



## Exercise 2: Change Detection with Supervised Classification

### Objectives

- Understand how to stack bands from two dates
- Generate training sites for a supervised classification
- Use the R software to run the Random Forest Algorithm for conducting a supervised classification for change detection
- Apply a color scheme to interpret changes over time
- Conduct post-processing of an image to improve the accuracy of the image classification

### Overview of Topics

- Band stacking and image enhancement
- Developing training sites
- Random forest classification algorithm
- Post-processing of the classified image

### Tools Needed

- QGIS 3.2 for Windows OS
- R statistical program (version 3.3.0 or higher)
- R Studio

### Associated Data

For this exercise, you will need the image files from the week 1 exercise:

- Tanzania1993\_Clip.tif
- Tanzania2016\_Clip.tif
- And you will need to download the R script classification.R from the ARSET website here: <https://arset.gsfc.nasa.gov/land/webinars/adv-change18>
- Create a folder called Exercise2, and save the R script to this folder

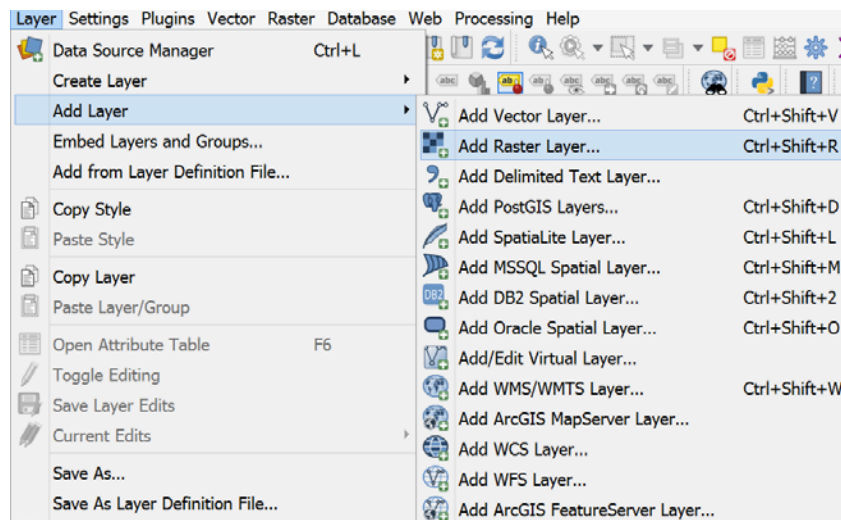


## Introduction

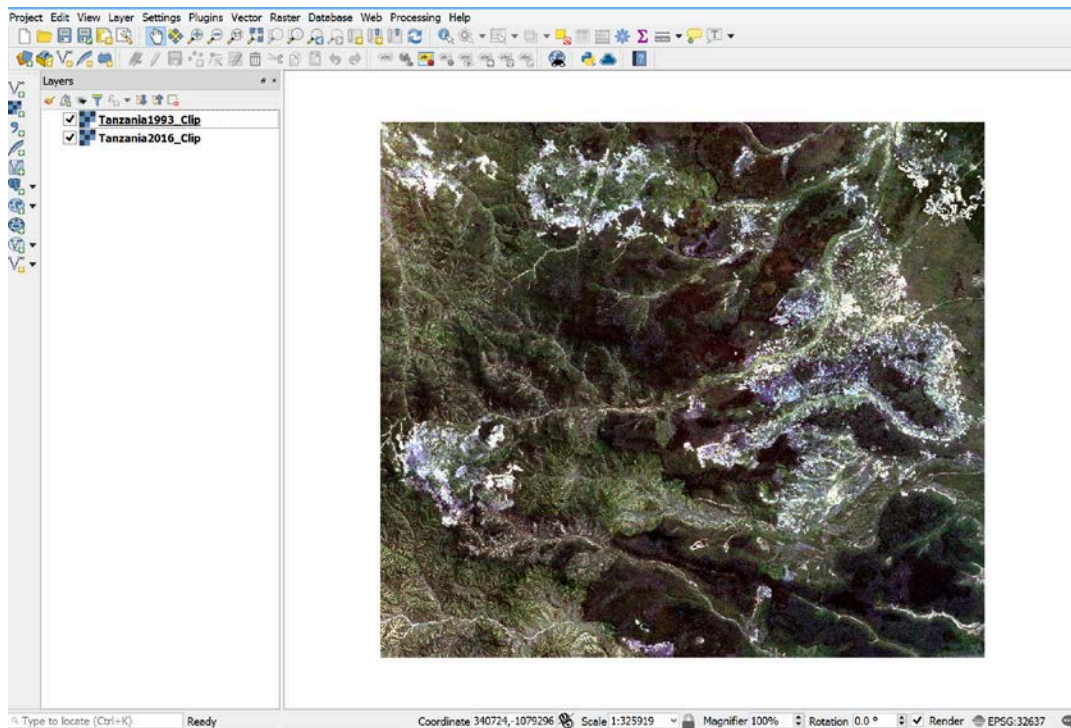
For this exercise we will be examining vegetation changes in Tanzania from 1993 to 2016 using a supervised classification technique. We will use Landsat data from these two dates. For the first image in 1993 we will use data from Landsat 5, and the 2016 imagery will be from Landsat 8. We will stack the bands from both images together and use the random forest algorithm to generate a change image. This image will indicate regions where forest cover has increased, decreased, or remained the same. This technique can be applied to many different regions and could include different types of image classification for your region of interest.

## Part 1: Band Stacking and Image Enhancement

1. Open QGIS
2. On the top panel click on **Layer > Add Layer > Add Raster Layer**



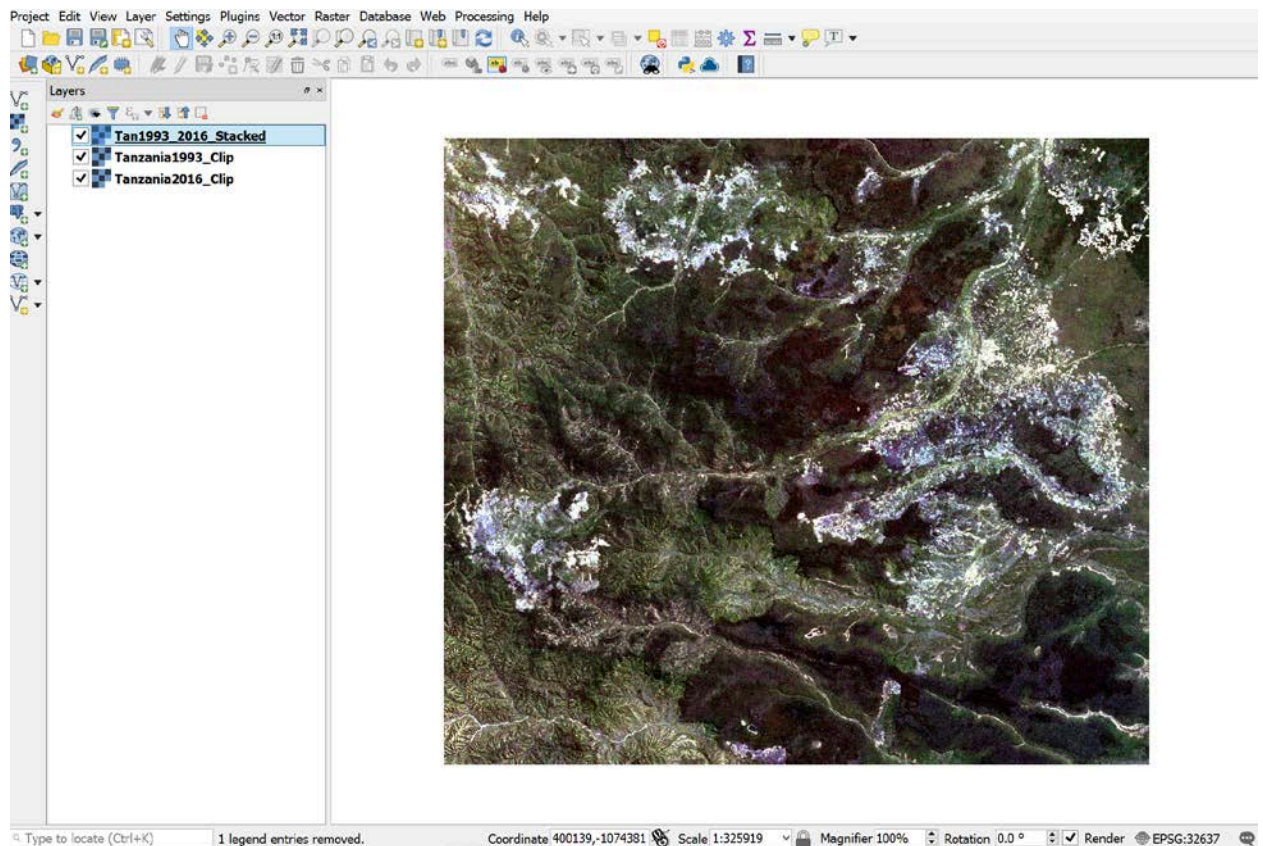
3. Navigate to your **Exercise1** folder and add the Tanzania1993\_Clip.tif and the Tanzania2016\_Clip.tif click **Add**, then **OK**, then **Close**. Make sure the 1993 image is on top of the 2016 image in the **Layers** panel.



4. Click on **Raster > Miscellaneous > Merge** to stack the bands of the two images together into one image
  - a. Click on the  next to **Input Layers** and click on **Select all** to select each image. You must also ensure that the 1993 image is on top of the 2016 image in the **Input Layers**. Click **OK**.
  - b. Select the **Place each input file into a separate band** button
  - c. Keep the **Output data type** as the **Float32** default
  - d. Click on the  next to **Merged** and select **Save to File**. Save the new file as **Tan1993\_2016\_Stacked.tif** in your **Exercise2** folder.
  - e. Click on **Run in Background**
  - f. Once the processing is complete, click **Close**
5. The new temporary merged image will be added to the map. Navigate to your **Exercise2** folder and click on **Layer > Add Layer > Add Raster Layer** and select the **Tan1993\_2016\_Stacked.tif** click **Add**, then **OK**, then **Close**.
6. Remove the temporary merged image by right clicking on it in the **Layers Panel** and selecting **Remove Layer**. Click **OK** when the pop-up window appears.



## Advanced Webinar: Change Detection for Land Cover Mapping September 28 & October 5, 2018



7. Click on the **Save Project As**  icon to save your project as **Exercise2.qgz**

Next, we will enhance the image like we did in the week 1 exercise. Note that this new image will have all the bands from the 1993 image and the 2016 image. Different band combinations enable you to visualize the imagery in different ways. There are several different band combinations that are typically used for looking at vegetation change including true color, color infrared and false color composite (see link below).

<https://landsat.usgs.gov/how-do-landsat-8-band-combinations-differ-landsat-7-or-landsat-5-satellite-data>

In this exercise we will view the image using the false color composite band combination. It is helpful if we create one band combination for the first time period, one for the second time period and one that contains bands from both dates to highlight changes. In QGIS you can create duplicates of your 2-date image and then assign the unique band combinations to each.

8. To create a duplicate, right-click on **Tan1993\_2016\_Stacked.tif** and choose **Duplicate**. This will create a new image with the **suffix \_copy**, to which you can

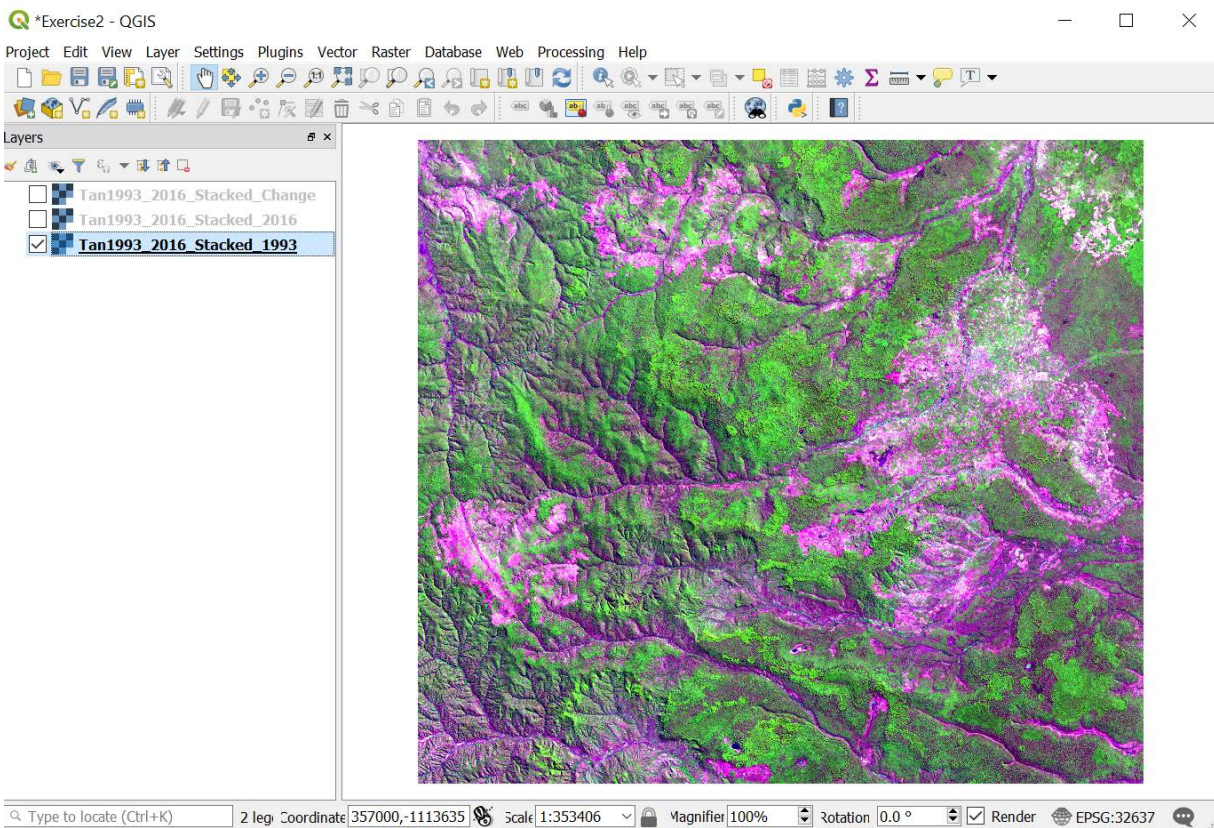




- assign different band combinations.
9. Rename the first duplicate by right clicking on the copy and selecting **Rename Layer**. Change the name to **Tan1993\_2016\_Stacked\_1993.tif**.
  10. Create another duplicate and change the name to **Tan1993\_2016\_Stacked\_2016.tif**.
  11. Rename the original **Tan1993\_2016\_Stacked** image to **Tan1993\_2016\_Stacked\_Change**
  12. Remove the **Tanzania1993\_Clip** and the **Tanzania2016\_Clip** images.
    - a. You can remove multiple images by clicking on them in the **Layers Panel** and right click and select **Remove Layer**
  13. Turn off all layers except **Tan\_1993\_2016\_Stacked\_1993**
  14. Right click on the **Tan\_1993\_2016\_Stacked\_1993** image in the **Layers Panel** and click on **Properties**. A window will pop up and you should be taken directly to the **Symbology** tab.
    - a. Leave the **Render type** as the default **Multiband color**
    - b. Change the **Red band** to **Band 5**
    - c. Change the **Green band** to **Band 4**
    - d. Change the **Blue band** to **Band 3**
  15. Click **Apply**
  16. Click on the drop down arrow next to **Min/max value settings**
  17. Turn on the **Mean +/- standard deviation** and ensure the value is set to 2.00 (this should be the default)
  18. In the drop down next to **Statistics extent**, you can leave the default as **Whole raster**
    - a. Click **Apply**, then **OK**.



## Advanced Webinar: Change Detection for Land Cover Mapping September 28 & October 5, 2018

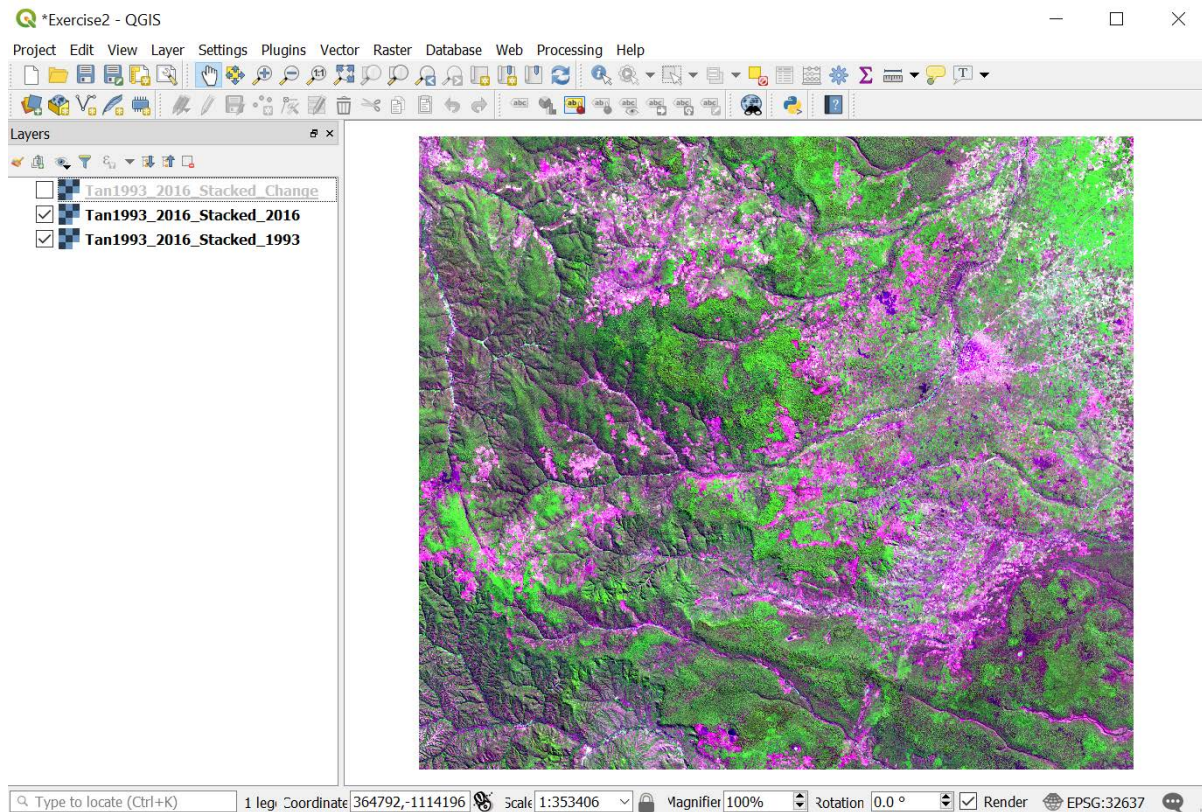


19. Repeat steps 13-17 for the **Tan\_1993\_2016\_Stacked\_2016** image. All steps will be the same, except for the color combinations. Remember the 2016 bands are stacked after the 1993 bands, so use:
- Change the **Red band** to **Band 12**
  - Change the **Green band** to **Band 11**
  - Change the **Blue band** to **Band 10**

Note that the bands 12, 11, and 10 correspond to bands 6, 5, and 4 of the 2016 Landsat 8 image. This is because we have six bands from our 1993 Landsat 5 image stacked on top of the 2016 Landsat 8 bands. This is the same false color display.



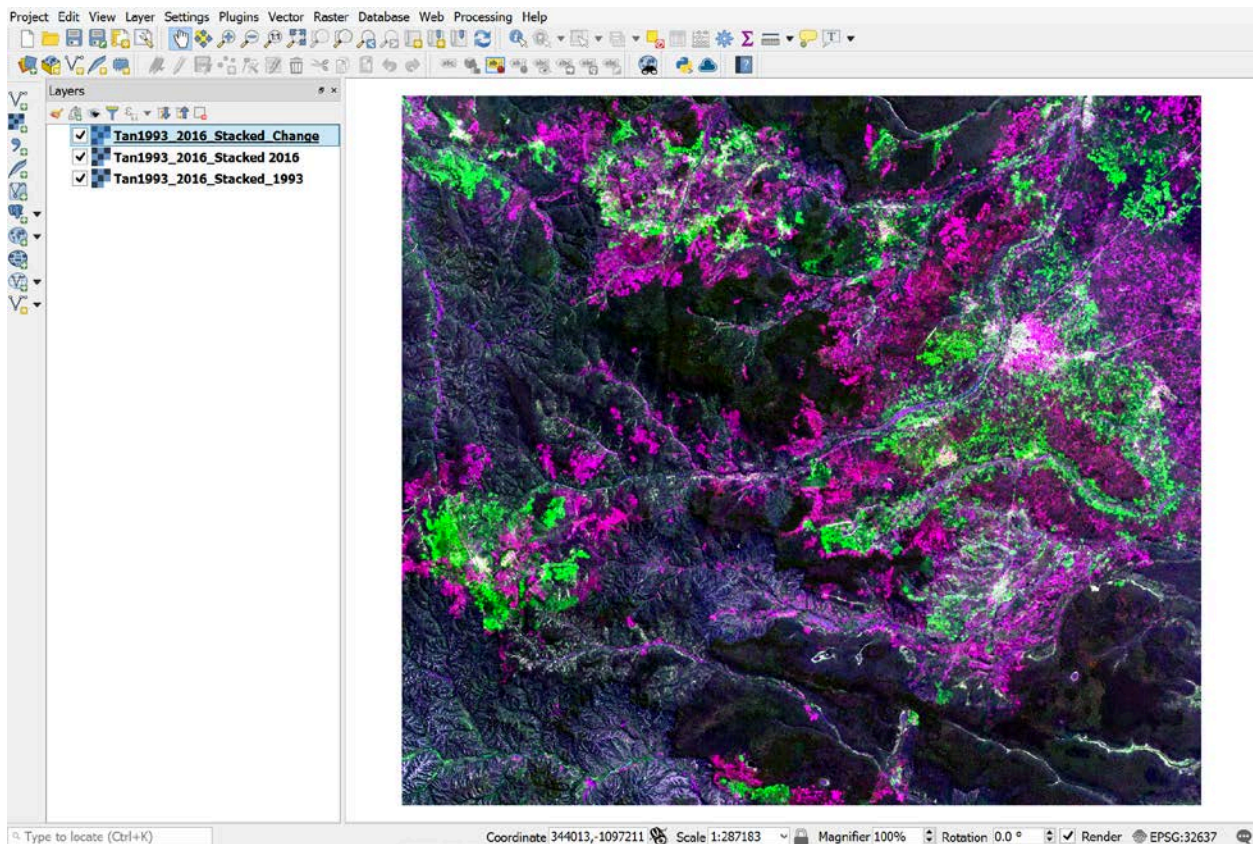
## Advanced Webinar: Change Detection for Land Cover Mapping September 28 & October 5, 2018





20. Repeat steps 13-15 for the **Tan\_1993\_2016\_Stacked\_Change** image. This time, change the bands to:
  - a. Change the **Red band** to **Band 13**
  - b. Change the **Green band** to **Band 5**
  - c. Change the **Blue band** to **Band 8**
21. This time choose the **Cumulative count cut** option under **Min/ max values settings** and leave the default as **2.0** and **98.0%**.
22. In the drop down next to **Statistics extent** choose you can leave the default as **Whole raster**.
23. Click **Apply**, then **OK**.

You will notice that the deforested areas are highlighted in magenta/purple and the areas that have increased vegetation are highlighted in green. This band combination uses bands from both dates in a composite to highlight changes in the Shortwave Infrared (SWIR) bands 13 (or the corresponding band 7 from the 2016 Landsat 8 image) and 5 (from the 1993 Landsat 5 image). Using this combination, you can see changes in the vegetation from the first date to the second date.





## Part 2: Developing Training Sites

1. Create a new polygon shapefile by clicking on **Layer** along the top of menu. Click on **Layer > Create Layer > New Shapefile Layer**. Or click on the **New Shapefile Layer**  icon from along the left side.
  - a. **File name:** Click on the  next to **File name** and navigate to your **Exercise2** folder and name the file **training.shp**. Click **Save**.
  - b. **File encoding:** **UTF-8** (may be the default)
  - c. **Geometry type:** Use the drop-down menu to choose **Polygon**
    - i. Under **Geometry type:** **EPSG:32637 - WGS 84 / UTM zone 37N – Projected** OR **Project CRS: EPSG:32637 – WGS 84 / UTM zone 37 N** (the name might differ slightly depending on if you are working on a Mac or PC).
    - ii. This is done to ensure our reference system is the same as the Landsat image.
  - d. Under **New field > Name:** type **Notes**
  - e. **Type:** Text data





- f. **Length:** 20
- g. Click **Add to fields list**
- h. Click **OK**

**New Shapefile Layer**

File name: ments\ARSET\ChangeDetection\Week2\training.shp

File encoding: UTF-8

Geometry type: Polygon

Include Z dimension  Include M values

Project CRS: EPSG:32637 - WGS 84 / UTM zone 37N

**New field**

Name: [ ]

Type: abcText data

Length: 20 Precision: [ ]

[Add to fields list]

**Fields list**

Name	Type	Length	Precision
id	Integer	10	
Notes	String	20	

[Remove field]

OK Cancel Help

In order to create training sites, we want to identify areas where we know the land cover category and the type of change observed from 1993 to 2016. For this exercise, we will use these categories and ID code:



Cover Class	Code
Forest to Forest	11
Forest to Non-Forest	12
Non-Forest to Non-Forest	22
Non-Forest to Forest	21



To keep the exercise simple, we will use two types of land cover: forest and non-forest. The categories will either remain the same from 1993 to 2016 (forest to forest and non-forest to non-forest), or they will change to the other land cover category (forest to non-forest or non-forest to forest). For future change detection of different regions, you will likely have more categories of change (e.g. bare ground, water, or urban).

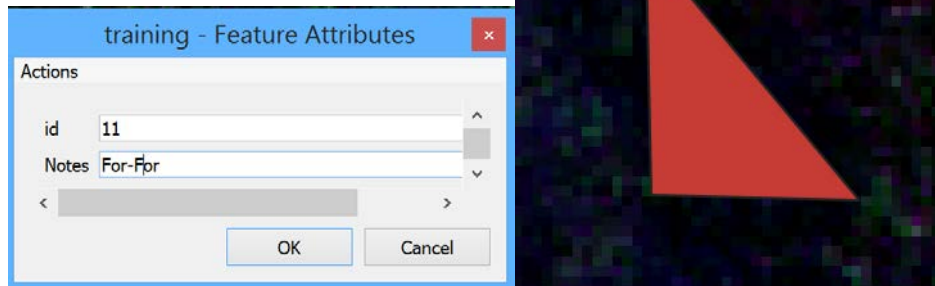
1. Make sure the layers are in this order in the layers Panel: (1) training polygon, (2) the change image, (3) the 1993 image, and (4) the 2016 image.
2. First we will delineate training sites where there is forest in 1993 and 2016 (no change). Look at the change image first. Areas that are dark are generally areas that have not changed. Confirm that this is true by clicking off the change image and toggling between the 1993 and 2016 images.
  - a. You can turn the layers on and off by clicking on the check mark next to each image in the **Layers Panel**.

Once you have found a region with no change from forest to forest, you will draw a polygon around those pixels to create the training site. You can zoom in and out of the images and toggle them on and off until you are confident that you have found a region with no change from forest to forest from 1993 to 2016.

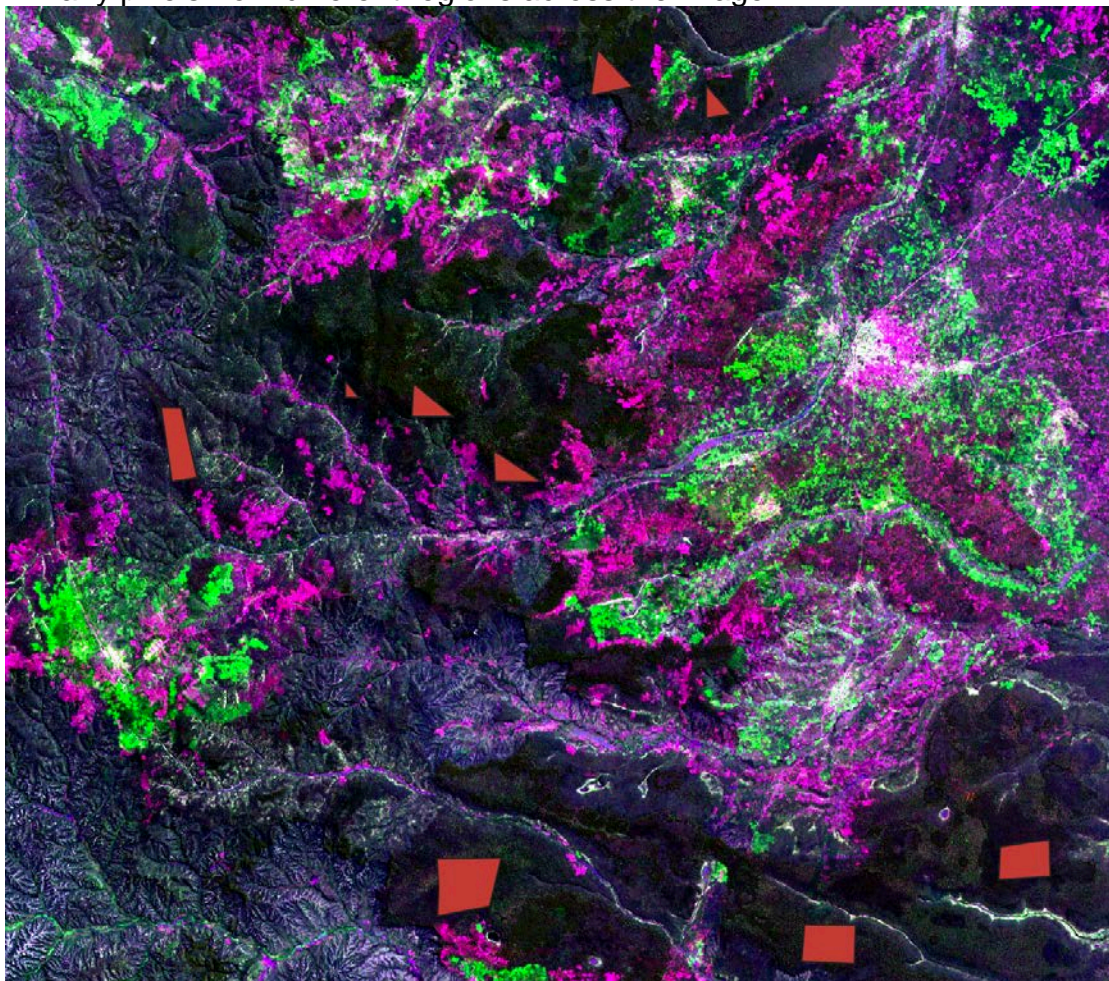
3. To draw polygons, right click on the **training** layer in the **Layers Panel** and select **Toggle editing** or click on the **Toggle editing**  icon while the vector layer is highlighted in the **Layers Panel**. This will allow you to edit the polygon shapefile. These polygons will be used as the training sites for the classification.
4. To add features to the layer (which will be used as training sites), click the **Add Polygon Feature**  icon in the **Digitizing Toolbar** or click on **Edit** and select **Add Polygon Feature**. After you select **Add Polygon Feature** a crosshair will appear when you hover your mouse over the image.
5. Click on the image to draw an outline of the polygon that only includes pixels with no change from forest to forest. Right click to close the polygon.
6. A dialog box will appear to provide attribute information. Give this polygon the **ID 11**. In the **Notes** put **For-For**. Click **OK**.



Advanced Webinar: Change Detection for Land Cover Mapping  
September 28 & October 5, 2018



7. Repeat steps 4-6 this for other unchanged forested areas in the image until you have at least 10 trainings sites with **ID 11**. Ensure that you create polygons with many pixels from different regions across the image.







Next, we will create training sites that were forest in 1993, but were bare ground (non-forest) in 2016. These areas will be purple/magenta in the change image. Again, you can confirm this type of change by clicking on and off the 1993 and 2016 image.

8. Repeat steps 4-6 for the forest to non-forest category. Give each of these training polygons the **ID 12** and in the notes type **For-NonFor**.
9. If you are having trouble seeing the polygons you are creating, you can change the color to red. Right click on the training layer in the **Layer Panel**. Click on **Styles** and click on the red area of the color wheel. This will automatically change all the polygons to red.



Next, we will create training sites that were non-forest in both 1993 and 2016. These areas will be indicated in white on the change image. You will notice a town in the north central region of the image; this is a good area for indicating this category of non-forest to non-forest. Again, you can confirm this type of change by clicking on and off the 1993 and 2016 image.

10. Repeat steps 4-6 for the non-forest to non-forest category. Give each of these training polygons the **ID 22** and in the notes type **NonFor-NonFor**.

Finally, we will create training sites that were non-forest in 1993, but were forest in 2016. These areas will be indicated in green on the change image. Again, you can confirm this type of change by clicking on and off the 1993 and 2016 image.

11. Repeat steps 4-6 for the non-forest to non-forest category. Give each of these training polygons the **ID 21** and in the notes type **NonFor-For**.

After you have created all your training sites for the four categories, you need to save the layer edits.

12. Click on the **Save Layer Edits**  icon or right click on the training layer in the **Layers Panel** and click on **Save Layer Edits**
13. Turn Toggle editing off by clicking on the **Toggle editing**  icon so that it is no longer highlighted
14. Remember to always **Save** your map along the way



## Part 3: Random Forest Classification Algorithm

Now we run the random forest classification algorithm to generate the classified change map. We will do this using the R and R Studio software.

1. Open R Studio
2. Along the top panel, click on **File > Open File**. Navigate to your **Exercise2** folder and open the **classification.R** file.

```
1 | #####
2 | # The script reads an ESRI Shapefile (defined by the "shapefile" variable) with
3 | # training polygons and then either selects all pixels or randomly selects a
4 | # user-determined number of samples (defined using classNums and classSampNums)
5 | # for each land cover type. A multilayer image that contains spectral, other
6 | # continuous data or categorical data is also input (defined by the inImage
7 | # variable). For each randomly selected sample the data values for that pixel
8 | # are determined and these data are used to run the Random Forest model.
9 |
10 | # After building the model the multilayer image is read, and up to three output
11 | # images (classImage, probImage, threshImage) can be selected.
12 | # "classImage" classifies all of the pixels.
13 |
14 | # "probImage" outputs the class probability of the class that got the most vot
15 | # (i.e., the class that was selected for the classImage layer).
16 |
17 | # "threshImage" is the same as "classImage" except all pixels with a class pro
18 | # of the class that got the most votes below the "probThreshold" parameter ar
19 | # this is useful to identify pixels with inter-class confusion.
20 |
21 | # The images are written out using the GeoTIFF format and the file name is creat
22 | <
```

```
R version 3.5.1 (2018-07-02) -- "Feather Spray"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

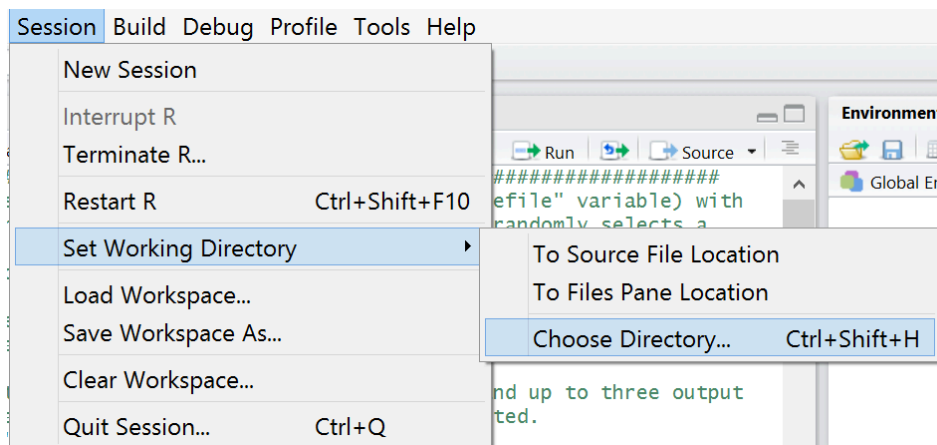
> |
```

As a first step it is good to set the working directory to the folder that contains the training site shapefile and the 2-date image stack

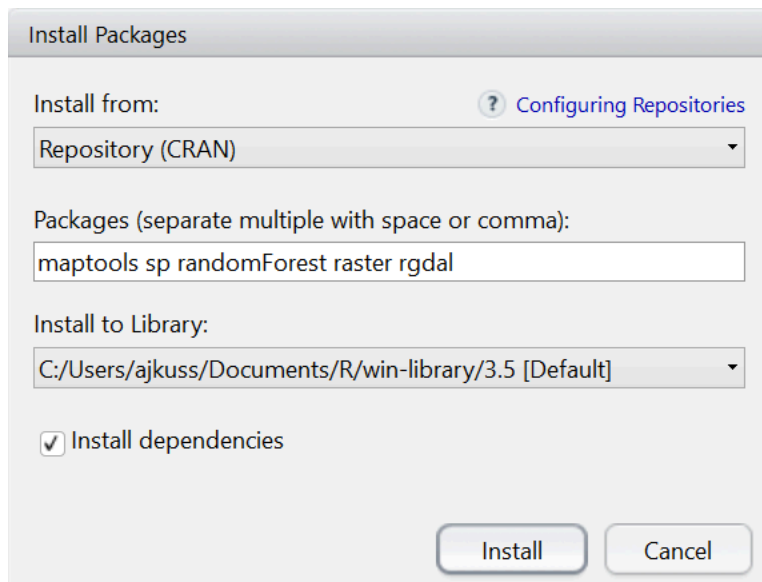
3. Go to **Session > Set working directory > Choose Directory** and set it to your **Exercise2** folder
  - a. You will need to ensure that all of your images and your **training.shp** file are located in this folder without any subfolders.



## Advanced Webinar: Change Detection for Land Cover Mapping September 28 & October 5, 2018



4. Next, we need to install a few specific packages. Go to **Tools > Install Packages**.
5. Type in following packages separate by spaces: **mapproj, sp, randomForest, raster, rgdal**
  - a. **Note:** as you type these package names, they will likely appear in a generated drop-down list, which you can select to ensure the name is correct. Also, if you are working on a Mac, you may also need to install the **rgeos** package.
6. Click **Install**
  - a. The process will run and may take a few minutes to ensure each package is installed. Once the packages are installed, the R console will display a blue **>**.







Next, we will need to change a few lines of the R code specific to the location of the images and the training.shp for your individual computer.

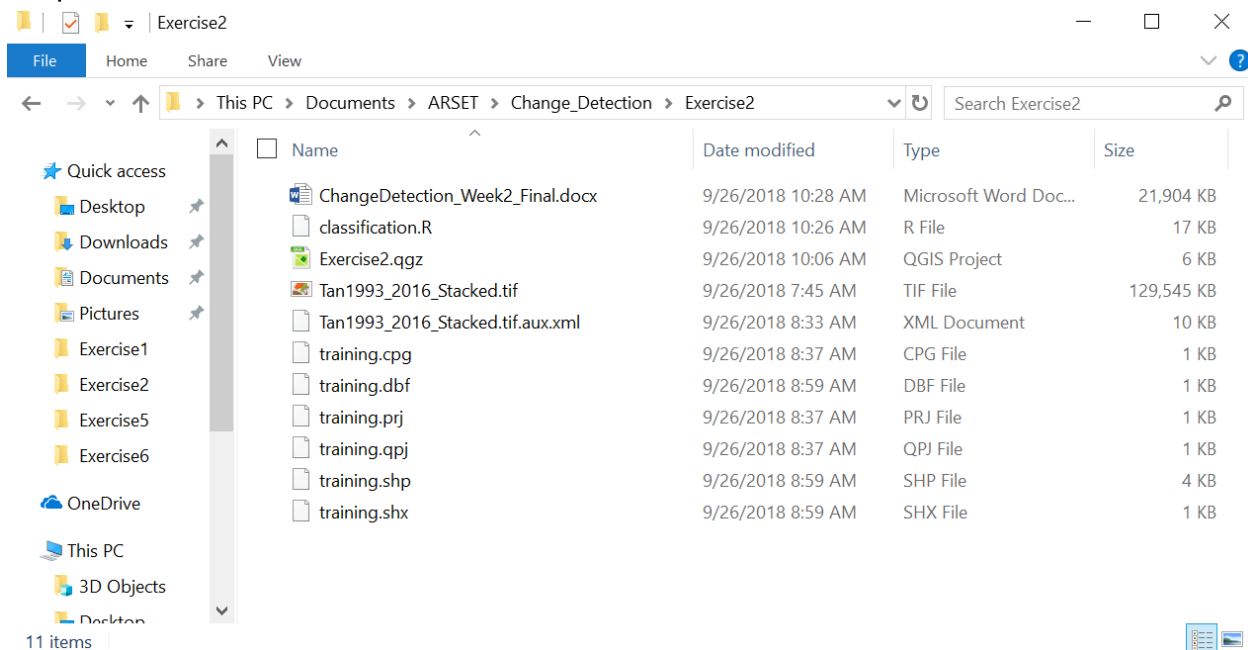
First, set the working directory:

1. Scroll down to line 65 of the R code that starts with setwd. You will need to change this line to identify your working directory.
2. Change the parts in red to navigate to your **Exercise2** folder on your computer. You can see what this path looks like by going to your windows folder or your Finder window if you are working on a Mac. Also note that on a Mac, you can navigate to your **Exercise2** folder, click on **File > Get Info**, and copy the path next to **Where**.

Initial Directory:

```
setwd("C:/Users/amccullu/Documents/ARSET/Change_Detection/Exercise2")
```

**\*Note:** that in this example, this path navigates to a folder called Exercise2. It does not necessarily matter what you name your folder, so long as it is correctly listed in the script.



Next, change the name and path for your training shapefile:



3. Scroll down to line 67 of the R code. Change the parts in red to navigate to your training shapefile in your Exercise2 folder.

Initial Directory:


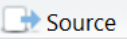
shapefile <-

`'/Users/amccullu/Documents/ARSET/Change_Detection/Exercise2/EX2_Practice/trainin  
g.shp'`

You will leave all other parts of the R code as default. If you are interested, you can scroll through the R code to learn about what the steps are. For example:

- Line 69: `classNums <- c(11,12,21,22)`: These are the classes we identified in the training shapefile
- Line 72: `classSampNums <- c(100,100,100,100)`: These are the approximate number of training samples to be randomly selected for running the Random Forest algorithm
- Line 74: `attName <- 'id'`: This is the attribute that the code uses to run the classification
- Line 76: `nd <- -9999`: This is the non-data value
- Line 78: `inImageName <- 'Tan1993_2016_Stacked.tif'`: This is the input image that the code is using for the classification

You can read through the remainder of the code and read the notes (with the # in green) to see what the code is doing.

7. Save the R code, be clicking on the Save  icon along the top.
8. Click on the **Source**  icon along the top to run the entire code
  - a. You will see each of the steps run in the R Console along the bottom
9. A message will come up that says: **Type n to stop, c to change feature space bands, s to define a rectangle to locate gaps in feature space, or any other key to continue with random forests model creation and prediction:**
10. Click on the **K** key, and the remainder of the code will run. As the process runs, it will show this along the bottom as each block runs:

```
calculating random forest object  
starting predictions  
processing block 4 of 4
```

The R code will create a confusion matrix and will display it in the R console as well. When the entire process is complete, the R console will display the processing time and




you may see a warning message that says: **readShapePoly is deprecated; use rgdal::readOGR or sf::st\_read**. This is OK. The process should have run properly.

11. You can double check to see if Random Forest Algorithm ran. Navigate to your **Exercise2** folder and find the **Tan\_1993\_2016\_Stacked\_Class.tif**.
12. Close R Studio. A pop-up will come up to ask if you would like to **Save the Workspace Image**. Click **Yes** or **Save** on a Mac.

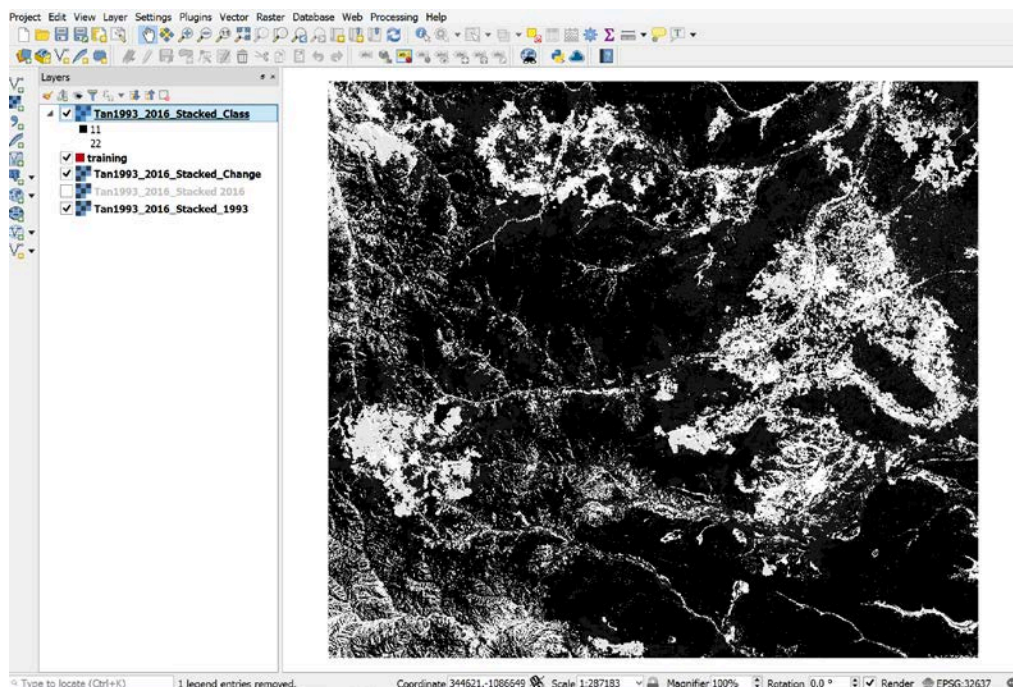
## Part 4: Post-Processing of the Classified Image

In this final step, we will go back to QGIS and view, color, and “clean up” our classified image

1. Go back to QGIS. Click on the **Add Raster Layer**  icon.
2. Navigate to your **Exercise2** folder and add the **Tan1993\_2016\_Stacked\_Class.tif** click **Add**, then **OK**, then **Close**.

You will notice that the image has values from 11 to 22, which are actually the categories that we defined in part 3 of this exercise.

**\*Note:** If your image contains values from 1 to 4, assume that 1=11, 2=12, 3=21, and 4=22. You can input the correct values into the file in step 4 below. This might be a difference in the way QGIS behaves on a Mac or a PC.



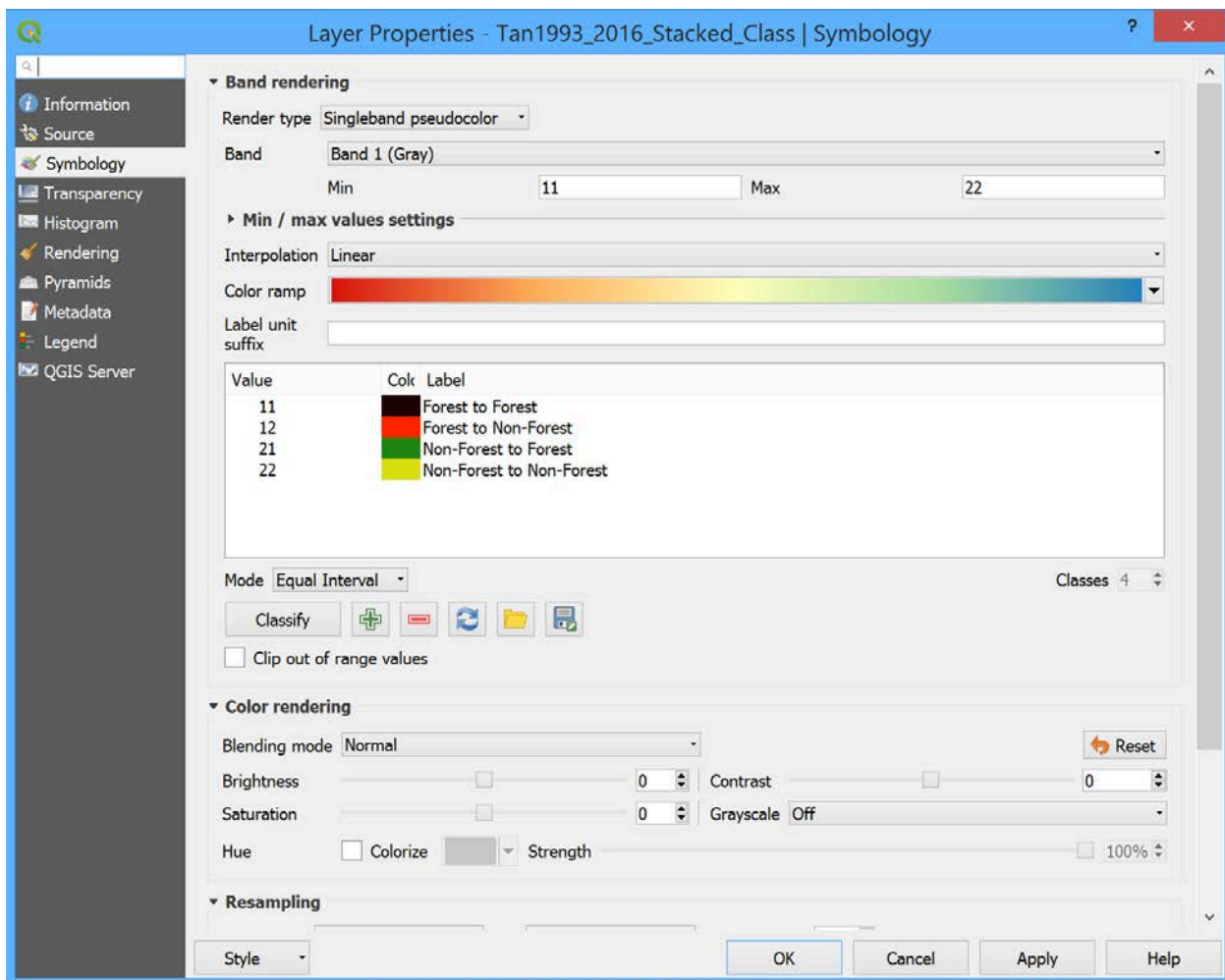




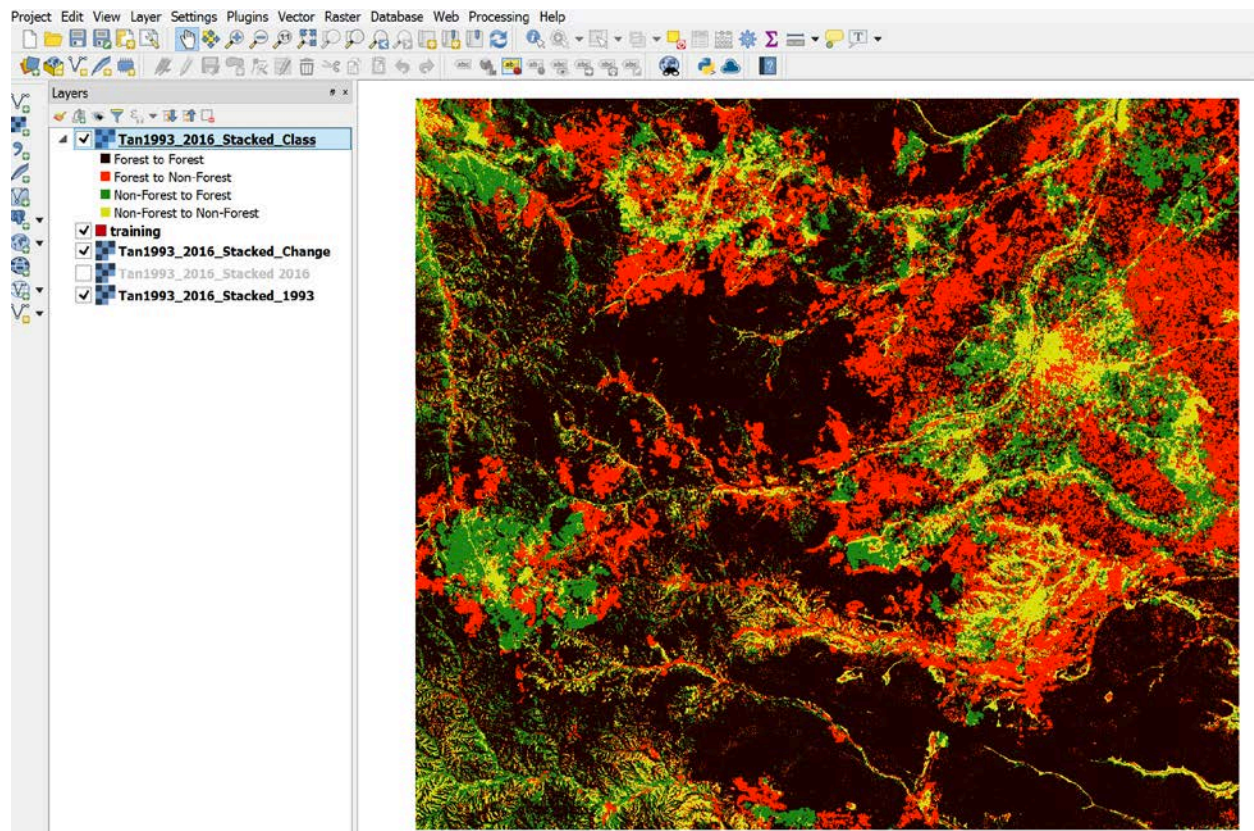
3. Right click on the **Tan1993\_2016\_Stacked\_Class** image in the **Layers Panel**.  
Click on **Properties > Symbology**.
  - a. Render Type: Singleband Pseudocolor
  - b. Interpolation: Linear or Exact (both will work)
  - c. Mode: Equal Interval
  - d. Change the Number of Classes to 4
  - e. Click on Classify.
4. Change the Values and Labels to:
  - a. Value: 11 or 1; Color: black; Label: Forest to Forest
  - b. Value: 12 or 2; Color: red; Label: Forest to Non-Forest
  - c. Value: 21 or 3; Color: green; Label: Non-Forest to Forest
  - d. Value: 22 or 4; Color: yellow; Label: Non-Forest to Non-Forest
5. To change the colors, you can double click on the color icon to select colors that you think are appropriate to identify the categories. You can choose different colors than the example provided if you would like.
6. Click **Apply**



## Advanced Webinar: Change Detection for Land Cover Mapping September 28 & October 5, 2018



7. Click on the **Export Color Map to File**  icon
8. Navigate to **Exercise2** folder and save the file as **classcolors.txt**
9. Click OK to close the **Layer Properties** window

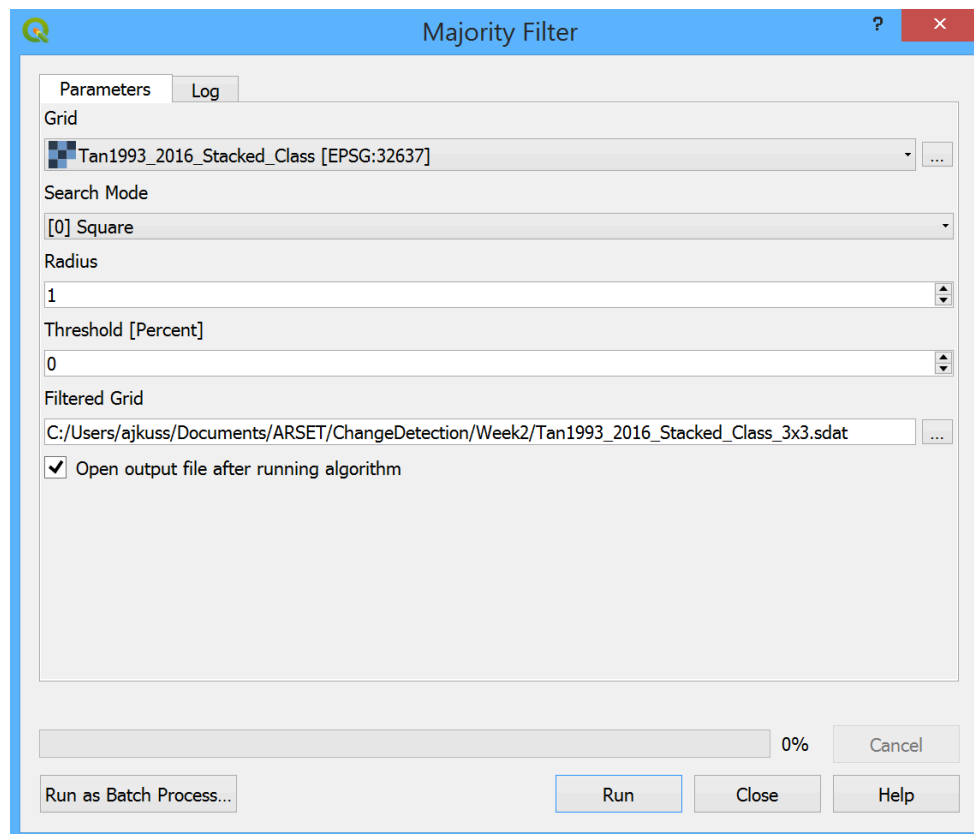



Next we will use the **Majority Filter** to remove some of the “noise” or speckles of the image.

10. Click on **Processing > Toolbox**. The Toolbox should appear on the right.

11. Type in **Majority Filter** into the search box and double click on the **Majority Filter** tool.

- a. Grid: Ensure that the **Tan1993\_2016\_Stacked\_Class [EPSG:32637]** is selected. This should be the default.
- b. Leave the other attributes as the default and click on the  next to **Filtered Grid**. Navigate to your **Exercise2** folder and save the new file as **Tan1993\_2016\_Stacked\_Class\_3x3.sdat**.



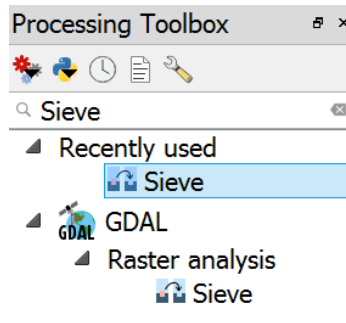
12. Click **Run**
13. The process will bring up a new pop-up window to run and when the processing is complete the new temporary **Filtered Grid** will be added to the map
14. Click **Close** to close the **Majority Filter**.
15. Remove the temporary **Filtered Grid** file from the map and add in **Tan1993\_2016\_Stacked\_Class\_3x3.sdat** using the **Add Raster Layer** function
16. Right click on the **Tan1993\_2016\_Stacked\_Class\_3x3.sdat** and click on **Properties > Symbology**
  - a. Render Type: **Singleband Pseudocolor**
  - b. Under Style, select **Load color map from file** 
  - c. Open the **colorclass.txt** file
  - d. Click **Apply** and **OK**
17. Close the **Properties** window. Click on and off the **Tan1993\_2016\_Stacked\_Class\_3x3** image and you can see that the speckles or noise has been removed from the new image.

Now we will conduct one final post-processing step called Sieve. This will remove the small clumps of pixels that are incorrectly classified.

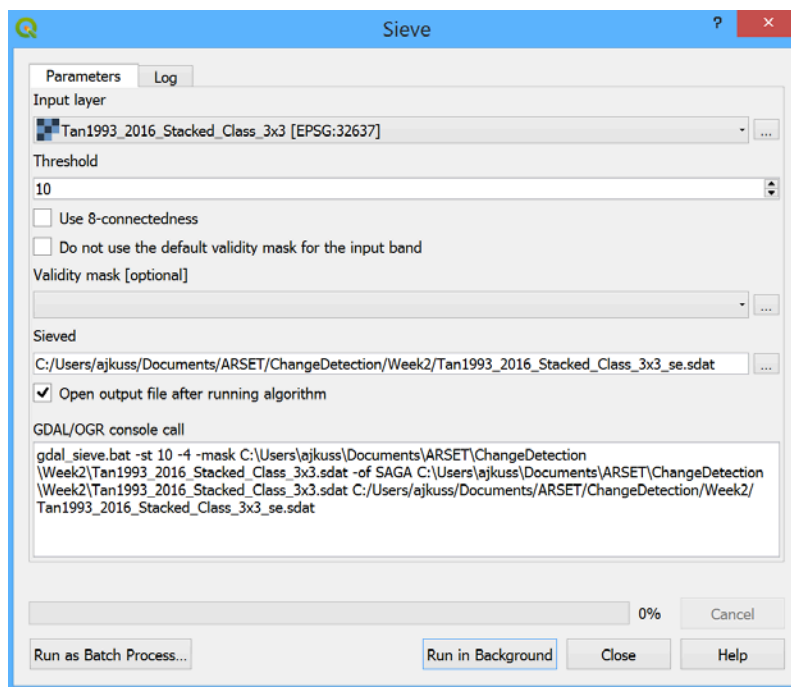




18. Click on **Processing > Toolbox**. The Toolbox should appear on the right.
19. Type in **Sieve** into the search box and double click on the **Sieve** tool.



20. Input these parameters into the Sieve tool:
  - a. Input Layer: **Tan1993\_2016\_Stacked\_Class\_3x3.sdat**
  - b. Threshold: **10** (this should be the default)
  - c. Click on the **Use 8-connectedness** box
  - d. Click on the [...] next to **Sieved**: Save to file as **Tan1993\_2016\_Stacked\_Class\_3x3\_se.sdat**
  - e. Click on: **Run in background**



21. Click **Close** to close the **Sieve** tool
22. Remove the temporary **Sieved** file from the map and add in **Tan1993\_2016\_Stacked\_Class\_3x3\_se.sdat** using the **Add Raster Layer**



function

23. Right click on the **Tan1993\_2016\_Stacked\_Class\_3x3\_se.sdat** and click on **Properties > Symbology**

a. Render Type: **Singleband Pseudocolor**

b. Under Style, select **Load color map from file** 

c. Open the **colorclass.txt** file


d. Click **Apply** and **OK**

24. Close the **Properties** window. Click on and off the **Tan1993\_2016\_Stacked\_Class\_3x3\_se** image and you can see that this was another improvement in the classification.

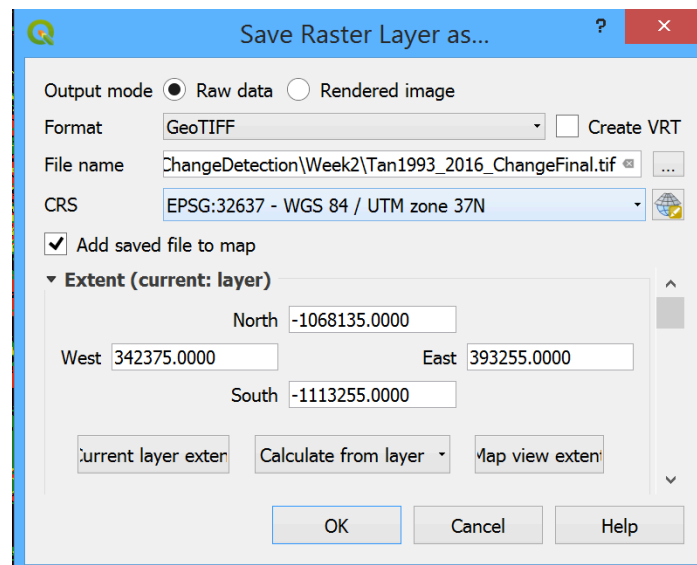
In order to have an accurate classification it often takes multiple iterations of creating training sites and re-running the Random Forest Algorithm. While we will not step through this process again, it will be useful for you in your specific study area of interest.

25. Export the final map product. Right click on the

**Tan1993\_2016\_Stacked\_Class\_3x3\_se.sdat** and click on **Export > Save As**

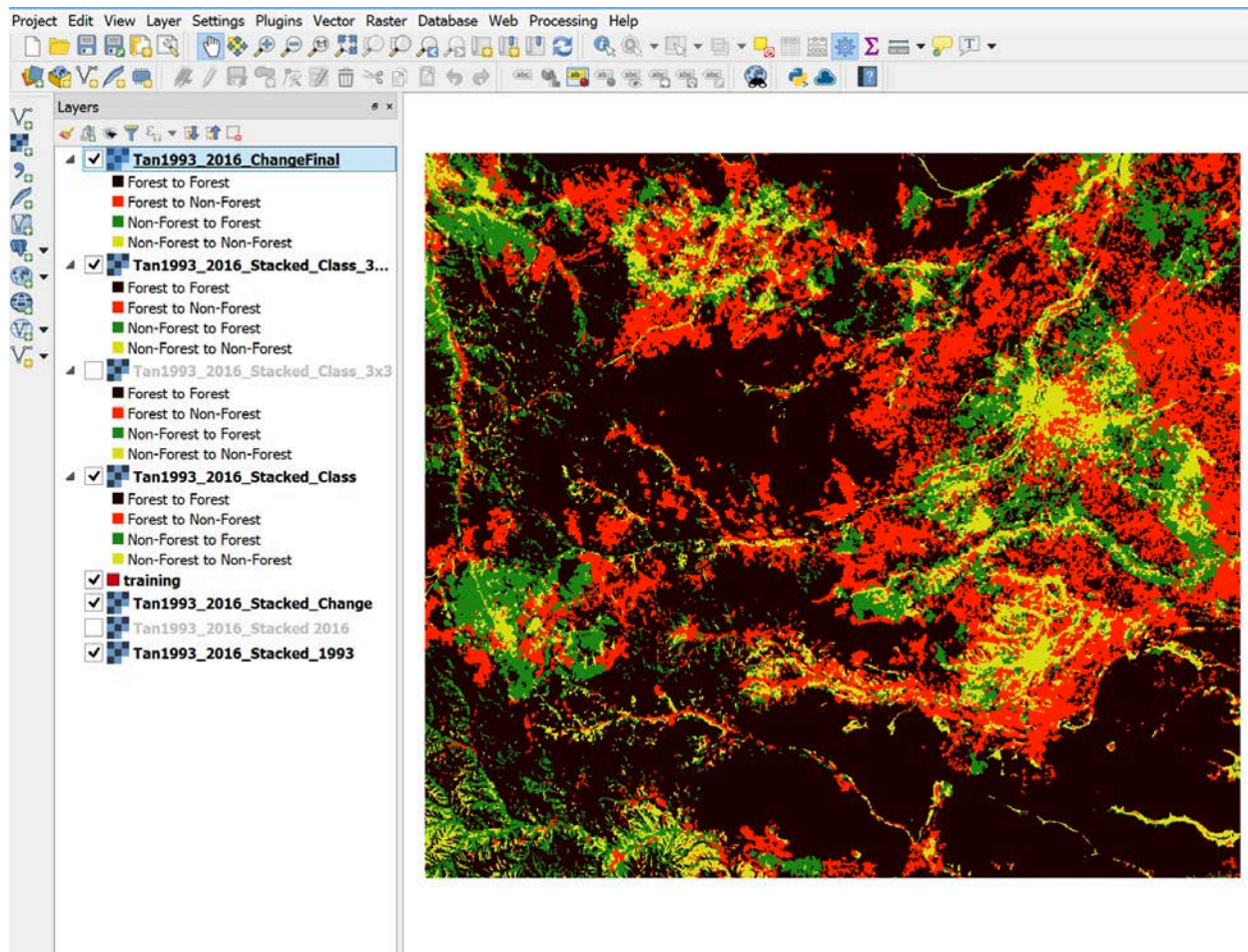
26. Click on the  next to File name. Navigate to your **Exercise2** folder and save the file as **Tan1993\_2016\_ChangeFinal.tif**

27. Click **OK**



28. Your new image will be added to your map. You can change the color scheme to the same as the other if you would like.

29. **Save** your map and quit QGIS.



## Conclusion

In this exercise, we used two images (from 1993 and 2016) to identify change in the landscape. First, we stacked the two images to use the bands from each image to identify regions where forest had increased, decreased, or did not change. Then, we generated training sites that were used in the Random Forest classification algorithm to create a classified image that highlights change or no-change. Finally, we used filtering techniques to reduce noise or uncertainty in the image. These techniques can be applied to many different landscapes that may have different land cover types and changes.



## References

Much of these exercises were modified from previous training materials created by Ned Horning, the Director of Applied Biodiversity Informatics, at the American Museum of Natural History (AMNH) and adopted by Conservation International (CI) for remote sensing training. Please see the links and references below for more information:

AMNH Biodiversity Informatics: <https://www.amnh.org/our-research/center-for-biodiversity-conservation/capacity-development/biodiversity-informatics>

Horning, N. 2013. Training Guide for Using Random Forests to Classify Satellite Images - v9. American Museum of Natural History, Center for Biodiversity and Conservation. Available from <http://biodiversityinformatics.amnh.org/>. (1/3/2014).

Conservation International: <https://www.conservation.org/Pages/default.aspx>