

CEQUEAUphysiography Toolbox: Workflow guide

Stephen Dugdale, 2015-06-05

1. Introduction

The CEQUEAUphysiography Toolbox is designed to speed up and automate the extraction of flow routing and land use data necessary to run the CEQUEAU model. In the past, the extraction of physiographic data necessary to run the CEQUEAU model was relatively laborious, particularly with regards to computing the flow routing between neighbouring whole squares (CEs) and partial squares (CPs) which was done by hand. The CEQUEAUphysiography Toolbox uses a combination of ArcGIS and Matlab to complete this task with a minimum of user input, allowing for the physiography data to be extracting relatively simply. The purpose of this workflow guide is to provide the user with a step-by-step guide to generating the CEQUEAU 'bassinVersant' structure using the CEQUEAUphysiography Toolbox.

2. How it works

The CEQUEAUphysiography Toolbox requires data from Arc Hydro Tools in order to replicate the steps used to manually calculate flow routing. First, Arc Hydro Tools is applied to a digital elevation model (DEM) of the watershed in order to calculate the flow direction between each DEM pixel and its neighbouring pixels based on their relative elevations, resulting in a raster grid of flow directions. Next, a flow accumulation raster is computed by assigning each pixel a value based on the cumulative number of upstream pixels that drain into it as a function of the flow direction raster. This process essentially delineates river channels present within the DEM, as rivers and streams will inevitably comprise highest flow accumulation values. The flow accumulation raster is next used to create a raster of watershed sub-basins by finding all pixels upstream of each point at where the flow accumulation jumps by a user-defined threshold.

The resulting flow accumulation and sub-basin rasters are then loaded into Matlab and used as inputs to the CEQUEAUphysiography.m function which automates the partial square flow routing process. First, the sub-basin raster is intersected with a shapefile grid representing the whole squares (CEs) in order to divide each whole square into partial squares. Next, partial squares having a surface area smaller than a pre-defined threshold (eg. 1% of whole square size) are merged with their neighbours in order to ensure a maximum of four partial squares per whole square. Flow routing is subsequently achieved using the flow accumulation raster, whereby the outlet of each partial square is identified based on a partial square's maximum flow accumulation value. The location of these outlet points allows for the determination of a partial square's downstream neighbour, defined as the new partial square into which the outlet point of the first partial square drains. Through recording the downstream neighbour of each partial square, the downstream route between each partial square and the watershed mouth is recorded. A recursive algorithm is then

applied to this data to find all partial squares upstream of each partial square within the watershed, allowing for the tabulation of the flow routing data necessary to run CEQUEAU.

For each whole and partial square, the CEQUEAUphysiography Toolbox also calculates other required physiographic parameters. The percentage area of each partial square within each whole square is computed as a function of the raster pixel dimensions. Partial square and whole square elevation is computed from a DEM of the watershed. Partial and whole square land cover is achieved through intersecting each square with a land-use raster of the study area. Lake and wetland cover is computed by intersecting each whole/partial square with vector polygons describing water bodies and wetlands. CEQUEAU also requires values comprising the cumulative area of each of these parameters upstream of each partial square; these are achieved using the automatic flow routing table to sum the data for all partial squares upstream of a given point.

3. Toolbox location

The CEQUEAUphysiography Toolbox can be found on the SVN under /Steve/physiography. A readme file within this folder explains the purpose of each Matlab function and sub-function.

4. System requirements

- ArcGIS 9.1 or higher (although version 10.1 or higher is much faster)
- Arc Hydro Tools module (free download from here: <http://downloads.esri.com/archydro/archydro/Setup/>. Ensure that the Arc Hydro Tools version matches the ArcGIS version)
- Matlab 2014a or higher (may work on older versions but as yet untested)
- Matlab Mapping Toolbox
- Matlab Image Processing Toolbox
- 12 – 16Gb RAM (if working in large watersheds or with very high resolution DEMs)

5. Toolbox location

In order to run the CEQUEAUphysiography toolbox, you will need the following:

- SRTM ~30 m digital elevation model of watershed. Available for free download here: <http://eros.usgs.gov/>

- A copy of the North American Land Change Monitoring System raster dataset (250m resolution), which can be freely downloaded from here:
<http://www.cec.org/Page.asp?PageID=122&ContentID=2819>
- Shapefile of waterbodies (lakes, rivers) present within the watershed. Can be downloaded from a range of online sources (eg. Geogratis in Canada, NHD in the US, etc).
- Shapefile of marshlands/wetlands. Again, can be downloaded from a range of online sources (Various provincial sources in Canada, NHD in the US).
- Shapefile of the watershed extent.

6. Important considerations

- This workflow assumes that the user is familiar with the basics of ArcGIS, Matlab and the CEQUEAU hydrological model.
- This workflow is specific to rivers in North America because it requires use of the SRTM digital elevation dataset and the North American Land Change Monitoring System raster dataset. It will be necessary to modify the CEQUEAUphysiography toolbox functions should you need to use data from different sources.
- All data (shapefiles, rasters, etc) MUST be projected in the UTM coordinate system and be in the same UTM zone.
- All raster datasets (eg. DEM, land cover raster) must have EXACTLY the same spatial extent. This can be achieved by using the DEM to clip other raster datasets to the same dimensions using the ArcGIS 'Clip Raster' function: Data Management Tools -> Raster -> Raster Processing -> Clip. Be careful to ensure that the 'Maintain Clipping Extent' option is checked.
- All raster datasets should be in *.TIF format and have accompanying worldfiles which contain their georeferencing data. If your rasters do not have accompanying worldfiles, they can be created using the ArcGIS 'Export Raster World File' tool: Data Management Tools -> Raster -> Raster Properties -> Export Raster World File
- Because the Matlab tools for processing shapefiles are slow, it is often helpful to ensure that polygons in the waterbodies/wetlands shapefiles are composed of less than 500 vertices each. This can be achieved by using the ArcGIS 'Dice' function: Data Management Tools -> Features -> Dice, setting the 'Vertex Limit' field to 500. This will ensure that any polygons with large numbers of vertices will be split into several pieces. This will not affect the output of any land use calculations for CEQUEAU.

- For more detailed instructions on the Arc Hydro Tools elements of this workflow, please refer to Venkatesh Merwade's useful tutorials here:
<http://web.ics.purdue.edu/~vmerwade/tutorial.html>
- The CEQUEAUphysiography Toolbox does not currently assemble the data necessary to run the 'Qualité' component of the CEQUEAU model which are normally found in the 'bassinVersant' structure. However, this functionality will be added in a later release.
- The CEQUEAUphysiography Toolbox does not currently assemble any of the other data pertaining to the CEQUEAU structure (eg. 'meteoStation', 'parametres', etc). At the current time, this must be accomplished manually.
- Although the CEQUEAUphysiography Toolbox has been thoroughly tested on several watersheds, it is likely that bugs may still be present. It is therefore advisable to have a firm understanding of Matlab, in order to be able to solve any problems that arise during the use of the toolbox.
- Despite extensive testing, it is still possible that errors (eg. flow routing, land use, etc) might arise during the extraction of the data. It is therefore necessary to closely inspect the data contained within an automatically extracted 'bassinVersant' structure to ensure that all values make sense and are within realistic 'real world' limits.

7. ArcGIS raw data preparation workflow

Step 1. Load the DEM, land cover raster, watershed extent shapefile, waterbodies shapefile and wetlands shapefiles into ArcGIS.

Step 2. Ensure that all files are in UTM projection, WGS84 datum and within the same UTM zone.

Step 3. Reduce the vertex count of the waterbodies and wetlands shapefiles using the ArcGIS 'Dice' function: Data Management Tools -> Features -> Dice, setting the 'Vertex Limit' field to 500. This will dramatically speed up the processing when using the CEQUEAUphysiography Toolbox in Matlab.

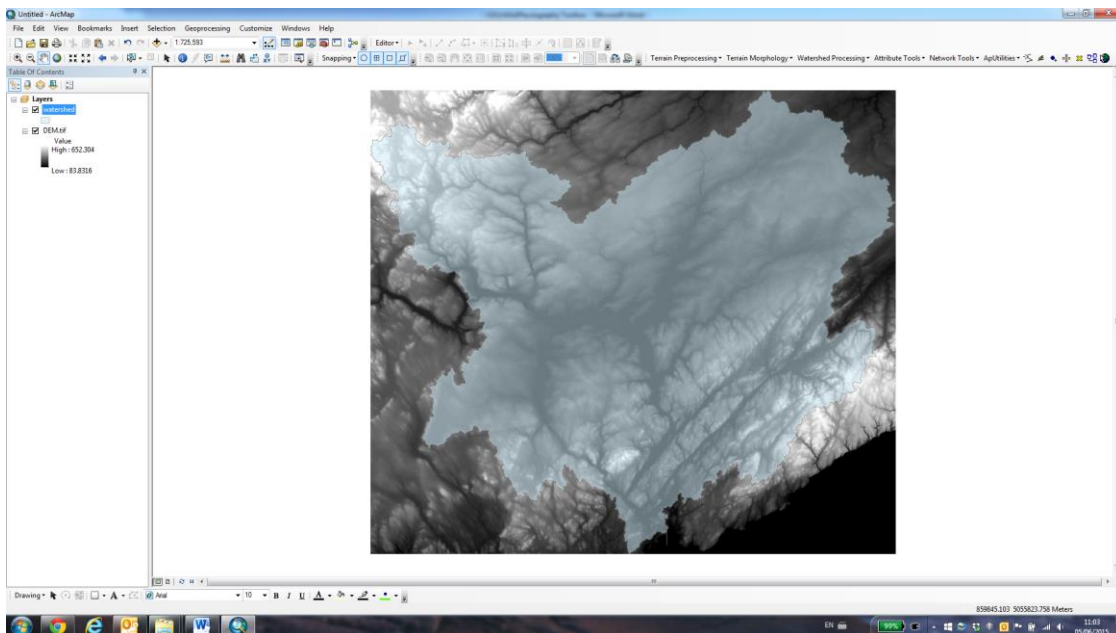
Step 4. Clip the DEM by the watershed extent shapefile using the ArcGIS 'Clip Raster' function, selecting the watershed extent shapefile as the 'Output Extent': Data Management Tools -> Raster -> Raster Processing -> Clip. This will ensure that the DEM is the smallest size possible, thus speeding up subsequent processing.

Step 5. Next, clip the land cover raster by the newly clipped DEM, using the ArcGIS 'Clip Raster' function: Data Management Tools -> Raster -> Raster Processing -> Clip. Ensure that the 'Maintain Clipping Extent' option is checked so that the land cover raster covers exactly the same extent as the new DEM.

Step 6. Export all newly created files (Clipped DEM, clipped land cover raster, diced waterbodies shapefile, diced wetlands shapefile) to a new folder.

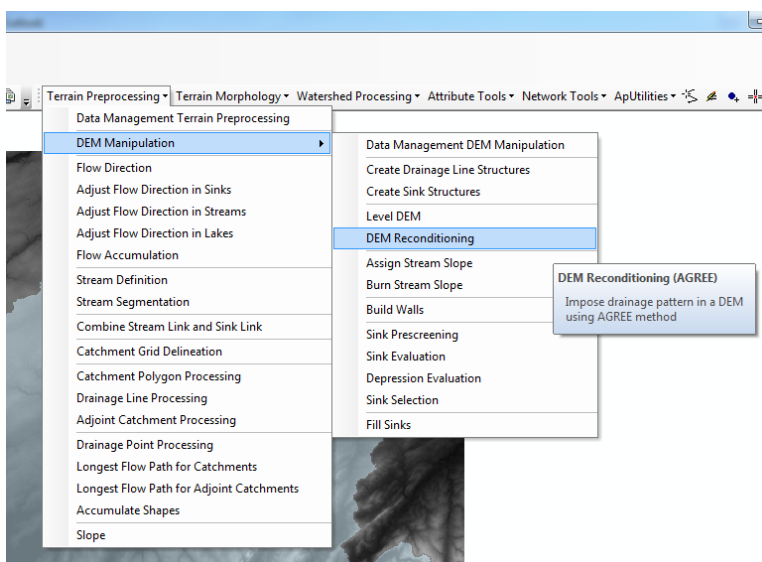
8. Arc Hydro Tools workflow – creation of flow accumulation and sub-basin rasters

Step 1. Load your DEM and watershed extent shapefile into ArcGIS:

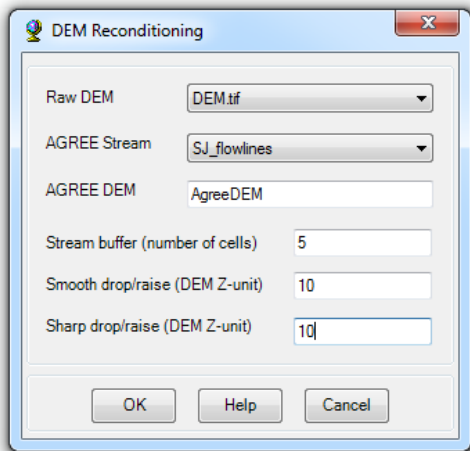


Step 2. Make sure that the Arc Hydro tools extension is visible by right-clicking on the ArcGIS toolbar and selecting 'Arc Hydro Tools'

Step 3. (Optional step). If you already have a detailed shapefile of the river network, you can use this data to 'burn' water courses into the DEM. This can (in some circumstances) improve the quality of the Arc Hydro tools-generated flow network. To do this, click on 'Terrain Preprocessing -> DEM manipulation -> DEM reconditioning':

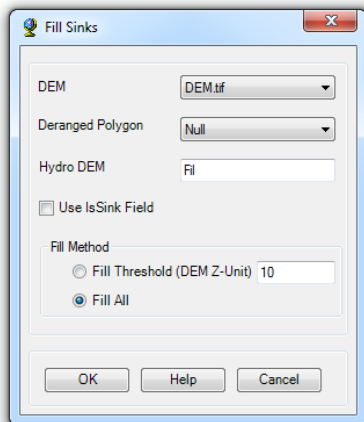


In the resulting dialogue box, specify your DEM file as the 'Raw DEM' and the river network shapefile as the 'AGREE Stream'. Ensure that the stream buffer is 5 and the Smooth and Sharp drop/raise values are both set to 10, and press 'OK' to run DEM reconditioning:



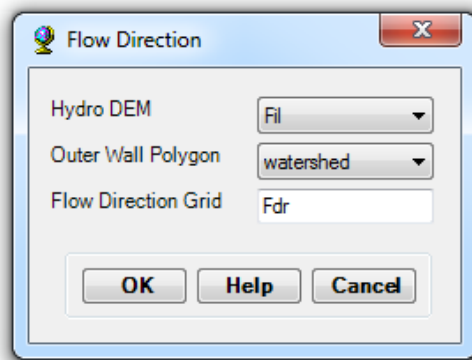
This will output a new raster called 'AgreeDEM' – this should be visible in the ArcGIS table of contents. *Remember that this step is optional and is not necessary to successfully run Arc Hydro tools.*

Step 4. We now need to ensure that the DEM is hydrologically correct (ie. it does not contain data holes or sinks). To do this, click on 'Terrain Preprocessing -> DEM Manipulation -> Fill Sinks':

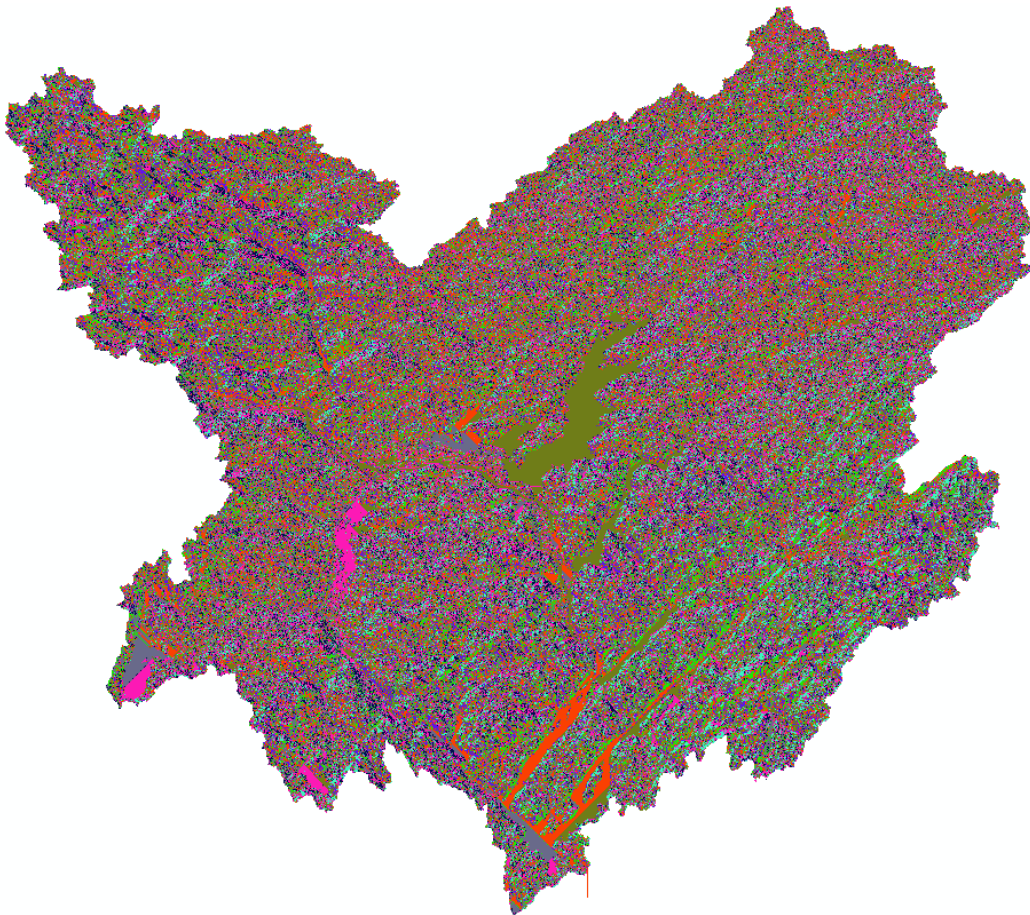


Depending on whether you completed the optional step 3, you will need to specify either your raw DEM or the AgreeDEM in the 'DEM' box. Leave all of the other values as they are, and press 'OK'. You will see that a new raster has appeared in the table of contents. It should be called 'fil'.

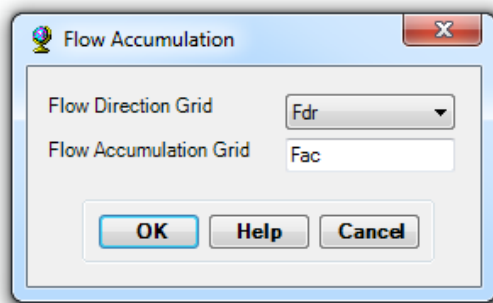
Step 5. Next, compute the flow direction raster by clicking on 'Terrain Preprocessing -> Flow direction':



In the 'Outer Wall Polygon', specify the name of the shapefile that contains the polygon of your watershed extent. Do not change any of the other values, then click on 'OK' to calculate the flow direction grid. This will output a new raster called 'fdr'. It should look similar to this:

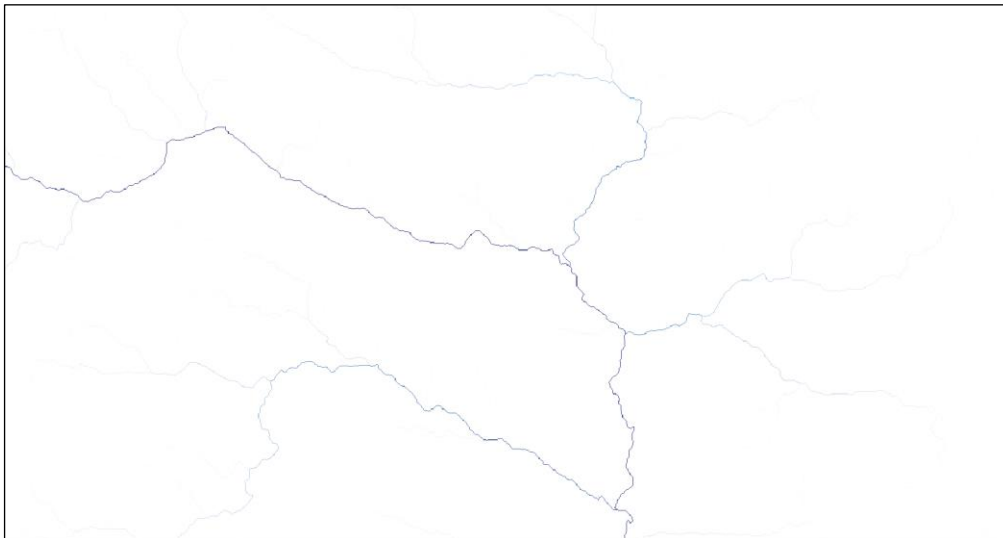


Step 6. Next, compute the flow direction raster by clicking on 'Terrain Preprocessing -> Flow accumulation':

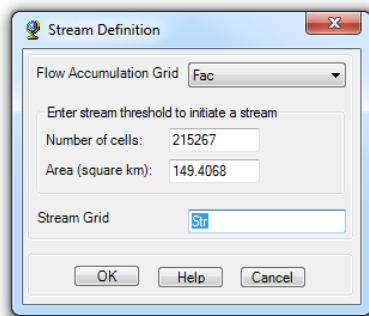


Don't change any of the fields - just press 'OK'. This will output a new raster called 'fac' which contains the flow accumulation data necessary to run the CEQUEAUphysiography toolbox in Matlab.

If you want, you can zoom in and inspect the raster to look at the flow accumulation data. The value of each pixel in the raster is equivalent to the number of pixels upstream of that particular point:



Step 7. Next, define a fixed stream network by clicking on 'Terrain Preprocessing -> Stream definition'. This step is very important and is what will govern the size of the sub-basins within your watershed (see step 9):



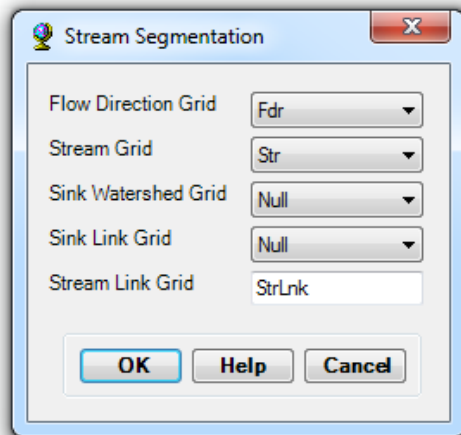
Here, you must enter a 'stream threshold' value. This value essentially defines the minimum size of sub-basin needed to create a stream, in km^2 . In essence, by defining a stream threshold, you are determining the size of the sub-basins that ArcGIS will generate. For a large watershed (eg. 20 000 km^2), you may want to choose a relatively large threshold (eg. 100 km^2). For smaller watersheds, this value can be much lower (eg. 10 km^2).

The stream threshold value will also be governed by the dimensions of your chosen whole squares (CEs). In order to ensure that there are no more than 4 partial squares (CP) per whole square, you must choose a 'stream threshold' value that does not result in too many small sub-basins.

Unfortunately, this is a trial-and-error process, and may take a few repetitions of steps 7 – 10 to ensure that the sub-basin size is appropriate.

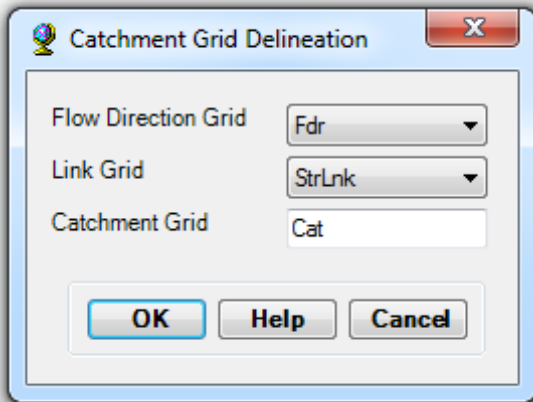
Once run, the stream definition tool will output a new raster called 'str'.

Step 8. The next step is to perform stream segmentation by clicking on 'Terrain Preprocessing -> Stream definition'. This stage essentially splits up the stream that you defined in step 7, creating a new segment each time the stream bifurcates:



Leave all of the fields as they are, and press OK. The resulting raster will be called 'StrLnk'.

Step 9. The final operation using ArchHydro tools is to define the watershed sub-basins using the 'Terrain Preprocessing -> Catchment Grid Delineation' tool:



Leave all of the fields as they are, and press OK. The resulting raster will be called 'Cat'. It should look similar to this (although it may be necessary to change the symbology to highlight the sub-basins in different colours. The results should be similar to this:



Step 10. Once the sub-basin raster is complete, you will need to create a grid of whole squares (CEs). This can be done using the ArcGIS 'Create Fishnet' tool: Data Management Tools -> Feature Class -> Create Fishnet.

Create Fishnet

Output Feature Class

Template Extent (optional)
Same as layer watershed

Left: 633073.831700, Top: 5160862.892800, Right: 806133.429000, Bottom: 5008378.656300

Fishnet Origin Coordinate
X Coordinate: 633073.8317000009, Y Coordinate: 5008378.656300003

Cell Size Width: 10000, Cell Size Height: 10000

Number of Rows, Number of Columns

Opposite corner of Fishnet (optional)
X Coordinate: 806133.4290000005, Y Coordinate: 5160862.8928

☒ Create Label Points (optional)

Geometry Type (optional): POLYGON

OK, Cancel, Environments..., Show Help >>

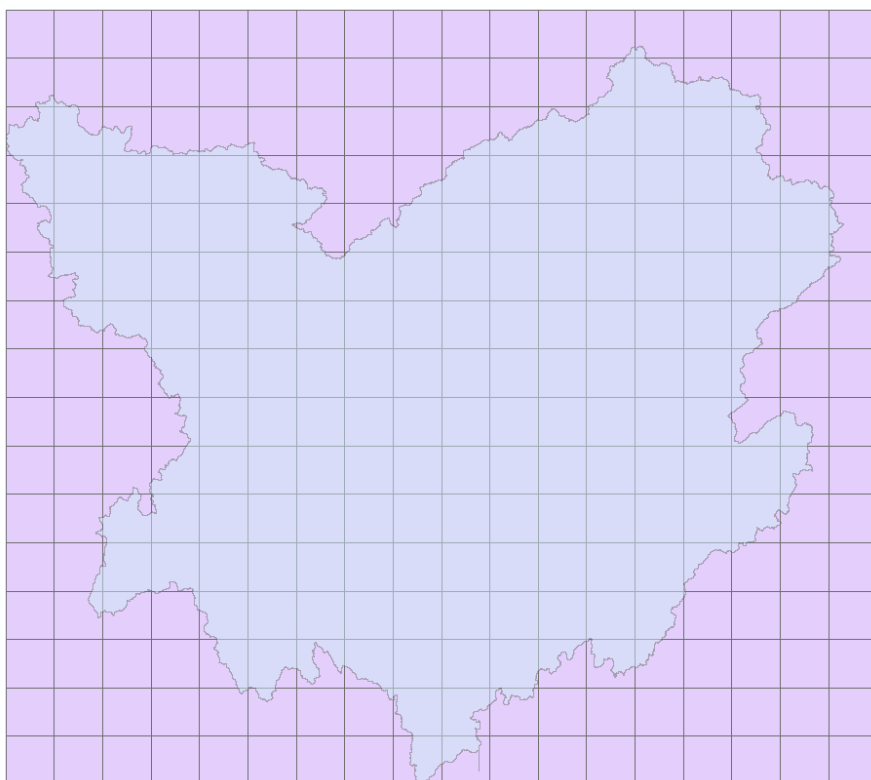
The output feature class should be saved in '.shp' format (not geodatabase), in the same folder where you saved the clipped DEM, land cover raster, waterbodies shapefile and wetlands shapefile.

For the 'template extent' field, select your watershed extent shapefile. ArcGIS will then automatically fill in the spaces for the fishnet coordinates.

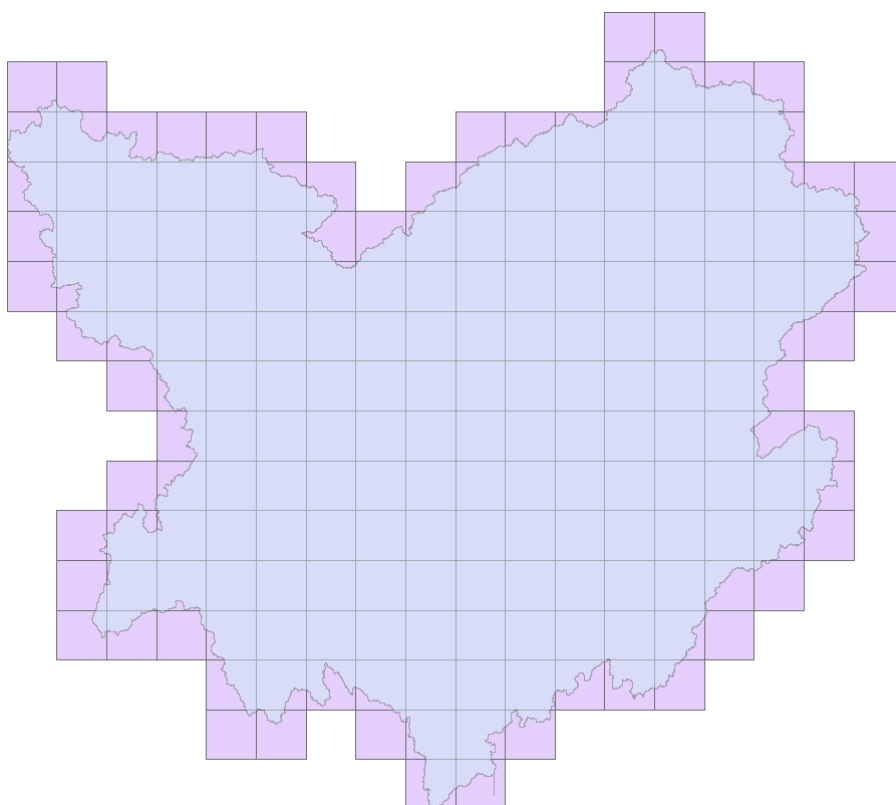
The 'Cell Size Width' and 'Cell Size Height' fields will determine the size of your whole squares (CEs). This will be determined by the size of the watershed sub-basins (see step 7), and it may be advisable to try creating several fishnet grids with different cell sizes to see which works best to ensure a maximum of 4 partial squares per whole square.

Ensure that the selected geometry type is 'polygon'. The 'Number of Rows' and 'Number of Columns' fields can be left blank; ArcGIS will automatically calculate the optimum no. of rows/columns to fit the spatial extent.

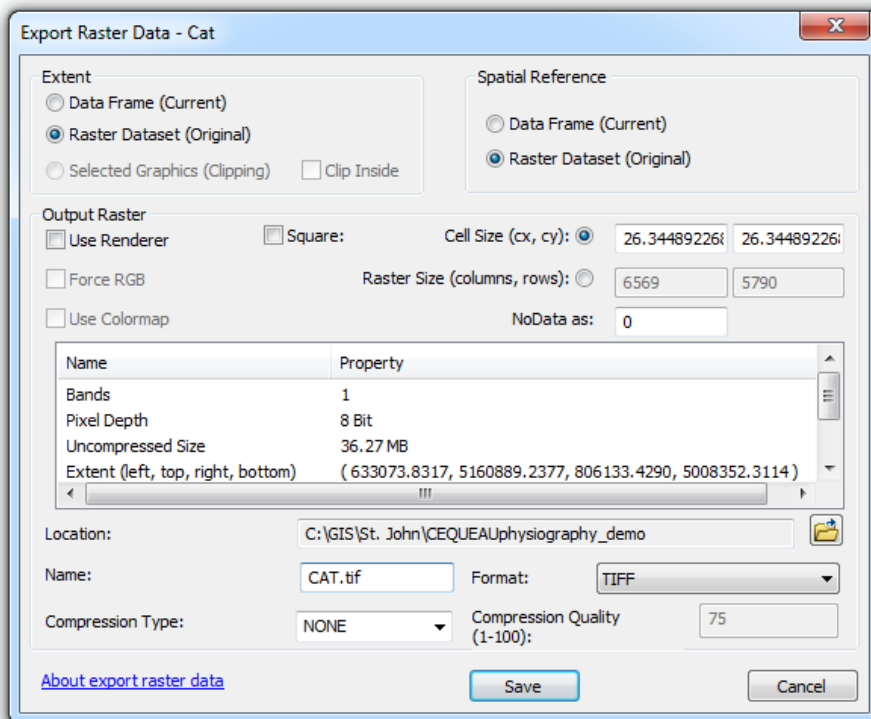
Finally, press 'OK' to create the whole square grid. It should look similar to this:



Step 11. It is advisable to eliminate whole squares that do not cover any part of the watershed. This can be accomplished using the ArcGIS editor tool to delete 'empty' squares from the fishnet shapefile. The process can be achieved manually simply by deleting 'empty' squares one-by-one, or by using the 'Select by Location' tool to highlight squares that intersect the watershed extent shapefile. The resulting fishnet should be similar to this:



Step 12. Finally, export the flow accumulation raster and sub-basin rasters to *.TIF files by right clicking on them in the table of contents and selecting Data -> Export Data:



It is advisable to give the rasters simple names (ie. 'FAC.tif' for the flow accumulation raster, 'CAT.tif' for the sub-basin raster') for ease of opening the files in Matlab.

Ensure that you save these new files in the same folder as the clipped DEM, land cover raster and waterbodies/wetlands shapefiles that you created earlier.

Your data should now be ready for use with the CEQUEAUphysiography Toolbox.

9. Using the CEQUEAUphysiography toolbox to create the 'bassinVersant' structure

Step 1. Load Matlab

Step 2. Load the appropriate data sources into the Matlab workspace using the following commands:

Digital Elevation Model (DEM):

```
DEM = imread('[my folder]\DEM.tif');
```

DEM worldfile:

```
R = worldfileread('[my folder]\DEM.tfw');
```

Land cover raster:

```
LC = imread('[my folder]\LC.tif');
```

Flow accumulation raster:

```
FAC = imread('[my folder]\FAC.tif');
```

Sub-basin raster:

```
CAT = imread('[my folder]\CAT.tif');
```

Fishnet grid shapefile:

```
Fishnet = shaperead('[my folder]\fishnet.shp');
```

Waterbodies shapefile:

```
waterbodies=shaperead('[my folder]\waterbodies.shp');
```

Wetlands shapefile:

```
wetlands=shaperead('[my folder]\wetlands.shp');
```

Step 3. You are now ready to run the CEQUEAUphysiography Toolbox. To do this, simply type:

```
[bassinVersant,CEgrid,CPgrid,routing_data,CPcount,CEgrid_warning]=CEQUEAUphysiography(CAT,FAC,R,fishnet,1,DEM,LC,waterbodies,wetlands);
```

Make sure not to omit the '1' between 'fishnet' and 'DEM'. This will tell the CEQUEAUphysiography.m function to ensure that there are no more than 4 partial squares per whole square by forcing partial squares together. This process is done in a hydrologically correct manner, and should not influence the flow routing. However, if you do not want to do this, simply replace the '1' with a '0'.

Step 4. Once started, the CEQUEAUphysiography Toolbox will display progress bars appear on the screen to inform the user about the current state of the physiography extraction process. When these progress bars have stopped appearing and the output variables are visible in the Matlab workspace, the process will be complete.

The CEQUEAUphysiography Toolbox will run for a period of time relative to the size of your watershed. Tests with a large watershed (~60 000 km²) show that it takes around 3 hours to fully compute the

physiography. The bulk of this processing time is made up the vector land cover data extraction process (waterbodies and wetlands). The processing time can however greatly vary based on the resolution/quality of the input data.

Step 5. When complete, the output variables from the CEQUEAU of the CEQUEAUphysiography toolbox will be follows:

`bassinVersant`: the completed CEQUEAU `bassinVersant` structure

`CEgrid`: a raster image showing the whole squares

`CPgrid`: a raster image showing the partial squares

`routing_data`: a structure containing more information about the flow routing (upstream and downstream neighbouring partial squares, all partial squares upstream of a given point, routes from each partial square to basin exit, etc)

`CPcount`: a raster showing the number of partial squares per whole square

`CEgrid_warning`: a raster showing any whole squares that might have problems when running CEQUEAU (ie. not hydrologically correct, too many partial squares)

The variable of most interest is the 'bassinVersant' structure, which can be directly added to a CEQUEAU structure. Please note that the `bassinVersant` structure will not include the data necessary for running the 'Qualité' element of CEQUEAU.