



Questions & Answers Part 1

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 Sean McCartney (sean.mcartney@nasa.gov) or Erika Podest (erika.podest@jpl.nasa.gov).

Question 1: Is it possible to follow the PolSAR approach for urban classification, or is it only recommended for different vegetation types areas?

Answer 1: Absolutely, yes. There are several studies that show PolSAR approaches to urban areas. One of the most famous freely available dataset for PolSAR is in fact over the city of San Francisco. PolSAR in urban areas is powerful, especially in combination with Interferometric data (PolInSAR approaches). Some of the common urban applications include change detection and classification.

Question 2: There are some similar pixels in the RGB SAR image in your slide. How do you differentiate agricultural fields with that of marsh, even wetlands?

Answer 2: Very good question. As a general rule when using RGB images made with polarimetric features, please keep in mind that these are only 3 images, one for each color. PolSAR data are generally more informative, with more independent observables that can be derived. So please don't be fooled by the RGB, that is not all that PolSAR offers. :) That being said, you are absolutely right that some phenological stages of a specific crop can be indistinguishable to other targets or stages for other crops. Unfortunately things may look very similar to the radar (sometimes this happens even by eye where even with an optical image we cannot say what crop is in a parcel). The core of this training is to show that we can harness the temporal information in the time series of PolSAR to try to discriminate between crops and other vegetation. That is to say, if the two plants do not evolve in the same way in time, we can use this to discriminate them.

Question 3: Can we map crop damage using Sentinel-1?

Answer 3: This is a very valuable application. What we often try to do is to detect temporal anomalies in crop evolution to flag conditions that may be related to crop damage. This requires us building a model for the time series of PolSAR data (i.e., trends for the PolSAR observables). Once we have those we can use some tracking/filtering methodology (e.g., Gaussian Filtering) to spot anomalies and update



models. This is what we did for instance regarding tracking evolution:

<https://www.mdpi.com/2072-4292/12/12/1993>.

Question 4: Is it the same definition of the dual polarimetry decomposition as coded into SNAP?

Answer 4: Yes, the code you see in this tutorial applies the same decomposition that is developed in SNAP.

Question 5: Can you please explain saving the input image stack as binary ENVI format? If possible, please briefly show how?

Answer 5: To do this you need the library spectral. I didn't include it to not leave out attendees that could not install this library. More info on how to read and save data in ENVI format is in:

<https://stackoverflow.com/questions/45178549/reading-img-envi-file-with-spectral-pythhon>.

You see it is a few lines of code to save in ENVI, but you need to make sure the data have been previously read from an ENVI format, and do not change the image size or the code gets more complex to write. If you need to do more complex operations, the suggestion would be to install GDAL and use that.

Question 6: Can you explain the data preparation stage please? Also, the format of data?

Answer 6: I guess you mean the preparation of data using the SNAP graph. This may take a bit of time to show in this answer, and my suggestion would be to look at a SNAP tutorial, where you can see where you need to click in SNAP. In this training we use simple pre-processing. A tutorial on this can be found in the following link:

<https://www.youtube.com/watch?v=30x6XE6U4CU> (until about minute 30) which is the ARSET training: [*SAR Polarimetry with Sentinel-1, RCM, & SAOCOM Imagery for Agriculture, Part 2/4*](#).

Question 7: Which SAR indices will you recommend to map and monitor crops from Sentinel-1 data?

Answer 7: There are several and I believe this will be covered in Part 3 of this webinar series. You can also find some info in the previous ARSET training from last year: [*SAR Polarimetry for Agriculture \(Theory and Practice\), Part 1/4*](#)

<https://www.youtube.com/watch?v=-Mq5A9aXluc> around 1h 20min in the video. My perspective on this is that some indices work better than others depending on the



Crop Mapping using Synthetic Aperture Radar (SAR) and Optical Remote Sensing

April 4 - 11, 2023

specific crop, that is depending on the scattering model that best characterizes the crop type. When you want to select a vegetation index, you will need to make sure it is the one that gives the best performance for your specific target.

Question 8: How to run the SNAP GPT graph using Python? Can you please provide any small example or used in this for generating inputs for this tutorial?

Answer 8: Interesting point, the SNAP Graph Processing Tool (GPT) can really automate nicely the processing stack keeping manual intervention to a minimum. This is a good introduction to this <https://www.youtube.com/watch?v=CD5Fvoy4PWU>. Make sure your graph works properly in SNAP before using the GPT, because it is much harder (I find it very frustrating) to debug graphs from Python.

Question 9: Is it possible to extract these parameters using PolSARPro software using a batch processing tool?

Answer 9: Nice question, absolutely you can use PolSARPro for doing this. PolSARPro also can be run with batch processing using the command line, so you should be able to do it. Recently we are working on translating the PolSARPro routines in Python to sit within MAAP <https://www.earthdata.nasa.gov/esds/maap> when we will have finished (it will take some time :) you will be able to access the routines straight in Python.

Question 10: Can we use Earth Engine Python API to execute this exercise? I mean using the GEE Sentinel-1 catalog and not downloading all images.

Answer 10: Yes and no :) You will NOT be able to run this processing because we need Single Look Complex (SLC) images while GEE only stores Ground Range Detected (GRD) images. We don't have the phase information in the GRD. But YES, if you upload the SLC yourself, then you will be able to use GEE in case the rest of your processing chain is in GEE. If you want to harness the power of collaborative cloud processing in Google, then I would suggest you do it in Google Colab instead, where you can use Python directly... although you will still need to upload the SLC data.

Question 11: In the code I have got an error on np.int() using numpy version '1.24.1'. This can be fixed by replacing np.int with np.int64.

Answer 11: Yes, different versions may require different commands. Well done debugging this.

Question 12: Is there a way to directly download data from the ESA website using APIs?



Crop Mapping using Synthetic Aperture Radar (SAR) and Optical Remote Sensing

April 4 - 11, 2023

Answer 12: Yes indeed, you can download using APIs. I find it even easier to batch download data from the Alaska Satellite Facility (ASF)
<https://asf.alaska.edu/how-to/data-tools/asf-bulk-data-download-options/>. You can find several tutorials online to do this.

Question 13: Can you please specify which versions of the libraries you are using for sklearn, numpy, scipy etc.?

Answer 13: sklearn I use is '0.19.1', numpy is '1.14.2'. Anything above that will work, below that I didn't test it.

Question 14: How do you clip the exact same subset area, because I think there is some shift (sometimes 1 pixel) between different images?

Answer 14: One of the steps you will need to do to stack time series is by doing a co-registration. The orbit data gives a good location for the images, but they are seldom perfect and there may be small shifts, as you mention. Here there is an example using a different satellite

<https://forum.step.esa.int/t/cosmo-skymed-coregistration-with-snap/23450>. You need to make sure this is adequate for your analysis. In PolSAR change detection and accuracy of a bit less than 1 pixel is acceptable, because we do averages. But if you apply any interferometric analysis, then the accuracy has to be much better than 1 pixel. Although SNAP is very good, it is sometimes a bit fiddly to get images well overlapped, but it is necessary and if SNAP does not want to work properly, you may need to use some Python code to refine this.

Question 15: I'm not sure about defining the file paths in Part 3.1 of the PolSAR_TimeSeries document. Should I set it up with the Cubes or the Imagery?

Answer 15: Those are the images that you received, not the cubes. The cubes come in the second file. Be sure to use the double slashes in the code if you are getting errors.

Question 16: Is there a maximum number of vegetation classes?

Answer 16: In principle you can have several and the exact number depends on your case study, the amount of crops you can expect. The more classes you put there, the harder will be the task. In some instances you may need to cluster different crops inside the same class, for instance you may struggle differentiating between different cereals.



Question 17: Can we create Land Use & Land Cover (LULC) maps just by using vegetation indices values using Python?

Answer 17: That is too hard a question to answer here :) I would not suggest you limit your observable to vegetation indexes, PolSAR data are much more informative than vegetation indices. In this tutorial we also use Cloude-Pottier, but there are other observables you could use too. My very vague answer would be, I guess you will manage to get a good LULC map as long as the targets in your scene are separable. If you want to separate city, urban, crops, and forest, then you can do a very good job. If you want to have a map with 20 types of agricultural crops there, then I doubt you can do it easily unless you have quad-pol data (and maybe some optical too when clouds allow it). I hope this is not too generic and please let me know if you have further questions specifying the types of targets you want to separate.

Question 18: How do you estimate the coordinate extent for a Region of Interest (ROI) for each set of training datasets?

Answer 18: In this very simple experiment, you can look at the grid on the side of the image and read the pixel index there. If you want to do this more operationally, then you should extract ROI using shapefiles (installing GDAL) or using some more powerful image tools (using OpenCV). Choosing a small ROI like this will inflate the accuracy, as you have seen at the end of this tutorial. So the code here explains what you have to do in a simple way, but this simple way may get fiddly (hard to see) when you will do it on your dataset... sorry there was not enough time to show this in today's practical.

Question 19: How did you prepare the training samples? Which software did you use for preparing the training samples (e.g., QGIS, SNAP, ArcGIS)?

Answer 19: Please read the answer to question 18. Here I just use Python so that everybody can do it quickly during the practical and we can concentrate on the PolSAR and ML part of it. Sorry there was not enough time to explain how to use other libraries (and some may have struggled installing them).

Question 20: Can we use a different POLSAR dataset for the time series analysis?

Answer 20: Absolutely yes, any PolSAR time series will do. Just make sure the images are from the same orbit and they are well coregistered with each other. If you have quad-pol, then the suggestion is that you see our previous tutorial on how to get more from quad-pol (you will have many more features).



Question 21: This looks more complex than the GEE classification, which was straightforward. What is the effort for? More precision?

Answer 21: I guess the point here is, why use Python instead of GEE, and Python functions seem more difficult to use than Java functions. The extra effort of using Python is about being in control of the data and procedures (in this practical you know what is going on and you set all parameters... you are more in control of it). For instance in GEE you cannot use SLC (doing the PolSAR analysis), or only apply the functions that GEE provides with the data formatting that is supported. Also, you will see the hardest bit is to produce the feature extracted from the temporal analysis. So, to answer your question, the extra complexity pays back in accuracy, since you can include more features, use more methods, with more variants of ML and parameter settings.

Question 22: Why didn't you apply train_test_split from sklearn, and use the long way? Is there any difference? Could we use an extra shapefile as a training area?

Answer 22: I am not sure I got this point, we did use train_test_split. Do you mean why I don't use the black box method and I train the model again? :) Well that is didactic, so you can see what is happening, but you can use any shortcut you can find. As I mentioned here I reduced the "shortcuts" as much as possible, but by any means, you should use them if you know what is going on in the background. :)

Question 23: Assume I have one SAR image of a given location X captured at time Y and prepared training samples from X. Can I use these training samples for image x_1, x_2, \dots, x_N for given location y at a different time?

Answer 23: If I understood well, the answer is "not necessarily". The polarimetric characteristics of a target can change in time, so the model you build in December may be different from the one you build in June. This is the problem of using a single date in a time series to do classification. For some targets it will be fine, but others will show misclassification.

Question 24: Can we use the same code to classify global crop types in a HPC system? What are the challenges and how do you overcome them?

Answer 24: In principle yes, but the first challenge I see is that a data model valid in Scotland may not be good for Brazil, and so on. You may need to retrain based on targets and phenologies you find there. In principle it is feasible, but in practice it will require a lot of work in updating the models. Transfer learning can help with this, but still I believe it would be a monumental task. :)



Question 25: Is it not better to spare some validation (test) ROIs that are not used to develop the model?

Answer 25: Excellent point. If we had time and more processing power (for everybody doing it on laptops at home), I would have shown how it is better to produce more ROIs for the same target. Then use one set of those for training and one set for testing. One issue in what we are doing is that pixels inside the same ROI are correlated (they are too close spatially) and this inflates the accuracy... that is why you saw such ridiculously high accuracies. Sorry, if there was not enough time/resources to do this here, but please do it (and maybe also take advantage of GDAL and OpenCV).

Question 26: With regards to SAR data for mountainous/highly uneven terrain, what challenges can be expected in recreating this exercise for such data with aberrations? Would it be possible to briefly go over some potential mitigating measures for the same as well in this or upcoming sessions?

Answer 26: This is an excellent question, thanks for asking. SAR data are impacted by terrain and one needs to correct this before applying classification. Here we do terrain correction by extracting the gamma naught. That is generally not enough, because it corrects the data radiometrically but not polarimetrically (the phase is unchanged by the correction). If you had quad-pol, then you should correct for orientation angle. If not, another trick that often helps is to include the local incidence angle as a feature in your model. The machine will learn that a specific target behaves a certain way in certain slope conditions and use this info to correct for terrain-induced errors. You get the local incidence angle from SNAP. It's not an easy task to deal with mountainous terrain, but lots of progress has been made recently and we are getting there. :)

Question 27: Why are we creating 1d features?

Answer 27: That is the way that the ML methods like the data. If you look at the tutorial you will see we need to create these X and y arrays so that the machine can build the model to link both. Once you train using these 1d vectors, you also need to test converting your full image in 1d-fashion. I hope this answers the question. Please ask for more clarification if it does not.

Question 28: Does this method apply in Google Earth Engine?

Answer 28: Please see answer to question 10.



Question 29: What is the difference between the [Jupyter PolSAR] approach and the conventional method like downloading images and using software like GIS?

Answer 29: Python allows you to personalize your algorithm to make it best for the data, that is, it will provide better accuracy. In other software (I guess you mean QGIS) you can apply off-the-shelf toolboxes, but you will find it hard to create the layers to feed the ML black box when you deal with multitemporal SAR data, unless of course you run python in QGIS. For instance you will need to apply this manually layer by layer. Multispectral single-date classification in QGIS is easy because the toolbox already knows what to do (it does it in the background) but as soon as you want to do something different, you will struggle and Python allows automation, which you may want to use if you have a large project.

Question 30: Is integration of images from different sources (like, for example, satellite and Landsat) possible in the case of long term analysis?

Answer 30: It is possible and also a very good idea. You will see more related to fusing SAR and optical later in this training series. There are different approaches to do this, but in principle you can use the different sensors to enlarge your feature vector. The complexity is in dealing with missing optical images (when clouds did not allow). But there are several solutions that try to solve this. Fusion is very powerful and I expect there will be more and more in the future.

Question 31: Is identification of crop species possible using the current method after classification?

Answer 31: It really depends on the crop type you are interested in. You can read more in my previous answer to question 17.

Question 32: Do we have another webinar where actual ground-observed data is used for training and testing? How do you maintain the integrity of a pixel on a particular date in a particular class?

Answer 32: Grounds observed data are often not easy to produce. In this training we used the LULC produced for the Scottish government that I showed at the start. But this does not include the type of crop. This is why we only focus on land types that can be assessed using that map. In case you have info on the crop type, you can try to separate them. This training was aimed at showing how to apply ML with PolSAR data (bridging the gap between the two), and not to show results from a specific study, but indeed you will need to know which crop is in the field to train properly your model (if



Crop Mapping using Synthetic Aperture Radar (SAR) and Optical Remote Sensing

April 4 - 11, 2023

you are interested in separating different crops). The second question I think is related to eventual changes of land type. Those pixels should not be used for training and they can indeed create misclassification. If you are interested in change detection, the time series should not be too long and you may decide to use a tracking algorithm instead. Please see answer 3 for more details.

Question 33: The K means, is that what GIS applications use?

Answer 33: Yes, you can find K means in GIS software. The complexity there is that they normally don't allow you to personalize the feature vector easily, so if you want to do something like PolSAR and multitemporal easily, you will need to construct layers manually. Please read more in my answer to question 29.

Question 34: What about using median instead of mean for synthesizing multitemporal information?

Answer 34: Good point, indeed you can use that if you prefer. Also people often add the min and max in the time series. The bottom line is that you will need to test this on your dataset and see what operation/observable best encapsulates the temporal information.

Question 35: Accuracy was 0.9949 with non cross-validation, would you comment regarding overfitting? Meaning, is accuracy the best metric to evaluate the model performance?

Answer 35: As I mentioned, these values are ridiculously high because we only use very small portions as ROIs. See answer to question 25 for instance. So to answer your question, I would say, yes we have overfitting because we don't select several ROIs. Once you are familiar with this code, if you have the chance to install GDAL or OpenCV please do things more properly.

Question 36: Is there a plan to perform this task on cloud computing?

Answer 36: You can absolutely do this using Google Colab. It is very simple to implement. Please read more on my answer to question 10.

Question 37: I would like to know, are the ROIs created from one image used to train and test all the images? How do we handle that for the case of many years?

Answer 37: The ROI are located in the same spot in time. This is essential because you need to capture the time evolution of the same target (i.e., same place) over time. I hope this answers the question, otherwise please ask again.



Question 38: What about pixel-based 1D deep learning (not 2D trained by image), is it better than random forest?

Answer 38: Good point, indeed you can apply that too, I forgot to mention. There are some that classify the trends and indeed they are quite powerful. My suggestion and experience with these types of approaches is to reduce the field to a single trend (you average all the pixels in the field together), otherwise the noisiness of the data will overwhelm the DL algorithms unless you have a really large amount of training data. When it comes to these types of techniques (trend detection), I am more of a fan of Gaussian filtering, which I think can deal better with the noise of SAR data, but indeed there are studies that use DL that will beg to differ. :) The bottom line is that you need to see what works best in your scenario.

Question 39: Is there any use of DInSAR on crop monitoring?

Answer 39: Fantastic question, DInSAR as such is not much used, but there is a lot on PolInSAR for crop biophysical estimation. This is an example of what we tried <https://ieeexplore.ieee.org/document/9508494>. DInSAR has been shown useful to soil moisture (although there is still lots of research to be carried out).

Question 40: Can we correctly classify those crops which have a similar phenology cycle using PolSAR data and the approach shown in the session?

Answer 40: Yes, using time series you can focus on the evolution of PolSAR features in time and use this info to separate different crops. You may need to be careful in using a technique that is able to deal with different sowing times, but this may also be info for you if you want to discriminate between practices.

Question 41: Can Sentinel-1 detect infested crop fields before optical sensors can?

Answer 41: That is a million-dollar question :) One thing that SAR can do better than optical is: a) get images regardless of cloud cover, therefore having more chances to see it sooner, and b) look deeper under the canopy cover, which may be an advantage when the stress starts happening from the soil and not the top of the canopy. However, SAR does not see the change in color of the plant. What we see is mostly the change in water content which produces a change in dielectric constant and changes in shape/orientations. My best bet for a faster system will be to build it using SAR and optical together; when you have an available optical image that will help, and when you don't SAR will do the job.



Question 42: Can we estimate the low, medium, and high crop yield pixel wise areas using PolSAR data and, if yes, how do we estimate it?

Answer 42: Great question, yield indeed is the information that most stakeholders want to access. Estimating potential yield is possible although depending on the type of crops you may have different accuracies. The way it is often done is by using some ML methods trained with several types of data, PolSAR, but also time of the year, temperature/rain and other relevant information for the growth. This is an example of what we tried but there are many more examples out there <https://www.mdpi.com/2072-4292/12/12/1993>. Also, don't forget that the real yield is sometimes different from the potential one, because it depends on the harvest practices.

Question 43: Is it possible to calculate NDVI from Sentinel-1 data? Can we classify vegetable fields using Sentinel-1 data?

Answer 43: More information on this will be presented during Part 2 of this training series.

Question 44: Could this methodology be customized for groundwater storage classification from GLDAS data or Modified Normalized Difference Water Index?

Answer 44: I don't have experience with groundwater estimation. I did work on surface soil moisture (SSM) and if this is what you are interested in then, yes, you can do this using SAR. Copernicus also has some low resolution products available: <https://land.copernicus.eu/global/products/ssm>. For higher resolution, I would suggest using L-band to improve accuracy. Groundwater is not visible as such because SAR does not penetrate enough into the soil. But if in your test site this modifies SSM or vegetation cover, then you have a good chance to use SAR to monitor this.

Question 45: Is it possible to perform classification at a farm size, for example a size of less than 1 ha for pasture field? How can we use Sentinel-1 data for regions where field sizes are less than 1 hectare? Is it possible to use Sentinel-1 and ML models to identify maize in smallholder farms (0.5 to 2 ha) with multiple crops?

Answer 45: Currently, working with smallholder fields is very challenging. This is especially because often the smallhold contains several crops in the same acre. The future is bright, because we will have higher resolution images, but right now, with Sentinel-1 you will need to add some extra information into the model. Something



Crop Mapping using Synthetic Aperture Radar (SAR) and Optical Remote Sensing

April 4 - 11, 2023

related to practices, boundaries, temperature, rain, etc. Those models will be very specific to the place you apply them (you will need to train again in a different area) but they can help give some overall information that is very important to managing the region. This is very much a research topic and I am not sure there are operational solutions using Sentinel-1 so far, but we are not too far off and new satellites will help a lot.

Question 46: How do you usually define the minimum size of training datasets?

Answer 46: This is hard to do, and there are no one-size-fits-all solutions. For random forest, thousands of samples may be enough to train, but it depends on the separability of classes. There are procedures to check this and basically you try different regions and see when overfitting seems to disappear (the accuracy goes down). But it is hard to say when you can stop and especially if you are adding different targets in the ROI (so it is not misclassification, but mis-selection of regions). A hard question to give a straight answer, sorry.

Question 47: In PolSARPro, when we apply any decomposition we get the output in 8 bit output. But it should be in dB, my question is how do you rescale it into a relevant range for that particular model?

Answer 47: I would like to start answering this saying that you do not need data in dB to run the code I showed. The covariance matrices are not in logarithm scale (cross correlation do not have a logarithm). So you don't need to have them in dB. Some people use dB because this helps with the scaling, but the random forest is quite robust regarding this. The output coming from PolSARPro are still in ENVI, but indeed the data type can change. This is reported in the header and you will need to check this before reading. But in principle, as long as you fix eventual data type changes, PolSARPro binary outputs can be read in Python. I hope this answers the question, otherwise please ask.

Question 48: BOXCAR that performs clear averaging, or R.Lee that preserves edges: which one is better to implement when working with crops?

Answer 48: Fantastic question, I love it. :) Unfortunately there is no straight answer, this is more my opinion. I prefer to use BOXCAR when I know that the area I am using for training is rather homogeneous. The BOXCAR will smooth, "reduce" outliers and do simple math on the stats. R.Lee will help dealing with edges, but it does an MMSE, which gives some peculiar characteristics to the filtered image. In practice, you can see Lee seems to provide slightly higher accuracies in some cases. I guess this **may also**



happen due to edge effects when doing the validation. I normally use BOXCAR for this, but I understand if somebody prefers Lee.

Question 49: I am seeking the proper tools and techniques to monitor lake shore changes. Do you think that your approach would fit?

Answer 49: Interesting application. My suggestion would be to apply some change detection methodology that is able to use time series. For instance, we used Cumulative Sums for this (applied to forest in this paper

<https://www.mdpi.com/2072-4292/12/18/3061>), but there are many other methods as the following

<https://www.tandfonline.com/doi/abs/10.1080/01431161.2017.1376128?journalCode=tres20>. In principle, you can take any change detector and apply image after image, then evaluate the detection in time. It would be interesting to see your results.

Question 50: Are there any tools to process SAR imagery to date besides SNAP? Are there Python packages you'd recommend?

Answer 50: If you want to skip the SNAP step all together (and don't substitute it with some other software you pay for, like Gamma Remote Sensing, ENVI, SARscape, etc.), and do it all in Python, then my best bet would be to use GDAL. The most complex bit would be to read the format of the data and "import" the images as binary files. This may be quite challenging if GDAL does not have a function to cover that specific sensor and mode. Fortunately, SNAP is quite updated to current sensors and it is free to use.

Question 51: Is it necessary to go through all the radiometric and geometric error corrections for the SAR data before considering the ML classification step?

Answer 51: My suggestion is generally yes, but if you have no topography (very flat area), you only use one single orbit, and the area of interest is quite small, then you may get away without those steps. But that model will not be exportable to another location. Unfortunately you will have a fluctuation of backscattering due to incidence angle and topography. One step trying to assess this may include adding information about the local incidence angle as a feature and let the machine work out the pattern of how the backscattering changes with this. But I would personally not trust this. I would prefer doing the calibration, although this takes some time.



Crop Mapping using Synthetic Aperture Radar (SAR) and Optical Remote Sensing

April 4 - 11, 2023

Question 52: Does it matter the type of SAR value stored in the pixel? For example, would Sigma Nought, Gamma Nought, etc. produce the same results when used as covariates?

Answer 52: Great question. If you have no topography, then sigma naught is enough. But if you have slope you should consider gamma naught. Please read more on this in my answers to questions 26 and 51.