# Drone imagery classification with ArcGIS Pro

Jeffrey Gillan, Ph.D.

Sept. 2020

## Objectives

- Become familiar with ArcGIS Pro software by following a step-by-step tutorial to classify high-resolution drone imagery
- Create a canopy height model by subtracting a digital terrain model from a digital surface model on a cell-by-cell bases
- Create training data for a supervised classification
- Execute a supervised classification using a *random forests* algorithm
- Assess the accuracy of the classification

## Software Used

- ESRI ArcGIS Pro 2.5.0

## Data Used

- 21_2_ortho_clip.tif

  This is a raster orthomosaic created from many overlapping aerial images collected from DJI Phantom 4 RTK quad-rotor drone. The imagery has blue, green, and red bands, each with 8 bit radiometric resolution. The orthomosaic was created using structure-from-motion software Agisoft Metashape. The ground sampling distance (i.e., spatial resolution) is ~1 cm.
- 21_2_dsm_clip

  This is a grid format raster depicting the elevation of the study area. It is referred to as a digital surface model (DSM). It includes all features on the surface including topography, vegetation, and any manmade features. It was created by interpolating a point cloud using the software Agisoft Metashape. The spatial resolution is ~1 cm.
- 21_2_dtm_clip

  This is a grid format raster depicting the elevation of only the bare ground in the study area. It is referred to as a digital terrain model (DTM). Vegetation features have been removed using Agisoft Metashape. The spatial resolution is ~1 cm.

## Part 1: Set-up Map Project and Explore Data Layers

1. Start a new project in ArcGIS Pro using the Map template. Save it in your working directory or wherever you would like to store it on your local computer.

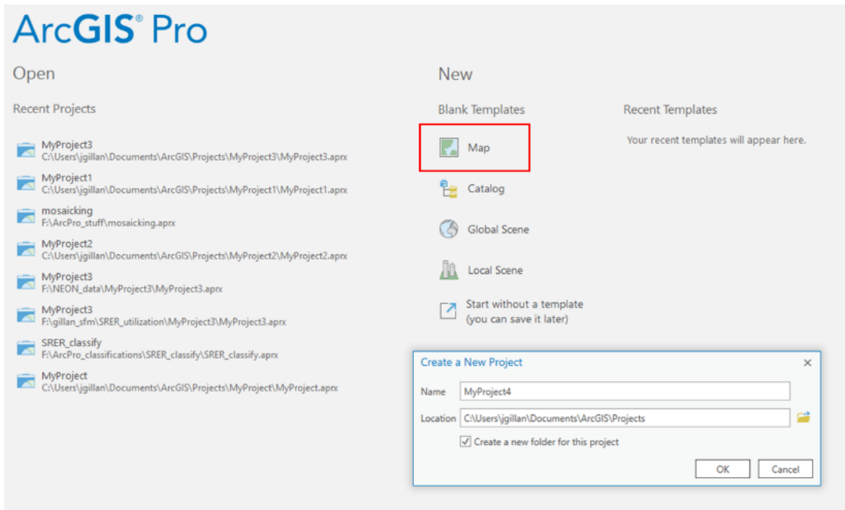
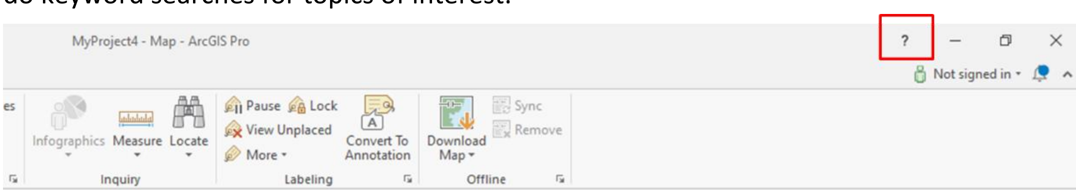



Fig. 1

2. When using ArcGIS Pro, you can access help documentation by clicking on the ? in the upper right corner of the graphical user interface. Clicking on the ? icon will bring up a web browser where you can do keyword searches for topics of interest.




3. Add data layers to the map. Add the orthomosaic, the DSM, and the DTM.

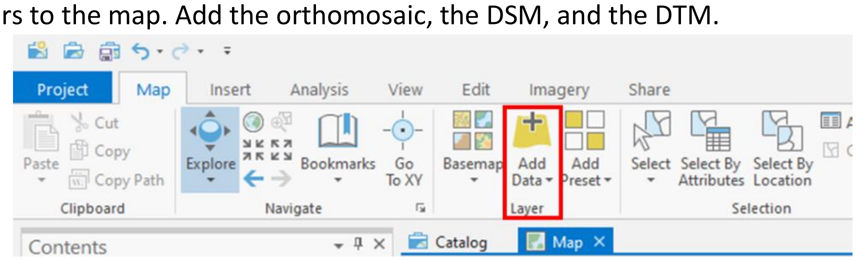



Fig. 3a

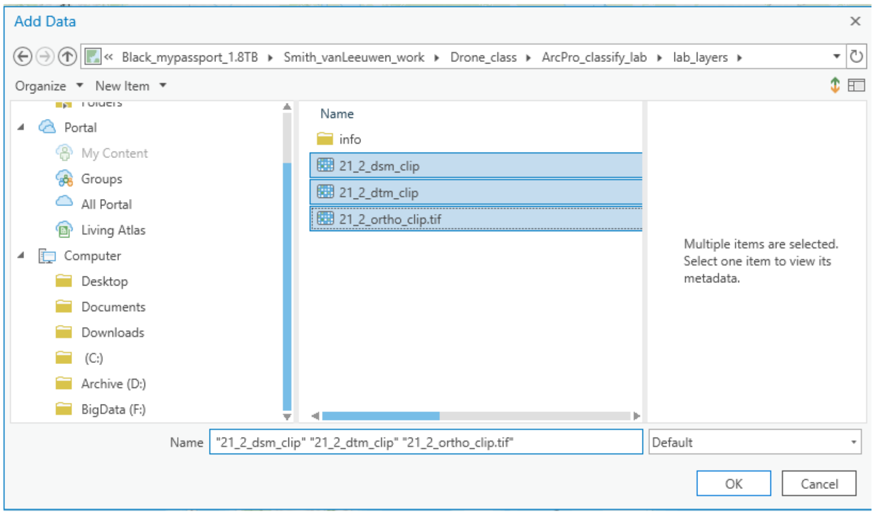



Fig. 3b

Click *yes* to generate pyramids for all 3 raster layers. This will make display faster when you are scrolling around on the map. It may take a few minutes to complete the pyramid building.

4. Explore the data layers using the *Explore* tool located on the main toolbar.

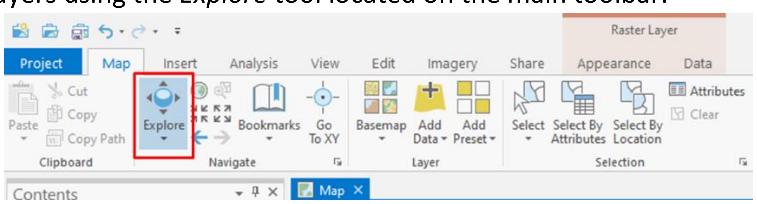



Fig. 4a

With this tool, you can pan across the data layers and zoom in and out (mouse scroll wheel). Left clicking on the data layers in the map view will bring up a pop-up window describing the pixel values of all data layers (if they are checked on).

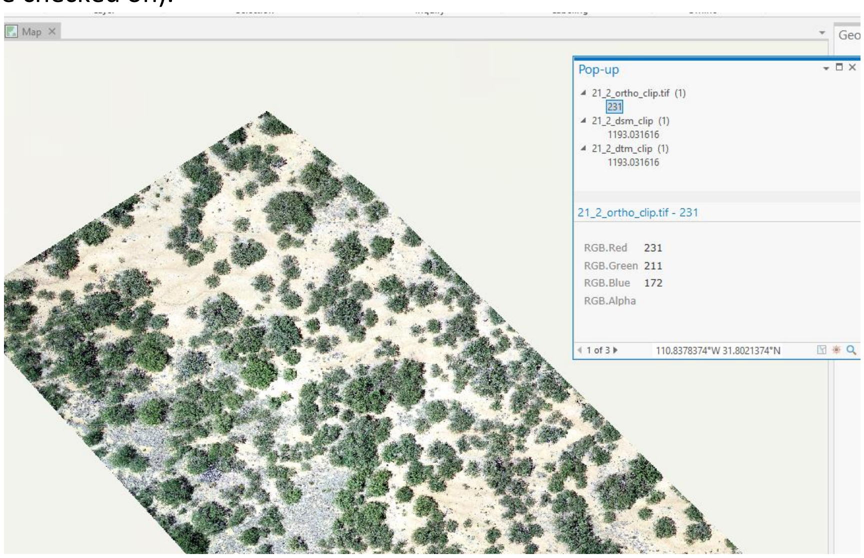



Fig. 3b

Q1: Using the *Explorer* tool, left click in the middle of a shrub. Please report the elevation value from the DSM and the elevation value from the DTM. Which one is higher and why? What units is elevation being reported in?

5. Zoom out in the map (using mouse scroll-wheel) to see where the study area is located. The imagery was collected at the base of the Santa Rita Mountains which is about 50 km south of Tucson. The imagery was collected in May of 2019. Santa Rita Experimental Range is a Sonoran desert savanna. The vegetation composition is dominated by mesquite trees, cactus such as cholla, and perennial grasses such as Lehman's Lovegrass. There are also significant amounts of bare ground.

6. Let's look at the properties of the orthomosaic by right-clicking on the layer in the table of contents.

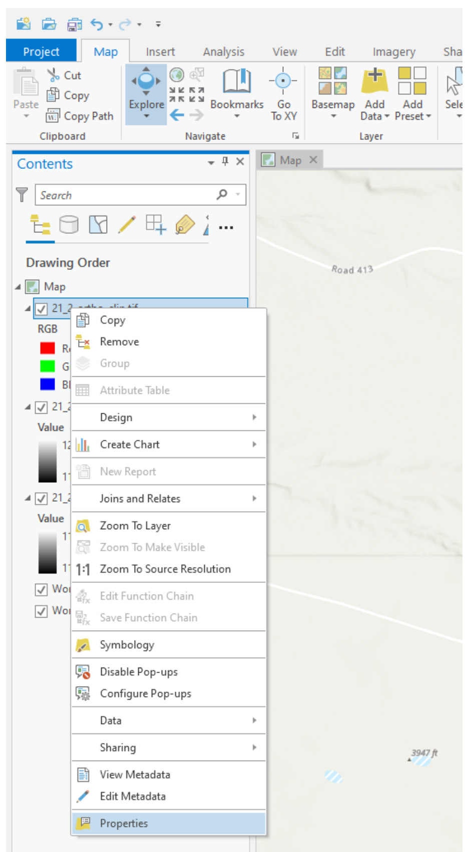


Fig. 6

7. Within the *Layer Properties* window, click on *Source* and then *Spatial Reference*.

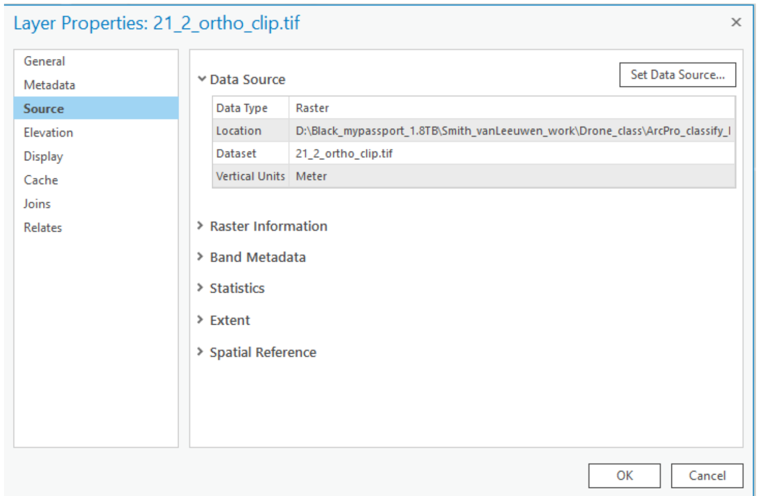



Fig. 7

Q2: What is the geographic coordinate system of this raster layer?
Q3. What projection does the orthomosaic have?
Q4. Please provide screenshots of the DSM, DTM, and orthomosaic at an extent where the entire data layers can be seen.

## Part 2: Create Canopy Height Model

A canopy height model depicts vegetation height. Later in the lab, we will use vegetation height as a feature used in the classification. First, we need to create the canopy height model by subtracting the digital terrain model from the digital surface model. With this procedure, we are measuring the vertical distance between the elevation of vegetation (DSM) and the elevation of the bare-ground (DTM).

8. Open the *Geoprocessing* window which is found on the main toolbar under the *View* tab

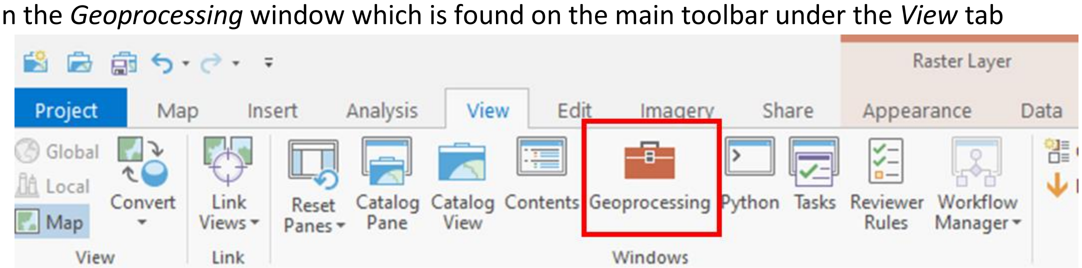



Fig. 8

9. In the search box, type *raster calculator*. Click *Raster Calculator* tool. The tool found in *Image Analyst Tools* and *Spatial Analyst Tools* should be the same so it does not matter which one you select.

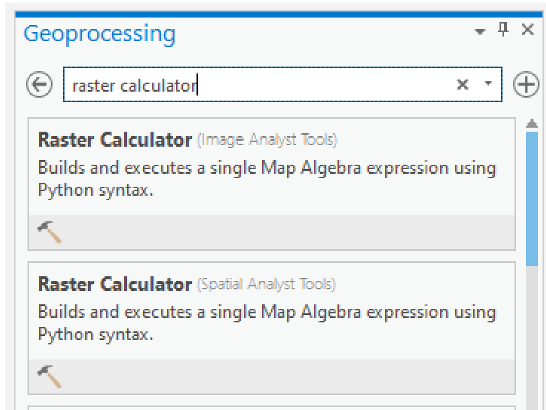



Fig. 9

10. Create the map algebra expression as shown in Fig. 9a. Double click ‘21_2_dsm_clip’. Then double click the minus sign. Then double click ’21_2_dtm_clip’. You are subtracting the DTM from the DSM on a cell by cell basis. Save it as a .tif in your working directory.

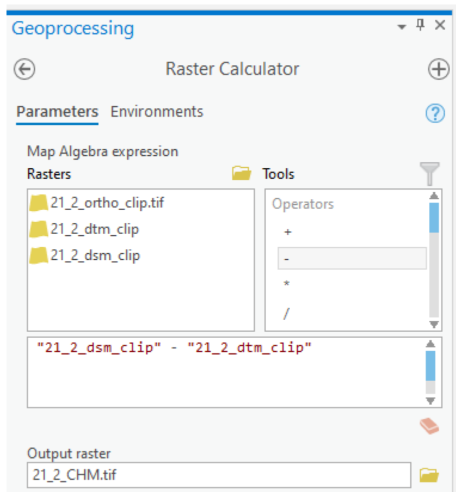
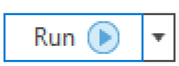



Fig. 10a

Click 


The Raster Calculator processing output will be found in the table of contents on the left side of the GUI. The canopy height model should appear on the map.

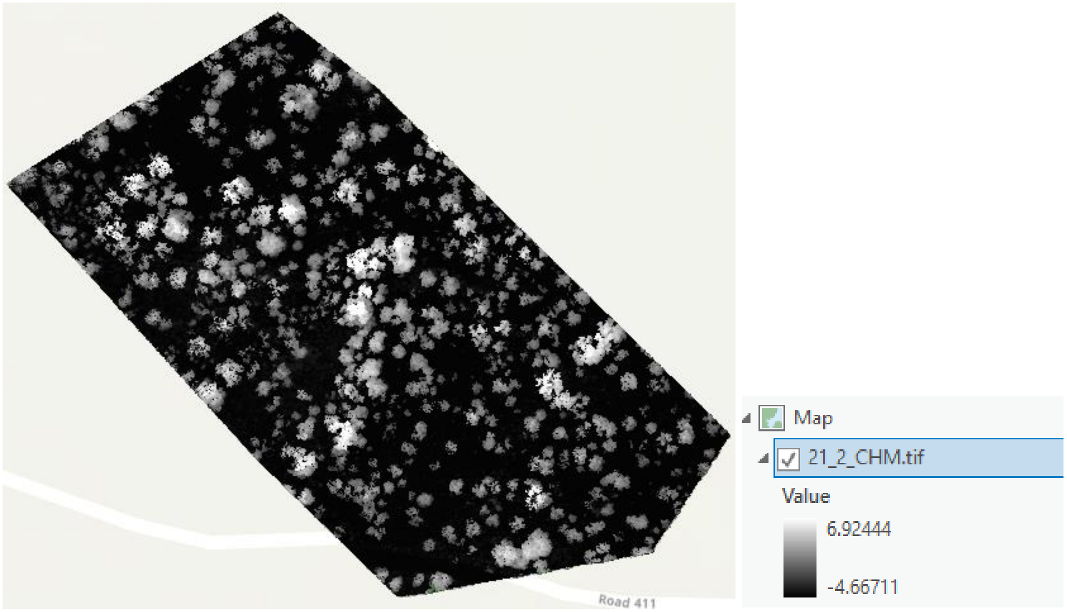



Fig. 10b

You will notice there are some negative values in the canopy height model. It doesn't make much sense to have negative vegetation heights. The creation of DTMs and DSMs is not perfect and can sometimes lead to erroneous data. Let's delete these negative value errors from the canopy height model.

11. In the *Geoprocessing* search box, type conditional, and select the tool *Con*. It should not matter if you use the *Con* in the Image Analyst Toolbox or Spatial Analyst Toolbox, they are the same.

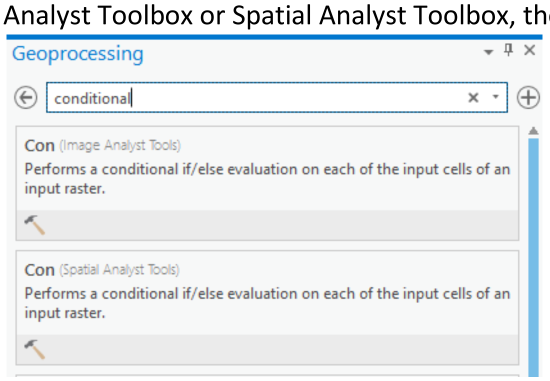



Fig. 11

12. Fill out the parameters you see in Fig. 12. We are going to set any value less than 0.05 m to zero, otherwise the pixels will retain their height value. Output the product as a .tif in your working directory.

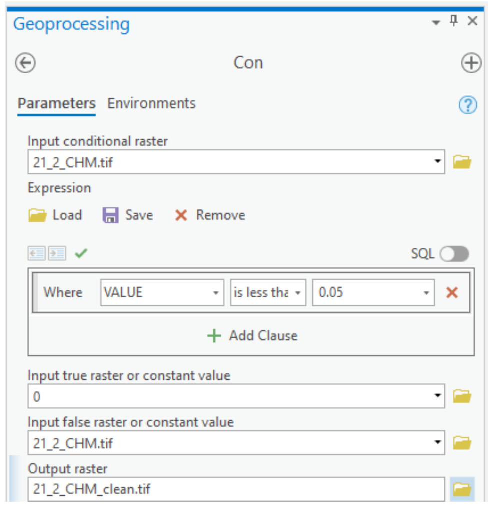



Fig. 12

Q5. Provide a screen shot of the canopy height model with negative values removed. Also provide a screenshot of '21_2_CHM_clean' and color ramp as shown in the table of contents

# Part 3: Classification Training

13. Create a new shapefile to hold the training samples. In the *Geoprocessing* search box, type *create feature class* and open that tool.

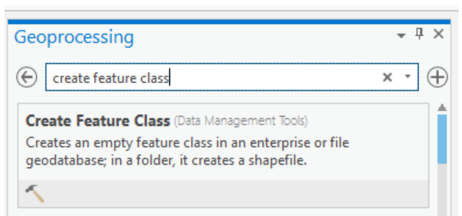



Fig. 13

14. Fill out the parameters as shown in Fig. 14. Make sure the coordinate system is the same as the canopy height model (WGS 84)

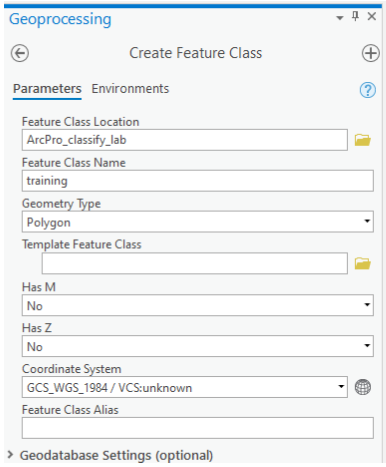



Fig. 14

15. We are going to use the *Training Samples Manager*. To find this tool, make sure the orthomosaic is highlighted in the table of contents, then click on the *Imagery* tab >>>> *Classification Tools* >>>> *Training Samples Manager*

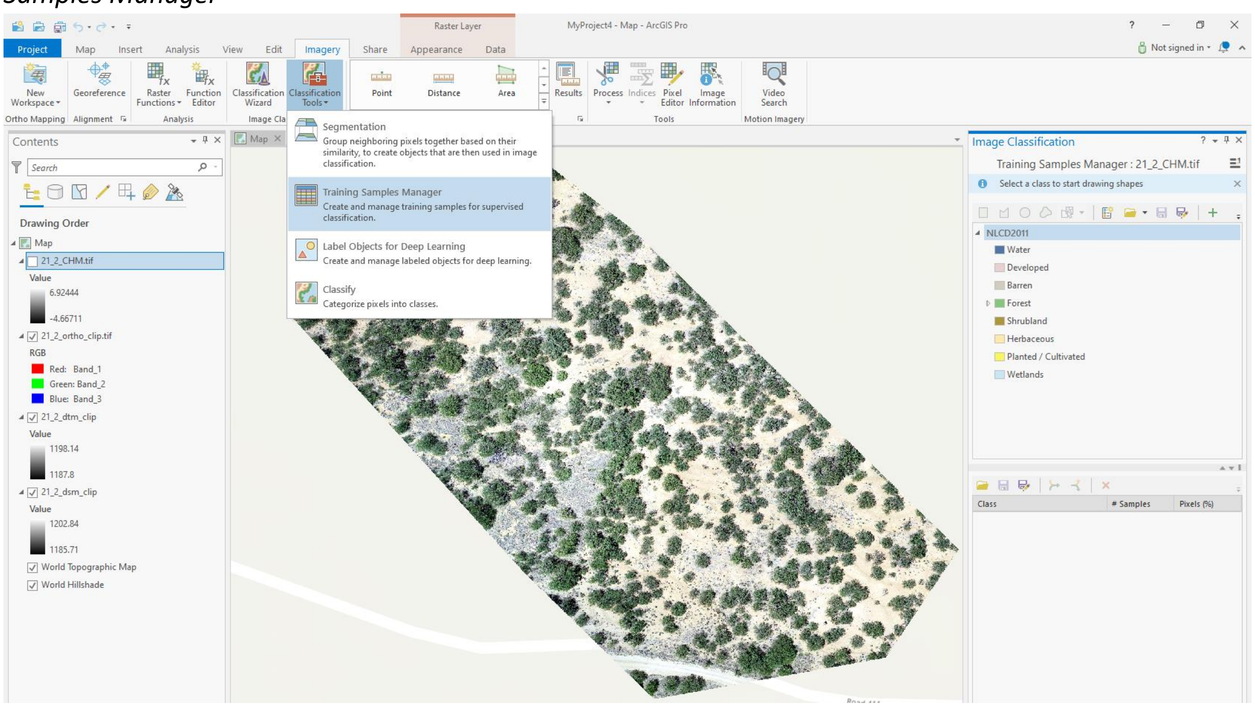



Fig. 15

16. In the *Training Samples Manager*, click *Create New Schema*

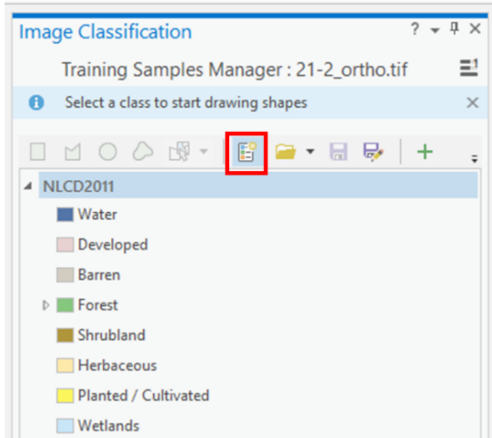



Fig. 16

17. Click on *Add New Class*

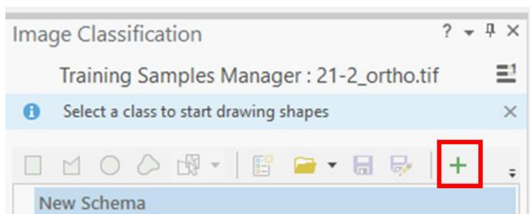



Fig. 17

18. Call this class 'woody', give it a value of 1 and make it dark green

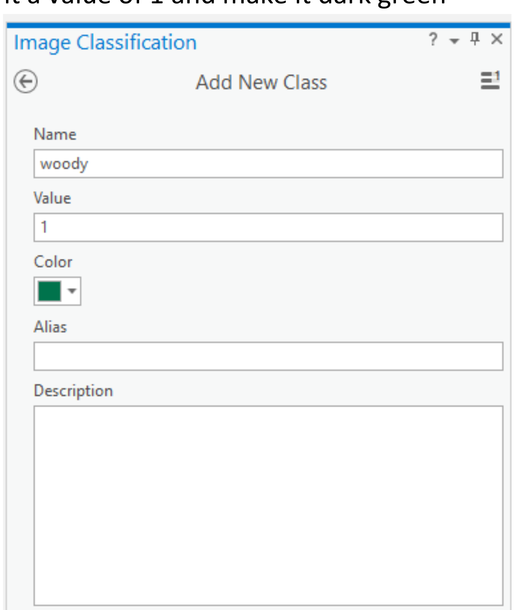



Fig. 18

19. Add two more classes to the schema. Be sure to highlight 'New Schema' before clicking on *Add New Class*. Otherwise it will make the new classes as children of 'woody'.

Class Name: herb; Value: 2; Color: light green

Class Name: bareground; Value: 3; Color: brown

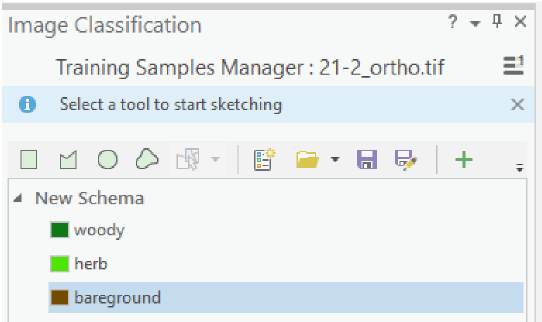



Fig. 19

20. Highlight the bareground class, then click on the *polygon* tool

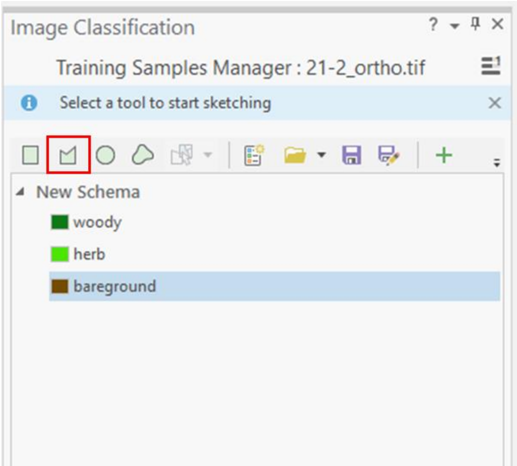



Fig. 20

21. Scroll around the orthomosaic and find areas that represent bareground. Draw 5 total polygons that capture the variability of bareground colors you see in the study area. For example, the dirt road has a white look to it and much of the rest of the bareground is light yellowish brown. Make the size of the polygons about 2 m diameter, though there is not a hard rule on how large the training samples should be. Try to keep these samples as pure as possible, meaning try not to have any grasses or shrubs within your bareground polygons.

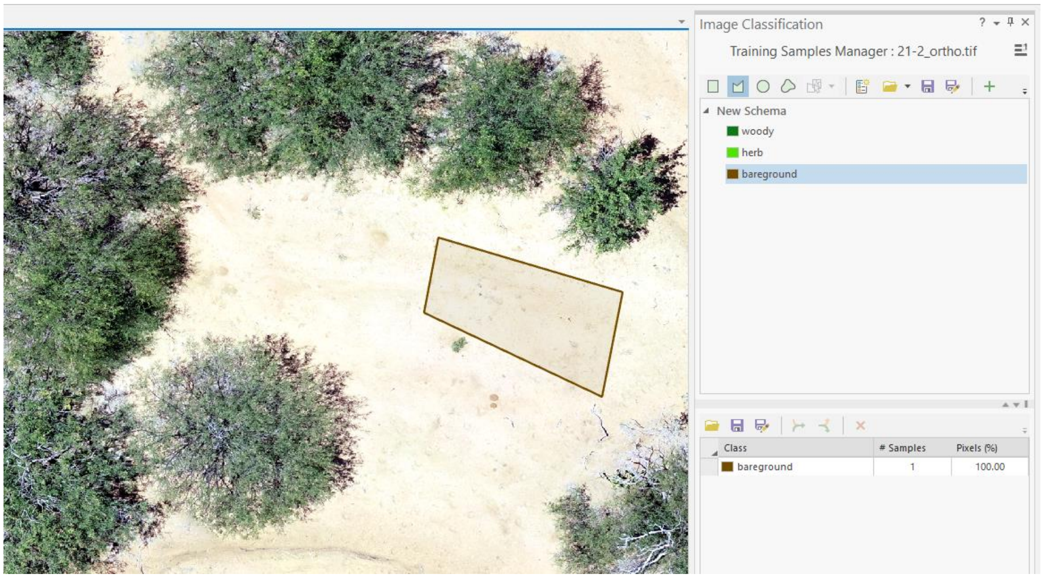


Fig. 21

22. Make 5 training polygons for the herb class. Most of the herbs in this study area are senesced perennial grasses such as Lehman's Lovegrass. The imagery was collected in May when the grasses are still mostly dormant and have not turned green. In the image, the grasses have a grey color to them and should be easy to distinguish from bareground.

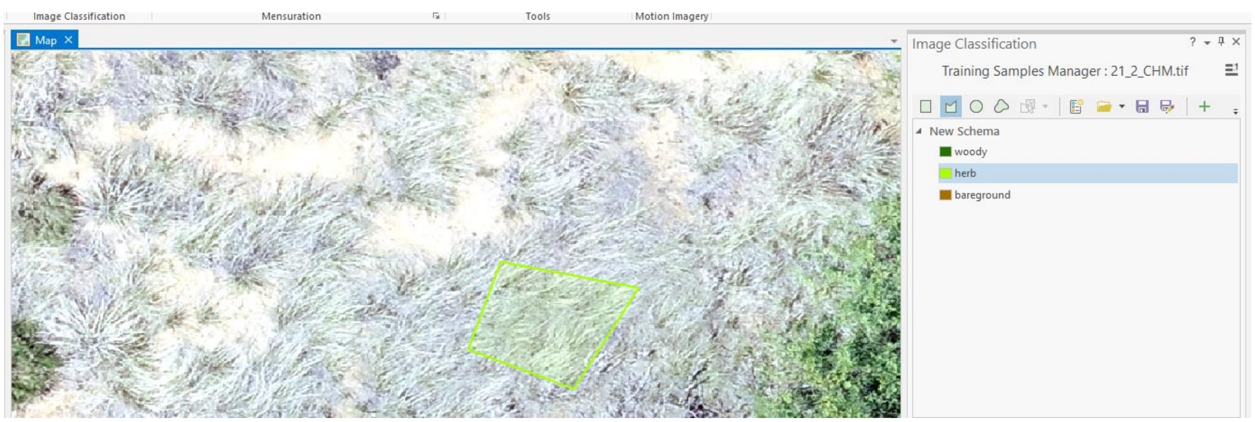


Fig. 22

23. Make 5 training polygons for the woody class. All of the green vegetation you see in this orthomosaic are likely woody vegetation (mostly velvet mesquite *Prosopis velutina*).


Fig. 23

24. After creating 5 training polygons for each class, *Save* the samples.

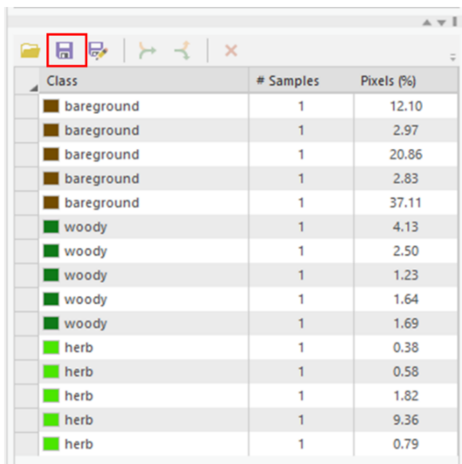



Fig. 24a

You are going to save the samples as the empty shapefile you created earlier in step 13. Click yes to overwrite the existing file.

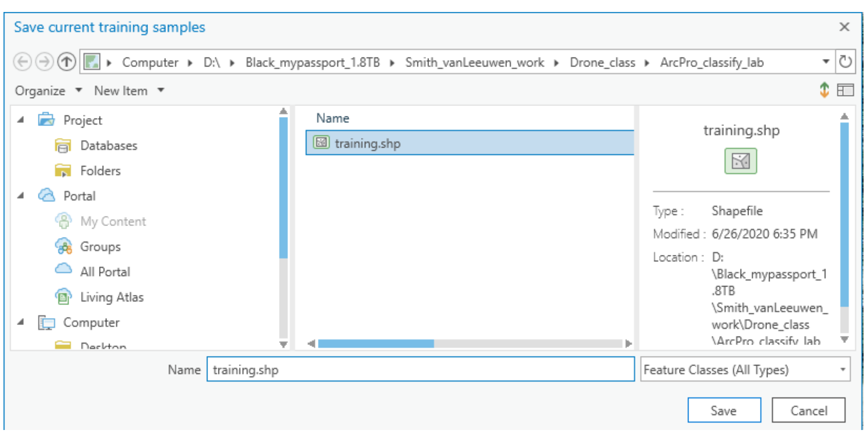



Fig. 24b

# Part 4: Classification

25. Go to the *Geoprocessing* search box to find the tool *Train Random Trees Classifier*. Random trees is a supervised classification algorithm that uses many decision trees.

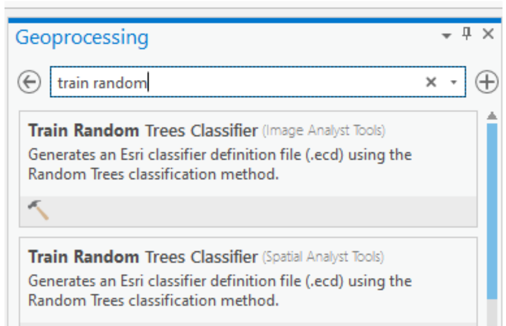



Fig. 25

26. Fill out the parameters as show in Fig. 26. We are using the red, green, and blue bands from the orthomosaic and canopy height values to characterize each class. The output of this tool is something called a classifier definition file. It has an .ecd file extension.

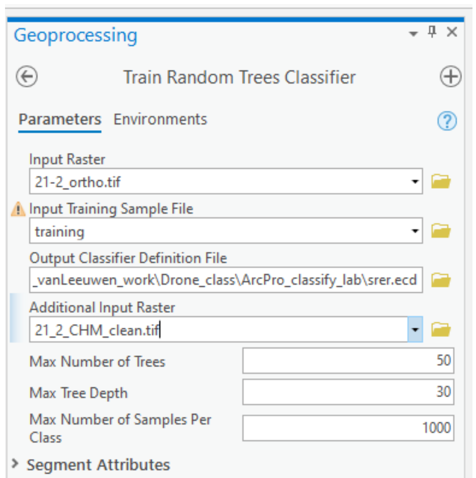
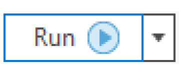



Fig. 26

Click 


27. Go back to the *Geoprocessing* search to find the *Classify Raster* tool.

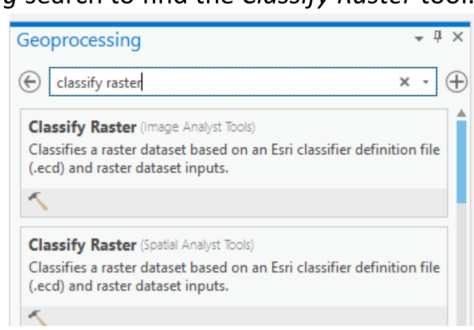


Fig. 27

28. We are now going to execute the classification of the orthomosaic using the *Classify Raster* tool. Fill out the parameters as shown in Fig. 28. The .ecd file you created in Step 26 will be the *Input Classifier Definition File*. Be sure to add the canopy height model as an additional raster layer. Also, go into *Environments* where you can increase the processing speed with *Parallel Processing Factor*. This option lets you specify the percentage of CPUs you would like to allocate to the classification. Type 95%. The runtime will depend on the local computing resources.

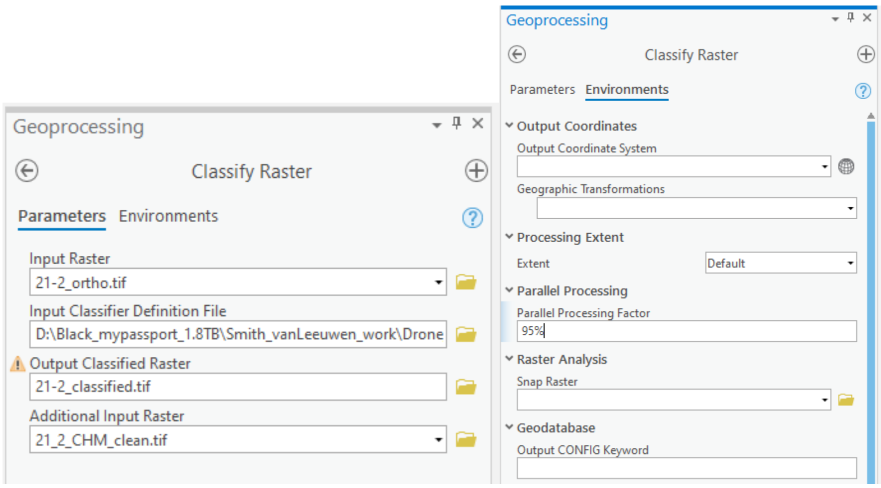



Fig. 28

29. After processing, pan around the image and compare it with the orthomosaic.

Q6. Provide a screen shot of your classified map.

Q7. Upon visual inspection, how accurate do you feel the classification is?

## Part 5: Accuracy Assessment

30. We are going to quantify the accuracy of the classified image by generating many random validation points. At each of these points, you will compare the classified image with your visual interpretation of the correct class. In the *Geoprocessing* search, find and open *Create Accuracy Assessment Points*.

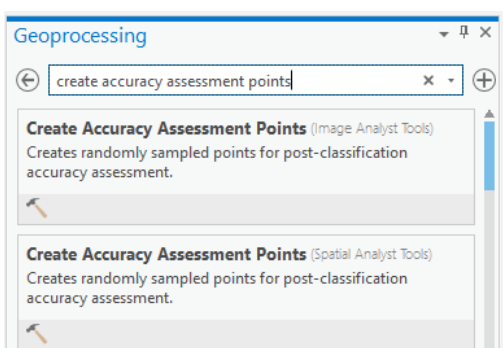


Fig. 30

31. In the *Create Accuracy Assessment Points* tool, fill out the parameters as shown in Fig. 31. The tool will randomly generate 99 points throughout the classified map; 33 points will fall in woody, 33 in herb, and 33 in bare ground. The output will be a shapefile.

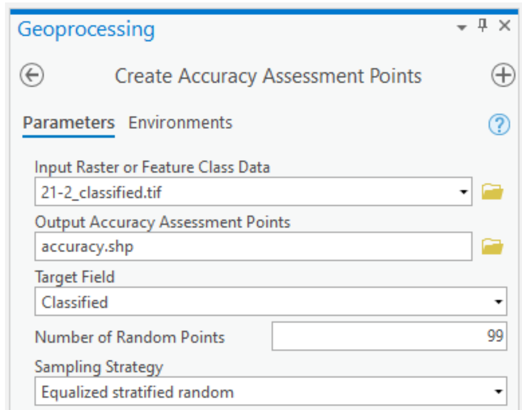



Fig. 31

32. The new shapefile will now appear in the table of contents. Right click on the shapefile to open the *Attribute Table*.

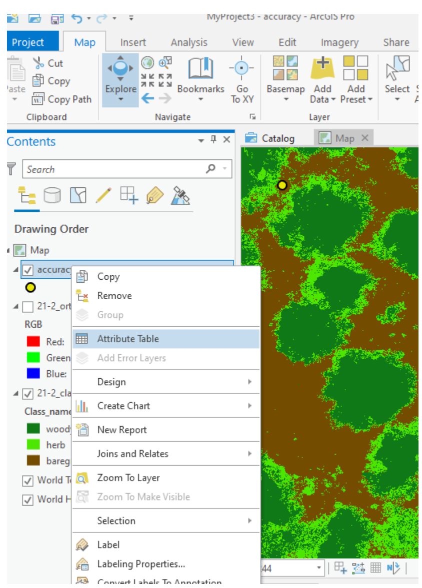



Fig. 32

33. The attribute table shows a column called 'Classified' with values of 1, 2, or 3. These numbers correspond to classes (1 = woody; 2 = herb; 3 = bare ground) that were assigned by the algorithm. We are going to hide this column while we visually interpret what the class assignment of each validation point should be. This is a good protocol to reduce bias influence from the classified map. To hide the column, right-click on the column heading for 'Classified'. It will bring up a menu where you should select *Hide Field*.

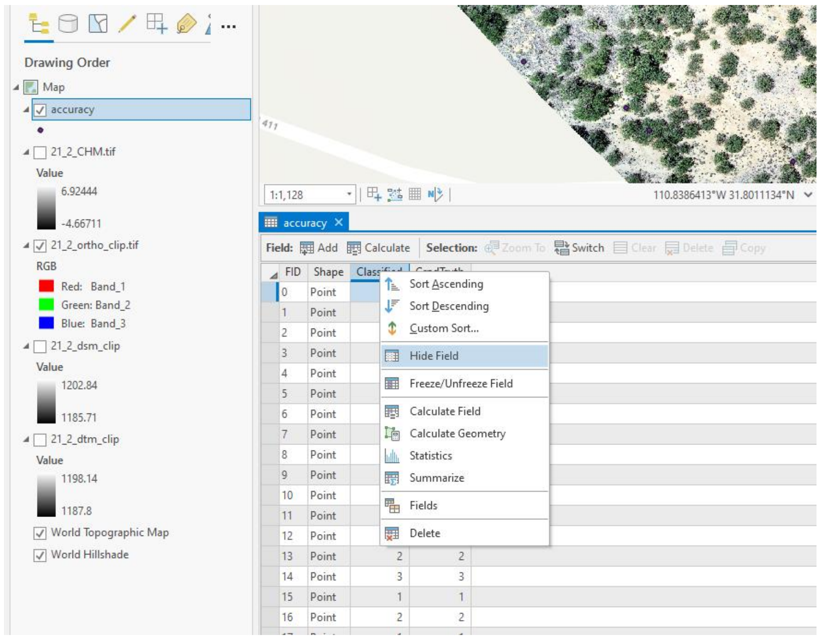


Fig. 33

34. Go to the location of one of the validation points by double clicking on the area to the left of the 'FID' column (shown in Fig 34). Make sure the orthomosaic layer is turned on and the classified layer is turned off.

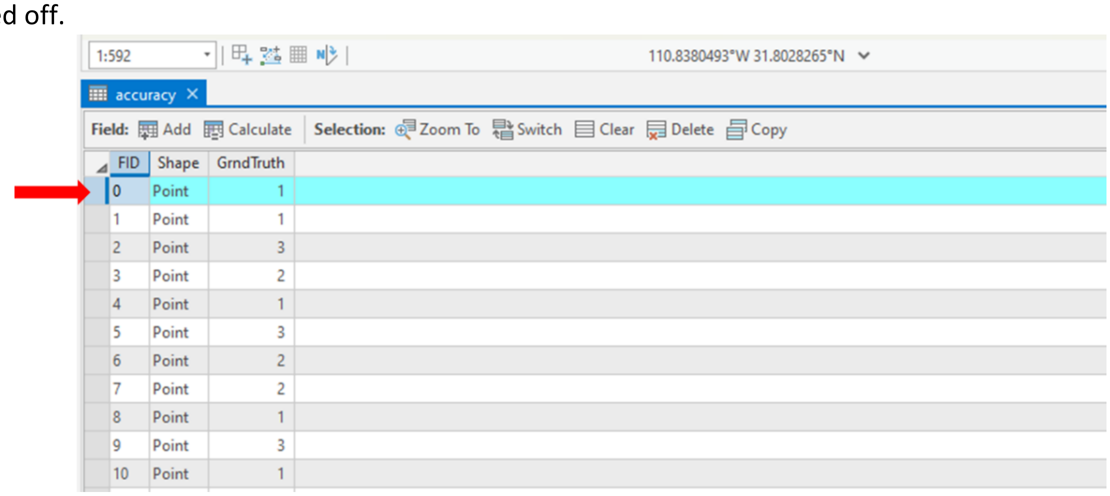


Fig. 34

35. Look at the orthomosaic directly behind the validation. Do you think it is woody vegetation, herb vegetation, or bareground? The 'GrndTruth' column is where you will input your visual interpretation of

the correct class value. You can directly edit the cell values in the 'GrndTruth' column to input your interpretation of the correct class value. If you think it is woody, type a '1' in the 'GrndTruth' column. Type '2' for herb, or type '3' for bareground.

You are going to do this for all 99 points.

36. When you have interpreted all of the validation points, please remember to *Save* your edits.

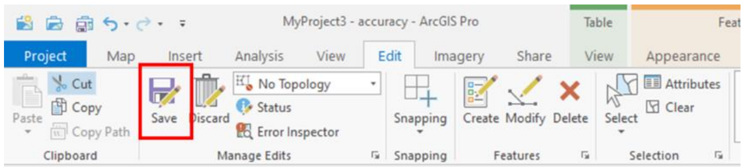



Fig. 36

37. To make the 'Classified' column visible again, click on the 3-line button on the upper right corner of the attribute table as shown in Fig. 37. In the menu select *Show All Fields*.

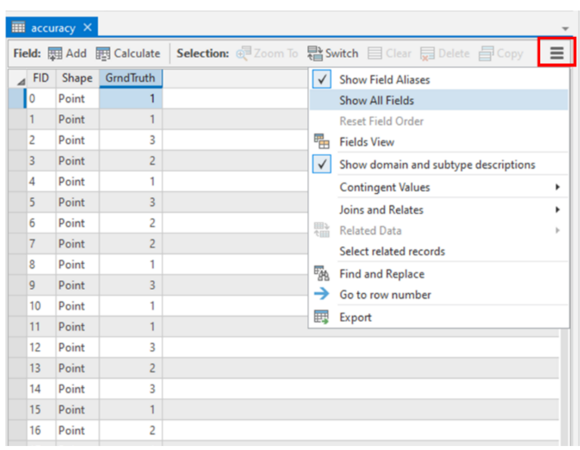



Fig. 37

38. In the Geoprocessing search, find and open the *Compute Confusion Matrix tool*.

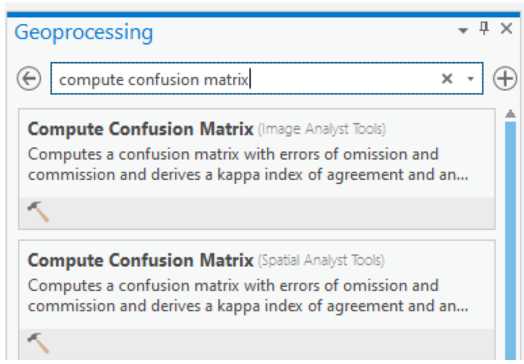



Fig. 38

39. You will input the accuracy points shapefile and output a confusion matrix table. Make sure the output has .dbf extension.

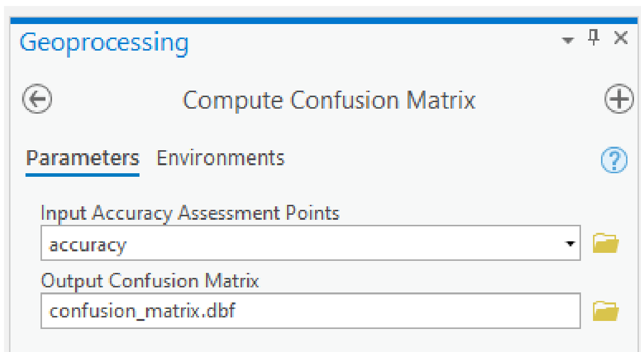
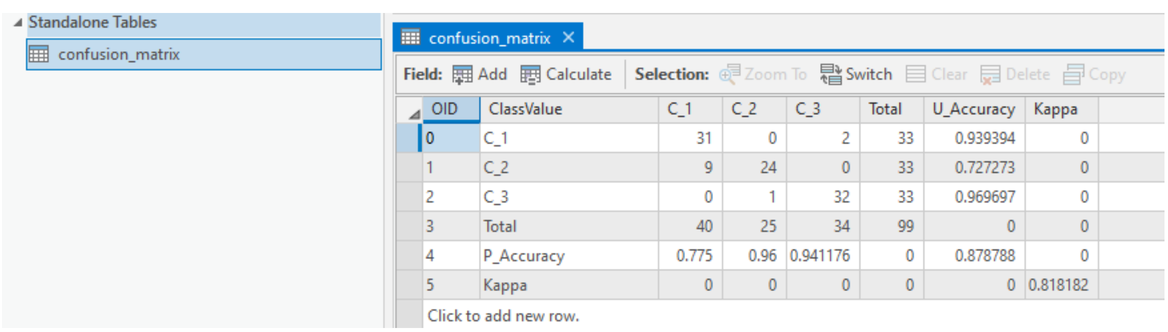



Fig. 39

40. The confusion matrix table will appear in the table of contents under *Standalone Tables*. Right click on the layer to open the attribute table. Fig. 40 shows the accuracy of the random trees classified map compared with your visual interpretation.




| OID | ClassValue | C_1 | C_2 | C_3 | Total | U_Accuracy | Kappa |
|---|---|---|---|---|---|---|---|
| 0 | C_1 | 31 | 0 | 2 | 33 | 0.939394 | 0 |
| 1 | C_2 | 9 | 24 | 0 | 33 | 0.727273 | 0 |
| 2 | C_3 | 0 | 1 | 32 | 33 | 0.969697 | 0 |
| 3 | Total | 40 | 25 | 34 | 99 | 0 | 0 |
| 4 | P_Accuracy | 0.775 | 0.96 | 0.941176 | 0 | 0.878788 | 0 |
| 5 | Kappa | 0 | 0 | 0 | 0 | 0 | 0.818182 |

Fig. 40

Q8. Provide a screenshot of your confusion matrix.

Q9. What is the overall accuracy of your classified map?

Q10. Which class had the best user accuracy?

Q11. Which class had the worst user accuracy? Which classes were confused the most? For example, Fig. 40 shows that herb class the lowest user accuracy (72%). It most often (9 times) classified the pixel as woody when it should have been classified as herb.

Q12. Which class had the worst producer' accuracy in your classification? In Fig. 40, the woody class had the worst because 9 points were assigned as woody when they should have classified as herb.