

UKSA International Partnership Programme (IPP)



EO4cultivar

Project No 417000001416

WP7 Sustainable Livelihoods: Peruvian Case Study

Mapping the Viru Valley



Report prepared by:

G.A. Jones,
D. Colson
P. Robinson,
M. Harris
M.A.E. Smith
C. Cheffings

Joint Nature Conservation Committee

Monkstone House
City Road
Peterborough
PE1 1JY
UK

Tel: +44 (0)1733 562626
<https://jncc.gov.uk/>

and

G. Bell
J.A. Parker

Environment Systems Ltd.

9 Cefn Llan Science Park
Aberystwyth
Ceredigion
SY23 3AH
UK

Tel: +44 (0)1970 626688
www.envsys.co.uk

UKSA International Partnership Programme (IPP)

Contents

1	Summary	1
	Outline of method	1
2	Introduction	2
3	Processing Sentinel-1 and Sentinel-2 data.....	3
3.1	Identification of Sentinel-2 data	3
3.2	Identification of Sentinel-1 data	4
3.3	Pre-processing of Sentinel data	4
4	Very High Resolution (VHR) data.....	4
4.1	Purchasing of VHR optical data.....	5
4.2	Pre-processing of VHR data.....	6
4.3	Purchase of VHR Digital Terrain Model (DTM).....	6
5	Classification of broad habitat map.....	6
6	Field visit to the Viru Valley to collect validation data	7
7	Final iteration of classification for broad habitat map.....	7
7.1	Mapped Outputs	8
7.2	Accuracy.....	10
8	Classification of features of interest using VHR optical data.....	10
8.1	Mapped Outputs	12
8.2	Accuracy.....	12
Annex 1	Implementing the Living Maps Method (LMM) in the Viru Valley.....	13
Annex 2	Feature extraction using VHR optical data.....	21
Annex 3	Additional work after completion of map.....	23
References.....		25

1 Summary

This paper outlines the process of generating a habitat map for the Viru Valley which will be used as a key input into the Spatial Evidence for Natural Capital Evaluation (SENCE) Approach to ecosystem services mapping.

Outline of method

(according to timeline)

- **Processing Sentinel-1 and Sentinel-2 data.** Identifying suitable Sentinel-2 scenes with minimal cloud cover of the area of interest and to cover seasonal differences. Sentinel-1 scenes that coincide with dates of Sentinel-2 imagery are chosen. All scenes processed by Environment System Ltd.'s Data Services (WP6).
- **Purchase of Very High Resolution (VHR) optical data.** To identify some features, such as linear water courses, which are very important for the SENCE mapping but cannot be identified to the desired precision with Sentinel data, commercial VHR data was sourced.
- **Classification (first iteration) of broad habitat map produced using the Living Maps Method (LMM).** Using the Sentinel data, a first iteration of the habitat map was produced. The LMM was chosen as a suitable method due to the repeatability of the process using machine learning and efficiency for reiteration. The biggest limitation was the unavailability of field data to inform the classes present and the characteristics of the classes to aid separation.
- **Field visit to collect validation points of first iteration of habitat map.** To ensure the habitat map represented the landscape as accurately as possible, and with no field data sourced, a field visit was arranged with local experts to collect validation points.
- **Purchase of VHR Digital Terrain Model (DTM).** Due to the complex terrain, and gaps in the SRTM 30 m product used until this point, the project identified a need to obtain a more precise and higher spatial resolution DTM. Required for orthorectification of VHR optical data and for SENCE mapping.
- **Classification (final iteration) of broad habitat map.** To produce a more accurate map to reflect any changes in classes required after field visit, and to incorporate the VHR DTM into the classification. Overall accuracy of habitat map is 74%.
- **Classification of features of interest using the VHR optical data.** Extraction of linear features such as trees and scrub surrounding field boundaries and water features of importance to SENCE mapping.
- **[Additional work after completion of map]** Following the production of the first habitat map it became apparent that some minor revisions were required to a small number of classes. A further two additional classes were also identified to improve ecosystem service mapping outputs. Subsequent revisions are described in Annex 3.

2 Introduction

Habitat maps are a key component to understanding the distribution and extent of features across the landscape, and are a requirement in facilitating better management practices, ecosystem service mapping, interpreting and targeting biodiversity monitoring and delivering policies such as those related to sustainable development goals. There are many global land cover and land use products available, but these do not provide enough detail to perform analyses on a regional level. No known accessible habitat map was available for the Viru Valley in Peru; therefore, a habitat map was required to carry out the ecosystem service mapping and to help inform the research required to shape the sustainable livelihoods case study outputs. These aim to introduce options for informing sustainable ecosystem-based management options within the Peruvian agricultural land matrix.

Earth Observation (EO) has been extensively used to provide a synoptic view of land use, cover and change at a variety of scales. New sensors are being developed and launched at an increasing rate, with some missions making data accessible through open source licensing (such as the Copernicus Programme's Sentinel data). EO is a valuable resource when no other data are available but is most powerful when combined with field data and a variety of other data sources to create products that provide critical information, particularly for decision making.

The EO4Cultivar Project key objective is delivering analysis-ready and value-added data to the South American agricultural supply chain and farmer advisory services. The sustainable livelihoods work package was tasked to deliver ecosystem service case studies; these provide analysis and context on agricultural systems within the wider environment to help stakeholders better understand the functioning and benefits delivered by sustainable use of soil, water and biodiversity. In order to provide results that would benefit both commercial and smallholder operation the Peruvian case study focused on the Viru Valley in the La Libertad region of Peru (Figure 1), one of six key agricultural regions in Peru (World Bank, 2017). Viru supports growing operations of varying sizes, including considerable extent of operations managed by the project's International Partner Organisations. This area of interest was selected in agreement with project partners and stakeholders prior to the data collection phases commencing.

3 Processing Sentinel-1 and Sentinel-2 data

Copernicus datasets have been identified as a key data source for this project because of their open access. The EO4Cultivar project aims to enhance the uptake of this open access resource and these data are utilised to deliver many elements of the different work packages. For the mapping work, both Sentinel-1 and Sentinel-2 were used to capture as many features of the landscape as possible. The area of interest for the habitat map was agreed upon through discussions with project partners (Figure 1).

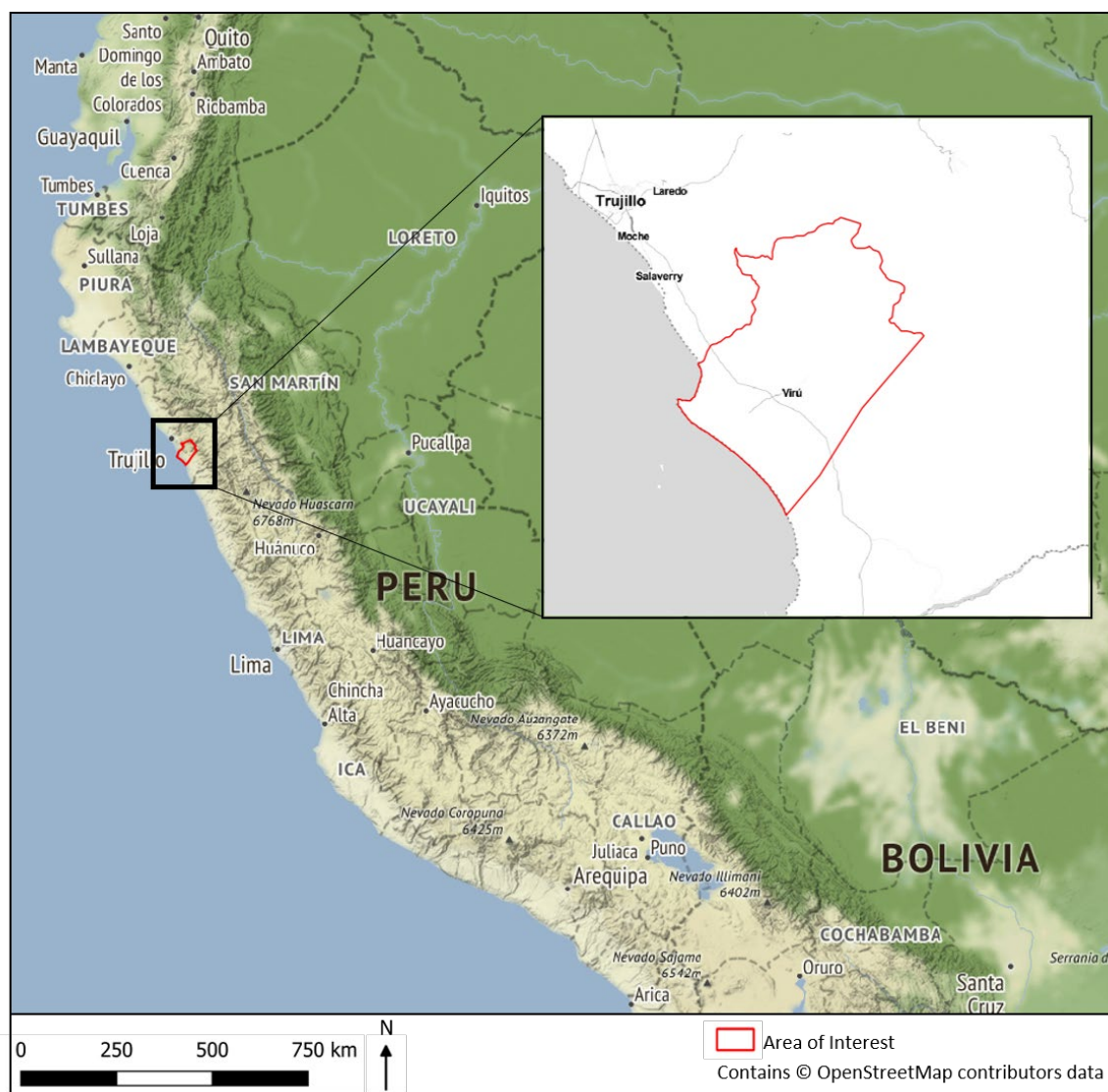


Figure 1: Area of interest for the Viru Valley habitat mapping.

3.1 Identification of Sentinel-2 data

To secure cloud-free data of the area of interest an investigation was required to identify suitable scenes. Seasonal changes and differences were considered as these variations can be critical to habitat identification and separation of classes. For example, leaf flushes showed a differentiation of habitat structures in the upper valley reaches. The mapping year is 2017 and the scenes from the following dates were identified:

- 4th January 2017
- 14th April 2017
- 13th June 2017
- 15th November 2017
- 20th November 2017

The 15th and 20th of November scenes were merged to create one November scene. This is to fill any gaps in the November scene left by cloud cover. This provided approximately quarterly scenes and hence were considered sufficient to identify most seasonal changes.

3.2 Identification of Sentinel-1 data

Sentinel-1 is not constrained to the same limitations as Sentinel-2, because this radar remote sensing tool can image Earth's surface through cloud cover and during periods of darkness. Not every available scene is required to produce a habitat map, therefore Sentinel-1 scenes are chosen to capture the seasonal changes and differences only. As Sentinel-1 and Sentinel-2 satellites are not orbiting the same place at the same time, to align with the Sentinel-2 images scenes with dates that corresponded as closely as possible were identified:

- 14th January 2017
- 20th April 2017
- 10th June 2017
- 13th November 2017

3.3 Pre-processing of Sentinel data

Work package 6 of the EO4Cultivar project provides data services which include the pre-processing of Sentinel data. Once the required scenes were identified, the resource available from data services, provided by Environment Systems Ltd, was used to pre-process these data.

Sentinel-1 data were delivered to JNCC as a backscatter product generated from data collected in the Interferometric Wide swath (IW) mode and processed from the Ground Range Detected (GRD) version of the data made available by ESA.

Sentinel-2 data were delivered to JNCC as a bottom-of-atmosphere product with clouds masked out. 10 m and 20 m bands were stacked as separate images. As part of the services carried out by Environment Systems' data services, vegetation indices were also delivered (Enhanced Vegetation Index (EVI), Normalised Difference Vegetation Index (NDVI), and Normalised Difference Water Index (NDWI)).

For more information on the details of pre-processing methods see Environment Systems documentation for Work Package 6.

4 Very High Resolution (VHR) data.

It was realised during the initial phases of habitat mapping that some landscape features could not be classified due to the low spatial resolution of the Sentinel data (10-20 m). These features were deemed important to understand the ecological and geological functions and

processes that were operating that are key for the SENCE ecosystem service mapping. It was agreed that commercial very high resolution optical data would be purchased.

4.1 Purchasing of VHR optical data.

To make the process of ecosystem service mapping as repeatable and sustainable, in terms of financial cost of data, the following factors were considered when deciding upon image purchase:

- Only archive imagery from various providers would be considered, as these are cheaper and allows the user to choose from a variety of available images, which commissioning a new image would not;
- The spatial resolution needs to be high enough to separate and classify the features;
- No cloud, or as little cloud as possible;
- Spectral requirements are minimum of Red, Green, Blue and Infra-red bands;
- Only needs to cover areas where finer features exist, which excludes large parts of the Andes included in the broad habitat map Area of Interest (AOI).

Following the above criteria, the most suitable images were identified as Pleiades images (Figure 2).

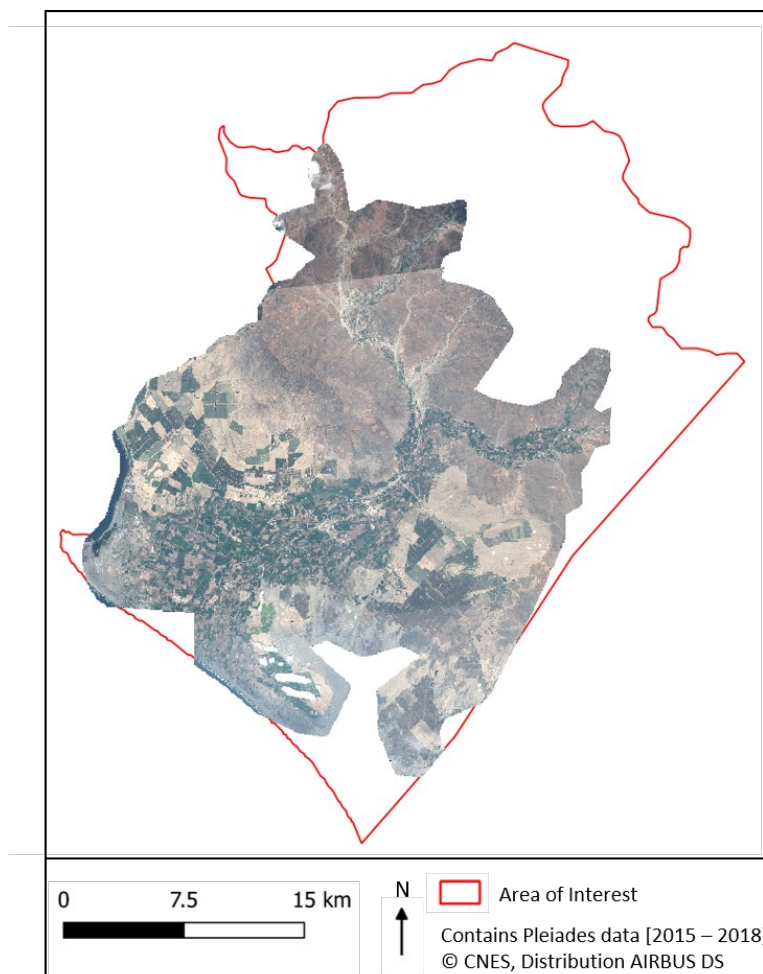


Figure 2: Mosaic of Pleiades images purchased for project covering smaller AOI. Includes all 5 images purchased from Airbus.

4.2 Pre-processing of VHR data.

The Viru Valley has extreme variations of elevation due to the geographic proximity to the Andes. The topographically diverse nature of the landscape led to inherent distortion in the VHR data. For example, some areas had approximately 500 metres of horizontal shift due to terrain displacement. To prepare the data for use in feature extraction and classification, an orthorectification step was performed. This enables the topography to be accounted for and attempts to accurately remove image distortions.

An attempt was made to orthorectify the VHR data using the Shuttle Radar Topography Mission (SRTM) 30 m product, however the complex terrain and gaps in the imagery did not markedly improve the original terrain displacement of ~500 m. Subsequently, Peru geoids were also employed in conjunction with the SRTM, but this did not significantly improve the outputs. Therefore, the project identified the requirement to employ the use of a VHR Digital Terrain Model (DTM). Orthorectification using the WorldDEM™ product was successful and the terrain-corrected VHR corresponded to the Sentinel-2 data and Bing Aerial data, available in the native coordinate reference system in QGIS 2.14.5.

4.3 Purchase of VHR Digital Terrain Model (DTM)

Due to the complex terrain, and gaps in the Shuttle Radar Topography Mission (SRTM) 30 m data, the project recognised the need to obtain a more precise and higher spatial resolution Digital Terrain Model (DTM). This data assisted orthorectification of VHR optical data and SENCE ecosystem service mapping. WorldDEM™ product was used as other alternatives had licensing restrictions which would mean other work packages in EO4Cultivar could not utilise the data for their own requirements.

5 Classification of broad habitat map

The classification method chosen for the generation of the habitat map was the Living Map Method (LMM) (Kilcoyne *et al.*, 2017). The LMM developed a code base that allowed users to implement a machine learning approach to classification, with the flexibility to integrate numerous different data types, including earth observation data. The availability of a ready-made code base that is straight forward to adapt provides an advantage over other similar open source tools, as it removes the time needed to write code for implementation. Other reasons for choosing the method include:

- Utilises machine learning algorithms which are increasingly used in mapping applications utilising earth observation data;
- Can be carried out on pixel-based and object-based analyses;
- All steps can be carried out using open-source software;
- The code is available under an Open Government License;
- JNCC had a minor role in supporting the development of the code, and hence had good familiarity with the method;
- Can be repeated across different areas and a variety of datasets with little adaption to code;
- Iterations of maps can be generated quickly, efficiently, and it is highly repeatable.

For details on the datasets, classes and method, including data preparation, segmentation, training dataset creation and outputs, see Annex 1 of this report.

6 Field visit to the Viru Valley to collect validation data

Following the completion of the first draft of the broad habitat map, data were required to validate the outputs and re-iterate the classification process if necessary. To generate a field survey, a set of points from each class was randomly generated using the outputs of the first draft habitat map. Figure 3 shows the location of these points, which were manually checked to ensure that no point overlapped with any training points used to generate the classification algorithm. However, due to the terrain, access restrictions, and health and safety considerations, only a subset of the points were visited and recorded. Data were collected for all broad habitat classes except 'erosion gully – dense scrub'. Figure 3 shows the points randomly generated and those that the field team were able to visit feasibly during the five-day field campaign. This introduces some bias and spatial correlation to the validation set, which becomes more evident in section 7.2.

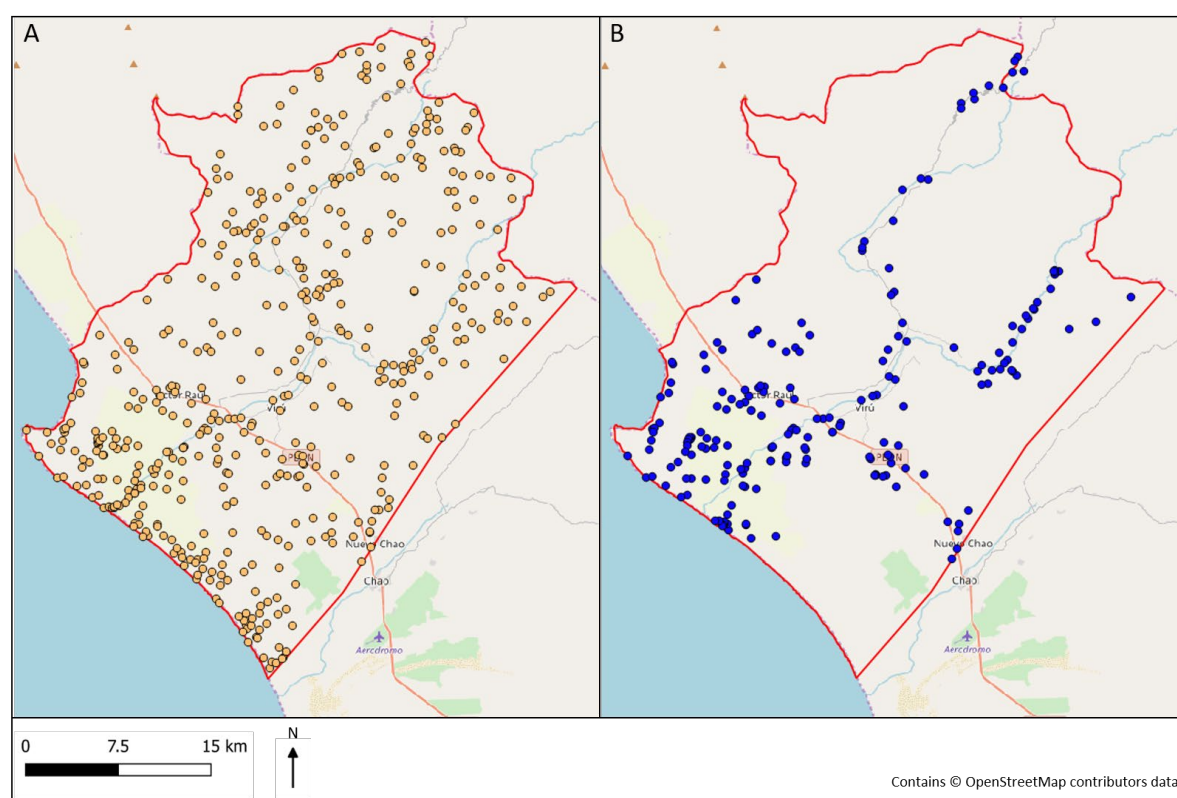


Figure 3: Location of points to visit during the Viru Valley field campaign. A) Points to visit B) Points visited.

7 Final iteration of classification for broad habitat map

Following the field visit, it was decided that two new classes were required in the classification scheme. These were sand dunes with vegetation, and reservoir. Additional training data were

created for these classes, based on the information collected in the field, and the classification process was re-iterated.

7.1 Mapped Outputs

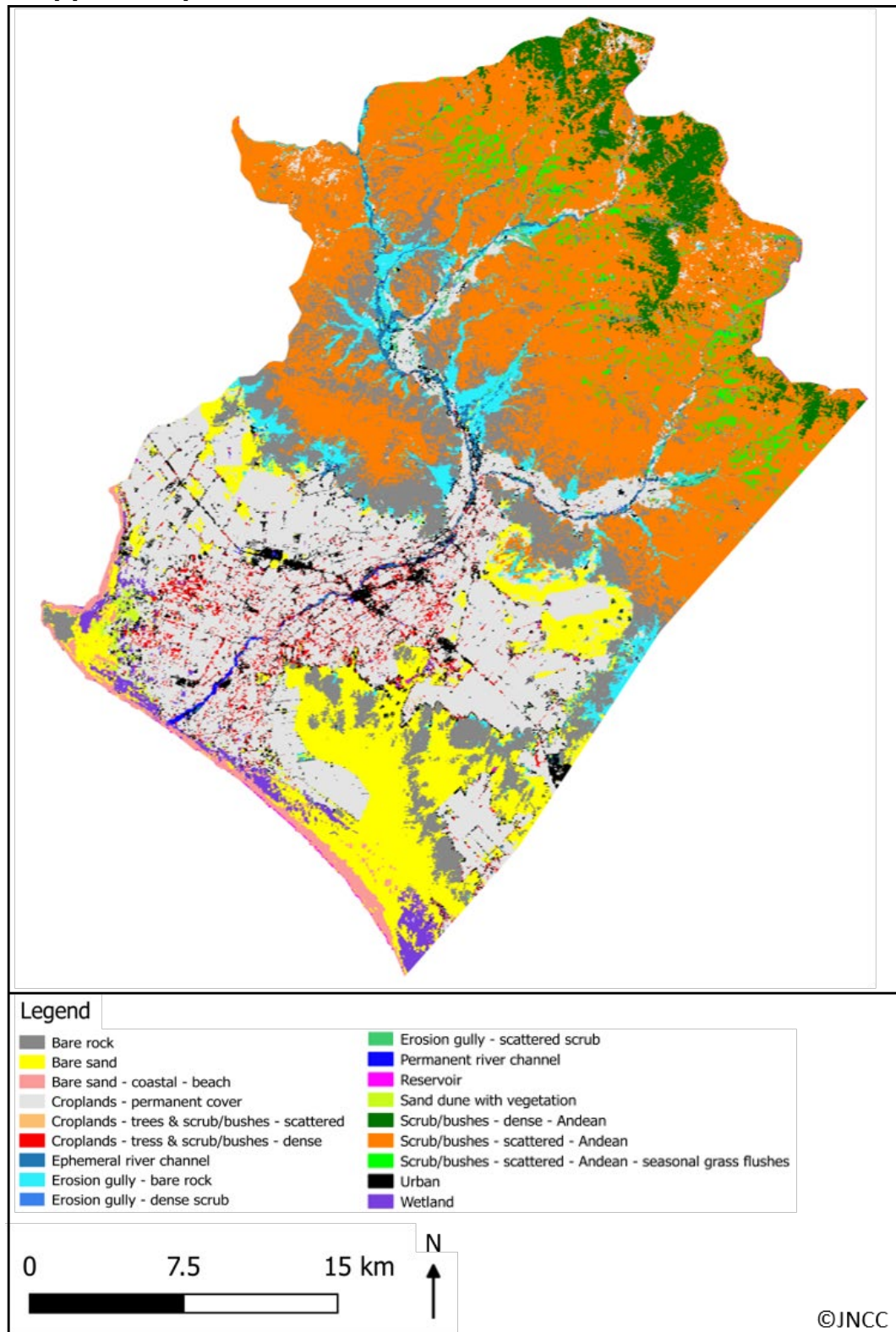


Figure 4: Shows a representation of the broad habitat map output.

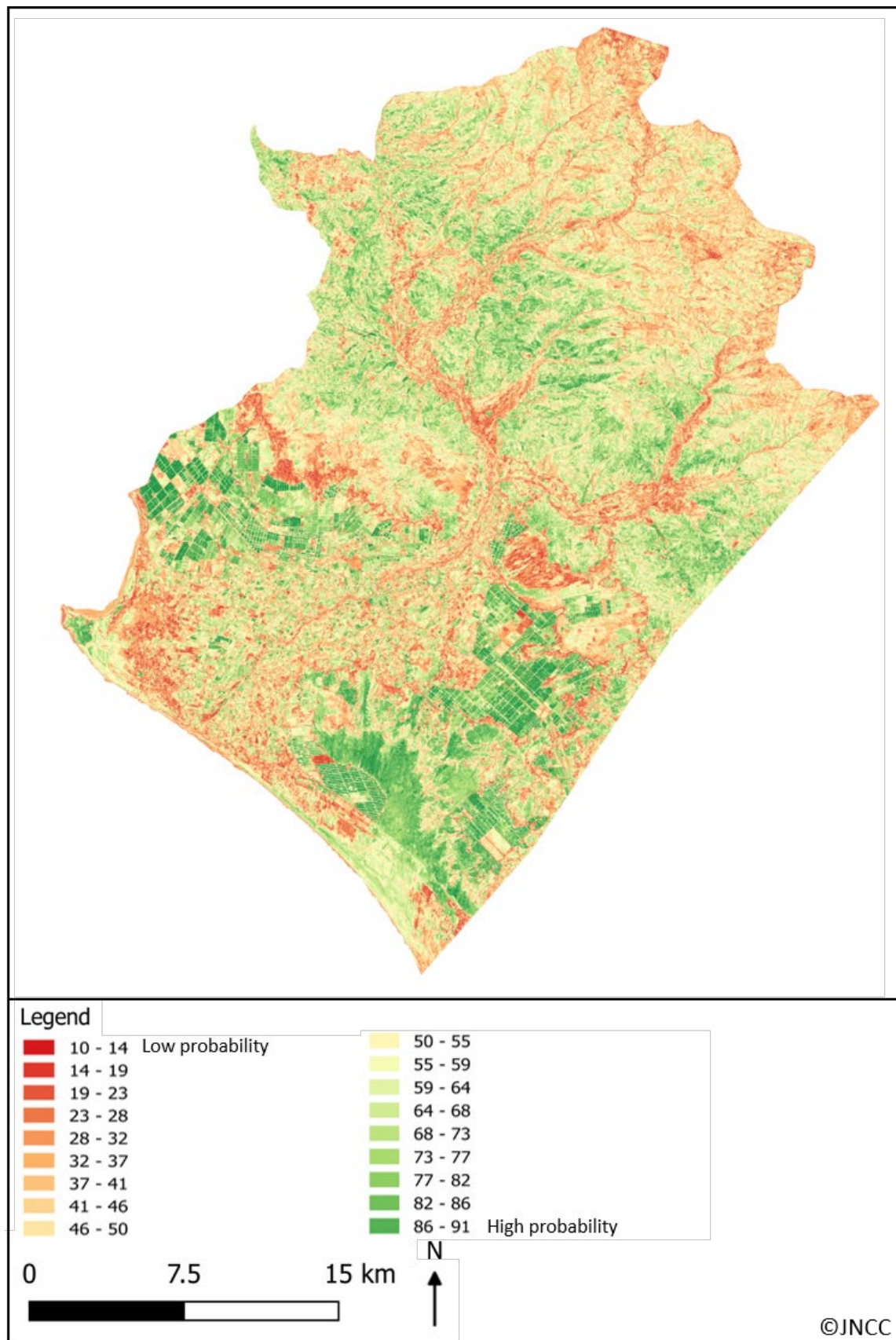


Figure 5: Probability of classes being the identified class output, generated from the random forest algorithm.

7.2 Accuracy

To analyse the accuracy of the outputs, the data collected during the field visit only were used to generate an error matrix of the habitat map. The overall estimated accuracy of the map, which gives us an indication of how many of the field data were classified correctly, is 74%.

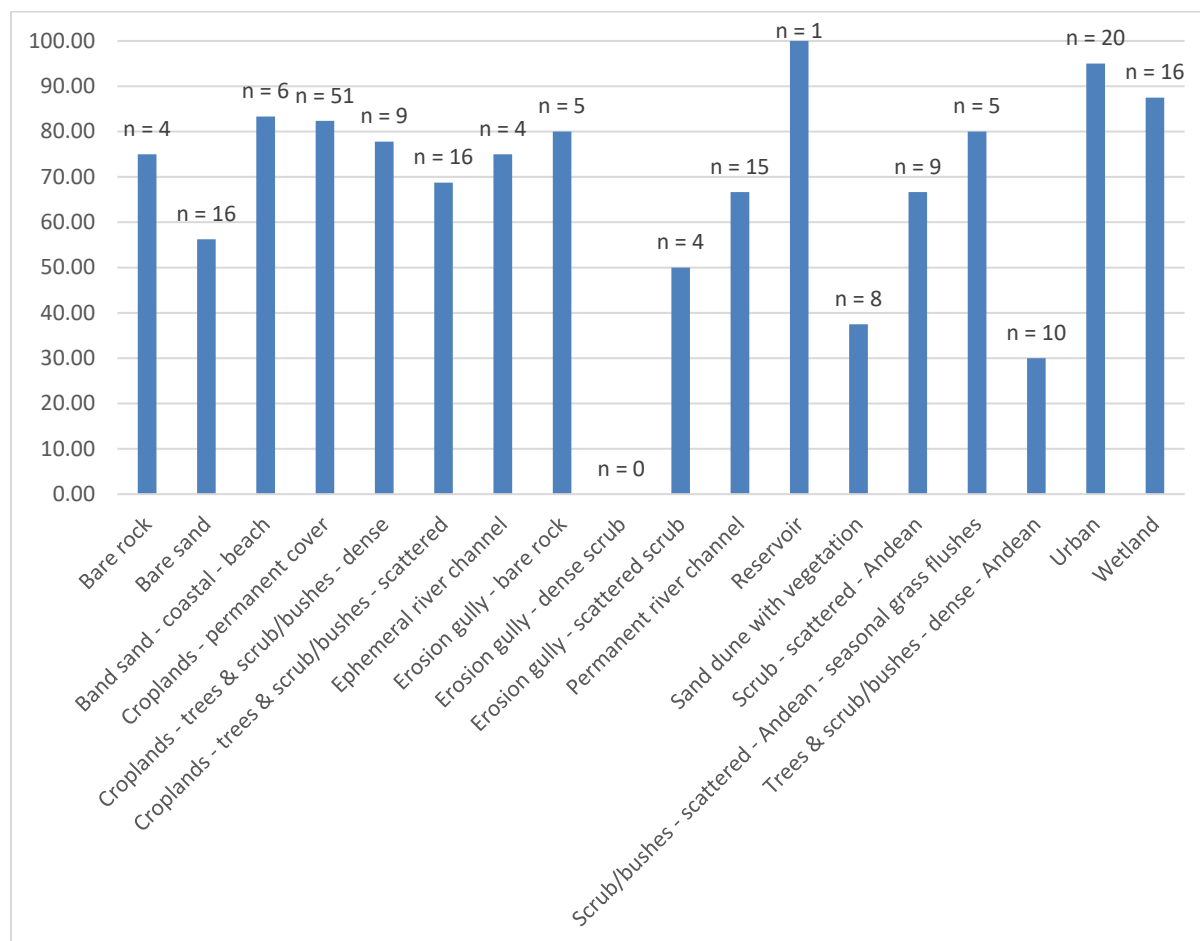


Figure 6: A graphical output of the user's accuracy from the error matrix, where 'n' is the number of field validation points available per class.

In order to interpret these results, an assessment of the error matrix is required. It should be noted that these are estimates of accuracy which have variable certainty associated with them due to differing field data sample numbers. For example, as no validation points were collected for the class 'erosion gully – dense scrub' the estimated accuracy is unknown but also impacts the overall accuracy, which is reduced as a result. There are several other classes with only one to four points (for example Ephemeral river channel, Reservoir, Bare rock) requiring careful interpretation of the accuracy of the map. This means that a more representative sample of these classes is required to fully understand how successful the map is at representing the landscape of the area.

8 Classification of features of interest using VHR optical data.

The use of VHR optical data complemented features identified in the broad habitat mapping and contributed important aspects to the sustainable livelihoods work package. Elements of

the landscape that the low spatial resolution of the Sentinel data (10-20 m) struggled to identify included linear vegetation and water features; several of the artificial water channels are approximately 8m wide. The features can introduce spectral mixing and are therefore not easily identifiable in Sentinel data when performing an object-based image classification (Figure 7).

The linear vegetation and water channels feature regularly in the agricultural regions of the Viru Valley. An understanding of their geographical spread is important to comprehend the ecological functions provided and assist the SENCE ecosystem services mapping.

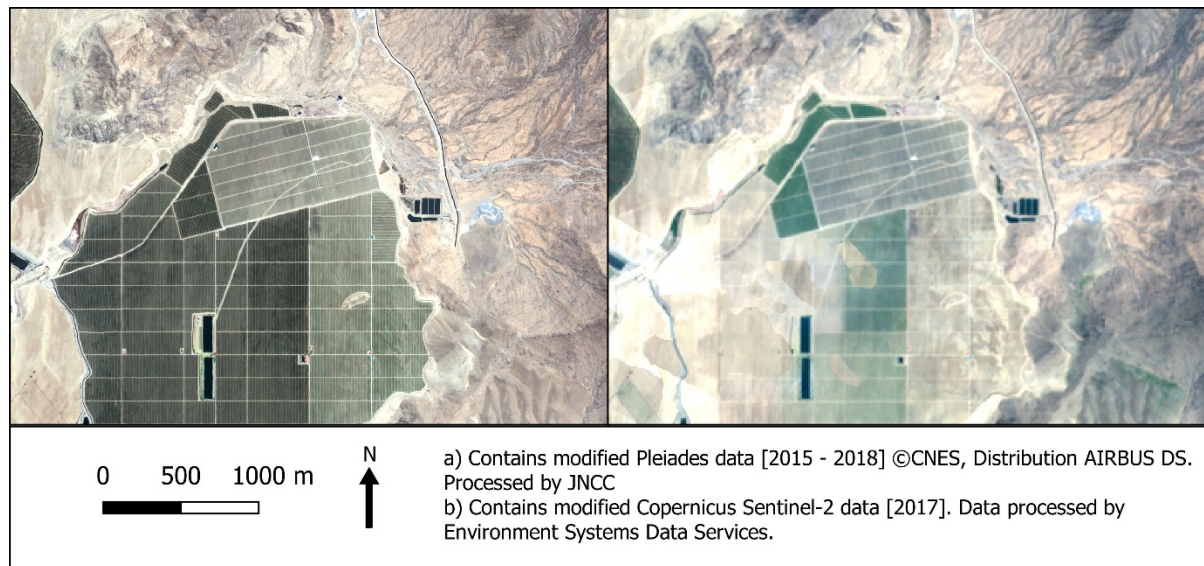


Figure 7: Example of value added from the (a) VHR data when compared to the (b) Sentinel-2 data. Water features are clearly discernible in the VHR data, and, whilst visible in Sentinel-2, are not so well-defined.

The methods chosen to delineate the features of interest were implemented using the Orfeo Toolbox (CNES 2018) suite of tools. This was justified due to the open nature of the toolbox allowing for repeatability without incurring significant licensing costs.

For details on the datasets, outputs and method, including data preparation, see Annex 2 of this report.

8.1 Mapped Outputs

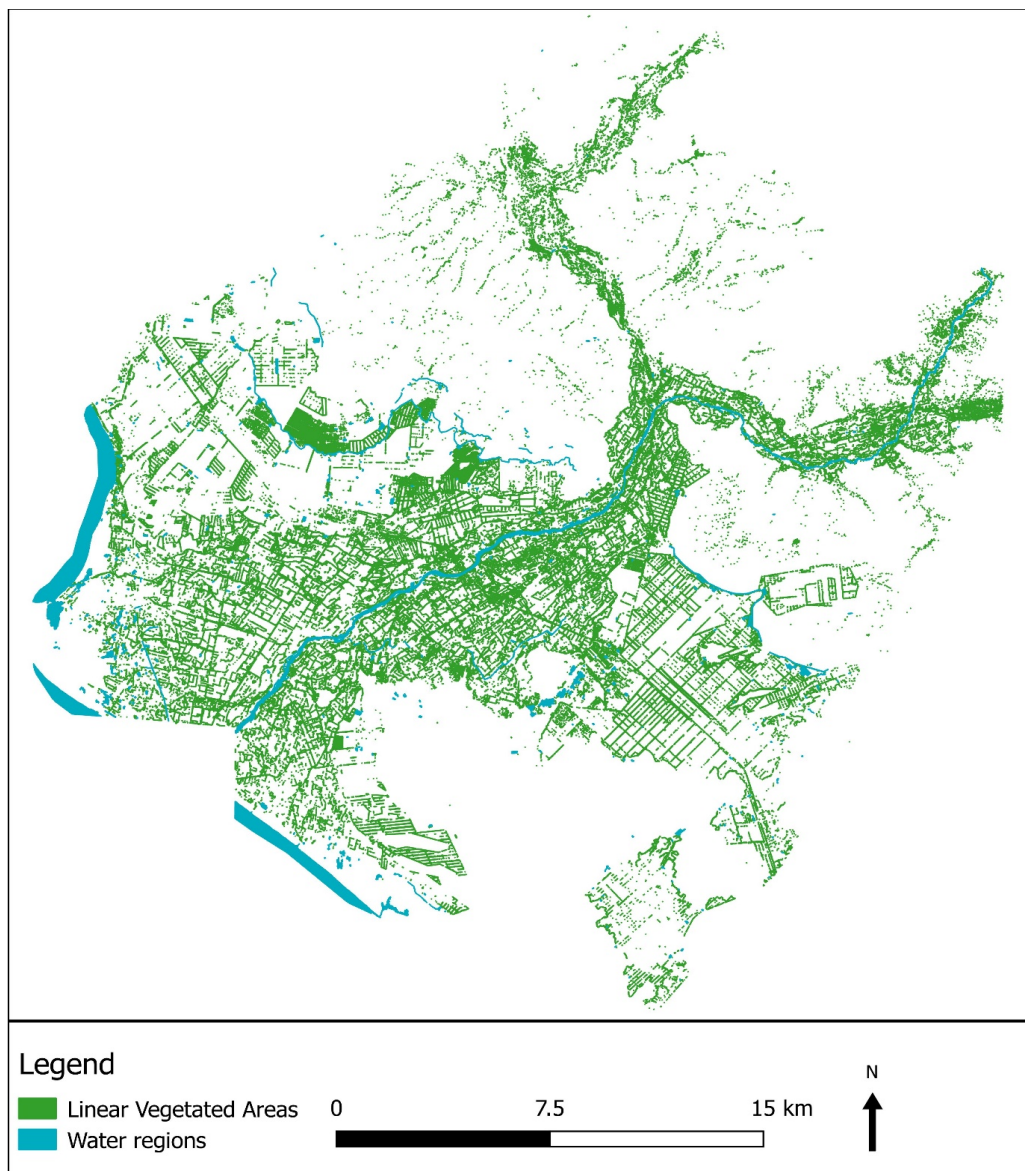


Figure 8: Shows a representation of the feature extraction output

8.2 Accuracy

In order to analyse the accuracy of the outputs, a visual comparison with OpenStreetMap derived water zones & Google Earth Imagery and was used, due to the lack of suitable available field data. There is a very good overall agreement between the water features present and both reference images. Due to the nature of the linear vegetated features, the overall agreement of these features are harder to interpret, however, many of the features associated with field boundaries are present.

Annex 1

Implementing the Living Maps Method (LMM) in the Viru Valley

1. Datasets

A summary of each of the datasets used for input to the habitat classification is listed below. All data were georeferenced to WGS 84 / UTM zone 17S (EPSG:32717) and were resampled to 10 m unless otherwise specified.

1.1 Sentinel-1

Sentinel-1 is a polar-orbiting, all weather, day-and-night radar imaging mission for land and ocean services. The mission is part of the European Union (EU) Copernicus Programme and is operated by the European Space Agency (ESA). The mission consists of a two-satellite constellation providing orbit revisit times of six days. The radar instrument transmits and receives in C-band (5.405 GHz) at a resolution of 5 x 20 m in Interferometric wide-swath mode. This is the mode that is used over land masses in Europe, including England. The data is processed and stored as Level 1 Single Look Complex (SLC) and Level 1 Ground Range Detected (GRD) products by the ground segment of ESA. The SLC product contains the intensity of returns and phase information, whereas the GRD product does not contain the phase information, due to the enhanced processing that it receives. This GRD data is multi-looked and projected to ground range using the earth ellipsoid model.

As stated in the report, Sentinel-1 data were delivered to JNCC as a backscatter product generated from Interferometric Wide (IW) Ground Range Detected (GRD) data by Environment Systems Ltd. Products included VV and VH polarisations as well as a band of the ratio of the two polarisations.

1.2 Sentinel-2

Sentinel-2 is a polar-orbiting, multispectral high-resolution imaging mission for land monitoring. The mission is also part of the EU's Copernicus Programme and operated by ESA. The mission consists of a two-satellite constellation providing orbit revisit times 10 days at the equator with one satellite, and 5 days with 2 satellites under cloud free conditions which results in 2-3 days at mid-latitudes. The optical instrument payload samples 13 spectral bands: four bands at 10 m, six bands at 20 m and three bands at 60 m spatial resolution. The orbital swath width is 290 km.

As stated in the report, Sentinel-2 data were delivered to JNCC as a bottom-of-atmosphere product with clouds masked out. 10 m and 20 m bands were stacked as separate images. As part of the services carried out by Environment Systems' data services, vegetation indices were also delivered (Enhanced Vegetation Index (EVI), Normalised Difference Vegetation Index (NDVI), and Normalised Difference Water Index (NDWI)).

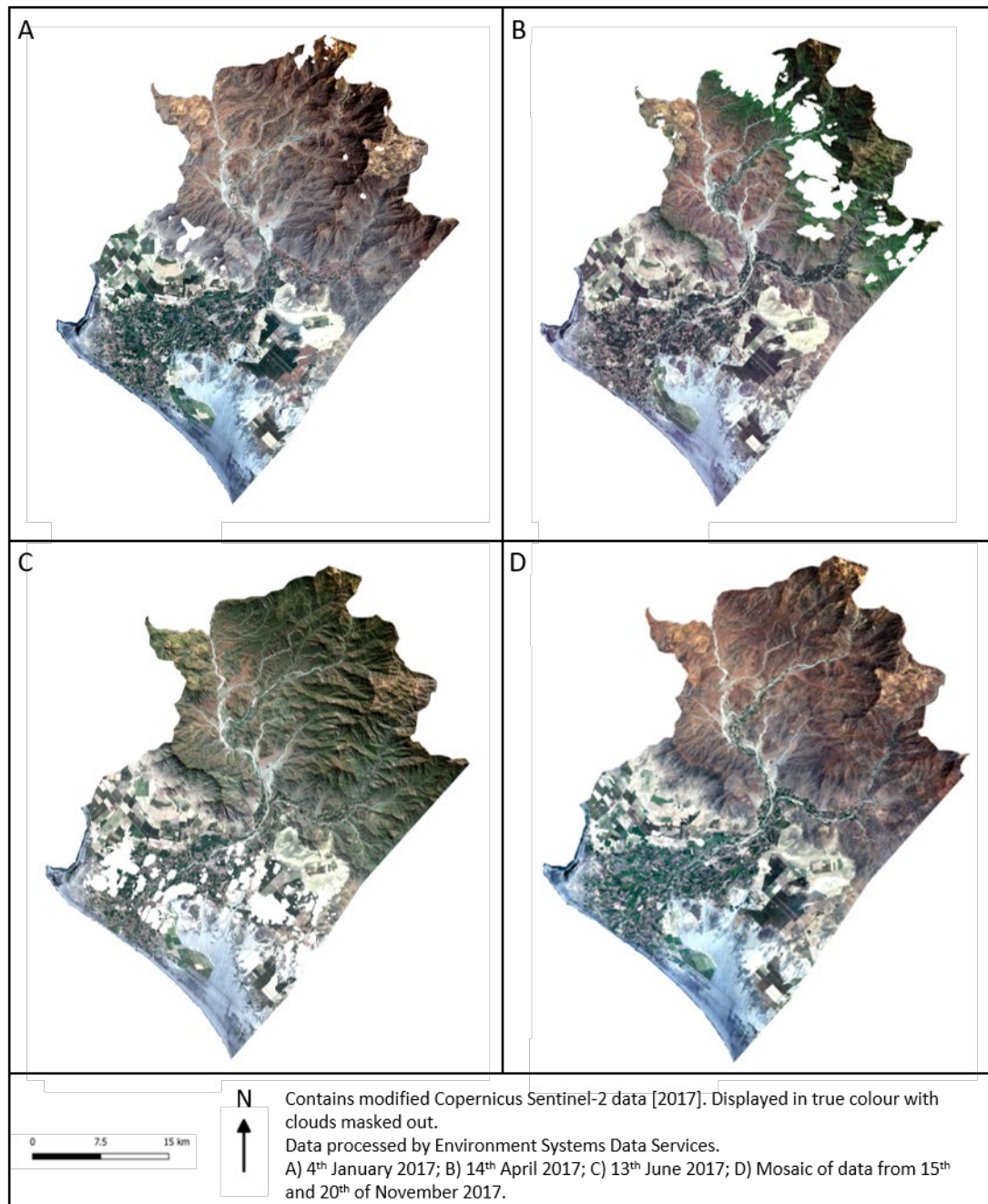


Figure 1.1: Sentinel-2 images which have been cropped to the Viru Valley AOI.

Additional processing includes resampling of the 20 m bands to 10 m and stacking all the bands together to create one image per date. ESA's SNAP toolbox (<http://step.esa.int/main/toolboxes/snap/>) was used to create Linear Spectral Unmixing (LSU) layers for productive vegetation, non-productive vegetation and shade / water for all four Sentinel-2 scenes. By identifying pure pixels of desired criteria, each pixel is attributed with affinity to the training pixels. For more detail on the LSU process see Living Map Method Report (Kilcoyne *et al.*, 2017).

1.3 Digital Elevation Model (DEM)

A DEM provided by NASA's Shuttle Radar Topography Mission (SRTM) in 2000 was used for the first draft of the habitat map (before the WorldDEM product was purchased). The SRTM data is open source and is available at a resolution of 1 arc-second, or about 30 m. Data were downloaded from Earth Resources Observation and Science (EROS) Center website (<https://www.usgs.gov/centers/eros>). Layers of slope and aspect were also created, and all layers were resampled to 10 m spatial resolution.

For the final draft of the map, the purchased WorldDEM product was used instead of the SRTM data. However, the WorldDEM does not cover all the area of interest, so the SRTM product was used to fill in the gaps.

1.4 Bioclim

Maximum temperature, minimum temperature and annual rainfall was downloaded as freely available raster layers at ~1km spatial resolution (30 arc-seconds) from WorldClim (<http://worldclim.org/bioclim>).

1.5 Crop mask

A version of the crop mask, generated in Work Package 5 was delivered as a shapefile by Environment Systems Ltd. The shapefile included boundaries of all the crop areas in the Viru valley and generated using a time-series of Sentinel-1 data. For more information see Work Package 5 documentation.

1.6 Open Street Map data

OpenStreetMap® are open data, licensed under the Open Data Commons Open Database License (ODbL) by the OpenStreetMap Foundation (OSMF). The data were downloaded through a QGIS plugin (https://wiki.openstreetmap.org/wiki/QGIS_OSM_Plugin) and included layers on buildings and roads.

2. Training Data

There were no data available from the field for use in this project for habitat mapping. Training data were therefore created by interpreting the landscape using Google Earth (https://www.google.co.uk/intl/en_uk/earth/). Visual interpretation of the Sentinel-2 imagery and derived products such as Normalised Difference Vegetation Index (NDVI) were also used in the generation of training data. Figure 1.2 shows the stages to generate training data.

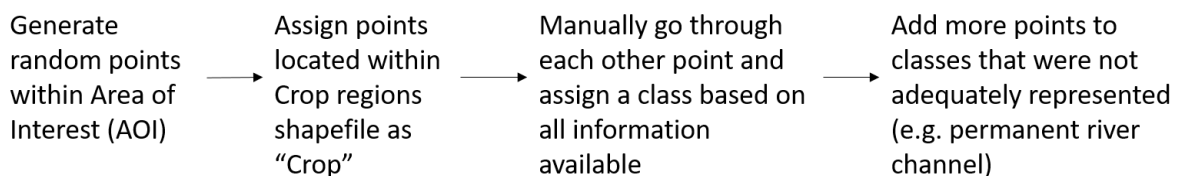


Figure 1.2: Method to identify habitat classes for use within a classification.

3. Classes

Classes are defined in relation to habitats of interest for the SENCE ecosystem service mapping.

Table 3.1: Class names and descriptions.

Class Name	Class Description
Bare Rock	Defined as an area of bedrock that is exposed. Covers varying types of rock types.
Bare Sand	Includes areas of bare ground that are sandy. Does not include coastal areas or areas of sandy deposits within erosion gullies.
Bare sand – coastal - beach	Bare sand that is located near the coast.
Croplands – permanent cover	Areas that are permanently covered with agricultural activity of some kind. Includes a variety of crop types including grasslands used for grazing.
Croplands – trees & scrub/bushes - dense	Areas surrounding field boundaries that include trees and scrub bushes in a dense format. Includes hedgerows.
Croplands – trees & scrub/bushes - scattered	Areas surrounding field boundaries that include trees and scrub bushes but that are scattered, i.e. individual trees and bushes are not connecting but are scattered across the field boundary.
Ephemeral river channel	Defines the channel of a river where water flows for short periods typically after snow melts or heavy rainfall.
Erosion gully – bare rock	A landform created by running water, eroding the landscape and includes no vegetation.
Erosion gully – dense scrub	A landform created by running water, eroding the landscape, typically on a steep hillside and includes areas of dense scrub along the formed gullies.
Erosion gully – scattered scrub	A landform created by running water, eroding the landscape, and includes scattered scrub, grass or bushes.
Permanent river channel	A channel that includes a permanent flow of water.
Reservoir	Areas of artificial lakes created as a source of water supply.
Sand dune with vegetation	A ridge of sand created by wind, typically located behind coastal areas, and includes some areas of scattered vegetation types.
Scrub/bushes – dense - Andean	Defined as areas of Andean steppe that are covered in dense scrub bushes.
Scrub/bushes – scattered - Andean	Defined as areas of Andean steppe that are covered in scattered scrub bushes.

Scrub/bushes – scattered – Andean – seasonal grass flushes	Defined as areas of Andean steppe that are covered in scattered scrub bushes that include areas of seasonal grass flushes.
Urban	This class includes any built structure, including houses, farm buildings, roads and any other built structures, tarmac, concrete or gravel.
Wetland	Defined as land consisting of marshes, swamps or any other form of saturated land. Includes salt marshes.

4. Method

A summary of steps is outlined in the flowchart in Figure 1.4.

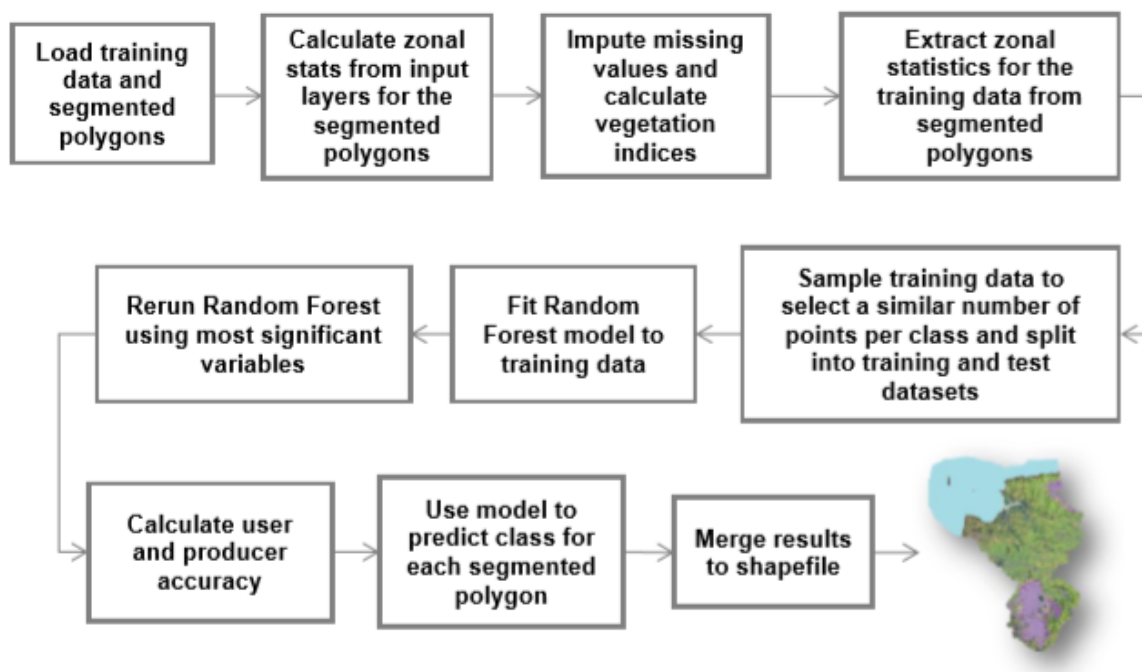


Figure 1.4: Flowchart of classification process (Kilcoyne *et al.*, 2017).

The method follows the steps of the Living Maps Method. More detail on the steps can be seen in the technical report (Kilcoyne *et al.*, 2017). The only difference to the method outlined in this report is the segmentation process, and the number of significant variables identified to improve the model fit of Random Forest, which is outlined below.

4.1 Segmentation

The segmentation was applied in Trimble eCognition version 9.2.1. The processing of the segmentation used the Sentinel-2 images from June the 13th 2017. The segmentation algorithm used was a combination of the multispectral segmentation and spectral differences algorithm. The multispectral segmentation algorithm was first used to generate objects, while the spectral difference algorithm was used to eliminate small objects by combining spectrally similar objects into one. The parameters used are outlined in Table 1.

Table 4.1: Parameters used for object generation.

Multiresolution Segmentation	Parameters
Image band weights	Blue = 1; Green = 1; Red = 5; NIR = 5
Scale parameter	120
Shape	0.1
Compactness	0.8

4.2 Rerun Random Forest using most significant variables

The Viru Valley script uses the most significant 100 variables to rerun the Random Forest Model. The number of variables varies per project and the final number is determined by the levelling of overall accuracy i.e. the inclusion of more variables will not have a significant effect on the accuracy of the final map. For example, only 40 significant variables were used in the example provided in the LMM report. The final list of variables is available in Table 4.2.

Table 4.2: Final list of variables used in the mapping and their significance values as determined by the random forest algorithm.

Variable name	Significance Value
Height	14.508365
Slope	14.4282998
Topographic Wetness Index	13.4462596
Annual Rainfall	11.24304918
Linear Spectral Unmixing – June - Photosynthetic Vegetation	11.11098651
Sentinel-2 – January – NDVI	10.3573368
Sentinel-2 – June – NDVI	9.505599504
Sentinel-2 – January – NDWI (Green)	9.407039183
Sentinel-2 – June – NDWI (SWIR)	9.013546553
Sentinel-2 – January – NDWI (SWIR)	8.633625807
Difference Index – Sentinel-2 – June – Green Band – SWIR1 Band	7.722319946
Sentinel-2 – November – Blue Band – Median	7.54919104
Sentinel-2 – November – NDWI (Green)	7.544144683
Sentinel-2 – November – Blue Band – Mean	7.458022925
Sentinel-2 – January – Red Edge [8a] Band – Mean	7.252278879
Linear Spectral Unmixing – November - Photosynthetic Vegetation	6.923945885
Minimum Annual Temperature	6.808263892
Difference Index – Sentinel-2 – November – Green Band – SWIR Band	6.802561197
Sentinel-2 – January – Blue Band – Standard Deviation	6.794954515
Sentinel-2 – January – Red Edge [8a] Band – Median	6.650129808
Sentinel-2 – January – Red Band – Standard Deviation	6.638592631
Difference Index – Sentinel-2 – January – Green Band – SWIR1 Band	6.59597065
Sentinel-1 – January – Backscatter – VV – Mean	6.460818662
Difference Index – Sentinel-2 – November – Blue Band – Red Band	6.383894411
Sentinel-2 – January – Green Band – Standard Deviation	6.379678813

Sentinel-2 – June – NDWI (Green)	6.342999089
Sentinel-2 – November – NDWI (SWIR)	6.259285302
Sentinel-2 – November – NDVI	6.161794386
Sentinel-2 – November – Green Band – Median	5.955621831
Sentinel-2 – November Red Band – Standard Deviation	5.839130322
Difference Index – Sentinel-2 – June Red Edge [6] Band – Red Edge [7] Band	5.562267471
Difference Index – Sentinel-2 – November – Red Band – Red Edge [5] Band	5.550925833
Sentinel-2 – June – Blue Band – Mean	5.542571443
Sentinel-2 – November – Green Band – Mean	5.535249546
Difference Index – Sentinel-2 – January – Red Edge [6] Band – Red Edge [7] Band	5.516922826
Sentinel-2 – November – Blue Band – Standard Deviation	5.472885541
Sentinel-2 – November – Red Edge [8a] Band – Mean	5.470600147
Difference Index – Sentinel-2 – November – Red Band – Red Edge [5] Band	5.419546232
Difference Index – Sentinel-2 – January – Red Edge [6] Band – Red Edge [7] Band	5.415443497
Sentinel-2 – January – NIR Band – Mean	5.385597116
Sentinel-2 – January – Red Edge [7] Band – Median	5.3691745
Sentinel-1 – November – Backscatter – VH – Median	5.353599415
Sentinel-2 – January – Blue Band – Median	5.33917009
Sentinel-1 – January – Backscatter – VH – Mean	5.32792712
Difference Index – Sentinel-2 – November – Red Band – SWIR2 Band	5.265643388
Sentinel-2 – November – NIR Band – Mean	5.2586895
Sentinel-2 – November – Red Band – Median	5.236769019
Sentinel-1 – June – Backscatter – VV – Median	5.233798684
Sentinel-1 – November – Backscatter – VV – Median	5.186841387
Sentinel-1 – January – Backscatter – VV – Median	5.109189285
Sentinel-1 – June – Backscatter – VH – Median	5.107040478
Sentinel-2 – June – Green Band – Mean	5.090522144
Linear Spectral Unmixing – November - Non-Photosynthetic Vegetation	5.026433662
Sentinel-2 – June – Blue Band – Median	5.000160511
Linear Spectral Unmixing – November - Water/Shade	4.974094374
Difference Index – Sentinel-2 – June – Red Edge [6] Band – Red Edge [7] Band	4.97333849
Sentinel-1 – January – Backscatter – VH – Median	4.962866071
Sentinel-2 – November – NIR – Standard Deviation	4.950790771
Sentinel-2 – November – Red Edge [7] Band – Mean	4.946051003
Sentinel-2 – January – Red Edge [7] Band – Mean	4.931888672
Sentinel-2 – January – NIR Band – Median	4.92616703
Sentinel-2 – November – NIR Band – Median	4.912698035
Sentinel-2 – January – Red Edge [6] Band – Median	4.903687983
Sentinel-2 – November – Red Edge [7] – Median	4.891308649
Difference Index – Sentinel-2 – January – Red Band – Red Edge [5] Band	4.890170944
Difference Index – Sentinel-2 – January – Red Band – Blue Band – Standard Deviation	4.882346073
Sentinel-2 – January – Blue Band – Mean	4.845987779
Sentinel-2 – November – Red Edge [8a] Band – Median	4.784024082

Difference Index – Sentinel-2 – January – Blue Band – Red Band	4.78039349
Difference Index – Sentinel-2 – November – Blue Band – Green Band	4.774076138
Sentinel-2 – November – Red Edge [6] Band – Mean	4.743808029
Sentinel-2 – November – Red Band – Mean	4.731707427
Sentinel-1 – January – Backscatter – VH – Standard Deviation	4.710746445
Difference Index – Sentinel-2 – November – Red Band – SWIR2 Band	4.654216336
Sentinel-2 – June – Green Band – Median	4.604973846
Difference Index – Sentinel-2 – January – NIR Band – Sentinel-2 June – Red Edge [7] Band – Standard Deviation	4.593414692
Sentinel-1 – June – Backscatter – VV – Mean	4.567169203
Sentinel-2 – November – Red Edge [5] Band – Median	4.5493638
Sentinel-2 – January – Red Edge [6] Band – Mean	4.544814896
Sentinel-2 – November – SWIR2 Band – Mean	4.537863319
Sentinel-1 – November – Backscatter – VV – Mean	4.40890232
Sentinel-2 – November – Green Band – Standard Deviation	4.404754025
Sentinel-2 – January – Green Band – Median	4.337397356
Sentinel-2 – November – SWIR2 Band – Median	4.309974897
Sentinel-1 – January – Backscatter – VV – Standard Deviation	4.278797877
Sentinel-2 – June Blue Band – Standard Deviation	4.240859833
Difference Index – Sentinel-2 – January – Red Band – Red Edge [5] Band	4.229966299
Difference Index – Sentinel-2 – January – Red Band – SWIR2 Band	4.223776046
Sentinel-2 – June – NIR Band – Standard Deviation	4.160549921
Sentinel-2 – November – SWIR1 Band – Standard Deviation	4.150151163
Difference Index – Sentinel-2 – June – Red Band – Red Edge [5] Band	4.138624888
Sentinel-1 – November – Backscatter – VH – Standard Deviation	4.074244562
Difference Index – Sentinel-2 – January – Red Edge [6] Band – Red Edge [7] Band	4.068290203
Difference Index – Sentinel-2 – November – Red Band – SWIR1 Band	4.02805603
Sentinel-2 – January – Red Edge [5] Band – Standard Deviation	4.005571548
Sentinel-2 – November – Red Edge [5] Band – Standard Deviation	3.980220586
Sentinel-2 – November – Red Edge [6] Band – Median	3.928022443
Sentinel-1 – November – Backscatter – VH – Mean	3.868710355
Sentinel-2 – June – Green Band – Standard Deviation	3.628683295
Difference Index – Sentinel-2 – January – Red Band – SWIR1 Band	3.5056466

Annex 2

Feature extraction using VHR optical data

1. Datasets

A summary of each of the datasets used for input to the feature extraction is listed below. All data were georeferenced to WGS 84 / UTM zone 17S (EPSG:32717).

1.1 Pleiades

Pleiades is an optical observation system consisting of two identical satellites that deliver 2m multispectral (Red, Green, Blue and NIR) products. A pan-sharpening band is also available, that can provide 50cm colour products.

As stated in the report, Pleiades data were purchased for the project and pre-processed by JNCC. The topographical distortions were accounted for using the Orthorectification tool in Orfeo ToolBox 6.6.1 (https://www.orfeo-toolbox.org/CookBook/Applications/app_OrthoRectification.html). The application uses inverse sensor modelling with a user-defined interpolation function to resample the sensor data to a geometrical grid, in this instance the WorldDEM™. The method was implemented through the command line using the bicubic interpolation method with RPC modelling enabled.

2. Method for delineating features

In order to enhance the delineation of features, several indices and products were derived using the multispectral Pleiades imagery:

- NDWI2: Normalized difference water index 2 (McFeeters, 1996) (Green, NIR)
- SAVI: Soil adjusted vegetation index (Red, NIR)
- Energy (from NDWI2)
- Haralick's Correlation (from NDWI2)
- Haralick's "Sum of Squares: Variance" (from NDWI2)
- Haralick's Correlation (from SAVI)
- Haralick's "Sum of Squares: Variance" (from SAVI)

NDWI2 was chosen due to the lack of MIR bands in Pleiades imagery and SAVI was used to account for the climate and geomorphology of the Viru Valley. When compared against NDVI, SAVI delineated features to a greater extent. The above Haralick's features are generated from the Feature Extraction tool accessed through the command line in Orfeo ToolBox 6.6.1 (<https://www.orfeo-toolbox.org/CookBook/recipes/featextract.html>). An example of one of these outputs can be seen in Figure 2.1. They represent textures in the image channel (in this instance NDWI2 and SAVI) and are generated using a Grey Level Co-occurrence Matrix (GLCM).

Once the Pleiades imagery was prepared, an 8 band GeoTIFF was created using the original 4 bands and the NDWI2 derived data. This was segmented using the Mean-Shift algorithm in Orfeo ToolBox 6.6.1 (https://www.orfeo-toolbox.org/CookBook/Applications/app_Segmentation.html).

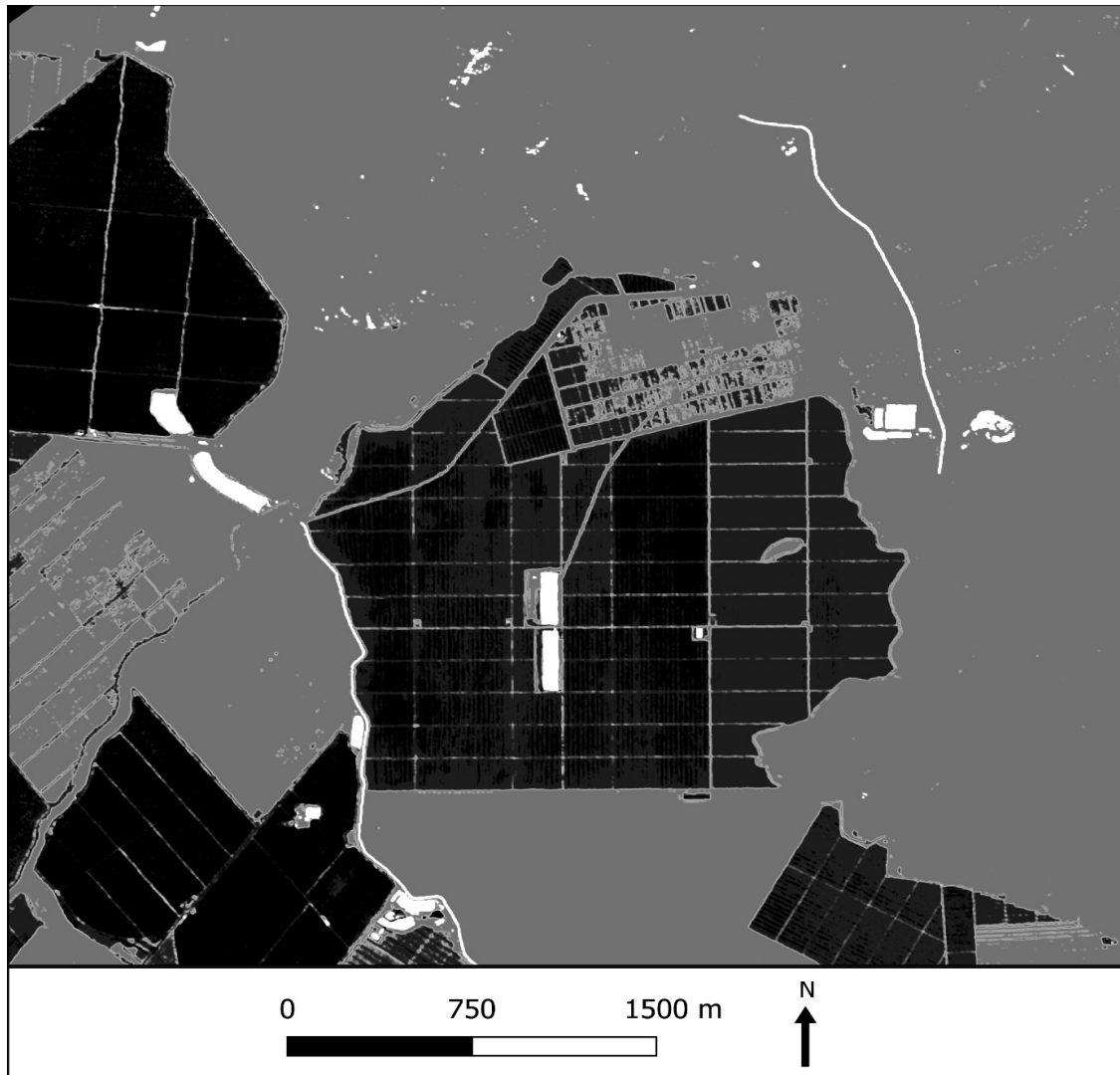


Figure 2.1: Example region of the Haralick's Correlation output showing the texture features identified. Bare soil/scree is shown in grey and regions of water are defined in white.

This algorithm combines spectral and spatial differences with the texture features derived considered. A spatial radius of 30 m was implemented. The resulting segmentation clearly defined areas of water. Statistics were run on the segmentation, extracting the mean values of each band. A threshold was performed, and the outputs defined as water regions.

A similar process was applied to define the linear vegetated areas, however issues occurred with the segmentation and thresholding process due to linear vegetated areas of interest not being spectrally different from the other vegetated areas. This was a challenging aspect of the feature classification due to the nature of the landscape and the distribution of the linear vegetated features. In order to attempt to overcome this, a threshold was performed on the two Haralick's outputs. The resultant regions were grouped based on the Haralick's outputs and region size, and the outputs are included in the final product.

Annex 3

Additional work after completion of map

Following the production of the first habitat map it became apparent that some minor revisions were required to a small number of classes. A further two additional classes were also identified to improve ecosystem service mapping outputs. Subsequent revisions are described below.

1. *Acacia tree belts – new class*

Acacia tree belts were defined as linear vegetated features running alongside roads. These were identified visually from Pleiades VHR data aided by an Open Street Maps road layer. Where a tree belt was located on one or both sides of the road, it was manually changed to the new *Acacia tree belt* habitat class. The previous class is recorded in the Class_Befo field of the attribute table. It appears there are more roads present in the area of interest than the Open Street Maps layer shows. Only those roads that could be identified with confidence were assessed for the addition of acacia tree belt.

2. *Shelter belts – new class*

Shelter belts were defined as linear vegetated features that are found within or alongside fields. These were identified visually from Pleiades VHR data, with the aid of linear features picked out in Habitat Map Version 1 (V1). Where Habitat Map V1 did not distinguish a linear feature, even where this could be seen using the visual data, no edits were made. For example, a field of 'Permanent croplands' that appeared dissected by a line of trees on the VHR data but classified as 'Permanent croplands' for the entire field. Where Habitat Map V1 did distinguish a linear feature and classified it separately from the background habitat (for example a line of 'Croplands – trees and bushes – dense' within a field of 'Permanent croplands'), the line was changed to 'Shelter belts.'

Shelter belts were easily identified within large commercial growing areas, but nearer the Viru river the heterogeneity of the land use matrix made it harder to identify shelter belts from mixed scrub; this includes most of the areas managed by small holders. There are too many features that could be potentially classified as shelter belts within the area of interest to identify them all manually within the time available; due to the class being added in later to facilitate ecosystem service mapping. Those that were most easily visually identified with a greater extent of coverage were prioritised over those that were harder to identify. More effort applied to distinguishing this class would likely lead to greater extent and accuracy of this class.

The previous class is recorded in the Class_Befo field of the attribute table. Where any uncertainty was involved surrounding whether the feature was a shelter belt or not, the original habitat class from Habitat Map V1 was retained.

3. Wetlands

There was some confusion between wetlands, described as naturally forming land consisting of marshes, swamps or any other form of saturated land, and reservoirs. There were also some edge artefacts surrounding the boundary of the AOI, where objects were incorrectly classified as wetland. As this class is seen as a priority for ecosystem service mapping, these objects were manually changed to remove outliers. The class changes were identified using Pleiades VHR data and Google Earth imagery.

4. Bare Sand – coastal – beach

Some areas inland were being identified as *beach* when they should have been classed as '*bare sand*'. These were manually changed based on interpretation of Pleiades VHR and Google Earth imagery.

5. Croplands – trees & scrub/bushes – dense

This class refers to areas surrounding field boundaries that include trees and scrub bushes in a dense format. These areas were overclassified, particularly in areas with smaller fields of crops, where the crop type have similar characteristics to trees and scrub. As this class is also identified as important for ecosystem service mapping, the largest outliers were identified and manually changed based on visual interpretation of Pleiades VHR data and Google Earth imagery.

Summary

In total, 0.78% of the objects were changed manually from the first version of the habitat map. None of the field data collected for the project overlapped with any of the changed polygons so overall accuracy of the map does not change. However, the quality of the map has improved for the purpose of ecosystem service mapping. All changes were recorded in an additional column provided in the attributes of the final habitat map product.

References

KILCOYNE, A.M., ALEXANDER, R., COX, P. & BROWNETT, J. (2017). Living Maps: Satellite based Habitat Classification. Evidence Project SD1705

McFEETERS, S. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. International Journal of Remote Sensing, 17(7), pp.1425-1432.

WORLD BANK (2017). Gaining Momentum in Peruvian Agriculture: Opportunities to Increase Productivity and Enhance Competitiveness. Available at: <http://documents.worldbank.org/curated/en/107451498513689693/pdf/P162084-06-26-2017-1498513685623.pdf> [Accessed 6 February 2018]