

Open Source GIS-based Lake Basin Delineation Procedure: a Tutorial

Khila Dahal

1. Introduction:

A comprehensive lake (or river) management plan generally includes an indicator-based assessment of the status of, and threats to, the lake under consideration. Stresses to the lake emanate from its drainage basin, which ultimately determine the health of that lake, i.e. quality and quantity of its water. Therefore, it is important to determine an accurate areal extent of the watershed of the lake before making any threat assessment. Only then, we can start identifying all the threat indicators relevant to the lake and initiate their evaluations. This document provides a detailed GIS-based procedure to delineate lake watersheds (other names include drainage basin and catchment area). The purpose here is to illustrate a procedure for lake basin and stream network delineation based on digital elevation model (DEM) in a freely available open source GIS platform.

In terms of ownership, there are two general types of GIS software: proprietary (commercial) and open source. The software and its source codes are freely available for public use (i.e. to download, modify and use as needed) in open source GIS, whereas users should pay to acquire and use proprietary GIS such as ESRI's ArcGIS. There are multiple open source GIS platforms to choose. QGIS, GRASS, SAGA, GeoDa, DIVA-GIS and gVSI are some commonly used open source Desktop GISs. You can find a comparative description of the software on this link: <https://unearthlabs.com/blog/free-gis-software/>. As each of these software have been developed with specific objectives, they have their own strengths and are more appropriate for certain purposes. Particularly, QGIS (<https://www.qgis.org/en/site/>) is a more comprehensive platform outperforming other existing open source GIS packages (GIS Geography 2020; Menke 2017). On top of being an open source platform, it is user-friendly, highly reliable (i.e. not buggy), faster in geo(processing), presents a better documentation along with elaborate illustrations and tutorials, has one of the largest GIS user-communities, allows for the widest range of geoprocessing operations, and consumes all GIS data types (QGIS 2018). It comes with most regular and timely updates. Moreover, it is treasured with an ever-growing number of native as well as user-developed plugins. It integrates external packages such as GRASS, GDAL and SAGA, thereby adding to the repository of available tools and operations.

In this tutorial, we use QGIS to generate basin boundary and stream networks. Generally, hydrology and channel tools such as Fill Sink, Flow Direction, Flow Accumulation, Stream Orders and Upslope Area are run in a special sequence in GIS for the delineation purpose (Figure 1). In the QGIS desktop interface, these tools are available under GDAL, GRASS and SAGA toolboxes as QGIS well integrates these packages and other plugins. Although the Hydrology tools under each of these packages/plugins work in their own special way, we will use SAGA tools simply because SAGA has an edge in terrain analyses and spatial statistics. SAGA has a more sophisticated array of Hydrology and Channels tools with comparatively more accurate algorithms and their smooth (bugless) implementation (Hengl et al. 2009). Thus, we will run SAGA tools from within the QGIS interface in this exercise. This also allows us to integrate SAGA tools with the tools developed by other platforms such as GRASS and QGIS thereby making our geoprocessing tasks more robust. Here, we will run different hydrology operations such flow direction, flow accumulation, streams, stream segments, and watersheds. We will use both raster and vector datasets for the purpose.

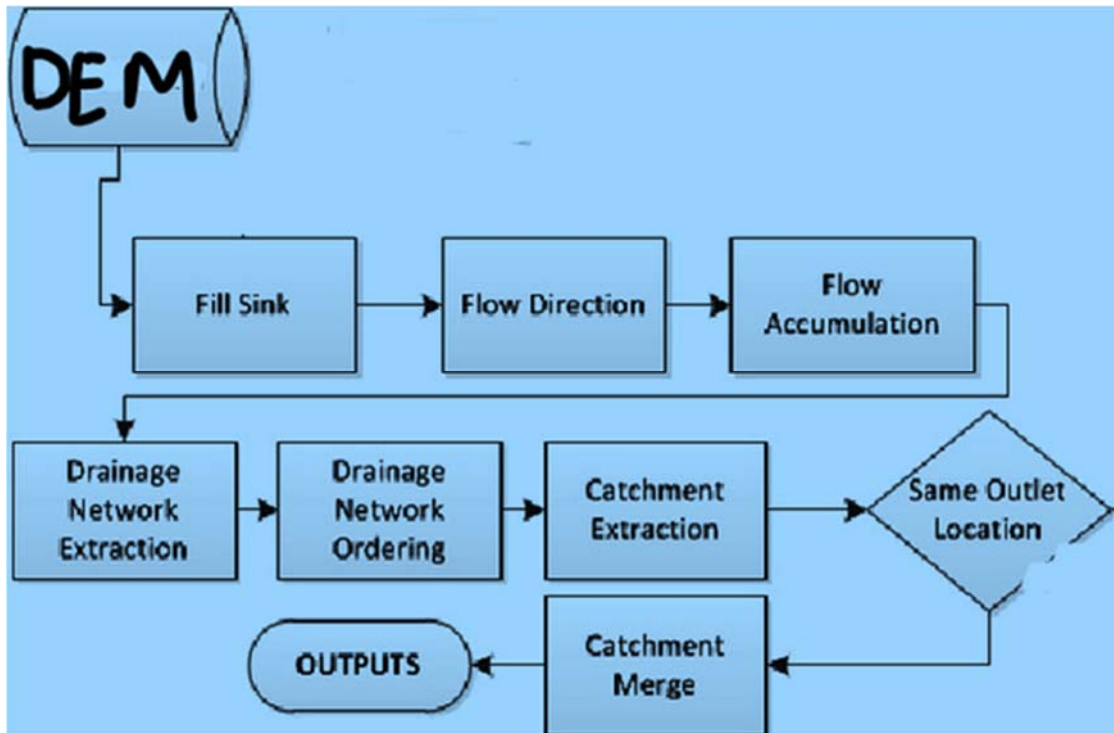


Figure 1: A generic lake/river basin delineation workflow (adapted from Anornu et al. 2012)

LEARNING OBJECTIVES:

- Enhance familiarity with open source GIS packages, specifically SAGA and QGIS functionalities.
- Explore various online data sources to find necessary geographical datasets for lakes, elevation and other socio-demographic indicators
- Identify and run geoprocessing tools involved in DEM reconditioning.
- Understand and implement some raster local and focal functions related to hydrologic analysis.
- Create and edit different layers in a GIS
- Recognize and properly execute the sequence of Hydrology tools required to delineate drainage basin and stream networks from the DEM.
- Understand and interpret the results of different Terrain Analysis results.
- Develop watershed and river stream maps.

SOFTWARE AND DATA:

Required software: QGIS. You should download and install QGIS on your computer (if you have not already) before we start the delineation. If needed you can refer to any of various tutorials on how to download and install QGIS found on the internet. GIYF, i.e. just google ‘how to download QGIS’ and follow the instructions or watch the YouTube tutorial. It is better to download QGIS version 3.14 so that you can see the same menus and buttons as in the tutorial although other versions are also OK.

Necessary datasets: Lake (optional) and elevation data. Both datasets will be downloaded from web sources.

NOTE: This tutorial is accompanied by a video tutorial, which explains some of the steps in a more meaningful way with visual graphics and screen-shares. Therefore, it is a better idea to refer to the video version of this tutorial as needed. These two versions are complementary to each other.

2. Basin Delineation Methodology:

2.1 Obtaining and preparing the lake and DEM data

The very first step in the process of watershed delineation is to obtain GIS data for the lake for which we want to create the watershed. We can obtain lakes polygon data (i.e. GIS vector layer of lakes) from different online repositories or from local agencies. Different databases including Shuttle Radar Topography Mission's (SRTM) Water Body Data (SWBD) (http://dds.cr.usgs.gov/srtm/version2_1/SWBD/), WWF Global Lakes and Wetlands Data Base (GLWD) (<https://www.worldwildlife.org/pages/global-lakes-and-wetlands-database>), and USGS HydroSHED (<https://www.hydrosheds.org/>) can be used to download lakes data. Alternatively, we can create lake polygons by tracing over a base-map or google map in a GIS. For the illustration purpose, we are going to use GLWD data. Let us follow these steps:

- 1.1.1 **Saving and organizing data files:** You will be working with many different types of files so it is important to organize and manage your files well. For this purpose, create a workspace (i.e. a folder) in the Desktop, Documents folder or in a flash-drive. Within the folder, you can create subfolders to save files for different tasks of the exercise as needed. As an example, let's create a folder something called 'Basin_Delineation' in Documents by clicking Start Menu (i.e. the icon on the very left bottom of your computer) > File Explorer > Documents > New Folder (or Right click on the white space > New > Folder) > name it Basin_Delineation. Create subfolders as needed such as 'Lake' and 'Elevation' within the folder, or you can create additional ones as you go through.
 - 1.1.2 Click the above GLWD link.
 - 1.1.3 Click 'Access Level 1 Data' link. (You can choose Access Level 2 Data if your lake of interest is of comparatively smaller size.)
 - 1.1.4 Click 'Download' link > save the zipped file in your workspace (i.e. in the 'Lake' folder > unzip it by using 7-zip or Extract All tool.
- (Note: steps 1.1.2 – 1.1.4 are optional as we can alternatively use a basemap to locate the 'lake of interest' in QGIS)

Next, we will download the Digital Elevation Model (DEM) data. (If you already have the DEM, you can directly go to Step 2.2.5). Depending on the tentative size of the basin, we can choose elevation data of different spatial resolutions (i.e. pixel/cell sizes). There are multiple sources. We can obtain high-resolution DEMs from local sources. Alternatively, we can download from global online sources such as SRTM data (<http://earthexplorer.usgs.gov/>) and Aster GDEM (<http://gdex.cr.usgs.gov/gdex/>). Let us use the USGS website for our example. We will take Lake Volta – a West African lake – for illustration purpose.

- 1.1.5 Go to <https://earthexplorer.usgs.gov/>
- 1.1.6 Create an account with USGS, and log in.
- 1.1.7 Click on 'Search Criteria' tab on the top-left side of the page.
- 1.1.8 Click Geocoder and Polygon tabs below if they are not already selected.
- 1.1.9 Zoom into your target area (i.e. tentative basin).
- 1.1.10 Click and create multiple points on the map canvas to cover the basin area. In Figure 2a you can see a tentative drainage area for Lake Volta (in red color).

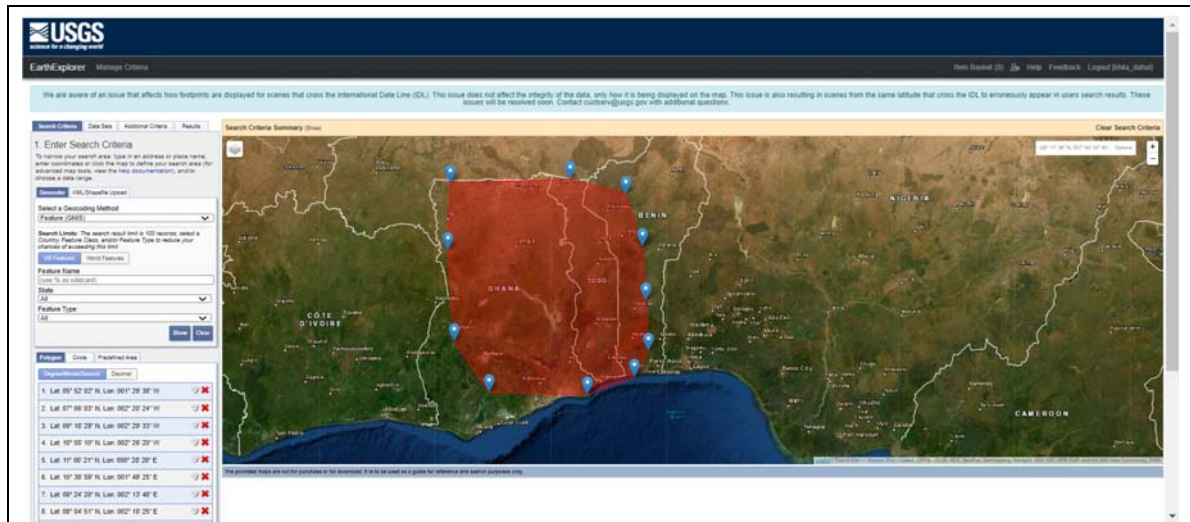


Figure 2a. Tentative drainage area for Lake Volta

- 1.1.11 Click 'Data Sets' tab next to 'Search Criteria' to select the proper DEM > expand Digital Elevation link > choose an appropriate DEM to download. (In this illustration, GMTED2010 is selected as it is already void-free, has different resolution options, requires comparatively smaller storage space and hence is faster for processing)
- 1.1.12 Click on 'Results' tab on the top of the table of content window. This selects four elevation rasters that cover the area as shown on the left hand side of Figure 2b.

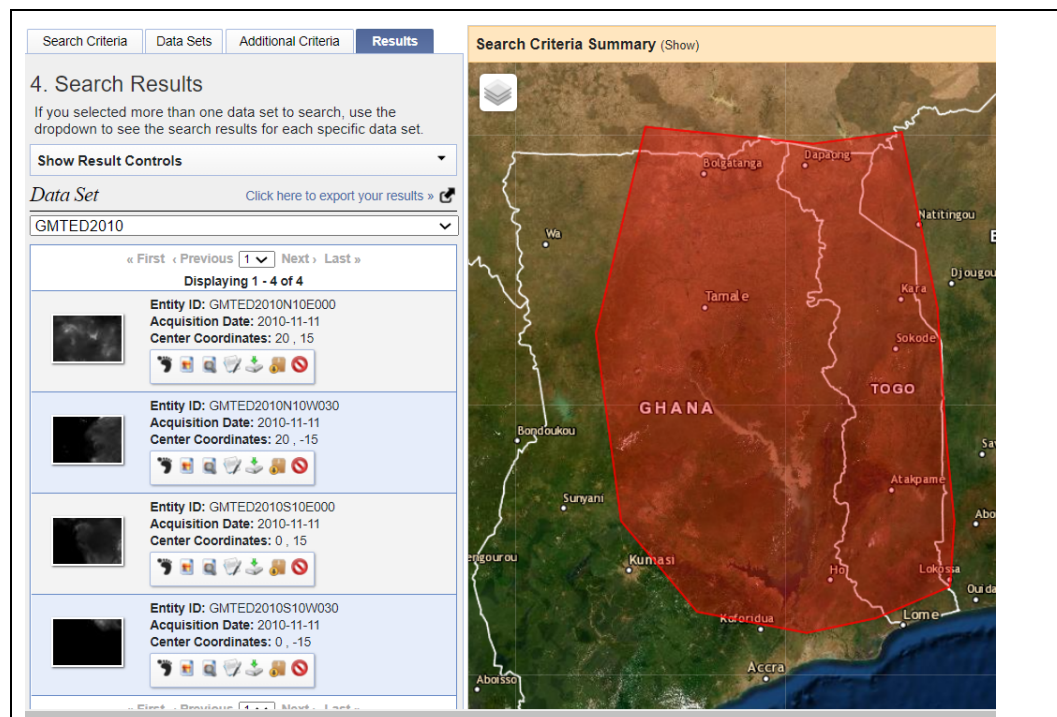


Figure 2b. Selected four elevation rasters covering Lake Volta drainage basin

- 1.1.13 Download the four image tiles one by one and save in your workspace (e.g. 'Elevation' folder) > unzip them.

For further details on how to download elevation data from USGS webpage, you can refer to YouTube videos such as this: <https://www.youtube.com/watch?v=A4ZEMqPXmEw>.

2.2 Delineation of lake basins


Once we have downloaded the necessary datasets, our next step is to add the datasets we downloaded into QGIS and run different Hydrology tools to create the lake drainage basin and stream networks. As mentioned earlier, we will use SAGA tools from within the QGIS package. Therefore, the first step is to open QGIS interface.

2.2.1 Adding data layers:

- Open QGIS
- Add the GLWD-Level1 layer, i.e. global lakes (this step is optional)
- Add the four GMTED data (i.e. the DEM) layers downloaded earlier one by one (by choosing 'mean' elevation surfaces for better results.)

After you have added the data, the interface should look something like Figure 3.

BTW: It is a good habit to save the project (i.e. the QGIS or the .qgz file) and continue clicking the Save button time to time. It is useful in case your computer unexpectedly shuts down, the QGIS crashes, or you want to stop now and want to continue later. For this,

- Click Project tab from the Menu bar > Save (or click on Save Project icon directly).
- Navigate to your work-folder (e.g. 'Basin_Delineation')
- Give a descriptive name such as 'basin_project_file'. This saves your .qgz file.
- Continue hitting the Save () button time to time throughout the exercise.
- If you have to open the same project again after closing, just go to your workspace and double click the project file (e.g. basin_project_file.qgz), and it opens.

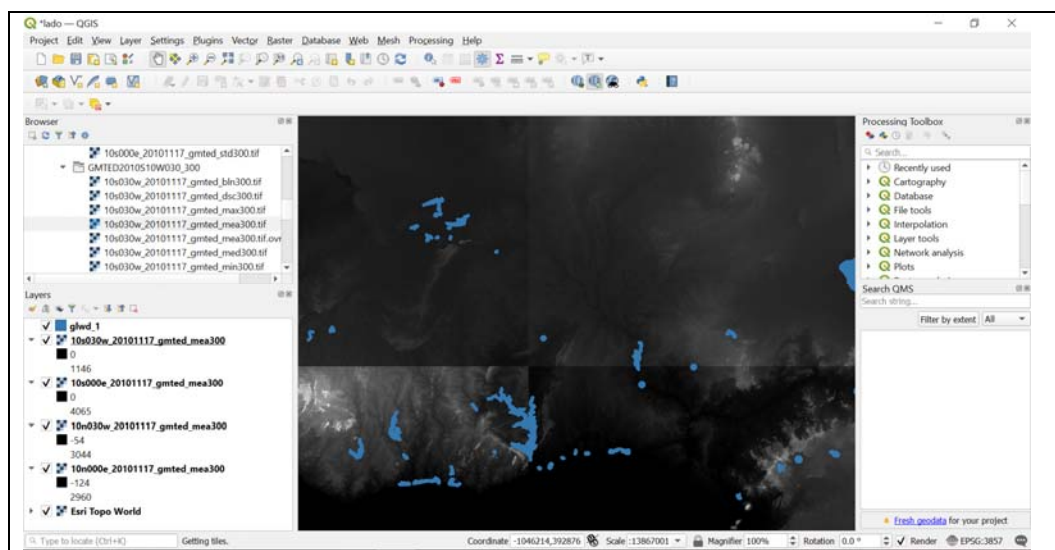



Figure 3. QGIS interface with GLWD and GMTED data added

2.2.2 Extracting the lake of interest (optional step)

As our focus is on Lake Volta, we want to make a layer that shows only Lake Volta by

selecting it from the GLWD data.

- a. Select Lake Volta from the GLWD layer by opening the attribute table.

Alternatively, click 'Select by Features by area or single click' icon () > zoom in to Lake Volta polygon and click. The lake polygon gets selected.

- b. Right click on the GLWD layer > Export > Save Selected Features As > select any format (e.g. ESRI Shapefile)
- c. In the resultant dialog box, navigate to your workspace (e.g. 'Lake' folder) and give a file name such as lake_volta > OK.

This creates a new layer just for Lake Volta.

2.2.3 Merging the DEMs

Next, we need to mosaic (or merge) the four DEM tiles into a single layer. To mosaic the four GMTED tiles:

- a. Click Raster from the Menu bar > Miscellaneous > Merge.
- b. In the Merge dialog box, click the three dots under Input Layers (the red highlight in Figure 4 below) > select the four GMTED layers.
- c. Save to your workspace by giving a descriptive name such as merged_dem under the Merged parameter.
- d. Hit OK with all other default parameter values.
- e. Once the processing completes, it creates a mosaicked layer. It looks something like Figure 5 given below.

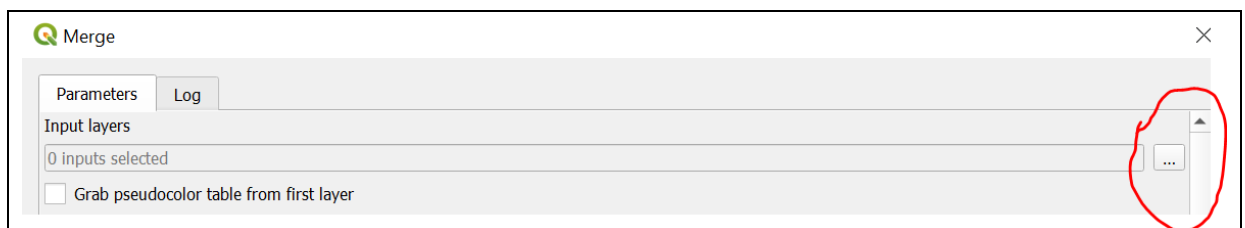


Figure 4. The top part of Merge dialog box

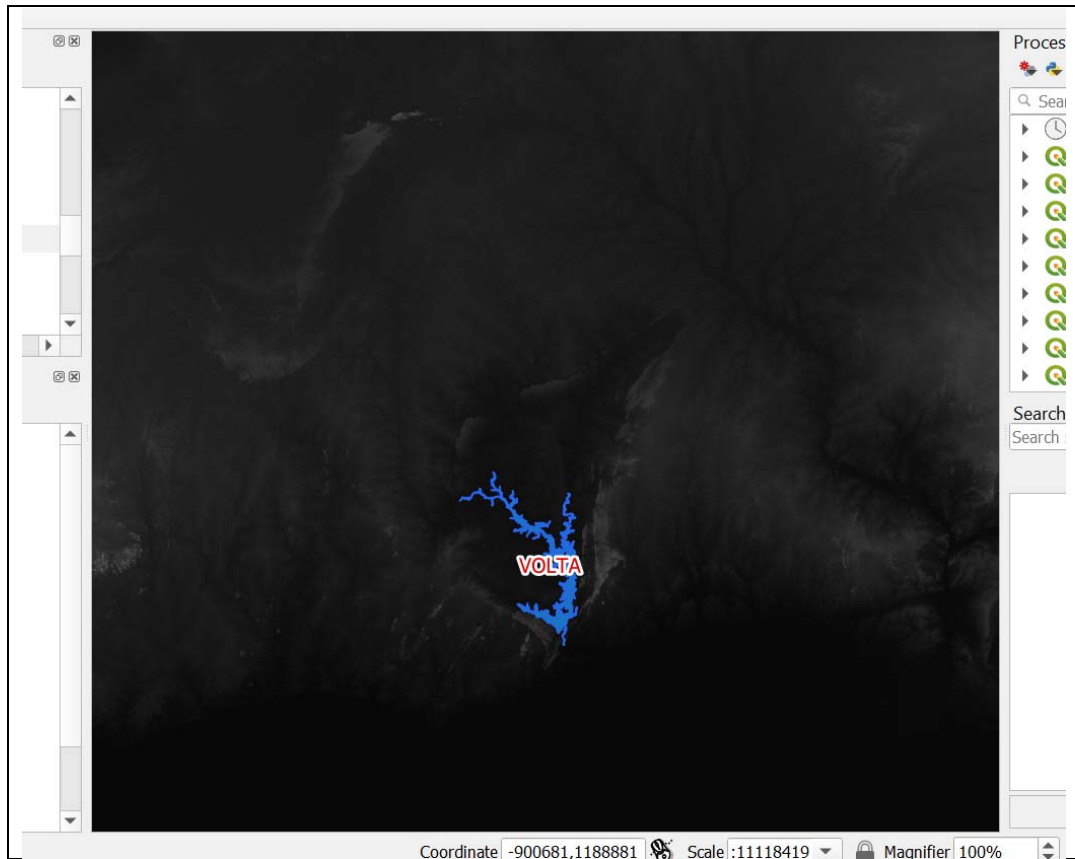


Figure 5. Mosaicked (merged) DEM layer

2.2.4 Slicing the raster

As the mosaicked raster covers a large area and so takes longer processing time in upcoming steps, we want to downsize it by clipping to focus only in the Lake Volta area.

- a. Click Raster > Extraction > Clip Raster by Extent
- b. In the Clip Raster by Extent window, input mosaicked layer > click Clipping extent dropdown arrow > select Draw on Canvas > draw a box that covers the tentative drainage basin area > leave all other parameters as defaults > save the Clipped (extent) to your workspace. The window with necessary parameters should look like Figure 6.
- c. Run the tool. The resultant clipped elevation raster appears something like the one that is at the center of Figure 7.

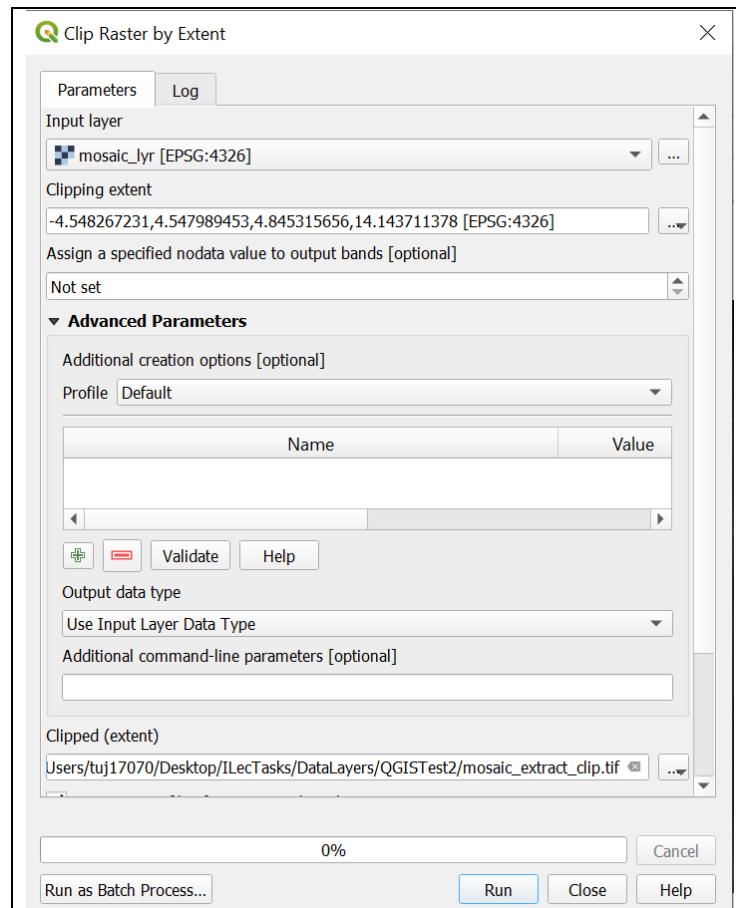


Figure 6. Clip Raster by Extent dialog box

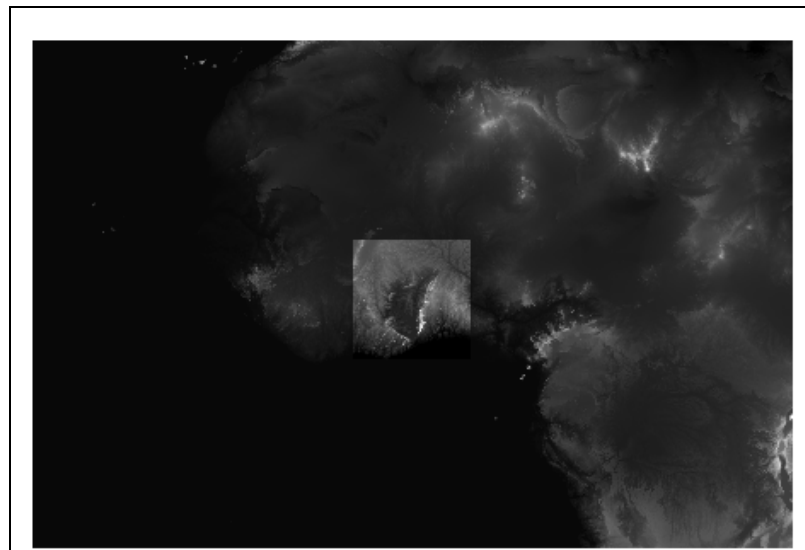


Figure 7. Clipped elevation raster

2.2.5 NOTE: if you already have a DEM for your area from the local sources or downloaded from any other websites, you can start from this step; there is no need to go through the above steps dealing with the DEM. You can just open QGIS, add your elevation data and start to follow from Step 2.2.6.

(Optional: Sometimes for better results and smooth functioning of the tools in QGIS, it may require that we project (convert) both the lake and DEM layers to an appropriate coordinate reference system (CRS) that is accurate for the local region. For Lake Volta

region, we will use UTM Zone 30N. There are multiple ways to project the layers, but we will export these layers to project. Right click the mosaicked raster (i.e. the DEM) > Export > Save As > in the Save Raster Layer As dialog box, navigate to your folder and name the projected raster properly > select UTM Zone 30N > leave other parameters as defaults and click OK. Similarly, project the Lake Volta layer to UTM Zone 30N.)

2.2.6 Starting the actual delineation procedure

Now, let us start running different Hydrology tools (or functions) such as Fill and Basin. However, skip this step (i.e. a-d points here) if you did not project the raster in the previous step.

- a. Close the current QGIS Project or the QGIS file (you can save it if you like).
- b. Open a new empty QGIS (project)
- c. Add the (projected) mosaicked layer, i.e. the DEM and the (projected) Lake Volta layer.
- d. Save the project with a new name such as lake_volta_tutorial. Continue saving from time to time.

2.2.7 Hydrologically correcting the DEM

- a. The DEM thus prepared typically may contain a good number of sinks or depressions, which are single or multiple pixels surrounded by pixels of higher elevation. Some of these sinks are naturally occurring landscape features, representing closed (inland) basins with no outlets. Others are spurious, often produced during the process of DEM production. The spurious sinks are critical problems in hydrological applications, since they interrupt continuous flow across the DEM surface. Thus, it is essential to remove the anomalous sinks. For this, we use Fill Sinks tool.
- b. Click Processing from the Menu bar > Toolbox (if the Processing Toolbox window is not already open in the right hand side of QGIS interface).
- c. In the Processing Toolbox window, scroll down > expand SAGA > click Terrain Analysis – Hydrology > Fill sinks (wang & liu)
- d. In the resultant dialog box, fill out the parameters accordingly (i.e. DEM should be the projected mosaicked DEM > save the output Filled DEM to your workspace > give a descriptive name > uncheck the boxes for Flow Directions and Watershed Basins and leave the default value for Minimum Slope) > Run.
- e. The tool window should look something like Figure 8.

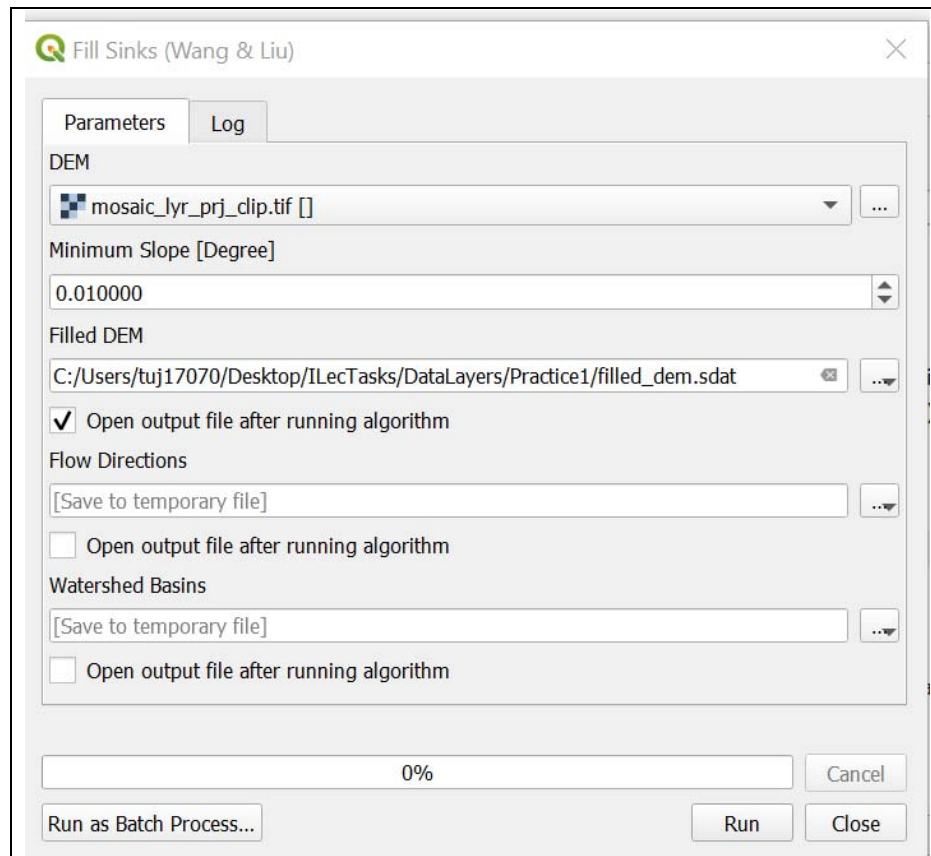


Figure 8. Fill Sinks dialog box

2.2.8 Running Hydrology and Channels tools

This hydrologically conditioned (i.e. with all the sinks filled) elevation surface is used as input for different other Channels and Hydrology tools. As our main purpose is to create a watershed that drains through the outlet of Lake Volta, we will use Channels and Hydrology toolboxes from SAGA.

- a. Expand Terrain Analysis – Channels toolbox > Channel network and drainage basins.
- b. In the Channel Network and Drainage Basins window, input the filled_dem layer under Elevation > keep 5 for Threshold > uncheck the boxes for Flow Direction and Flow Connectivity > save to your location the Strahler Order surface with an appropriate name > uncheck the box for Drainage Basins > save to your location the Channels as a Shapefile layer > uncheck the boxes for Drainage Basins and Junctions > Run.
- c. The operation generates two layers: Strahler Orders (a raster layer) and stream network (a vector layer). The result looks something like Figure 9. You can refer to the tool documentation pages for details on how the operations work.

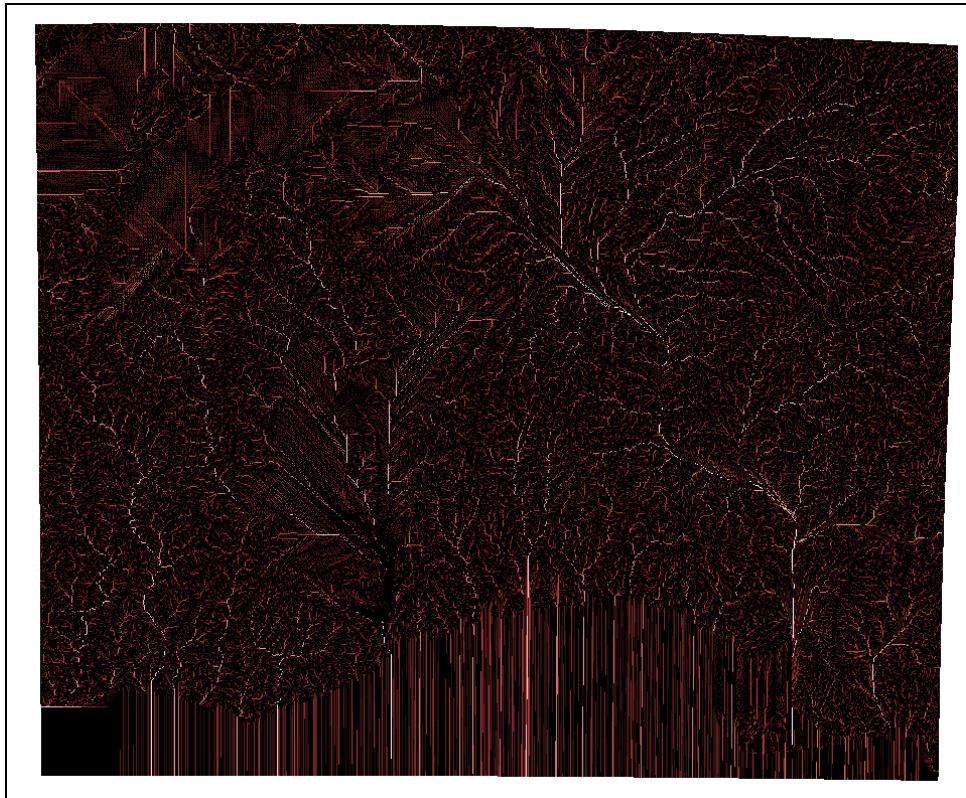


Figure 9. Strahler Orders and stream network

2.2.9 Pinpointing the pour point

- a. Now, with the help of the Volta Lake layer (you can use the base-map if you haven't created the lake layer) and the Strahler Orders layer, we need to pinpoint the pour point at the lake outlet through which all the water from the basin drains away.
- b. To add a base map, click Web > Quick Map Services > select any option such as OSM and then OSM Standard. (**Note:** If Quick Map Services link is not active (i.e. if the base-maps are not available) in your QGIS, you need to add this plug-in. For this: click Plugins > Manage and Install Plugins... > Search 'quick' in the search area of the 'Plugins' window > select Quick Map Services > Install Plugin > Close. Now, the base-maps are available. If it doesn't work, please go to this link and follow the steps to activate/add the plug-in. The link is: <https://opengislab.com/blog/2018/4/15/add-basemaps-in-qgis-30>. **Or**, simply google 'how to open a basemap in qgis' and watch one of the quick/short video tutorials.)
- c. Zoom to the Lake Volta and its dam outlet (you can take help from the base map) > identify the pour point (you need to turn on and off the Strahler layer or make the lake layer hollow).
- d. Identify/pinpoint the brightest pixel (i.e. a white pixel that indicates high flow accumulation) on the Strahler layer that is the closest to the pour point. Remember these white (high accumulation) cells make the river route/path.
- e. Since our purpose is to capture the coordinate values of this location/pixel, right click on the identified bright pixel (which is on the river path) > hover the cursor over Copy Coordinates > click the appropriate coordinate (e.g. UTM zone 30 N)
- f. Paste in a notepad or any other document/place because we need these coordinate (X,Y) values when we run the Upslope Area tool, which captures all the pixels

(area) that drain through this pour pixel to determine the spatial boundary of the basin.

2.2.10 Creating the basin (upslope area)

- a. To create the drainage basin, go to Terrain Analysis – Hydrology > Upslope area
- b. Complete parameterization in this way: leave Target Area as it is > copy and paste the X coordinate value from the notepad under Target X coordinate > Y coordinate value under Target Y coordinate > filled_dem under Elevation > keep Sink Routes untouched > select Deterministic 8 under Method > leave the default value for Convergence > save the Upslope Area raster into your workspace > Run.
- c. The completed Upslope Area window should like this.

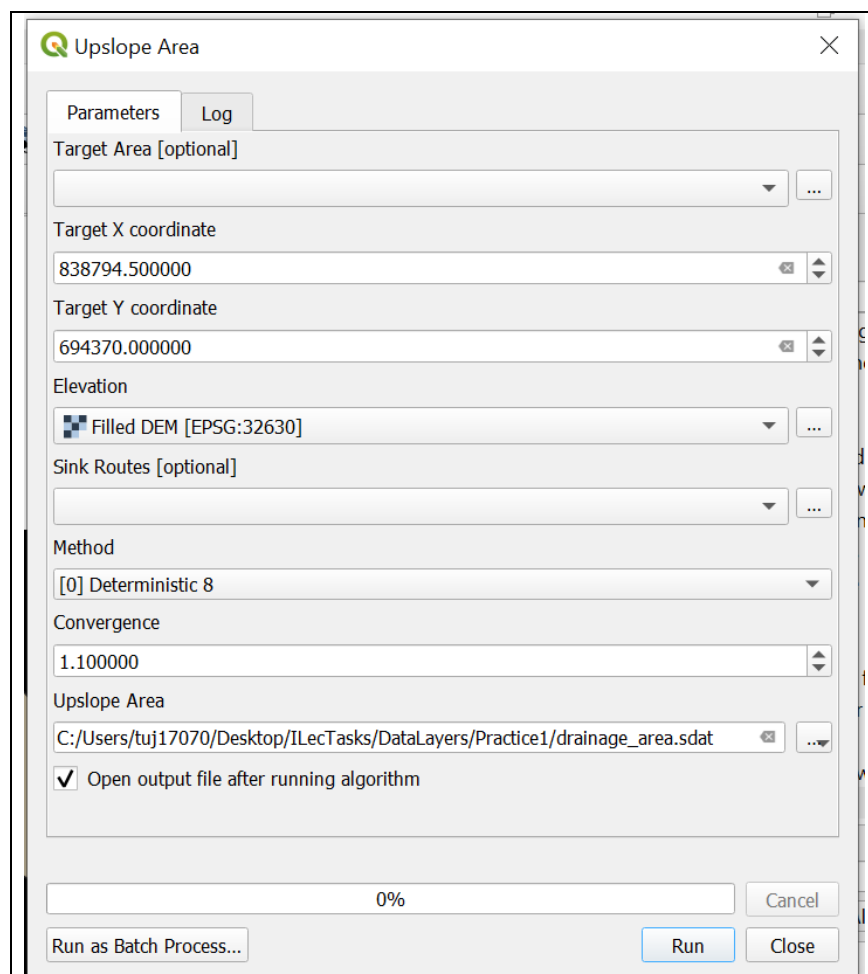


Figure 10. Upslope Area dialog box

When the processing completes, the Lake Volta basin area and river channels look something similar to the ones shown in Figure 11.

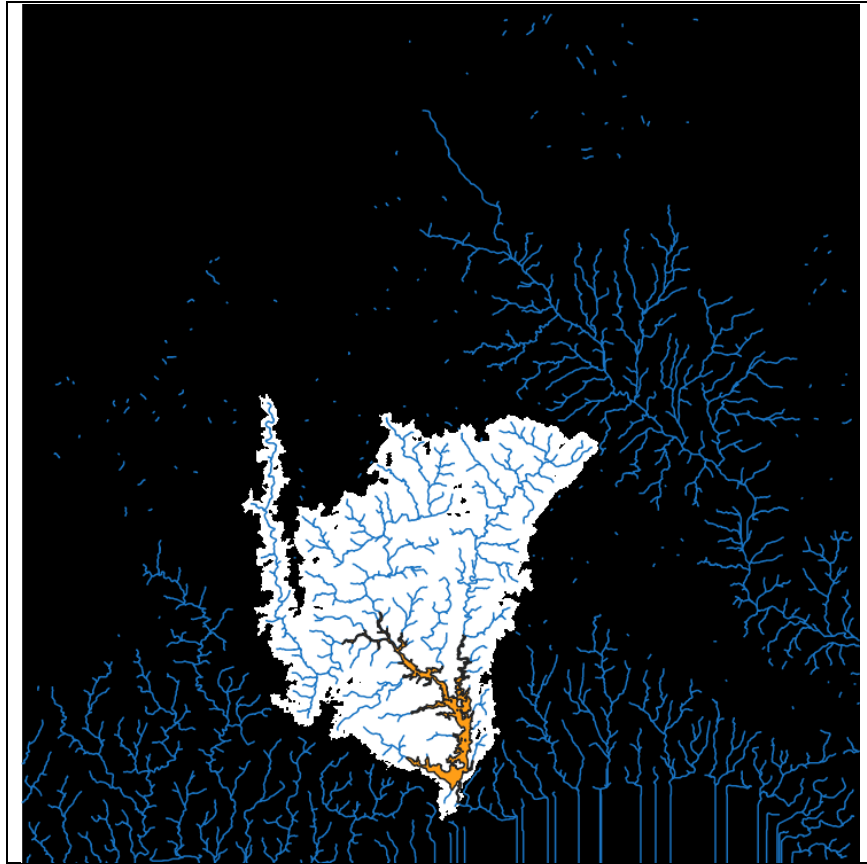


Figure 11. Lake Volta basin and river channels

2.2.11 Vectorization of the resultant basin layer

- We can convert the upslope (i.e. basin) layer in to a Shapefile by clicking Raster > Conversion > Polygonize (Raster to Vector).
- In the resultant vector layer, select only the basin polygon by using attribute table or simply using Select Features button > right click on the layer > Export > Save Selected Features As > complete the parameters in the resultant window > save it as a Shapefile with an appropriate name in your workspace.
- Then, we can clip the river channels only within the basin boundary for a better visualization. Search for 'polygon clipping' from the Processing Toolbox > double click and open the Polygon Clipping
- In the Polygon Clipping window, input the vectorized upslope area layer as the Clip Features and channels as the Input features > save it to your folder.

2.2.12 Final visualization

Finally, we can visualize the basin and river channels by using appropriate color scheme, transparency etc. and over the hillshade (as shown in Figure 12) or on a base-map (as shown in Figure 13).

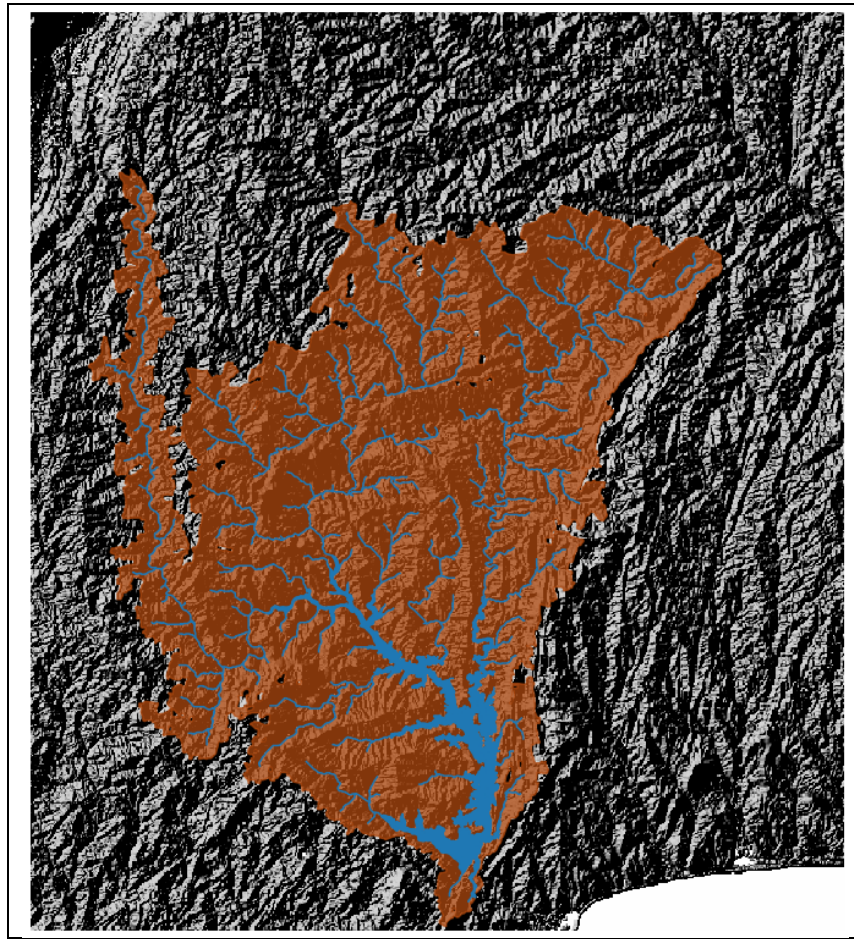


Figure 12. Lake Volta basin visualized on a hillshade

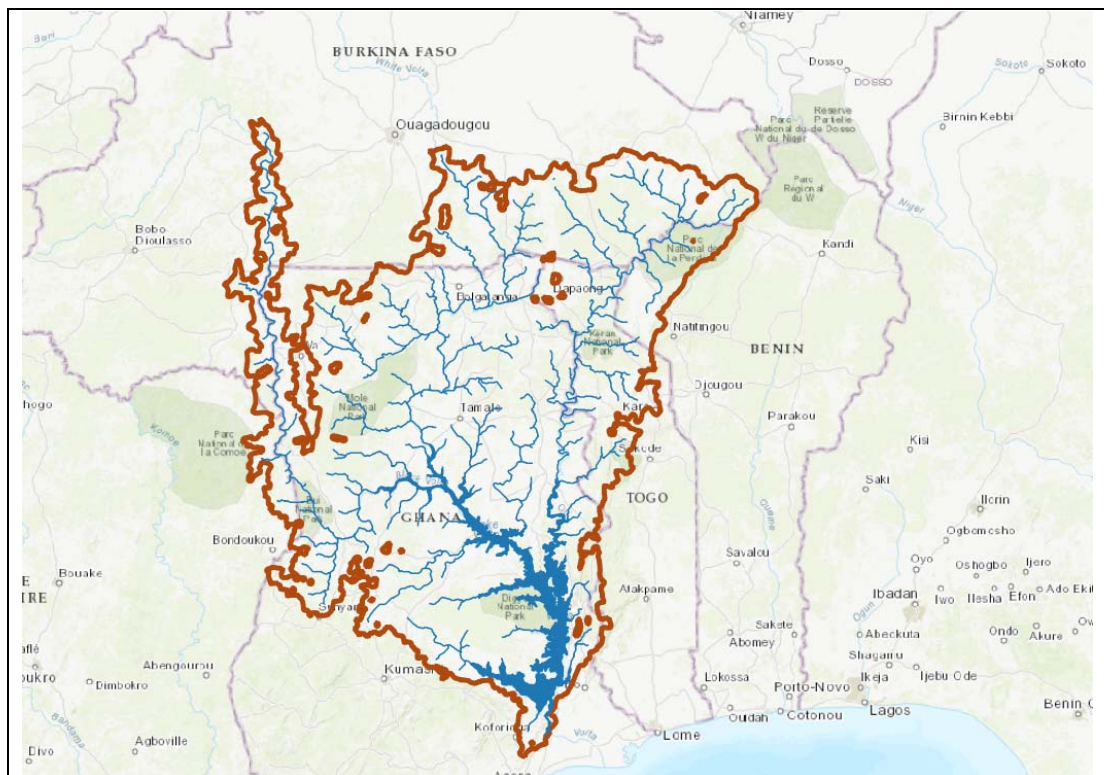


Figure 13. Lake Volta basin visualized on a base map

3. Conclusion:

The quality (specifically the positional accuracy) of GIS outputs is very important because the results are often used as decision-support tools. Therefore, it is always a good idea to make an accuracy assessment of the modelling outputs. The results can be statistically compared with the reference data by using different numerical measures or indices (e.g. RMSE) to determine the accuracy. It is also customary to make a visual comparison of the GIS outputs (mostly in the form of maps) with the reference data (if available). Figure 14 presents a visual comparison of the Volta Lake basin created as part of this tutorial with the same lake boundary but generated by using different other datasets and software. Figure 14a shows our result. Figure 14b shows the product of a finer-resolution DEM from SRTM source and by using ESRI' ArcMap tools. Figure 14c uses additional ancillary data such as field observation points and information derived from the mapmaker's familiarity with the local geography.



Figure 14: Volta basin map a) created as part of this tutorial, b) generated in ArcMap using SRTM elevation data, and c) authored by Ndehedehe et al. (2017).

As illustrated in the figure, the overall accuracy of the delineated lake basin boundary depends on multiple factors including the quality and resolution of the DEM, the algorithms of the tools, and user dexterity (e.g. identification of the right pixel to use as the pour point). Thus, the users should be careful in the selection of the DEM and available GIS tools. Sometimes, it is useful to run different versions of the same tool (e.g. there are three different versions of the Fill Sinks tool in SAGA toolbox) to see which might result comparatively a better output. Users can also play around with the tools' functioning and observe outputs by changing different parameter values. In addition, learners should start developing a habit of referring to the tool help link or the tool documentation pages from the web to understand the proper functioning.

Thus, this tutorial walks the learners through the generic steps of a GIS-based hydrological modeling procedure to generate a lake drainage basin boundary and stream networks. There are other relevant steps and tools as well that are worth exploring to achieve specialized goals. As this document provides the most basic and a common procedure to achieve the goal of basin delineation, individual needs and issues may require additional effort on the part of the learners to do some internet 'searching' and some individual 'troubleshooting'.

ACKNOWLEDGEMENTS

Special thanks go to the Japan ILEC team, particularly Dr. Victor Muhandiki for the constructive comments and feedback, which were vital to enhance the quality of this document.

REFERENCES

- Anornu, G., Kabo-bah, A., and Kortatsi, B. 2012. Comparability studies of high and low resolution digital elevation models for watershed delineation in the Tropics: A case of Densu River Basin of Ghana. *International Journal of Cooperative Studies*, Vol 1, No 1, 9-14.
- GIS Geography, 2020. 27 Differences Between ArcGIS and QGIS – The Most Epic GIS Software Battle in GIS History. Accessed October 25th, 2020 from <https://gisgeography.com/qgis-arcgis-differences/>
- Hengl, T., Grohmann, C., Bivand, R., Conrad, O., and Lobo, A. 2009. SAGA vs GRASS: A comparative analysis of the two open source Desktop GIS for the automated analysis of elevation data. *Proceedings of Geomorphometry*, Zurich, Switzerland, 31 August-2 September, 2009. Accessed October 27th 2020 from <http://www.geomorphometry.org/system/files/hengl2009geomorphometry.pdf>
- Menke, Kurt, 2017. At least 10 reasons you should be using QGIS. Accessed October 25th 2020 from <https://www.birdseyeviewgis.com/blog/2017/5/1/at-least-10-reasons-you-should-be-using-qgis>
- Ndehedehe, C., Awange, J., Kuhn, M., Agutu, N., and Yoichi, F. 2017. Analysis of hydrological variability over the Volta river basin using in-situ data and satellite observations. *Journal of Hydrology: Regional Studies*, Volume 12, Pages 88-110
- QGIS Blog, 2018. Ten Advantages of QGIS software. Accessed October 1st 2020 from <https://grindgis.com/software/10-advantages-of-qgis-software#:~:text=Access%20and%20Freedom%3A%20one%20of,use%20on%20their%20personal%20computers.>