# Sentinel 1 and Forestry Change Detection Techniques Progress Report

## July 2018

In July I was able to attend a 3 day introductory training course on Synthetic Aperture Radar (SAR) using examples from Sentinel 1. The course was not just for those working in forestry, but other applications, such as flood detection and even applications outside of the UK such as the impact of earthquakes using interferometry techniques. The course was very useful as background and taught many useful techniques, but also made me realise the challenges of using SAR in forestry. The course instructor had prepared a forestry demo as beforehand we were able to submit requests for what we want to learn, but using an individual Sentinel 1 image to detect a recent clearfelling was not really possible in that example. As the instructor advised at the time and as I out found myself later when I started using Google Earth Engine a temporal composite of multiple images may be necessary to help reduce the noise when trying to detect forestry changes.

The instructor advised for further background information to study the MOOC called ‘Echoes in Space’. This was produced in part by Iain Woodhouse of Edinburgh University and from this I discovered that he had written a book ‘Introduction to microwave remote sensing’, which I then downloaded and tried to read / understand as background information.

## August 2018

In August I started to use Google Earth Engine (GEE) for the first time. I already knew that Sentinel 1 and Sentinel 2 data were available on GEE shortly after being made available by the European Space Agency (ESA). However, I had not yet tried to use GEE to access the Sentinel data to see how feasible it would be to use it for my project. GEE has both a Javascript and a Python API. I had initially hoped to only use the Python API as this would hopefully allow the GEE interaction to work closely to other data science modules and functionality in Python. Initially I had to set up the GEE Python API and checked I could connect to the cloud based service. Despite Google’s warnings about Python versions and the difficulties of running GEE this way, I found the ‘minimal installation’ worked fine, so I did not have to use a Docker Container [1]. I managed to get it to work both on my Ubuntu Python 2.7.15 installation and on my Windows 10 using Python 3.5. Once I got it to connect, I next wanted to get started with accessing some imagery, but I discovered that there was little in the way of instructions available for Python on the website. I eventually managed to track down a few basic examples on Github, but little to get going. As the GEE website mostly shows examples and provides a few tutorials for the Javascript code editor I decided I might be best to use this initially and I had seen indications that beyond the basic differences between the two languages, using the GEE methods and functions was essentially the same syntax whether in JS or Python. Having run through some of the basic tutorials [3] to get going with the JS code editor, I then looked at some example Sentinel 1 visualisations I could find online. Without too much difficulty I was able to produce a monthly composite, mosaicking multiple images spatially for GB. I had learned in the course in July that the most common way to visualise Sentinel 1 is to use VV polarisation in the red band, VH in the green and VV/VH in the blue band. An alternative is to use VH in the red band, VV in the green band and VV – VH in the blue band. In both cases this produces a ‘sort of’ true colour visualisation of the Sentinel 1 data. Below shows the later option for September 2018 across GB. This makes the image easier on the eye than just visualising one polarisation (either VV or VH) in black and white, but it seems there is also some derived data of value for modelling by adding the proportion or difference between the two bands.

As well as exploring options with Sentinel 1, the GEE Code Editor was used to investigate Sentinel 2 optical imagery. One thing I had always heard is that the problem with Sentinel 2 and optical imagery in general is that cloud inhibits its use for near real time monitoring of change on the ground. In the Sentinel 2 data they serve, Google have added a band called QA60, which is a bitmap cloud mask. According to the GEE Data Catalog, this QA band was created as outlined by ESA [2]. In addition, metadata is added to each image of the percentage of cloud. This information can be used in a few different ways:

-If working with a large image stack covering weeks or months, the images for a given location can be sorted to show the least cloudy first of all

-A cloud mask can be created, so any pixels covered by the QA cloud band are shown as no data. This can be used to build a temporal cloud free mosaic over a large area.

However, the Google QA cloud mask was not perfect and it appeared to be missing many of the less dense clouds still obscuring what was happening on the ground, as shown below. In addition it does nothing to address cloud shadow (darker areas in the shadow of the cloud, even if not blocked by the cloud directly).

An alternative option was found on the GEE message board created by the effort of a few different users. This technique, using picked up more of the lighter cloud, but it was too effective and it would mask out white building rooves, mistaking them for cloud. It might be OK for Forestry use, however.

To put some metrics to the cloud cover problem, I produced a GEE script, that would use the QA60 band to analyse whether a point on the ground was inside the cloud mask. For each available image over a range of dates (for example a year), the script would determine if the point on the ground was obscured by cloud or not. The output was a csv table with each image capture date and whether or not Sentinel 2 data was obscured by cloud. This was a conservative estimate because, as described above, the QA60 band does not pick up all cloud. The results are shown below for one year for a location in a forest close to Peebles in the Scottish Borders.

## September 2018

For two weeks of September I was away on holiday. On returning I continued to look at ways to display and analyse Sentinel 1 data. A screenshot is shown below of a particular visualisation of Sentinel 1 data using a 3 month composite for summer 2018 for part of Thetford Forest in East Anglia. This is done using an example on the GEE website, separating images captured in ascending and descending orbit into different bands. It can be seen within the forest boundary, how areas of trees and non-trees are distinguishable.

Time was also spent in preparing training data to later use for supervised classifications or assessing results of unsupervised techniques. The training data available to be used for this purpose are Open Data published by the Forestry Commission. There are two main data sources to be used:

### The National Forest Inventory Map (NFI Map)

The NFI Map is a spatial dataset of all woodland in Great Britain, whether on public or privately owned land. It should be noted that by international definition a woodland must be over 0.5 ha in size and 20 metres or more in width and this is the rule by which woodland polygons are added to or removed from the NFI Map. The NFI Map was originally captured from aerial photography in 2009 and the Forestry Commission now maintain it using information from planting grants and felling licence data, using NDVI techniques to detect clear-felling from remote sensed imagery and by reviewing newer aerial photography when it comes available. It is often more than a year before change on the ground is updated on the map and it is hoped that the Sentinel 1 techniques reviewed in this project may assist in the process of detecting change and applying it to the map. Part of the time taken in publishing updates is to process change into a clean and tidy spatial product, for example the polygons are snapped to Ordnance Survey mapping boundaries where possible. The most recent changes published in the NFI Map likely occurred on the ground in 2015 – 2016, so the main use of the NFI Map for this project is to produce a boundary of interest. Note that forest clearfelling remains on the woodland map until it is later discerned that either replanting has occurred (most commonly) or permanent deforestation has occurred, at which time the area is removed from the NFI Map.

In preparing the data to use in GEE, it was possible to upload the NFI Map to GEE. However, to make it easier to work with, internal boundaries were dissolved (as its main use was to be an external boundary of woodland) and it’s geometry was simplified. This work was done using QGIS software.

The other source of training data is the Forest Commission’s Sub Compartment Database. A simplified form of this dataset is available for the public to download and this Open Data product was adequate for this project. The Sub-Compartment data is the Forest Commission’s inventory of land use and tree species on its own estate. As approximately one third of woodland in Great Britain is Forestry Commission land, this provides a large source of training data. Sub-compartments are the smallest spatial division of what is occurring on the ground. Individual stands of trees of a particular age, species, planting regime, storey structure will be recorded as a separate sub-compartment polygon. Against each polygon, up to nine non-spatial tabular records, known as components, are recorded for each tree species, planting year and land use occurring in the sub-compartment. The relative proportion of each component within a sub-compartment is recorded as a percentage of area on the ground. For example, a sub-compartment could consist of 70% Sitka spruce and 30% birch. In producing the Open Data product, the Forestry Commission flatten out the component data so the component records are stored as attributes of the sub-compartment polygons themselves. This makes the data easier to work with for the general public, but it means only the top three component records (by proportion of area) are recorded against each Sub-compartment. For most of the purposes of this project a pure sub-compartment of only one land use is adequate. The advantage of the Sub-compartment data for this project is that it is ground-truthed data, maintained by the staff working locally to each forest. The most recent sub-compartment data was acquired and processed into the Open Data product, meaning that fellings occurring in summer 2018 were included.

Once the Sub-compartment data had been acquired, it was determined that it should be processed into a product suitable for using in GEE for analysing patterns in the Sentinel 1 SAR data and ultimately for training and testing models. In order to prepare this the following steps were taken:

1. The sub-compartments were clipped to the NFI Map woodland boundary. This should remove any non-woodland land owned by the Forestry Commission from the dataset and means all analyses are occurring inside this defined boundary.
2. The R Spatial package ‘sf’ was used to prepare polygons to use for data discovery.
3. The sub-compartment shapefile output from step 1 was read into R.
4. The attributes of the polygons were used to give each polygon one of the following labels:

Broadleaved

Conifer

Mixed (mixture of conifer and broadleaves)

Young

Felled

Open (open space internal to the woodland, but mapped out as sub-compartments)

Other (where none of the above categories applied).

1 <https://developers.google.com/earth-engine/python_install_manual>

2 Cloud Mask <https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-2-msi/level-1c/cloud-masks>