GEE Classification Implementation Guide

July 5th 2023

Updated Aug 29 2024

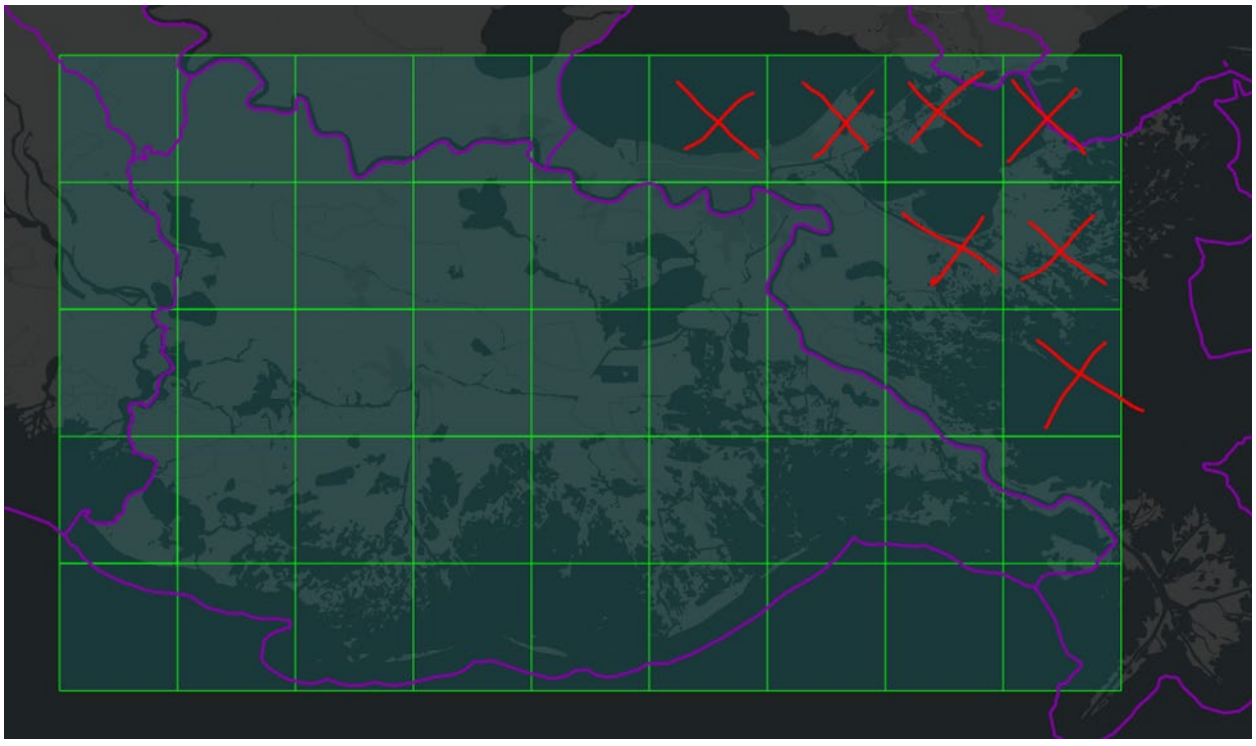
Courtney Di Vittorio

Example shown for Barataria, LA watershed

Step 1: **Identify watershed sub-grids that should not be included in the GEE classification.**

a) In ArcPro load the watershed extents and add the sub-grid kml files for your watershed (located in “Coastwide_Classification/watershed_grids/grid_kml”. An example for Barataria (watershed #8) is shown below.

b) Identify the rows and columns that should be removed because they are outside the wetland area.



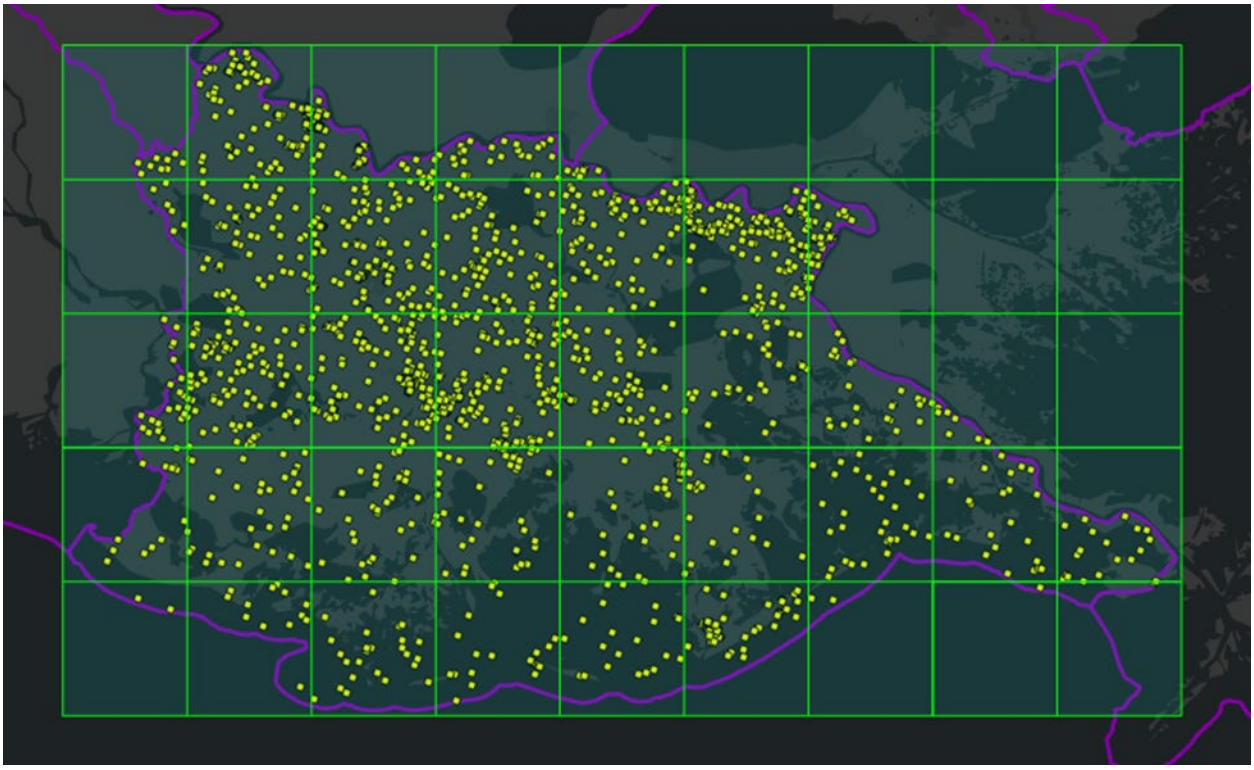
Step 2: **Identify the grids that do not have training/validation data**

a) Under the map tab of ArcPro, click the drop-down arrow under add data and then click “XY Point Data”. Select the *unchanged* training data csv file for your watershed from “G:\My Drive\Research\NASA_OBB_Coastal_Marshes\Coastwide_Classification\training_data\gee_allWS\csv_unchanged”.

b) On the geoprocessing side bar, make sure Longitude is populated for the “X Field” and Latitude is populated for the “Y Field” (should be automatic). Click Run.

c) Similar to step 1, add the sub-grid kml files for your watershed to the map.

d) Identify grids that overlap with the watershed that do not have any training data. trainingBARcv1.csv is shown below, and there are no grids that meet this criteria, but it is likely that this criteria will be met for other datasets and other watersheds.



Step 3: Run GEE Segment Export

a) Open up GEE script ccdc_segments_template_part1

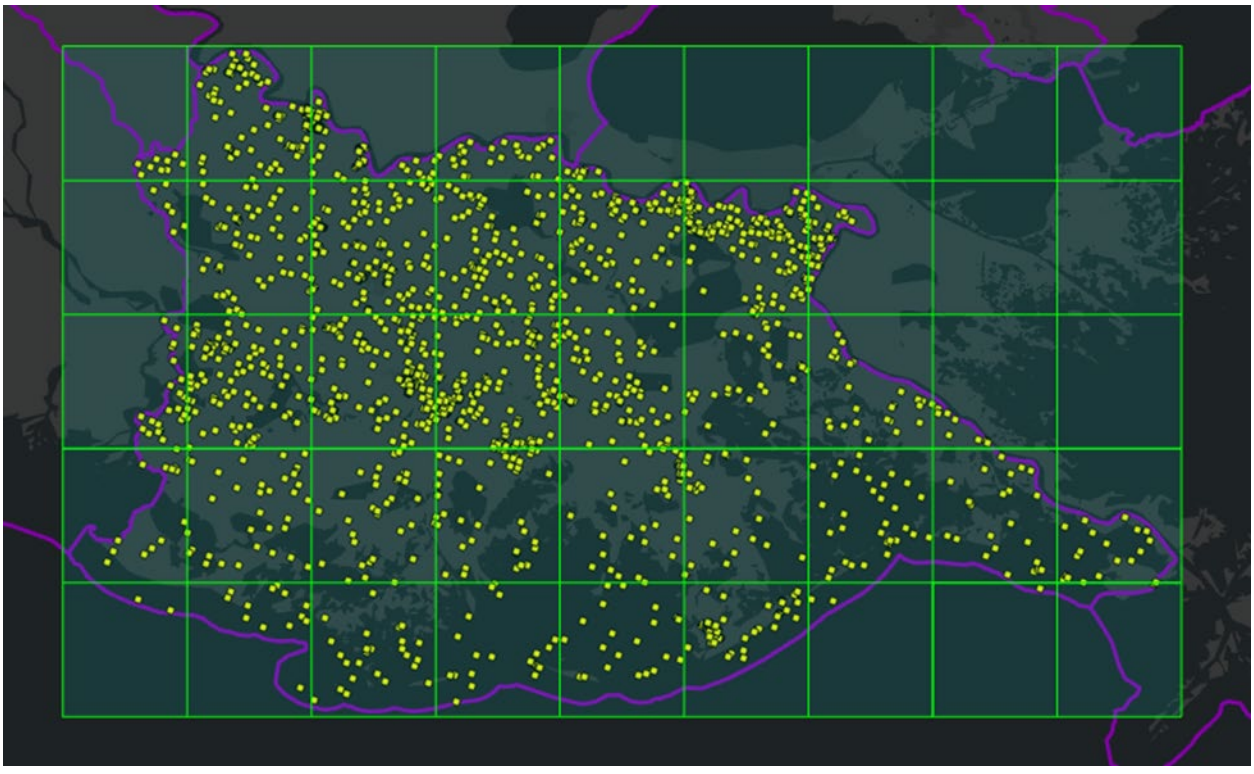
b) Add your name and date to the top and use “save as” to change the name to ccdc_segments_[watershed_number]_[your_initials]

c) Open txt file that contains lat and lon coordinates for your watershed, located in “Coastwide_Classification\watershed_grids\subGrid_txt” folder

d) Copy/paste the grid coordinates for your watershed at the top of the GEE script (replacing those for Barataria in the template)

e) Delete any grids that are outside of the watershed boundaries

f) IF the segment does not export then this is likely because the grid is located in a place where there are 3 overlapping Landsat tiles, which creates a very dense stack of images with duplicate information and causes GEE to run out of memory. Go through Step 3 supplement for that grid to confirm whether issues is 3 overlapping tiles and identify the 2 tiles you would like to keep. These two tiles should have the same path or row number. Take note of the common number. Open the ccdc_segments_redMemory_rev1_template GEE script and save a copy, adding the watershed number and your initials to the end. Copy/paste the grid you are processing in the GEE script and enter the common path and/or row number when filtering. Export the results as a GEE asset.

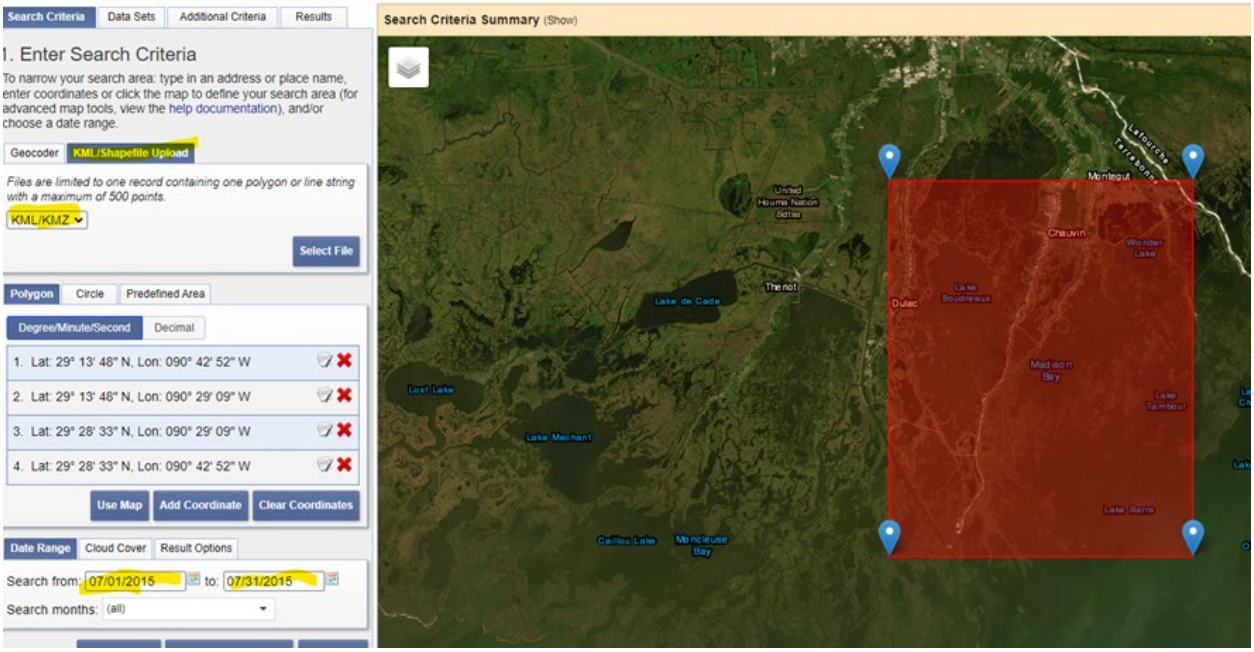


Step 3 - (only when needed - out of memory) Evaluate overlapping Landsat tiles

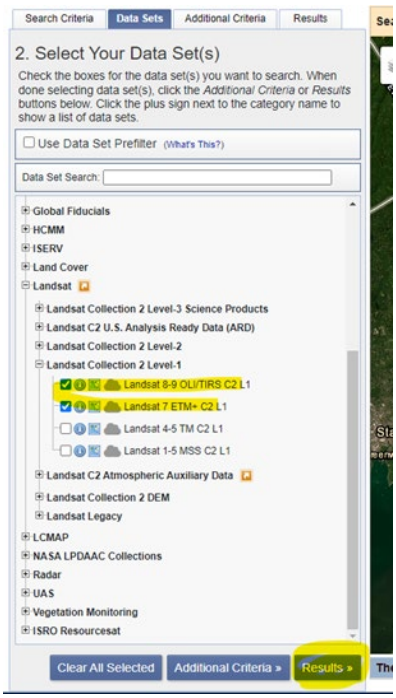
a) Go to https://earthexplorer.usgs.gov/ You will need to create an account and login.

b) Click the shapefile upload option with KMZ/KML. Click on “select file” and navigate to “watershed_grids/grid_kml” folder. Select a grid from the watershed that you think might have 3 overlapping tiles.

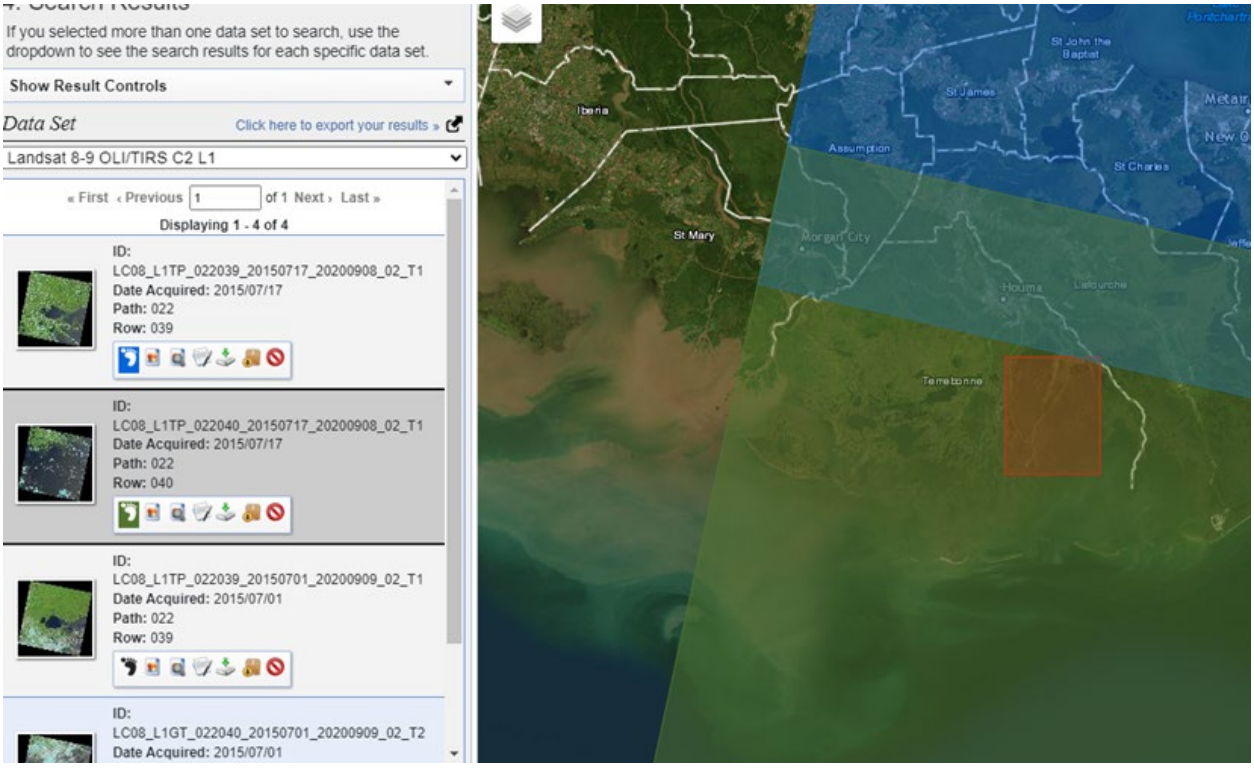
c) Add dates for the search criteria – 1 month will be sufficient (it does not really matter) I suggest choosing a month in 2015 to ensure it aligns with Landsat 7 and 8.



d) Under Data Sets Tab select Landsat Collection 2 Level-1 and click both L7 and L8-9. Then click on results.



e) Click the footprints on for all tiles and count number of overlapping tiles. In this case there are 2, so this is not a concern, but if there are 3 tiles then record the watershed, row, and column number. I believe the tiles are the same for landsat 7 and 8/9, but double check this with a few grids.



f) Upload the next sub-grid that is suspect and repeat until all sub-grids have been checked

Step 4: **Upload training data as GEE Asset**

a) Go to the training data folder (Coastwide_Classification\training_data\gee_allWS\csv_unchanged) and upload the unchanged training dataset (csv file) associated with your watershed as a GEE Asset.

b) When the task is complete you can “refresh” under the asset tab and check that it is there. Click on it to make sure it imported properly

Step 5: **Run GEE classification script**

a) Open the “**ccdc_classify_template_part2_allWatersheds**” script

b) Add your name, date, and comment at the top and hit “save as”. Save the script as “ccdc_classify_[watershed_number]_[your_initials]

c) Run the script and export 1 grid at a time, for each cross-validation dataset. After each run, you must submit the task for the file to export to your drive folder. (you need to set up a folder in your drive account, mine is called “EarthEngineExports”

Step 6: Run Annual Classification in GEE

a) For the annual classification, we will train the model on the full dataset (as opposed to unchanged only). Go to the full training data folder (**Coastwide_Classification\training_data\gee_allWS\csv_full**) and upload the full training dataset (csv file) associated with your watershed as a GEE Asset.
b) Open the GEE script "annual_classification_template_allWatersheds"
c) We can run this classification on the entire watershed as opposed to 1 grid at a time. To define your grid, go back to your segments script (ccdc_segments_ws2_LH) and identify the upper left and lower right grids. Copy/paste these grids onto the current script and take the first and last coordinates from the upper left grid and the middle coordinates from the lower right grid to make the new grid extents, called fullAOI.
d) Make sure you are using the training data you uploaded in step a) and run the script.
e) Instead of exporting each grid separately, you will need to export each year separately.

Step 7: Image post-processing in MATLAB/Pyhton/GIS

1. Run export_images.m on "classified_smoothed" results to combine individual grids and make annual geotifs of each year and change type. Note you will need to change the gridNums variable and directories within the MATLAB script.
2. Create a new folder for the mosaiced images and run "python_mosaic" within ArcPro python notebook to mosaic images for each year. Make sure to change the workspace and output folder in the script.
3. Create mask for "others" that stay others for the full time series, using mask_AgDev.m. This will create a maskAgDev.tif file.
4. Project and Snap the maskAgDev.tif file to CCAPmask.tif in arcPro "project raster" tool. Set x and y dimensions to CCAPmask.tif and in the environment variables tab "snap" to CCAP mask. Save as a new file, "maskAgDev_CCAP.tif"
5. Export watershed of interest as a single polygon using the "export feature" tool (need to have updatedMap.tif layer highlighted in workspace and watershed being exported selected to use this tool. Project to CCAPmask.tif using the "project" tool.
6. Use arcPro "project_mask_clip" model (in GEE_training toolbox) to project, mask, snap, and clip all images to the CCAP. Within this tool, will need to update the maskAgDev_CCAP mask and the watershed polygon for clipping rsvh time. Will also need to check the "use input features as clipping geometry" box (but do not maintain clipping extent) within the Clip tool each time.