



## Questions & Answers Session 2

Please type your questions in the Question Box. We will try our best to get to all your questions. If we don't, feel free to email Cynthia Schmidt ([cynthia.l.schmidt@nasa.gov](mailto:cynthia.l.schmidt@nasa.gov)) or Amber McCullum ([amberjean.mccullum@nasa.gov](mailto:amberjean.mccullum@nasa.gov)).

Question 1: How can I apply the cloud elimination algorithm?

Answer 1: There is no set algorithm. You will need to use the QA image provided with the Landsat SR product and then you can use Raster calculator to eliminate those pixels in the image that correspond with clouds and cloud shadows. What we may do is offer a short webinar on how to eliminate clouds. Here is the QGIS plugin for cloud masking: <https://plugins.qgis.org/plugins/CloudMasking/>

Question 2: Does cloud mask address cloud shadows?

Answer 2: Yes, the QA data from the Landsat SR product includes cloud shadows

Question 3: Which of the two ways do you recommend to improve the visualization of the image? The stretch or the combination of bands? How to select one or the other way?

Answer 3: Usually you use both. You need to stretch the image to see the features better and you need different combinations of bands to see the image in color. The band combination that you choose depends on your preference and what features you are interested in. Here's a great link that describes different band combinations (for Landsat 8) for looking at different features : <https://www.esri.com/arcgis-blog/products/product/imagery/band-combinations-for-landsat-8/>

Question 4: How can we reduce spectral mixing of different classes during unsupervised classification (eg: two different types of vegetation having similar spectral reflectance). Is it an inherent property of satellite data?

Answer 4: A Landsat pixel averages the spectral reflectance of everything on the ground within that pixel. In classification processes, there's no way to get around that. However, there are some unmixing algorithms that can, to some extent, give you an idea of what percentage of landcover is in each pixel



Question 5: In the Random Forest it is not required to select pure training areas ... Does this mean that you can select training areas without specific care? For example, can a class of latifolia be mixed with soil and other types of vegetation?

Answer 5: You can do that if you don't have a separate class for soil or other types of vegetation.

Question 6: Where can be found more information on the random forests algorithm?

Answer 6: If you want general information about RF, you can do a google search. If you want specific information about this code, you can go to the link that I provided in the chat from AMNH. You can find the original script and instructions for R and Random Forest, which was written by Ned Horning at the American Museum of Natural History, along with a lot of other great resources here: <https://www.amnh.org/our-research/center-for-biodiversity-conservation/capacity-development/biodiversity-informatics/open-source-software-and-scripts>

Question 7: Is it important to have roughly the same number of pixels in each class, when using this method?

Answer 7: No you don't need the same number of pixels in each class. You probably want to have a minimum number of pixels. Here this is 100 in the R script, so you will need to make sure to have at least that many in each class for this exercise.

Question 8: if I want to classify the type of crops in agriculture(ex: rice,wheat,soya), which one is best among pixel base and object based classification?  
and is it possible to classify crops at field level using sentinel data, can you please give suggestions or material links to know about crop type classification

Answer 8: This webinar is about change detection, not land cover classification. Send email, we might be able to direct you to agricultural classification.

Question 9: How do I use land change detection to map biological diversity and species distribution of animals with respect to change in their microhabitats?

Answer 9: There are papers for this. We cannot easily answer this here.

Question 10: can we do this process if we have 3 or more images of different years?

Answer 10: Yes, you can do this with multiple images (not just two dates). We will be doing a time series analysis webinar sometime in 2019. This allows you to do multiple



dates of imagery and process them. You can also check out the LandTrendr program:  
<http://landtrendr.forestry.oregonstate.edu/>

Question 11: Could you please briefly explain how the first loop results can be used to iterate the process?

Answer 11: In this code, you can look at your accuracy and evaluate the training sites. You can re-create your training sites and re-run the algorithm.

Question 12: Could you explain if we could use ground truth points to classify image in QGIS?

Answer 12: Yes, you can create a shapefile based on those ground points and go through the same process with points rather than polygon.

Question 13: what is the difference between .sdatt and tiff?

Answer 13: The .sdatt file is a different type of gridded file that the SAGA tools use in QGIS. You will need to create .sdatt files in the those final post-processing steps we did in Exercise 2, but you can always export (Save File As) the files as a tiff for future use. Here is a stack exchange explanation too:

<https://gis.stackexchange.com/questions/174785/saga-tools-in-qgis-only-allow-save-as-sdatt-file>

Question 14: Besides Random Forest, which other machine learning methods are out there for image classification?

Answer 14: We will need to investigate that. Here is a link from Research Gate:  
[https://www.researchgate.net/post/Different\\_deep\\_learning\\_methods\\_for\\_image\\_classification](https://www.researchgate.net/post/Different_deep_learning_methods_for_image_classification)

Question 15: Can I combine images with different spatial resolution (so for example Landsat with Sentinel 2) and run the randomForest algorithm for change classification?

Answer 15: You need same spatial resolution for change detection. It is looking at difference in value between two pixels (from different dates).

Question 16: You mentioned in Random Forest for training data it is important for covering the entire brightness range- how do you ensure your training data is doing that?



Answer 16: There are products produced from the R code that allow you to look at training site stats. This outlines how well you did in characterizing the spectral variability.

Question 17: why did you choose to do the classification on R studio not qGis or ENVI?

Answer 17: We wanted to use RandomForest, this is available via the R code. We use QGIS because it is free and open source, thus we try to use these for webinars, not ENVI which is commercial.

Question 18: How to deal with the same RF process if there is multiple years map ?

What will be the different band combinations for displaying the change such as (5,4,3/13,5,8....) ?

Answer 18: You can add a third year if you want, you just stack them all together. For processing you can use multiple dates. Check back for Time series webinar in 2019.

Question 19: does cloud elimination algorithm affects the reflection after removing clouds from image?

Answer 19: No. it does not affect the pixel values that don't have clouds after removing the clouds. You just pull out cloudy pixels from the image.

Question 20: Is it required to classify all spatial variation in an image that you are not interested in? Let's say for example I am interested in different forest types in wetlands, but I am not interested in other forest types

Answer 20: To use satellite imagery, you need to classify all the variation in the image. You can clip the image to decrease the number of vegetation types.

Question 21: If I don't have ground truth points, is there any other way to access the accuracy of classified results?

Answer 21: You can use high spatial resolution imagery (Google Earth) or other Landsat imagery, or DigitalGlobe (WorldView-2) imagery. There are varying levels of accuracy, depending on how high the spatial resolution is and how close to the date the image is to the image you are classifying.

Question 22: how much good is Molusce plugin of qgis for change detection

Answer 22: We have not used this plugin.



Question 23: Is the spatial resolution of the images important for picking pixel based classification? If one processes UAV images, maybe object based is better? Is the RF algorithm correct for such application?

Answer 23: Typically for high spatial resolution people like to use object-based, pure spectral based classification can be noisy and confusing.

Question 24: for which reason you have chosen standard deviation 2 DURING image processing ?

Answer 24: SD of 2 is just a typical stretch that captures most of the spectral range of the imagery. This is just a standard way to do this, you can choose something different.

Question 25: Can you use a tool in QGIS to determine the percentage of area in each classification, ie. the percent of Forest to Non-Forest change in the image?

Answer 25: Yes. We didn't demonstrate this. In QGIS you can look at the tools that include GRASS there is a r.stats tool. You can use this to calculate the statistics.

Grass> Raster > r.stats

Question 26: How far are we with Geospatial intelligence, in relation to change detection. Do you have any research being done on that or it has proved to be difficult to perform.

Answer 26: Sorry, but that is really outside of our field of expertise. All NASA Earth science data, technology and tools are available to the public. We typically do not do any Intelligence work.

Question 27: I would like to know more about moving window analysis and change vector analysis (Tasseled cap analysis).

Answer 27: That is a much more complicated process that we will not address during this webinar series. Here are a few papers on this more advanced processing:

- <https://www.mdpi.com/2072-4292/9/2/147/htm>
- [https://www.researchgate.net/publication/258547219\\_A\\_New\\_Approach\\_to\\_Change\\_Vector\\_Analysis\\_Using\\_Distance\\_and\\_Similarity\\_Measures](https://www.researchgate.net/publication/258547219_A_New_Approach_to_Change_Vector_Analysis_Using_Distance_and_Similarity_Measures)

Question 28: can we just perform staking process using R?

Answer 28: Yes, many of these processes can be completed entirely in R. Python and other programming packages have many of these capabilities.



- Here is a good resource for conducting image analysis and for band stacking in R: <https://www.earthdatascience.org/courses/earth-analytics/multispectral-remote-sensing-data/landsat-data-in-r-geotiff/>
- Here is another resource from the NEON group that has a training where you can use R for working with multi-band rasters: <https://www.neonscience.org/dc-multiband-rasters-r>

Question 29: How would a cloud mask be incorporated in this process?

Answer 29: A cloud mask would be used at the beginning of the process because you need to eliminate cloud/cloud shadow before you do any image subtraction. QGIS has a cloud masking plugin you can use. We are also considering doing a short webinar on how to do cloud masking sometime in the future. Here is the Cloud Mask plugin: <https://plugins.qgis.org/plugins/CloudMasking/>

Question 30: What band combination is preferred to detect water and built up area?

Answer 30: A band combination that used the SWIR band would be helpful to detect water and built up area. Also this is a good resource for the usefulness of different band combinations: <https://www.esri.com/arcgis-blog/products/product/imagery/band-combinations-for-landsat-8/>

Question 31: What band combinations are best for differentiating between forest and agriculture/crop?

Answer 31: Using the NIR and SWIR bands would be most helpful for this.

Question 32: Does the working directory (for running R) have to be on the root level of the C:\drive?

Answer 32: I don't think so, the R code should work properly so long as you have the image files and the training file in the same folder (without any subfolders) and you reference the correct folder in the R code.

Question 33: What is an idea to quantify the change?

Answer 33: Yes. We didn't demonstrate this. In QGIS you can look at the tools that include GRASS there is a r.stats tool. You can use this to calculate the statistics. Processing > Grass > Raster > r.stats. This will ask for your classified raster, you can print the cell count, print the area totals, and get the categories.



Question 34: Are there any scripts that automate this procedure? And is there a QGIS plugin that allows you to use R from within QGIS? except the manual training sites, of course.

Answer 34: There are likely many resources available for you. You can take a look at these resources here:

- Here is a good resource for conducting image analysis and for band stacking in R: <https://www.earthdatascience.org/courses/earth-analytics/multispectral-remote-sensing-data/landsat-data-in-r-geotiff/>
- Here is another resource from the NEON group that has a training where you can use R for working with multi-band rasters: <https://www.neonscience.org/dc-multiband-rasters-r>
- Here is another about using R and GIS. There appears to be a RQGIS tool, which we are not familiar with but it is here: <https://jannes-m.github.io/RQGIS/index.html>

Question 35: Any suggestion on the area of the training polygons? I guess that if the area is homogeneous, the bigger the better? What's the recommended minimum size of a training polygon?

Answer 35: For the training polygon you want to make sure you are capturing all the spectral variability within the image. Your training polygon needs to be big enough to capture that variability. Random Forest combines your training polygons into one statistic, it combines each change class. Thus, if the total amount is 50, or 100, you don't have to worry about the size of the individual training site.

Question 36: Are atmospheric corrections mandatory? What method do you recommend for Landsat?

Answer 36: Yes, for many reasons. You don't want change due to atmospheric effects in two images. If you are using Landsat 5 and Landsat 8, the pixel values are different (different bit images). Using the surface reflectance images, this is already done, so it makes that easier. You can also use a vegetation index on each image separately, and then compare the images. But you still might have some differences. We used the SAP (Semi-Automatic Classification Plugin). Check that out.

Question 37: How did you define that 10 training samples/polygons per class was a good number? would having as much training samples as possible be beneficial?



Answer 37: We just did 10 per class because we wanted to do it for demonstration. The number of samples per class is dependent on the size of the class. Check out the Land Cover Classification webinar: <https://arset.gsfc.nasa.gov/land/webinars/18adv-land-classification>

Question 38: Is reflectance calculation necessary to perform NDVI?

Answer 38: Yes. It is recommended you do a reflectance calculation. You don't need to do it to perform NDVI, but reflectance is highly recommended.

Question 39: Is it possible to do a land cover change analysis with a combination of Landsat and sentinel images? This would be useful where the landsat images might have extreme cloud cover (which is common over the rainforest)

Answer 39: You need images with the same spatial resolution. You could resample Sentinel to Landsat to make them the same.

Question 40: Can I use sentinel 2 images for this class of classification, If you answer is YES, Which bands should I use?

Answer 40: Yes. You can check the bands of Sentinel and use the same equivalent bands to Landsat. You can check some of the ESA Sentinel documentation here: <https://earth.esa.int/web/sentinel/user-guides>

Question 41: You were talking about validating classification course. How to get an access into that ?

Answer 41: [Advanced Webinar: Accuracy Assessment of a Land Cover Classification, https://arset.gsfc.nasa.gov/land/webinars/18adv-land-classification](https://arset.gsfc.nasa.gov/land/webinars/18adv-land-classification)

Question 42: How did R create the class.tif image?

Answer 42: You can check the code to see the notes. You can find the original script and instructions for R and Random Forest, which was written by Ned Horning at the American Museum of Natural History, along with a lot of other great resources here: <https://www.amnh.org/our-research/center-for-biodiversity-conservation/capacity-development/biodiversity-informatics/open-source-software-and-scripts>