BIO_SOS

Project Title:	**BIO_SOS Biodiversity Multisource Monitoring System: from Space TO Species**
Contract No:	FP7-SPA-2010-1-263435
Instrument:	Collaborative project
Thematic Priority:	
Start of project:	1 December 2010
Duration:	36 months

Deliverable No: D5.5

Report on RS-IUS second–stage modules software description

Due date of deliverable:	30 November 2012
Actual submission date:	12 December 2012
Version:	1st version of D5.5
Main Authors:	Richard Lucas and Gwawr Jones (Aberystwyth University), Vas iliki Kosmidou and Ziss Petrou (CERTH), Jordi Inglada (CESBIO), Maria Patrizia Adamo and Palma Blonda (CNR), Damien Arvor (IRD)

Project ref. number	**263435**
Project title	**BIO_SOS: Biodiversity Multisource Monitoring System: from Space to Species**

Deliverable title	Report on RS_IUS second-stage modules software description.
Deliverable number	D5.5
Deliverable version	Version 1
Previous version(s)	
Contractual date of delivery	30 November 2012
Actual date of delivery	12th December 2012
Deliverable filename	BIO_SOS_D5.5_RS_IUS_second-stage modules_v1
Nature of deliverable	R = **Report**
Dissemination level	PU = Public
Number of pages	71
Workpackage	WP5
Partner responsible	P12
Author(s)	Richard Lucas and Gwawr Jones (Aberystwyth University), Vasiliki Kosmidou and Ziss Petrou (CERTH), Jordi Inglada (CESBIO), Sander Mucher (ALTERRA), Maria Patrizia Adamo and Palma Blonda (CNR), Damien Arvor (IRD)
Editor	Richard Lucas
EC Project Officer	Florence Beroud

COOPERATION

Abstract	D5.5 describes the EODHaM 2nd stage and the translation of the EODHaM 1st, 2nd and 3rd stages to an open source environment and the sequence of processing, including feature extraction and segmentation, classification of land covers and translation to GHCs. Texture and Dempster-Shafer uncertainty analysis are also outlined. Land cover maps are presented for BIO_SOS test sites with estimates of accuracy.
Keywords	EODHaM system, pre-processing, segmentation, feature extraction, classification, land cover, habitat

Signatures

Written by	**Responsibility- Company**	**Date**	**Signature**
Richard Lucas	EHODaM System development responsible (Abery)	30/11/2012	
Verified by			
Laurent Durieux	WP5 leader up to 30/11/2012 (IRD)	12/11/2012	
Approved by			
Palma Blonda	Project Coordinator, CNR	12/11/2012	
Fifamè Koudogbo	Quality Team, AI	11/11/2012	

Table of Contents

1 Executive summary

In previous WP5 deliverables, the pre-processing and segmentation of very high resolution (VHR) satellite sensor optical and, where available, airborne LiDAR data and subsequent classification of objects into FAO LCCS (Land Cover Classification Scheme) classes was undertaken within commercial software using a rule-based classification. In D5.5, focus has been on translating the developed rules and procedures (image pre-processing, segmentation, classification and accuracy assessment) into open source software in such a way that the EO Data for Habitat Monitoring (EODHaM) system can be more easily implemented by end-users and within user organisations.
The open source software used also facilitates automated translation of LCCS classes to General Habitat Categories (GHCs).

The pre-processing is crucial and involves radiometric and correction to Top of Atmosphere (TOA) reflectance, with capability for surface reflectance retrieval and topographic correction. Two primary layers representing 'large' and 'small' objects are generated from multi-temporal imagery, with these ideally acquired during periods of peak flush and senescence (i.e., winter/summer or wet/dry). Each layer object is then associated with a raster attribute table, the columns of which are populated with spectral and derived data (e.g., vegetation indices, object length/width ratio, area etc.) with these then used in the subsequent classifications. Within the large objects, broad features of the landscape are classified, including agricultural and forestry units of varying size and both urban and water features. Within the small object layer, a range of independent segmentations is implemented to specifically detect shape features including those that are linear (e.g., ditches, hedgerows, roads), rectangular (e.g., permanent and temporary dwellings) or circular/elliptical (e.g., individual tree crowns), with these classified using spectral and contextual rules. Remaining objects in the landscape are then generated using a generic segmentation based only on spectral information. Where appropriate, ancillary data layers (e.g., cadastral information) are integrated. Information contained within the large and small object layers is then combined to describe more complex landscape components (e.g., olive groves, caravan parks, field units). A rule-based classification (as described in the D5.3 Report) is then applied to over 30 separate data columns which describe, according to the LCCS concept, the various land cover characteristics associated with lifeform (e.g., cover, height, leaf type, phenology) and non-lifeform (e.g., urban density, water dynamics, surface aspect) categories, as defined by the LCCS. The LCCS codes (e.g., A1 for woody, A2 for herbaceous) generated for each layer are then combined to identify the final LCCS category at both Levels 1-3 and 4, with the latter automatically translated to a GHC category. Assessments of classification accuracy can be generated for each layer and also for the final output products (i.e., the LCCS and GHC maps). For the primary test sites of Wales, Italy, the Netherlands and Portugal, LCCS and GHC classifications were generated with assessments of accuracy for the GHC categories. The mapping approach, which is now relatively flexible and rapid to implement because of the consistency of rules used, is being applied to the test sites in Greece, Brazil and India and modified to consider change (to be delivered in D5.6).

The research undertaken within D5.5 has lead to the development of the EODHaM system as a comprehensive open-source and consistent tool for land cover and habitat mapping from VHR imagery, which can be applied anywhere worldwide given appropriate data. The system benefits from the inclusion of LiDAR data, which facilitates improved discrimination of lifeform and some non-form categories. Height information of the vegetation is crucial in the discrimination of the LCCS and GHC categories and can be directly derived from LiDAR data although, if necessary, indirectly from spectral information in the VHR optical imagery. Future work in WP5 is focusing on refining components of the system (and further operationalization for end-users) and integrating a change detection module, with both to be reported in the final deliverable for WP5, D5.6.

2 Introduction

In D5.3, the approach to the classification of Food and Agricultural Organisation (FAO) Land Cover Classification Scheme (LCCS) categories from very high resolution (VHR) Worldview and Quickbird data was outlined, with this forming the basis of the EODHaM system (Figure 1). In the EODHaM 1st stage, the classification to LCCS Levels 1 and 2 was performed using rules based on spectral data and derived indices alone, with the main focus being the separation of vegetated and non-vegetated areas and aquatic and terrestrial environments. The EODHaM 2nd stage integrated both spectral and contextual features and focused first on the classification of LCCS Level 3, which established whether the landscape was either cultivated, managed or artificial or natural or semi-natural. Classes beyond Level 3 were then classified by first generating over 30 separate layers relating to key descriptors within the LCCS classification and assigning codes within these. These codes were then combined to generate a summary class code and an associated descriptor (e.g., permanently cropped area: Graminoid crops). In EODHaM 3rd stage, the translation of the LCCS categories to General Habitat Categories (GHCs) was performed. Particular benefits of the process were that a) the classification could be performed regardless of scale and b) a globally recognised land cover classification scheme was being applied.

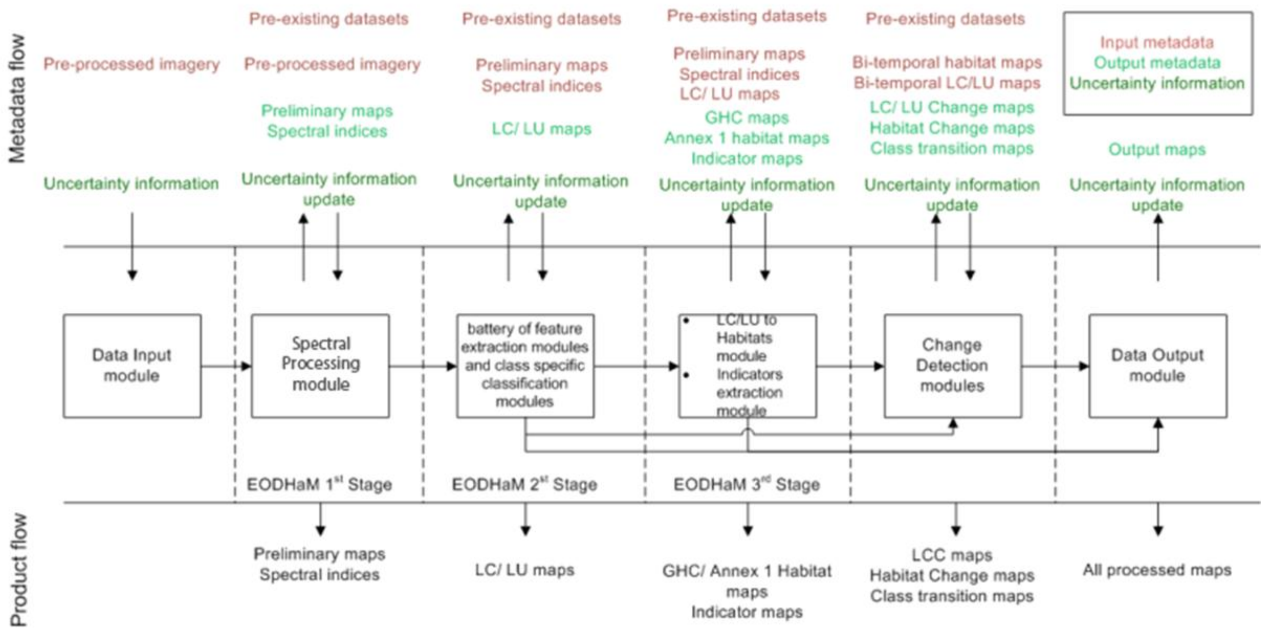


Figure 1. Overview of the EODHaM System

The entire process was conducted within the framework of commercial object-orientated software and LCCS maps were subsequently generated for Le Cesine (Italy), Cors Fochno (Wales) and Veluwe (the Netherlands). However, to increase efficiencies and transparency and provide a low cost solution, attention within D5.5 focused on translating the approach to open source code, which could then be incorporated readily within the EODHaM system (as described in D3.2). The following sections therefore provide an overview of advances in implementation and the sequence of processing. Maps of LCCS classes (beyond Level 3) are then illustrated with preliminary estimates of classification accuracy provided, noting that further field validation will take place in the spring/summer of 2013.

3 EODHaM second stage modules software description

3.1 Advances in implementation

In D5.3, the approach to segmenting landscapes and classifying LCCS categories to Level 3 was demonstrated using examples from Cors Fochno (Wales), Le Cesine (Italy) and Veluwe (the Netherlands). Mapping beyond Level 3 (e.g., life form, physical status and surface aspect) was demonstrated for Cors Fochno, by considering the key building blocks of the LCCS classification (e.g., woody/herbaceous, height, cover, leaf type and phenology for lifeform categories) separately and combining the codes to generate the final LCCS codes. Throughout, rules based on reflectance data and derived indices (primarily those quantifying relative amounts of photosynthetic and non-photosynthetic vegetation and moisture), textural features and digital elevation models (DEMs) and derived products (e.g., terrain slope and aspect and vegetation height) were used. The rules were developed by combining the knowledge of ecologists and remote sensing scientists.

The segmentation and classification software eCognition provided an invaluable environment in which EODHaM Stages 1 and 2 were developed. However, several limitations to the use of this software were identified.

a) Whilst eCognition provided capability in processing large datasets (based on tiled input and subsequent merging of classification outputs), the speed of processing was often limited within the partner organisations because of limiting computing power but also lack of availability of engine server licenses.

b) The rules for classification were typically developed on small subsets of data but rules that performed well in these subsets sometimes generated errors in other subsets where new classes were encountered. Working on subsets was considered to be less effective since it was not evident that thresholds applied led to confusion with other classes in the surrounding area until the final classification had been generated.

c) Within eCognition, over 30 key LCCS discriminative layers were classified separately and then the codes of each layer were combined into a final LCCS code (concatenation of LCCS codes). However, this approach was difficult to automate and was compromised when a large number of LCCS classes occurred (e.g., representing combinations of height and cover for woody lifeforms).

For these reasons, an alternative route was sought within an open source environment. Python programming language with the RIOS library developed by Gillingham and Flood (2011), as well as RSGISLib codes developed by Bunting et al. (2012) in combination with the ORFEO Toolbox developed by Inglada et al. (2012) were selected as the most powerful combination of open source software. The key features of this software suite that advanced the implementation of EODHAM 1st and 2nd stages were:

a) The development of remote sensing and GIS processing capability within RSGISLib (Bunting et al, 2012). and LiDAR processing within SPDLib (Bunting et al, 2012a,b), undertaken at Aberystwyth University.

b) The generation of raster attribute tables, such that each object (ranging in size from an individual pixel to an entire image) could be linked with values listed in columns.

c) The development of a new image format (KEA) developed in New Zealand by Bunting and Gillingham (2012).

d) Capability for viewing raster data and querying attributes within the viewer, again developed by Gillingham et al., (2012).

e) Development of image segmentation (Shepherd et al., 2012) and feature extraction procedures (Inglada et al., 2012).

f) The Raster I/O Simplification (RIOS) python library (Gillingham and Flood 2011) provides a simple to use python interface using numpy to the raster attribute table columns.

The combination of these processing steps provided a new framework in which the EODHaM 1st, 2nd and (elements of) 3rd stages were implemented within an open source environment. The primary benefits were:

a) The segmentation of the imagery, classification of LCCS categories and translation to General Habitat Categories (GHCs) could be performed on an entire image with no requirement for tiling or development of rules on subsets.

b) The code for classification (which was developed within Python) was easier to understand and to amend. In the majority of cases, only changes in a parameter card defining values (e.g., thresholds for reflectance data and indices) were needed, with the exception being areas where new land covers were encountered.

c) Accessibility of the software to all users in a faster and more flexible environment and a better understanding of the decision rules in each step of the identification of the final classes.

3.2 Overview of procedures

The sequence of modules and associated procedures relevant to D5.5 is outlined in Figure 2. The ultimate objective of the processing was to generate a classification of LCCS classes and GHCs from a raster attribute table, the format of which is represented diagrammatically in Figure 3. Typically, the raster attribute table consists of reflectance bands of data acquired ideally in the pre-, peak- and post- flush periods and derived indices. Additional columns describe the characteristics of small and large objects within the landscape as well as information obtained from ancillary information (e.g., terrain slope and aspect). From these column data, rules are applied to generate the codes representing LCCS land cover categories 1-3 and also beyond and then to translate these to a GHC habitat code. More detailed information on these procedures is listed in the following sections.

3.2.1 Pre-processing

The primary pre-processing steps are radiometric calibration and correction to Top of Atmosphere (TOA) reflectance. The option to correct to surface reflectance using the 6S code is available within ORFEO toolbox and RSGISLib, with local or regional parameterisation achievable where other data (e.g., MODIS-derived optical depth) are available. The data can also be topographically corrected using the procedures outlined by Shepherd and Dymond (2003). Orthorectification can be undertaken within ORFEO toolbox or commercial software such as ERDAS imagine, although this component is considered to be the most time-consuming as manual selection of ground control points is often necessary. Ideally, all data should be delivered by commercial provides georeferenced within < 1 pixel error as this would avoid the requirement for this step. For orthorectification, an accurate Digital Elevation Model (DEM) of appropriate spatial resolution is desirable if not essential.

3.2.2 Conversion to kea format

The KEA format as developed in New Zealand by Bunting and Gillingham (2012) allows for the inclusion of raster attribute information, as indicated in Figure 3. Each layer, which is associated with projection (e.g., .WKT) information, is initially converted to a kea file, and where required is linked with a raster attribute table. The KEA format also includes the use of compression (zlib) throughout with no file size limitations, minimising the storage requirements.

3.2.3 Generation of pixel-based layers

Several layers in the LCCS scheme require, for their generation, classification at the pixel level. These include measures of vegetation cover, phenology and surface type (e.g., gravels, stones

etc.) based on a range of spectral indices as well as texture (e.g., based on the measures of Haralick; 1980).

DATA INPUT MODULE

Module 1: Data preparation

Module 2: Preprocessing

Orthorectification, radiometric, atmospheric & topographic correction

Image format conversion

EODHAM 1ST STAGE

Module 3: Calculation of derived products

Spectral Indices, textural features, pixel-based layers

Module 4: Segmentation

Generate small objects

Generate large objects

Calculate segment metrics (e.g., shape)

Build and populate statistics and colour tables

Module 5: Preliminary spectral map extraction

Run classification of LCCS Levels 1 and 2

EODHAM 2ND STAGE

Module 6: LC/LU final classification

Extract context features

Run classification of LCCS Level 3 and beyond

Dempster-Shafer fuzzy classifications

Apply standardized colour scheme

Module 7: Land Cover/Use Change Detection

EODHAM 3rd STAGE

Module 8: LC to GHC Translation

Module 9: GHC to Annex 1 Translation

Module 10: LC to Annex 1 translation

Module 11: Habitat change detection

Module 12: Accuracy and uncertainty assessment

Module 13: Ecological modelling at landscape level

Module 14: Biodiversity indicators extraction

Module 15: Biodiversity indicators change detection

Figure 2. Modules within the comprehensive EODHAM system

3.2.4 Segmentation and feature extraction

Two essential steps in the EODHaM system are a) extraction of discrete features in the landscape and b) segmentation of the remaining areas to divide the landscape into units that are spectrally homogeneous. In the case of a), algorithms have been developed or enhanced to allow routine

detection of, for example, tree crowns, buildings, hedgerows and fields. For b), algorithms based on clustering and object elimination (Shepherd et al., 2012) and the mean shift (Comaniciu and Meer, 2002) have been implemented (see D5.3). Using both sets of procedures, two layers are generated with these representing small and large objects, which are then described on the basis of within-object and between object-characteristics. The segmentation procedure is based entirely on spectral rules (and hence is a component of the EODHaM 1st stage), although it enables the integration of existing ancillary data layers such as those representing field boundaries or urban infrastructure (e.g. cadastral data).

Obj. ID	Pre 1 … n	Peak 1 … n	Post 1…n	Small Obj. Desc 1…n	Large Obj. Desc. 1…n	Ancillary 1…n	LCCS L1-3	Layer code 1..30	LCCS code	GHC code
1 . . . n										

Figure 3. The format of the raster attribute tables used for the classification of LCCS classes and GHCs. Values 1..n represent the reflectance bands and derived indices, object descriptors and ancillary data.

Generation of small objects: Small objects refer to those that are discrete within the landscape and which can be delineated on the basis of spectral rules alone. These include objects corresponding to individual tree crowns, buildings and water bodies. These objects are those that are used for the classification of LCCS Levels 1-2 to provide as output semantic strata: A1_ terrestrial_vegetated,A2_Acquatic_vegetated,B1_Terrestrial_Non_Vegetated, B2_Acquatic_Non_Vegetated (see D5.3)

Generation of large objects*:* Large objects are first extracted in the image segmentation phase (the mean-shift algorithm is used), which is purely spectral (EODHaM 1st stage) and then identified (labelled) after LCCS level 3 through expert knowledge rules and context-sensitive features (EODHaM 2nd stage). Large objects are defined as those that:

a) Represent expansive objects within the image such as individual agricultural fields, large water bodies, and forestry plantations. These can be described from knowledge of the characteristics of the small objects contained. As an example, olive groves are comprised of a large unit containing numerous trees with crowns often or similar size and orientation.

b) Are combined from knowledge of the number and characteristics of smaller objects. An example is tourist parks that are represented by individual caravans or tents, buildings, minor roads, playgrounds and amenity grasslands.

Calculating segment metrics: Each object that is defined within the separated layers of 'small' and 'large' objects is populated with statistics from the imagery and ancillary layers and associated with descriptors, with these inserted as columns in the raster attribute table. Within-object descriptors include:

a) Length/width ratio
b) Area
c) Perimeter length
d) Shape (e.g., roundness, rectangularness, elongation, elliptical fit).

Between-object descriptors relate to neighbours that are immediate or some distance away and include:

a) Within object descriptors (listed above)
b) Class
c) Distance to/from
d) Enclosure

Once defined, these descriptors can be used to enhance the classification of objects, either within the ‘small’ or ‘large’ object layers (with these then added as additional column attributes), or the LCCS categories themselves.

3.2.5 Classification of LCCS Levels 1 and 2

The EODHAM 1st stage uses spectral information only to classify vegetated and non-vegetated and terrestrial or aquatic land covers, with a standardised suite of indices describing vegetative state (e.g., photosynthetic or non-photosynthetic) and relative moisture content. Key indices (listed in D5.3) include the Normalised Difference Vegetation Index (NDVI), Greenness Index, Forest Discrimination Index (FDI), the Plant Senescence Reflectance Index (PSRI) and the Water Band Index (normalised or otherwise). The classification is performed using a relatively simple ruleset that has proved to be consistent between sites and provide classifications of the Level 1-2 classes to acceptable levels of accuracy (typically > 80 % for users’, producers’ and overall). The rules are applied within the raster attribute table, resulting in the generation of new columns with class codes.

3.2.6 Classification of LCCS Level 3

To classify the LCCS Level 3 classes that relate to whether the landscape is managed, cultivated, artificial or semi-natural/natural, information contained within the large and small object layers is used. For example, fields comprised of single crops or grassland are classified based on spectral information alone whilst those associated with plantations or tree crops make reference to the characteristics of the smaller objects contained or are classified by considering the groups of smaller objects that form the larger objects (industrial areas are an example, with these comprised of large buildings, roads and amenity grasslands). The system has the option to bring in cadastral and infrastructure layers that have been independently generated.

3.2.7 Beyond Level 3: Assignment of codes to LCCS layers

To classify beyond Level 4, the EODHAM 2nd stage uses spectral information but also class description provided by experts (ontologies) including context-sensitive relations between the small and large object layers to assign class-specific codes to column attributes representing the layers listed in Table 1, with the latter representing the building blocks of the LCCS classification. Phenological information (temporal relations) is also used.

Table 1. Main attribute columns populated for subsequent classification to LCCS Level 3 and beyond

Vegetation	Physical status	Artificial	Bare Areas	Cultivation
Lifeform	Water state[1]	Surface aspect	Surface aspect	Crop lifeform
Cover	Dynamics	Urban density	Materials	Field size
Height	Persistence		Macropattern	Water seasonality
Leaf type	Duration			Cultural
Phenology	Tidal/not tidal			Crop combinations
Stratification	Substrate material			
Distribution	Depth			
Floating/rooted	Sediment loads			

[1]Water, ice or snow

3.2.8 Combining LCCS codes

Once each column has been assigned a code, these are combined to form the full LCCS code by joining these in a string. Each code is also associated with a name, which is similarly combined to provide a more interpretable description of the land cover being classified.

3.2.9 Application of an appropriate colour scheme

The LCCS scheme provides the opportunity to generate a very detailed classification of the landscape, which can compromise the interpretability of the map. This is particularly the case for woody vegetation where a large number of height and cover combinations result in a proliferation of the LCCS codes. For this reason, a generalized colour scheme has been developed which is universal such that new legends do not need to be generated when new LCCS combinations are generated for a site. The colour scheme has been designed based on the attributes of the LCCS. In particular, each LCCS qualifier in the individual layers corresponds to an attribute of colouring that is applied sequentially and reflects the appearance of the land cover in the landscape. For example, greens are applied to vegetation, browns to bare ground, greys to artificial and blues to aquatic. Gradients within these then reflect, for example, the life form as a function of height (high to low saturation for short and tall vegetation respectively) whilst shading reflects vegetation phenological and leaf type attributes (high to medium and low hue for needle-leaved and broad-leaved, evergreen and deciduous or annual and perennial). In the same notion, a colour scheme for the presentation of the GHC classes has been developed. The standardized colour maps for the LCCS and the GHC are provided in Appendix I.

To reduce the number of colours, colour codes are not modified for information relating to, for example, field size (which can be deduced by the size of the coloured patch in the image) or spatial distribution (e.g., fragmented or continuous). To deal with the representation of patches with multiple GHC classes assigned to them, the colour corresponding to the mean of the RGB values of the unique GHCs has been used to form the colour of the combinations.

Translation to General Habitat Categories (GHCs)

As a component of the EODHaM 3rd stage, the LCCS land cover codes are automatically translated to GHC habitat codes (using semantics) with several options provided where one to many relationships occur.

3.2.10 Accuracy assessment

To assess accuracy as a component of the EODHaM 3rd stage, data on LCCS classes and GHCs collected in the field are compared against the classification outputs. A complication is that the LCCS codes generated from the mapping often far exceed those recorded in the field because of the greater ability to differentiate (e.g., different cover categories over areas of varying spatial extent). Hence, the approach to accuracy assessment is as follows:

a) Use available data to assess the accuracy in the classification of each layer (e.g., height, cover, leaf type) and subsequently provide a combined estimate of accuracy based on the accumulation of errors.

b) Use partial versions of the final LCCS categories (i.e., exclude field size, spatial distribution) and compare against ground data where common attributes have been recorded (e.g., lifeform, height, cover, leaf type, phenology, water sediment loads and surface material). From these, standard measures of classification accuracy can be applied, with consideration given to error tolerances, as outlined in D5.2.

3.2.11 Summary

The processing of remote sensing data and the mapping of LCCS categories and their subsequent translation to GHCs within the EODHAM 1st and 2nd stages has been transferred to an open source environment, thereby facilitating application across a range of platforms by a wide range of users at minimal cost. Within the EODHAM 1st stage, the generation of objects (through

feature extraction and segmentation) and subsequent classification to LCCS Levels 1-2 is based entirely on spectral information. Classes associated with LCCS Level 3 and beyond are classified using a combination of spectral information and context-sensitive features as small and large object descriptors, with the inclusion of existing ancillary information allowed. The classification is undertaken within a raster attribute table with many columns and the accuracy of classification is based on that of each discriminative attribute column (or through a manageable set of codes derived from these) that contributes to the final assignment of a land cover class. The classification approach also facilitates the detection of change based on a change in codes following reclassification of areas observed at a later date. The specific change in a code can also help to indicate the driver and direction of change associated with natural (e.g. vegetation succession) or human induced events and processes (e.g., clearance).

3.3 Key advances in EODHaM 2nd stage classification

The EODHaM system has been designed and demonstrated for selected sites, with the approach described in D5.1 to D5.5. Several components of the system are new and innovative including a) the design of a logical system for generating LCCS codes and combining these into classes using raster attribute tables, b) the development of new segmentation and feature extraction procedures and the application of Dempster-Shafer rules. The generation of LCCS classes has been described in detail in D5.3 and hence the following sections focus on providing more detail on the feature extraction and textural analysis procedures that have been implemented as well as the Dempster-Shafer procedures.

3.3.1 Feature extraction

Many objects within a landscape (e.g., buildings, trees) are difficult to discriminate using pixel-based information only. However, the extraction of such features is critical for the classification of LCCS classes at all levels, including in the EODHaM Stage I where their identification improves overall accuracies in the classification of, for example, bare ground versus urban areas or trees in cultivated areas versus those in semi-natural/natural situations. For these reasons, BIOSOS has focused on implementing region-growing segmentation techniques that allow the reliable extraction of buildings, hedgerows and tree crowns using spectral or textural information, with implementation within open source code.

For segmentation, the *connected* component technique (Grizonnet *et al*., 2012) is adopted, which is a region-growing algorithm that uses pixel-based information only. The starting point of the segmentation is a binary image, obtained usually by applying a simple threshold to an index (e.g., the Normalised Difference Vegetation Index or NDVI) to identify candidate regions where targets can be grown. An iterative procedure is then applied which adds connected pixels, with this relying on the so-called connected component formula. This can be either an absolute condition based on the pixel values (e.g., the green reflectance is greater than a specified threshold) or a relative condition with respect to adjacent pixels already including in the region (e.g., difference in green reflectance below a certain threshold). Complex formulae combining several of these conditions can also be used. The approach can be applied to multi-band images, which can also include computed values such as vegetation indices and textures. Once the region growing has completed, regions can be discarded on the basis of simple criteria such as size or shape or statistics within the region (mean and standard deviation of a band). This step allows discrimination of some surfaces (e.g., corresponding to buildings, bare soil). Examples of how the procedures are implemented are provided below.

3.3.1.1 Building extraction

In many VHR optical images, buildings and similar man-made objects are confused with bare ground, particularly if only spectral information is used. To avoid this, a combination of spectral and textural information in used in a simple object analysis to reliably detect built up objects. In EODHaM 2nd stage, the Haralick texture feature is computed on the blue band, as the contrast with vegetated areas is greater, and using a 3 x 3 window size. The initial mask for the region

growing was built by thresholding the entropy image (values > 2) and then all connected pixels with a blue reflectance > 20 % within this region added. Buildings are then identified as those objects belonging to LCCS B1 strata (i.e. primarily non-vegetated terrestrial) smaller than a given threshold (300 pixels) and with an elongation factor of < 4. An example of using this procedure for extracting individual caravans is given in Figure 4.

3.3.1.2 Hedgerow extraction

To extract hedgerows (Figure 5), two different approaches were adopted. In the first, texture features were not required but forest borders and other features (e.g., individual trees) were included. In the second, texture information was integrated to reduce the confusion. In both cases, the connected component segmentation was applied alongside simple mathematical morphology functions such as erosion and dilation. In the first case (Figure 5), mathematical morphological processing was applied whereby thin objects were retained by first removing these by erosion, dilating thick objects and then subtracting the original thin objects. A 3 x 3 structuring element was used with 4 iterations applied.

Figure 4. Examples of caravans extracted using the connected component technique, Cors Fochno. a) Worldview image, b) entropy image calculated using a 3 x 3 window and the blue reflectance channel and c) extracted objects.

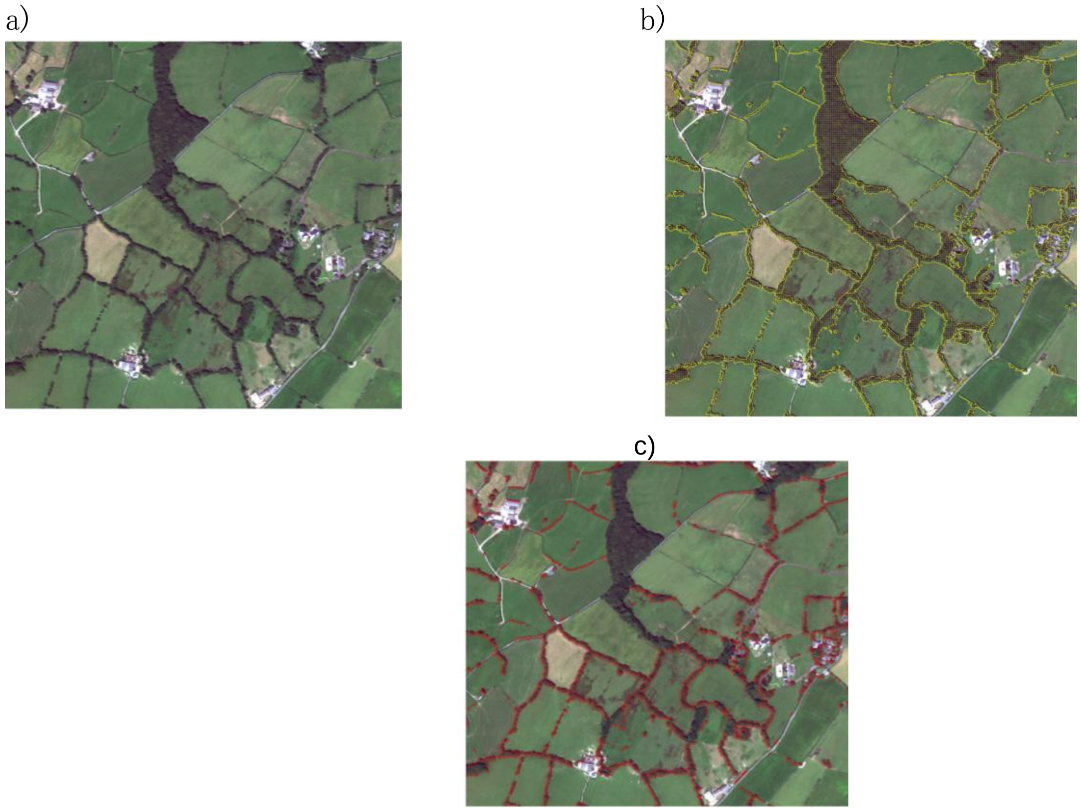


Figure 5. a) Worldview image of July, 2011, with the extent of woody vegetation mapped using a simple threshold (red reflectance between 8 and 11.8 %) and a connected component formula (near infrared > 15 % and the red/blue ratio of < 0.55, NIR standard deviation > 0.35). c) Extracted hedgerows.

In the second case, a mean shift smoothing (Comaniciu and Meer, 2002) was applied to the input image (Figures 6a and b) to remove small variations inside the region of interest and texture features were calculated subsequently. The parameters used were the spatial radius (10), spectral radius (10) and threshold (0.05) and 40 iterations were applied. The Haralick variance texture feature was then calculated (using a window size of 5 x 5, with 64 bins) for the green reflectance band (Figure 6c) to discriminate more expansive areas of, for example, woodland or buildings, with the variance value of the latter (Figure 6d) typically being lower than that of the linear and thin objects (i.e., the hedgerows). Using the connected component segmentation, areas of woody vegetation were extracted, as described in the previous subsection. The woody mask was obtained by thresholding green reflectance values between 8.1 % and 11.5 % and the connected component formula was based on a NIR reflectance > 15 % and a red/blue ratio of < 0.55. The OBIA expression was based on a standard deviation of the variance feature of > 0.8. The final classification is presented in Figure 6e.

For woody vegetation (e.g., hedgerows, small forests, olive groves), the variance of the mean value is also low and the use of erosion and dilation of the objects (based on a 3 x 3 window, 4 iterations) allows elongated features to be better discriminated (Figure 7a). Again, the connected component approach can be applied, with the threshold used for the binary eroded/dilated image being > 0 and the OBIA expression being a standard deviation of the variance feature of > 0.8 and a size > 400.

The outputs from the connected component segmentation and the mathematical morphological processing are then concatenated and the objects discriminated on the basis of an elongation factor of > 4 and size (> or < 100) to discriminate large and small compact objects). The approach was found to be particularly useful for identifying individual tree, which may form a hedgerow and separating these from trees that were isolated (Figure 7c and d). The processing sequence is

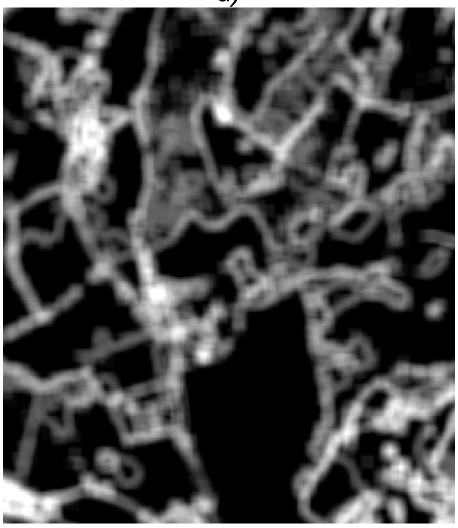
summarized in Figure 8.

a)

b)

c)

d)



e)

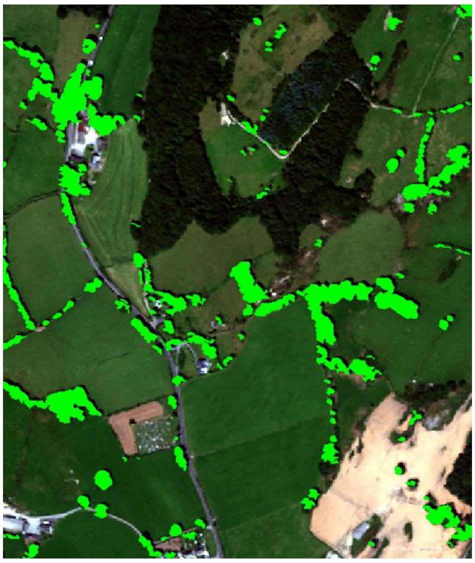


Figure 6. a) The original Worldview image, b) smoothed using the mean shift algorithm. c) the Haralick texture feature and d) the variance image. e) the final hedgerow layer.

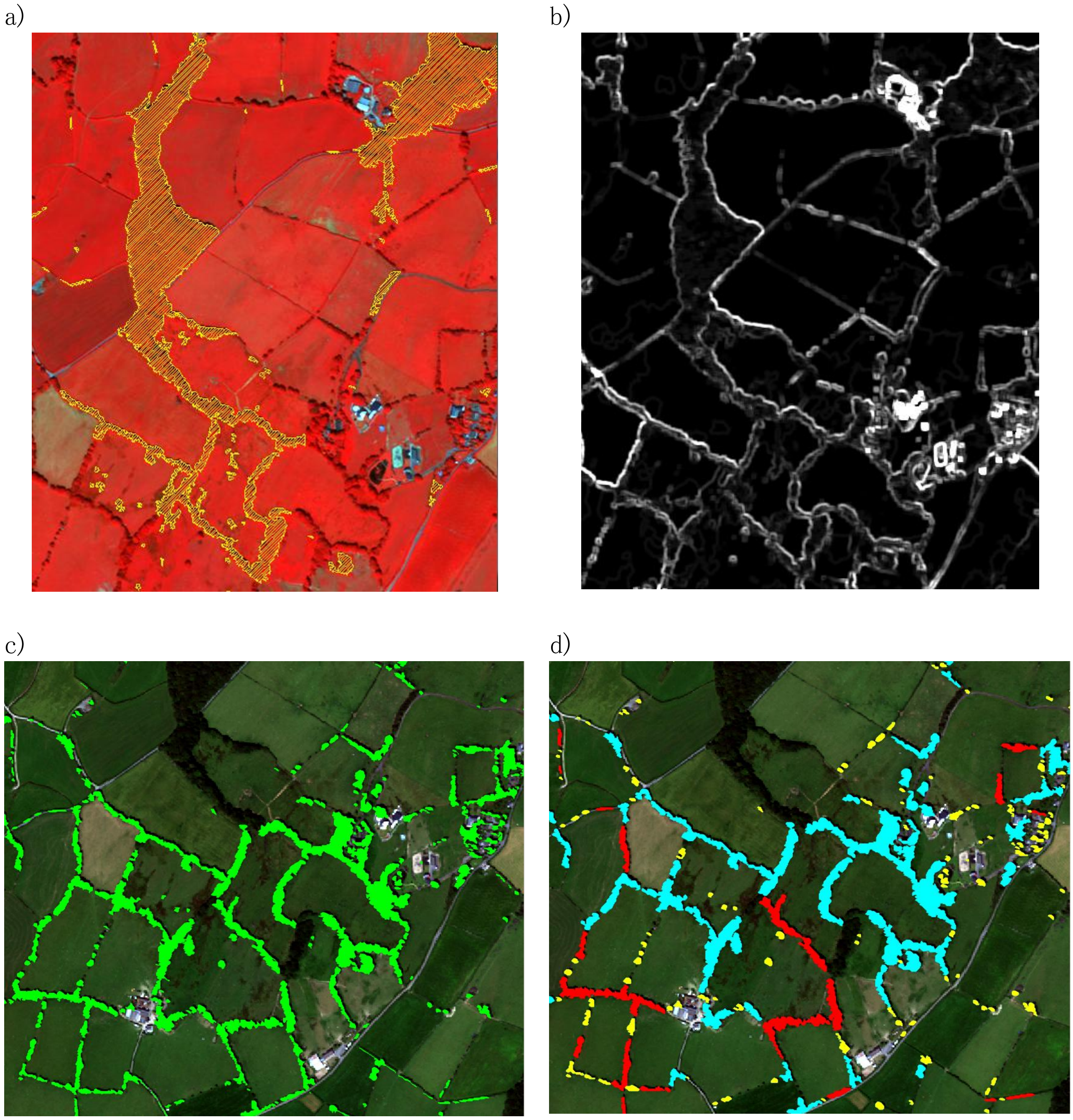


Figure 7. Examples of hedgerow and tree objects (variance < 0.8) and the resulting hedgerow layers.

Elongated

Compact large (Size >100)

Compact small (Size <100)

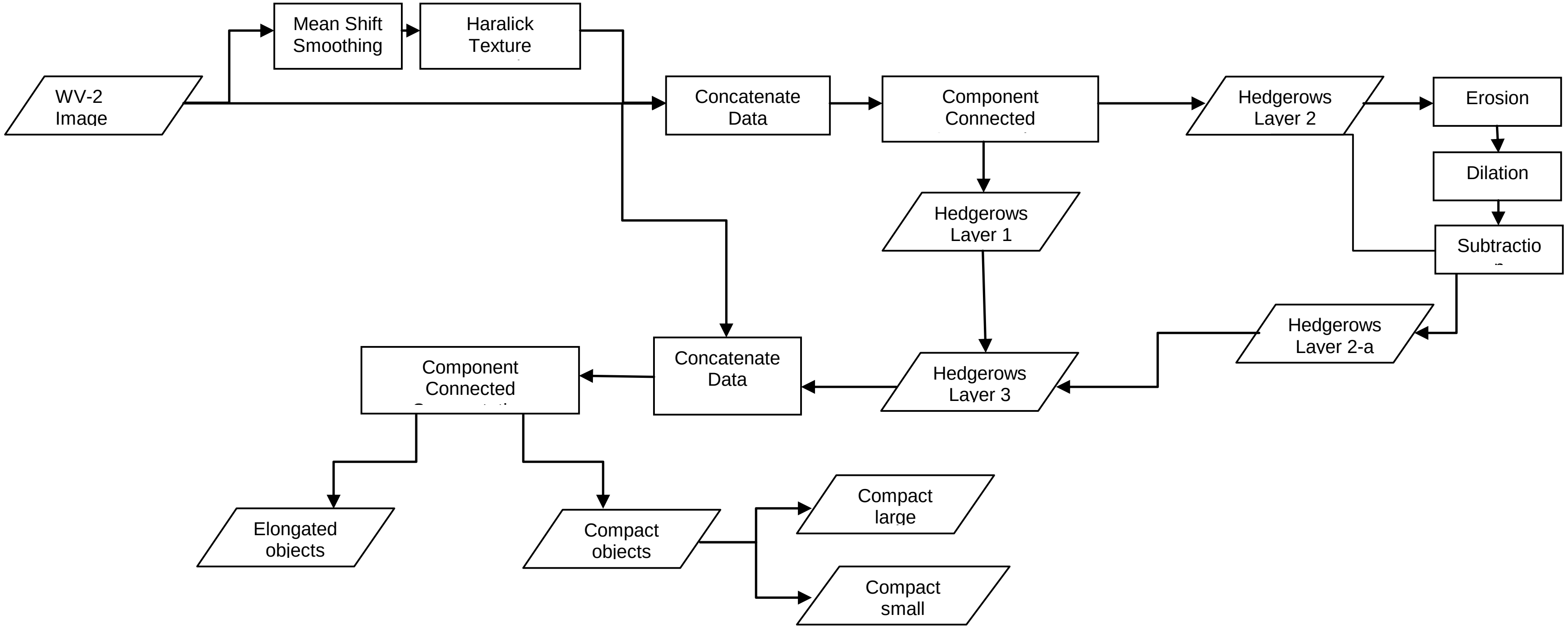


Figure 8. Sequence of processing for extraction of hedgerows, trees and isolated forest stands.

3.3.2 Vegetation height estimation through texture analysis

Texture measures may be applied to facilitate discrimination of some LCCS categories, through characterization of their morphological patterns. Texture is commonly described on the basis of variance within kernels or objects and is often represented by first order (occurrence) and second–order (co-occurrence) statistics, including homogeneity, entropy and correlation (Petrou, 2006).

In D5.1 and D5.2 a preliminary effort was undertaken to use texture properties (i.e. first-order entropy from the green band) extracted from VHR optical data to retrieve information on vegetation height and discriminate among certain GHC and LCCS categories, especially trees and shrubs respectively. Although primary height information, such as that derived from LiDAR data, exists for several BIO_SOS training sites, the requirement for the generic applicability of BIO_SOS EODHaM system to sites where such data are not always available dictates the need for alternative approaches to approximating vegetation height.

New experiments have been conducted on the Le Cesine site, Italy, (already used in D5.1 and D5.2) where various texture measures have been evaluated for their ability to differentiate landscape patches with vegetation of various heights. The measures are calculated locally, initially on a per-pixel basis, and then averaged for and assigned to each patch. Based on these values, patches are labelled depending on whether the vegetation is either below or above 2 m in height. The proposed measures are based on estimations of local variance, local entropy and local binary patterns as well as proposed variations, with these calculated for each band of a Quickbird image of Le Cesine. The ability of each texture measure and the band from which it was calculated has been assessed in terms of its ability to differentiate between low (< 2 m) and high (>= 2 m) vegetation through both statistical tests and classification using decision trees (Breiman, 1984). For most measures, the accuracy of classification was high and was greatest (98.6 %) when a variation of the local binary pattern measures calculated in the green band was used. Further details can be found in (Petrou et al., 2012). Thresholds from the different texture measures can be fed into the main LCCS classification algorithm to assist discrimination amongst height categories when LiDAR or other measures of vegetation height are not available.

An example of the followed process is demonstrated in Figure 9, for a variation of the local binary patterns approach calculated for the green band of the Quickbird image. Figure 9a shows a subset of the true colour Quickbird image acquired over Le Cesine. In Figure 9b, tree and shrub patches with vegetation lower and higher than 2m are indicated in the olive green and bright green shades respectively. A texture value is calculated for every pixel of each patch, with these illustrated in Figure 9c. Lighter colours represent higher values, with these ranging from 0 to 255. Pixel values are averaged for each patch, with the mean values shown in Figure 9d where the patches with higher vegetation have higher texture values compared to shorter vegetation.

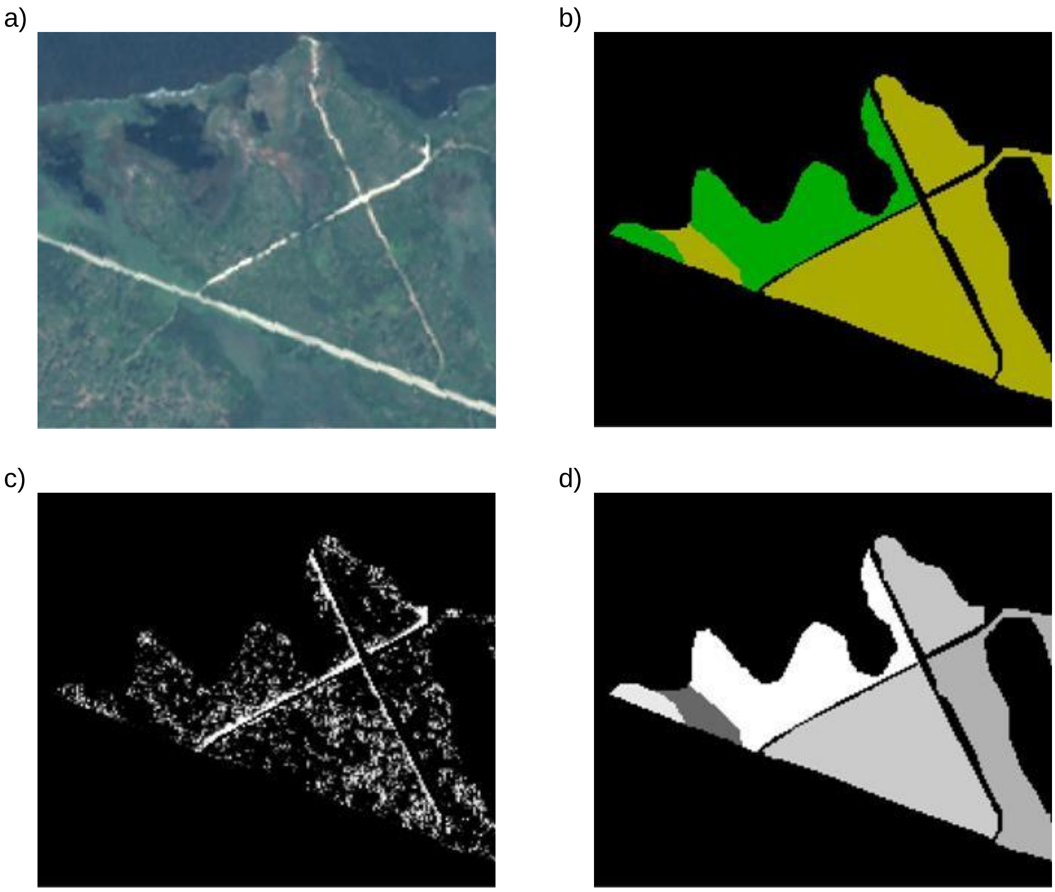


Figure 9. Discrimination between vegetation of lower and higher than 2m in height. a) Original Quickbird image. b) Patches lower and higher than 2m, drawn in olive and bright green respectively. c) Texture measure estimation per pixel. d) Average texture measure per patch.

3.3.3 Dempster-Shafer analysis

Differences in the types and availability of data from site to site are commonly observed, especially when studying diverse regions and landscapes. For instance, data might be sourced from different sensors, resulting in the low availability of certain spectral bands used for LCCS mapping in some sites, while LiDAR information might only be available for some sites. Where such information is absent, the ability to map LCCS classes can be compromised because further disambiguation between certain classes is not possible. In addition, the data may be afflicted by noise or inaccurate thresholds in the expert rules may be applied, with both leading to possible misclassifications. All of these shortcomings can introduce uncertainty in the final classification result, which needs to be handled in order to increase the robustness of the EODHaM 1st and 2nd stages.

The Dempster-Shafer (D-S) theory, providing a framework able to handle uncertainty and missing information, is adopted in BIO_SOS to improve system robustness, both up to Level 2 of LCCS classification and beyond. The theory, introduced by Dempster (1967) and Shafer (1976), is a mathematical theory of evidence, considered as a generalization of the Bayesian theory. The main difference with the Bayesian concept is the assignment of a belief interval to each event (e.g., a potential LCCS class for a specific patch) instead of a single probability. The belief interval consists of two values: The lower limit, referred as ‘belief’, denotes the confidence that an event holds, based on the available evidence supporting it; the upper limit, or ‘plausibility’, indicates the highest confidence on the event, in the event that all missing information were in favour of this event. This formalism allows a classification scheme based on this theory to express the natural uncertainty in the different events, providing the lowest and highest degree of confidence that they hold, with this being similar to a coverage interval for uncertainty evaluation (JCGM/WG 1, 2008).

The framework allows the assignment of belief intervals to both single and multiple events. Initially, a mass function, or basic probability assignment function, value *m*, is assigned to each single or multiple event for which there is information, indicating how probable this event will hold. Following the common probability principles, all *m* values sum up to 1. Then, for each single and multiple event, a belief value, *bel*, is calculated as the sum of the *m* values of all its subset events. For instance, if *A* and *B* are two possible single events, $bel(A) = m(A)$, $bel(B) = m(B)$ and $bel(\{A,B\}) = m(A) + m(B) + m(\{A,B\})$, where $\{A,B\}$ stands for the case where either *A* or *B* event happens. In addition, a plausibility value, *pls*, is calculated for each such event *A*, by removing from the absolute probability 1, the belief in the complement of event *A*, $\bar{A}$, as $pls(A) = 1 – bel(\bar{A})$. In the Bayesian framework, the plausibility of an event *A* would be equal to the belief in it, since disbelief in one event implies belief in its complement, and *vice versa*. However, in the Dempster-Shafer framework, the belief in an event *A* can be smaller than its plausibility, $bel(A) \leq pls(A)$, their absolute difference, or belief interval, with this representing uncertainty and missing information (Ahmadzadeh, 2001).

The D-S theory provides a concrete framework for data fusion, for combination of evidence that might come from different sources, e.g. RS sensors, expert opinions, etc. The *m* function values of the different events are combined using the so-called Dempster rule of combination, performing a normalization for empty events. Providing an example of the data fusion functionality of the D-S theory, let us assume that two experts express their opinion on land cover characterization of a certain patch during photo-interpretation, as

- Expert 1: $m_1(\{c_1\}) = 0.5$, $m_1(\{c_3,c_4\}) = 0.3$, $m_1(\{\Theta\}) = 0.2$,
- Expert 2: $m_2(\{c_1\}) = 0.6$, $m_2(\{c_2,c_3\}) = 0.3$, $m_2(\{\Theta\}) = 0.1$,

where m_1 and m_2 represent the opinions of the first and second expert, respectively, c_i, $i = 1,\ldots,4$, stand for LC classes, while Θ for the entire set of all possible classes, i.e. $m_1(\{\Theta\}) = 0.2$ can be interpreted as “probability of 0.2 that the patch can be of any LC class”.

The opinions of the two experts are merged into a final *m* function, calculated as illustrated in Table 2. When the mass functions of the two sets are combined, a new mass function value *m* is produced for the intersection of the sets, by multiplying the respective m_1 and m_2 values.

Table 2. Dempster rule of combination.

		Expert 1		
		$m_1(\{c_1\}) = 0.5$	$m_1(\{c_3,c_4\}) = 0.3$	$m_1(\{\Theta\}) = 0.2$
Expert 2	$m_2(\{c_1\}) = 0.6$	$m(\{c_1\}) = 0.3$	$m(\{\varnothing\}) = 0.18$	$m(\{c_1\}) = 0.12$
	$m_2(\{c_2,c_3\}) = 0.3$	$m(\{\varnothing\}) = 0.15$	$m(\{c_3\}) = 0.09$	$m(\{c_2,c_3\}) = 0.06$
	$m_2(\{\Theta\}) = 0.1$	$m(\{c_1\}) = 0.05$	$m(\{c_3,c_4\}) = 0.03$	$m(\{\Theta\}) = 0.02$

The sum of the mass function values *m* in the table is 1, including values assigned to the empty set. The total value assigned to the empty set is $K = 0.18+0.15 = 0.33$ and should be removed from the process by dividing all *m* values referring to non-empty sets by $(1\text{-}K)$, in order to force them to sum up to 1. After summing the values referring to the same sets, the final mass function values become:

$$m(\{c_1\}) = (0.3+0.12+0.5)/(1-0.33) = 0.701, \quad m(\{c_3\}) = 0.134,$$

$$m(\{c_2,c_3\}) = 0.09, \qquad m(\{c_3,c_4\}) = 0.045, \qquad m(\Theta) = 0.03.$$

As deduced from the text, $m(\{c_2,c_3\})$ indicates the degree of a specific partial ignorance; the degree of certainty that the area belongs to either c_2 or c_3 land cover class, and at the same time the degree of lack of information for the distinction between the two. $m(\Theta)$ indicates the degree of complete ignorance and complete incapability in distinguishing among any class. Based on the *m* function values, the belief values for the different events are calculated as:

$$bel(\{c_1\}) = 0.701, \qquad bel(\{c_3\}) = 0.134,$$

$$bel(\{c_2,c_3\}) = 0.224, \quad bel(\{c_3,c_4\}) = 0.179, \quad bel(\Theta) = 1.$$

Based on the definition of the belief function, $bel(\{c_2,c_3\})$ expresses the belief that one of the two classes c_2 and c_3, no matter which, is assigned to the under investigation area. The respective plausibility values are then:

$$pls(\{c_1\}) = 1 - bel(\{c_2,c_3,c_4\}) = 0.731,$$

$$pls(\{c_3\}) = pls(\{c_2,c_3\}) = pls(\{c_3,c_4\}) = 1 - 0.701 = 0.299, \qquad pls(\Theta) = 1.$$

The differences between belief and plausibility (i.e., the belief intervals) represent the uncertainty of a specific event. This uncertainty arises from the previous events, including uncertainty in data, conditions and expert rules, with these mainly caused by a lack of adequate information. The D-S theory provides a framework from the propagation of uncertainty from event to event, up to the final resulted event under study. As an example of demonstrating uncertainty creation and handling in the D-S framework, consider a landscape patch that, in order to be classified as aquatic, class c_A, or terrestrial, class c_T, a rule stating that “if the mean Water Band Index (WBI) is above a threshold *t*, then the patch is classified as aquatic with a probability of 80%” is used. In the case of no information on WBI being available, no probabilities can be given to either the c_A or c_T classes. However, belief intervals can be assigned describing the lower and higher limits of such probabilities. Therefore, c_A class is assigned a belief interval [0, 0.8], while c_T a belief interval [0,0.2]. Belief on both individual events-classes is 0, stating that no information exists to support either event. In addition, the plausibility of the patch belonging in an aquatic class, c_A, is 0.8, which is the larger probability it would have in the case that the WBI values were both available and larger than the pre-defined threshold *t*; the respective plausibility of class c_T is 0.2, in the case that both WBI is known and less than *t*. Furthermore, uncertainty can be expressed in both the data and expert rules themselves. As an example, consider a landscape patch being assigned a mean Forest Discrimination Index (FDI) value of more than –10 with a belief interval [0.6, 0.8] and a subsequent rule stating that “patches with FDI> –10 are classified as photosynthetic vegetation, class c_{PV}, with uncertainty [0.7,0.9]”. According to the D-S theory uncertainty propagation framework, the belief interval of class c_{PV} is calculated as [0.6x0.7,1-0.8(1-0.9)] = [0.42,0.92], denoting the minimum belief based on the given data and the maximum confidence in the event if all uncertain information were to support the photosynthetic class (Hau and Kashyap, 1990).

In BIO_SOS, the ability of the D-S theory in allowing multiple events is particularly useful when handling missing information which prevents the discrimination of two or more potential LCCS (or habitat) classes, such as in cases where specific indices used in the classification cannot be calculated (e.g. the WBI, calculated as the ratio of 990nm to 900nm, from a Quickbird image, since no band near the 990nm spectral region is captured by the respective sensor). In such cases, where a decision in favour of one class is impossible, a belief interval is assigned to the complex event that the patch belongs in one of the potential classes (which are not specified) and the classification process continues accordingly. If appropriate evidence appears, the classification can be refined through the assignment of belief intervals to the individual classes. In order to mitigate the effects of potential noise affliction of the input data or inaccurate thresholds provided by the experts in the form of classification rules and surpass the limitations of a crisp classification framework, several fuzzification approaches are proposed and evaluated. The approaches include both generic fuzzification solutions and individual ones for each feature involved in the classification process, with these based on theoretical analysis of its expected properties. The D-S framework offers great flexibility in incorporating such fuzzification approaches, in order to improve both the system’s tolerance to noise and its applicability and transferability among similar sites, even in different geographical regions. The D-S framework is currently being implemented to include the crisp LCCS classification rules developed for the EODHaM second stage and to incorporate the appropriate fuzzification approaches.

In summary, the Dempster-Shafer theory offers great flexibility in rule-based classification and is expected to be particularly effective in handling uncertainty in the BIO_SOS case, where diversity in the geographical regions, sensors and image acquisition times makes the design of a unified land cover classification framework challenging.

4 Product quality evaluation procedures (validation)

For each of the test sites in Italy, Wales, the Netherlands and Portugal, validation data in the form of LCCS and/or GHC surveys were conducted. Further field data will however be collected during 2013 given the large diversity and complexity of classes (particularly LCCS) generated from the classification approach.

4.1 Reference unit sampling strategy

For validation at sites, a stratified random sampling approach was taken based on the best available existing vegetation mapping (for example, the Phase 1 Habitat Survey of Wales). From these maps, squares of standard dimensions (e.g., 250 or 1 km) with origins aligned with a 1 km national grid were generated, with a number selected for sampling within each of the broad categories identified (e.g., broadleaved forest, improved grasslands, water). Within each of these, a standard GHC survey was conducted using the protocols outlined in Bunce et al. (2011) and also D6.2. From the field data and also available VHR remote sensing data (e.g., aerial photography), LCCS categories were assigned. Further details on the sampling have been provided in Deliverable D6.3.

4.2 Validation of classifications

For validation, standard confusion matrices were used for LC and/or GHC maps, with the latter often providing comparatively less detail and more options in terms of class allocation. A difficulty in the assessment of accuracy was a) the complex descriptions of LCCS classes and b) ambiguities in ground and satellite-based classifications because of similarities between some classes (e.g., height and cover categories). Hence, consideration needs to be given to standardized approaches to assessing accuracy, which will be reported in D5.6. However, a preliminary assessment of accuracy in the classification of LCCS and GHC categories is reported for several sites and options for describing and standardising accuracy are considered.

One option for the validation of LCCS maps is to consider each of the contributing layers separately and then combine the error estimates to produce an assessment of overall accuracy. A particular benefit of the approach is that the where accuracies are reduced for some components (e.g., leaf type), with these contributing to a reduction in the users, producers and/or overall accuracy, the reasons for this can be established and hence recognised by the user. Accuracies may be particularly low for some categories but offset by higher accuracies for others (e.g., height classes derived from LIDAR). Procedures are being implemented to provide an accuracy assessment based on this approach but will reported in the final deliverable for WP5 (D5.6) alongside estimates of accuracy in the detection of change.

5 LCCS and GHC maps, BIOSOS Study Sites

5.1 Wales (Cors Fochno)

The Cors Fochno site in Wales was classified using a combination of Worldview surface reflectance data that captured the 'stable' pre and peak flush periods (March 2012 and July 2011 respectively) as well as the post flush transition period (November, 2011). In addition, LiDAR data were also available, which were used primarily to quantify vegetation height.

For the site, key features in the landscape that required identification within the small object layer included buildings and temporary structures (e.g., caravans) and hedgerows. For these, specific tools for feature extraction outlined in the previous section were used, with the remaining landscape segmented using the procedures of Bunting et al. (2012). Larger objects were associated with fields comprised primarily of improved grasslands and managed forestry plantations, with the latter associated with land units that were largely occupied by woody vegetation (in the small object layers).

The classification of LCCS categories is provided in Figure 10 with the legend provided in Appendix I. The classes (including the saltmarsh and sand dune complex) are listed in Tables 3a and b, with the large number generated resulting from the inclusion of LiDAR data but also the diversity of habitats occurring at the site. The classification reflected the distribution of land covers observed within the area, including the active bog, perennial closed grasslands, the cultivated areas, different forest types and urban areas. The accuracy in the classification of the majority of LCCS classes exceeded 70 %, but further validation will be required with reference to aerial photography, unmanned airborne system (UAS) and field data to be collected in 2013.

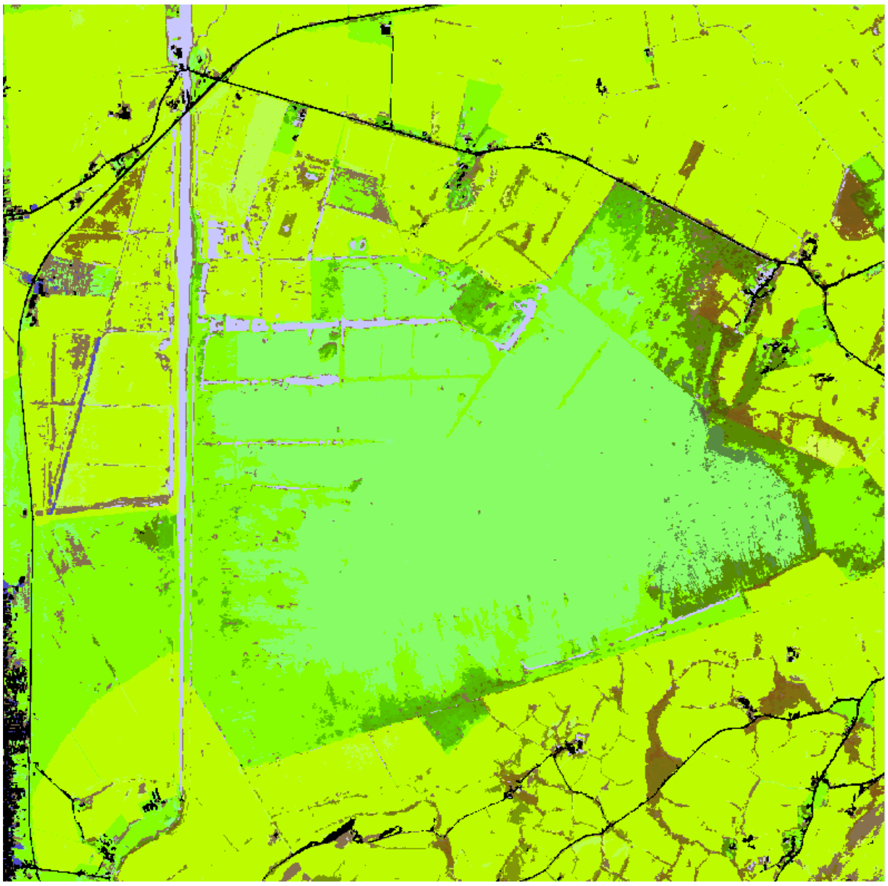


Figure 10. Classification of LCCS categories, Cors Fochno and surrounds.

Table 3a. Descriptions of LCCS classes for Cors Fochno

LCCS Class Description	
Broadleaved Deciduous Closed Low Trees	Graminoid crops
Broadleaved Deciduous Closed Medium High Shrubland (thicket)	Graminoids on Flooded land
Broadleaved Deciduous Closed Medium Trees	Gravel, Stones and Boulders
Broadleaved Deciduous Closed Shrubland (thicket)	Large sized field of Broadleaved Deciduous Shrub crops
Broadleaved Deciduous Closed Woody vegetation	Large sized field of Broadleaved Deciduous Tree crops
Broadleaved Deciduous Low Trees on Flooded land	Large sized field of Broadleaved Evergreen Shrub crops
Broadleaved Deciduous Medium High Shrubland (thicket) on Flooded land	Large sized field of Broadleaved Evergreen Tree crops
Broadleaved Deciduous Medium Trees on Flooded land	Large sized field of Needleleaved Deciduous Shrub crops
Broadleaved Deciduous Open (15-40%) Low Trees on Flooded land	Large sized field of Needleleaved Deciduous Tree crops
Broadleaved Deciduous Open (15-40%) Medium High Shrubland (thicket) on Flooded land	Large sized field of Needleleaved Evergreen Tree crops
Broadleaved Deciduous Open (15-40%) Shrubland (thicket) on Flooded land	Medium sized field of Broadleaved Deciduous Tree crops
Broadleaved Deciduous Open (15-40%) Woody vegetation on Flooded land	Needleleaved Deciduous Closed Low Trees
Broadleaved Deciduous Open (40-65%) Low Trees	Needleleaved Deciduous Closed Medium High Shrubland (thicket)
Broadleaved Deciduous Open (40-65%) Low Trees on Flooded land	Needleleaved Deciduous Closed Medium Trees
Broadleaved Deciduous Open (40-65%) Medium High Shrubland (thicket)	Needleleaved Deciduous Closed Shrubland (thicket)
Broadleaved Deciduous Open (40-65%) Medium High Shrubland (thicket) on Flooded land	Needleleaved Deciduous Closed Woody vegetation
Broadleaved Deciduous Open (40-65%) Medium Trees	Needleleaved Deciduous Open (15-40%) Low Trees on Flooded land
Broadleaved Deciduous Open (40-65%) Medium Trees on Flooded land	Needleleaved Deciduous Open (15-40%) Medium Trees on Flooded land
Broadleaved Deciduous Open (40-65%) Shrubland (thicket)	Needleleaved Deciduous Open (15-40%) Woody vegetation on Flooded land
Broadleaved Deciduous Open (40-65%) Shrubland (thicket) on Flooded land	Needleleaved Deciduous Open (40-65%) Low Trees
Broadleaved Deciduous Open (40-65%) Woody vegetation	Needleleaved Deciduous Open (40-65%) Low Trees on Flooded land
Broadleaved Deciduous Open (40-65%) Woody vegetation on Flooded land	Needleleaved Deciduous Open (40-65%) Medium High Shrubland (thicket)
Broadleaved Deciduous Shrub crops	Needleleaved Deciduous Open (40-65%) Medium High
Broadleaved Deciduous Shrubland (thicket) on Flooded land	Needleleaved Deciduous Open (40-65%) Shrubland (thicket) on Flooded land
Broadleaved Deciduous Tree crops	Needleleaved Deciduous Open (40-65%) Woody vegetation
Broadleaved Deciduous Woody vegetation on Flooded land	Needleleaved Deciduous Open (40-65%) Woody vegetation on Flooded land
Broadleaved Evergreen Shrub crops	Needleleaved Deciduous Shrub crops
Broadleaved Evergreen Tree crops	

Table 3b. Description of LCCS classes, Cors Fochno, Wales

LCCS Class description	
Needleleaved Deciduous Tree crops	Small sized field of Broadleaved Deciduous Tree crops
Needleleaved Deciduous Woody vegetation on Flooded land	Small sized field of Broadleaved Evergreen Shrub crops
Needleleaved Evergreen Shrub crops	Small sized field of Broadleaved Evergreen Tree crops
Needleleaved Evergreen Tree crops	Small sized field of Graminoid crops
Non-Graminoid crops	Small sized field of Needleleaved Deciduous Shrub crops
Open (15-40%) Graminoids on Flooded land	Small sized field of Needleleaved Deciduous Tree crops
Open (40-65%) Forbs	Small sized field of Needleleaved Evergreen Shrub crops
Open (40-65%) Graminoids	Small sized field of Needleleaved Evergreen Tree crops
Open (40-65%) Graminoids on Flooded land	Small sized field of Non-Graminoid crops
Short Closed Graminoids	Sparse (1-4%) Graminoids on Flooded land
Short Graminoids on Flooded land	Turbid Shallow water bodies
Short Open (15-40%) Graminoids on Flooded land	Closed (> 65 %) Graminoids
Short Open (40-65%) Graminoids	Turbid Shallow Artificial Perennial waterbodies
Short Open (40-65%) Graminoids on Flooded land	Turbid Shallow Artificial snow
Small sized field of Broadleaved Deciduous Shrub crops	Turbid Shallow Artificial waterbodies

5.2 The Netherlands, Veluwe

The classification of Veluwe was based on WV-2 data acquired in June and September 2011, with the former acquired during the period of peak flush. However, the vegetation was still highly productive, although transitional, in September and hence contrasts in the landscape were not as great compared to the pre-flush period. Nevertheless, many of the LCCS categories could be discriminated using the combination of the two dates. The classification also benefited from the inclusion of DEMs representing surface (DTM) and canopy height (CHM) for the majority of site. The larger objects consisted primarily of the agricultural fields and also the managed forestry plantations, with smaller objects largely associated with buildings as tree crops and hedgerows were generally absent. The LCCS classification (Figure 11a; legend in Appendix I and classes in Table 4) reflected the spatial arrangement of land covers, with these being largely cultivated or managed with a near equal division into agricultural land and forestry and about 20 % of the landscape consisting of urban development. Separate maps reflecting the height and cover of woody vegetation are provided in Figure 11b.

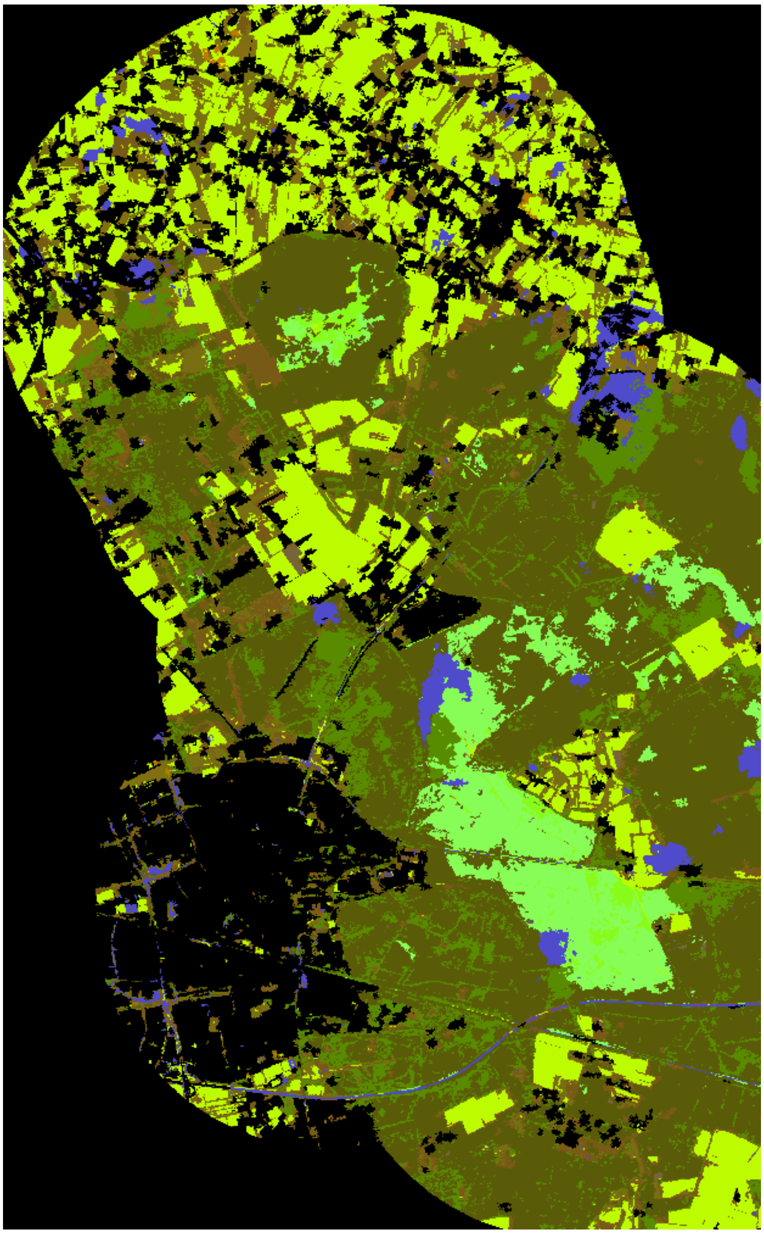


Figure 11a. Classification of LCCS categories, Veluwe, the Netherlands.

Table 4. Subset of LCCS class descriptions, Veluwe, the Netherlands

LCCS Class Description	
Annual Closed Graminoids	Needleleaved Deciduous Closed Woody vegetation
Annual Graminoids	Needleleaved Deciduous Open (15-40%) Woody vegetation on Flooded land
Annual Open (40-65%) Graminoids	Needleleaved Deciduous Open (40-65%) Woody vegetation
Annual Short Closed Graminoids	Needleleaved Deciduous Tree crops
Annual Short Open (40-65%) Graminoids	Needleleaved Evergreen Tree crops
Graminoid crops	Non-Graminoid crops
Herbaceous crops	Closed Forbs
Broadleaved Deciduous Closed Woody vegetation	Open (40-65%) Forbs
Broadleaved Deciduous Open (40-65%) Woody vegetation	Short Closed Forbs
Broadleaved Deciduous Tree crops	Short Open (40-65%) Forbs
Broadleaved Evergreen Tree crops	Turbid Shallow
Unconsolidated materials	Turbid Shallow Artificial Perennial waterbodies
Consolidated materials	Turbid Shallow Artificial waterbodies
Deep to Medium Artificial Perennial waterbodies	Deep to Medium Artificial waterbodies

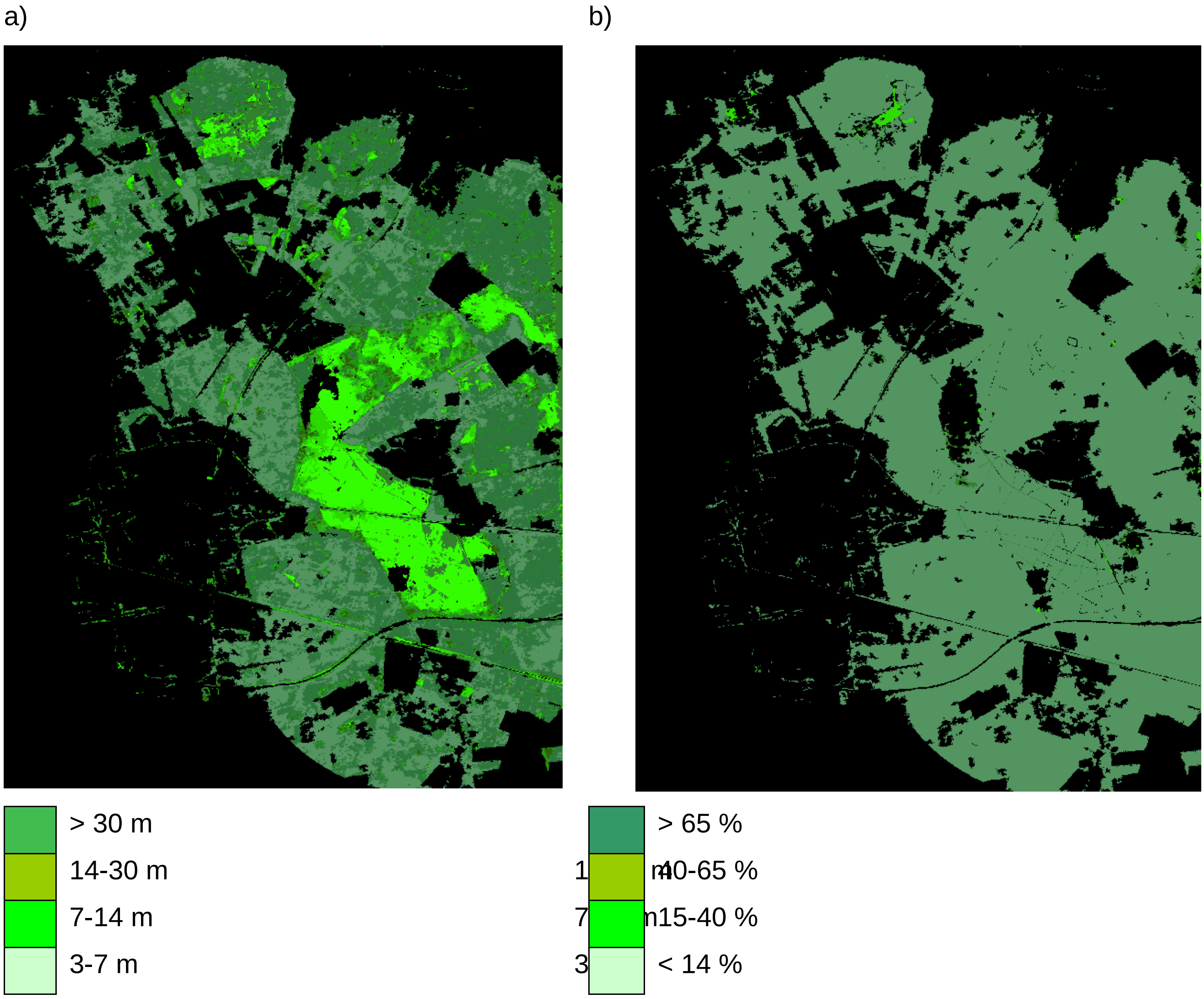


Figure 11b. Classification of a) height and b) cover categories for vegetated areas occurring within the Veluwe study site. Note that only the western and central parts of the site are displayed because of incomplete coverage of LiDAR data.

To support validation of the classification of land covers and GHCs, fieldwork was conducted in June and July 2011. For twelve sites, GHC maps were generated with the locations selected according to a stratified random sampling approach. Half of the sites were located in the protected Natura 2000 site, he Veluwe, whilst the other were outside. The sample sites inside the National Park were characterised primarily by forest, heathland and inland sand dunes whilst those outside were mainly characterised by agricultural fields and farms as well as forest. Each sample site covered 6 ha and the GHCs were mapped according to the protocol in the GHC handbook (Bunce *et al*., 2011). In short, each mapping unit covered a minimum area of 400 m^2 and was characterised in its composition in terms of major life forms (and percentage cover of each totalling 100) and plant species (fauna were also recorded). For all mapping units within a specific GHC, vegetation relevés were made in nested plots although these were not used in the validation, which focused purely on recorded GHCs and associated land covers. An example of the mapping generated, with the classification of GHCs (translated from LCCS categories) as a backdrop, is given in Figure 12, with the associated GHCs listed in Table 5.

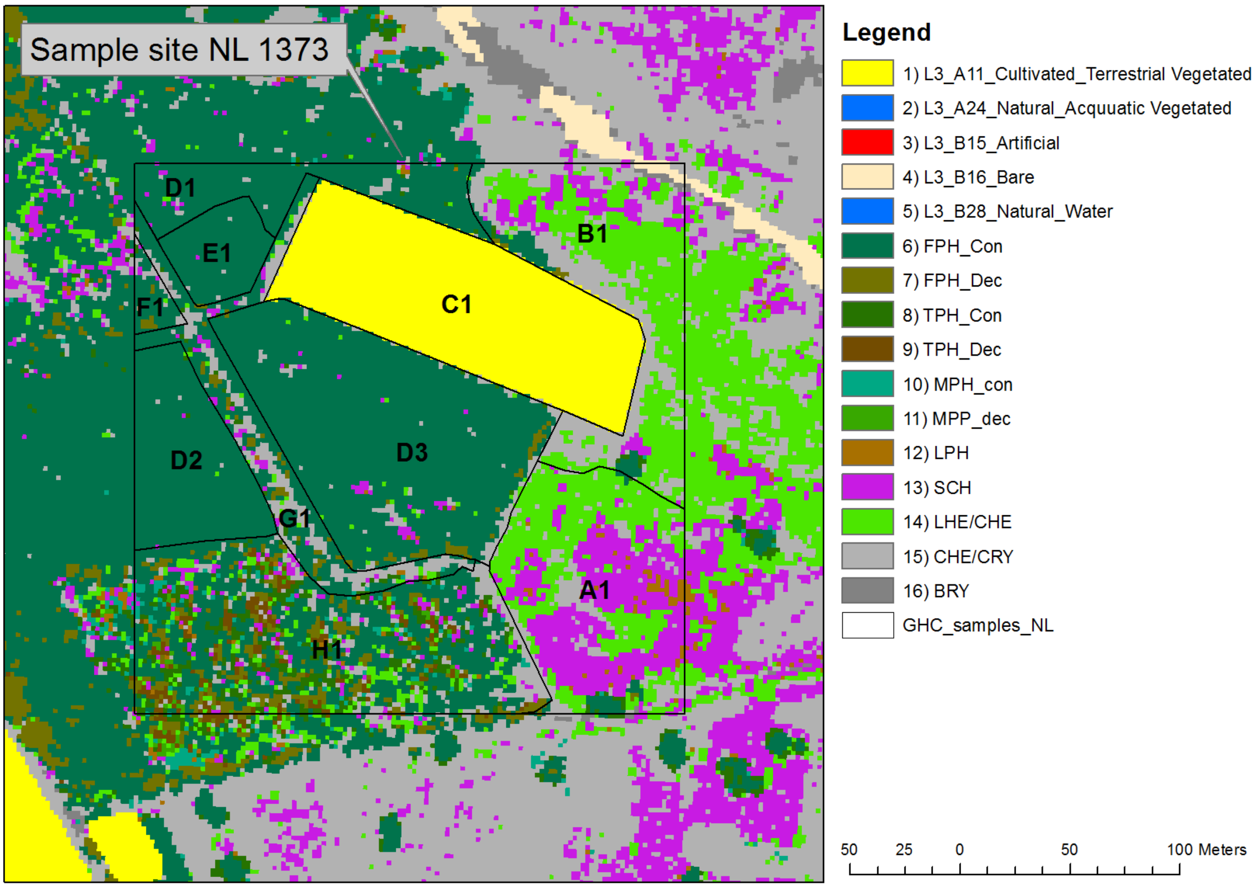


Figure 12. The EODHaM classification overlain with a GHC grid (sample NL 1373)

Table 5. Major attributes of the GHC recorded in the field for NL1373

Alpha	GHC	Date	LF1	% cov	LF2	% cov	LF3	% cov	LF4	% cov
G1	TRS/FPH/DEC	30-6-11	SPV/SAN	30	TRS/FPH/DEC	70		0		0
C1	CUL/CRO	30-6-11	CUL/CRO	100		0		0		0
B1	TRS/SCH/EVR	30-6-11	HER/CHE	40	TRS/SCH/EVR	30	TRS/LPH/EVR	20	HER/CRY	10
H1	TRS/FPH/CON	30-6-11	TRS/TPH/DEC	40	TRS/FPH/CON	50	HER/CHE	10		0
F1	TRS/FPH/CON	30-6-11	TRS/FPH/CON	100		0		0		0
D1	TRS/FPH/CON	30-6-11	TRS/FPH/CON	100		0		0		0
E1	TRS/FPH/CON	30-6-11	TRS/FPH/CON	100		0		0		0
B1	TRS/SCH/EVR	30-6-11	HER/CHE	40	TRS/SCH/EVR	30	TRS/LPH/EVR	20	HER/CRY	10
A1	TRS/LPH/EVR	30-6-11	TRS/SCH/EVR	50	TRS/LPH/EVR	50		0		0
D1	TRS/FPH/CON	30-6-11	TRS/FPH/CON	100		0		0		0
H1	TRS/FPH/CON	30-6-11	TRS/TPH/DEC	40	TRS/FPH/CON	50	HER/CHE	10		0

Using a standard confusion matrix comparing the dominant GHC recorded in the field and mapped from the VHR data, the overall accuracy of classification was 69 % (Table 5). As there is some ambiguity in the classes, the classification was assumed to be correct when both cells in dark and light grey were shaded. However, the accuracy was increased to 76 % when the following was considered:

a) Shrubby chameophytes and low phanerophytes differ by 10 cm in height and this can be misinterpreted in the field; hence, classification as either was considered correct.

b) The GHC is determined mainly by the dominant lifeform, even though the types occurring can be very close in coverage. Hence, the second and third lifeform were considered as correct if classified as such (Table 6).

Table 6. The different life forms and their estimated coverage per mapping unit and their final GHC code. Often, more life forms are present within a mapping unit and the second life form can be close to the first in terms of coverage. The last column shows the dominant life form that has been classified based on the majority rule for all pixels within the mapping unit.

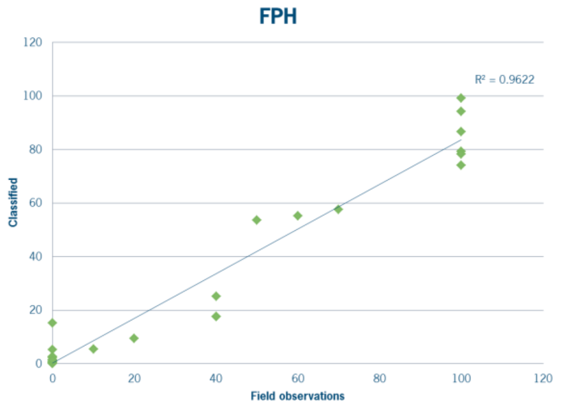
	FIELD MEASUREMENTS									RS
MU	GHCCode_field	LF1	LF1 % COV	LF2	LF2 %COV	LF3	LF3 %COV	LF4	LF4 %COV	GHC_clas
NL510B1	HER/CHE	TRS/SCH/EVR	20	HER/CHE	60	HER/LHE	10	SPV/EAR	10	CHE/CRY
NL510E1	TRS/LPH/EVR	HER/CHE	50	TRS/LPH/EVR	50		0		0	LHE/CHE
NL510C1	TRS/LPH/EVR	HER/CHE	20	TRS/LPH/EVR	70	HER/CRY	10		0	SCH
NL510D1	HER/CHE	HER/CHE	60	TRS/SCH/EVR	20	HER/CRY	10	SPV/EAR	10	SCH
NL510F1	TRS/FPH/DEC	HER/CHE	10	TRS/FPH/DEC	40	TRS/TPH/DEC	10	TRS/SCH/EVR	40	LHE/CHE
NL52J1	TRS/FPH/DEC/CON	TRS/FPH/DEC	60	TRS/FPH/CON	40		0		0	FPH_Con
NL52M1	TRS/FPH/DEC	TRS/FPH/DEC	80	TRS/FPH/CON	20		0		0	FPH_Dec
NL52C1	TRS/FPH/DEC	TRS/FPH/DEC	70	TRS/FPH/CON	30		0		0	FPH_Dec

Table 7. Error matrix comparing the classification of dominant GHC recorded in the field against those classified from VHR data.

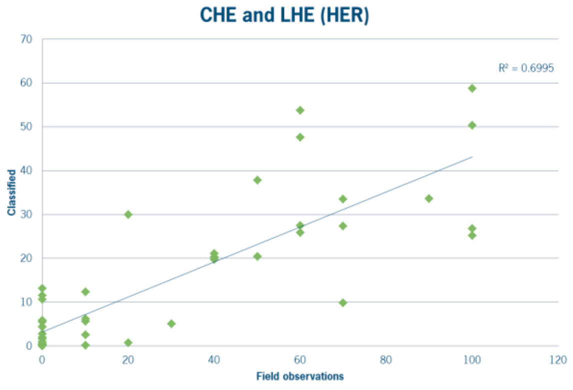
Sum of Count MU	GHC clasified																
GHC Field	CUL/ CRO	HER/ CHE	HER/ CHE/ CRY	HER/ CRY	BRY	HER/ LHE	HER/ LHE/ CHE	SPV/ SAN	TRS/ FPH/ CON	TRS/ FPH/ DEC	TRS/ FPH/DEC/ CON	TRS/ FPH/ DEC	TRS/ TPH/ CON	TRS/ LPH/ EVR	TRS/ SCH/ EVR	URB/ ART	Total
CUL/CRO	6																6
HER/CHE	11		5		1		1								4		22
HER/CHE/CRY			3														3
HER/CRY			1														1
BRY																	0
HER/LHE	1																1
HER/LHE/CHE	1		1														2
SPV/SAN								2									2
TRS/FPH/CON							1		11								12
TRS/FPH/DEC	2						1		3	6			1				13
TRS/FPH/DEC/CON									3								3
TRS/TPH/DEC							1										1
TRS/TPH/CON																	0
TRS/LPH/EVR			1				1								10		12
TRS/SCH/EVR			3				2										5
URB/ART	2		3		1											4	10
Grand Total	23	0	17	0	2	0	7	2	17	6	0	0	1	0	14	4	93
Mapping units	93																
OA	0.69																

A second approach to assessing classification accuracy is to compare the percentage coverage for a specific life form measurement in the field against that classified from the VHR imagery, with this achieved by computing the histogram of classified pixels for each mapping unit. For FPH, a correspondence of 96 % was observed (Figure 13) whilst this was 70 % for CHE and LHE (HER). These categories are difficult to discriminate from VHR data and hence were combined. The correspondence for SCH was lower, with this being the consequence of heathlands dominated by *Calluna vulgaris* being classified as LPH (low phanerophytes) in the field but as SCH in the imagery. The percentage coverage of shrubs is also more difficult to assess from field measurements.

a)



b)



c)

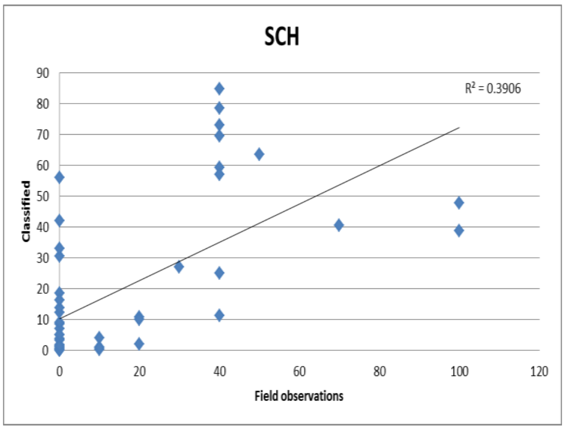


Figure 13. Correlation in the estimates of the percentage coverage for a) FPH, CHE and LHE (HER) and c) SCH as measured in the field and mapped from VHR imagery.

5.3 Italy

5.3.1 Le Cesine

The classification of Le Cesine was undertaken using a Quickbird and Worldview image acquired in April and October 2011 respectively. A limitation was that only a four Quickbird channels were available, which prevented the use of some indices (e.g., PSRI) and hence compromised the classification of, for example, non-photosynthetic vegetation. The classification of olive groves was based on counts of delineated trees within larger objects. The classification in Figure 14 (with legend in Appendix I and classes in Table 8) reflects that extensive area of olive groves and the semi-natural/natural habitats occurring in the Natura 2000 site.

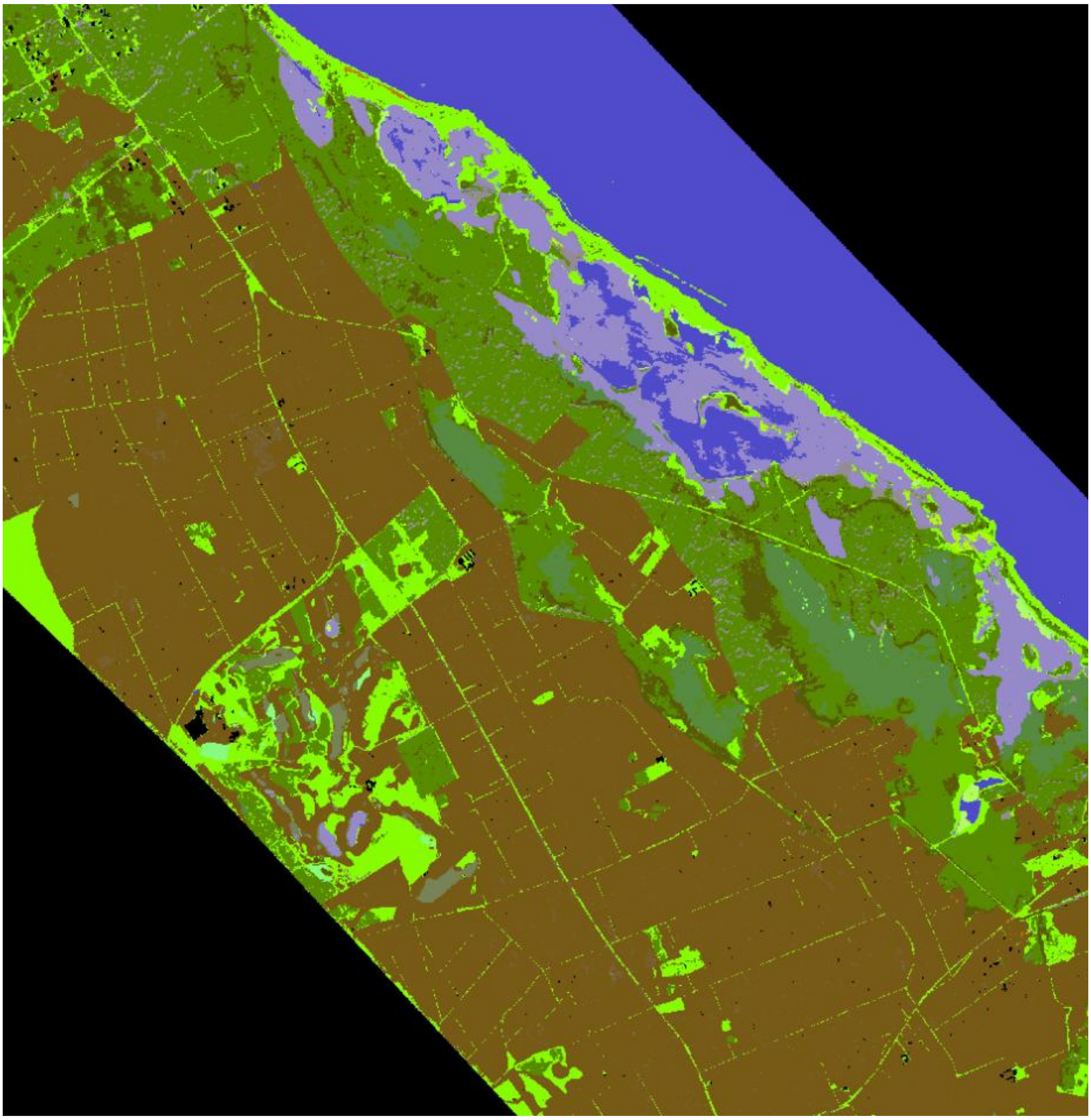


Figure 14. LCCS classification of Le Cesine, Italy.

Table 8. Subset of LCCS Class descriptions, Le Cesine, Italy

LCCS Class Description.	
Broadleaved Deciduous Closed Woody vegetation	Deep to Medium Artificial Perennial waterbodies
Clear Deep to Medium	Deep to Medium Artificial waterbodies
Clear Deep to Medium Artificial Perennial waterbodies	Needleleaved Deciduous Closed Woody vegetation
Clear Deep to Medium Artificial waterbodies	Short Closed Graminoids
Clear Shallow Artificial Perennial waterbodies	Turbid Shallow
Clear Shallow Artificial waterbodies	Turbid Shallow Artificial Perennial waterbodies
Closed Forbs	Turbid Shallow Artificial waterbodies
Consolidated materials	Deep to Medium Natural waterbodies
Unconsolidated materials	

5.3.2 Murgia Alta

The classification of Murgia Alta ("Murgia Alta" IT9120007 SCI/SPA) site was undertaken using two WorldView2 images acquired in 19th April and 10th October, 2011 respectively (Figure 15). Additional images have been collected but not yet used to improve the two-season classification. Additional WorldView2 images available for future classification refinements were acquired on the 22nd January and 6th July, 2012. An archival image was also available for the same area, with this acquired on the 16th April, 2009. The site is characterised by a complex system of agricultural fields which are not of regular shape and highly fragmented. Many of the farmers carved up the soil (an activity which is now illegal) between 1990 and 2000 as a consequence of EU incentives promoting durum wheat production, which complicates the spectral appearance of land covers within the imagery compromise the use of spatial (both topological and non-topological) and morphological relationships.

A classification beyond Level 3 (Figure 15c) has been made as a preliminary study using e-cognition, with a view to generating using the open source software described above. The classification of cultivated areas has been carried out using the convergence of evidence of the a-priori knowledge concerning:

- The agricultural practices in the site;
- The different morphological features of the agricultural fields depending on the quota (Digital Elevation Model - DEM). For example, in flat areas, agricultural fields have regular shape, but in hilly areas they are mixed with natural grassland and do not have regular shape.

For this site, a first GHC Map was also obtained by considering as input, a previous LC/LU map provided by Puglia Region at a scale of 1:5000. The map, originally in the CORINE taxonomy, was firstly converted into FAO-LCCS taxonomy and then used to produce the GHC map shown in Figure 16. The disambiguation rules were based on both spatial relations and phenological knowledge extracted by the two images more close to the LC/LU map validation made by Puglia Region. Whilst the Alta Murgia site has not been classified using the updated EODHaM system, the classification provides an indication of GHCs occurring across the side, which will be used to inform the EODHaM classification once undertaken.

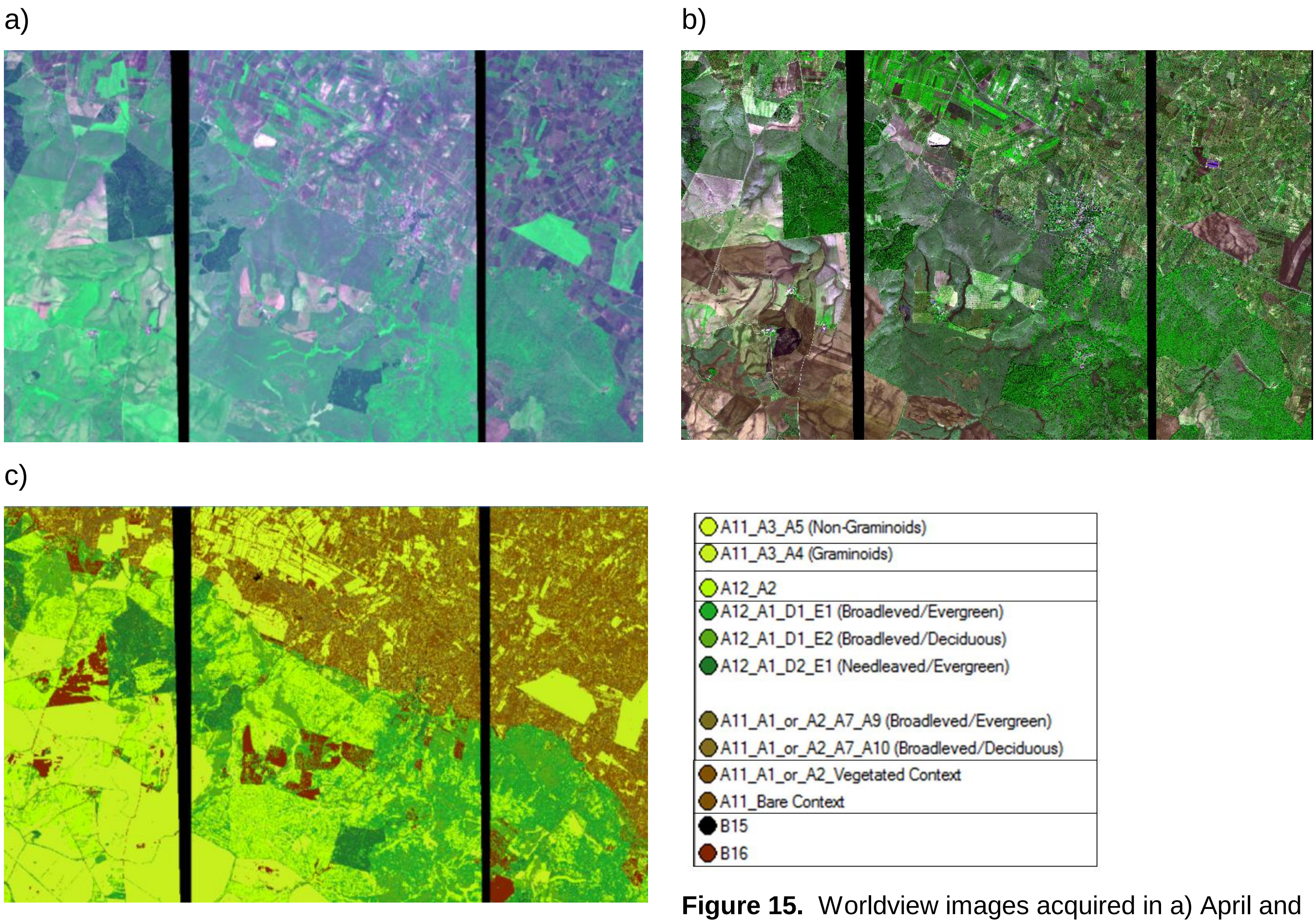


Figure 15. Worldview images acquired in a) April and b) October, 2011. c) LCCS classification undertaken in eCognition.

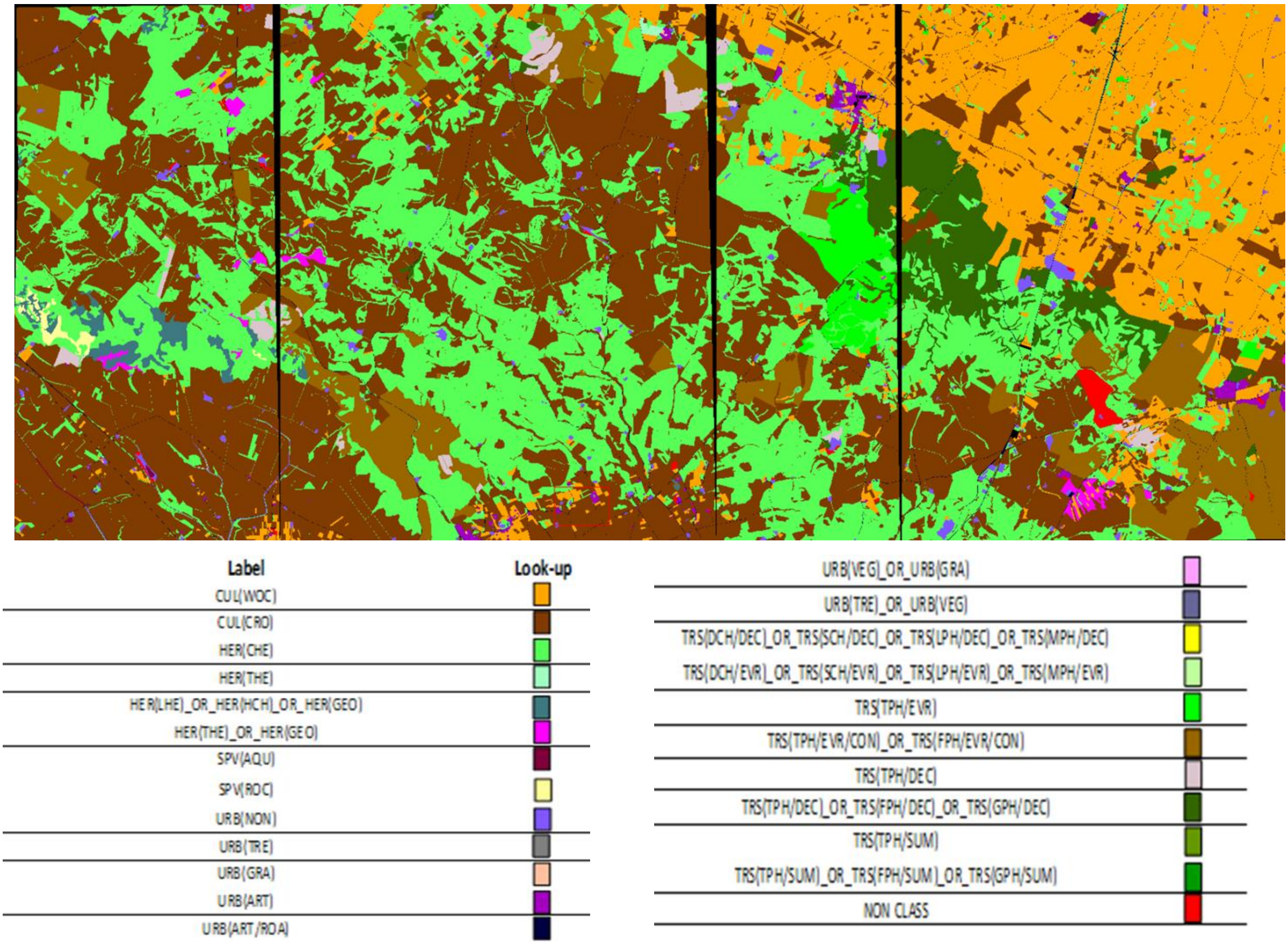


Figure 16. Preliminary GHC classification of Alta Murgia translated from the LCCS classification of the site.

5.4 Portugal

For Portugal (PT1 test site), two Worldview scenes were available from the 28th October, 2012 and the 17th June, 2017. The area observed was mountainous and the two difficulties encountered with the October image were that a) the orthorectification was insufficiently reliable and hence pixels were misaligned with those of the June image, with this attributed in part to an acquisition at a different and too large angle, and b) shadowing was common in the forest areas which compromised the classification. On this basis, the classification was undertaken using a single image (Figure 17), with this justified on the basis of a relatively small spectral change between the October and June images.

Compared to other sites, only 9 LCCS categories were defined (Table 9), with these comprised of agricultural tree plantations (primarily olive and almond groves) and vineyards. The natural/semi-natural vegetation consisted of large areas of broadleaved evergreen shrublands (< 5 m) and both broadleaved and/or evergreen woodlands (> 3 m), with these comprised of trees such as *Quercus rotundifolia*, *Juniperus oxycedrus* and *Quercus suber*. Non-vegetated surfaces were largely associated with urban infrastructure and aquatic areas (open water in the form of rivers and small seasonal ponds in the river bed).

To classify the olive and almond groves, individual or clusters of trees were extracted and described on the basis of roundness and then counts of these made within larger objects (generated using the segmentation algorithm of Bunting *et al*., 2012). These larger objects were then associated with a cultivated/managed landscape. The accuracy in the classification is being assessed using field data collected during 2012.

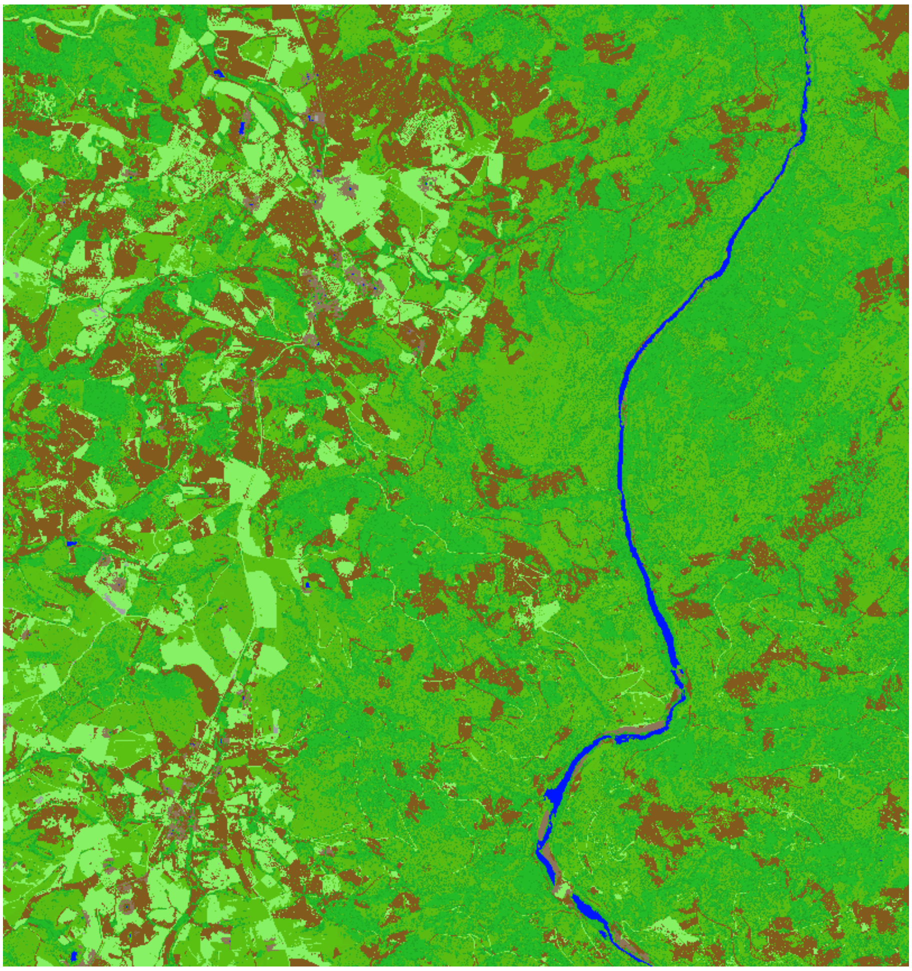


Figure 17. Classification of LCCS classes, Portugal (PT1 site)

Table 9. LCCS classes, Portugal (GHC classes also included).

Broadleaved evergreen shrubland (<0.5 - 5m)
Trees broadleaved evergreen woodland (3 - >30m)
Artificial surfaces and associated areas
Unpaved roads
Permantly cropped area of rainfed broadleaved deciduous tree crops
Permantly cropped area of rainfed broadleaved evergreen tree crops
Sparse vegetation
Natural perennial waterbodies
Cultivated and managed terrestrial

.

Field data for the BIO_SOS Portuguese Sabor test-site (PT1), intended for validation of LCCS/GHC/Annex I habitat maps and development/validation of LCCS/GHC/Annex I classification and conversion rules was collected by the ICETA/CIBIO/FCUP team members from September to November 2012. Four campaigns were performed in this period collecting a total of 200 validation points for this test-site.

Sampling locations were selected using land cover patches with an area higher than 20x20m (following EBONE protocol guidelines) identifiable from ancillary aerial photography (Bing Maps) available for the PT1 test-site as well as from WorldView-2 imagery. The sample set size was fixed to 200 due to cost and access limitations. Point records were collected at each sampled location (at patch-level) by inspecting and registering all the different LCCS classes that the patch comprised. On-site data were acquired for three different thematic legends: LCCS, EUNIS and Annex I, and were almost always accompanied by additional surveyor notes and photographic records (for further reference and validation). An additional GHC field was later added to the dataset by employing conversion rules previously devised for PT1 test-site. Point data records were acquired using two similar (in terms of features) handheld GPS devices: Garmin GPSmap 60CSx and Magellan SporTrack PRO both with WAAS/EGNOS correction. At best these devices can attain a ca. 5m positional accuracy with some fluctuations due essentially to tree cover and satellite coverage. In the final version (v4) the field validation dataset contained the following fields:

- uid01 - a unique identifier for each point containing the session number and the waypoint identifier
- lat - latitude in decimal degrees (WGS 1984 GCS)
- lon - longitude in decimal degrees (WGS 1984 GCS)
- lccs_cod - LCCS full code
- lccs_dsc - LCCS code description
- eun_annex - EUNIS code/description and Annex I habitat code (if such exists)
- GHC – general habitat category code
- obs - Field observations for the point record

5.5 Santarem, Brazil

Both Landsat and SPOT sensor data are available for mapping the land cover and habitat classes in the Tapajos National Forest and its surroundings as well as detecting change. Over the period from 1984 to 2011, 39 Landsat 5 TM images (path-row 227-062) were acquired from two databases: INPE and USGS (Table 10). The USGS database was firstly considered because the metadata are easier to analyse in order to pre-process the images and because the georeferencing is more accurate. The INPE data were acquired to complete the database for years where no data were available on the USGS server. The data were selected based on their cloud cover. All the images were pre-processed to TOA reflectance.

Table 10. List of Landsat TM data available for the Brazilian study area.

Date	Source	Scene	Date	Source	Scene
19840621	USGS	227-062	19990802	INPE	227-062
19840909	USGS	227-062	20000905	USGS	227-062
19860729	USGS	227-062	20011111	INPE	227-062
19860814	USGS	227-062	20030829	INPE	227-062
19860830	USGS	227-062	20031016	INPE	227-062
19880718	USGS	227-062	20041103	USGS	227-062
19880803	USGS	227-062	20050701	USGS	227-062
19890822	INPE	227-062	20050717	USGS	227-062
19900809	INPE	227-062	20050802	USGS	227-062
19910625	USGS	227-062	20060805	USGS	227-062
19910727	USGS	227-062	20061109	USGS	227-062
19920729	USGS	227-062	20070621	USGS	227-062
19930529	INPE	227-062	20080623	USGS	227-062
19931020	INPE	227-062	20081130	USGS	227-062
19950604	INPE	227-062	20090712	USGS	227-062
19951010	INPE	227-062	20090728	USGS	227-062
19960708	USGS	227-062	20100629	USGS	227-062
19970625	USGS	227-062	20100731	USGS	227-062
19970727	USGS	227-062	20110616	USGS	227-062
19980612	USGS	227-062			

Although Landsat images provide an opportunity to consistently monitor land cover changes over decadal periods, the 30m spatial resolution is often too coarse for monitoring small changes that occur in areas occupied by small-scale farmers, as it is the case of the riverine communities of the Tapajos National Forest. Consequently, both SPOT-4 and SPOT-% data were acquired through the SEAS Guyane antenna located in French Guyana. The SPOT5 data were acquired at a 10m spatial resolution (2.5 m in panchromatic) whereas as the SPOT4 data were initially available at 20m spatial resolution but were pan-sharpened to 10m. Up to date, 29 images were acquired (Table 11) but new acquisitions are expected since the satellite is programmed to capture new images of the study area. The dataset is basically composed of 3 sub-datasets:

Table 11. List of SPOT 4 and SPOT5 data available for the Brazilian study area.

product	sat	k/J	date	Cloud cover
10m C	Spot5	689/355	11/07/2009	21%
10m C	Spot5	689/356	11/07/2009	4%
10m C	Spot5	689/357	11/07/2009	2%
10m C	Spot5	689/358	11/07/2009	2%
2.5m BW	Spot5	689/355	11/07/2009	15%
2.5m BW	Spot5	689/356	11/07/2009	4%
2.5m BW	Spot5	689/357	11/07/2009	2%
2.5m BW	Spot5	689/358	11/07/2009	2%
10m C	Spot4	689/355	28/07/2009	14%
10m C	Spot4	689/356	28/07/2009	8%
10m C	Spot4	689/357	28/07/2009	4%

10m C	Spot4	689/358	28/07/2009	2%
10m C	Spot4	689/359	28/07/2009	1%
10m C	Spot4	690/355	27/07/2009	23%
10m C	Spot4	690/356	27/07/2009	10%
10m C	Spot4	690/358	27/07/2009	5%
2.5m BW	Spot5	689/356	14/08/2011	4%
2.5m BW	Spot5	689/357	14/08/2011	3%
2.5m BW	Spot5	689/357	14/08/2011	3%
2.5m BW	Spot5	690/355	30/07/2011	0%
2.5m BW	Spot5	690/356	30/07/2011	0%
2.5m BW	Spot5	690/357	30/07/2011	0%
2.5m BW	Spot5	690/358	30/07/2011	0%
10m C	Spot5	689/357	14/08/2011	5%
10m C	Spot5	689/358	14/08/2011	5%
10m C	Spot5	690/355	30/07/2011	0%
10m C	Spot5	690/356	30/07/2011	0%
10m C	Spot5	690/357	30/07/2011	0%
10m C	Spot5	690/358	30/07/2011	0%

a) Complete coverage of SPOT5 at 2.5m black and white (BW) and 10m colour © of the right bank of the Tapajos river (not including the interior part of the Tapajos National Forest, along the BR163 highway)

b) Complete coverage of SPOT4 pan-sharpened 10m C, with all scenes acquired on the 27th July, 2009.

c) Complete coverage of SPOT5 at 2.5m BW and 10m C of the interior part of the Tapajos National Forest (along the BR163) acquired on the 30th July, 2011. Other data of the right bank of the river acquired on the 14th August, 2011, are also available but there is no coverage for the northern part.

The Digital Elevation Model provided by SRTM is available for the entire study area at a 90m spatial resolution.

5.5.1 List of classes and their description based on ontologies

The Brazilian study area is located in the State of Para, in the Brazilian Amazon. The focal study area is the Tapajos National Forest, which is a protected area on the eastern border of the Tapajos River. Since we are also interested in the human threats for this protected area, the study area also encompasses the surroundings of the area. This study area has been chosen due to its large variety of occupation processes, which are representative of the main landscapes encountered in the Amazon. Indeed, the study area is at the crossing between the Amazon of Rivers and the Amazon of Roads. It means that both small-scale traditional agriculture and small-scale agriculture installed through public colonization programs are represented in the area. Large-scale agriculture dedicated to crops (soybean and maize) or pasture is also present. Finally, the second major city in Para (i.e., Santarem), is included in the study area.

In order to identify the classes to be mapped, land use classes mentioned in the TerraClass project (http://www.inpe.br/cra/projetos_pesquisas/terraclass.php) were considered. This project is led by Embrapa and INPE-CRA (which is the Brazilian end-user of the EODHaM system) and aims at mapping land use in cleared areas. In this project, eleven land cover classes are considered:

- “Forest”: Woody (trees) vegetation little altered or without alteration, with continuous canopy.

- “Secondary vegetation”: Woody (with shrub and trees) secondary vegetation with forest appearance and with native or exotic species. Represented per areas in natural regeneration forest process or plantation areas with native or exotic species.
- “Non Forest”: Non forest natural vegetation with characteristics of cerrado, campinas or campinaranas (savannah). (N.B.: this land cover class is not present in the study area)
- “Mechanized agriculture”: Areas with predominance of crops of annual cycles (e.g. soybean, maize, etc).
- “Agropecuaria”: Areas with agricultural activities dedicated to crops and pastures, with predominance of use of traditional technologies.
- “Clean pasture”: Pasture areas in productive process with string predominance of herbaceous vegetation.
- “Dirty pasture”: Pasture areas in productive process with predominance of herbaceous vegetation and sparse presence of shrub species.
- “Pasture with bare soil”: Pasture areas with soil partially exposed.
- “Regeneration with pasture”: Pasture areas in productive process with presence of herbaceous vegetation, shrubs and trees.
- “Urban area”: Areas with presence and concentration of buildings.
- “Mining”: Areas with evidence of mineral exploration and strong alteration of the vegetation cover.
- “Hydrography”: Waterbodies
- “Clouds”: Cloud and shadow areas that impede the observation of land cover. (N.B.: this is not a land cover class).
- “Other”: Areas that does not match any of the above definitions.

As it was the initial focus of the TerraClass project, these classes were mainly focused on the anthropic land cover classes. For the BIOSOS, distinguishing natural and semi-natural areas was also considered. For this purpose, sub classes of forests and secondary vegetation were distinguished based on the definitions proposed by IBGE. The IBGE defines five main types of forests in the classification system of Brazilian vegetation: Dense Ombrophylous Forest, Open Ombrophylous Forest, Mixed ombrophylous Forest, Semi-deciduous Seasonal Forest, and Deciduous Seasonal Forest. In the study area, only the two first classes are encountered.

- Dense Ombrophylous Forest:

This vegetation is characterized by macro- and mesophanerophytes and the presence of lianes and epiphytes that distinguish it from other forest classes. However, the main characteristic of this kind of forest is linked to its ecological environment, i.e. mean temperatures higher than 25°C and high precipitation rates all over the year (from 0 to 60 dry days).

- Open Ombrophylous Forest

This type of forest is differentiated from dense forests using rainfall information (more than 60 dry days per year). Moreover, four types of plants (palm tree, *cipo, sorococa* or *bambu*) are contained, which alter the ecological physiognomy of the dense ombrophylous forest. In the study area, the most common represented open ombrophylous forest is the Sub-mountainous Open Ombrophylous Forest with presence of palm trees (*babaçu*).

For each forest type, five sub-types have been defined according to topographic hierarchy. For the areas located between 4°N and 16°S, the topographic thresholds are:

- Alluvial vegetation is located in alluvial plains of rivers.

- Lowland vegetation from 5 to 100 m above sea level. (N.B.: mainly coastal areas)
- Sub-mountainous vegetation from 100 to 600 m above sea level.
- Mountainous vegetation from 600 to 2000 m
- High-mountainous vegetation higher than 2000 m

In the study area, the main interest is the alluvial and sub-mountainous vegetation.

With regards to secondary vegetation, two situations can be considered: secondary vegetation after pasture (*juquira*) and secondary vegetation after small crops (*capoeira*). For each class, various types can be defined basically based on the vegetation height and vegetation cover. Finally, in cultivated areas, mechanized agriculture (mainly soybean and maize) should be distinguished from small agriculture (*roça*).

Converting these classes of interest into LCCS and GHC classes highlighted a few issues, which are summarized below:

1) One of the major issue refers to the fact that "capoeira" can be considered as a sub-class of "Cultivated or managed terrestrial vegetation" (since it is part of a fallow agricultural system) or as a "natural and semi-natural vegetation". The decision was to describe it as a sub-class of "natural or semi-natural vegetation".

2) Similarly, "juquira" is a pasture so it should appear in the "cultivated and managed vegetation" class. However, it then becomes impossible to add a second layer (we can add a second crop, but its is not the case here) and a vegetation cover so it is defined as "natural and semi natural" vegetation.

3) "Capoeira" cannot be distinguished from "juquira" in the LCCS and GHC and a land cover change analysis is therefore required to allow classification.

4) Within the LCCS, vegetation heights above 14m are not differentiated. Differentiation at 30m would be useful for distinguishing old capoeira and mature forest. A similar case occurs with the GHCs, where forest phanerophytes refer to trees between 5 and 40m.

5) In the LCCS, vegetation between 3 and 7m refers trees but it might correspond to shrubs in the Amazon.

6) There are many different types of roças (with palm trees in the middle or only manioc, etc) so that it is difficult to assign a single LCCS class.

7) Alluvial forests cannot be distinguished from sub-mountainous; hence, all forests are in the same class: Broadleaved Evergreen High Trees.

10) In the GHC, two classes that are not from the same super category (i.e., CUL/TRS for pasture with trees) can be combined.

11) Pastures can be considered as both a cultivated class (CUL/CRO) and an herbaceous class (HER).

Based on this knowledge, the land cover classes and habitats to be mapped in BIOSOS are listed in Appendix II. The informal name column refers to names given internally. Maps of LCCS classes and GHCs are being generated for Santarem but with reference to the time-series of these data and will be reported in D5

5.6 Description of LCCS and GHC classes with ontologies

Data standards construction is a complex process that implies a strong community concern. A review (Pinto et al. 2004) highlights different methodologies (TOVE, Methontology) for ontology construction that rely on the reuse of top-level ontologies. These top-level ontologies guide the integration of different domain ontologies by facilitating the links between different concepts from different domain ontologies. Top-level ontology reuse is then the key of interoperability.

OBOE (Extensible Observation Ontology) (Madin et al. 2007, Madin et al. 2008, Saunders et al. 2011) is a top domain ontology used to describe scientific observations and measurements. The OBOE ontology is based on concepts of Observation, Entity, Measurement, Characteristic, Protocol and Standard (cf. **Figure 1**8). An observation is of some entity (i.e. geographic place, biological organism), and gathers a set of measurements in a unique observational event. A measurement values a characteristic of the observed entity (i.e. the field surface, the tree height). The value is also seen as an entity in the OBOE ontology, since entity concept is specialized to achieve this by the primitive value concept (integer, or string for instance). Measurements are associated to measurement standards (units) and can be enriched with additional information linked to the measurement protocol, the methods used or the measurement precision or accuracy. Moreover, an observation can be made in the context of zero or more information. Context implies dependence between observations. For instance, the observation of a specific tree leads to the measurement of its height. This observation is measured in a given geographic place that can be seen as the context of the observation of the tree. The geographic place carries important information for the observation of the tree height interpretation, since the tree height value depends on the tree environment.

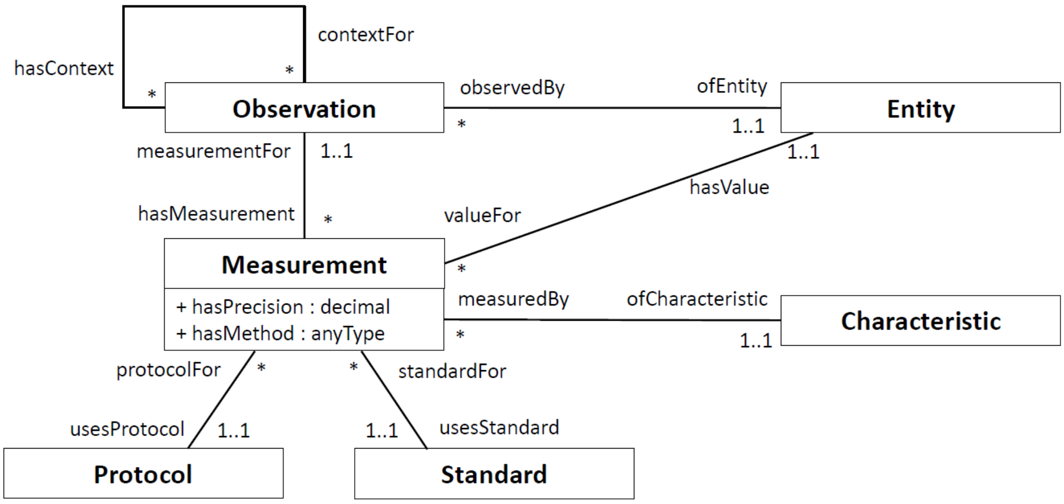


Figure 18. UML diagram of the OBOE ontology: The OBOE ontology is used to describe scientific observation and its related measurements. A given entity is observed and its characteristics are measured.

As an ontology building good practice, we chose to rely the LCCS and GHC ontologies on the OBOE ontology. In the OBOE model, a LCCS class has to be defined in terms of entity and characteristic. A LCCS class is then defined as a measurement type made during the observation event of a given area (the geographic entity) with specific characteristics (the geographic entity properties). The geographic entity observation is the context for a second observation, which is about a specific entity (the substance) described through its own characteristics (substance property). Consequently, as a first step of the building of the LCCS ontology, we extended the OBOE ontology with entities and characteristics specific to the LCCS:

- the OBOE entity class has been specialized by the geographic entity class and the substance class;
- the OBOE characteristic class has been specialized by the geographic entity property class and the substance property class;

Then, the OBOE measurement class has been extended with the LCCS class to define. Each LCCS class is defined using existing geographic entity, geographic entity property, substance and substance property.

These five classes have then been specialized. As an example, the Tree class is a subclass of the Substance class. A tree can be described by its height. The Height class is defined as a subclass of the Substance property class.

Similarly, a GHC class is also defined as a measurement type of an observation of a given substance with some properties made in the context of an observation of a geographic entity with specific properties.

The LCCS and GHC ontologies have been implemented using the LCCS and GHC classes from the Le Cesine site. Figure 19 and the Figure 20 give a list of the implemented classes. Thereafter, we illustrate the LCCS ontology with the example of the Olive Grove class, and the GHC ontology with the example of the woody crops (CUL/WOC) class. An olive grove is field with a regular geometric shape, characterized by broad-leaved evergreen woody crops (e.g., an aggregation of at least 20 olive trees of height 2-4 m, which are spaced approximately 10 m, and arranged in rows that are orthogonal and typically oriented in two directions). From March to June, herbaceous vegetation and/or bare soil can be observed depending on farming practices. The translation of this definition in description logics, Protégé syntax and OWL/XML serialization are presented in Table 12, Table 13 and Table 14. Similarly, the woody crops class definitions are illustrated on Table 15, Table 16 and Table 17.

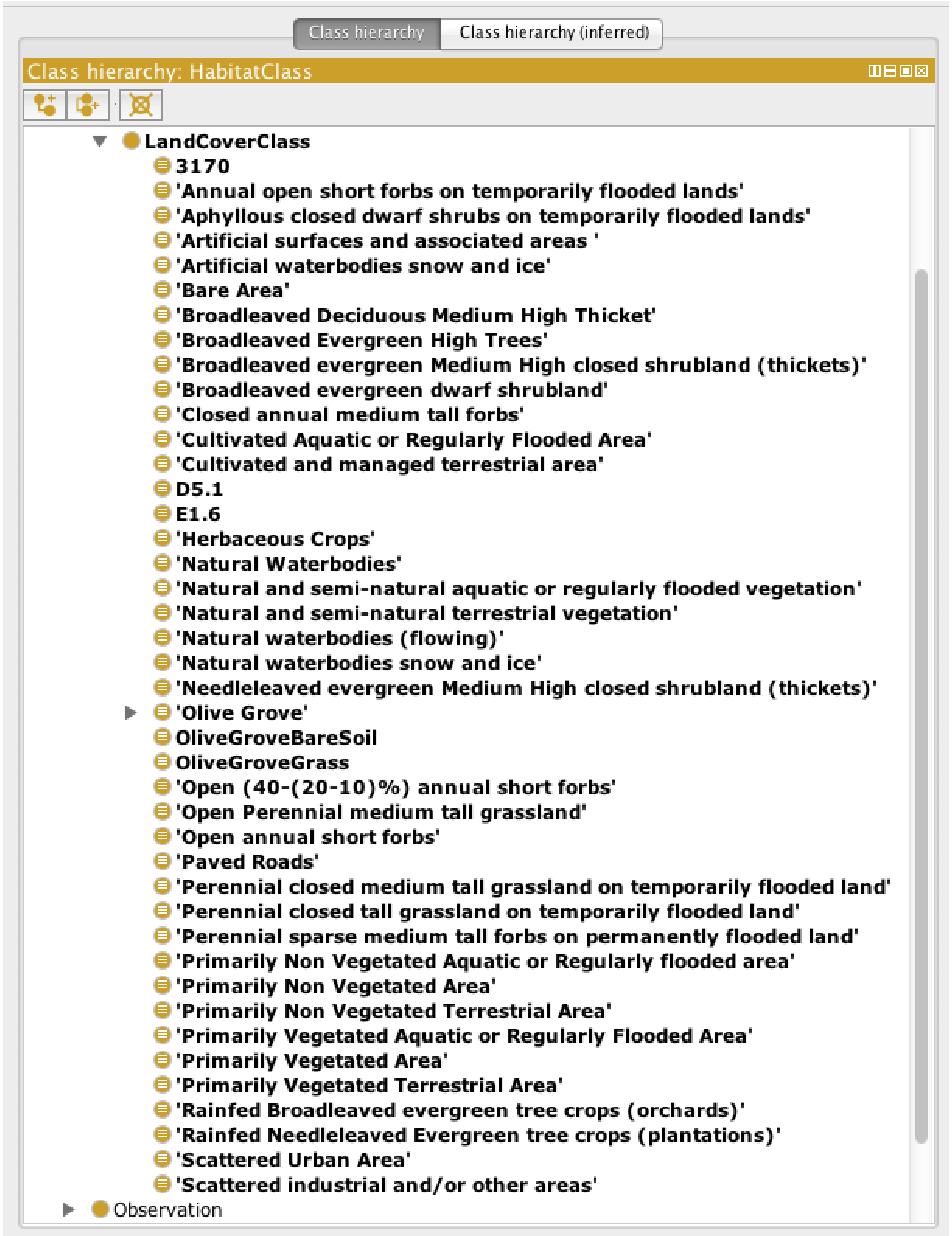


Figure 19. List of the LCCS classes from the LCCS ontology

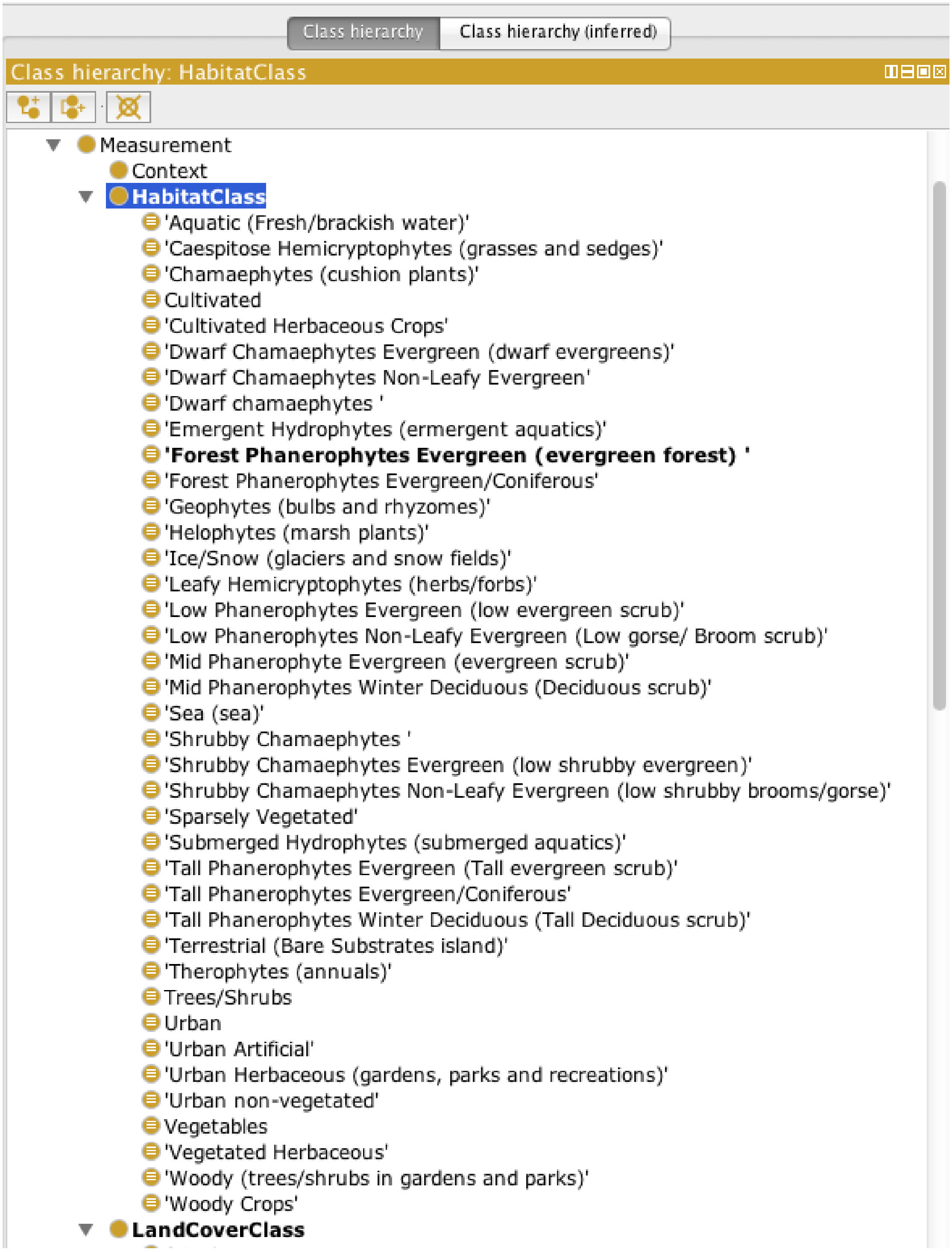


Figure 20. List of the GHC classes from the GHC ontology

Table 12. Definition of the LCCS class Olive grove using description logics

Olive Grove ≡ LCCS Class
∩ ∀ measurementFor . (
Observation ∩ ∀ ofEntity . Orchard
∩ ∀ adjacentTo . D5.1 Observation
∩ ∃ contextFor . (Observation ∩ ∀ ofEntity . (Olive Tree ∩ ∃ hasPart . (Crown ∩ ∃ hasShape . Circular))
∩ ∃ hasMeasurement . (Measurement ∩ ∀ ofCharacteristic . (Broad ∩ Evergreen ∩ (∃ hasHeightValue . >= 2 ∩ <=4)))))
∩ ∀ ofCharacteristic . (Terrestrial ∩ More than Twenty ∩ Rainfed ∩ Regular ∩ Single Crop ∩ (Medium ∪ Scattered)

Table 13. Protégé syntax of the Olive Grove Class

```
Description: 'Olive Grove'
Equivalent To
LandCoverClass
 and (measurementFor only
   (Observation
    and (contextFor some
      (Observation
       and (hasMeasurement some
         (Measurement
          and (ofCharacteristic only
            (Broad
             and Evergreen
             and (hasHeightValue some float[>= 2.0f , <= 4.0f])))))
       and (ofEntity only
         (OliveTree
          and (hasPart some
            (Crown
             and (hasShape some Circular)))))))
    and (adjacentTo some D5.1OBS1)
    and (ofEntity only Orchard)))
 and (ofCharacteristic only
   (Terrestrial
    and MoreThanTwenty
    and Rainfed
    and Regular
    and SingleCrop
    and (Medium
    or Scattered)))
```

Table 14. Abstract of the LCCS ontology: RDF description of the olive grove class (OWL/XML serialization)

```
<!-- http://www.purl.org/EspaceDev/LCCSOntology#OliveGrove -->


    <Class rdf:about="&LCCSOntology; OliveGrove ">
        <rdfs:label rdf:datatype="&rdfs;Literal">Olive Grove</rdfs:label>
        <equivalentClass>
            <Class>
                <intersectionOf rdf:parseType="Collection">
                    <rdf:Description rdf:about="&LCCSOntology;LandCoverClass"/>
                    <Restriction>
                        <onProperty rdf:resource="&oboe-core;measurementFor"/>
                        <allValuesFrom>
                            <Class>
                                <intersectionOf rdf:parseType="Collection">
                                    <rdf:Description rdf:about="&oboe-core;Observation"/>
                                    <Restriction>
                                        <onProperty rdf:resource="&oboe-core;contextFor"/>
                                        <someValuesFrom>
                                            <Class>
                                                <intersectionOf rdf:parseType="Collection">
                                                    <rdf:Description rdf:about="&oboe-core;Observation"/>
                                                    <Restriction>
                                                        <onProperty rdf:resource="&oboe-core;hasMeasurement"/>
                                                        <someValuesFrom>
                                                            <Class>
                                                                <intersectionOf rdf:parseType="Collection">
                                                                    <rdf:Description rdf:about="&oboe-core;Measurement"/>
                                                                    <Restriction>
                                                                        <onProperty rdf:resource="&oboe-core;ofCharacteristic"/>
                                                                        <allValuesFrom>
                                                                            <Class>
                                                                                <intersectionOf rdf:parseType="Collection">
                                                                                    <rdf:Description rdf:about="&LCCSOntology;Broad"/>
                                                                                    <rdf:Description rdf:about="&LCCSOntology;Evergreen"/>
                                                                                    <Restriction>
                                                                                        <onProperty rdf:resource="&LCCSOntology;hasHeightValue"/>
                                                                                        <someValuesFrom>
                                                                                            <rdfs:Datatype>
                                                                                                <onDatatype rdf:resource="&xsd;float"/>
                                                                                                <withRestrictions rdf:parseType="Collection">
                                                                                                    <rdf:Description>
                                                                                                        <xsd:maxInclusive
rdf:datatype="&xsd;float">4.0</xsd:maxInclusive>
                                                                                                    </rdf:Description>
                                                                                                    <rdf:Description>
                                                                                                        <xsd:minInclusive
rdf:datatype="&xsd;float">2.0</xsd:minInclusive>
```

```
                                </rdf:Description>
                              </withRestrictions>
                            </rdfs:Datatype>
                          </someValuesFrom>
                        </Restriction>
                      </intersectionOf>
                    </Class>
                  </allValuesFrom>
                </Restriction>
              </intersectionOf>
            </Class>
          </someValuesFrom>
        </Restriction>
        <Restriction>
          <onProperty rdf:resource="&oboe-core;ofEntity"/>
          <allValuesFrom>
            <Class>
              <intersectionOf rdf:parseType="Collection">
                <rdf:Description rdf:about="&LCCSOntology;OliveTree"/>
                <Restriction>
                  <onProperty rdf:resource="&LCCSOntology;hasPart"/>
                  <someValuesFrom>
                    <Class>
                      <intersectionOf rdf:parseType="Collection">
                        <rdf:Description rdf:about="&LCCSOntology;Crown"/>
                        <Restriction>
                          <onProperty rdf:resource="&LCCSOntology;hasShape"/>
                          <someValuesFrom rdf:resource="&LCCSOntology;Circular"/>
                        </Restriction>
                      </intersectionOf>
                    </Class>
                  </someValuesFrom>
                </Restriction>
              </intersectionOf>
            </Class>
          </allValuesFrom>
        </Restriction>
      </intersectionOf>
    </Class>
  </someValuesFrom>
</Restriction>
<Restriction>
  <onProperty rdf:resource="&LCCSOntology;adjacentTo"/>
  <someValuesFrom rdf:resource="&LCCSOntology;D5.1OBS1"/>
</Restriction>
```

```
                        <Restriction>
                            <onProperty rdf:resource="&oboe-core;ofEntity"/>
                            <allValuesFrom rdf:resource="&LCCSOntology;Orchard"/>
                        </Restriction>
                    </intersectionOf>
                </Class>
            </allValuesFrom>
        </Restriction>
        <Restriction>
            <onProperty rdf:resource="&oboe-core;ofCharacteristic"/>
            <allValuesFrom>
                <Class>
                    <intersectionOf rdf:parseType="Collection">
                        <rdf:Description rdf:about="http://www.purl.org/EspaceDev/GHCOntology#Terrestrial"/>
                        <rdf:Description rdf:about="&LCCSOntology;MoreThanTwenty"/>
                        <rdf:Description rdf:about="&LCCSOntology;Rainfed"/>
                        <rdf:Description rdf:about="&LCCSOntology;Regular"/>
                        <rdf:Description rdf:about="&LCCSOntology;SingleCrop"/>
                        <Class>
                            <unionOf rdf:parseType="Collection">
                                <rdf:Description rdf:about="&LCCSOntology;Medium"/>
                                <rdf:Description rdf:about="&LCCSOntology;Scattered"/>
                            </unionOf>
                        </Class>
                    </intersectionOf>
                </Class>
            </allValuesFrom>
        </Restriction>
    </intersectionOf>
  </Class>
 </equivalentClass>
 <rdfs:subClassOf rdf:resource="&LCCSOntology;LandCoverClass"/>
</Class>
```

Table 15. Definition of the LCCS class Woody Crops using description logics

$$\text{Woody crops} \equiv \text{Habitat Class}$$

$$\cap \ \forall \ \text{measurementFor} \ . \ ($$

$$\text{Observation} \cap \forall \ \text{ofEntity} \ . \ \text{Cultivated Area}$$

$$\cap \ \exists \ \text{contextFor} \ . \ (\text{Observation} \cap \forall \ \text{ofEntity} \ . \ \text{Woody}))$$

$$\cap \ \forall \ \text{ofCharacteristic} \ . \ \text{Cultivated and Managed}$$

Table 16. Protégé syntax of the Woody Crops Class

```
Description: 'Woody Crops'

Equivalent To
  HabitatClass
   and (measurementFor only
     (Observation
      and (contextFor some
        (Observation
         and (ofEntity only Woody)))
      and (ofEntity only CultivatedArea)))
   and (ofCharacteristic only CultivatedAndManaged)
```

Table 17. Abstract of the GHC ontology : RDF description of the woody crops class (OWL/XML serialization)

```
<!-- http://www.purl.org/EspaceDev/GHCOntology#CUL/WOC -->

  <Class rdf:about="&GHCOntology;CUL/WOC">
    <rdfs:label rdf:datatype="&rdfs;Literal">Woody Crops</rdfs:label>
    <equivalentClass>
      <Class>
        <intersectionOf rdf:parseType="Collection">
          <rdf:Description rdf:about="&GHCOntology;HabitatClass"/>
          <Restriction>
            <onProperty rdf:resource="&oboe-core;measurementFor"/>
            <allValuesFrom>
              <Class>
                <intersectionOf rdf:parseType="Collection">
                  <rdf:Description rdf:about="&oboe-core;Observation"/>
                  <Restriction>
                    <onProperty rdf:resource="&oboe-core;contextFor"/>
                    <someValuesFrom>
                      <Class>
```

```
                        <intersectionOf rdf:parseType="Collection">
                            <rdf:Description rdf:about="&oboe-core;Observation"/>
                            <Restriction>
                                <onProperty rdf:resource="&oboe-core;ofEntity"/>
                                <allValuesFrom rdf:resource="&GHCOntology;Woody"/>
                            </Restriction>
                        </intersectionOf>
                    </Class>
                </someValuesFrom>
            </Restriction>
            <Restriction>
                <onProperty rdf:resource="&oboe-core;ofEntity"/>
                <allValuesFrom rdf:resource="&GHCOntology;CultivatedArea"/>
            </Restriction>
        </intersectionOf>
    </Class>
</allValuesFrom>
</Restriction>
<Restriction>
    <onProperty rdf:resource="&oboe-core;ofCharacteristic"/>
    <allValuesFrom rdf:resource="&GHCOntology;CultivatedAndManaged"/>
</Restriction>
</intersectionOf>
</Class>
</equivalentClass>
<rdfs:subClassOf rdf:resource="&GHCOntology;HabitatClass"/>
</Class>
```

Since both LCCS and GHC ontologies are based on the OBOE ontology, the building of bridges between the two ontologies is facilitated. Consequently, conversion from a LCCS class to a GHC class and reciprocally can be performed automatically thanks to the use of a reasoner. The mapping between the two ontologies results from the manual alignment of both entities and characteristics from the LCCS ontology and from GHC ontology. Two kinds of such alignments have been used: the equivalent class link and the subclass link. For instance, about the mapping between the Orchard class from the LCCS and the Cultivated area class from the GHC, we manually created a subclass link between these classes in order to imply that an orchard is a cultivated area. In the same way, an olive tree has been defined as a subclass of woody. Once all the mappings have been added, an inferred hierarchy results from the reasoning mechanism apply on both ontologies. This inferred hierarchy groups all the LCCS classes and the GHC classes in a single “new ontology” (cf. Figure 4) that allows the automatic transformation from a given class of the LCCS in its equivalent one in the GHC, and reciprocally.

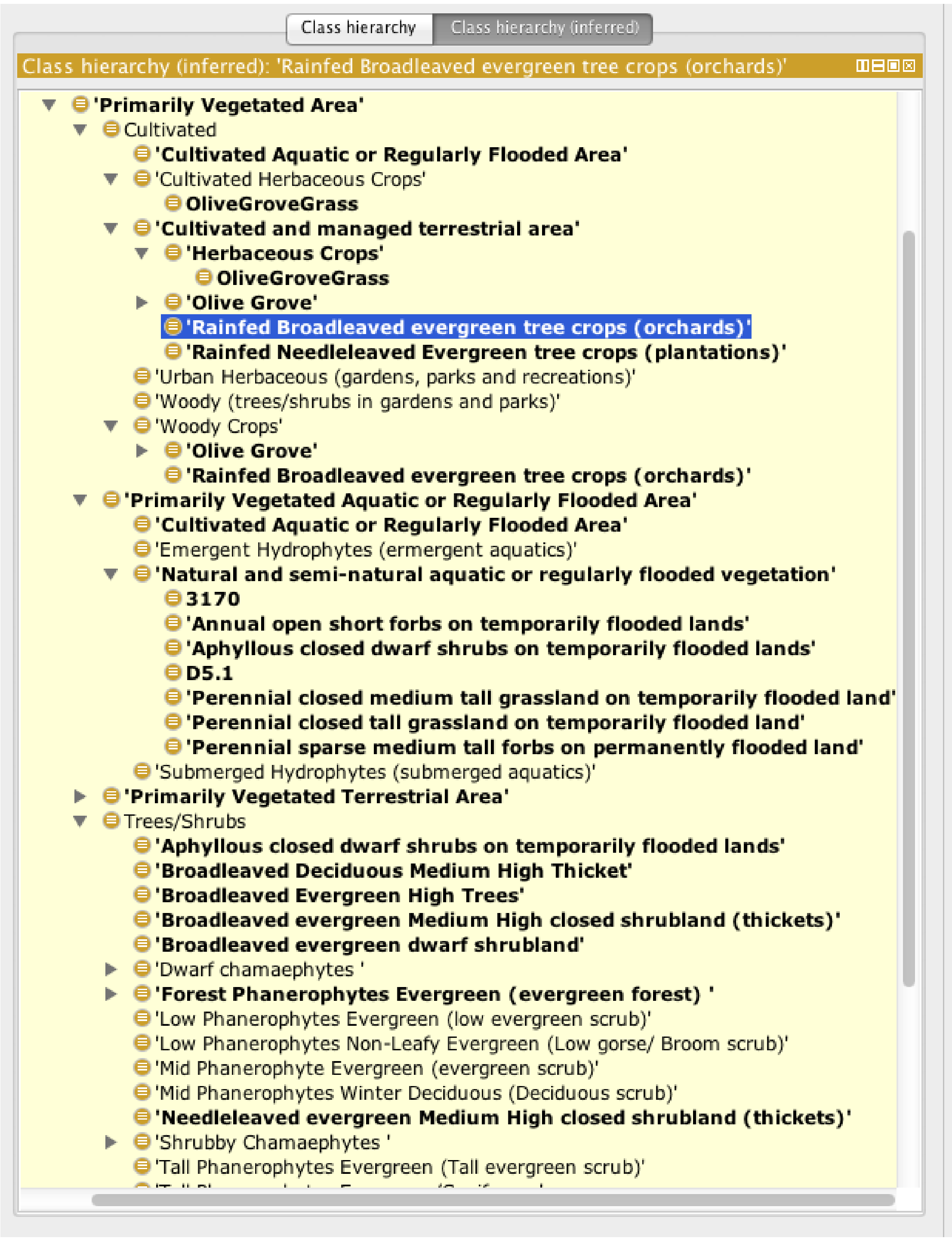


Figure 21. Part of the inferred hierarchy of both LCCS and GHC classes after the use of the reasoner. The transformation from a LCCS class (bold classes) to a GHC class is then automatic.

6 Summary and conclusions

Within D5.5, significant advances have been made in the development of the EODHaM system. More specifically:

a) Advances in feature extraction have allowed automated detection of buildings and temporary structures (e.g., caravans), hedgerows, tree crowns and field boundaries. Textural analyses have also allowed for the estimation of forest height where LiDAR data are not available. A new segmentation algorithm (Shepherd et al., 2012) allows the remaining landscape to be subdivided into discrete units that reflect the distribution of spectrally distinct areas.

b) The division of objects into 'small' and 'large' and the association of each with a descriptor (e.g., length width ratio, roundness, border to neighbours) prior to classification to land covers has allowed better discrimination of areas that are cultivated, managed or artificial.

c) The classification of LCCS Levels 1-2 in the EODHaM 1st stage has been refined through comparisons between sites such that the indices used are more consistent and focus on the classification of vegetative stages (e.g., photosynthetic/non-photosynthetic) and surface moisture.

d) For classification of LCCS Level 3, less reliance is placed on existing cadastral and infrastructure layers, although these can still be incorporated.

e) For the classification of LCCS classes beyond Level 3, methods for discriminating components of the system (e.g., lifeform, water dynamics) have been refined.

f) A new image format, which allows for the creation of and processing within a raster attribute table has been used. The new format has facilitated the implementation of the EODHaM classification for all sites and is also open source.

g) The sequence of processing has been programmed within open source software to allow greater and free use by the user community.

h) Approaches to validation of the classification of both land covers and habitats have been advanced, with final assessments to be provided in D5.6.

i) The classification scheme can be easily modified (to be undertaken in D5.6) to allow from changes in land covers and habitats from one class to another.

D5.5 represents a major advance towards the generation of the EODHAM system, with the complete processing chain for classification and translation of LCCS to GHC and subsequently Annex I classes placed into a user friendly, intuitive and open source environment.

7 References

Ahmadzadeh, M. R., 2001. Reasoning with uncertainty in remote sensing. Ph.D. dissertation. Dept. Electronic Engineering, Information Technology and Mathematics, University of Surrey, Surrey.

Breiman, L., J. Friedman, C.J. Stone and R.A. Olshen, 1984. Classification and Regression Trees. Chapman & Hall/CRC. Boca Raton, FL.

Bunce, R.G.H., Bogers, M.M.B., Roche, P., Walczak, M., Geijzendorffer I.R. and Jongman, R.H.G. (2011). Manual for Habitat and Vegetation Surveillance and Monitoring: Temperate, Mediterranean and Desert Biomes, Wageningen, NL.

Bunting P., Armston., J., Lucas R.M., and Clewley D. (2012a) Sorted Pulse Data (SPD) Library. Part I: A generic file format for LiDAR data from pulsed laser systems in terrestrial environments. Geoscience and Computing. Submitted.

Bunting P., Armston., J., Clewley D., and Lucas R.M. (2012b) Sorted Pulse Data (SPD) Library. Part II: A processing framework for LiDAR data from pulsed laser systems in terrestrial environments. Geoscience and Computing. Submitted.

Bunting, P. Clewley, D. and Lucas R.M., (2012). The Remote Sensing and GIS Software Library. Computers and Geosciences. In Prep.

Bunting P. and Gillingham (2012) The KEA image format. Computers and Geosciences. Submitted.

Comaniciu, D. and Meer, P., (2002). Mean shift: a robust approach toward feature space analysis," Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.24, no.5, pp.603-619.

Dempster, A.P. (1967). Upper and lower probabilities induced by a multivalued mapping. Ann. Math. Stat. 38(2): 325-339.

Gillingham S. and Flood N. (2011) The Raster I/O Simplification (RIOS) python library. https://bitbucket.org/chchrsc/rios/overview (Accessed Dec 2012).

Gillingham S., Flood N., Bunting P., Shepherd J.D., Roudlier P., and Gill T. (2012) "A Simple raster viewer for use by nerdy types" https://bitbucket.org/chchrsc/viewer (Accessed Dec 2012).

Grizonnet, M., Fontannaz, D., Nasser, G. and Mangin, A. (2012). Study of coastal monitoring indicators from Pleiades-like data: Detection of boats mooring areas and coastline monitoring, Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International , vol., no., pp.7106-7109, 22-27 July 2012

Gruijter, J. d., D. Brus, M. Bierkens, and M. Knotters. 2006. Sampling for Natural Resource Monitoring. Springer-Verlag, Berlin Heidelberg.

Hau, H. and Kashyap, R., 1990. Belief combination and propagation in a lattice-structured inference network. IEEE Trans. Syst., Man, Cybern., 20(1): 45-58.

Inglada, J. and Christophe, E. (2009). The Orfeo Toolbox remote sensing image processing software," Geoscience and Remote Sensing Symposium, 2009, IEEE International, IGARSS 2009 , vol.4, no., pp.IV-733-IV-736, 12-17 July 2009.

JCGM/WG 1, 2008. Evaluation of measurement data – Guide to the expression of uncertainty in measurement, Joint Committee for Guides in Metrology.

Petrou M. and Garcia-Sevilla P., (2006), Image processing: Dealing with texture, John Wiley editors ISBN-10 0-470-02628-6.

Petrou, Z., C. Tarantino, M., Adamo, P., Blonda and M. Petrou, 2012. Estimation of Vegetation Height through Satellite Image Texture Analysis. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume 39-B8, XXII ISPRS Congress, Melbourne, Australia, pp. 321–326.

Shafer, G. (1976). A Mathematical Theory of Evidence. Princeton University Press. Princeton, NJ.

Shepherd, J. D., & Dymond, J. R. (2003). Correcting satellite imagery for the variance of reflectance and illumination with topography. *International Journal of Remote Sensing*, *24*(17), 3503–3514. doi:10.1080/01431160210154029

Shepherd J. D., Bunting P., Dymond J.R. (2012) Segmentation of imagery based on iterative elimination. Remote Sensing. Submitted.

8 Appendixes

8.1 Appendix I: The generic colour map for the illustration of the a) LCCS and b) GHC maps.

LCCS Maps

	A11-Cultivated terrestrial vegetation	evergreen	needleleaved	orchards	trees
				plantation	
			broadleaved	orchards	
				plantation	
		deciduous	needleleaved	orchards	
				plantation	
			broadleaved	orchards	
				plantation	
		evergreen	needleleaved	orchards	shrubs
				plantation	
			broadleaved	orchards	
				plantation	
		deciduous	needleleaved	orchards	
				plantation	
			broadleaved	orchards	
				plantation	
		herbaceous			
		forbs			
		graminoids			
		parks			
		parklands			
		lawns			

	A12-Natural terrestrial vegetation				woody
		aphyllous		high	trees
				medium	
				low	
				high	shrubs
				medium-high	
				medium	
				dwarf	
		needleleaved	evergreen	high	trees
				medium	
				low	
				high	shrubs
				medium-high	
				medium	
				dwarf	

			deciduous	high	trees
				medium	
				low	
				high	shrubs
				medium-high	
				medium	
				dwarf	
		broadleaved	evergreen	high	trees
				medium	
				low	
				high	shrubs
				medium-high	
				medium	
				dwarf	
			deciduous	high	trees
				medium	
				low	
				high	shrubs
				medium-high	
				medium	
				dwarf	
		herbaceous	annual	high	
				medium-high	
				medium	
				dwarf	
			Perennial	high	
				medium-high	
				medium	
				dwarf	
		forbs	annual	high	
				medium-high	
				medium	
				dwarf	
			Perennial	high	
				medium-high	
				medium	
				dwarf	
		graminoids	annual	high	
				medium-high	
				medium	
				dwarf	
			Perennial	high	
				medium-high	
				medium	
				dwarf	

		lichens/mosses			
		lichens			
		mosses			

	A23-cultivated aquatic	woody
		forbs
		graminoids

	A24-Natural aquatic vegetation	aphyllous		high	trees
				medium	
				low	
				high	shrubs
				medium-high	
				medium	
				dwarf	
		needleleaved	evergreen	high	trees
				medium	
				low	
				high	shrubs
				medium-high	
				medium	
				dwarf	
			deciduous	high	trees
				medium	
				low	
				high	shrubs
				medium-high	
				medium	
				dwarf	
		broadleaved	evergreen	high	trees
				medium	
				low	
				high	shrubs
				medium-high	
				medium	
				dwarf	
			deciduous	high	trees
				medium	
				low	
				high	shrubs
				medium-high	
				medium	
				dwarf	
		herbaceous	annual	high	

				medium-high	
				medium	
				dwarf	
			Perennial	high	
				medium-high	
				medium	
				dwarf	
		forbs	free-floating	annual	high
					medium-high
					medium
					dwarf
				Perennial	high
					medium-high
					medium
					dwarf
			rooted	annual	high
					medium-high
					medium
					dwarf
				Perennial	high
					medium-high
					medium
					dwarf
		graminoids	annual	high	
				medium-high	
				medium	
				dwarf	
			Perennial		
				medium-high	
				medium	
				dwarf	
		lichens/mosses			
		lichens			
		mosses			

	B15 - Artificial Terrestrial	Built up	linear	roads	paved
					unpaved
				railways	
				com lines pipelines	
			nonlinear	industrial	
				urban	high dens
					medium dens
					low dens
					scattered

		non built up	waste dumps		
			extraction sites		

	B16-Bare Areas	consolidated	bare rock & fragments	
			rock	
			gravel stones boulders	
			gravel	
			stones	
			boulders	
			hardpans	
		unconsolidated	bare soil and other	stony
				very stony
			loose & shifting sands	stony
				very stony

	B27 - Artificial aquatic non vegetated	water	flowing
			standing
		snow	
		ice	

	B28 - Natural aquatic non vegetated	water	flowing
			standing
		snow	
		ice	

GHC maps

	TRS	NLE		GPH
				FPH
				TPH
				MPH
				LPH
				SCH
				DCH
		CON	EVR	GPH
				FPH
				TPH
				MPH
				LPH
				SCH
				DCH
			DEC	GPH
				FPH
				TPH
				MPH
				LPH
				SCH
				DCH
			EVR	GPH
				FPH
				TPH
				MPH
				LPH
				SCH
				DCH
			DEC	GPH
				FPH
				TPH
				MPH
				LPH
				SCH
				DCH

	HER	LHE	
		THE	
		GEO	
		CHE	
		HCH	
		CRY	
		EHY	BRY

			LIC
		SHY	FLO
			LEA
		HEL	

	CUL	SPA
		CRO
		WOC

	URB	NON
		ART
		GRA
		VEG
		TRE

	SPV	SEA
		TID
		AQU
		ICE
		ROC
		BOU
		STO
		GRV
		SAN
		EAR

8.2 Appendix II: Common land covers observed at the study sites.

Table A1a. LCCS categories and General Habitat Categories, Cors Fochno, mid Wales.

Cat.	LCCS Code Modifer	Description	GHC
A11	A3.A4.B1.B5.C1.D1.D9_B4	Permanently cropped area: Graminoid crops	CUL/CRO or URB/GRA
A11	A1.B1.B5.C1.D1.D9_A8.B4	Permanently cropped area with rainfed needle-leaved tree crops (plantations).	CUL/WOC or URB/TRE or TRS/TPH/(EVR, DEC)/CON or TRS/FPH/(EVR,DEC)/CON
A11	A1.B1.B5.C1.D1.D9_A7.B4	Permanently cropped area with rainfed broad-leaved tree crops (plantations).	CUL/WOC or URB/TRE or TRS/TPH/DEC or TRS/FPH/DEC
A12	A1.A3.A10.B2.C2.D1.E2.B5	Broad-leaved deciduous fragmented high trees	URB/TRE or TRS/TPH/DEC or TRS/FPH/DEC
A12	A1.A4.A11.B3.C2.D1.E2.B14	Broad-leaved deciduous medium to high shrubland	URB/TRE or TRS/MPH/DEC, or TRS/TPH/DEC
A12	A1.A4.A11.B3.C2.D1.E1	Broad-leaved Evergreen Fragmented Shrubland single layer.Heathland (uplands)	TRS/SCH/EVR
A12	A2.A6.A10.B4.C1.E5_B12.E6	Closed Perennial Medium T all Grassland (e.g., *Molinia/Juncus)*	HER/CHE
A12	A2.A6.A11.B4.XX.E5_A12.B12.E6	Open ((70-60)-40 %) Perennial Medium Tall Grassland (e.g., *Eriophorum)*	HER/CHE
A12	A2.A6.A10.B4.C2.E5_B13	Closed short grassland	HER/CHE
A12	A2.A5.A10.B4_B11	Closed medium tall forbs (3.0-0.8 m)	HER/LHE
A12	A2.A5.A10.B4_B12	Closed medium tall forbs (0.8-0.3 m)	HER/LHE
A24	A1.A4.A20.B3.C1.D1.E1 .F2.F4.F7.G4_C4	Closed to Open Broad-leaved Evergreen Shrubs with Herbaceous Vegetation on Permanently Flooded Land (Persistent) (Active Bog)	TRS/DCH/EVR or TRS/SCH/EVR or TRS/LPH/EVR, or TRS/MPH-EVR HER/EHY/HEL/SHY-FLO/LEA)
A24	A2.A6.A12.B4.C1.E5_B11.C4.E6	Perennial closed tall grassland on permanently flooded land (persistent)	HER/HEL
A24	A2.A6.A13.B4.C1_B13.C5	Open short grassland on permanently flooded land (with daily variations) (Unmanaged *Saltmarsh)*	HER/HEL

Table A1b. LCCS categories and General Habitat Categories, Cors Fochno, mid Wales.

Cat.	LCCS Code Modifer	Description	GHC
B15	A3_A8	Paved road(s)	URB/ART
B15	A3_A10	Railway(s)	URB/ART
B15	A4_A13	Urban areas	URB/ART/NON
B16	A3_A7	Bare rock	SPV/ROC
B16	A6.B6	Shifting Sands.Saturated Parabolic Dunes	SPV/SAN
B16	A6_A12	Stony loose and shifting sands	SPV/STO
B16	A5_A13	Very stony bare soil and unconsolidated material(s)	SPV/STO/GRV
B27	A1.B1.C2.D1.A5	Clear shallow artificial waterbody (Standing)	SPV/AQU
B27	A1.B1.C1_A4	Turbid Deep to Medium Deep Artificial Perennial waterbodies (Flowing)	SPV/AQU
B27	A1.B1.C1_A5	Deep to Medium Perennial Artificial Waterbodies (Standing)	SPV/AQU
B28	A1.B1.C1_A5	Deep to Medium Perennial Natural Waterbodies (Standing)	SPV/AQU
B28	A1.B3_A4.B6	Tidal Area (Flowing); Surface Aspect (sand)	SPV/AQU(TID)
B28	A1_A4	Natural waterbodies, flowing (ocean/sea)	SPV/AQU/SEA

Table A2a. LCCS categories and General Habitat Categories, the Netherlands.

Cat.	LCCS Code Modifer	Description	GHC
A11	(A3.)A4.B1.B5.C1D1.D9	Permanently cropped area Graminoid crops	HER/CHE or HER/LHE/CHE
A11	(A3.)A4.B1.B5.C1.D1.D9-B4-S7	Permanently cropped area Graminoid crops, Crop Type: Fodder	CUL/CRO
A11	(A3.)A5.B1.B5.C1.D1.D9-B4-S4	Permanently cropped area Non-Graminoid crops, Crop Type: Roots and Tubers	CUL/CRO
A11	(A3.)A5.B1.B5.C1.D1.D9-B4-S9	Permanently cropped area Non-Graminoid crops, Crop Type: Industrial Crops	CUL/CRO
A11	(A3.)A5.B1.B5.C1.D1.D9-B4-S3	Permanently cropped area Non-Graminoid crops, Crop Type: Cerals	CUL/CRO
A11	(A3.)A5.B1.B5.C1.D1.D9-B4	Permanently cropped area Non-Graminoid crops	CUL/CRO
A11	A1.B1.B5.C1.D1.D9-W8	Permanently Cropped Area with small sized field of rainfed tree crops, crop cover: Orchards	CUL/WOC
A12	A3.A10.B2.C1.D1.E2.F2.F6.F7.G3.F1-B5F9G8	Broad-leaved Deciduous high trees with open high shrubs	TRS/FPH/DEC
A12	A3.A10.B2.C1.D2.E1.F2.F6.F7.G3.F1-B5F9G8	Needle-leaved evergreen high trees with open high shrubs	TRS/FPH/CON
A11	A6.A11	Vegetated Urban Area(s) / Park(s)	URB/TRE
A11	A6.A11	Vegetated Urban Area(s) / Park(s)	URB/TRE
A11	A6.A13	Vegetated Urban Area(s) / Lawn(s)	URB/GRA
A11	A2-A6	Extraction site	URB/NON
A11	A6.A13	Vegetated Urban Area(s) / Lawn(s)	URB/GRA
A12	A4.A10.B3.C1.D1.E1.F1-B10	Broad-leaved Evergreen Dwarf Thicket, Single Layer	TRS/SCH or TRS/LPH
A12	A4.A10.B3.C1.D1.E1.F1-B10	Broad-leaved Evergreen Dwarf Thicket, Single Layer	TRS/SCH or TRS/LPH
A12	A6.A10.B4.C2.XXXXF1-B12	Interrupted closed medium tall grassland, single layer	HER/CHE
A24	A6.A12.B4.C3.XXXXF1-B11	Closed Tall grassland on waterlogged soil	HER/HEL
A24	A3.A12.B2.C1.D1.E2.F1	Mixed medium high trees on permanently flooded land (persistent)	TRS/TPH or TRS/FPH
A12	A6.A10.B4.C1.XXXXF1-B12	Closed medium tall grassland, single layer	HER/CHE or HER/LHE/CHE
A11	A1.B1.B5.C1.D1.D9-W8	Permanently Cropped Area with small sized field of rainfed tree crops, crop cover: Orchards	CUL/WOC
A11	A1.B1.B5.C1.D1.D9-W8	Permanently Cropped Area with small sized field of rainfed tree crops, crop cover: Orchards	CUL/WOC
A12	A2.A6.A10.B4.C2.E5_B13	Closed short grassland	HER/CHE

Table A2b. LCCS categories and General Habitat Categories, the Netherlands.

Cat.	LCCS Code Modifer	Description	GHC
B27	A1.B1C1.D1-A4	Clear deep to medium deep artificial perennial watrebodies (flowing)	SPV/AQU
B15	A4_A13	Urban areas	URB/ART
B15	A4-A12A15	Medium density industrial and or other areas	URB/ART or URB/ART/GRA or URB/ART/TRE
B15	A3-A7	Road(s)	URB/ART
B15	A4-A13A17	Scattered Urban Areas(s)	URB/ART
B16	(B16).A6	Loose and shifting sands	SPV/SAN

Table A3. LCCS categories occurring within Le Cesine, Italy.

Cat.	LCCS Code Modifer	Description	GHC
A11	A3	Herbaceous crops	CUL/CRO or URB/VEG or URB/GRA or HER/LHE or HER/THE or HER/HCH or HER/GEO
A11	A1.B1.C1.D1.W7.A8.A9.B3	Monocultured fields of rainfed evergreen needle-leaved tree crops (plantations).	URB/TRE or TRS/TPH/EVR/CON or TRS/FPH/EVR/CON
A11	A1.B1.C1.D1.W8.A7.A9.B4	Monoculture fields of rainfed broad-leaved tree crops orchards (olive groves)	CUL/WOC or URB/TRE
A12	A1.A4.A10.B3.D1.E2.B9	Broad-leaved deciduous medium/high closed shrubland (thickets)	URB/TRE or TRS/MPH/DEC or TRS/TPH/DEC
A12	A1.A4.A10.B3.D2.E1.B9	Needleaved evergreen medium/high closed shrubland (thickets)	URB/TRE or TRS/MPH/EVR/CON or TRS/TPH/EVR/CON
A12	A1.A4.A11.B3.D1.E1.B10	Broad-leaved evergreen open dwarf shrublands	URB/VEG or TRS/DCH/EVR or TRS/SCH/EVR or TRS/LPH/EVR
A12	A1.A4.A10.B3.D1.E1.B9	Broad-leaved evergreen medium/high closed shrubland (thickets)	URB/TRE or TRS/MPH/EVR or TRS/TPH/EVR
A12	A2.A6.A11.B4.E5.B12.E6	Open Perennial Medium Tall Grassland	HER/CHE or URB/GRA
A12	A2.A5.A11.B4.E5.A13.B13.E7	Open (40-(20- 10)%) annual short forbs	URB/GRA or HER/THE or HER/LHE or HER/HCH or HER/GEO or weak TRS/DCH or weak TRS/SCH
A12	A2.A5.A11.B4.E5.B13.E7	Open annual short forbs	URB/GRA or HER/THE or HER/LHE or HER/HCH or HER/GEO or weak TRS/DCH or weak TRS/SCH
A12	A2.A5.A10.B4.E5.B12.E7	Closed annual medium/tall forbs	URB/GRA or HER/THE or HER/LHE or HER/HCH or HER/GEO
A24	A2.A5.A13.B4.C2.E5.B13.E7	Open annual short herbaceous vegetation on temporarily flooded land	URB/GRA or HER/EHY or HER/SHY-FLO or HER/HEL or HER/SHY/LEA
A24	A1.A4.A12.B3.C2.D3.B10	Aphyllous closed dwarf shrubs on temporarily flooded land	URB/VEG or TRS/SCH/NLE or TRS/LPH/NLE or TRS/DCH/NLE
A24	A2.A6.A12.B4.C2.E5.B11.E6	Perennial closed tall (3-0.8m) grasslands on temporarily flooded land	URB/GRA or HER/EHY or HER/SHY or HER/HEL
A24	A2.A6.A12.B4.C2.E5.B12.E6	Perennial closed medium-tall (0.8- 0.3m) grasslands on temporarily flooded land	URB/GRA or HER/EHY or HER/SHY or HER/HEL
A24	A2.A5.A16.B4.C1.E5.A15.B12.E6	Perennial sparse medium tall herbaceous vegetation on permanently flooded land	URB/GRA or HER/EHY or HER/SHY-FLO or HER/HEL or HER/SHY/LEA
B15	A1.A3.A7.A8	Paved road(s)	URB/ART-ROA
B15	A1.A4.A13.A17	Scattered industrial or other areas	URB/NON
B15	A1	Natural waterbodies, flowing (ocean/sea)	SPV/AQU/SEA

Table A4. List of LCCS and GHC categories, Alta Murgia, Italy.

Cat.	LCCS Code Modifier	Description	GHC
A11	A3	Herbaceous crops	CUL/CRO or URB/VEG or URB/GRA or HER/LHE or HER/THE or HER/HCH or HER/GEO
A11	A2.A7.A10.W8	Shrubs.Broadleaved.Deciduous.Orchards	CUL/WOC or URB/TRE or URB/VEG
A11	A1.A7.A10.W8	Trees.Broadleaved.Deciduous.Orchards	CUL/WOC or URB/TRE
A11	A1.A7.A9.W8	Trees.Broadleaved.Evergreen.Orchards	CUL/WOC or URB/TRE
A11	A1.A8.A9.W7	Trees. Needle-leaved Evergreen (plantations).	URB/TRE or TRS/TPH/EVR/CON or TRS/FPH/EVR/CON
A12	A1.A4.A10.D1.E2.B9	Woody. Broad-leaved deciduous medium/high closed shrubland (0.5-3m)	URB/TRE or URB/VEG or TRS/MPH/DEC or TRS/LPH/EVR
A12	A1.A4.A10.D1.E1.B9	Woody. Broad-leaved evergreen closed shrubland (0.5-3m)	URB/TRE or URB/VEG or TRS/MPH/DEC or TRS/LPH/EVR
A12	A1.A3.A10.D1.E2.B7	Woody.Trees.Closed.Broadleaved.Deciduous.3-7m	TRS(FPH/DEC) or TRS(TPH/DEC) or URB(TRE) or TRS(GPH/DEC)
A12	A2.A6.A10.F2.F5.F10.G2.B12	Herbaceous.Graminoids.Closed.Secondlayerpresent.Shrubs.Sparse.3-30m.0.8-0.3m	O.HER(CHE) or URB(GRA)
A12	A2.A5.A10.B12.E7	Herbaceous.Forbs.Closed.0.8-0.3m.Annual	O.HER(THE) or URB/GRA OR O.HER(LHE) OR O.HER(HCH) OR O.HER(GEO) OR weak TRS(DCH) OR TRS(SCH)
A12	A2.A10.B12.E6	Herbaceous.Closed.0.8-0.3m.Perennial	O.HER(CHE) or O.HER(LHE) OR O.HER(THE) OR O.HER(HCH) OR O.HER(GEO) or URB(GRA) OR weak TRS(DCH)-TRS(SCH)
B16	A1.A7	Consolidated.Barerock	SPV(ROC)
B15	A1.A4.A13.A16	Builtup.Nonlinear.Urbanareas.Lowdensity	URB/NON
B15	A1.A4.A13.A17	Builtup.Nonlinear.Urbanareas.Scattered	URB/NON
B15	A1.A3.A7.A8	Builtup.Linear.Roads.Paved	URB/NON
B15	A2.A5	Nonbuiltup.Wastedumpdeposits	URB/NON
B28	A1.A5	Water standing	SPV/AQU

Table A5. LCCS classes and GHCs, Santarem, Brazil

Cat.	LCCS Code Modifer	Description	GHC
A12	A3.A10.B2.XX.D1.E1(B5)	Broadleaved Evergreen High Trees	TRS/GPH/EVR or TRS/FPH/EVR
A12	A4.A10.B3.XX.D1.E1(B9)	Broadleaved Evergreen Medium high Thicket	TRS/MPH/EVR or TRS/TPH/EVR
A12	A3.A10.B2.XX.D1.E1(B7)	Broadleaved Evergreen Low Trees	TRS/TPH/EVR or TRS/FPH/EVR
A12	A3.A10.B2.XX.D1.E1(B6)	Broadleaved Evergreen Medium High Trees	TRS/FPH/EVR
A11	A3.B1.XXXX.C1.D1.D9	Permanently Cropped Area with Herbaceous Crop(s)	CUL/CRO
A12	A2.A10.B4.XXXXXX.F2.F5.F10.G2	Herbaceous Vegetation with Trees	CUL/TRS
A12	A3.A11.B2.XXXXXX.F2.F4.F7. G4	Woodland with herbaceous layer	CUL/CRO
A11	A3.B2.XX.C1.D1	Small Sized Field(s) Of Rainfed Herbaceous Crop(s)	CUL/CRO
B15	A1.A4.13.A16	Artificial surfaces and associated areas	URB/ART
B15	A3.A9	Unpaved roads	URB/ART
B15.	A2.A6	Extraction site(s)	URB/NON
B28	A1.A4.B1	Natural perennial waterbodies (flowing)	SPV/AQU