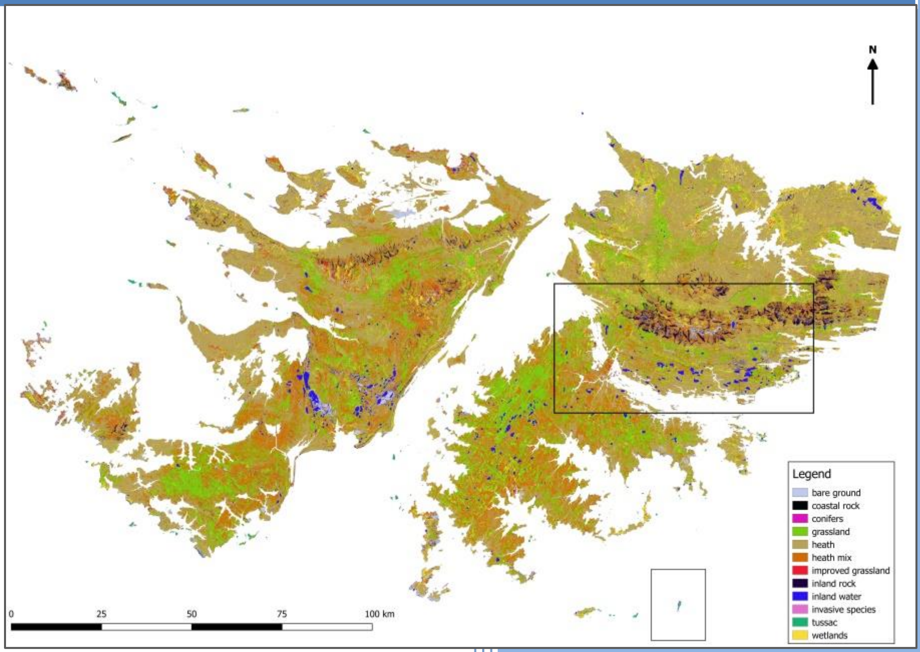
2018

# Falkland Islands broad scale habitat map from Earth Observation techniques





ILaria Marengo


27th March 2018

## Background

The UK Government, through the FCO managed Conflict, Stability and Security Fund, is supporting a suite of natural capital projects across the UK's South Atlantic and Caribbean Overseas Territories. This work is designed to improve economic stability in the Territories through enhanced environmental resilience as part of a programme led by the UK's Department for Environment and Rural Affairs (Defra). The natural capital project began in September 2016 and will be completed by March 2019 with the Joint Nature Conservation Committee (JNCC) as the Implementing Body.

In the South Atlantic, the Natural Capital Project work is being undertaken by the South Atlantic Environmental Research Institute (SAERI) under a Memorandum of Agreement with the JNCC. The project will assist the UK's Overseas Territories in the South Atlantic to assess and map natural capital, value priority assets and deploy decisions support tools to secure long-term economic benefits from the sustainable management of the territories' natural assets. This support will be provided through the development and collation of spatial (mapped) evidence, and a Territory-to-Territory partnership for technical exchange and capacity building within the UK's Overseas Territories in the region. The outcome will be a framework for the South Atlantic UK Overseas Territories to assess the value of the environmental goods and services available and integrate this information into marine and terrestrial spatial planning, economic planning and environmental protection.

SAERI will be providing an evidence base for the South Atlantic Overseas Territories to make decisions on the areas identified as a priority in this consultation. The project focuses on four key deliverables:

1. Spatial data on the distribution of selected natural capital assets, both marine and terrestrial, derived from satellite imagery and other existing resources, as relevant to each Territory;
2. Valuation of priority natural capital assets (value mapping integrated into national GIS) and the assessment of economic and societal benefits arising from them;
3. Application of analytical tools that will support decision making in the context of environmental management and economic development (e.g. scenarios);
4. Methods for monitoring changes to priority natural capital over time using appropriate attributes (e.g. indicators).

This report sits under deliverable 1 and outlines the development of a new habitat map for the Falkland Islands.

## Introduction

The Natural Capital Assessment (NCA) project offered the opportunity to carry out a broad scale habitat mapping exercise using the latest open source Earth Observation (EO) imagery (SENTINEL2), R open source statistical language for the classification model (Random Forest) and a habitat classification generated by an expert terrestrial ecologist, Dr Rebecca

Upson[1], who worked in the Falklands for four years. The habitat classification devised by Upson (2012) is currently used in the Falkland Island Government State of the Environment report[2] (2008) and on Falklands Conservation's website[3] (2018) to describe the terrestrial habitats of the Falkland Islands.

Since 2002, Earth Observation (EO) techniques have improved and a new 'family' of open access satellites has been made available to the public by the European Space Agency (ESA). The SENTINEL satellites have been launched with five specific missions. The objective of SENTINEL-2 is land monitoring, and the mission is composed of two polar-orbiting satellites providing high-resolution optical imagery. Vegetation, soil and coastal areas are among the monitoring objectives. The first SENTINEL-2 satellite was launched in June 2015[4] and, with its much higher resolution (10 metres against 30 metres of Landsat), provided the opportunity to develop a new habitat map. The resolution of the image is important for any classification as it implies the size of the objects on the Earth's surface that can be 'detected and seen' by the satellite. Hence SENTINEL-2 allows distinguishing objects with a size up to 10 metres. The disadvantage of all SENTINEL satellites is that, due to their recent launches, they cannot provide a long time series for historic comparisons. This document sets out the approach taken to obtain a broad scale habitat map from SENTINEL-2 data, using the habitat classification by Upson (2012) as a starting point.

**Methodology**

To develop a broad scale habitat map it was necessary to identify suitable satellite imageries and the appropriate level of aggregation of habitat classes that could be identified through the analysis of the remotely sensed imageries. The Joint Nature Conservation Committee (JNCC), helped with the first task. Various SENTINEL-2 imageries with less than 20% cloud coverage had been selected from December 2015 to February 2017, but only one image proved optimal for the analyses as it showed no cloud cover and included 95% of the land. This imagery was taken on the 16th of December 2016.

After acquisition of the imagery, a series of pre-processing steps were conducted to prepare it for the habitat classification analyses. Atmospheric correction, using an interim Dark Object Subtraction (DOS) method to estimate atmospheric conditions, and resampling of the SENTINEL-2 20 m image bands to 10 m spatial resolution were the main pre-processing steps. An interim atmospheric correction method was used while stable software to enable a

---

[1] Upson R., 2012, Important Plant Areas of the Falkland Islands. Unpublished Report, Falklands Conservation. 80 pp.

[2] http://www.fig.gov.fk/epd/index.php/environment/19-environment/60-state-of-the-environment-report-2008

[3] http://www.falklandsconservation.com/wildlife/plants/37-wildlife/about-falklands-wildlife/97-habitat-types-of-the-falkland-islands

[4] https://sentinel.esa.int/web/sentinel/missions

higher standard of pre-processed surface reflectance products were being developed for SENTINEL-2[5].

In Earth Observation techniques, there are two approaches to classification. The unsupervised is when the image analyst doesn't input pre-defined classification rules and groups but chose a classification algorithm and a number of classes provided by the image processing software. The supervised is when the image analyst attributes to sample pixels or segments specific classification rules or groups and train the software to use the information provided as references for the classification of the entire image.

For the NCA project, the supervised classification was the preferred approach due to the presence of officially recognized and used habitat classes. In fact, in parallel to preparation of the satellite imagery, the habitat classification by Upson (2012) was revisited (table 1) with the aim of defining which classes of habitat and levels of aggregation (from 1-broad to 4-fine) could be mapped using SENTINEL-2. The first level of habitat classes listed in table 1 was considered 'mappable' through EO techniques because they are very distinct from each other, well distributed in the Falklands and wide enough to cover large areas across the Islands. Within the second level, there were habitats that potentially could be distinguished from each other e.g. grassland vs improved grassland vs marshy grassland and other habitats that, due to their low occurrence, would be more difficult to detect (e.g. flushes, marginal, fens and swamps, scrub). Levels three and four habitat classes provided too much detail that could not be detected by SENTINEL-2 imagery analyses, and a field survey would have been more appropriate. Additionally, because the project's aim was to achieve a broad scale habitat map, levels three and four were not considered further.

Habitat levels one and two were compared to understand if further aggregation or movement through levels could be carried out to define the final habitat classes for the mapping exercise. Table one shows the comparison between Level one and two and the decision making process that occurred, while table two provides a summary of all classes used for the analysis of SENTINEL-2 imagery. Photos of each habitat have been added to the table as visual aid to better understand what these habitats looks like.

**Table 1: Comparison between Level one and two habitat classes**

| **1st level** | **2nd level** |
|---|---|
| Tussac | |
| Grassland | 1. Acid grassland<br>2. Neutral (including 'greens') grassland<br>3. Improved/ reseeded (THIS CLASS WAS MOVED TO LEVEL 1) |
| Dwarf shrub heath (BECAME HEATH) | 1. Dry dwarf shrub heath<br>2. Wet dwarf shrub heath<br>3. Dry dwarf shrub heath/ acidic grass mosaic |

[5] Gwawr Jones is the Earth Observation specialist at the Joint Nature Conservation Committee. She carried out the pre-processing operations that prepared the satellite imagery for the habitat classification.

| | |
|---|---|
| | 4. Wet dwarf shrub heath/ acidic grass mosaic |
| Fern beds (INCLUDED IN DWARF SHRUB HEATH) | 1. *Blechnum magellanicum*<br>2. *Blechnum penna-marina*<br>3. *Blechnum cordatum*<br>4. *Gleichenia cryptocarpa* |
| Cushion heath (INCLUDED IN DWARF SHRUB HEATH) | 1. Cushion heath - coastal<br>2. Cushion heath - inland |
| Montane/ Feldmark (THE CLASS WAS NOT CONSIDERED) | 1. Cushion plant dominated<br>2. Moss and lichen dominated |
| Heath mix (NEW CLASS INTRODUCED TO IDENTIFY MOSAIC) | Mosaic of wet/dry dwarf shrubs heath and grassland. Cushion plants and Christmas bush will occur as well |
| Bog and Flush (THIS CLASS WAS MERGED WITH FEN, MARSH AND SWAMP AND IT BECAME WETLANDS) | 1. Bog<br>2. Flush |
| Fen, marsh and swamp/ marginal communities | 1. Fen and swamp<br>2. Marginals<br>3. Marsh/marshy grassland |
| Open Water | 4. Standing water<br>5. Running water |
| Coastland (DIFFICULT TO MAP AND THE CLASS WAS NOT BROUGHT FORWARD) | 1. Littoral sediment<br>2. Saltmarsh<br>3. Rock/ boulders<br>4. Strandline vegetation<br>5. Sand dunes<br>6. Maritime cliff<br>7. Coastal cushion heath<br>8. Coastal (saline) grassland<br>9. Coastal dwarf shrub heath<br>10. Coastal feldmark |
| Inland rock | 1. Natural rock exposure<br>2. Artificial rock exposures |
| Scrub (DIFFICULT TO MAP AND THE CLASS WAS NOT BROUGHT FORWARD) | 1. Dense<br>2. Scattered |
| Woodland | 1. Coniferous<br>2. Broadleaved<br>3. Mixed |
| Other | 1. Arable and horticulture<br>2. Built-up areas and garden<br>3. Bare ground (THIS CLASS WAS MOVED TO LEVEL 1)<br>4. Introduced vegetation (THIS CLASS WAS MOVED TO LEVEL 1) |

**Table 2 Summary of all classes used for analysis of SENTINEL-2 imagery**

**1st habitat level for EO analyses and matching pictures**

Tussac

Grass
land

Modified
grass
land

Heath

Heath mix

Wetland s

Open water - fresh

Inland rock

Conifers

Bare ground

Invasive species

Open water-marine

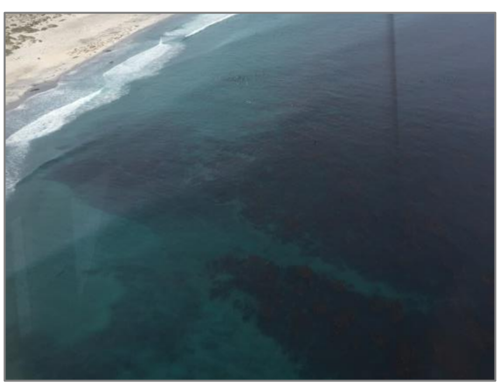


Table three shows how habitat classes used for generating the habitat map through EO techniques relate to the classes cited in the 2008 State of the Environment document.

**Table 3 Comparison of EO analysis and State of the Environment habitat classes**

| Habitat Level 1 EO analysis | State of Environment report 2008 |
|---|---|
| Bare ground | Littoral sediments<br>Sand dunes<br>Maritime rock, shingle, cliff and slope<br>Built up areas and gardens |
| Conifers | Coniferous forest |
| Grassland | Greens and neutral grassland<br>Acid grassland |
| Heath | Dwarf shrub heath<br>Scrub<br>Fern beds |
| Heath mix | |
| Modified grassland | Improved grassland<br>Arable and horticulture |
| Inland rock | Inland rock |
| Inland water | Standing open water<br>Rivers and streams |
| Invasive species | |
| Tussac | Tussac |
| Wetlands | Bogs<br>Fen, marsh and swamp |
| | Montane habitats |

The broad classification of some of the habitats was challenging, even with the use of SENTINEL2 imagery. Tables 1 and 3 highlight that coastal and montane habitats could not be identified and mapped with the level of detail described by Upson (and the state of environment report). The reasons are multiple: rocky shores appear to be very difficult to distinguish from bare ground. Montane habitat is likely to be identified as inland rock and sometimes heath mix, especially when heath mix occurs on higher ground. The difficulty with montane habitat is that it occurs frequently amongst rocks and it is therefore hard to identify a precise signature and pixel colour. Sandy shores could be identified more easily and an attempt to classify this habitat could be made in further iterations of the map.

Mosaics are mentioned in Upson's habitat classification scheme, but there are no classes which describe these various mosaics. Based on the author's local knowledge of the Islands, which includes four and a half years of extensive walking, mosaics of grassland (particularly white grass) and heath are common and widespread across the Falklands. By looking at the false colour infrared imagery, which highlights vegetation (in red), this mosaic habitat class was recognisable and therefore it was decided to introduce it.

The EO technique used for the broad scale habitat map was a supervised pixel-based approach. This approach adds points to the satellite imagery map on pixels that match with specific habitat classes. The technique requires:

1. To spread the points for each class across the entire study area, so that there are no clusters but an evenly distributed series of points.
2. To keep the number of points per habitat proportionate, so that each habitat class is represented equally.

Figure 1 shows the overall spread of points per habitat class. The total number of sample points used to derive the habitat map was 9,674. Since the project did not include time for fieldwork, the sampling points were collected opportunistically and did not follow a specific scientific methodology. Hence, some of the points were added on the basis of local knowledge of the region or after a leisure walk (or trekking) occurred. GPS data were associated to the points, and photos have been taken to recall the habitat at broad scale rather than in a systematic way (e.g. by taking set photos at the four cardinal points). Thus the exercise of adding points to the satellite imagery of reference was made by matching the author's knowledge of the habitat and the colour/shape of the pixel in the imagery.

| | 1 | 2 |
|---|---|---|
| | **level1** | |
| 2 | conifers | 29 |
| 3 | invasive species | 60 |
| 4 | modified grassland | 277 |
| 5 | tussac | 460 |
| 6 | inland rock | 481 |
| 7 | heath mix | 652 |
| 8 | inland water | 794 |
| 9 | bare ground | 1086 |
| 10 | grassland | 1222 |
| 11 | wetlands | 1429 |
| 12 | sea water | 1454 |
| 13 | heath | 1753 |

**Figure 1: Spread of points added to satellite imagery map per habitat class**

The figure reveals that habitats such as conifers and invasive species present a lower number of points compared to heath and grassland. This is because some habitats, such as conifers, are not common in the Falklands, and others do not frequently occur in large patches, such as invasive species; their detection at ten metre resolution is therefore not as easy as for more widespread and common habitats such as heath and grassland .

Figure 2 shows how the points were distributed across the area of study (the Falklands are in blue in the background). It is worth mentioning that the large blank areas in the NE and SE of the study area are without points simply because there is no satellite imagery to cover them.

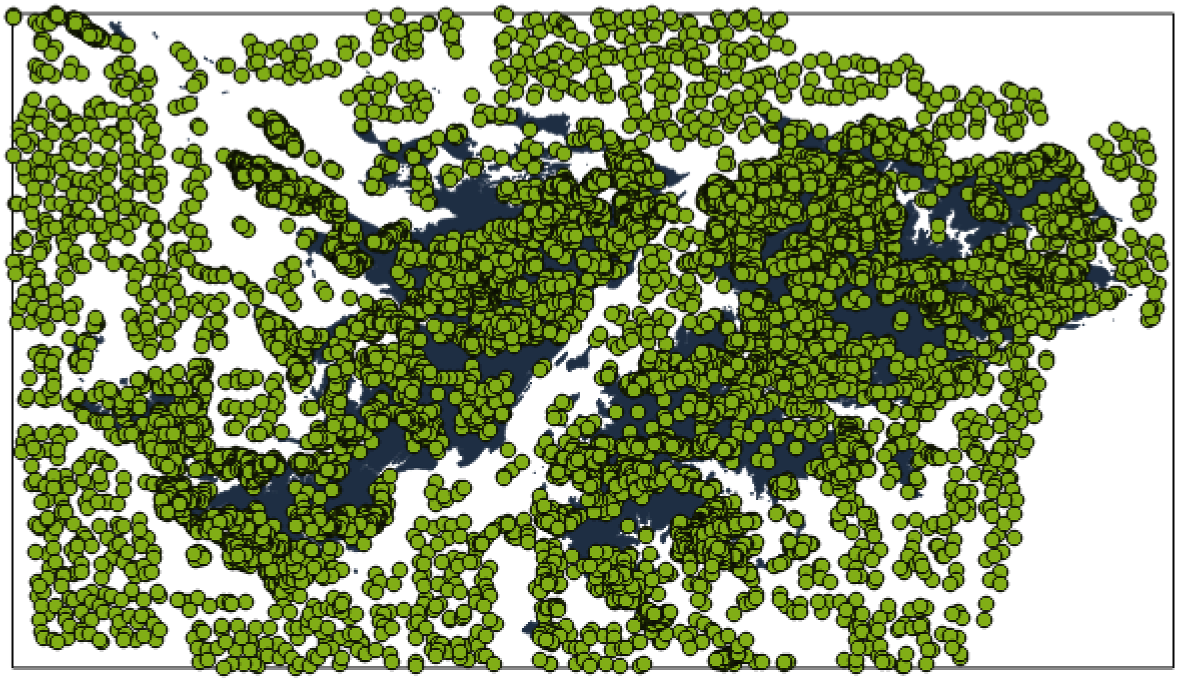


**Figure 2: Distribution of points added to satellite imagery map across the study area**

Following point data collection, satellite imagery was prepared for the computational model. An R script[6] was used as a tool to support the analyses. Initially the script splits the satellite imagery into ten different spectral bands: blue, green, red, red edge1, red edge2, red edge3, NIR1, NIR2, SWIR1 and SWIR2. For each band the reflectance value was calculated (simply dividing by 10,000) and saved individually as raster files.

Next, four environmental indices were calculated: Enhanced vegetation index (EVI), normalised difference vegetation index (NDVI), short wave infrared ratio (SWIR32) and plant senescence reflectance index (pSRI). Additionally, three textural features were derived by using Grey-Level Co-Occurrence Matrices (GLCM) library. These features were mean, variance and homogeneity. The GLMC algorithm obtains textural features by using a moving window across the entire imagery. It is possible to set the 'size' of the window and results will change accordingly. For this exercise the algorithm was run with a 3x3 and 5x5 window (Each pixel is 10 metres so the windows were 30x30 and 50x50 metres).

A new file was created by adding (stacking) the ten bands, four vegetation indices and three textural features. This new layer, comprised of 17 variables, was overlaid with the sample

[6] The script was written by Dr Sergio Godinho – University of Evora, Portugal

points and used as baseline for the function extract. The aim was to extract for each sample point the correspondent values of the 17 variables of the raster file.

The sample points were then split to create training (80% of the original dataset) and testing (20% of the original dataset) sets to generate random selection. The Random Forest model was used to create the habitat map. In simple terms, the model operates through a series of decision trees that are built using the information provided, in this case the 17 variables and the training points which classify the habitat. By looking at the variables and at the points, the model decides (and learns) how to classify the pixels in the image and which variables are leading this classification.

In more technical terms, the model first builds a 'grid search' to test different values for the parameter "mtry" which matches the number of independent variables (the 17 stacked files). Accuracy and confusion matrices were generated to check if the model is a good fit and for the overall performance of each habitat class. A plot showing the most important independent variables was also retrieved. Finally a prediction function was calculated so that the entire area of study could be classified. The resulting habitat map from the single 16/12/2016 imagery is depicted in figure 3.

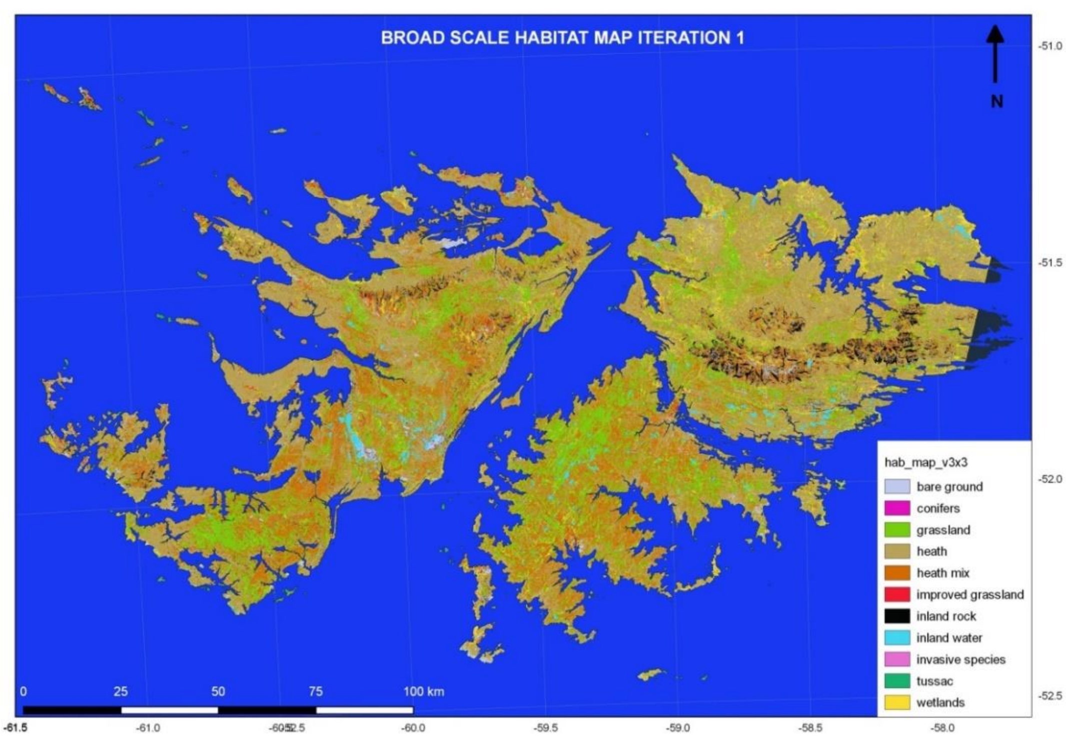



**Figure 3: Habitat map derived from imagery taken on 16/12/2016**

After mosaicking, further imagery covering the eastern most part of East Falkland (in dark blue in figure 3), and adding more sample points, the model was run again to generate a second habitat map. As shown in figure 4 the outcome was less neat and 'noise' appeared in all water-related habitat classes. This may be due to the presence of clouds in the second imagery set, or due to mistakes in the sampling points.

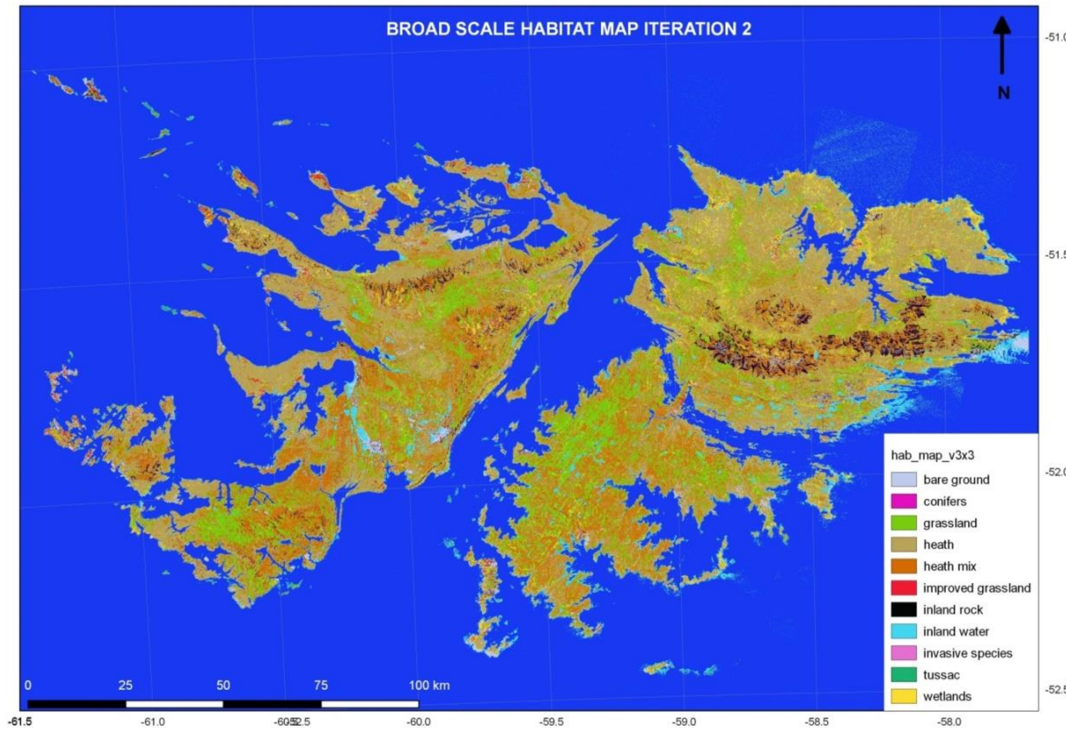



**Figure 4: Habitat map derived from the second iteration of the model**

The accuracy of the first iteration was:

```
Random Forest
6574 samples
  17 predictors
  13 classes: 'bare ground', 'coastal rock', 'conifers', 'grassland', 'heath', 'heath mix', 'improved
  grassland', 'inland rock', 'inland water', 'invasive species', 'sea water', 'tussac', 'wetlands'

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 5 times)
Summary of sample sizes: 5919, 5915, 5915, 5917, 5917, 5916, ...


Resampling results across tuning parameters:

  mtry  Accuracy   Kappa
   1    0.8517273  0.8298868
   2    0.8541283  0.8327571
   3    0.8537908  0.8324212
   4    0.8533080  0.8318987
   5    0.8531835  0.8317725
   6    0.8535516  0.8321981
   7    0.8528804  0.8314612
   8    0.8523937  0.8309107
   9    0.8512397  0.8295865
  10    0.8514507  0.8298542
  11    0.8507825  0.8290856
  12    0.8510868  0.8294462
  13    0.8496893  0.8278492
  14    0.8487741  0.8267982
  15    0.8484403  0.8264101
  16    0.8480753  0.8260130
  17    0.8480434  0.8259715

Accuracy was used to select the optimal model using  the largest value.
The final value used for the model was mtry = 2.
```

The accuracy of the second iteration was:

```
> rf_mask
Random Forest

7763 samples
  21 predictor
  12 classes: 'bare ground', 'conifers', 'grassland', 'heath', 'heath mix', 'inland rock', 'inland water',
  'invasive species', 'modified grassland', 'sea water', 'tussac', 'wetlands'

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 5 times)
Summary of sample sizes: 6957, 6954, 6956, 6954, 6955, 6952, ...
Resampling results across tuning parameters:

  mtry  Accuracy   Kappa
   1    0.8420853  0.8192003
   2    0.8451908  0.8229054
   3    0.8463541  0.8242664
   4    0.8472092  0.8252818
   5    0.8474152  0.8255514
   6    0.8474393  0.8256058
   7    0.8471821  0.8253233
   8    0.8468970  0.8249933
   9    0.8465094  0.8245625
  10    0.8454476  0.8233598
  11    0.8455250  0.8234716
  12    0.8444112  0.8221839
  13    0.8444895  0.8222832
  14    0.8430939  0.8207024
  15    0.8426783  0.8202265
  16    0.8407112  0.8179926
  17    0.8399861  0.8171595

Accuracy was used to select the optimal model using  the largest value.
The final value used for the model was mtry = 6.
```

The confusion matrix of the first iteration was:

```
Confusion Matrix and Statistics

                    Reference
Prediction           bare ground coastal rock conifers grassland heath heath mix improved grassland
  bare ground                158            3        0         1     3         2                  3
  coastal rock                 0            0        0         0     0         0                  0
  conifers                     0            0        2         0     0         0                  0
  grassland                    2            0        0       151     1        27                  5
  heath                        4            0        0         1   262        32                  2
  heath mix                    1            0        0        14    18        43                  6
  improved grassland           0            0        0         1     0         1                 11
  inland rock                 10            1        0         1     0         0                  1
  inland water                 1            1        1         0     0         0                  0
  invasive species             0            0        0         0     0         0                  0
  sea water                    0            0        0         0     0         0                  0
  tussac                       0            0        0         0     0         0                  0
  wetlands                     0            0        1        17     8         0                  9
                    Reference
Prediction           inland rock inland water invasive species sea water tussac wetlands
  bare ground                 15            4                0         0      1        0
  coastal rock                 0            0                0         0      0        0
  conifers                     0            0                0         0      0        1
  grassland                    0            0                2         0     13        5
  heath                        1            1                1         0      3       13
  heath mix                    0            0                0         0      3        1
  improved grassland           0            0                0         0      1        1
  inland rock                 64            0                0         0      0        0
  inland water                 0          143                0         1      0        1
  invasive species             0            0                1         0      0        0
  sea water                    0            0                0       276      0        0
  tussac                       0            0                2         0     51        0
  wetlands                     0            0                3         0      7      217

Overall Statistics

               Accuracy : 0.8424
                 95% CI : (0.8238, 0.8597)
    No Information Rate : 0.1784
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.819
 Mcnemar's Test P-Value : NA
```

The confusion matrix of the first iteration was:

```
> cm_mask
Confusion Matrix and Statistics

                    Reference
Prediction           bare ground conifers grassland heath heath mix inland rock
  bare ground                 94        0        24    36        11          15
  conifers                     1        1         1     0         0           0
  grassland                   26        1        71    52        18          16
  heath                       32        1        59   116        41          25
  heath mix                   12        0        17    32        24           4
  inland rock                 11        0         9    26         8          16
  inland water                 4        0         4     8         1           0
  invasive species             1        0         0     2         0           0
  modified grassland           7        0         7     8         3           2
  sea water                    1        0         2     1         1           3
  tussac                       1        0         3     4         3           1
  wetlands                    25        2        47    65        20          14
                    Reference
Prediction           inland water invasive species modified grassland sea water
  bare ground                   7                1                  3         3
  conifers                      0                0                  1         0
  grassland                     5                0                 16         1
  heath                         4                3                 12         3
  heath mix                     0                0                  3         1
  inland rock                   0                0                  4         0
  inland water                129                0                  0         3
  invasive species              0                1                  0         0
  modified grassland            0                1                  8         0
  sea water                     2                0                  1       271
  tussac                        3                0                  0         1
  wetlands                      5                6                  7         4
                    Reference
Prediction           tussac wetlands
  bare ground             5       22
  conifers                1        0
  grassland               5       38
  heath                   6       62
  heath mix               4       21
  inland rock             0       20
  inland water            2       10
  invasive species        2        1
  modified grassland      1        6
  sea water               0        1
  tussac                 55        2
  wetlands               10      102

Overall Statistics

               Accuracy : 0.4613
                 95% CI : (0.4388, 0.4839)
    No Information Rate : 0.1818
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.3844
 Mcnemar's Test P-Value : NA
```

**Conclusion**

The results of the second iteration did not show an improvement on the first iteration, with the map showing inaccuracies and unexpected noise. More work will be carried out to understand what caused the poor outcome of the second iteration.

For this exercise the SENTINEL-2 imagery was pre-processed using an 'interim' method. The method has since been finalised and standardised, and the imagery will be reprocessed. The R script will be run again using the imagery processed with this updated method.

The first SENTINEL-2 imagery was purchased over a year ago and since then more images were taken in 2017/2018; it would be valuable to use two sets of imagery for the habitat classification to compare temporal changes. The habitat map could also be improved by capturing more points and identifying classes such as sandy shores, cushion plants and possibly fern beds.

Finally, the pixel-based approach used in this study has almost been superseded by object-oriented analyses in the last few years. Instead of sample points, segments (areas) are used to extract values from the 17 variables. Segments, being areas, include more pixels and therefore more pixels are associated to a habitat class. This increases the likelihood of a better classification across the area of study. It is recommended that an object-oriented analyses should be conducted in the future and compared against the results described here from the pixel-based approach.