found in the Hydrological Analysis toolset. Flow direction rasters are referred to as D8 pointers in Whitebox tools.
2. Enter the arguments below, click Run, and then Close.



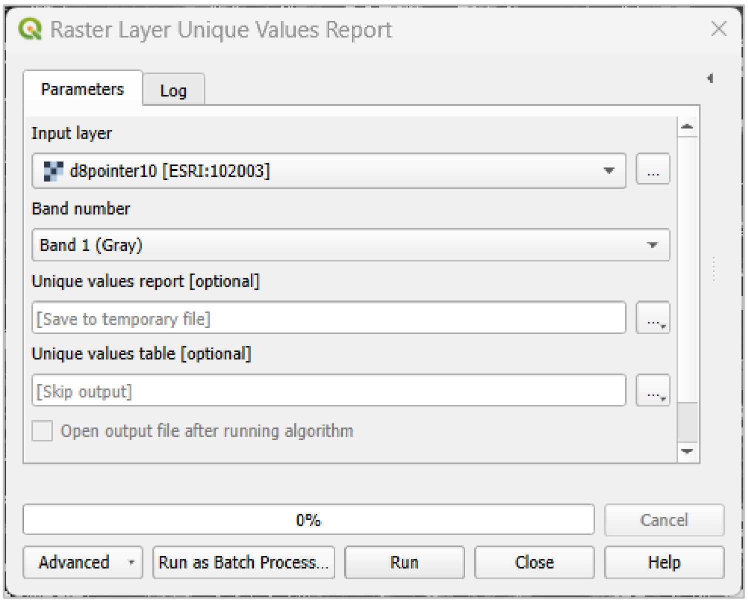
This tool produces a D8 flow direction raster, which is added to the Layers Panel. By default, WhiteboxTools D8 flow pointers use a different numeric index convention to represent the direction of flow out of a cell than ESRI flow direction rasters.

64	128	1
32	0	2
16	8	4

The D8 flow pointer should *only* contain these values. If a raster cell is assigned a value of 0, it means that it has no downslope neighbours.

3. Check that values in d8pointer10 are valid

    a. Open the Raster Layer Unique Values Report tool, enter the arguments shown below, click Run and then Close.



    b. View results

        i. The Results Viewer window should be open on the right side of the screen, with a link to an OUTPUT_HTML_FILE.html containing the results. Click on this link to open the results in your browser.

        The results show that the flow direction raster (D8Pointer) only contains valid values for the example dataset. If invalid values are present in your own data, you may want to try filling the hydrodem raster a second time. Notice that there are a small number of values equal to 0. These are not necessarily errors if they occur in natural depressions (e.g. lakes or reservoirs) or along the edges of the DEM. Close the browser window and return to QGIS.

c. Examine values of 0 in d8pointer10.

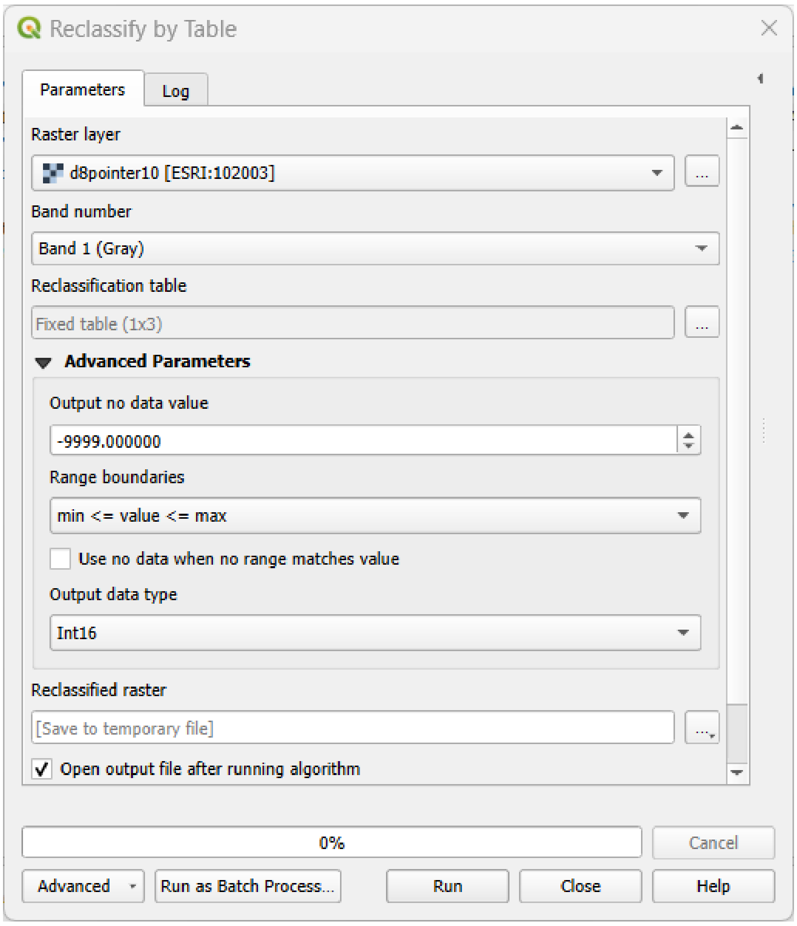
   i. Reclassify D8pointer10 to make 0 values more visible

      1. Open the Reclassify by table tool

      2. Set the argument values shown below.

      3. Reclassification table: Click the more options button, click Add Row, and set the Minimum, Maximum, and Value columns to 1, 128, and 1, respectively.

      4. Click Run and then Close.



   ii. Change the symbology of the temporary Reclassified raster

      1. Open the Properties for Reclassified raster and click on Symbology. Change the arguments in this order:

         a. Render type: Change to Singleband pseudocolor

b. Interpolation: Change to Exact

c. Mode: Change to Equal Interval

d. Classes: Set to 2.

e. If necessary, change the colours associated with each value by double clicking on the coloured box. They should look something like this, depending on your colour choices.

Value =	Color	Label
0		0
1		1

f. Click Ok.

iii. Identify where the 0 values are

1. Zoom into the boundary of the Reclassified raster until you can see the dark 0 cells, as shown below.

Pan around the raster and notice that all the depressions (0 values) are along the edges. This is ok and we can move onto the next processing step. If there were depressions in other areas that were not associated with natural features then it would be necessary to revisit the previous steps and refill, or reburn and refill the DEM.

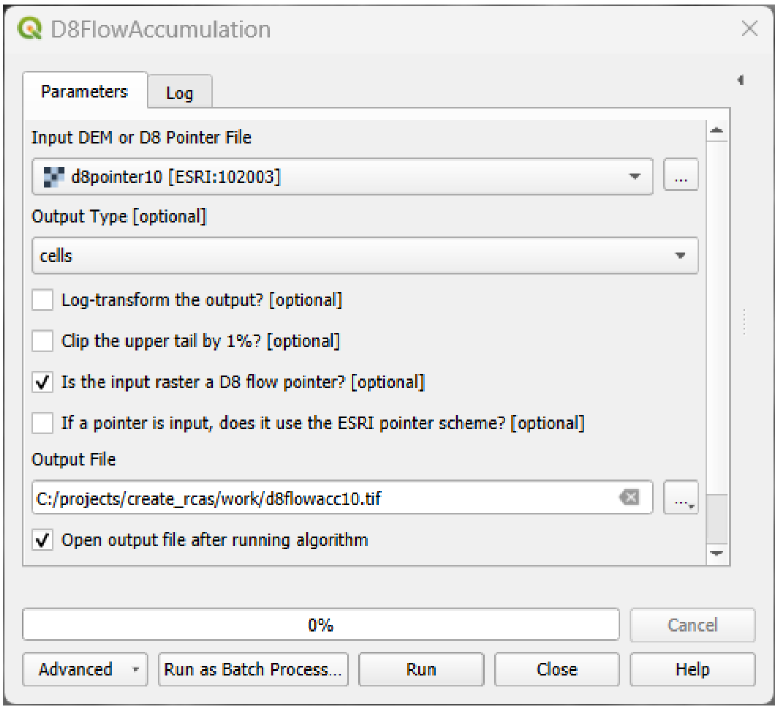
iv. Remove the Reclassified raster from the Layers Panel.

Flow Accumulation and Stream Delineation

Calculating the flow accumulation at this stage is not technically needed to calculate the RCAs. However, it is helpful to visually assess how closely flow paths are following the pour points.

1. Calculate Flow Accumulation

    a. Open the WhiteboxTools D8FlowAccumulation tool. Enter the arguments shown below, click Run, and then Close.



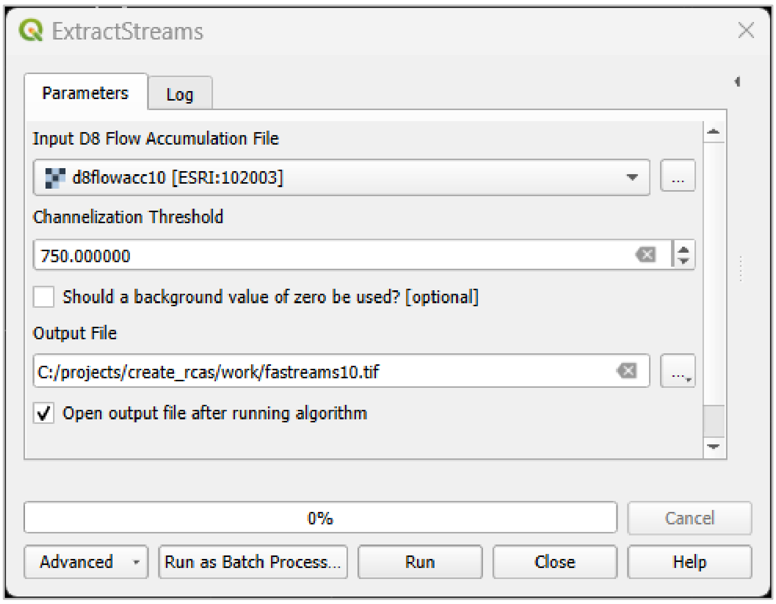
The flow paths can be difficult to see in the flow accumulation raster and so we will extract a set of streams to make them easier to see.

2. Extract streams

    a. Open the WhiteboxTools ExtractStreams tool in the Stream Network Analysis toolset.

    b. Enter the arguments below. The Channelization Threshold is the minimum flow accumulation value (number of cells flowing in) needed to delineate a stream. This value is subjective, and you may need to try different values when evaluating your own data.

    c. Click Run and then Close.



3. Examine the similarities and differences between fastreams10 versus the edges and waterbodies. You may need to turn off other layers in the Layers panel and/or rearrange the layers to make the features more visible. Zoom in and out to inspect different areas of the network.

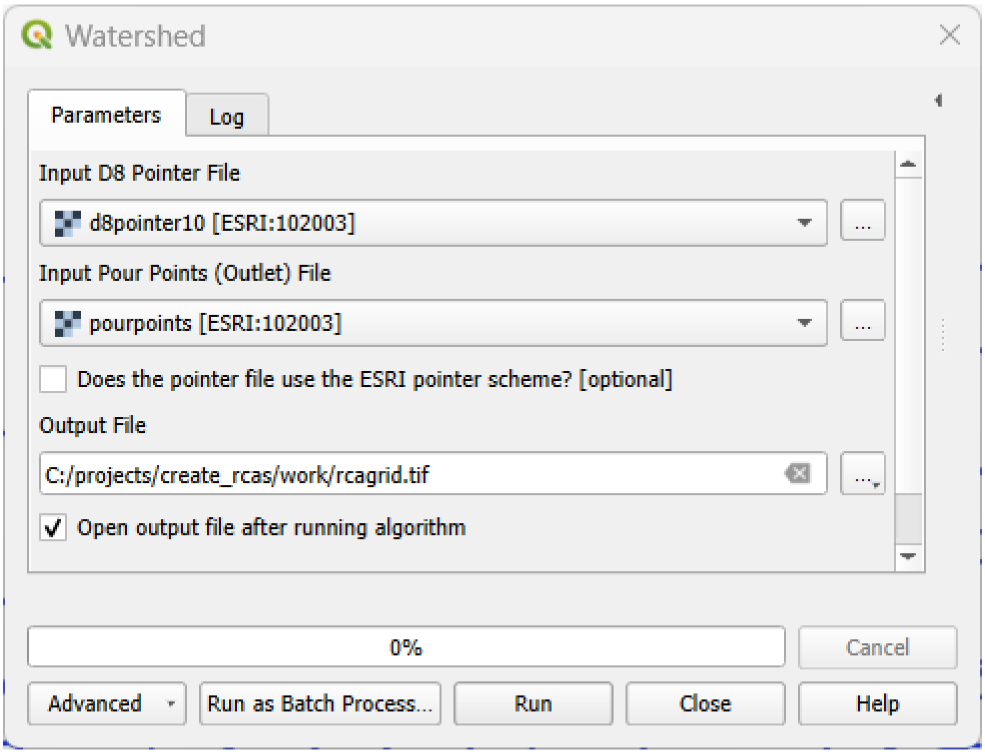
   You'll notice that the flow accumulation path follows the edges closely, but not perfectly. There are often deviations in areas where the streams are particularly sinuous or near confluences. Relatively small deviations such as these are to be expected, given the spatial resolution of the DEM. As expected, flow paths in fastreams10 often deviate from the edges within waterbodies; this is nothing to be concerned about. If you see large deviations from the edges, it could mean that the burn in value was too small and you should try a larger value. If there are relatively large flow paths that suddenly cease, then they have likely encountered a sink that was not filled properly. In that case, the hydrodem10 could be filled again more aggressively or the dem refilled using another depression filling algorithm.

6. Generate RCAs

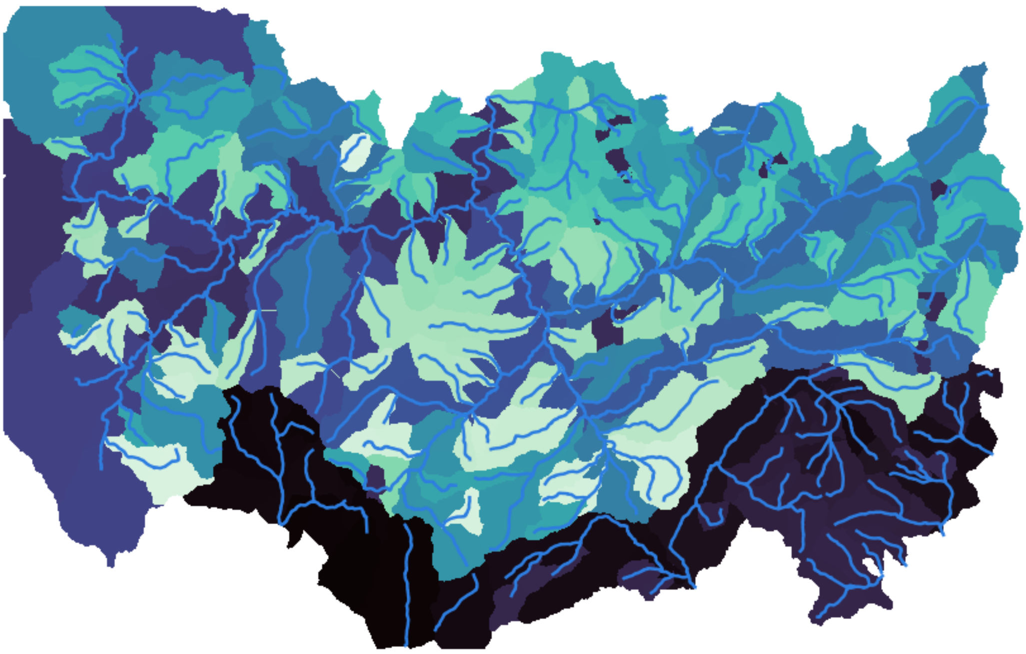
Once you are satisfied with the hydrologic conditioning results, it is time to generate the RCAs.

1. Open the WhiteboxTools Watershed tool found in the Hydrological Analysis toolset.

2. Enter the arguments shown below, click Run and then Close.



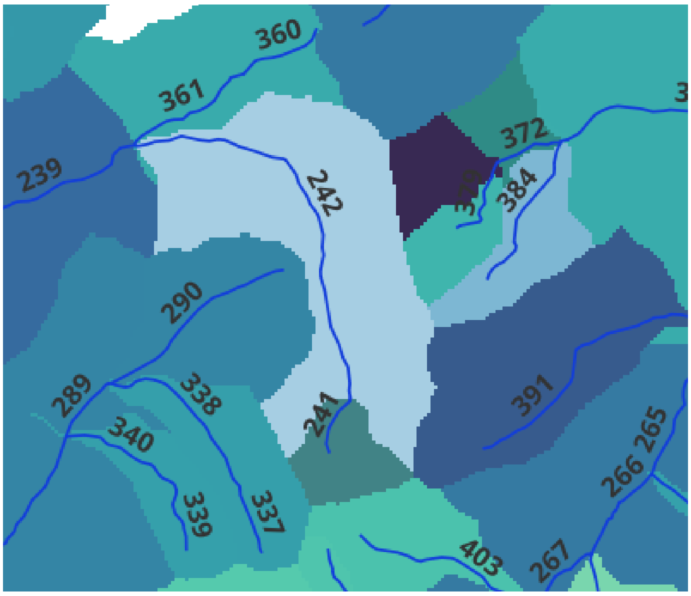
The rcagrid should look something like the image below when the edges are overlaid.



Zoom in and use the Identify tool to query the edges and the rcagrid at the same time. Notice that the rcagrid value in Band 1 is equivalent to the edges reachid. This attribute relationship enables each RCA to be directly linked to a single edge. Although there is generally a 1:1 relationship between edges and RCAs, this is not always the case. For example, it is not uncommon to find that a relatively small number of edges have not been assigned an RCA. This may occur if the length of the edge is short in relation to the spatial resolution of the DEM (e.g. stream reach length = 10m and the DEM spatial resolution = 30m). When this situation occurs, the RCA is essentially too small to delineate. When a group of edges is part of a single waterbody, such as a lake or reservoir, some of the edges will not be assigned RCAs.

Inspect the rcagrid boundaries to determine whether they adequately represent the topography before moving on to the next step. Turning the DEM layer on may help with this.

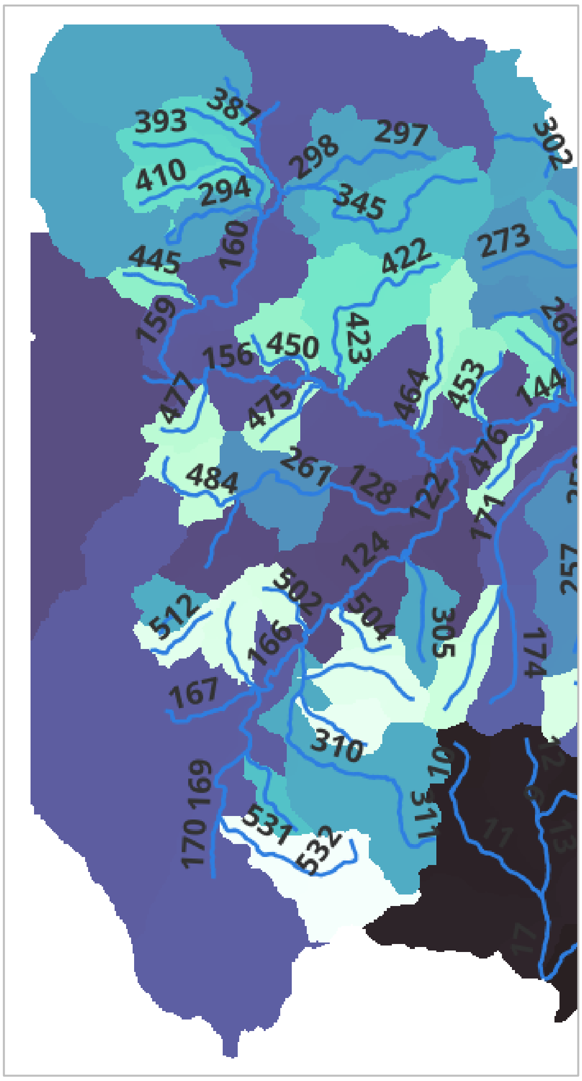
Zoom into the top right corner of the DEM, where additional edges outside of the study area were included (see below). In this case the RCA boundaries for the study area adhere to the ridgelines closely and provide a relatively accurate representation of drainage area.



*Black labels are edge reachids

Now zoom into the left portion of the DEM. Notice that the RCA boundaries for edge reachids are large and extend to the boundary of the DEM (as shown below). This occurred because there were no extra edge features located outside the boundary of the study area here. When the RCAs were delineated, the algorithm breached the ridgeline and continued until it reached the boundary of the DEM.

When this occurs, go back, and recreate the streams dataset, adding more edges to outside of the study area, and then repeat the workflow. Given that this is an example, we will continue with the remaining steps.



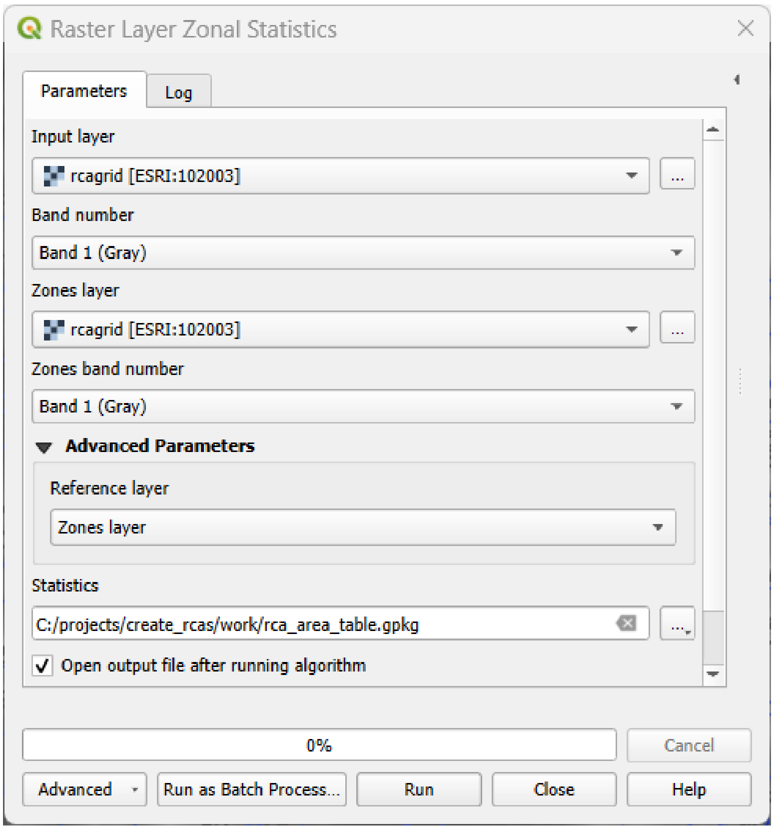
7. Calculate RCA area for each edge

1. Calculate area

   Here we use zonal statistics to calculate the area for each RCA. Although tools exist to convert the rcagrid to a polygon layer before generating the areas, it is safer to calculate area from the raster layer. Some of the raster to vector conversion tools in QGIS will create duplicate polygons in the conversion process, which can inflate area estimates when the area is summed.

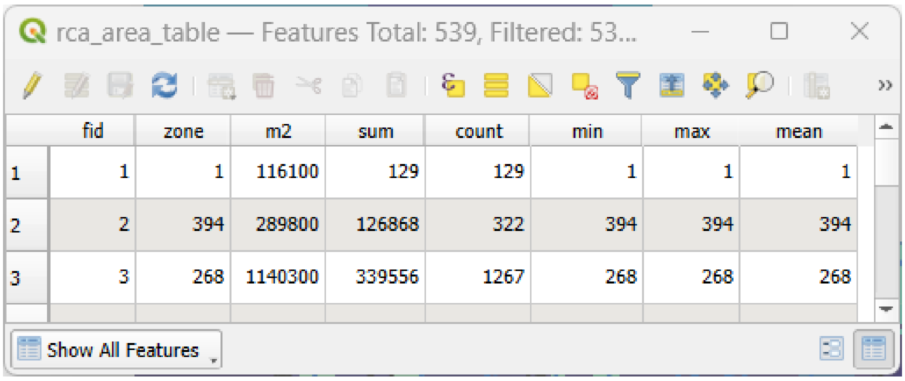
   a. Open the Raster Layer Zonal Statistics tool and enter the arguments described and shown below.

      Notice that rcagrid is used for both the Input layer and the Zone layer. This is a special case because we want to calculate statistics about the RCA itself. To calculate statistics about another raster layer in each RCA, such as Urban land use, the Input layer would be Urban, with rcagrid as the Zone layer.



b. Click Run and then Close.

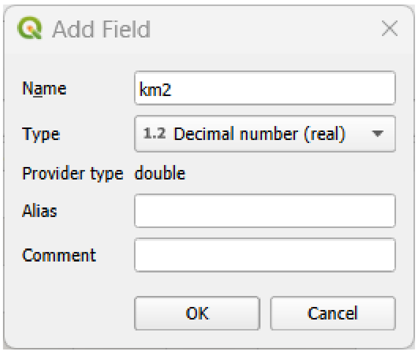
c. Open the rca_area_table in the Layers Panel. The zone column represents the rcagrid cell values, which are equivalent to the reachid values in the pourpoints dataset used as an input in the Watershed tool. The m2 column represents the area of each RCA in m2.



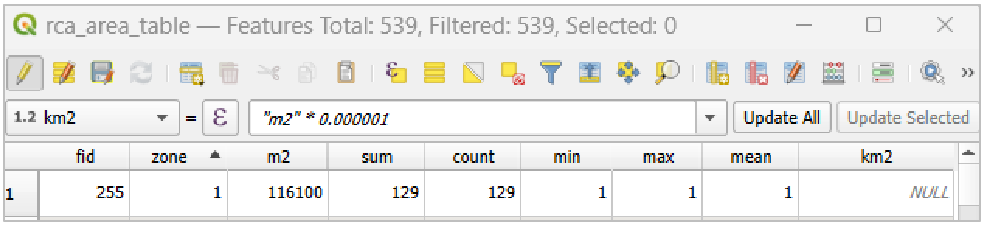
d. Convert area to square kilometres for convenience

  i. In the rca_area_table, Toggle Editing mode on.

  ii. Click the Add Field button, enter the arguments shown below, and click OK.



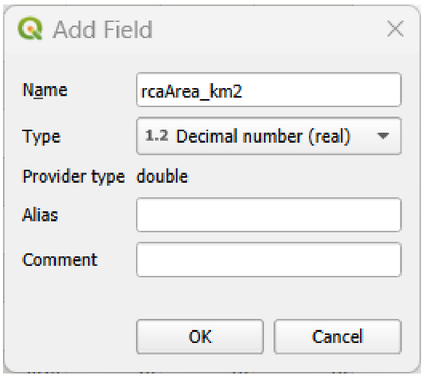
iii. In the Expression Editor at the top of the table, select the km2 column and enter the formula shown below. Then click the Update All button.



iv. Save the edits, Toggle Editing mode off, and close the rca_area_table.

2. Add the RCA area field to the edges attribute table

   a. Open the edges attribute table, Toggle Editing mode on, and add a new column named rcaArea_km2 as shown below. Click OK.



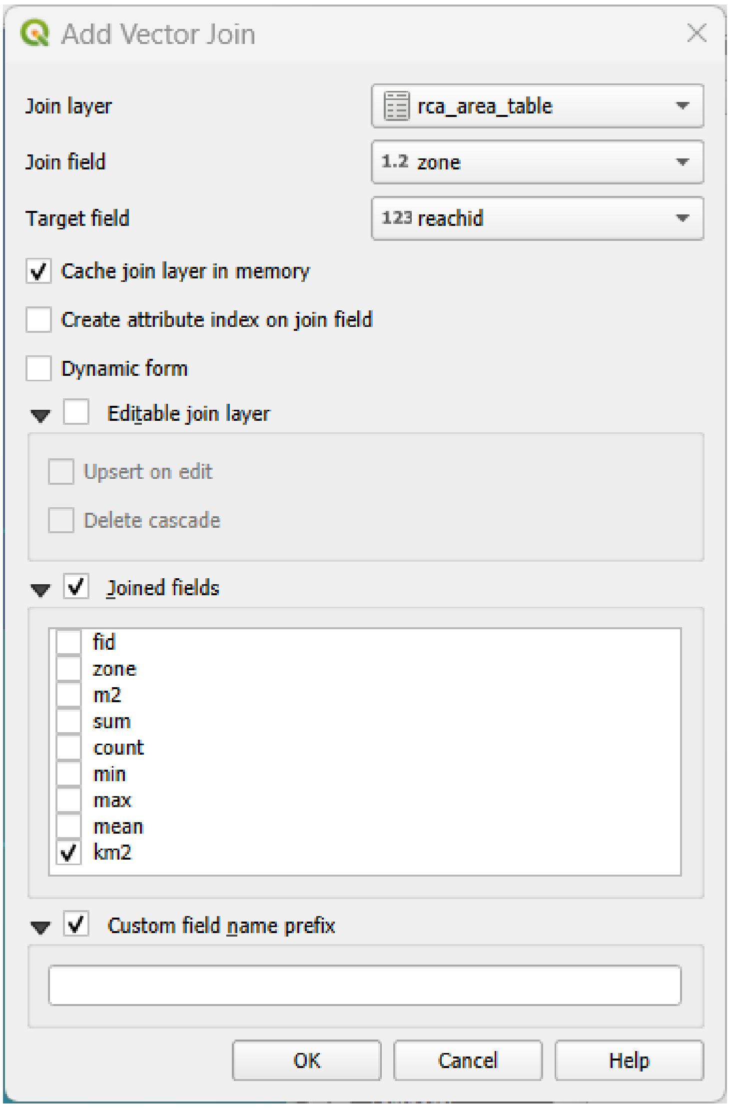
   b. Save the edits and Toggle Editing mode off.

3. Join edges and rca_area_table

   a. Open the edges layer Properties and select Joins on the left panel

   b. Click the  button in the lower left corner to Open the Add Vector Join Window.

   c. Enter the arguments shown below and click OK to close the window.

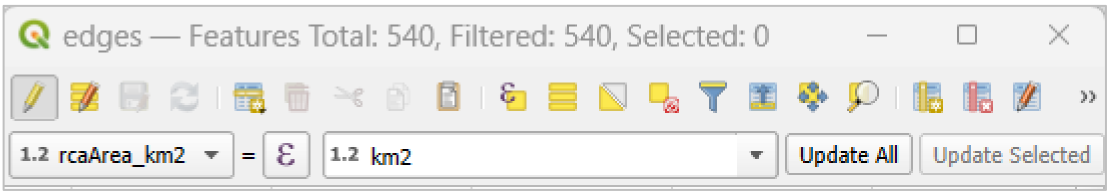


d. Click OK in the Layer Properties window.

e. Return to the edges attribute table. There should be a new column named km2 along with the empty rcaArea_km2 column created previously.

f. Toggle Editing mode on. Set the expression below and click Update All.

   Note: If rcaArea_km2 is not available in the dropdown menu, Toggle Editing mode off, and close the attribute table. Save the Project. Then reopen the edges attribute table, Toggle Editing mode on, and try again. If the attribute still isn't present, try shutting down and reopening the project before trying again.



g. Save edits and Toggle Editing mode off.

h. Remove the join

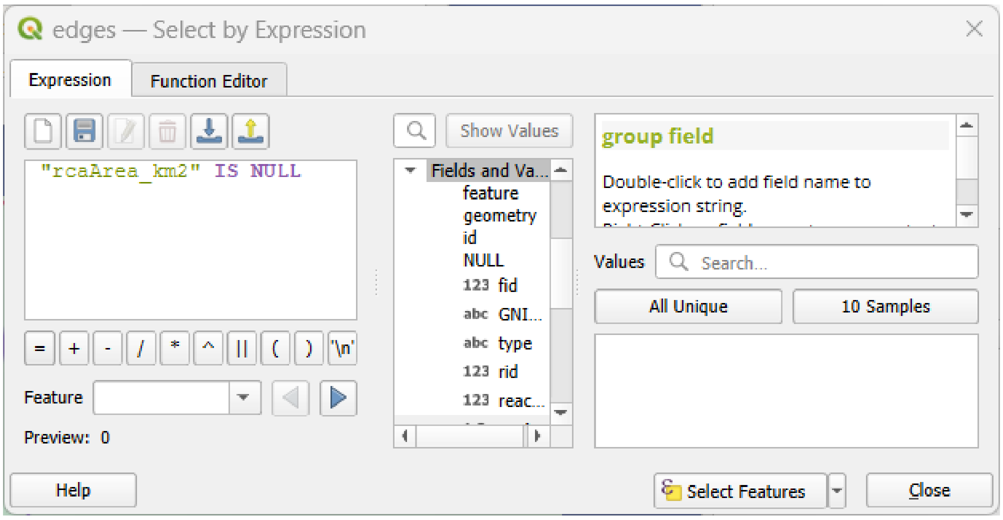
i. Open the edges Layer Properties and click on Joins

ii. Select Join layer rca_area_table and click on the  button on the lower left corner to remove the join. Click OK.

4. Find NULL rcaArea_km2 values
When the rcaArea_km2 value for an edge feature is NULL, it means that there was no rcagrid zone created for the edge.

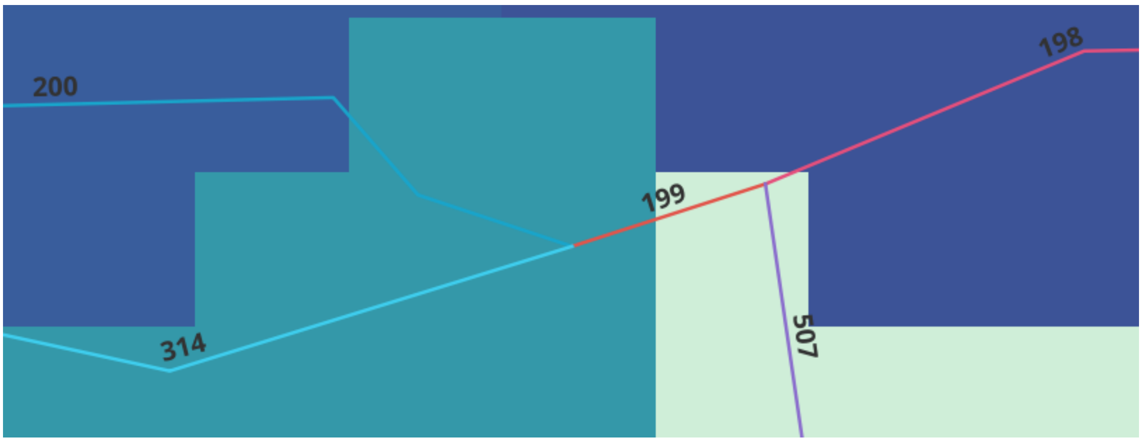
a. Return to the edges attribute table and click on the Select Features using an expression button 

b. Enter the following expression and click Select Features and then Close.



Notice that one feature is selected in the edges attribute table. Click on the  button to zoom to the selected feature, which has a reachid = 199.

Although edge 199 is longer than the spatial resolution of the DEM cells (30m), it is still relatively short and is located between two confluences. As a result, the two raster cells intersecting the edge are assigned to RCAs 314 and 507.



NULL values can create issues if RCA attributes are summed downstream to create watershed attributes using the SSNbler function accum_edges. Therefore, we need to replace the NULL rcaArea_km2 value with 0 in the edges attribute table.

5. Replace NULL value with 0
    a. Return to the edges attribute table and ensure that one feature is still selected.
    b. Ensure that Toggle Editing mode is on and enter the expression shown below. Click the Update Selected button. Notice that NULL values have now been set to 0.
    c. Save the edits and Toggle Editing mode off.
    d. Clear the selection and close the edges attribute table.

8. Wrap up

Congratulations! You have successfully generated an RCA raster and an RCA area value for each edge. These updates to edges are saved to the original edges.gpkg when you save edits and so the new data will be available for use in SSNbler.

The RCA raster can be used to do more than calculate RCA area. Other variables (e.g. land use area or number of point sources) and statistics (e.g. minimum, median, maximum, etc.) can also be generated for an RCA. Once these variables have been transferred to the edges attribute table, the SSNbler function accum_edges can be used to accumulate sums (e.g. counts and totals) downstream. This produces a watershed-scale variable for the downstream node of each edge feature. In addition, any of the variables residing in the edges attribute table can be transferred to observed and prediction sites in the LSN using the rid column as the join field. This provides users with the flexibility to calculate their own variables, which represent RCA and watershed characteristics that have a strong conceptual relationship with their response variable(s) in spatial statistical stream network models.

9. References

Ai-Ling J., Hsu K., Sanders B.F., and Soroosh Sorooshian S. (2023) Topographic hydro-conditioning to resolve surface depression storage and ponding in a fully distributed hydrologic model. Advances in Water Resources, 176, 104449. https://www.sciencedirect.com/science/article/pii/S0309170823000842

Bureau of Meteorology (2012). Australian Hydrological Geospatial Fabric (Geofabric) Product Guide. Australian Government, Bureau of Meteorology. http://www.bom.gov.au/water/geofabric/documentation.shtml

Frieden J.C., Peterson E.E., Webb J.A., and Negus P.M. (2014) Improving the predictive power of spatial statistical models of stream macroinvertebrates using weighted autocovariance functions. Environmental Modelling and Software, 60: 320-330.

Horizon Systems Corporation (2007) National Hydrography Dataset Plus: Documentation.

http://www.horizon-systems.com/NHDPlus/NHDPlusV2_documentation.php

Lindsay, JB. 2014. The Whitebox Geospatial Analysis Tools project and open-access GIS. Proceedings of the GIS Research UK 22nd Annual Conference, The University of Glasgow, 16-18 April, DOI: 10.13140/RG.2.1.1010.8962.

Moore R.B., McKay L.D., Rea A., Bondelid T.R., Price C.V., Dewald T.G., Johnston C.M. (2019) User guide for the national hydrography dataset plus (NHDPlus) high resolution. U.S. Geological Survey Open-File Report 2019–1096, 66 p. https://pubs.usgs.gov/publication/ofr20191096

QGIS Development Team (2024) QGIS Geographic Information System. Open-Source Geospatial Foundation Project. Version 3.28.7. Available at: https://qgis.org

R Core Team (2023) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org

Shreve, R. L. (1967), Infinite Topographically Random Channel Networks, Journal of Geology, 75, 178–186.

Stein JL, Hutchinson MF, Stein JA (2012). National Catchment and Stream Attributes Database Version 1.1.5. https://ecat.ga.gov.au/geonetwork/srv/eng/catalog.search#/metadata/73045

Strahler, A. N. (1957). Quantitative Analysis of Watershed Geomorphology. Geological Society of America Bulletin, Vol. 63, No. 11, 923-938.

Wieczorek M. and LaMotte A.E. (2010). Attributes for NHDPlus Catchments (Version 1.1): Basin characteristics, 2002-2010. U.S. Geological Survey. https://catalog.data.gov/dataset/attributes-for-nhdplus-catchments-version-1-1-for-the-conterminous-united-states-average-s