DPLUS065 Coastal Habitat Mapping: Developing broad scale habitat maps for South Georgia using a Random Forest classification in Google Earth Engine

Background

The Coastal Habitat Mapping project has utilised recently available, medium resolution, satellite imagery alongside other spatial data and local expert knowledge to develop the first island-wide, broad scale (10m resolution), satellite-derived habitat maps for South Georgia. This document provides a short narrative of the input layers used in the Random Forest model, workflow methodology, outputs and future challenges.

Model inputs

- Sentinel-1 (band 1)
- Sentinel-2[1] (all 10m bands including visible, various NIR and SWIR)
- Slope, aspect and elevation: derived from Shuttle Radar Topography Mission (SRTM)
- Landsat-8 (Band 1 – 30m coastal aerosol band)
- Normalized Difference Vegetation Index (NDVI)
- Normalized Difference Water Index (NDWI)
- Enhanced Vegetation Index (EVI)
- Geary's C (texture analysis)
- Various ground validation datasets (2017 South Georgia DPLUS065 Coastal Habitat Mapping expedition, 2005/06/07 Giant Petrel survey, manual point digitization (kelp, barren ground, ice and cloud), 2010 SMSG expedition))

Workflow summary

Sentinel imagery processed by JNCC was brought into Google Earth Engine, then clipped to the area of interest. Terrain correction, to help address the conflicting effects of shadows from relief and bright surfaces, was applied. Cloud masking would normally be applied at this stage, but was not required for the Feb 2018 Sentinel-2 imagery. Various indexes such as NDVI, NDWI and EVI were calculated. EVI is similar to NDVI but is less sensitive to error due to dry vs bare ground. Geary's C texture analysis was used to assist with identification of subtidal rock features on Landsat 8's coastal aerosol band (band 1). For each of the ground validation points, information from each input dataset (satellite band/index layer) is extracted, and saved to a new table, called "training data". The Random Forest classifier is applied to this table (100 trees, 80% of ground validation as model training data, 20% as validation). The classifier operates by looking at each pixel in the area being mapped, looking at what the combined imagery "looks like" for that site, and finding the best match to the various land cover types listed in the "training data" table. This match is then tested using the validation data. A confusion matrix is exported (this is a report that compares the land cover type listed by the validation points to the output classification map as a verification of the classifier's success) as a text file, along with a GeoTIFF of the classified map.

Planned improvements

- Clouds/bare grounds: some artefacts remain around areas of high cirrus clouds or at the edge of cumulus clouds where the classifier assigns a value of "bare" rather than "cloud"
- Further work delineating offshore substrate – determine if greater "level of detail" is possible
- Seasonal opacity of water to determine subtidal areas more likely impacted by glacial output
- Non-grass vegetation classes in particular would greatly benefit from additional ground validation e.g. Acaena or mire and bog
- Use of expert opinion to provide additional ground validation in addition to future field work?
- Object-based classification (as opposed to the current pixel-based technique) to improve model performance.

[1] Sentinel data acquired on February 22nd 2018.