

## HYDROLOGICAL INFORMATION MANAGEMENT SYSTEMS FOR WATER RESOURCE MODELLING

---

This section focuses on Geospatial Data for hydrological modelling. The learners will be able to:

- i. Understand the key geospatial data sets and fundamental data structures that applies to hydrological modeling
- ii. Learn about the available open sources of data for hydrological modelling and information management systems
- iii. Learn how to prepare and manipulate key available geospatial data for hydrological modelling

### 1. INTRODUCTION

There is a growing concern on sustainable water use due to competing high demands that results to diversion of large water volumes from natural sources without releasing enough for ecological conservation. Unsustainable water abstraction has led to considerable negative impacts such as stream flow regime alteration, risks of downstream flooding including the entire ecosystem shift. An appropriate way to find solution to the existing water resource problems is to adopt a single water management system built up by a river basin in an integrated manner (Peel and Blöschl, 2011). Water resource modelling provides viable opportunities to understand the functioning of the water resource systems on a catchment scale. Water managers are able to predict and forecast quality and quantity of water in the wake of increasing anthropogenic interferences and thus make proper decisions on water allocation. A fundamental characteristic of hydrological modelling is the representation of the spatial variability of a watershed in terms of elevation and topography, landcover, soils and the derivative parameters that govern key hydrological processes such as infiltration, evapotranspiration and runoff. Such open source geospatial data that can be accessed for model preparation and set-up exists in the online domain (Korres and Schneider, 2017). Therefore, representing the key watershed physical characteristics as well as meteorological parameters such as precipitation can be achieved using freely available GIS datasets. Each GIS data possess a structure that has a unique implication on the hydrological modelling process. The key data structures that are fundamentally vital with regard to hydrological modelling are *vector* and *raster*. Vector structure store spatial data and is comprises of polygons, lines and points that are defined by the start and end points, which meet at nodes. Raster data on the other hand is meant to display continuous information across an area with single values represented in the form of grid cells. In this tutorial we seek to address geospatial data structures, open/free sources of data for hydrological applications and manipulation and preparation of data (Projection, Scale, dimensionality etc.) for hydrological modelling purposes.

### 2. GEOSPATIAL DATA AND DATA STRUCTURES FOR HYDROLOGICAL MODELING

#### 2.1. GIS DATA TYPES

##### a) Topographic Data

Topographic data refers to information about elevation including the shape and features of the land surface. The most important GIS capabilities for hydrologic applications is the description of

topography for the catchment/watershed of interest. Topography is best represented by the Digital Elevation Models (DEMs) which are raster illustration of the continuous surface with a referencing of the earth's surface. DEMs as basic data and GIS software are used deriving the topologic data vital for describing physical properties of the watershed. An example of a hydrologic topology includes the delineated lines and points describing the stream network and watershed/catchment boundary.

#### b) Topologic Data

Topologic data provide information on location and relationships between spatial objects. topographic data is also part of the overall classification of topologic data, there are other attributes of hydrology that do not related to the elevation. The most common attributes include but not limited to catchment boundaries, slope, flow lengths, landcover, soil, landcover among others. Some of the hydrological topological features are dependent on topography. The attributes are critical in describing the ability of a specific catchment to effectively transmit and store water. Topographic data is used to derived some the attributes in GIS. For instance, watershed boundary can be derived from a DEM given the point of outlet. Most of the topographically derived attributes are used in determining other using attributes of a watershed such as time of concentration, flow potential energies and flow attenuations.

## 2.2.DATA STRUCTURES

#### a) Raster or Grid-Based Data Structure

Raster data is an abstract of the real world where spatial data is represented as a matrix of cells. The data is pixelated (gridded) where each individual pixel is associated with specific geographical location. The pixel values can either be categorical (land use) or continuous (elevation). The cells in a raster should be of the same size, determining the resolution.

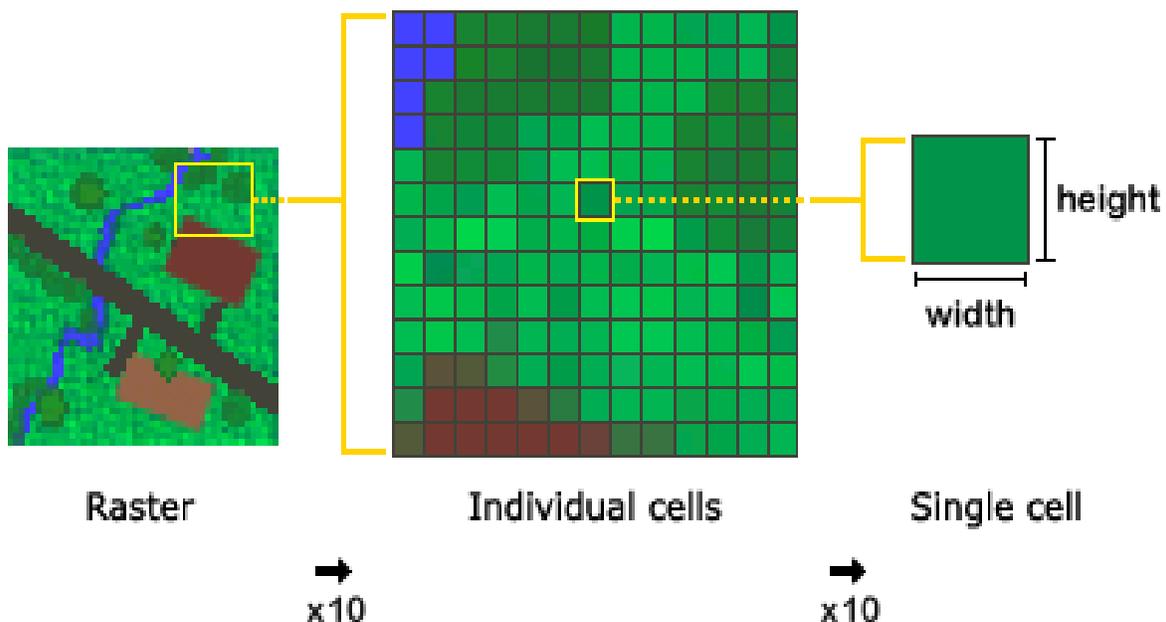


Figure 1: An illustration of the raster dataset. Source: ArcGIS Online Tutorial.

<https://desktop.arcgis.com/en/arcmap/latest/extensions/spatial-analyst/what-is-the-spatial-analyst-extension.htm>

The cells can be of any size (spatial resolution); however, the size should ensure that the desired analysis details are accomplished. For example, available resolution/pixel sizes for the DEMs are 250X250 m, 90X90 m, 30X30 m and 12X12 m. For purposes of assessing the general catchment hydrology courses resolution DEMs are suitable. Quite often, the catchment sizes are usually large and high resolution DEMs may require longer computational time. High resolution DEMs (12, 10 m or less) are best used in flood inundation mapping.

#### b) *Vector data*

Vector data provides a means of representing features of the real world in GIS. Three basic vector data types include points, lines and polygons. Points are commonly used to represent locations, lines are used to represent linear features such as streams or rivers, while polygons are used to represent areas such as land cover classes, watersheds and basins.

The shape of a vector feature is represented using geometry which is made of either single vertex or numerous interconnected vertices. Vertices define the specific position in space with reference to X, Y and an optional Z axis. A geometry feature that consists of a single vertex is referred to as a point feature. Polyline and polygon features are formed when a geometry consists of more than one vertex. Polygons are usually enclosed and usually have a minimum of three vertices. Geometries that have vertices with values in the Z-direction describe a depth/height at a specific vertex and are generally referred to as 2.5D.

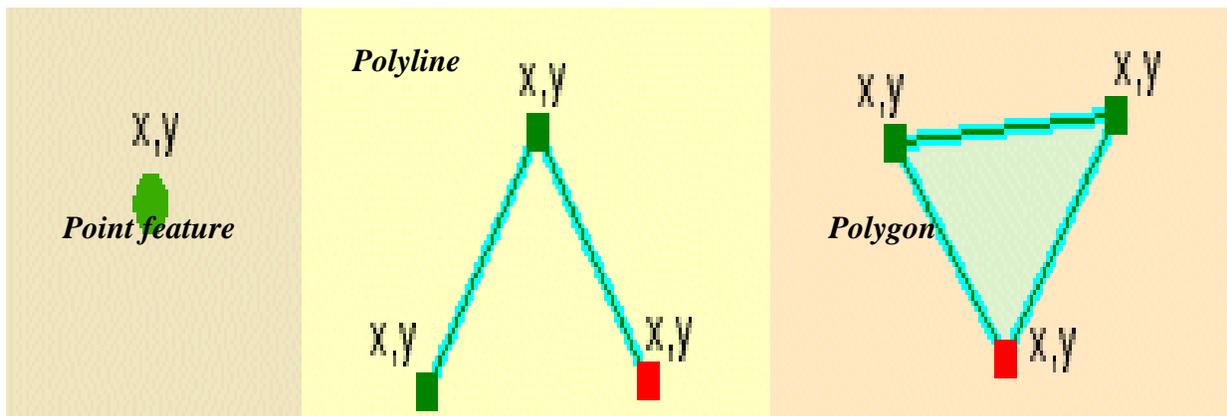


Figure 2: Illustration of point, polyline (e.g. river) and polygon (e.g. catchment) features described vertices in X and Y and an optional Z coordinate.

Polyline and polygon have a sequence of joined vertices with the polygon enclosed in that the last and first vertex are in the same position.

Triangular Irregular Networks (TIN) – TIN are digital means that represent surface morphology similar to DEM. They are vector-based geographic data constructed by triangulating set of vertices. Thus, as the name suggest, a network of triangles is formed by connecting the vertices with a series of edges. Different triangulation techniques are provided for in different GIS software packages. Considering the anatomy of TIN therefore, it is composed of faces, edges and points. Point represents data input values and defines triangle end-points. The lines drawn between points form edges which are outline of triangles. The face is the area/surface enclosed by each of the triangles.

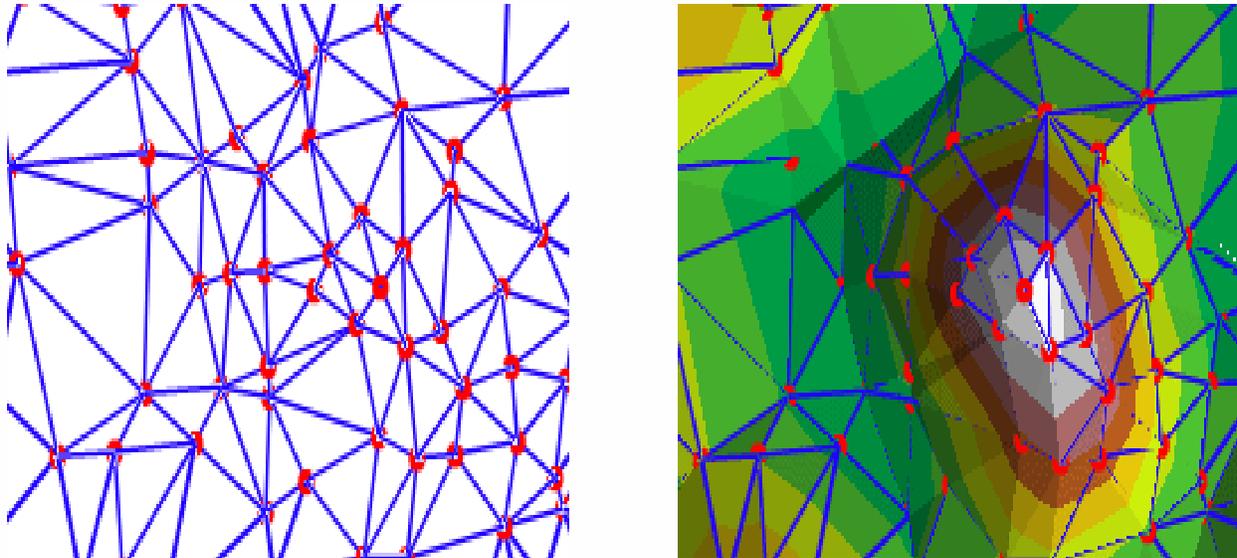


Figure 3: Illustration of the nodes and edges of a TIN (left) and the nodes, edges, and faces of a TIN (right). Source: <https://desktop.arcgis.com/en/arcmap/10.3/manage-data/tin/fundamentals-of-tin-surfaces.htm>

### 3. SOURCES OF GEOSPATIAL DATA FOR HYDROLOGICAL MODELING

The accuracy of a hydrological computation depends on the accuracy and resolution (temporal and spatial) of the input data. There are numerous parameters that are selected within the software/tool interface for the purpose of computation and analysis. Depending on the software/modelling tool, (distributed or lumped), the basic data requirement for the purpose of hydrological modelling is a listed below.

- Precipitation data - Temporal or spatial distribution data for gauges, distribution network
- Basin characteristics - Digital Elevation Models (DEM) from which basin slope, channels, sub-basins, among other properties are extracted in GIS; Landcover classification which can be derived from satellite imagery such as Landsat and Sentinel; and Soils.
- Losses - Evapotranspiration, infiltration, basin transfers, Withdrawals and consumptive water use differentiated according to source etc.
- Ground station discharge data – Validation and calibration data.

Most of the data listed above is freely available in the online domain. The choice of the most preferred data source is dependent on the resolution and accuracy of the data. Such information is usually provided as the meta-data. Table 1 below shows some of the open source data for hydrological modelling.

## Examples of data sources for hydrological applications

DATA TYPE	DATA SOURCE AND RANGE	SCALE	DESCRIPTION
Digital Elevation Model	Aster-GDEM (30 m) <a href="https://earthexplorer.usgs.gov/">https://earthexplorer.usgs.gov/</a>	Grid cell: 30 X 30 m	ASTER-GDEM of USGS
	ASF DEM (12.5 m) <a href="https://asf.alaska.edu/">https://asf.alaska.edu/</a>	Grid cell: 12.5 X 12.5 m	Radiometrically terrain-correct ALOS PALSAR data
Land-use	Land-use TM data <a href="https://earthexplorer.usgs.gov/">https://earthexplorer.usgs.gov/</a>	Grid cell: 30 X 30 m	Landsat TM 30X30 m-resolution data sets
	Sentinel data <a href="https://www.sentinel-hub.com/">https://www.sentinel-hub.com/</a>	Grid cell: 10 X 10 m	Land-use derived from 10 X 10 m resolution Sentinel datasets
Soils	FAO Soil <a href="http://www.fao.org/soils-portal/soil-survey/en/">http://www.fao.org/soils-portal/soil-survey/en/</a>	1:5,000,000	FAO Digital soil map
Weather Data	Global Weather Data <a href="https://globalweather.tamu.edu/">https://globalweather.tamu.edu/</a>	0.05° resolution	Climate Forecast System Reanalysis (CFSR)
	CHIRPS <a href="https://www.chc.ucsb.edu/data/chirps">https://www.chc.ucsb.edu/data/chirps</a>	0.05° resolution	Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS)
	TRMM <a href="https://gpm.nasa.gov/data">https://gpm.nasa.gov/data</a>	0.05° resolution	Tropical Rainfall Measuring Mission
Stream Network	HydroSHED <a href="http://www.hydrosheds.org">www.hydrosheds.org</a>	-	Mapping product that provides hydrographic information in a consistent format

## 4. PREPARATION OF GEOSPATIAL DATA IN OPEN SOURCE QGIS FOR HYDROLOGICAL MODELING

### 4.1. CATCHMENT AND STREAM NETWORK DELINEATION FROM A DIGITAL ELEVATION MODEL

Catchment delineations and extraction of stream networks is a fundamental process in hydrological analysis (Van der Kwast and Menke, 2019). GIS tools offer flexible capabilities to derive these topological characteristics, using DEM as main Data. The generic workflow for catchment delineation includes the following:

- i. Download DEM tiles for the study area in focus. Mosaic the titles if the study area is covered by multiple tiles. As indicated in Table 1 and depending on the desired resolution, the DEM tiles can be downloaded from either USGS or ASF website.
- ii. Downloaded DEM tiles are usually prepared using geographic coordinated and therefore must be re-projected into UTM coordinates. In case, the merged DEM tiles cover a larger area far beyond the study area, it is important to clip/subset into a smaller area to reduce the processing time. Usually larger and high resolution DEMs take longer to processes.
- iii. Due to a number of unknown factors, raw DEM may include errors and artifacts that may result to false derivatives such as the flow paths. Therefore, it is vital to hydrologically correct the DEM by filling the sinks that may alter flow and ensure that other derived topological factors are logical according to the hydrological rules.
- iv. Derive flow directions and flow accumulation raster. The amount of taken for processing will depend on the size of the study area and pixel sizes. This is critical when dealing with higher resolution extensive Digital Elevation Models. The flow accumulation and flow direction raster are automatically added to the map view after execution of the tool.
- v. Derive the drainage network and further calculate the catchment upstream of the selected outflow point. Note that when delineating catchments, the outflow locations are usually known sites such as discharge point, or hydrological station. With a known stream network, the process accuracy can be improved by “burning” the river network into the DEM. Figure 4 illustrates a summary of the describe workflow.

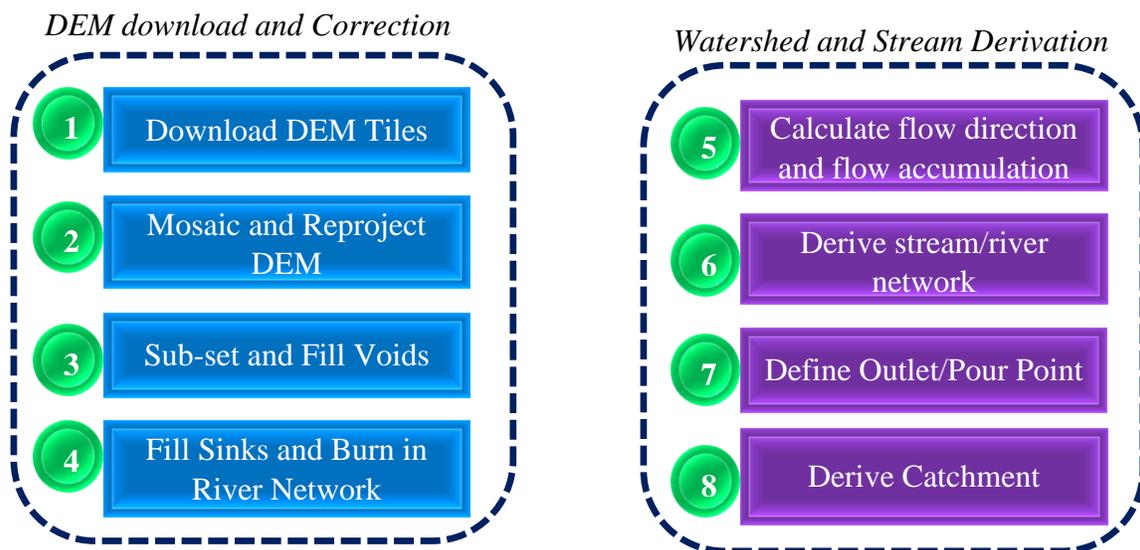


Figure 4: Workflow for stream and catchment delineation

A step by step illustration of the above described workflow follows below. Note that for the purpose, Olkeria Catchment in Kajiado County of Kenya has been adopted as a Case area for the illustration and 30 m ASTER-GDEM from USGS earth explorer has been adopted as the primary input data.

## 4.2.DEM DOWNLOAD AND CORRECTION

### i. Download DEM Tiles

We will download SRTM 1 Arc-Second (30 m) tiles from earth explorer website. (<https://earthexplorer.usgs.gov/>). You need to register in order to download the SRTM tiles. After registration and login into the USGS Earth Explorer zoom into the area of interest and click on the map to enclose the location of the watershed. It is necessary to have prior knowledge of the exact location of the catchment of interest. Upload of a KML or zipped shapefile of the area is also available as a search criterion.

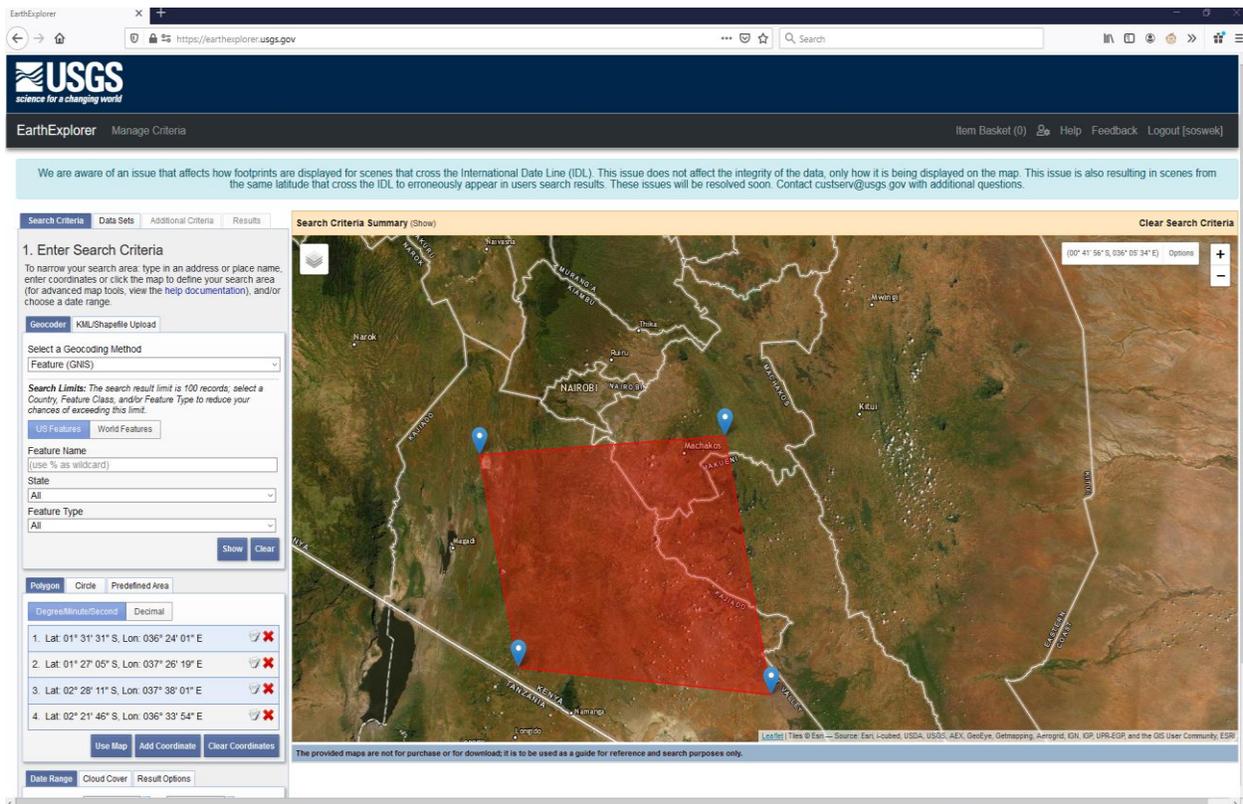


Figure 5: Search criteria with selected area of interest

After specifying the search criteria, click on Data sets and select the desired data for download. To download the DEM, click on the plus sign to select **Digital Elevation** and select **SRTM, SRTM 1 Arc-Second Global** and then click on the **Results** tab. You can see the search results displaying 4 tiles. Sometimes, you may define an area that results to tiles far outside the area of interest, you can show browse overfly of each tile by clicking on  for each search result to ensure that you download the tiles only need.

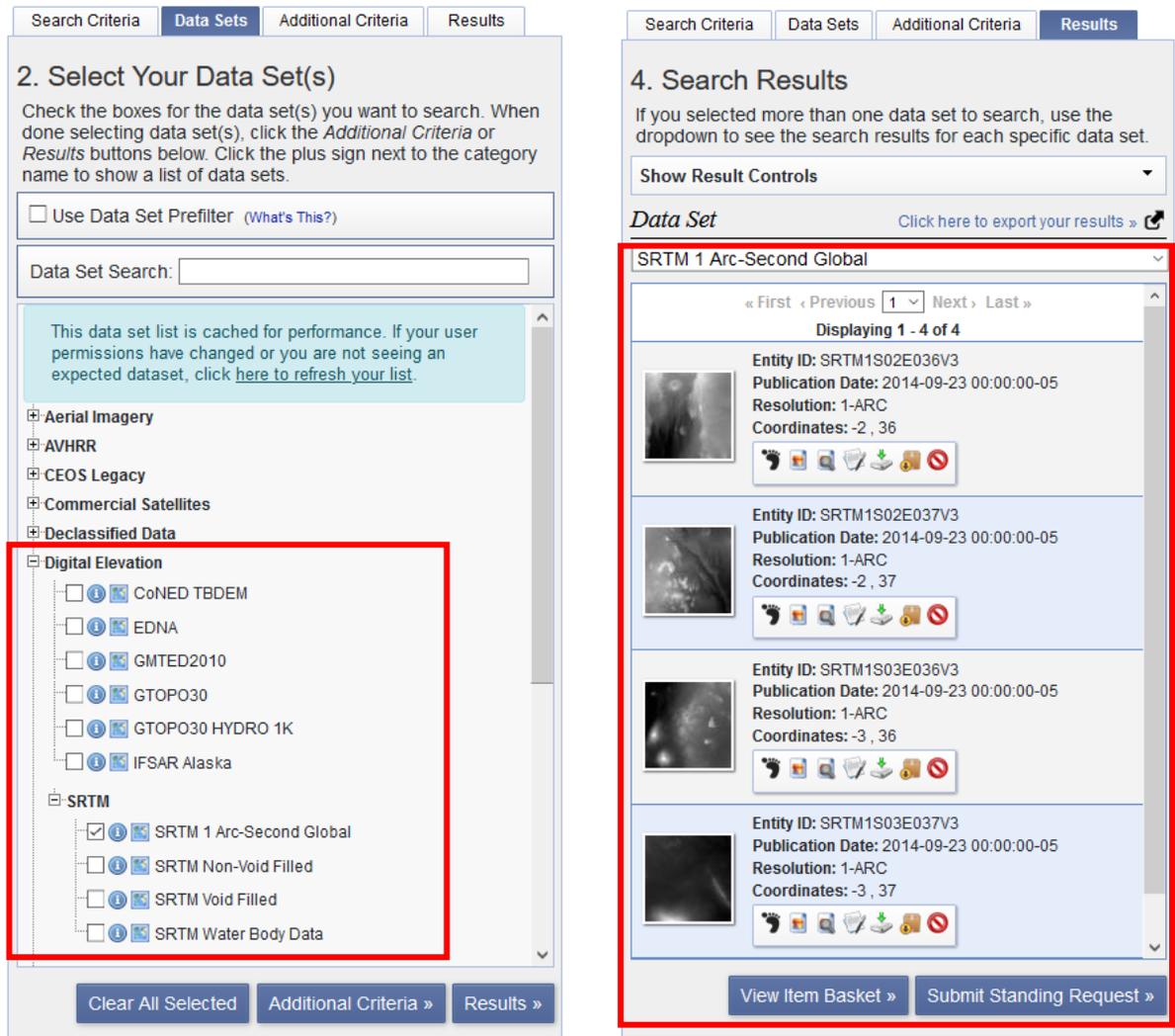


Figure 6: a) Selection of SRTM 1 Arc Second Global Data set and b) Search results with four (4) tiles in display.

Download the tiles by clicking the **Download Options button**  for each tile and select **GeoTIFF** file format in the **Download options dialogue**. Hurrah!!! You can download all the tiles and add them to you QGIS project.

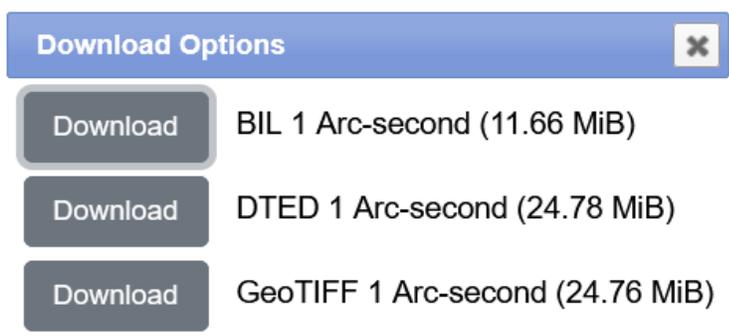


Figure 7: Download options dialogue

## ii. Mosaic and Reproject Dem

We assume that you have QGIS installed on your computer. If not, download and install the stable version of QGIS Desktop from <https://qgis.org/en/site/forusers/download.html>. Add the downloaded files to QGIS by dragging them to the map canvas from the browser tab in QGIS. We will merge the files by building a virtual file (.vrt). Although QGIS provides for merging the tiles in a physical file such as GeoTIFF, the option is slower especially when many tiles are involved and therefore creating files virtually is preferred.

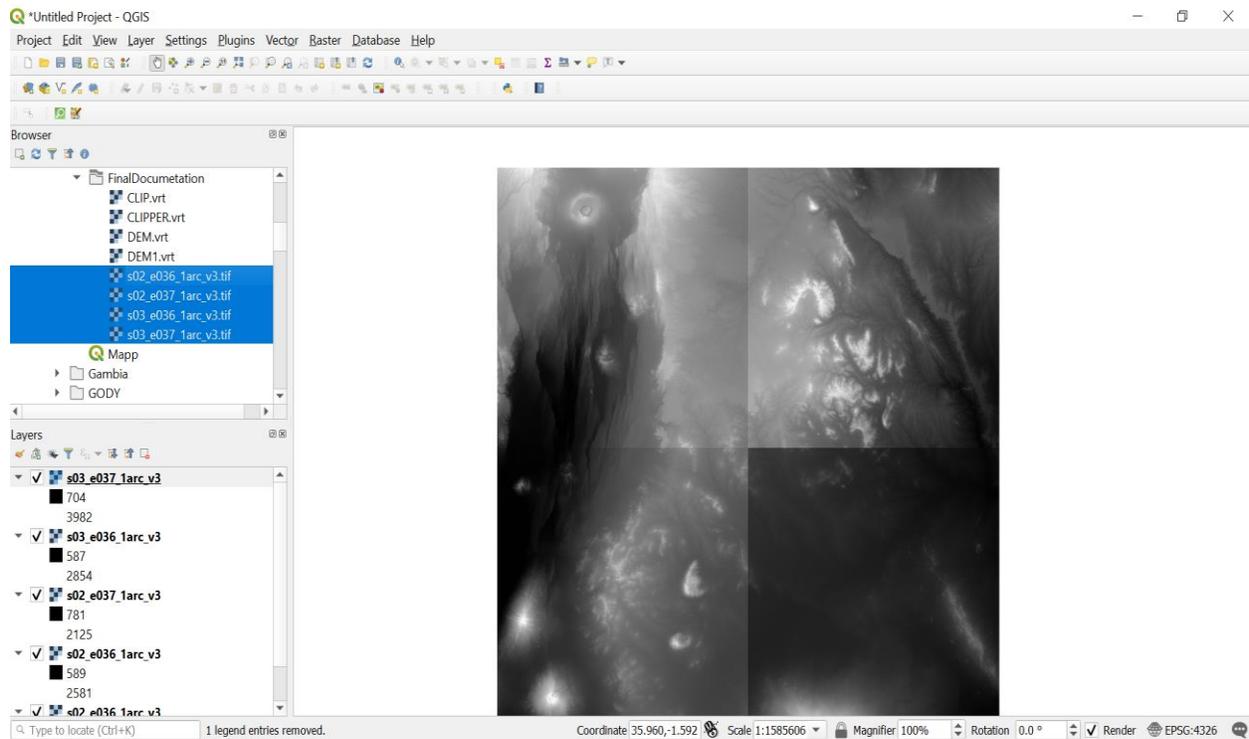


Figure 7: SRTM tiles added to QGIS map canvas.

In the main menu, select **Raster, Miscellaneous, Build Virtual Raster**. In dialogue box choose the files in the directory or click select all button to select all the files and click Ok. Browse to the working directory and assign the name to the output file. Remember to uncheck **Place each file into a separate band**. Maintain other settings as default and Click Run. A new layer will be added on the map canvas. At this stage, you can remove the individual tiles added and leave the mosaiced layer.

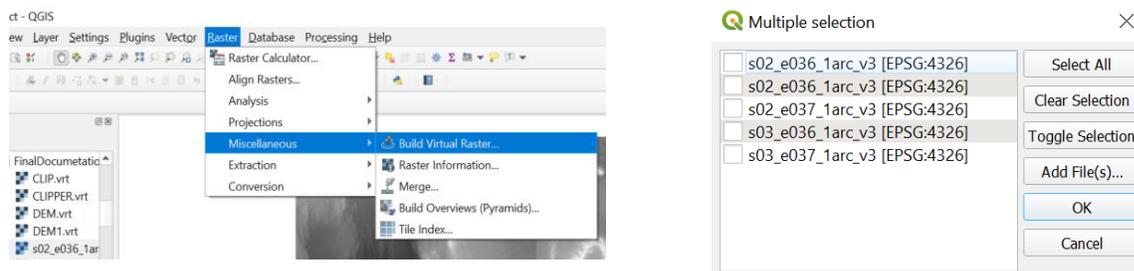


Figure 8: Build virtual raster option and multiple raster selection dialogue box.

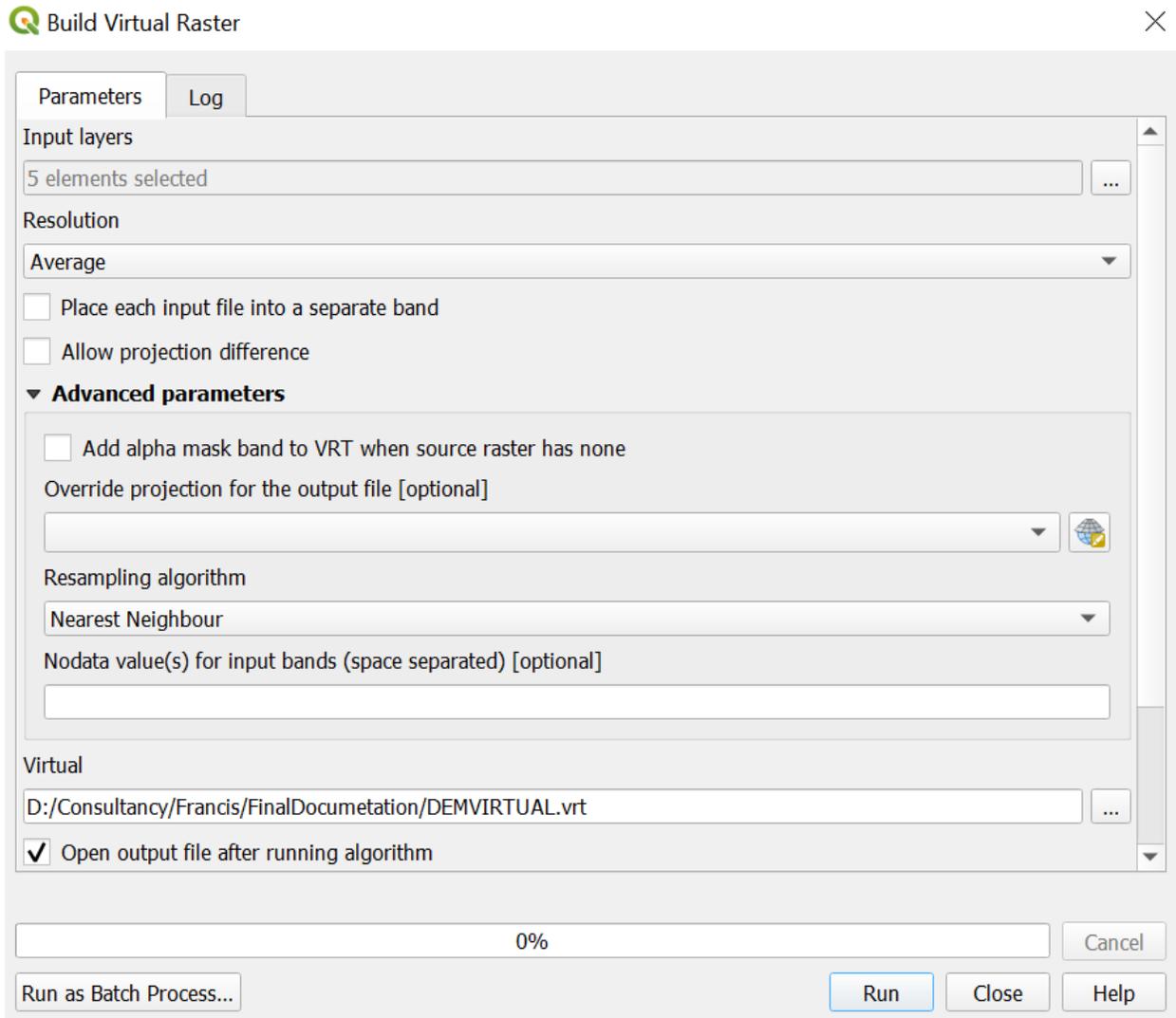


Figure 9: Build virtual raster dialogue

The DEM is still in its original Geographic Coordinate System (GCS) with WGS 84 (EPSG: 4326) datum. EPSG refers to the European Petroleum Survey Group that has published a database of coordinate system information including related documents on map projections and datums. If you are not sure with the projection of your area, then you find this on EPSG website by searching the area of interest (<https://epsg.io/?q=kenya>). If the project covers one country let's say Kenya, then we choose the national projection. Otherwise, you can use the global projection. In this case, we will use the global projection: UTM Zone 37 South with the datum as WGS-84. You can get the PSG code form the website highlighted above.

To re-project the DEM, select Raster, Projections, Wrap (Re-project) from the main menu. In the appearing dialogue window select Target CRS and browse to find the specific for your location using the EPSG code. Remember to set the No data Value to -9999 and set the output file resolution to 30 m. Leave other settings as default, save the file to your working directory and click run. To display the new dataset properly, change on-the-fly projection of the project on the lower right corner of the map canvas to match the projection of your study area.

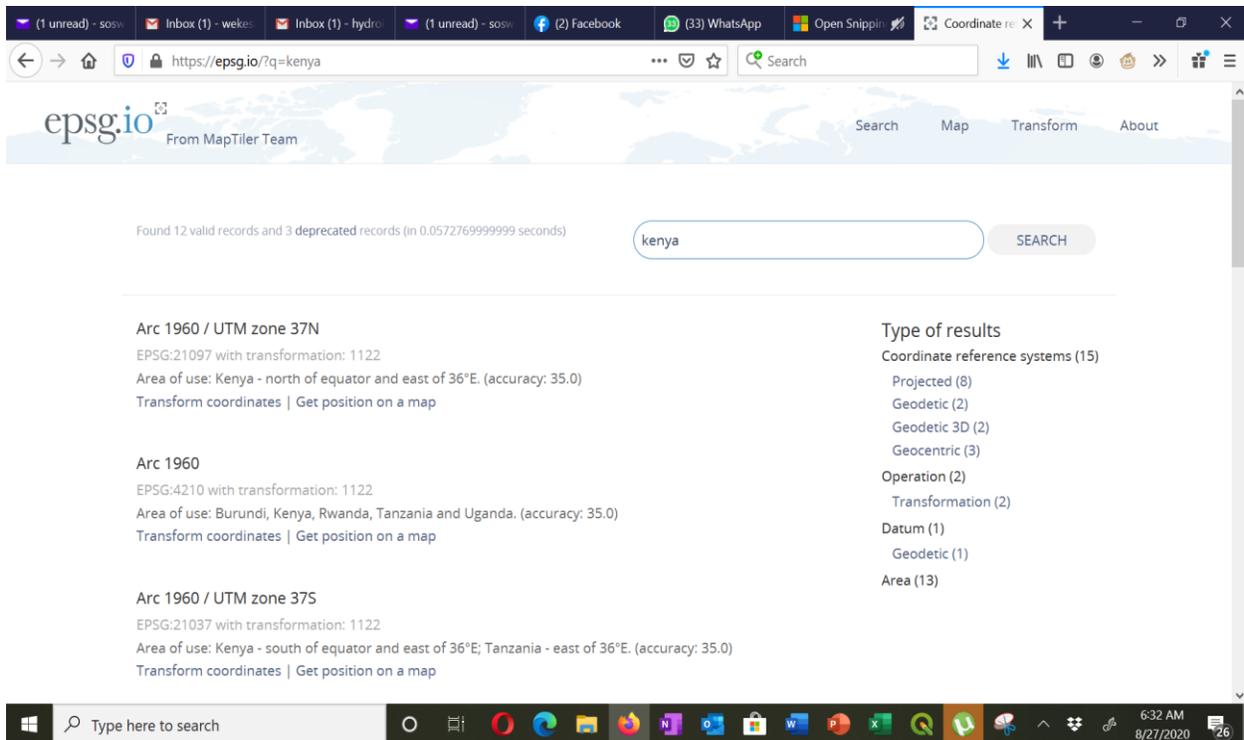


Figure 10: Result search for EPSG.IO

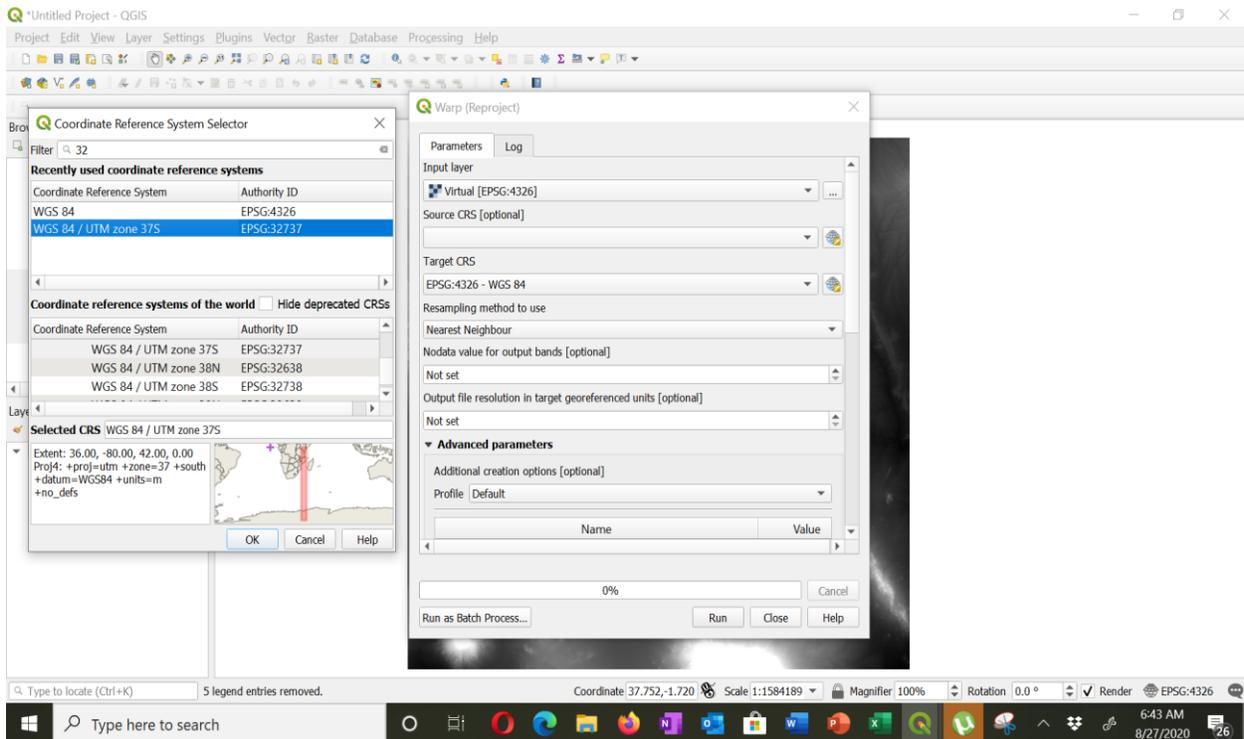


Figure 11: Wrap (Re-project) dialogue and select CRS using the EPSG code. The on-the-fly projection is on the lower right corner of the map canvas i.e. see EPSG 4326. This has to be changes so that it matches with your defined projection.

iii. Subset and Fill Voids

Will reduce the algorithms calculation time by clipping the raster layer to a smaller area. In the main menu, select Raster, Extraction, Clip Raster by Extent. Check whether correct input layer is selected, click on clipping extent and choose select Extent on Canvas. You can now click and drag on the map to highlight the clipping extent. Make sure that no data value is assigned to -9999. Specify the you working directory to save your output file and click run. We will assume that our DEM has no voids.

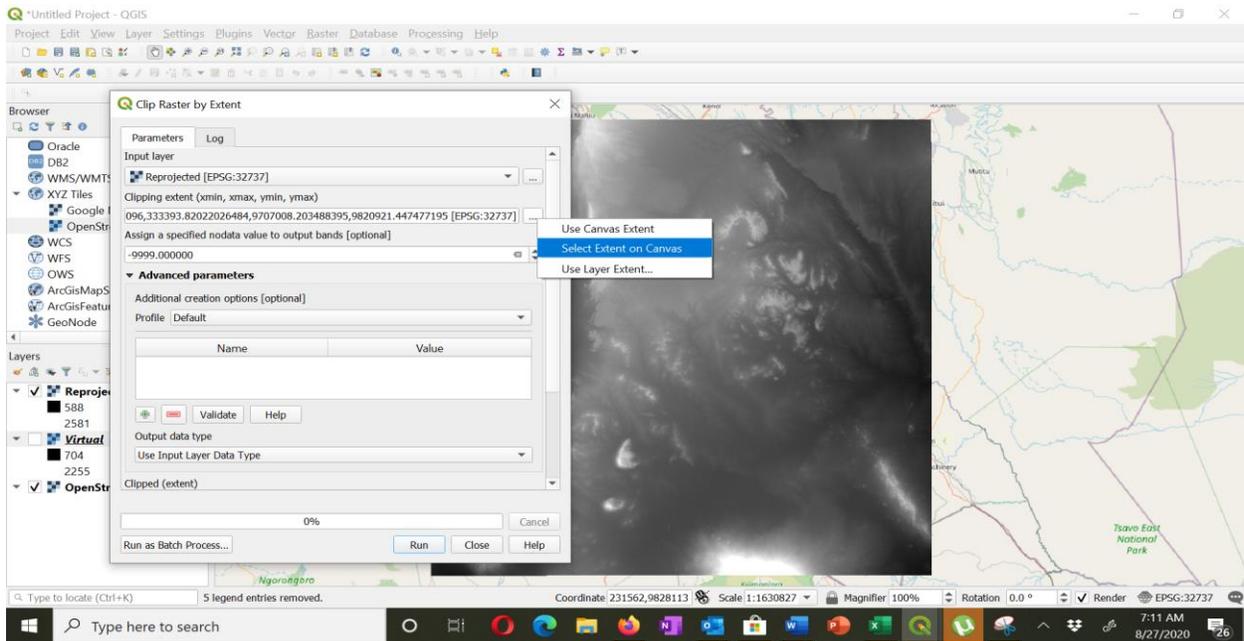


Figure 12: Clip raster by selecting extent on the canvas

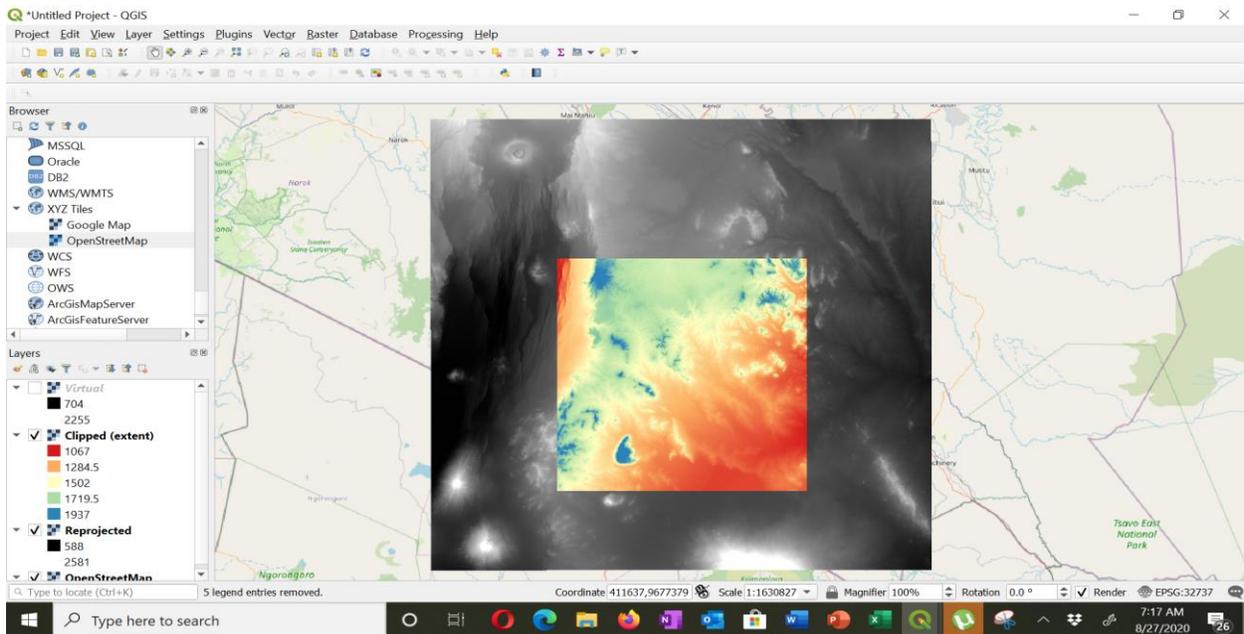


Figure 13: Clipped DEM extent.

## iv. Fill Sinks

Several algorithms have been developed for filling the sinks. For our case, we will use algorithm developed by Wang and Liu (2006) available in the QGIS Processing Toolbox. Choose Processing, Toolbox from the main menu to enable the Processing Tool box. The tool box will be now visible at the right side of your screen. Go to *SAGA, Terrain Analysis – Hydrology, Fill Sinks (Wang & Liu)*. In the dialogue box, maintain the default minimum slope and confirm that you have selected the correct input file. For this case, we have selected clipped (Extent) file. Click run and close when done.

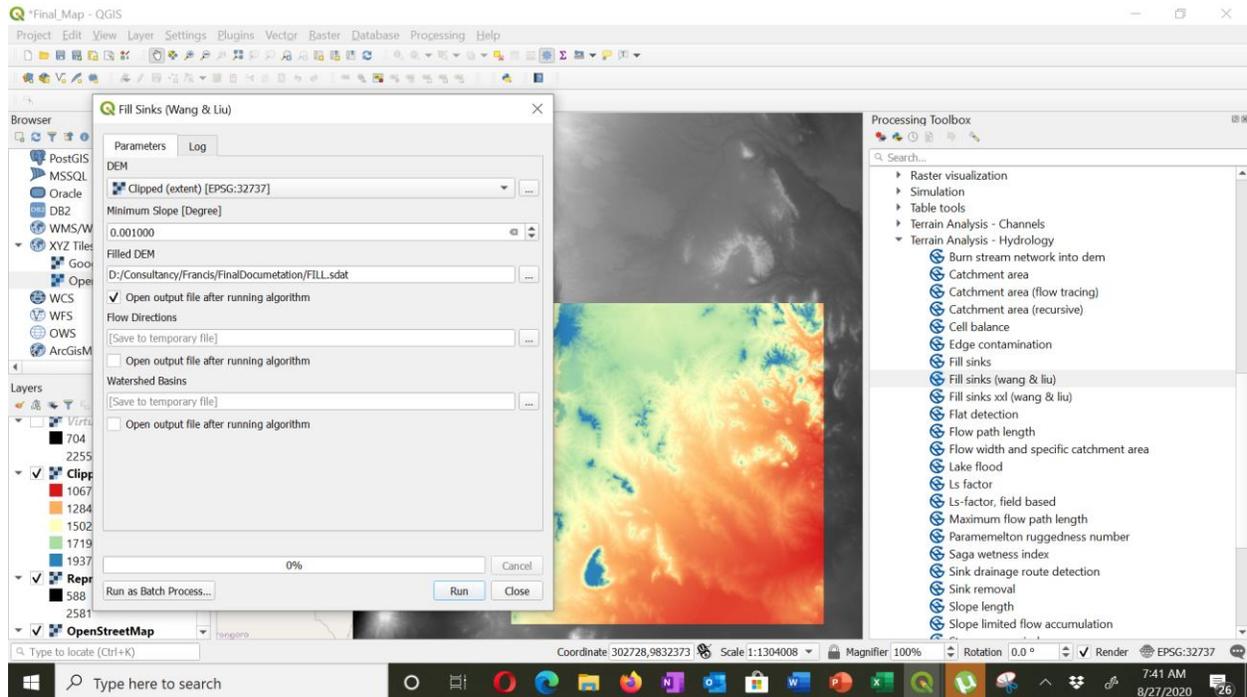


Figure 14: Search for fill sinks in the Processing toolbox and Fill sinks dialogue box.

N/B. The fill tool described above also process the flow direction and accumulation. We chose not to load the files on the map canvas for our case.

### 4.3.WATERSHED AND STREAM DELINEATION

#### Delineate Streams

Determining what you consider as streams is a pre-requisite step for deriving streams for a DEM. For our case, we will use Strahler classification. In Strahler, links with no tributaries are assigned order 1 with the order increasing when same order streams intersect. Note that an intersection of two streams of different order will result to no increase in order and always a higher order takes precedence. This is the most common approach for stream ordering, although GIS software provides for other methods of ordering such as the Shreve method. All links in the Shreve approach are accounted for. For instance, intersection between of 1<sup>st</sup> and 2<sup>nd</sup> order streams results to 3<sup>rd</sup> order streams.

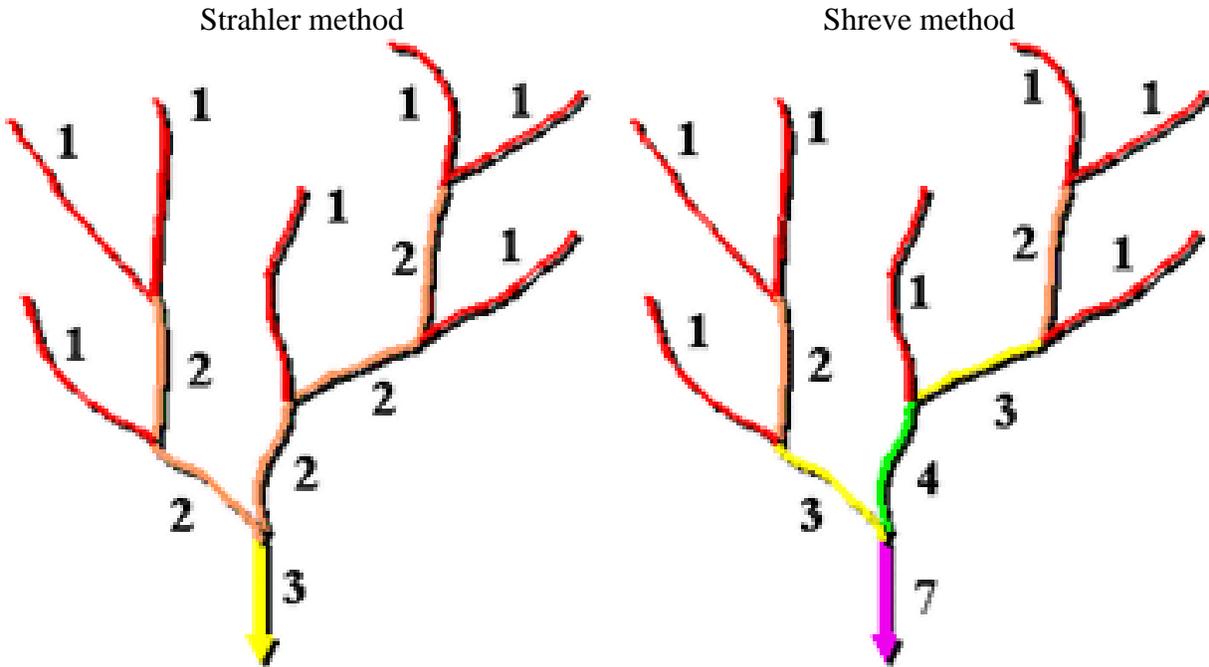


Figure 15: Methods of stream ordering (Image adapted from <https://pro.arcgis.com/en/pro-app/tool-reference/spatial-analyst/how-stream-order-works.htm>)

Find Strahler in the Processing Toolbox: SAGA, Terrain Analysis – Channels, Strahler. Select the correct input file as the filled DEM and click run.

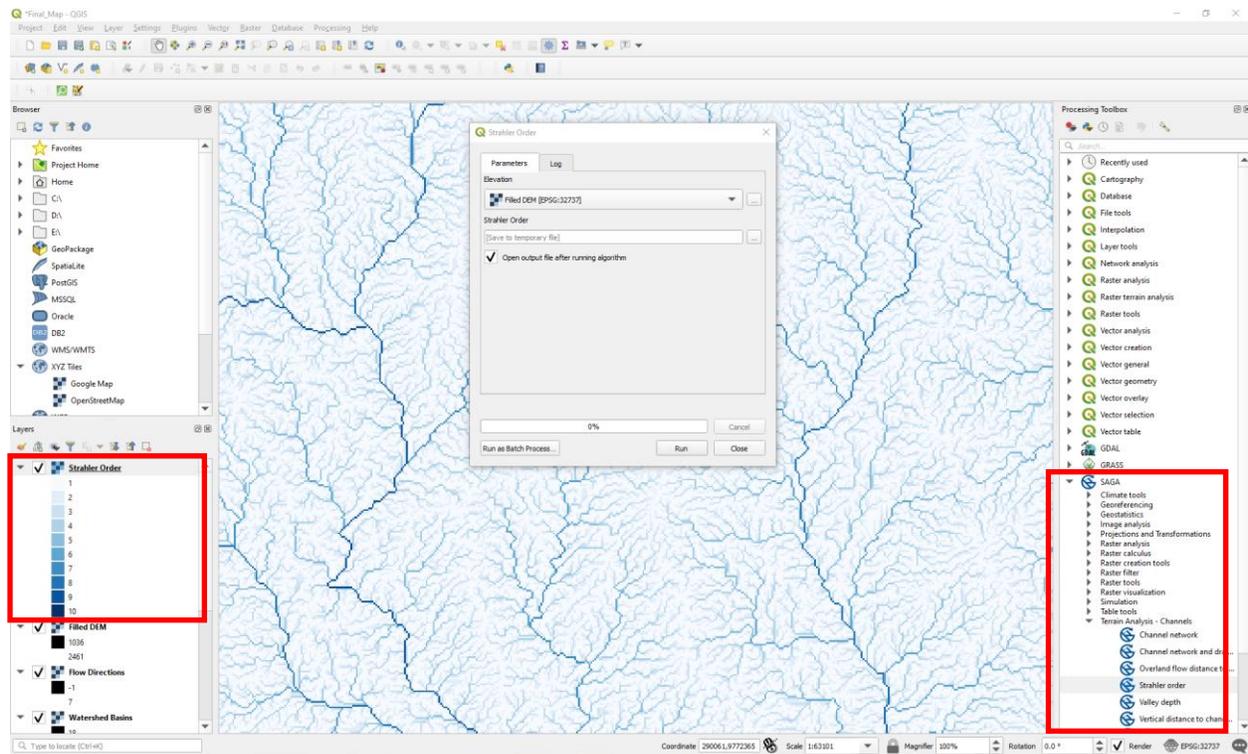


Figure 16: Strahler order tool, Strahler order dialogue, Strahler order styled with intuitive colors.

The stream order above shows that the highest order is 10. Defining stream network requires expert knowledge of the study area. To reduce the streams to match the density of network for the catchment in focus, we will apply the calibration procedure to determine the Strahler order streams that we consider streams. Create a Boolean map of the Strahler order using the Raster Calculator (**Raster, Raster Calculator**) for Strahler order  $\geq 5$  and zero (0) for other values.

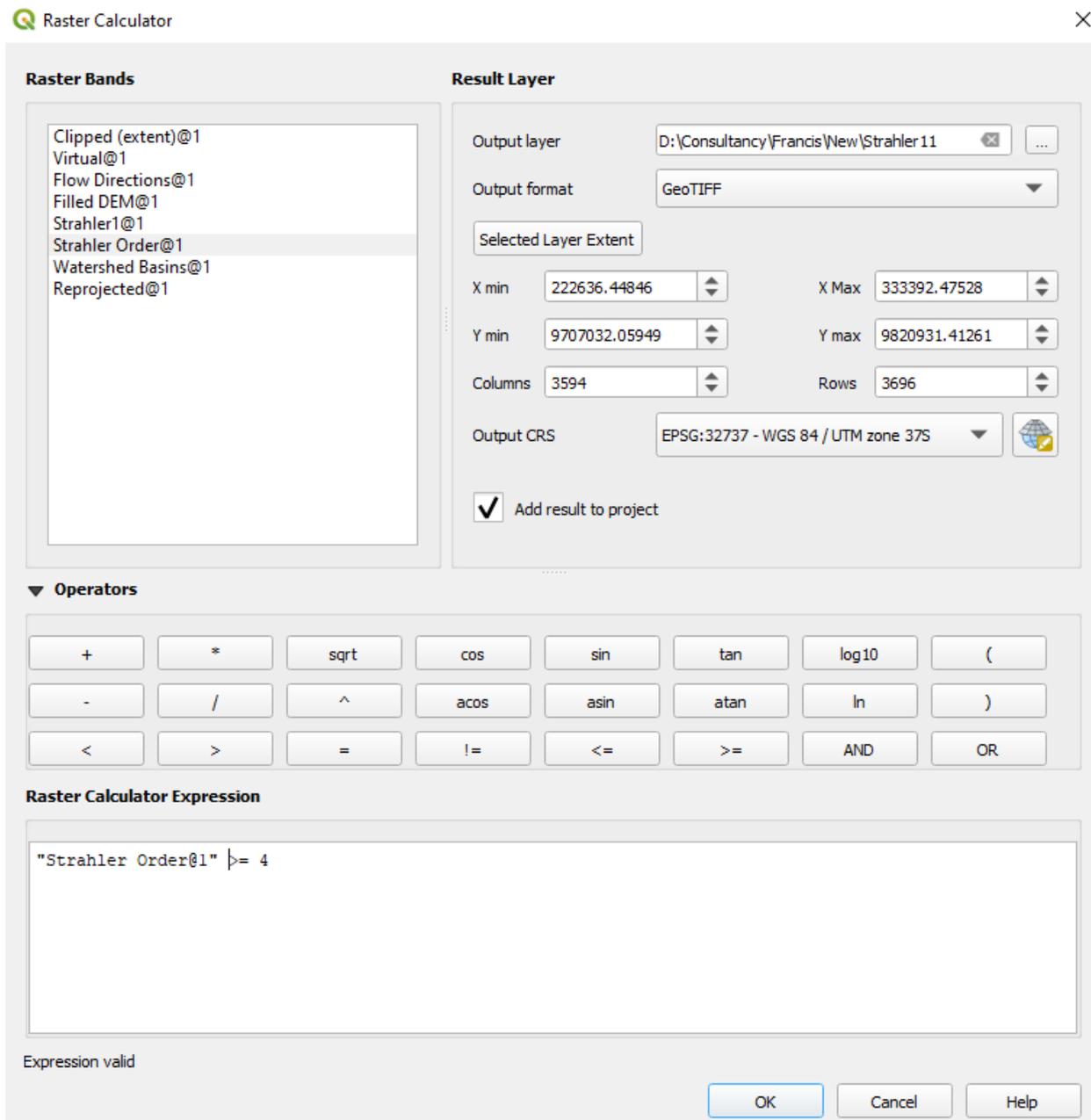


Figure 17: Raster Calculate Dialogue with a Boolean expression

The Boolean layer will have only pixels with values 0 and 1. If not sure on the nature and density of stream network in the catchment of interest, use OpenStreetMap or Google Satellite overlay on the Strahler raster.

With the Strahler Boolean threshold calculated, we can now calculate the channel stream network. We will also calculate the flow direction layer using the same tool. In the Processing Toolbox, choose SAGA, Terrain Analysis – Channels, Channel network and drainage basins.

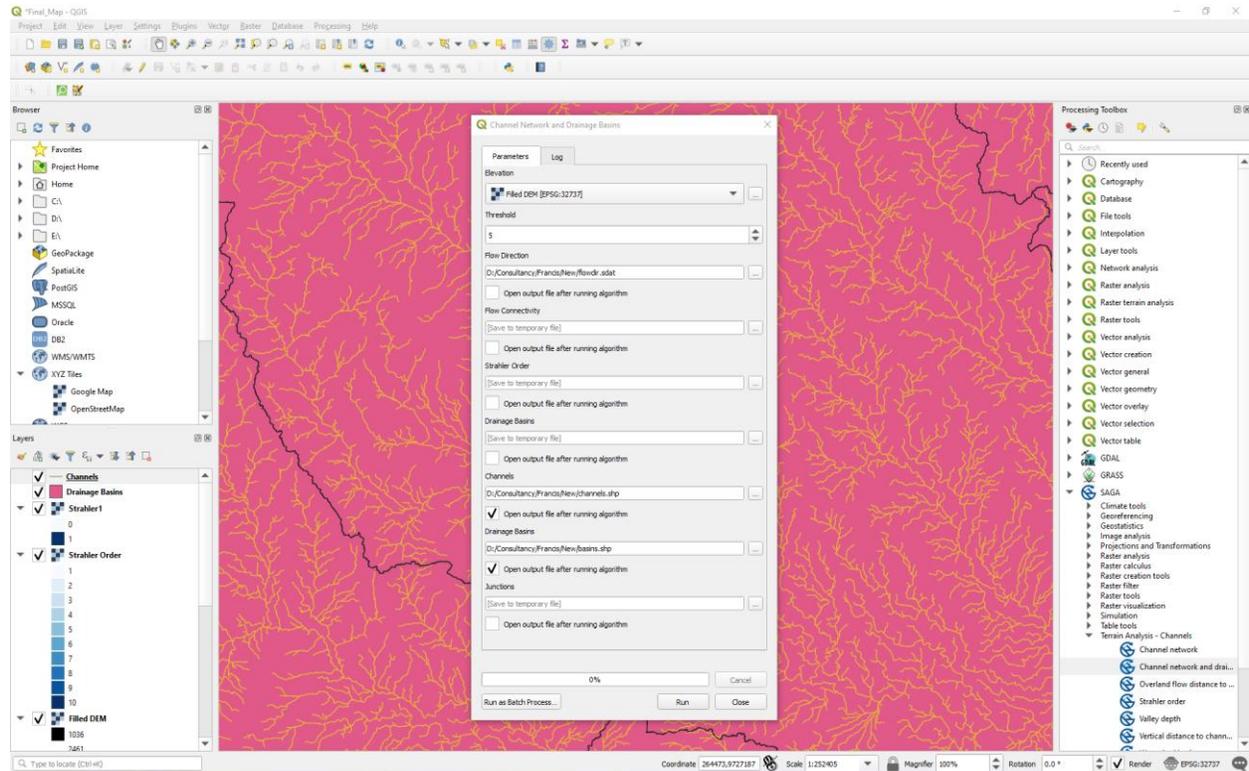


Figure 18: Channel Network and Drainage Basins dialogue.

### Define outlet/ Pour point

In order to extract the catchment boundary from a DEM, we need to have the following:

- Channel network that match the flow direction extracted from a DEM that is hydrologically corrected
- Outlet coordinates in the same coordinate system as the map canvas/project we are using.

The coordinate capture plugin will help get the coordinates of the outlet from the map canvas. If you do not have the tool, install it from manage plugins with searching coordinate capture then click install. If installed and you can't see it on the tool bar, active the plugin in the manage plugin dialogue box as well.



Figure 19: Tool bar showing activated Coordinate Capture Tool (Enclosed in the red box).

Note that with installation and activation, coordinate capture will appear on the side bar. Click Start capture click on the desired outlet point on the river.

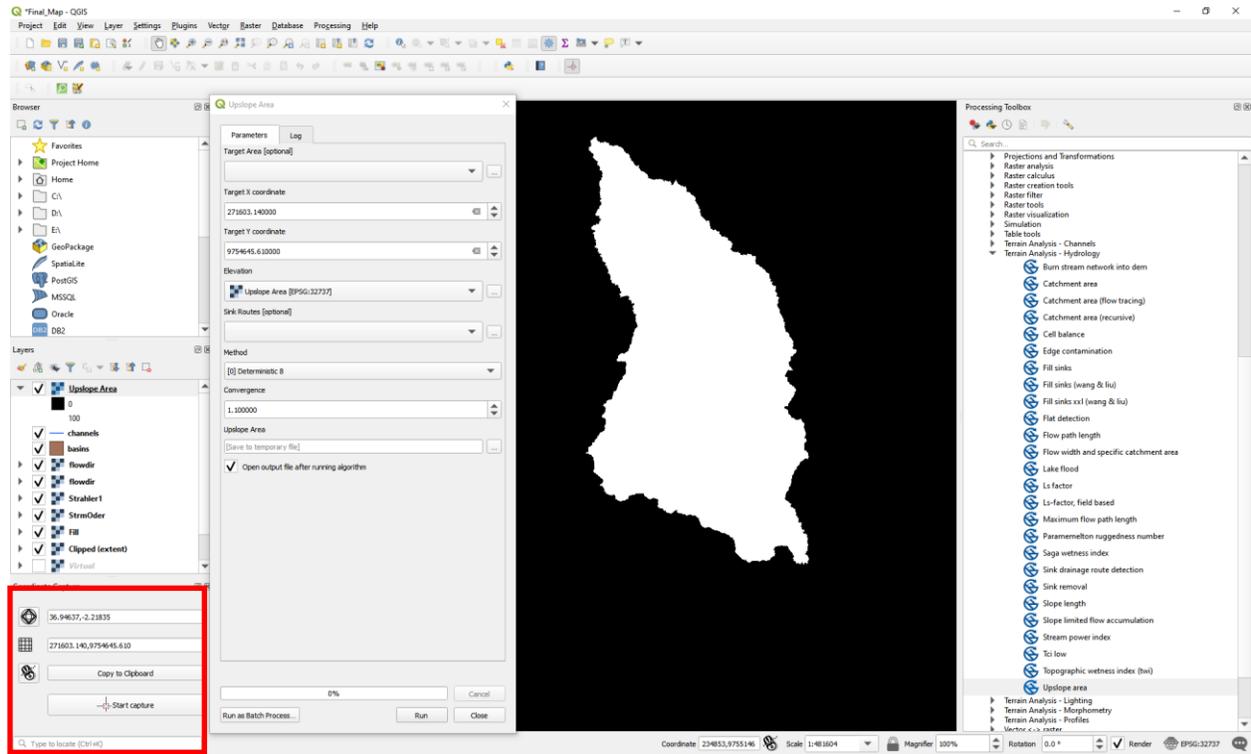


Figure 20: Upslope area tool dialogue and results. Note also the Coordinate capture tool.

We will now polygonise to convert the Catchment raster to vector file format. Select Raster, Conversion, Polygonize (Raster to Vector). Always make sure that the right input file is chosen.

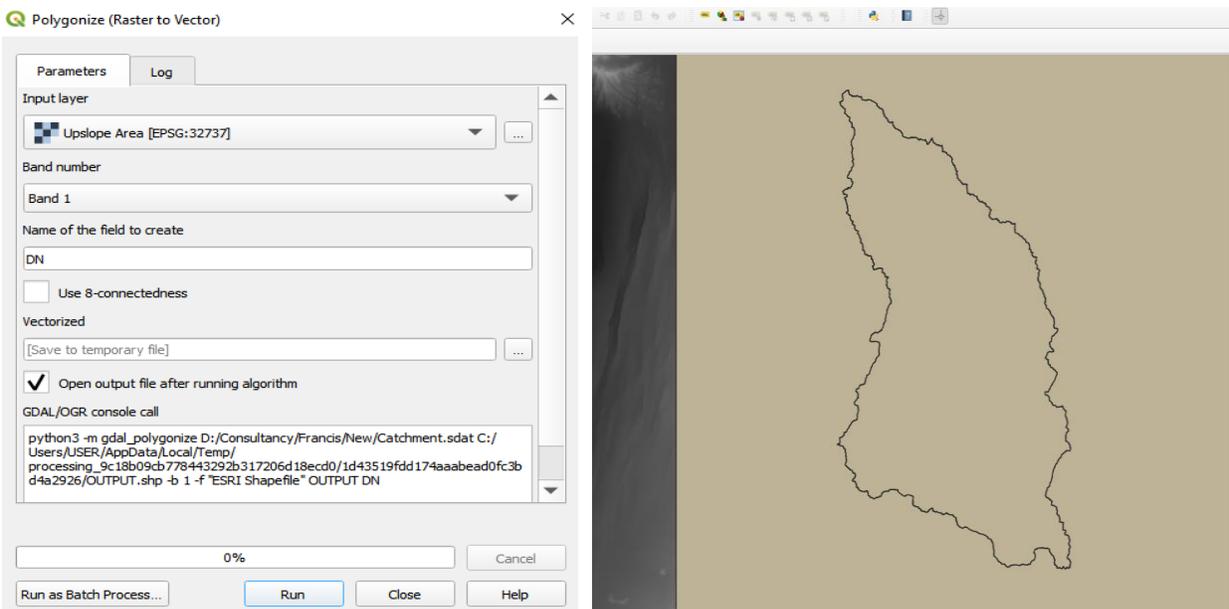


Figure 21: Polygonize dialogue and output of the conversion process.

We are only interested with the catchment boundary and therefore we will remove the outside polygon. Right click on the catchment vectorized layer and click open attribute table. In the attribute table, toggle to editing mode by selecting  and then select the record to remove. The selected area will always appear in yellow. Click button to  delete and toggle off editing by clicking  again then save the changes. Close the attribute table. Hurrah! We have our catchment boundary and stream network extracted.

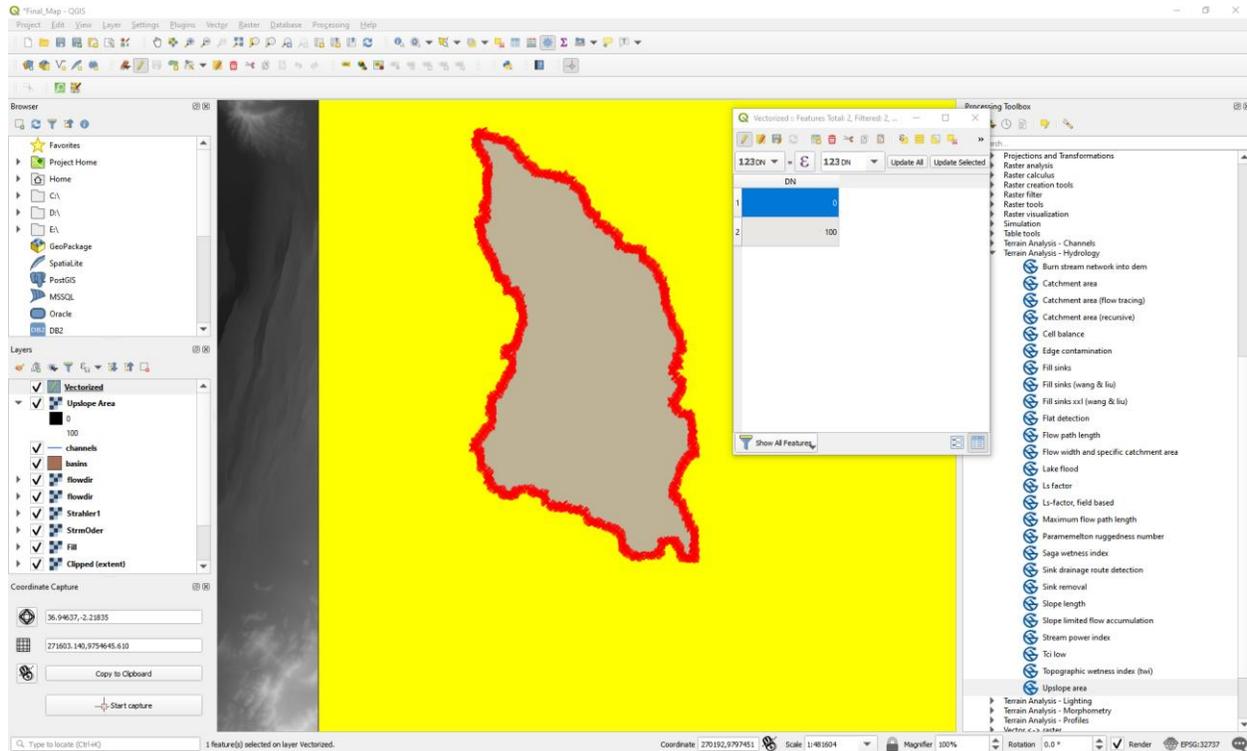


Figure 22. Delete polygon with Zero (0) value.

## REFERENCES

- Korres, W., Schneider, K., 2017. GIS for Hydrology, in: Comprehensive Geographic Information Systems. <https://doi.org/10.1016/B978-0-12-409548-9.09635-4>
- Peel, M.C., Blöschl, G., 2011. Hydrological modelling in a changing world. Prog. Phys. Geogr. <https://doi.org/10.1177/0309133311402550>
- Van der Kwast, H., Menke, K., 2011. QGIS for Hydrological Applications: Recipes For Catchment Hydrology and Water Resource Management