BIO_SOS

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

Deliverable No: 5.3

Report on Remote Sensing Image Understanding System (RS_IUS) 1st Stage Module

Due date of deliverable:	31-05-2012
Actual submission date:	28 June 2012
Version:	V1
Main Authors:	Richard Lucas (P11), Brian O'Connor (P11), Jordi Inglada (P16), Sander Mucher (P4), Laure Roupioz, Damien Arvor (P12), Valeria Tomaselli (P1), Zisis Petrou (P3), Madhura Niphadkar (P5), Harini Nagendra (P5), Maria Adamo(P1), Cristina Tarantino (P1), Stefano Carito (P1), Benedetto Biagi (P1), Francesco Lovergine (P1), Palma Blonda(P1)

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

Deliverable title	Remote Sensing Image Understanding System (RS_IUS) first-stage module
Deliverable number	D5.3
Deliverable version	V1
Previous version(s)	0
Contractual date of delivery	31 May, 2012
Actual date of delivery	28 June, 2012
Deliverable filename	D5.3
Nature of deliverable	R = Report
Dissemination level	PU
Number of pages	
Workpackage	WP5
Partner responsible	P1
Author(s)	Richard Lucas (P11), Brian O'Connor (P11), Jordi Inglada (P16), Sander Mucher (P4), Laure Roupioz, Damien Arvor (P12), Valeria Tomaselli (P1), Zisis Petrou (P3), Madhura Niphadkar (P5), Harini Nagendra (P5), Maria Adamo(P1), Cristina Tarantino (P1), Stefano Carito (P1), Benedetto Biagi (P1), Francesco Lovergine, Palma Blonda (P1)
Editor	Richard Lucas
EC Project Officer	Florence Beroud

Abstract	Provides Very High Resolution (VHR) maps of LCCS categories for four test sites in Italy, Wales and the Netherlands.
Keywords	VHR, land cover maps, habitats, semantic nets

COOPERATION

Signatures

Editor	**Responsibility-Company**	**Date**	**Signature**
Richard Lucas	Editor (P11)	12 June, 2012	
Verified by			
Palma Blonda	Deliverable Responsible (P1)	28 June, 2012	
Laurent Durieux	WP5 Leader (P12)	15 June, 2012	
Jens Stutte	WP3 responsible (P6)	21June ,2012	
Approved by			
Palma Blonda	Coordinator (P1)	28 June, 2012	
Fifamè Koudogbo	QAP (P7)	28 June 2012	

Table of Content

Executive Summary

D5.3, entitled *Report on Remote Sensing Image Understanding System (RS_IUS) 1st Stage Module* is the output of WP5 and outlines an amendment to the first stage of the RS_IUS component of the EODHaM system (Amendment N.1 to Grant Agreement, accepted by REA on 03/02/2012). The technique guarantees high performance and the cost-effectiveness of the whole system. The RS_IUS is equivalent to the EODHaM 1st and 2nd stages, which are devoted to Earth Observation (EO) data processing for Land Cover (LC) and Land Cover Change (LCC) map production.

Previous Deliverables (D6.1, D6.10, D5.1 and D5.2) have evidenced the effectiveness of the Food and Agricultural Organisation (FAO) Land Cover Classification System (LCCS) for both harmonizing different Land Cover (LC) classification systems and translating LC maps to General Habitat Categories (GHCs), from which Annex I Habitats can be defined. Accordingly, the FAO-LCCS hierarchical structure has been adopted as the schema for implementing the EODHaM 1st and 2nd stages. More specifically, EODHAM consists of a 1st stage robust classification of LCCS Levels 1 to 2, with this based primarily on spectral data, followed by a second stage that additionally utilises contextual information to discriminate and map classes in LCCS level 3 and beyond. As a component, approaches to optimise the selection of spectral data and indices used in the classifications have been developed. In both stages, a series of hierarchical layers (e.g., relating to lifeform, cultivation, physical status and surface aspect) that reflect the structure of the LCCS, are generated with each then combined appropriately to generate the LCCS category.

Whilst D5.3 should focus on the EODHAM 1st stage, through an informative Annex 1,it also conveys recent advances within WP5, Task 5.2 in the implementation of both the EODHAM 2nd, which includes classification of categories at LCCS Level 3 and beyond and EODHaM 3rd stage. This was undertaken as it was considered essential to test the feasibility of the FAO-LCCS based classification approach through implementation of the whole RS-IUS module (i.e., EODHaM 1st and 2nd stages). As a consequence, the approach that was proposed by P11 and approved by all the Partners involved in WP5 on the basis of D6.1's findings (as evidenced in D5.2) has been demonstrated.

The EODHaM system has been structured to allow use of a diverse range of spaceborne and airborne remote sensing data acquired at very high resolution (VHR; e.g., Quickbird or Worldview-2) or HR, with focus on optical data but with capacity to integrate other sensor modes (e.g., LiDAR, RADAR). Whilst development of the system has been undertaken and has been implemented within eCognition, the structure can ensure its transferability to open source code for wider use within future projects that might be linked to BIO_SOS.

In Deliverable D.5.3, the design and implementation of the system is conveyed. D5.3 is organized as described hereafter. The introduction (i.e., Section 1) briefly reviews the proposed modular system and evidences the peculiarities of the new EODHaM 1st stage which substitutes SIAM™ as per Amendment N.1.

Section 2 deals with RS_IUS 1st Stage Module (**EODHAM 1st stage)** corresponding to the *Spectral Processing Module* of the service chain provided in D3.1 and reported in Figure 1.1 of the Introduction.

The Data Flow Diagram (DFD) is shown and all EODHaM 1st stage sub-components are described. The segmentation procedure adopted is discussed but also future possible segmentation alternatives are illustrated. The generation of hierarchical layers is explained according to different processing module components. The reasoning behind the implementation approach is provided.

Section 3 describes the EODHaM 1st stage implementation and the consistency of the classification of the FAO-LCCS based approach to a range of test sites in Wales, Italy and the Netherlands as well as India is demonstrated in Section 3.

Section 4 deals with conclusions.

The informative Annex 1, in Section 5, includes actual achievements in WP5. It describes the integration of contextual information and implementation of procedures for classifying to LCCS level 3 and beyond (**EODHAM 2nd stage)**. This Annex also briefly describes the EODHaM 3rd stage, which focuses on the translation of land cover to habitat maps for demonstrating the compliance of the FAO-LCCS based

approach for the full processing chain with direct links to habitats. Finally, the description of ontologies is updated with respect to previous Deliverable D5.2.

Future work is focusing on refinement of the FAO_LCCS based approach from EODHaM 2nd stage to be fully described within D5.5 (due to the end of November) with the description of product quality procedures for the assessment of classification accuracy for the test sites.

1 Introduction

The BIOSOS project has focused on developing a robust system (see Figure 1.1) for habitat mapping and monitoring, named EODHaM, with this based on prior generation of LC and Land Cover Change map (LCC) maps and their subsequent translation to categories of habitat appropriate to support conservation agencies and land managers in decisions relating to protection of Natura 2000 sites. The input data sources are EO-based measurements and on-site data, including ancillary information and in-field measurements. For this purpose, the Food and Agricultural Organisation (FAO) Land Cover Classification Scheme (LCCS) and the General Habitat Categories (GHCs), from which Annex I Habitats can be defined, have been proposed for describing LC and habitat categories. Key criteria in the design of the system included a) ease of use by end users (e.g., land managers, conservation agencies), b) use with a defined range of satellite and, in some cases, airborne (e.g., LiDAR) data and c) low reliance on existing datasets (e.g., land cover maps, cadastral (if updated) and infrastructure layers).

In D3.1 the EODHaM's generic service chain has been mapped as in Figure 1.1

Specific features of the EODHaM system are described in detail below:

- The FAO-LCCS hierarchical classification scheme has been adopted as it provides a useful and logical framework for the integration of multi-source data. The LCCS classification has been designed to provide standardised classes that can be applied and are recognised internationally. Key benefits include the description of a wide range of land covers, including those associated with agriculture. The scheme also considers environmental attributes (e.g., altitude, climate) as well as other technical attributes (e.g., crop-type). The FAO Land Cover Classification System (LCCS) links to the EODHaM 1st and 2nd stages are shown in Figure 1.2.
- The design of a classification system strongly depends on the application at hand. In BIO_SOS, the target classes are mainly those associated with vegetation classes corresponding to habitats that should be monitored in a timely manner. As experienced in the first months of Task 5.1, when focusing on the output vegetation strata of VHR single-date single-sensor imagery, matches between each SIAM^{TM} vegetation sub-category (defined using single-date imagery) and vegetated land cover/use classes are one-to-many and many-to-many, largely because only a limited number of spectral bands are available. Hence, the fine spectral granularity of the SIAM^{TM} output maps was insufficient to deal with the class confusion that occurred in the second RS_IUS stage. This finding was discussed in D5.2 where the output map of the EODHAM 1st stage from single-date SIAM^{TM} processing was used as input to the EODHAM 2nd stage. Rules based on expert knowledge were also used to solve one-to-many and many-to-many relationships. In the revised EODHaM 1st stage, with respect to the original single-date SIAM^{TM-} based approach, multi-source (including multi-temporal and multi-sensor) data are used, as these have proved essential for stratified class description. In most cases, at least two dates of EO data (particularly in seasonal climates) are needed to characterise land surfaces, and particularly vegetation with marked phenologies
- The EODHaM 2nd stage provides a more detailed classification of land covers and is based on a series of hierarchical layers that reflect the dichotomous decisions undertaken within the LCCS classification Levels 1 and 2. Categories defined in these levels can be aggregated into LCCS classes within Level 3 and beyond. Within each layer, decision rules are used to differentiate between elements fundamental to the definition of the LCCS categories. So far, the decision rules are crisp but future work will be focusing on the fuzzification of these where appropriate.
- The EODHaM 2nd stage starts when context-sensitive features, selected on the basis of class description, are used for class discrimination. In some cases, this necessitates the use of ancillary and in-field data.

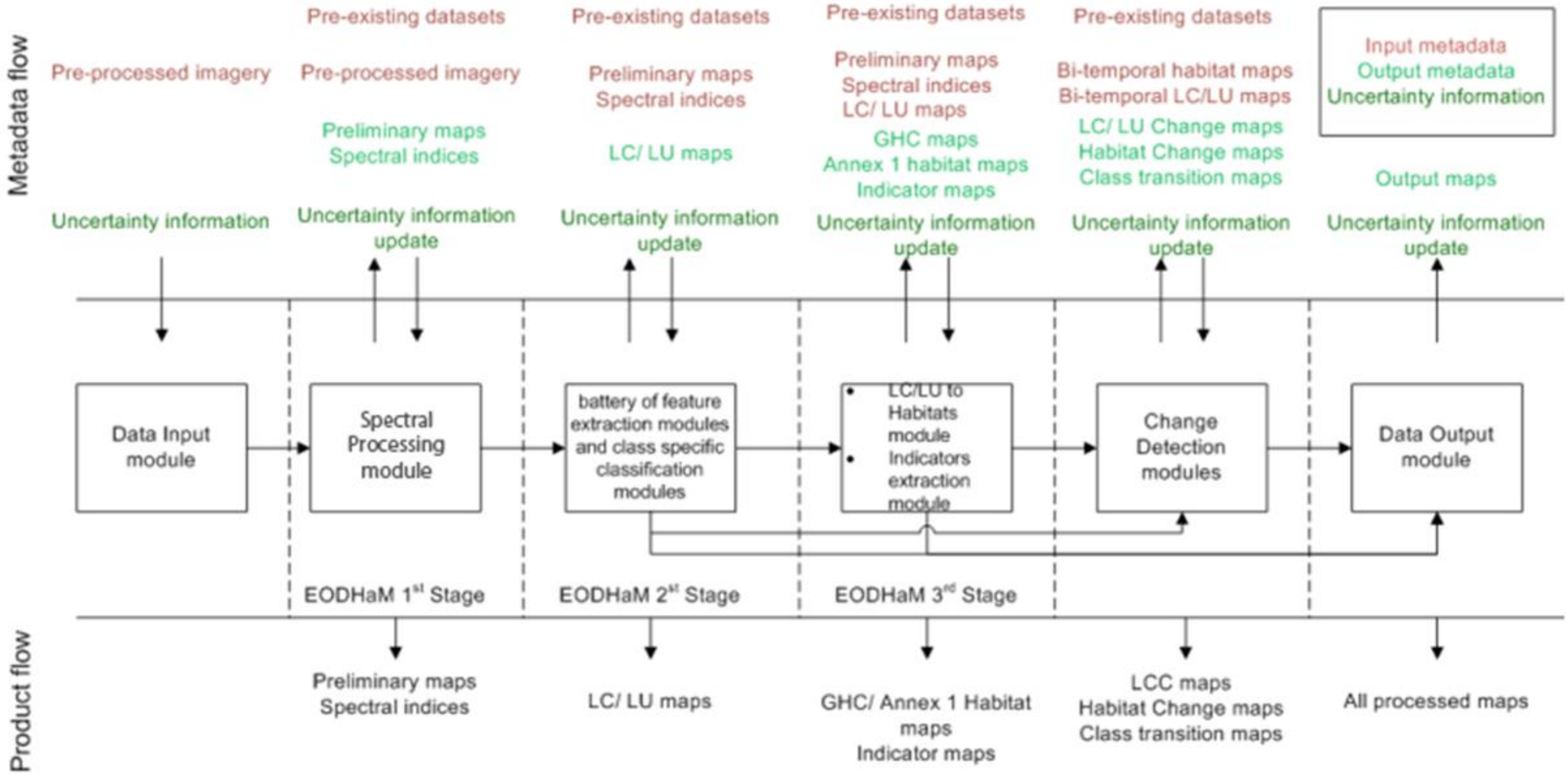


Figure 1.1. EODHaM's service chain (from D3.1)

.

- The classification scheme can make use of context-features that utilise information derived from blocks of pictorial data surrounding the area being analysed (Haralick at al., 1973). These include those based on shape (e.g., regular shapes for man-made objects). In addition, descriptions can go beyond this first level of context by taking into account the spatial organisation of objects with a scene. This list of context-sensitive features includes:
  - Stratified multi-scale texture features (e.g., co-occurrence matrix-derived texture indices).
  - Segment-based geometric attributes (e.g., area, perimeter, compactness, straightness of boundaries, elongatedness, rectangularity, number of vertices, etc.).
  - Stratified morphological attributes (e.g., differential morphological profiles, or DMP, which provide an estimate of the image autocorrelation; it is noteworthy that where autocorrelation is high, texture is low, and vice versa).
  - Spatial non-topological relationships between segments (e.g., distance, angle/orientation, etc.)
  - Spatial topological relationships between segments (e.g., adjacency, inclusion, etc.)

  These contextual features are input to the EODHaM 2nd-stage (see third Module of Figure 1.1) for class-specific classification modules (e.g., to discriminate, within the stratum vegetation, between the land cover classes of grassland and forest/trees or to discriminate within the strata of not-vegetated classes, such as urban).

The EODHAM System is designed to use EO data for the classification of land cover types, which can subsequently be translated to major categories of habitats to a level appropriate to support conservation agencies and land managers in decisions relating to protection of Natura 2000 sites.

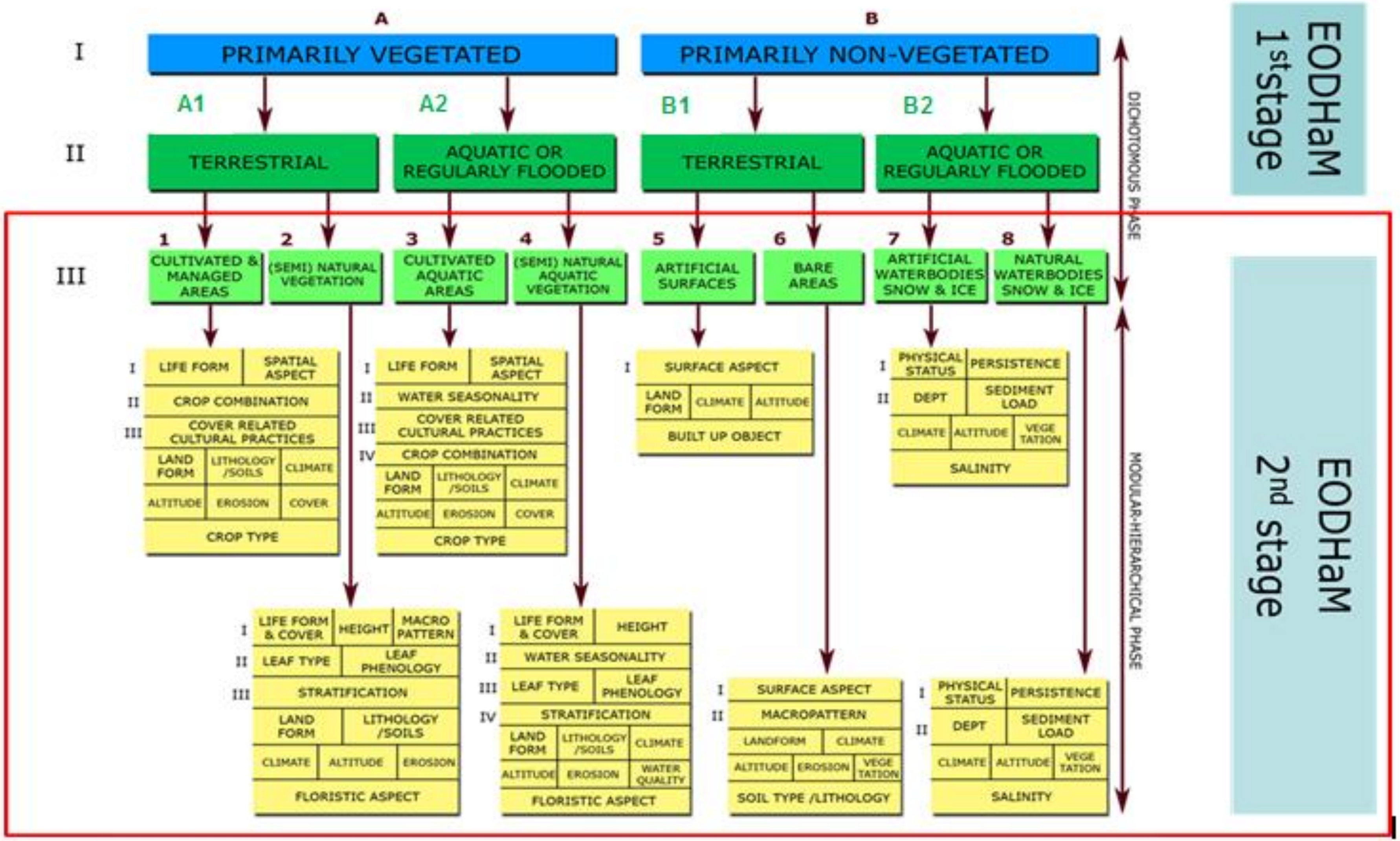


Figure 1.2. The FAO Land Cover Classification System (LCCS) with links to EODHaM 1st stage and EODHaM 2nd stage.

2 Development of EODHAM 1st Stage

2.1 Overview

In the previous Deliverables D5.1 and D5.2, the use of spectral categories for classification of land covers was conveyed for study sites in Italy, Wales and the Netherlands, with maps generated from moderate spatial resolution Landsat and SPOT sensor data (in Wales) and finer spatial resolution single-date imagery in Italy and the Netherlands. Rules for translating the LCCS to GHC categories for these test sites were developed and applied and, for the majority of the categories, a one-to-one matching was achieved in some cases but with exceptions. Protocols for sampling the landscape for developing and validating the classification were also developed and will be advanced in subsequent work packages. The approach to the initial classification was undertaken within a single layer, with land covers progressively categorised. The limitation of classifying land covers based on a single layer was highlighted and hence an alternative approach to classification was considered.

As outlined in D5.2, the initial classification focuses on the differentiation and mapping of the fundamental building blocks of the LCCS up to Level 2, with these being terrestrial or aquatic vegetated or non-vegetated land covers from EO data. Whilst the classification relies primarily, but not exclusively, on spectral data, a key component is appropriate segmentation of the imagery to capture the level of detail but also the spatial arrangement of land covers and contained elements (e.g., trees, buildings) across the landscape.

In this section we are describing the characteristics of the Spectral Processing Module shown in Figure 1.1

2.2 Data Flow Diagram (DFD)

The **EODHAM 1st-stage** implementation is *object oriented* either based on segments generated within software such eCognition or open source or ENVI) and uses a multi-dates approach. It consists first of a segmentation of the imagery based on spectral data acquired primarily by VHR optical sensors (e.g., Quickbird or Worldview-2). This 1st stage also allows inclusion of ancillary information including digital elevation models (DEM) and infrastructure layers, with these facilitating (where available) the mapping of LCCS Level 1 to 2 categories in some cases. In D5.1 and 5.2, the same ancillary datasets were necessarily and similarly used in conjunction with the SIAM™-generated segments to refine segmentation of complex landscapes or areas with high spectral variability (e.g., saltmarshes). Following segmentation, a robust assignment of segments to class primitives, with semantic meaning within or over the landscape, is carried out. These primitives are mapped primarily on the basis of *prior-spectral knowledge-based* information and include:

a) Clouds.
b) Vegetation, including photosynthetic (PV), non-photosynthetic (NPV), submerged/emergent aquatic vegetation (SV/EV) and burnt vegetation (BV).
c) Water or shadow, snow or ice
d) Not vegetated areas.

The primitives are then combined subsequently to generate the LCCS categories belonging to the dichotomous LCCS Levels 1 and 2, with the latter being (see also Figure 1.2.):

a) L2_A1_ Terrestrial_Vegetated,
b) L2_A2_ Aquatic_ Vegetated,
c) L2_B1_Terrestrial Non_Vegetated,
d) L2_B2_Aquatic_Non_vegetated

An additional output class not considered in the LCCS scheme is:

e) Cloud

The specific Data Flow Diagram (DFD) of EODHaM 1st- stage, which provides as output the LCCS Level 1 to 2 output semantic categories, is reported in Figure 2.1. The input images to EODHaM 1st stage are two images acquired at two dates t_1 and t_2 with these characterizing vegetation phenologies. The images are first ortho-rectified, calibrated and then co-registered, as discussed in D5.2.

EODHaM 1st stage is already applicable to any site. The selection of spectral indices is still site-dependent but a systematic analysis of indices is being carried out for generalization purposes. The EODHaM 1st stage encompasses the classification steps from LCCS Levels 1 and 2. The list of modules developed is reported hereafter with reference to DFD in Figure 2.1, with these corresponding to the sub-processes within the process tree generated in eCognition:

a) Bi-temporal image segmentation (Processing element P1).
b) Cloud masking (P2)
c) Spectral indices and 1st-order texture indices extractors (P3 and P4)
d) LCCS Level 1 intermediate and final sematic strata extractor (P5)
e) LCCS intermediate strata extractor (P6)
f) LCCS Level 2 final output (combined) semantic strata extractors (P7)

The following sections describe the multi-source EO images that have been used as input to the system, a more detailed overview of the sub-components of the EODHaM 1st stage, and particularly segmentation of the imagery and approach to classification of LCCS semantic strata (categories), and the input/output specifications of each sub-component of the EODHaM 1st stage.

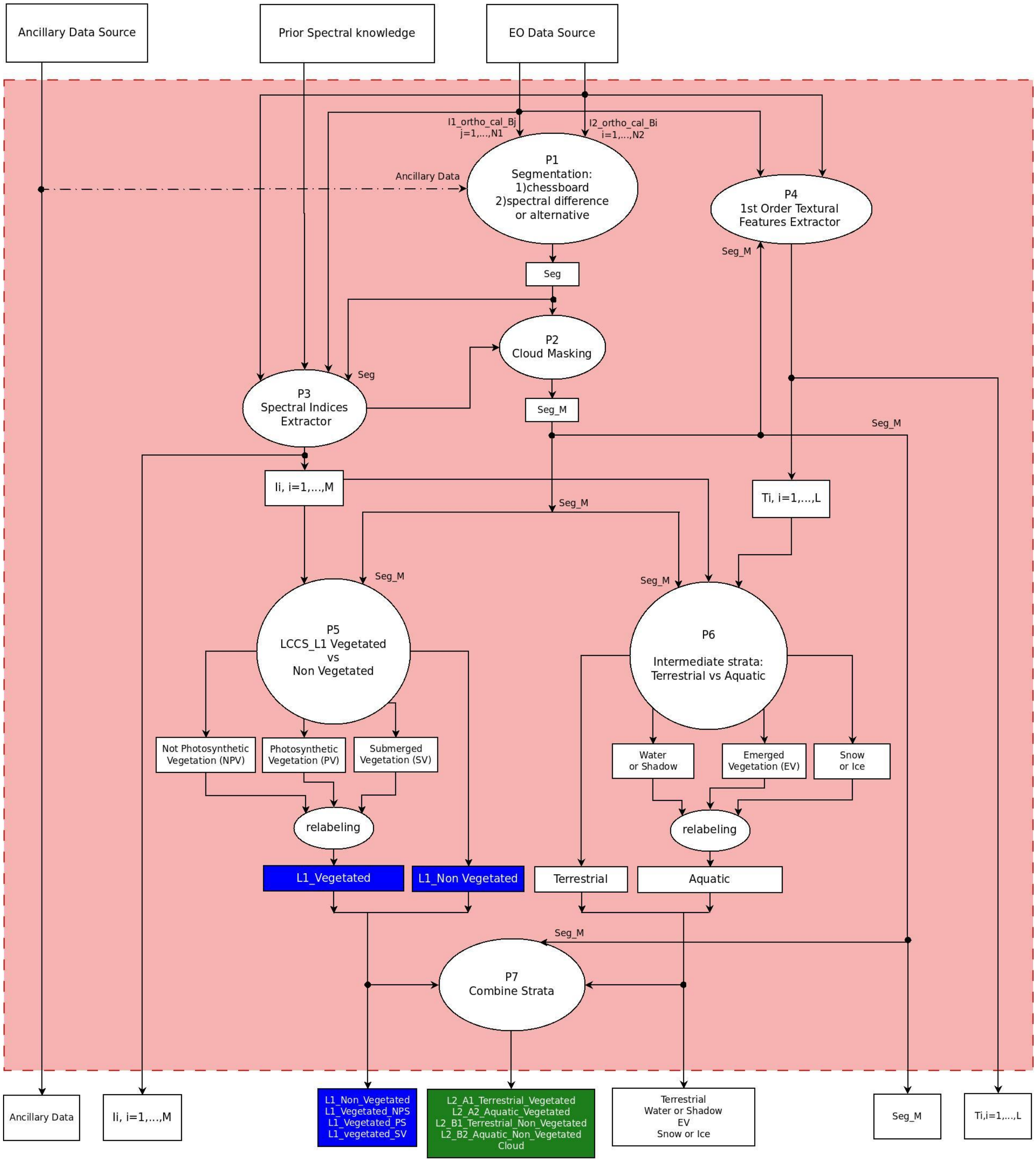


Figure 2.1 EODHaM 1st stage. Data Flow Diagram (DFD) outlining the approach to the classification of LCCS Levels 1 and 2 (in the coloured box). The box identifies the processor to be wrapped and integrated into the entire system. Context-sensitive features will be fed to EODHaM 2nd stage and used therein. The legend is given in Table 2.1.

Table 2.1 DFD's Legend

Pn, n=1…10	Processing module
I1_ortho_cal_Bj	Input ortho-rectified and calibrated image at time t_1 selected according to phenology
N1	Available input image bands at time t_1 selected according to phenology
I2_ortho_cal_Bi	Input ortho-rectified and calibrated image at time t_2
N2	Available input image bands at time t_2
Ancillary Data	These might include DEM, artificial roads network, updated cadastral (in same cases)
Seg	Segmented image from bi-temporal dataset
Seg_M	Segmented image masked for clouds
Not Photosynthetic Vegetation (NPV)	LCCS Level1 intermediate strata
Photosynthetic Vegetation (PV)	
Submerged/Emergent Vegetation (SV/EV)	
Burnt vegetation (BV)	
Terrestrial	LCCS Level 2 intermediate strata as categories in the output raster image from P6
Aquatic	
Water or Shadow	
Emerged Vegetation (EV)	
Snow or Ice	
L1_Non Vegetated	LCCS Level 1output image from P5: L1 categories
L1_Vegetated _NPV	
L1_Vegetated _PV	
L1_Vegetated_SB	
L2_A1_Terrestrial_Vegetated	LCCS Level 2 output image from P7: final categories
L2_A2_Aquatic_Vegetated	
L2_B1_Terrestrial_Vegetated	
L2_B2_Aquatic_Vegetated	
Cloud	
Ii, i=1,…,M	Spectral Indices
Ti, i=1,…,L	1^{st} order texture features

2.3 Use of Multi-source EO data

A particular advantage of the approach adopted for the classification of LCCS categories is that, in principle, as long as input layers are provided at the various levels of the scheme (e.g., lifeform, physical status), a classification can be provided. The first stage is designed in order to be able*: a)* to deal with multi-source EO data observing at a range of spatial and temporal resolutions as well as in different wavelength regions and operating mode as SIAM^{TM}; and *b)* to create a harmonized intermediate level of information as input for the subsequent stages. Multi-source EO data can be the input to sub-components P1, P4.

Comment on each of these sources is provided below with reference to future work and applications.

2.3.1 Spatial resolution

In the EODHAM System, focus has been on the use of VHR optical remote sensing data with consideration given to HR resolution systems (e.g. SPOT-5 HRG). The use of pan-sharpened VHR data provides advantages in the differentiation of some features (e.g., buildings) but can be detrimental in some cases (e.g., in tree crown delineation, as too much detail in terms of larger branches and clumping within crowns is provided). However, in the EODHaM system, data acquired at a range of spatial resolutions can be used, with the classification undertaken at the highest resolution. This allows for the inclusion of, for example, sensor data as coarse as 1 km (e.g., MODIS-derived snow cover products; not yet implemented) for assigning objects defined using VHR data to a duration category, as required by the LCCS.

.

2.3.2 Temporal frequency

The frequency of repeat coverage is the key to the classification of a range of LCCS categories and particularly those associated with inundation patterns and cropping. Intra-annual image acquisitions are also essential for establishing the phenology of vegetation but particularly the differentiation of vegetation that is evergreen, semi-evergreen or deciduous. For quantifying the extent and duration of inundation, imagery acquired on at least a monthly basis is desirable, particularly given the requirement to establish inundation frequencies ranging from with days to at least 9 months. For this purpose, data acquired by the high temporal frequency satellite sensors, including MODIS, may be needed. In some cases, the generation of a DEM coupled with estimates of river height may be used to predict the extent of flooding; hydrological models might also be used to consider inundation patterns as a function of precipitation regimes and events. In croplands, reference may be made to HR optical (e.g., Landsat) or higher frequency radar data (e.g., Tandem-X data) to quantify cycles of growth and harvesting in the EODHaM 2nd stage (not yet adopted).

2.3.3 Wavelength regions and operating modes.

The EODHaM system is primarily based on the use of data from optical sensors and time-series are needed in order to classify categories that are temporally dynamic. A number of VHR sensors operate only in the visible (blue, green and/or red) and near infrared channels (e.g., Quickbird). However, these are limited compared to the Worldview 2 (WV-2) sensor which, based on experience obtained through the BIOSOS project, is the sensor of choice. In addition to the standard visible and near infrared wavelength regions, this sensor also provides observations in the coastal, yellow and red edge bands and, by using stereo imagery, there is potential for the retrieval of vegetation canopy height (even if relative). WV-2 sensor has been widely used for LC map production in the training sites. However, data acquired in the short-wave infrared (SWIR), as in Landsat band 5 (1.55 – 1.75 µm), is desirable as this wavelength region has proved particularly useful for differentiating crop or vegetation types (e.g., as a function of water content and the structure of canopies). Whilst data acquired in this wavelength are not currently acquired by VHR sensors, these are likely to be available in the future, particularly following launch of the SENTINEL-2 and Worldview-3 (from mid-2014), which will provide panchromatic, multispectral resolution and SWIR data at 31 cm, 1.24 m and 3.7 m spatial resolutions respectively.

A limitation of using optical imagery is that information on the height of vegetation cannot be retrieved with sufficient reliability unless relationships with, for example, textural measures are provided. For this reason, LiDAR is very beneficial to the system and in addition can be used to establish the extent of layering within woody vegetation but also for providing detail on the topography of the terrain. For example, LiDAR-derived Canopy Height Models (CHM) can be used to estimate the height of shrubs and tree crowns from the ground surface, with this being critical for descriptions of woody life forms and stratification within the LCCS. The LiDAR signal is divided into ground and non-ground points, with the former assumed to be vegetation although confusion with man-made objects (e.g., communication infrastructure, cars and buildings) can still occur. However, most LiDAR systems also record the amount of energy reflected by target objects (i.e., the intensity) which can assist mapping. Discrimination of height classes within low vegetation is often difficult because of the level of noise range may be as large as that of height range of plants observed. LiDAR data are available for sites in Wales and Netherland whilst they will be acquired in Italy in the summer of 2012. SAR data provide options for the estimation of vegetation structural attributes, including biomass, and also for provision of temporal information on, for example, crop types and patterns as well as inundation. They could be adopted in Northern Europe sites due to frequent cloud coverage, and the use of Tandem-X data will be evaluated following acquisition of these data over Cors Fochno in Wales in June, 2012.

2.3.4 Use of ancillary information

In many regions, and particularly in northern Europe (e.g., the Netherlands and UK), frequently updated information on the extent of cultivation but also urban and communication infrastructure is available and can be used in the segmentation and classification process. Similarly, airborne LiDAR and DEMs are updated regularly. As an example, in the Netherlands, the Dutch Ministry of Economic Affairs, Agriculture and Innovation has invested several millions of Euro in a satellite data portal providing freely available imagery for a wide range of applications, including an updated cadastral map, relevant to the Dutch community in preparation for the Sentinel series of satellites. Where such updated ancillary data are available are updated regularly, these can be used in P1 as well as in P8 and P9 to reduce the complexity of the segmentation and classification tasks within EODHAM, which can be high when VHR data alone are considered. However, in many regions of the Mediterranean and also in tropical countries, such data are scarce or not available. Hence, the information needs to be extracted directly from VHR imagery.

2.3.5 Use of terrain information

There are numerous instances where topographic modelling of terrain can be useful (e.g., in the detection of railway line infrastructure, ditches and modified rivers). Whilst LiDAR is regarded as providing the most accurate estimates, alternative sources for retrieving height information are available, including stereo imaging and radar interferometry and can be used in place of LiDAR for mapping LCCS categories. However, discrimination of lower stature vegetation, as required by the LCCS, becomes more problematic and increases reliance on spectral information.

2.3.6 Overview

The EODHaM system design now allows the inclusion of data from any sensor but can also integrate ancillary information (e.g., thematic layers representing urban infrastructure or cadastral and cropping information, DEMs, or the modelled extent of flooding). Whilst the classification can benefit from the inclusion of these layers, there are regions of the world where such information is not available and so methods have been developed to derive information directly from the image data themselves, where feasible (e.g., cadastral maps from VHR data or DEMs from stereo imagery). This flexibility is one of the key strengths of the EODHaM system.

Whilst the approach to classification is based on the use of EO data, the levels to which the LCCS categories can be defined depend on the information content of the data used. As an example, in the later stages of the classification, the LCCS category Woody (A11) can be divided further into trees and shrubs and also into height categories and strata if LiDAR data are available. However, where only optical data are acquired, it may be that only woody vegetation can be discriminated from herbaceous vegetation and lichens/mosses. The lack of temporal data may also limit the ability to discriminate the

level of inundation within aquatic environments and hence only the class of water or a crop type at one point in time can be defined. In these cases, alternative sources of information may be accessed including, for example, hydrological models giving the extent of inundation or agricultural crop type and yield statistics (as a spatial dataset). By allowing such data to be included, flexibility in the ability to generate classifications of LCCS categories to varying degrees of detail is provided. For potential users, this aspect of the EODHaM systems also allows flexibility in the generation of LCCS categories; indeed, the user only has to have the inputs for each level to generate the classification without knowing the workings of how the layers are derived (although this is advantageous).

2.4 Processing module P1: Segmentation

In the initial development of the classification, segmentation was undertaken within eCognition using a chessboard segmentation that simply divides the scene into equal squares of a given size, followed by spectral difference segmentation. The segmentation can integrate ancillary information including urban infrastructure (buildings, roads) as well as land parcel (cadastral boundaries) but only if regularly updated. However, as such information is often not available, focus has been on extracting this directly from the available imagery and particularly those acquired at VHR (less than 5 meter spatial resolution).

An alternative approach to segmentation, developed by Shepherd *et al.* (2012), used an algorithm that could be applied to a range of EO data (e.g., optical, radar and gridded LiDAR), including multi-temporal datasets. The objective of the algorithm is to define spectrally homogeneous regions for classification. The approach first stretches the input data using a two standard deviation stretch to ensure all bands are equally weighted. A k-Means clustering is then used to identify the unique spectral regions of the image where the number of clusters, k, is user-defined; 60 clusters were used throughout these experiments. The image pixels are subsequently labelled based on the K-Means clusters before being clumped into discrete regions and an iterative elimination, removing regions under a minimum object size, is executed. The elimination process requires a user-defined minimum object size, specified in pixels, where features below this threshold are eliminated through assignment to their spectrally closest neighbour with a size larger than itself. The elimination starts with features of 1 pixel in size and incrementally eliminates the features of larger sizes in 1 pixels steps (i.e., 1 pixel, 2 pixels, …, n pixels). When all neighbours are of equivalent size to the current feature, no elimination takes place and the feature will be eliminated in a subsequent iteration. Following elimination, the resulting features are re-labelled to ensure they are numbered sequentially.

Where multiple images of different dates are available, a stacked composite image can be defined and used for the segmentation, where the spectrally unique regions (defined using data from each date) are identified. Commonly, regions of change between the two or more images are captured by the segmentation. The elimination is therefore executed on all dates simultaneously resulting in a segmentation that captures the features and variation in the input images. In the development, the segmentation was undertaken outside of eCognition and the resulting linework imported into eCognition and used in the subsequent classification procedure.

Within EODHAM, the preference is for the selective segmentation of particular categories (e.g., urban infrastructure, water bodies) within the landscape followed by a segmentation of remaining areas (e.g., vegetation mosaics) using, for example, the algorithm of Shepherd *et al.,* (2012). The spectral bands and indices for segmentation then need to be specific to the categories (strata or layers) being considered. For example, indices such as the Water Band Index; WBI) can be used to segment areas of water whilst those best representing non-vegetated surfaces (e.g., the NDVI) can be used to segment urban infrastructure. In each case, the segmentation needs to focus only on identifying and delineating those categories that are described by and make up the LCCS categories (e.g., high density industrial buildings).

2.5 Robust classification of LCCS Levels 1 to 2

Levels 1 to 2 of the LCCS focus on the differentiation of non-vegetated and vegetated terrestrial and aquatic surfaces. In the first implementation of the LCCS (Deliverable 5.1), these were classified within the same image layer by first classifying vegetated and non-vegetated areas and then, within each, classifying areas that were aquatic and terrestrial. The approach now adopted in the EODHAM 1st stage is to generate two separate layers representing a) vegetated and non-vegetated and b) terrestrial and aquatic surfaces and then to combine these into a four class image representing combinations of these categories (i.e., vegetated terrestrial, vegetated aquatic, non-vegetated terrestrial and non-vegetated aquatic). This first stage classification relies entirely on spectral indices, as outlined below. Whilst consideration is given to optical data, it should be emphasised that these thematic layers can potentially be generated from other sensors (e.g., RADAR, LiDAR).

2.5.1 Spectral bands and indices

Within the EODHAM system, differentiation of vegetated and aquatic vegetation and their complement (i.e., non-vegetated and terrestrial) has relied on the use of only a few key spectral indices rather than spectral bands, as outlined in the following sections. However, in some cases (e.g., mapping of green or photosynthetically active vegetation), several measures (indices but also spectral endmember fractions) are available. The following sections provide an overview of these measures and also some approaches used to ensure optimal selection in terms of accuracy, consistency and applicability to data from optical sensors. To optimize the selection of spectral indices used within EODHaM 1st stage, an analysis of the separability of LCCS level 3 key categories was carried out once the classification to this level was completed within EODHaM 2nd stage. The results can be found in the informative Annex 1, such findings provided insights into EODHaM 1st level processing chain.

Vegetation versus non-vegetation: For the classification of vegetated areas, commonly used indices include the Normalised Difference Vegetation Index (NDVI) and Simple Ratio (SR) whilst the endmember fraction representing photosynthetic vegetation has been considered (Table 3.2). These indices and fractions primarily identify productive (green) vegetation. However,, in many landscapes, the vegetation may be dead or senescent (particularly in the winter months or dry seasons) or submerged, either permanently (e.g., sphagnum pools on bogs) or temporarily (e.g., saltmarshes). Hence, a range of spectral data or derived measures needs to be exploited. Establishing the values of these indices that best represent the boundaries with non-vegetation but also other vegetative states (e.g., senescent) is difficult and requires reference to ground information or independent assessments based on, for example, aerial photography. A summary of additional measures providing best discrimination of vegetative states is given in Table 2.2. ***Non-vegetated*** *areas* are simply defined as those not classified as vegetation. To assist in the discrimination from vegetation, indices that can be used to detect non-vegetated areas include the Normalised Difference Built-up Index (NDBI), which is recoded (as 0 and 1) and used in combination with similarly recoded NDVI values.

Differentiation of terrestrial and aquatic surfaces requires identification of areas of non-vegetated open water, with a number of indices, including the Water Band Index (WBI) and Normalised Difference Water Index (NDWI), providing discrimination (Table 2.3a). However, in many areas supporting aquatic vegetation, the water surface may be partially or wholly obscured from the view of the sensor; reed beds and active raised bogs are notable examples. Aquatic vegetation is also inundated temporarily and at various levels of frequency, ranging from perennial (e.g. swamps and mires) to hourly (as in the case of saltmarshes and mangroves). For these reasons, the spectral variability across aquatic vegetated habitats is high. Nevertheless, their classification is essential in order to determine the extent of the area considered to be aquatic.

Table 2.2. Indices used for the detection of vegetation in various states

Index/fraction	Formula	Vegetative stage	Country applied	Sensor
NDVI1	$\frac{NIR1 - Red}{NIR1 + Red}$	Photosynthetic	Wales	WV
Greenness Index (GI)	$\frac{Red}{Blue}$	Photosynthetic	Italy	QB
SR[1]	$\frac{NIR1}{Red}$	Photosynthetic	Wales	WV
Vegetation Discrimination Index (VDI)	$\frac{NIR - Red}{Green - Red}$	Photosynthetic	Wales	WV/QB
PV fraction[2]	endmember	Photosynthetic	Wales	WV/QB
NDVI2	$\frac{NIR2 - Red\ edge}{NIR2 + Red\ edge}$	Photosynthetic	Wales	WV
Soil Adjusted Vegetation Index (SAVI)[8]	$(1 + L) * (\frac{NIR - Red}{NIR + red + L)}$	Photosynthetic with adjustment for soil		WV/QB
WDVI	$NIRv - gb * Redv$	Photosynthetic	Wales	WV/QB
PSRI	$\frac{Red - Blue}{Red\ Edge}$	Non-photosynthetic	Wales	WV
NPV fraction	endmember	Non-photosynthetic	Wales	WVQB
Normalised Difference Built-up[7] Index (NDBI)	$\frac{SWIR - NIR}{SWIR + NIR}$	Non-vegetation	NA	Landsat SPOT

[1]Huete et al. (2002); [2]Dennison and Roberts (2003; Spectral Mixture Analysis); [3]Rouse et al. (1974); [4]Merzlyak et al. (1999); [5]Clevers, 1991; [6]g represents the slope of the soil line, where v and b represent the reflectance for vegetation and bare ground respectively. The WDVI is proportional to the density of vegetation cover and number of leaf layers and the inverse indicates the amount of soil observed by the sensor. [7]Zha et al. (2003), used in combination with recoded NDVI values. [8]L = 0.5 for most conditions.

In the majority of cases, inundated areas supporting aquatic vegetation are typically located within flood plains or tidal areas of low slope. Whilst a DEM and derived data slope can be used to confine mapping to particular areas within the landscape, reference can also be made to prior classifications of vegetation and open water. As an example, consider the case of saltmarshes, which occur adjacent to tidal areas at low elevation and slope. A sequence of rules to identify saltmarsh from a segmented optical image would be as follows:

a) Map the extent of vegetation and then open water across the landscape and assume that all open water areas below the high tide level are associated with seawater (i.e., are tidal).

b) Where vegetated areas occur next to seawater, assign objects to saltmarsh.

c) Use region growing or distance rules to assign unclassified objects assigned previously to vegetation to saltmarsh if they fulfil the condition of occurring adjacent to objects already classified as saltmarsh and are below an elevation and slope threshold (e.g., 3m and 3^o respectively).

In this case, no spectral information is needed to map saltmarsh. In other cases (e.g., active raised bogs), the habitats are sufficiently distinct to allow discrimination. Examples of measures associated with commonly occurring aquatic vegetated ecosystems are listed in Table 2.3b.

Table 2.3a. Indices available for the detection of open water in various states

Index/fraction	Formula	State	Country applied	Sensor
Water Band Index (WBI1)	$\frac{NIR1}{NIR2}$	Open water	Wales	WV
Water Band Index (WBI2)	$\frac{Blue}{NIR1}$	Open water	Italy, Wales	WV/QB
Normalised Difference Pond Index (NDPI)	$\frac{SWIR - Green}{SWIR + Green}$	Open water		SPOT Landsat
Normalised Difference Water Index (Gao) (NDWI$_g$)	$\frac{NIR - SWIR}{NIR + SWIR}$	Open water		SPOT Landsat
Normalised Difference Water Index Mac Feeters) (NDWI$_m$)	$\frac{Green - NIR1}{Green + NIR1}$	Aquatic vegetation	Wales	WV/QB
Modified NDWI (Mac Feeters) (MNDWI)	$\frac{Green - SWIR}{Green + SWIR}$	Aquatic vegetation		SPOT Landsat
Normalised Difference Snow Index	$\frac{Green - SWIR}{Green + SWIR}$	Snow	Chile[3]	SPOT HRG
Normalised Difference Glacier Index	$\frac{Green - Red}{Green + Red}$	Ice	Chile	SPOT HRG
Shade/moisture fraction	Default photometric shade concept	Aquatic vegetation Shade in canopy	All	All

[1]Penuelas et al. (1993); [2]Dennison and Roberts (2003); [3]Evaluated as no test site had snow or ice cover

Table 2.3b. Indices used for the detection of aquatic vegetation

Index/fraction	Formula	State	Country applied	Sensor
WBI(1)	$See\ Table\ 2.3a$	Open water	Wales	WV
WBI(2)	$See\ Table\ 2.3a$	Open water	Italy, Wales	WV/QB
NDWI_g	$See\ Table\ 2.3a$	Aquatic vegetation	Wales.	WV/QB
SM fraction	$See\ Table\ 2.3a$	Shade in canopy		
NIR-difference	$NIR1 - NIR2$	Active bogs	Wales	WV
Elevation[1]	< 2.6 m Region growing Vegetation next to Open Water	Saltmarsh	Wales	DEM

[1]Contextual information if spectral indices are insufficient for discrimination

An evaluation of the optimal indices for discrimination of the categories of vegetated and aquatic landscapes from those that are non-vegetated and terrestrial is provided in a later section, together with examples of application to the main study sites.

2.6 EODHaM 1st stage's sub-components analysis.

The following sections provide a summary of the main components used in P1 through to P7

2.6.1 Sub-component name: P1 Segmentation

Input specs

Mandatory:
VHR_first_image_t_1
VHR_second_image_t_2

Auxiliary (if updated)
Cadastral layer (or similar)
Tide data
DEM (raster data, tiff format)
Urban layer (vector later, shapefile)
Canopy high layer

Output specs
Segmented map (raster file, tiff)

Dependencies (on other nodes)
VHR_first_image_t_1 (pre-processing node)
VHR_second_image_t_2 (pre-processing node)

Used technologies
eCognition: Chessboard and spectral difference algorithms

Alternatives
Shepherd *et al.* (2012) algorithm

Workflow collocation
Depends on pre-processing module
Output to P2, P3

Human interaction needs
None

DOW reference
WP5 Task 5.1

2.6.2 Sub-component name: P2 Cloud Masking

Input specs
Segmented map (raster file, tiff): Seg

Output specs
Cloud masked Segmented map (raster file, tiff): Seg_M

Dependencies (on other nodes)
P1 (Segmentation node)
P3 (Spectral Indices extractor)

Used technologies
eCognition

Alternatives
Many from literature

Workflow collocation
Depends on P1 and P3
Output to EODHaM 1st stage module P3, P4, P5, P6, P7
Output to EODHaM second stage

Human interaction needs
None

DOW reference
WP5 Task 5.1

2.6.3 Sub-component name: P3 Spectral Indices Extractor

Input specs
Segmented map (raster file, tiff): Seg
VHR_first_image_t_1 (pre-processing node)
VHR_second_image_t_2 (pre-processing node)
Prior Spectral knowledge

Output specs
Spectral Indices (raster file, tiff)

Dependencies (on other nodes)

P1 (Segmentation node)

Used technologies
eCognition

Alternatives
None

Workflow collocation
Depends on P1
Output to EODHaM 1st stage module P5, P6
Output to EODHaM second stage modules

Human interaction needs
None

DOW reference
WP5 Task 5.1

2.6.4 Sub-component name: P4, 1st Order Textural Features Extractor

Input specs
Cloud Masked Segmented map (raster file, tiff): Seg_M
VHR_first_image_t_1 (pre-processing node)
VHR_second_image_t_2 (pre-processing node)

Output specs
First Order textural indices (raster file, tiff)

Dependencies (on other nodes)
P2 (Cloud Masking node)

Used technologies
ENVI

Alternatives
None

Workflow collocation
Depends on P2
Output to EODHaM 1st stage module P6
Output to EODHaM second stage modules

Human interaction needs
None

DOW reference
WP5 Task 5.1

2.6.5 Sub-component name: P5, LCCS_L1 Vegetated vs Not Vegetated

Input specs
Cloud Masked Segmented map (raster file, tiff): Seg_M
Spectral Indices

Output specs
First Order textural indices (raster file, tiff)

Dependencies (on other nodes)
P2 (Cloud Masking node)
P3 (Spectral Indices Extractor)

Used technologies
ENVI

Alternatives
None

Workflow collocation
Depends on P2 and P3
Output to EODHaM 1st stage module P7
Output to EODHaM second stage modules

Human interaction needs
Threshold selection

DOW reference
WP5 Task 5.1

2.6.6 Sub-component name: P6, LCCS_L2 Terrestrial vs Aquatic

Input specs
Cloud Masked Segmented map (raster file, tiff): Seg_M
Spectral Indices
1st Order Textural Features

<u>*Optional (if updated)*</u>
Ancillary Data

Output specs
LCCS Level 2 intermediate strata (raster file, tiff)

Dependencies (on other nodes)
P2 (Cloud Masking node)
P4 (1st Order textural feature Extractor)

Used technologies
ENVI

Alternatives

None

Workflow collocation
Depends on P2 and P4
Output to EODHaM 1st stage module P7
Output to EODHaM second stage modules

Human interaction needs
Threshold selection

DOW reference
WP5 Task 5.1

2.6.7 Sub-component name: P7, Combine strata

Input specs
Cloud Masked Segmented map (raster file, tiff): Seg_M
L1 strata
L2 intermediate strata

Output specs
EODHaM 1st stage output categories (raster file, tiff)

Dependencies (on other nodes)
P2 (Cloud Masking node)
P5 (LCCS_L1 Vegetated vs Not Vegetated)
P6 (LCCS_L2 Terrestrial vs Aquatic)

Used technologies
eCognition

Alternatives
None

Workflow collocation
Depends on P2, P5, P6
Output to EODHaM second stage modules

Human interaction needs
None

DOW reference
WP5 Task 5.1

2.6.8 Summary

In the majority of cases, the EODHAM 1st stage classification can be performed by using only spectral indices, with particular focus on identifying vegetated and aquatic surfaces in two separate layers and assigning all unclassified objects to non-vegetated and terrestrial classes. Mapping of vegetation requires consideration of different states (primarily photosynthetic, senescent and submerged) whilst for aquatic areas (with the exception of open water) contextual information combined with spectral information is necessary depending on the categories occurring. The majority of aquatic habitats are

located in areas of shallow relief and hence the classification can be restricted to these parts of the landscape.

3 EODHAM 1st Stage (LCCS Levels 1-2): Practical implementation on training sites

As indicated earlier, the EODHaM 1st Stage is the sequential classification of vegetated and non-vegetated and aquatic and terrestrial surfaces and then the subsequent amalgamation to generate classes of vegetated terrestrial, vegetated aquatic, non-vegetated terrestrial and non-vegetated aquatic. The EODHAM 2nd Stage focuses first on the differentiation of semi-natural and natural/semi-natural surfaces from those that are cultivated, managed or artificial and then onto the more detailed categories relating to, for example, lifeform, surface aspect and physical status. The following sections provide an overview of the image data available and examples of the classification for EODHam 1st Stage within several of the study sites.

3.1 Overview of available and used imagery

To support the classification of LCCS categories, VHR imagery have been tasked for the selected study sites, with successful image acquisitions over those in Italy (Le Cesine), Wales (Cors Fochno and Cors Caron) and the Netherlands. In addition, moderate spatial resolution SPOT HRG have been acquired (either available from archive or through tasking) for several of the sites, including in Wales, the Netherlands and Portugal. The images currently used in this report for the validation of the classification approach are evidenced by an asterisk in Table 3.1 The remaining images will be used into the future

Table 3.1 VHR imagery available for the BIO_SOS test sites.

Italy (Le Cesine)	Wales (Cors Fochno)	Wales (Cors Caron)	The Netherlands
QB (July 2005)	WV-2 (July, 2011)*	WV-2 (April, 2011)	WV-2 (June 6th, 2011)*
QB (June 2009)*	WV-2 (November, 2011)*		WV-2 (September 29th, 2011)*
WV-2 (October 2010)*	WV-2 (March, 2011)*	WV-2 (March, 2012)	
WV-2 (February 2012)			

3.1.1.1 Worldview-2 data

For the Italian, Welsh and Dutch sites, WV-2 data have been acquired during the period of the BIOSOS project with further task requests submitted for 2013 The benefit of the WV-2 data is that eight spectral bands are acquired. As examples, five WV-2 images have been acquired for Wales; with four taken during periods of pre-flush (March and April for Cors Fochno and Cors Caron respectively) and peak flush (July, 2011 for Cors Fochno) (Figure 3.1). These two periods are relatively stable spectrally with, for example, most vegetation being senescent (e.g., natural grasslands, bracken, deciduous broadleaved forests) in the pre-flush period or with full leaf cover at peak flush. In the spring and autumn, images are acquired during periods of spectral transition, with different species leafing and flowering or experiencing leaf senescence at different times. This is illustrated in the November image (Figure 3.1 where many tree species can be differentiated because of differences in the reflectance characteristics of leaves during the fall.

Contrasts in the two stable periods allow differentiation of broad groupings of vegetation (i.e., photosynthetic and non-photosynthetic) whilst the transition periods facilitate differentiation of habitats, including some that might be rare. The recommended period for acquisition is from mid to late March through to late October as outside of these periods, the lower sun angle compromises the classification of land covers on sloped terrain and also forests. Panchromatic bands were also provided with the WV-2, allowing pan sharpening of the spectral channels.

3.1.1.2 Quickbird data

Quickbird data were available for the Italian and Dutch sites, with these supporting only four spectral bands in the visible red, green and blue.

3.1.1.3 GeoEye

A GeoEye-1 image of 7 January 2011 was acquired covering a 100 km^2 area of site IN-1 (BRT), in the pre-monsoon season. Further acquisitions are planned for the same area during another season (possibly post-monsoon).

3.1.1.4 SPOT-HRG

The SPOT-HRG data have been acquired for sites in Wales and Italy. The Welsh acquisitions are during the same periods as the WV-2, thereby allowing the implications of lower spatial and spectral resolution to be evaluated.

a)

c)

b)

d)

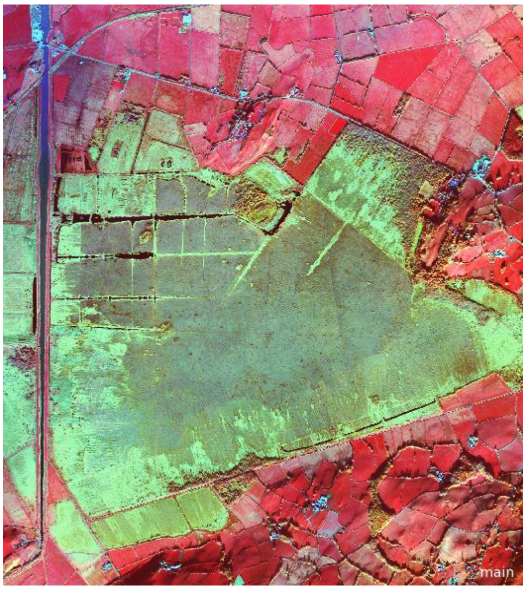


Figure 3.1. WV-2 data acquired for Cors Fochno and surrounds in a) July, 2011. Subsets for the active bog are provided for b) March, 2012, c) July, 2011 and d) November, 2011.

3.1.1.5 Landsat

For a number of sites (Wales, Brazil) time-series of Landsat sensor data have been acquired with these providing potential for understanding the dynamics of land covers (e.g., agricultural crops) and also changes in land use. In the Brazilian study area of the FLONA Tapajos (scene 227/62), 38 Landsat 5 – TM images were acquired for the study period from 1984 to 2011 on the image catalogues from INPE (Brazilian National Instituteof Space Research) and USGS-GLOVIS (United States Geological Survey Global Visualization Viewer; Table 3.2). Most images were acquired during the June to September period where there is a higher probability to get cloud free images. For some years (1984, 1986, 1988, 1991, 1993, 1997, 2003, 2005, 2006, 2008, 2009, 2010), more than one image has been acquired which can be useful for more accurate classifications, especially when vegetation seasonality must be considered to discriminate classes. On the contrary, no image was acquired for other years (1985, 1987, 1994, 2002) due to high cloud cover. All 38 acquired images were calibrated to Top-Of-Atmosphere reflectance.

Table 3.2: Landsat-5 TM acquisition dates, FLONA Tapajos study area (scene 227/62)

Year	Acquisition dates	Year	Acquisition dates
1984	21/6 ; 9/9	1998	12/6
1985		1999	8/2
1986	27/7 ; 12/8 ; 29/8	2000	6/9
1987		2001	11/11
1988	18/7 ; 3/8	2002	
1989	21/8	2003	29/8 ; 10/10
1990	7/9	2004	3/11
1991	23/6 ; 27/7	2005	1/7 ; 17/7 ; 2/8
1992	29/7	2006	5/8 ; 9/11
1993	28/5 ; 19/10	2007	21/6
1994		2008	23/6 ; 30/11
1995	9/10	2009	12/7 ; 28/7
1996	8/7	2010	29/6 ; 31/7
1997	25/6 ; 27/7	2011	16/6

3.1.1.6 Airborne data

As outlined in previous reports, airborne data included LiDAR (for Cors Fochno and the Netherlands) and hyperspectral (for Cors Fochno, the Netherlands and Italy (Le Cesine)).

3.1.1.7 Summary of optical data

For the study areas, VHR optical data have been acquired with these providing significant opportunities for land cover change and habitat mapping. The WV-2 data are considered optimal, largely because the 8 spectral channels include those in wavelength regions that benefit discrimination of vegetation communities and non-vegetated surfaces. The inclusion of panchromatic data also allows finer segmentation of spectral features.

3.1.1.8 SAR data when needed.

As written in Part B2 of DoW, section B2.1and subsection on risk analysis and mitigation strategies, SAR data are also to be acquired to establish the benefits of using in areas where cloud cover, haze or low sun angle, topographic shadowing occur or crop cycling is evident. The integration of the SAR specific processing chain with the LCCS based classification approach will be designed and implemented. P7 has competences to process SAR data and P11 and P1 for multi-source data integration.

Ideally imagery should be taken during periods of minimum surface moisture or else the moisture conditions should be known (e.g., through reference to rainfall records) at the time of the acquisition. X-band data offers a spatial resolution of 3 m and will be used during the project. In particular, a proposal has been submitted for the acquisition of TerraSAR-X data in response to the Announcement of Opportunity (AO; http://sss.terrasar-x.dlr.de/). TerraSAR-X is an X-band advanced Synthetic Aperture Radar (SAR)-satellite system for scientific and commercial applications and was launched mid-2007. The capabilities of radar imagery will be tested on the Welsh and Dutch sites where the cloud coverage sometimes prevents optical images being acquired. As no archive imagery is available over the Welsh sites, new TerraSAR-X imagery will be ordered. As the zone of interest is covered by different frames (Table 3.3); optimal acquisitions parameters will be set at a later stage. Ideally, data will be acquired in July and September but also late December and late March. This would allow the analyst to evaluate use in periods of low sun angle but also during major transition periods. Over the Dutch sites, a TerraSAR-X image acquired on December 05th, 2009 is available in archive. The parameters of this acquisition will be considered for future scenes. They are summarized in Table 3.3.

Table 3.3: Parameters of future TerraSAR-X acquisitions over the Welsh test sites

Mission	**Orbit**	**Start Date**	**Sensor Mode**	**Polarization Mode**	**Beam**	**Incidence Angle (º)**	**Pass Direction**
TSX-1	18	2012-06-04T06:42:37.733	Stripmap	Single	strip_004R	22.45 - 25.52	Descending
TSX-1	86	2012-06-08T17:52:43.733	Stripmap	Single	strip_006R	27.34 - 30.23	Ascending
TSX-1	109	2012-05-30T06:34:04.733	Stripmap	Single	strip_010R	36.12 - 38.54	Descending
TSX-1	162	2012-06-02T18:01:16.733	Stripmap	Single	strip_012R	39.95 - 42.20	Ascending

Table 3.4.: Parameters of the TerraSAR-X scene acquired over the Dutch test sites in December 2009

Mission	**Orbit**	**Start Date**	**Sensor Mode**	**Polarization Mode**	**Beam**	**Incidence Angle (º)**	**Pass Direction**
TSX-1	40	2009-12-05T17:09:40,892	Stripmap	HH	strip_003R	19.70 – 23.19	Ascending

It is important to note that it may be necessary to consider ordering polarimetric data. The additional information provided by the use of the second polarization channel should be valuable for classification purposes. However, the spatial resolution is reduced to 6 m. Furthermore, in June 2010, TerraSAR-X was supplemented in orbit by its twin, the TanDEM-X instrument. In a close formation flight, they will separately acquire data for the TerraSAR-X mission and jointly execute the TanDEM-X mission data collection. The TanDEM-X mission is anticipated to lead to the generation of a global DEM with a resolution of 12m. This DEM will be made available by the German Aerospace Center in more than a year but image pairs (TanDEM-X/TerraSAR-X) can also be provided via the submission of a proposal on the https://tandemx-science.dlr.de/ web interface; this will allow a DEM of the zones of interest using SAR Interferometry (InSAR).

3.2 EODHaM 1st Stage: (LCCS Levels 1-2): Case Studies

3.2.1 Italian sites (Le Cesine)

Le Cesine site is a coastal wetland situated in the south east of Italy in the region of Puglia with an area of 348 ha. The Reserve is one of the oldest protected areas in Puglia, declared as a Ramsar site in 1977 and a State Natural Reserve since 1980, and falls within the Natura 2000 network as pSCI (IT9150032) and SPA (IT9150014). Within the Natural 2000 site of Le Cesine, the natural/semi-natural vegetation consists primarily of the marshes and wet grasslands in proximity to the coastal lagoons and various channels, which form the retrodunal humid area (the most important in southern Italy), and the forests surrounding. Vegetation associated with cultivated/managed lands consists primarily of olive groves and coniferous plantations, with the latter included in the Reserve. The majority of the vegetation remains photosynthetic throughout the year, but in the drier months, dieback of the herbaceous vegetation provides a more extensive cover of non-photosynthetic vegetation. Submerged and other aquatic vegetation are found in the lagoon and proximal marshes. The naturally occurring non-vegetated areas consist primarily of the sand dunes but also the open sea and lagoons and artificial surfaces include the urban infrastructure, although this is located largely outside of the Reserve. The aquatic environments include open water and submerged vegetation, as indicated above.

Vegetation: To discriminate vegetation, the Greenness Index (GI) and Water Band Index (WBI) were used, with the former used to distinguish vegetation from areas of barren land and water. These two indices were exploited, as a Worldview-2 scene was not acquired during the summer months when vegetation biomass is generally at a maximum. To discriminate submerged vegetation, which corresponds mainly to the Annex I habitat 1150, the Normalised Difference Depth Index (NDDI) was used, as this relates to a combination of the turbidity, water depth and presence of vegetation. All remaining areas were assigned to a ***Non-Vegetated*** category (Figure 3.2a).

Aquatic: As illustrated in the DFD P8 processing module, intermediate layers corresponding to spectral semantic categories (e.g., open water, aquatic vegetation) are extracted in EODHaM 1st-stage and then merged into the LCCS Aquatic category. To distinguish open water (i.e., the lagoons), the WBI and GI were used. At Le Cesine, the aquatic vegetation consists of both helophytes and hydrophyes (i.e., *Emerged Vegetation* intermediate layer), with these distinguished using the WBI index extracted from both the two images and the GI from the Quickbird. Some differentiation was also achieved using the first order (occurrence) entropy texture index (Figure 3.2b), with this extracted from a 3 x 3 kernel applied to the green reflectance channel (October). This channel was selected based on band separability analysis and reflected the greater homogeneity of many areas of aquatic vegetation. All remaining areas were assigned to the ***Terrestrial*** category (Figure 3.2c). The spectral rules adopted for generating LCCS level 1 to 2 categories are reported in Annex 2

3.2.2 Welsh sites (Cors Fochno)

The Natura 2000 Cors Fochno site is associated largely with the Annex I habitats Active raised bog (7110) and Degraded raised bogs still capable of natural regeneration (7120). In addition, other main habitats include the saltmarsh and sand dune complexes within the lower Dyfi Estuary, with these being mainly vegetated with non-vegetated components. However, extensive areas of non-vegetated natural habitats occur including the sea, tidal waters and sand/mud flats of the estuary. Artificial non-vegetated surfaces are largely associated with the urban infrastructure (residential buildings, roads and railway).

Vegetation: Depending upon the time of year, the vegetation within and surrounding the Natura 2000 site is either green (photosynthetic), brown (non photosynthetic) or of low productivity, with the latter associated with woody shrubs, herbaceous vegetation (primarily graminoids) and lichens/mosses (e.g., on the active raised bog). To establish the extent of each of these vegetative states, a combination of the NDVI, PSRI and differences in the WV-2 near infrared channels were used respectively. An additional category of submerged vegetation (primarily sphagnum in bog pools) was also evident in the field but less so from the WV-2 bands, although could be inferred from knowledge of the distribution of water. Thresholds of these indices were used collectively to identify the vegetated area. All remaining areas were associated with ***Non-Vegetated*** *areas*.

Aquatic: Areas of open water included freshwater ponds (generally standing) and brackish and salt water (open sea). These were differentiated using the WBI, with the distinction made between standing and flowing water based on partial or complete surrounding by a non-water surface respectively. To map the extent of the aquatic class, consideration needed to be given to vegetation, which was submerged, emergent or floating (as in the case of active bog). The difference in the WBV NIR reflectance bands allowed differentiation of the active bog. However, the saltmarsh vegetation was spectrally diverse and hence reference was made to elevation datasets. In particular:

a) Areas identified previously as vegetation and immediately adjacent to flowing water were identified.

b) Region growing was performed to expand the area inland on the condition that the elevation remained below 2.5 m (with the growing effectively halted by the levees occurring on the landward margins).

As an alternative, the Normalised Difference Water Index (NDWI), which is sensitive to the canopy water content, allowed the identification of vegetation containing water including the active raised bog and the saltmarsh, although confusion with ploughed fields and sand dunes was evident (Figure 3.3). All remaining areas were assigned to the Terrestrial category. The classification of LCCS Levels 1 and 2 for a section of Cors Fochno is highlighted in Figure 3.4

3.2.3 The Netherlands

The Dutch site is located within the Natura 2000 site the Veluwe. The Veluwe is the largest end moraine in the Netherlands, an undulating sandy landscape that was created during penultimate glacial period, about 150,000 years ago. The final landscape of alternating sand dune areas, heathlands and dry forests were created by a long history of intensive land use. The heathland area Ginkelse and Ederheide covers an area of approximately 1000 ha in size and is known for its large area covered by Calluna heath vegetation. The main Annex Habitat types for this study area are: i) 4030 European dry heaths; ii) 2310 Dry sand heaths with Calluna and Genista; iii) 2330 Inland dunes with open Corynephorus and Agrostis grasslands. This terrain Ginkelse and Ederheide is managed by Ministry of Defense. The Wekeromse Zand, an active inland sand dune area, 3 km North of the Ginkelse & Ederheide, has a total area of approximately 500 ha and is managed by Geldersch Landschap. About 100 ha of this area is covered by open space with active inland sand dunes. The main Annex Habitat types for of the Wekeromse Zand are: i) 2310 Dry sand heaths with Calluna and Genista; ii) 2330 Inland dunes with open Corynephorus and Agrostis grasslands; iii) 6230 Species-rich Nardus grasslands; iv) 9120 Atlantic acidophilous beech forests with Ilex and sometimes also Taxus in the shrublayer; v) 9190 Old acidophilous oak woods with Quercus robur on sandy plains.

Aquatic. Most of the water bodies in the Dutch study areas are small fens or man-made water bodies near a farms. The fens can be quite shallow and are often surrounded by forests. Differences in depth and shadow effects can lead to spectral differences. However, most of the fens and small water bodies were successfully identified using the specific indices (WBI, WI, WBI diff). More problematic is that too much shadow and very dark surfaces are still confused with water, causing too many pixels identified as water (amongst others some roads). Several of the index strata used within the EODHaM 1st stage are shown in Annex 2.

.

a) b)

b) d)

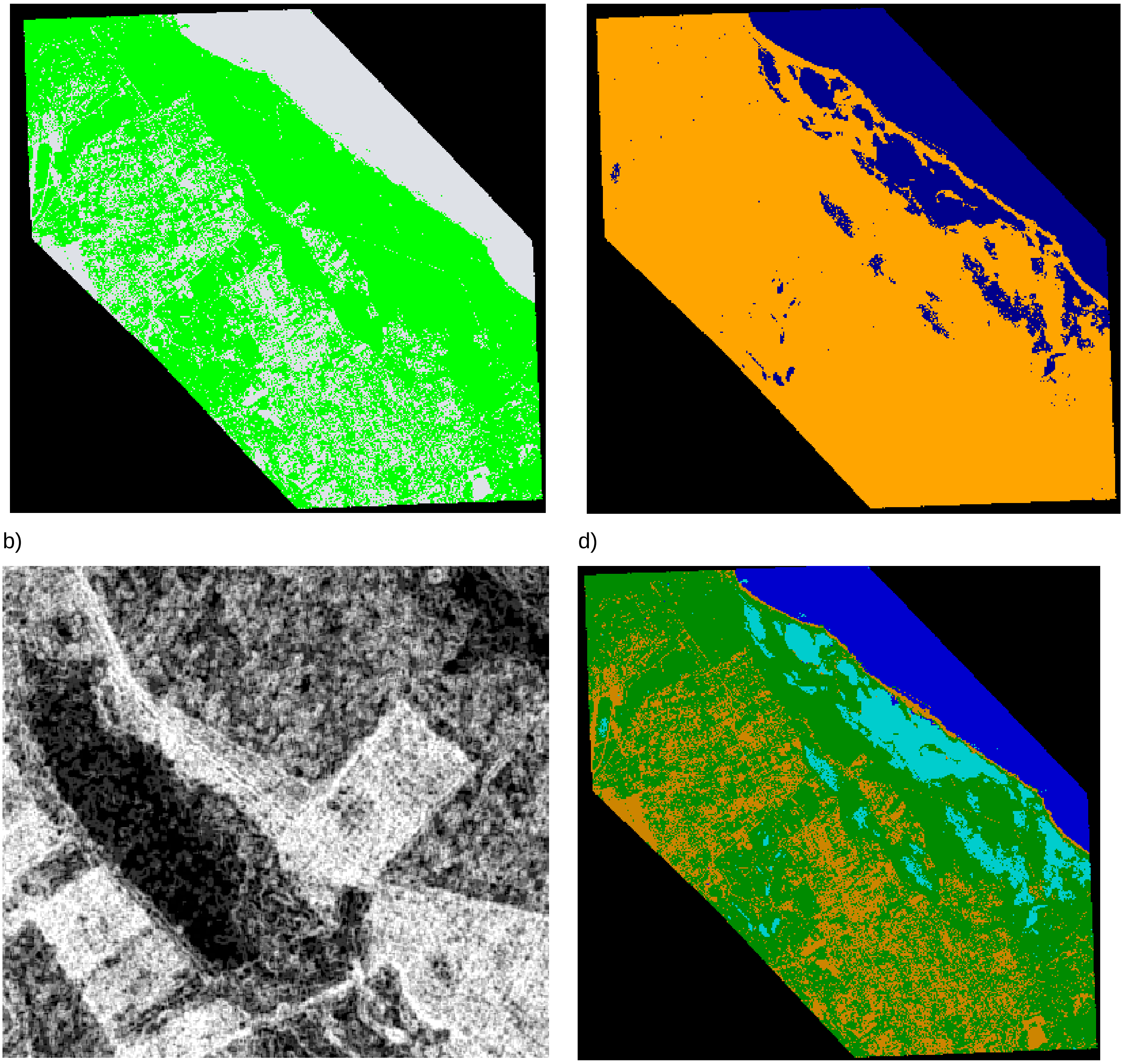


Figure 3.2. The extent of a) vegetated (green) and non-vegetated (grey) and b) aquatic (blue) and terrestrial (orange) within Le Cesine, Italy. c) Image of entropy, generated from a 3 x 3 kernel applied across the Summer Quickbird green reflectance channel. High values correspond to areas occupied by trees (e.g., olive groves, coniferous plantations) whilst low values are associated with homogenous fields (e.g., grasslands), low vegetation or aquatic vegetation. d) The Level 2 classification of terrestrial vegetated (A1; green), aquatic vegetated (A2; cyan), aquatic not vegetated (B1; blue) and terrestrial not vegetated (B2; orange).

a) b)

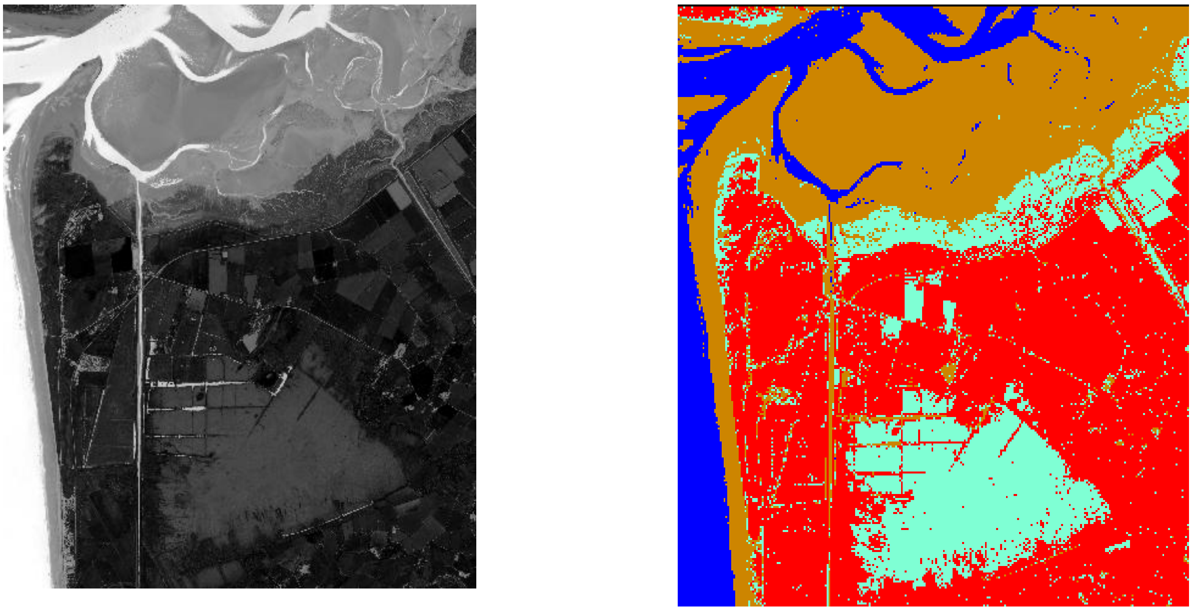


Figure 3.2. The NDWI for Cors Fochno and surrounds and b) a derived classification of aquatic vegetation

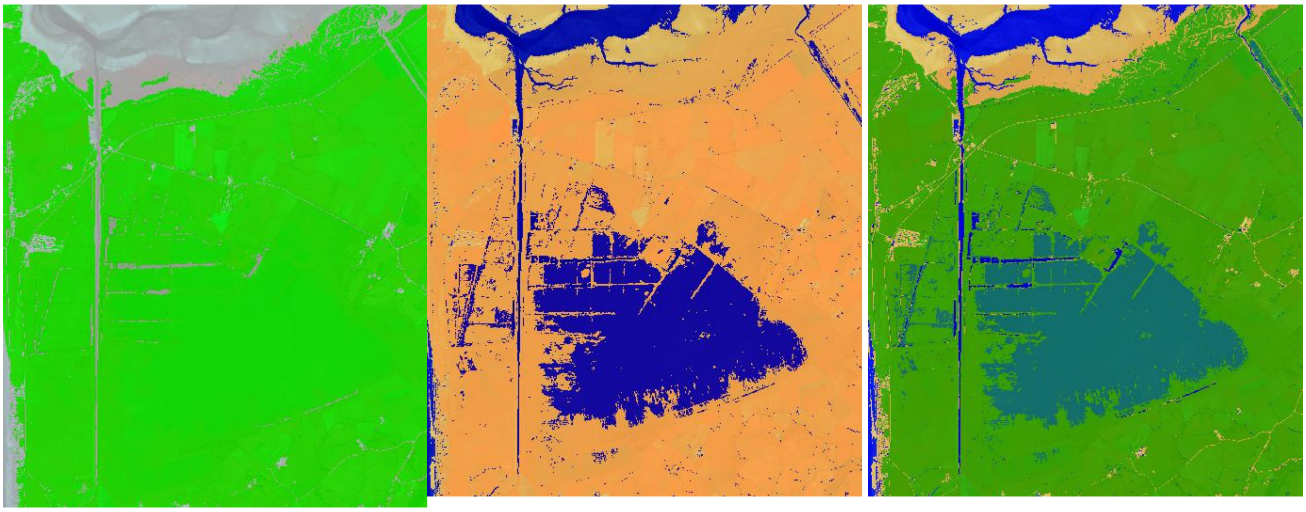


Figure 3.4. The classification of vegetation (green) and non-vegetation (grey), aquatic (blue) and terrestrial surfaces (orange) for a section of Cors Fochno and the lower Dyfi catchment.

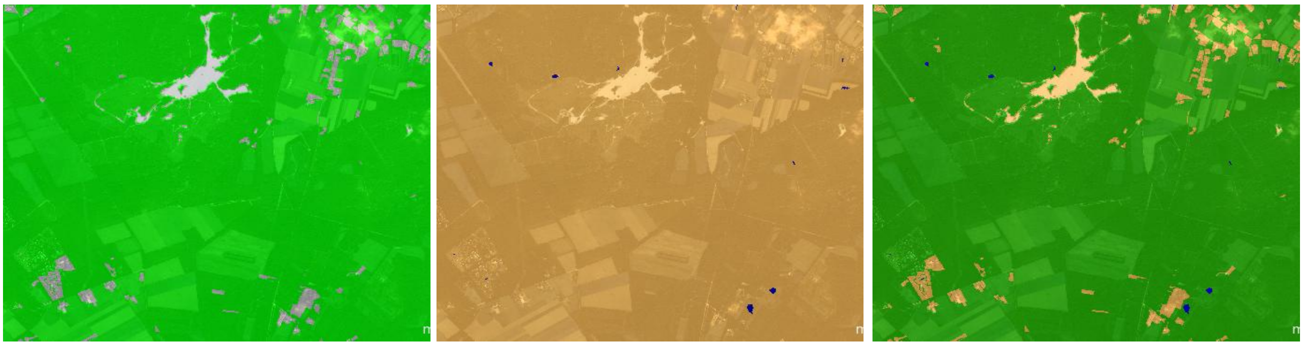
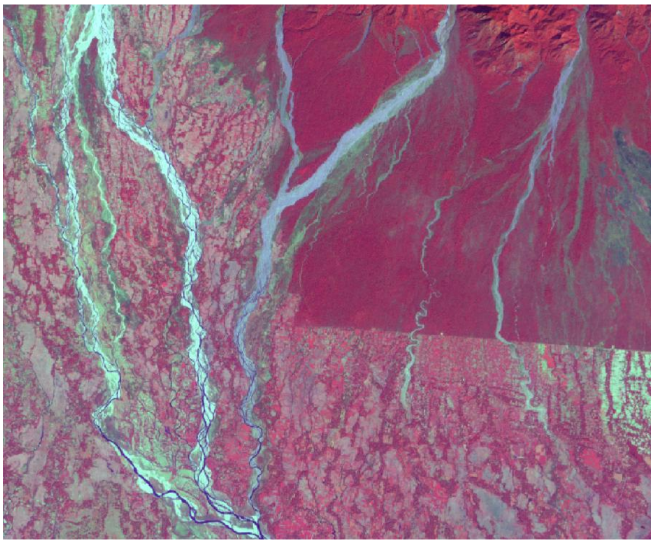


Figure 3.5. The classification of vegetation (green) and non-vegetation (grey), aquatic (blue) and terrestrial surfaces (orange) for a section of the Veluwe, the Netherlands.

3.2.4 Other sites

The classification to LCCS Level 2 has been achieved for a related site (Manas National Park in Assam, India, as well as in Patagonia, Chile. The classification for Manas (Figure 3.6) was based on a combination of SPOT-HRG and dual season Landsat Thematic Mapper data. Whilst the SPOT image exhibits less water than shown in the classification, this extended area was observed with the wet season Landsat image. To evaluate the use for identifying snow and ice, the classification of the Patagonian mountainous region was conducted using a single SPOT scene and exploiting the NDSI and NDGI (Figure 3.7). For mapping water bodies, ice, snow and open water were mapped separately using a combination of the NDWI, NDSI and NDGI and combined into a single class. In both cases, the approach to classification was logical and facilitated mapping of the broad land cover types needed for the EODHAM 1st stage.

a)



b)

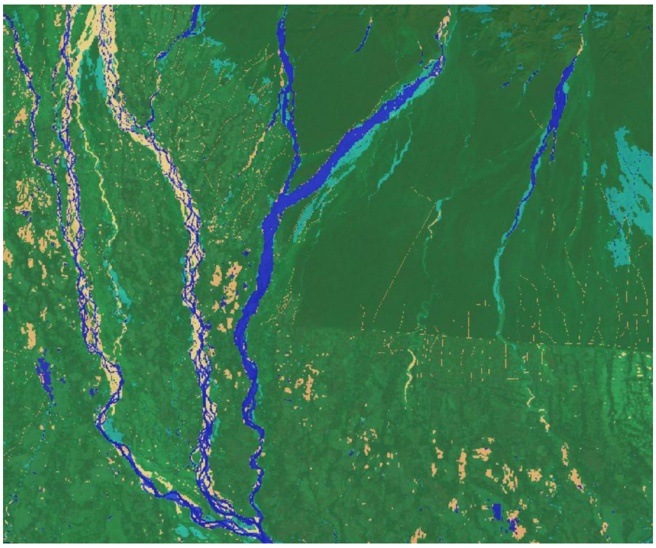


Figure 3.6. a) SPOT-HRG image of Manas National Park and b) LCCS Classification Level 2 classification showing areas of natural terrestrial (green) and aquatic (sea green) vegetation, water (blue) and bare areas (orange).

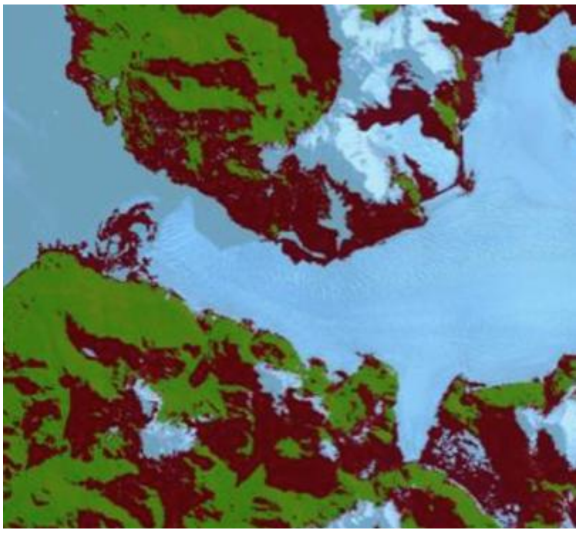


Figure 3.7. Classification of a single-date SPOT HRG image of Patagonia showing areas of natural waterbodies (mapped separately as open water, ice and snow; pale blue), natural vegetation and bare areas.

3.2.5 Accuracy assessment

3.2.5.1 Introduction

A high level of classification accuracy is essential at Level 2, as the extent of vegetation (e.g., photosynthetic, non-photosynthetic, non-submerged aquatic, submerged/emergent and burnt vegetation) and aquatic (open water and land, including with vegetation cover, that is inundated or regarded as wet) and their opposites (i.e., non-vegetated and terrestrial) determines the success in the classification of categories lower down in the class hierarchy.

A classification scheme has two critical components: a set of a) labels and b) rules or definitions, such as a dichotomous key for assigning labels (e.g., a “deciduous forest must have at least 75% crown closure of deciduous trees”) (Congalton, 2009). Without a clear set of rules, the assignment of labels to classes can be arbitrary and lack consistency. Without class definitions expressed as quantifiable rules, there can be little agreement on what area on the ground or the image should be labelled. In addition, a classification scheme should be a) mutually exclusive and b) totally exhaustive. Mutual exclusivity requires that each mapped area fall into one and only one category or class. A totally exhaustive classification scheme results in every area on the mapped landscape receiving a map label; no area can be left unlabelled. It is also advantageous to use a classification scheme that is *hierarchical*, as specific categories within the classification scheme can be collapsed to form more general categories. That said, reference data should be collected and labelled using the same classification scheme as that used to generate the map, with this being FAO-LCCS in the present work. However, as in the case of our study sites, existing maps were created using the same classification scheme. Any differences between the classification scheme of the map and the classification scheme of the reference data may result in discrepancies between map and reference accuracy assessment site labels. Consequently, as evidenced in Congalton (2009), the result will be an assessment of classification scheme differences, and not of map accuracy.

For the BIOSOS study sites, no pre-exiting Land Cover maps in FAO-LCCS taxonomy were available for use in stratified random sampling of reference points/areas for validation. A pre-existing LC map in CORINE taxonomy exists for Le Cesine but its translation in FAO-LCCS taxonomy (as undertaken in D5.1) does not ensure that the classification scheme used for labelling the classes in the original map is coherent with the hierarchical FAO-LCCS scheme adopted in EODHaM to produce LCCS level 1 to 2 outputs. Therefore, and to conclude, as in-field campaigns are still on-going and the final classified maps (i.e. LCCS Level 3 and beyond) are not yet completed, random sampling was adopted for validation with this stratified in the case of Cors Fochno because of the existence of Phase 1 Habitat Maps that were quite well related and could be translated to the LCCS categories. Initially, validation has been performed on the LCCS Level 1 to Level 2 output EODHaM products. The labelling of reference samples for both Cors Focno and Le Cesine was based on pan-sharpened photo-interpretation of input images, aerial photography and in-field visits. Validation is ongoing for the Dutch sites.

3.2.5.2 Cors Fochno

An example of the assessment of classification accuracy to Level 2 for Cors Fochno is given in Table 3.5, with this based a) on field data collected for the range of LCCS categories (and associated GHCs) occurring and b) interpretation of aerial photography through random points (circular polygons of 2 m diameter). A total of 555 points were assessed, with 255 collected in the field. The accuracy assessment, based on a standard confusion matrix, provided an Overall Accuracy of 81.8 % with and error tolerance of 3.2% (Spiegel 1961), with the User’s and Producer’s accuracies for three classes (terrestrial vegetated, aquatic vegetated and aquatic non-vegetated) being over 81.4 % and as high as 87.5 % (for Users). The accuracy was reduced to 60.4 % and 67.4 % for the User’s and Producer’s accuracy for terrestrial non-vegetated surfaces with the majority being attributed to the confusion with aquatic vegetation and to a lesser extent terrestrial vegetation. The main reason for this confusion was the with the area of saltmarsh which is comprised primarily of Spartina and Saliconia plants mixed with mudflats on the lower tidal edge and grading into terrestrial grasslands (e.g., with Couch grass; *Elymus*

repens) on the upper margins. When the WV-2 data were acquired, low and mid tides were observed and hence the mud and sand flats were exposed (some are only covered in the very high tides in the spring and autumn). As such, these were classified as terrestrial non-vegetated and could only be classified as aquatic non-vegetated if an image was acquired at high tide (i.e., the area was covered in water) or ancillary tidal information was obtained. On this basis, classification accuracies would be considerably higher and certainly well above the 80 % threshold.

Table 3.5. Assessment of classification accuracy for Cors Fochno LCCS Level 2.

Reference data (Field or photo-interpretation)

Classified data	A1 TERR VEG	A2 AQU VEG	B1 TERR NON-VEG	B2 AQU NON-VEG	Number	User's Accuracy(%)
A1 TERR	179	34	7		220	81.4
A2 AQU VEG	29	197	3	2	231	85.3
B1 TERR NON-VEG	2	10	29	7	48	60.4
B2 AQU NON-VEG		3	4	49	56	87.5
No.	210	244	43	58	555	
Producer's Accuracy(%)	85.2	80.7	67.4	84.5		
				Overall		81.80 (%)
				Error Tollerance		3.20 (%)

3.2.5.3 Le Cesine

The assessment of classification accuracy of LCCS Level 2 output map for Le Cesine is reported in Table 3.6. A total of 535 random sampling reference points were considered. A cluster of pixels regarded as a reference unit areas was associated to each sample point in order to best identify the point both on the image and in the field and to account for positional errors. The minimum mappable unit (MMU) area was 4 pixels to allow the crowns of olive trees to be considered. The reference areas were labelled by photo-interpretation and in-field campaigns according to the LCCS hierarchical classification scheme, as recommended in (Congalton, 2009). Each corresponding cluster of pixels (sampling area) was counted as one sample in the confusion matrix, with this ensuring that the set of samples in Table 3.6 are not correlated (Congalton, 2009, pag 72). On this basis, the accuracy assessment provided an Overall Accuracy of 90.84 % with and error tolerance of 2.44%.

Table 3.6 Reference data (Field or photo-interpretation)

Classified data	A1 TERR VEG	A2 AQU VEG	B1 TERR NON-VEG	B2 AQU NON-VEG	Number	User's Accuracy(%)
A1 TERR	264	23	23	0	310	85.2
A2 AQU VEG	0	79	0	0	79	100.0
B1 TERR NON-VEG	2	0	106	0	108	98.1
B2 AQU NON-VEG	0	1	0	37	38	97.4
No.	266	103	129	37	535	
Producer's Accuracy(%)	99.2	76.7	82.2	100.0		
				Overall		90.84 (%)
				Error Tolerance		2.44 (%)

(a) (b)

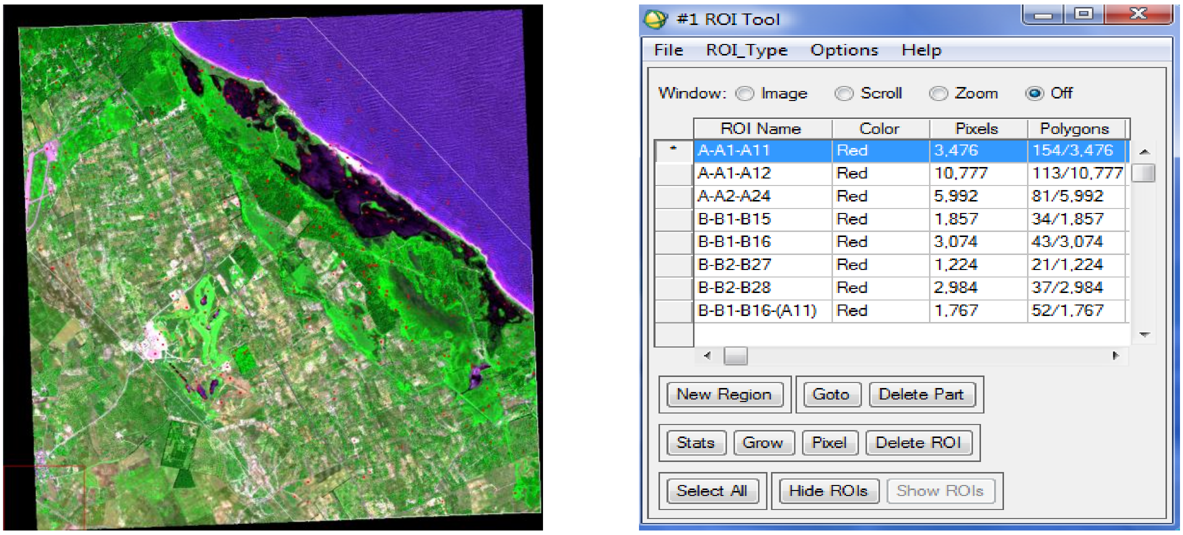


Figure 3.8. (a) LCCS random sample points in red; (b) cluster sample areas. Every area, with this composed by several pixels, is counted as one element of the confusion matrix.

3.2.5.4 Future work

The accuracy beyond Level 2 will be reported in full D5.5 for all test sites as further field campaigns are being conducted during the summer of 2012, with these considering all levels of the LCCS classification (to Level 3 and beyond; see Section 5.2). Furthermore, the approach to classification beyond Level 2, which culminates in a complete classification of LCCS categories (see Section 5), is considering an assessment of the final LCCS map (and its subsequent translation to GHCs) but also the individual layers used in its generation. For example, height and cover are key descriptors of woody vegetation (trees and shrubs) and high levels of accuracy can be obtained using LiDAR data, particularly for the former, although less so if these data are not available. By contrast, the classification of leaf type (needle-leaved, broad-leaved and aphyllous) is less reliable particularly in seasonal environments. For example, species of Larch (*Larix*) are needle-leaved but also deciduous and hence the ability to extract the area of cover will depend upon the time(s) of observation. Hence, the greatest error in the classification of LCCS categories will be linked to the layer leaf type. These errors will cumulate through the classification and focus is on assessing the impact of this but also that of the uncertainty and the optimisation of indices, as outlined in Section 5.2. Consideration is also being given to object-based assessment of accuracy and in evaluating sub-pixel proportions (e.g., as determined through spectral unmixing or fuzzy classification). The modules and protocols relating to error assessment will, however, be generated following the summer field campaigns and integrated shortly after within the EODHaM system.

3.3 Summary

The classification of VHR optical data to LCCS Levels 1 and 2 was based entirely on spectral indices. In each case, there was only the need to discriminate vegetation and aquatic surfaces as all other areas were logically assigned to non-vegetation and terrestrial. Within the vegetation category, which was discriminated largely using the NDVI, GI or FDI, a number of subcategories required identification with these being non-photosynthetic (brown), vegetation overtopping water (e.g., active bog) and submerged and/or emergent vegetation. In these cases, several additional indices were needed, namely the PSRI, NDWI and the difference in the WV-2 NIR channels. For mapping aquatic vegetation, the same rules were applied (with the exception on non-photosynthetic vegetation). Different water states were also mapped using the WBI or NDWI or, in the case of snow and ice, the NDSI and NDGI. All indices have

been published in the literature, with the exception of the WV-2 NIR difference, which was specific for discriminating the active bog, Cors Fochno. In the case of both vegetated and aquatic categories, all sub-categories were merged for their generation although the rules were reapplied at a later stage in the classification (i.e., beyond LCCS Level 3). In principle, as long as vegetation and aquatic surfaces are mapped to an acceptable level of accuracy, the generation of LCCS Level 2 categories can be readily performed as part of the EODHaM 1st stage.

4 Conclusions

The BIOSOS project seeks to develop a robust system for land cover and habitat mapping and monitoring to support the long-term protection of Natura 2000 sites and other areas of conservation importance, within Europe but with potential to expand to other biomes globally. Through WP5 and other associated work packages, the components of the EODHaM System have been designed and are progressively being implemented, first on the test areas in Italy, Wales and the Netherlands and subsequently on other areas in Greece, Portugal, India and Brazil. Key advances conveyed in WP5.3 and leading on from WP5.1 and WP5.2 are summarised below.

a) For classification of land cover, the FAO LCCS has been adopted, with this providing a dichotomous system that leads to the logical classification of land covers from EO data.

b) The EODHAM system is designed to take, as input, EO data from a range of sensors observing at different spatial and temporal resolutions and modes of operation. Whilst primarily designed to utilise data from VHR optical sensors, HR optical and also LIDAR and RADAR data and derived products can be exploited.

c) A key component of the EODHAM System is the segmentation procedures, which are being progressively developed to allow hierarchical segmentation of imagery, commencing with artificial, cultivated and managed areas and finalising with semi-natural natural land covers. An additional component considers within-pixel components of the landscape (e.g., different lifeforms).

d) The EODHaM 1st Stage classifies up to Level 2 of the LCCS using spectral information, with the main outputs being layers representing vegetation and non-vegetation (Layer 1) and aquatic versus terrestrial environments (Layer 2). These classes are combined in a subsequent process, allowing the creation of a map of terrestrial vegetated, aquatic vegetated, terrestrial non-vegetated and aquatic non-vegetated environments.

e) The EODHAM 2nd stage (described in Section 4) consists of two components. The first advances the classification to LCCS Level 3 and requires contextual as well as spectral information. The second component progresses beyond Level 3, providing more detailed classifications of land cover (e.g., in terms of life form, physical status) within hierarchical layers.

f) Within each of the hierarchical layers, separate classifications are undertaken and then these are combined subsequently to generate the LCCS classes. Additional environmental attributes can help class discrimination.

g) Following classification to LCCS categories, EODHaM3rd stage first facilitates translation of these categories to GHCs (i.e., a habitat map).

h) Within the broad habitat categories, more detailed classification of GHCs can be undertaken, with further classification to dominant species or genus level achievable in some cases.

The key benefits of the EODHAM System are:

a) The use of the LCCS provides a standardized framework for land cover classification that is globally recognised, logical in design and whose components can be generated from EO data.

b) The LCCS and subsequent GHC classes differ when change (e.g., in height, cover, frequency of inundation) occurs, and hence can be used for monitoring.

c) For end users, the building blocks of the system can be generated as background and the final LCCS categories then easily defined with reference to the LCCS components.

d) The approach is applicable to any area across a range of scales and, in the absence of some components from EO data, other information (e.g., thematic layers or modelled hydrology) can be included.

e) The EODHaM classification and the range of indices used in its generation can also be used as input to species distribution models, thereby supporting subsequent WPs and the monitoring of indicator species, which themselves can potentially be mapped from EO data.

The EODHaM system is being developed in collaboration with end users, thereby facilitating its subsequent uptake in the longer term.

5 Informative Annex 1: EODHAM 2nd and 3rd Stages

The EODHaM 2nd stage consists of classification to LCCS Level 3 and subsequent levels. The following sections provide a summary of the approach and its application to the BIOSOS test sites.

5.1 Classification to LCCS Level 3

5.1.1 Overview

To classify to Level 3, the maps of vegetated and non-vegetated terrestrial and aquatic land cover classes are combined with two additional layers representing a) artificial and natural bare areas and waterbodies and b) cultivated/managed and semi-natural/natural areas. Within each layer, all objects not assigned as one component are assigned to the opposite case (e.g., all areas not classified as artificial are assigned to natural). Once classified by layer, all available categories are combined to generate those listed in Figure 5.1.

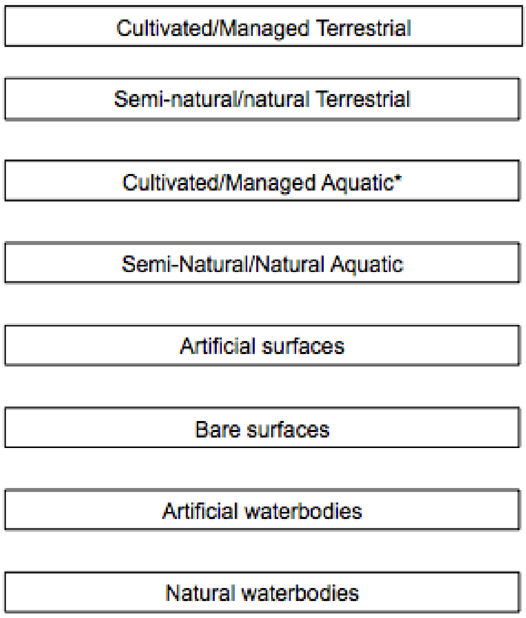


Figure 5.1 LCCS categories Level 1 to 3.

For the classification of LCCS Level 3, contextual as well as spectral information is required and a key component is to optimize the segmentation of the image such that object boundaries are the same dimensions as the objects being classified (e.g., individual buildings, water bodies). This optimization is achieved prior to the implementation of the classifications of LCCS Levels 1-2, and is being developed further through BIOSOS (to be reported on fully in D5.5). In undertaking the segmentation, a more targeted approach is being considered, with this focusing on delineating specific elements occurring within the landscape including those associated with urban infrastructure (e.g., roads, buildings), cultivated and managed agriculture (field boundaries) and water bodies (e.g., reservoirs). Following segmentation, classification using contextual as well as spectral information is undertaken for classes at Level 3 and beyond as outlined below.

5.1.2 Artificial surfaces (LCCS category B15):

Artificial bare surfaces are primarily represented by urban constructions, infrastructure and extraction or waste disposal sites. Within the urban environment, a wide range of structures exists with the LCCS defining these according to linear and non-linear features as well as buildings of variable density. However, within these categories (e.g., buildings), a wide range of materials is used and often these are sourced locally (e.g., roof tiles from nearby quarries) and hence exhibit spectral characteristics similar to

exposed consolidated or unconsolidated surfaces (e.g., rock, pebbles, base soils, mineral surfaces). As a consequence, mapping from the imagery themselves requires consideration of spectral indices or geometric features that provide discrimination and definition of these elements. Contextual information is particularly important as this can be used to extract, for example, buildings and roads. The concept of contextual information is nevertheless rather vague; it can refer to the use of the shape of the object (regular shapes for man-made objects) or can go beyond this first level of context by taking into account the spatial organization of the objects within the scene.

An example of the use of contextual information is provided for caravan parks (Figure 5.2). Caravans and associated dwellings (e.g., chalets) are constructed using different materials (e.g., tin or felt roofs) and hence there are no spectral features that can uniquely define these. However, a human is able to recognise caravan parks because of the spatial arrangement of objects (caravans but also other components such as amenity blocks) as well as the shape and size of the individual objects. On this basis, a simple workflow for the detection of caravan parks from EO data can be based on the following steps:

a) Perform a simple segmentation by thresholding based on the brightness of the image (caravans are typically brighter than their surroundings in the visible and near infrared wavelength regions); this is undertaken prior to classification at any LCCS level.
b) Retain only the regions that are compatible with the typical size of a caravan (lower and upper thresholds relating to size can be defined) and their compactness (i.e., caravans are typically not linear as would be in the case of roads).
c) Detect aligned objects and group these to define the concept of the caravan park.

The segmentation implemented in Figure 5.2 is based on that of Vanegas *et al*. (2009) and is implemented within the ORFEO Toolbox. The ORFEO Toolbox also executes algorithms for the extraction of roads from VHR images using the technique described by Christophe and Inglada (2007), an example of which is illustrated in Figure 5.3. The benefits of these algorithms are that they are fast to implement, have very few parameters and are only affected slightly by the image properties (spatial resolution, noise).

Other examples of feature extraction for artificial categories include the identification of extraction sites (quarries), which are associated with sinks (areas of internal drainage within a DEM), and also waste dumps (identified using a similar approach (if below the elevation of the surrounding terrain) or time-series of DEM data (if material is added over time, as in the case of waste dumps). Railways can also be identified as associated objects typically are of low slope and either raised above or sunk below the surrounding terrain.

5.1.3 Artificial waterbodies (B27).

Within the LCCS, water bodies may either be natural or artificial water, snow or ice. Where artificial water bodies occur, these are often distinct because of the presence of infrastructure (e.g., reservoir dams, water towers, car parks and roads leading up to and around the waterbody). However, confusion occurs where water bodies are associated with natural environments but the landscape has been modified to confine these to certain areas (e.g., dykes along straightened river bands, artificial ponds for attracting wildfowl). In the majority of cases, artificial waterbodies are standing or with gentle flows (as in the case of reservoirs). Areas of standing water can be identified by classifying all water areas within a scene and subsequently merging classified and also non-classified objects (assigned to a background). All water entirely enclosed by the background class is assigned to standing water whilst those that are connected to a larger body of water (e.g., the sea or major river) are assumed to be flowing (and by inference, natural). Confusion arises where a flowing water body lies between two bridges, for example, but distance rules can be applied allowing the algorithm to ‘jump’ over these in the classification. In the case of rivers and streams, many of which are obscured from the view of the sensor, the use of the slope surface generated from the DEM and hydrological modelling can be used in their identification (i.e., most

flowing rivers travel over sloping ground along a length). As in the case of caravan parks, artificial standing water bodies can be separated from natural water bodies because of the existence of infrastructure (e.g., dam walls and infrastructure, connecting roads and linear clearance associated with hydroelectric energy or water transfer.)

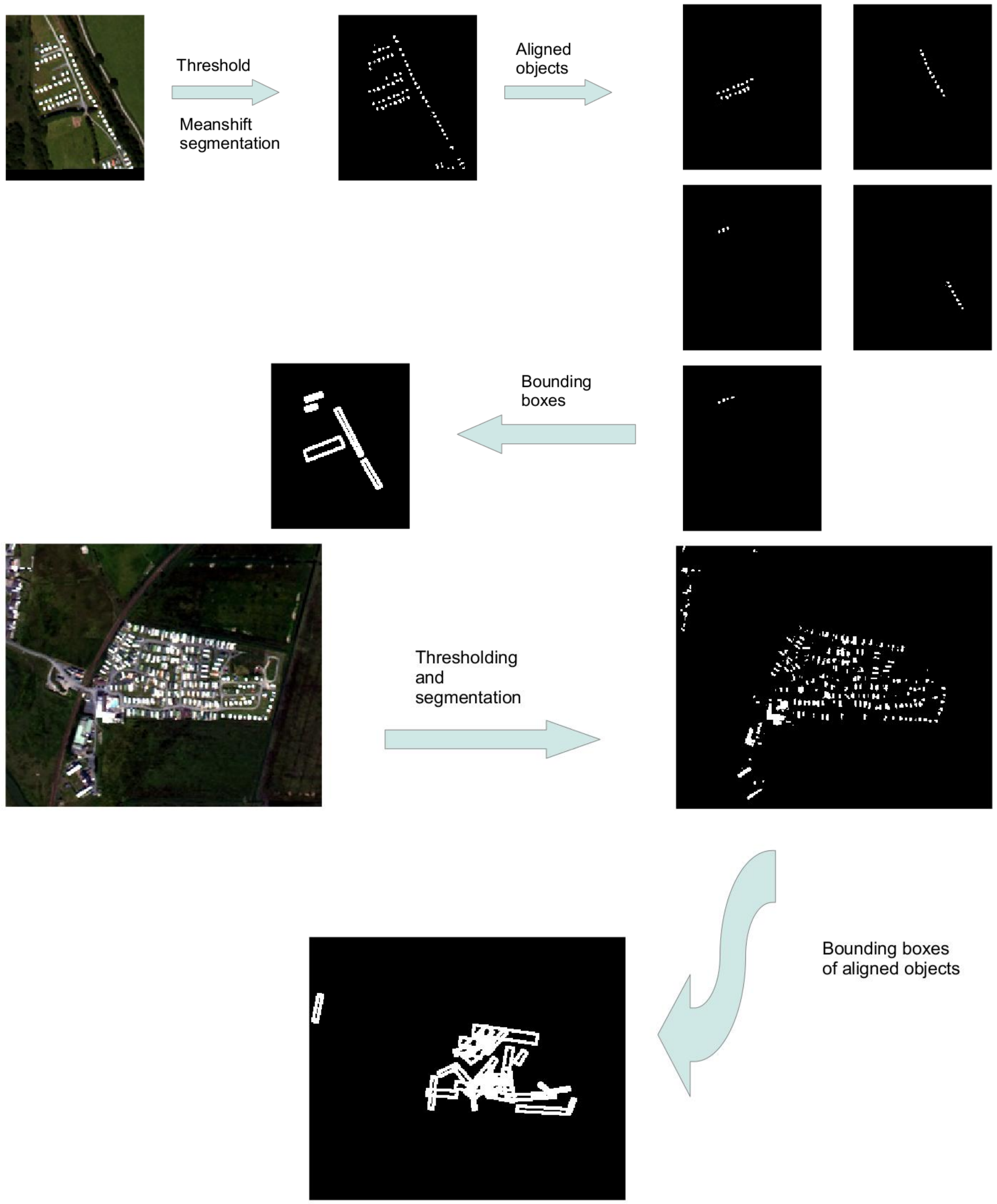


Figure 5.2. Approach to segmentation of caravan parks and associated features.

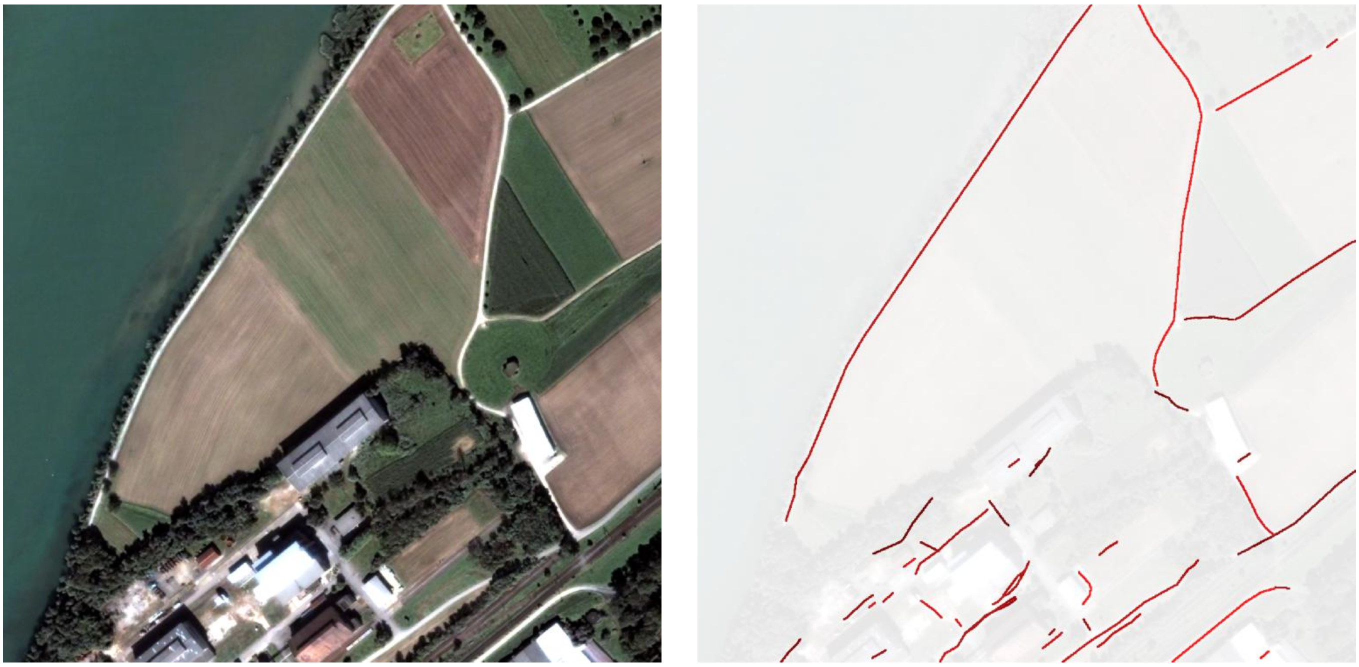


Figure 5.3. An example of road delineation based on the algorithm of Christophe and Inglada (2007).

5.1.4 Cultivated areas (A11 and A23)

Cultivated and managed landscapes are associated with a diversity of landscape elements (e.g., fields, plantations) and their components (individual trees, hedgerows and ditches). In many landscapes of northern Europe, these elements are quite distinct, largely because of their geometric structure, and have been mapped through manual interpretation of, for example, aerial photography and often to support reporting of agricultural area and yields. However, in many areas and particularly in the Mediterranean countries and in India and Brazil, such information is not available. Furthermore, the geometric arrangement of landscape elements is less distinct and different units (e.g., fields containing woody crops) may only be differentiated from others based on differences in, for example, crown size of individual trees making up the crops or row orientation). In such cases, the combined use of spectral and contextual information is essential. For example, in spectral terms, natural and semi-natural vegetation are very similar to managed vegetation (e.g., olive groves, vineyards and fruit orchards). Hence, different approaches to delineation are required. The following outlines approaches to the detection of features associated with cultivated areas including agricultural (field) units and woody crops (orchards and commercial plantations).

Agricultural units: Automated detection of cultivated areas can be relatively straightforward in regions where the agricultural unit (e.g., a field) is well defined and surrounded by containing elements such as hedgerows or ditches. In many cases, a large segmentation of the image can be undertaken which entirely captures the field unit (using only spectral information) and a finer segmentation can then capture the boundary elements, which are typically linear and relatively narrow. The ability to detect the boundary is dependent on the time of acquisition. For example, in Wales, differences between fields may be high in the summer months and in the visible imagery because the high reflectance from soil (e.g., ploughed fields) or shorter grass (e.g., because of grazing, cutting or drying) contrasts with the low values associated with hedgerow vegetation (because of higher foliage cover). In the winter, fields with graminoids are often highly productive and have a higher NIR reflectance compared to hedgerows where reflectance is low because of the deciduous nature of the vegetation. Hence, the wavebands used for classification will vary depending upon the season of acquisition but a similar result can be achieved.

The use of a larger segmentation to identify fields can however be compromised where there is spectral variability within the unit (e.g., as a result of encroachment of weeds, differences in fertiliser application or shadowing). Hence, several objects may be generated and these need to be grouped based on

proximity to a boundary feature and distance between and adjacency of objects. The approach to segmentation and merging of objects is, however, reasonably basic to implement, with examples provided in Figure 5.4

a) b)

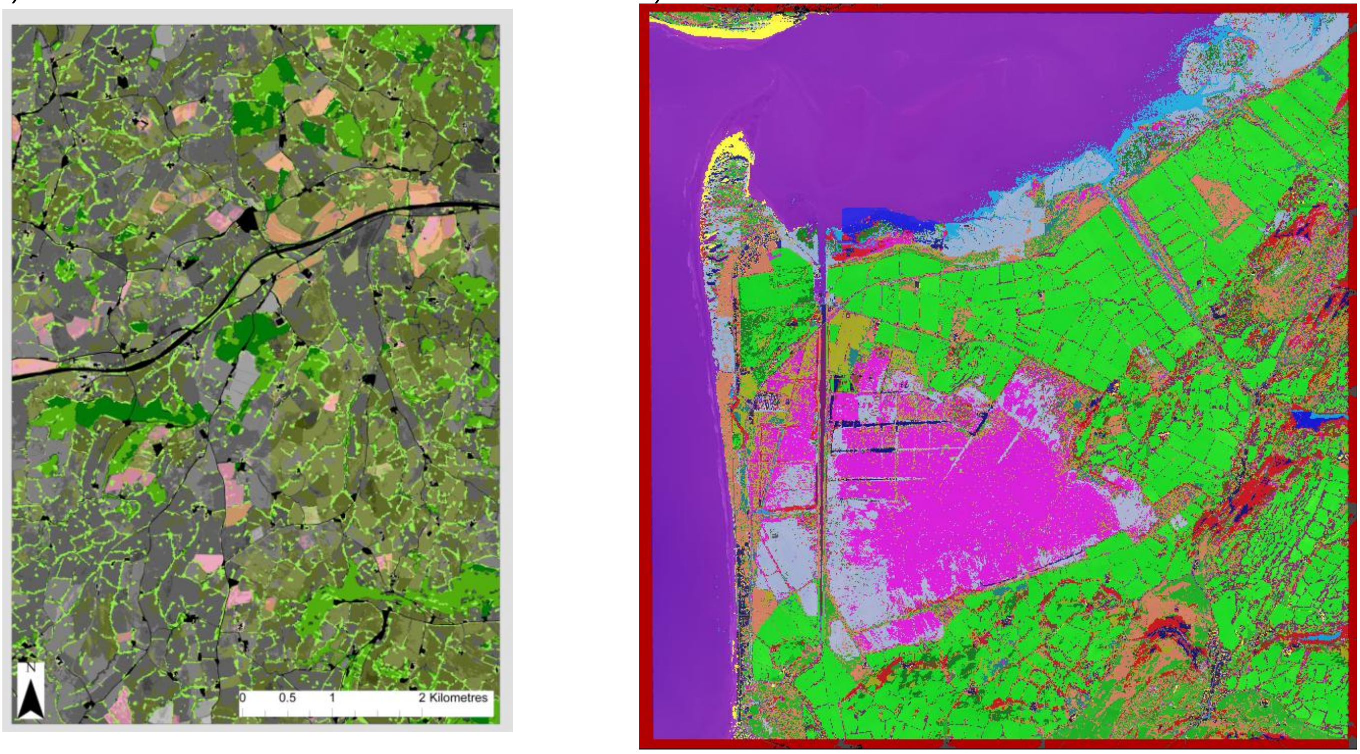


Figure 5.4. a) Delineation of field boundaries and associated agricultural units based on differential segmentation of the image (into larger and smaller objects for fields and boundaries respectively) and b) Cultivated areas defined through use of spectral and contextual rules.

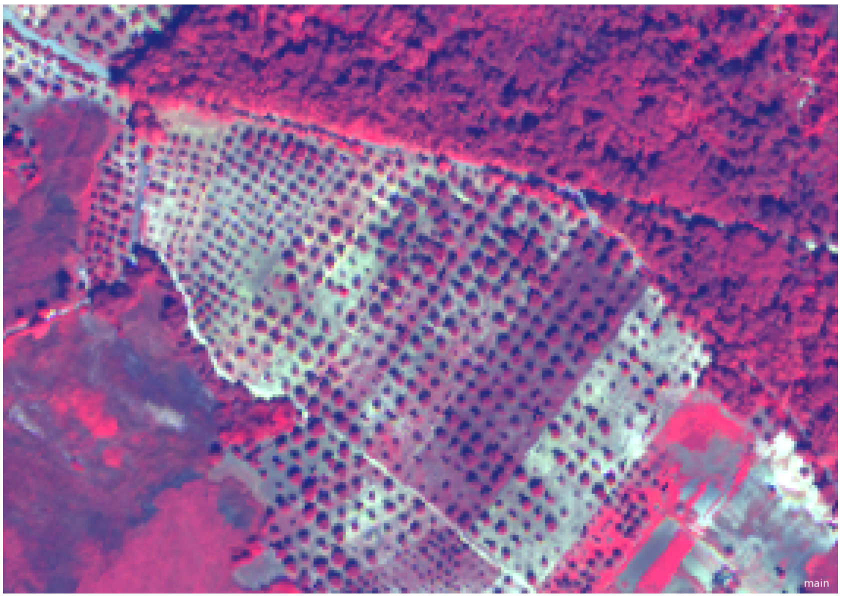
Tree crops: The detection of tree crops is often problematic because of a) the existence of a distinct foreground and background, as in the case of olive groves and orchards, and b) spectral similarity with natural and/or/semi-natural vegetation, with broadleaved plantations being an example.

Where individual trees are spaced apart such that the ground surface is visible from above (e.g., as in the case of olive groves), the variability in spectral reflectance of the background may be high, particularly where bare soil and vegetation at various states of moisture content and greenness respectively occur. This often compromises the detection of cadastral units, as does the lack of a distinct boundary element (i.e., a wire fence rather than a hedgerow). Hence, several approaches to extraction have been investigated, with these based on differences in row orientation but also consistency in the size, shape and proximity of tree crowns, as illustrated in Figure 5.5. In particular, where trees are of one orientation and tree size category, these can be distinguished from those of another, with the boundary placed mid-way and perpendicular to the trees occurring at each end of the orientation line. To delineate tree crowns, a range of algorithms are available including template matching and valley following. The delineation of tree crowns is, however, most successful where these are separated from each other (e.g., in olive groves) or are located within plantations and are of similar height and high density (e.g., coniferous forests). Where complex forest stands occur, delineation is more difficult particularly from optical data and where a diversity of tree species and layers occurs.

In the case of managed coniferous and broadleaved plantations of closed canopy, the orientation of high points identified in the forest block can be used to infer their existence. Many plantations are also similar in terms of their texture, being more homogenous spectrally compared to natural and semi-natural forests. Where units representing agricultural croplands, tree crops or managed forestry

plantations cannot be identified based on spectral signatures alone, cadastral information can be exploited if available. A limitation of these datasets is that unit boundaries may change over time, particularly in areas under pressure, and hence regular updates are needed. In India, many of the managed forests are those that are fragmented, being completely or partially surrounded by agriculture. Hence, these can be identified on the basis of having a relative border (e.g., > 50 %) compared to more intact and contiguous areas of natural forests. The presence of buildings and other infrastructure, which are commonly contained within the forest area, can be determined through reference to larger objects representing the extent of the managed forest.

a)



b)

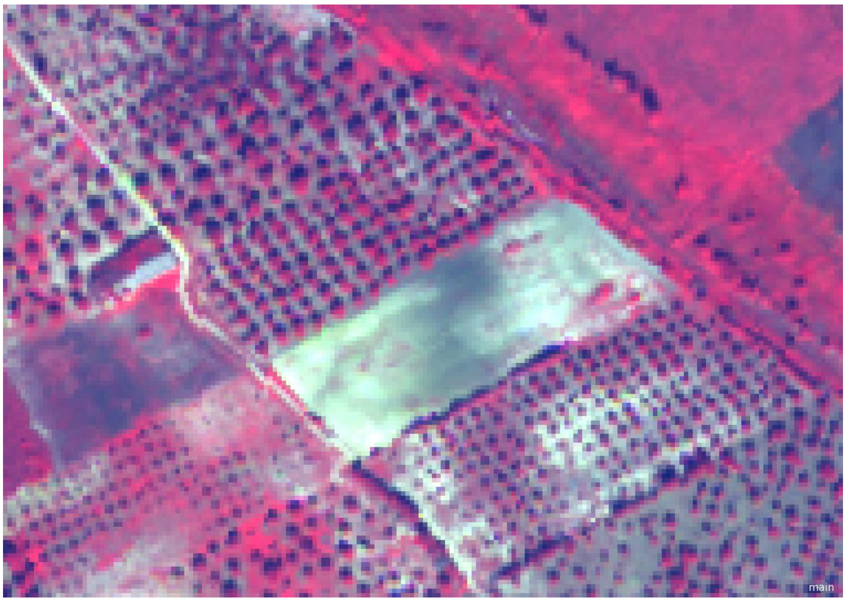


Figure 5.5. Differences in a) the orientation and b) the crown size of olive trees.

5.1.5 Natural/semi-natural vegetation (A12 and A24)

In some cases, feature extraction and contextual information can be combined to identify semi-natural or natural vegetation. As an example, Figure 5.6 highlights the existence of elevated banks with gullies in

between representing peat cuttings within the active raised bog, with these best visualised using minimum curvatures, as generated from the DTM. These features can be used with spectral information to separate the two Annex I habitats of the active raised bog and modified raised bog. Similarly, sinks within the DEM can be associated with bog pools within the active raised bog whilst locally high points correspond to hummocks, each of which contain distinct floral communities.

a) b)

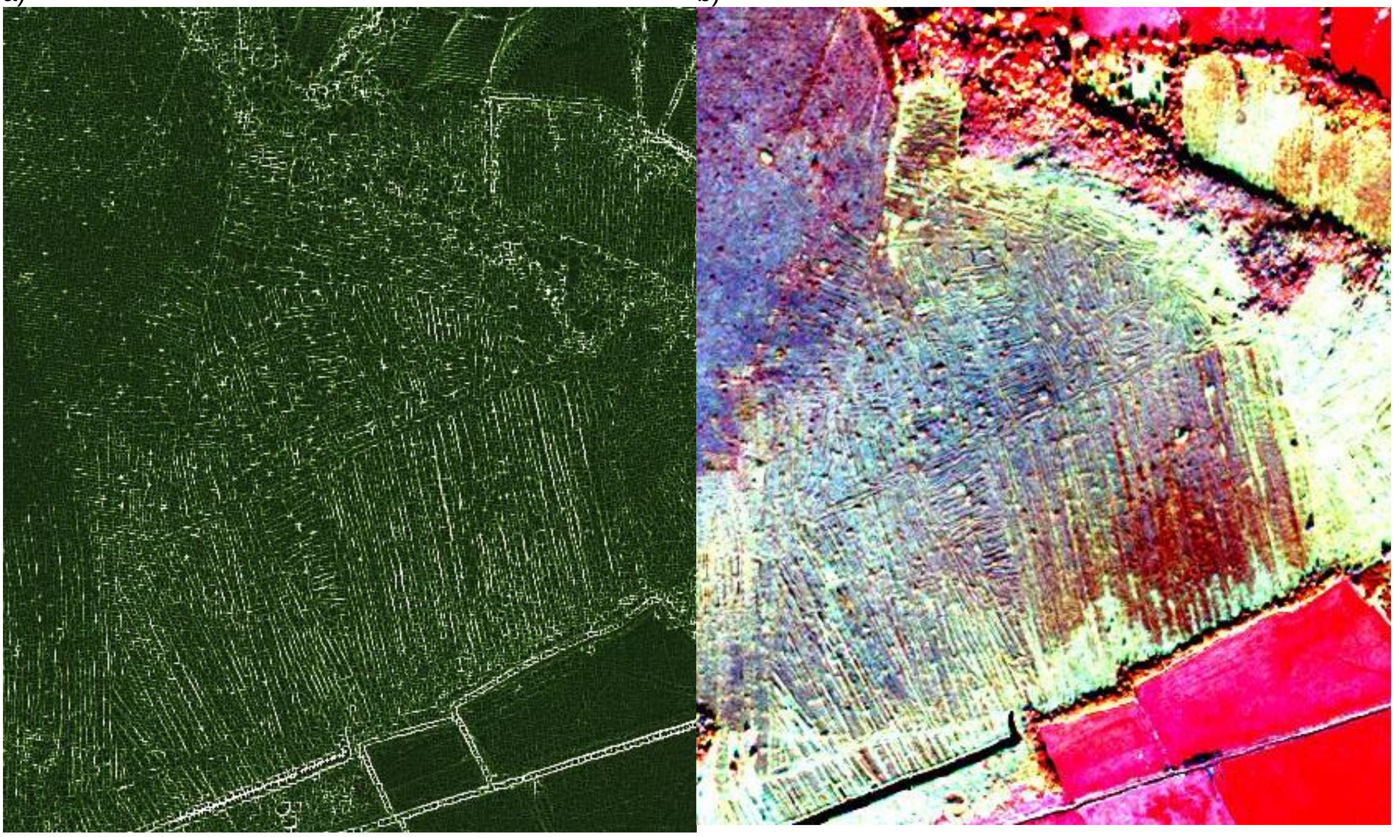


Figure 6.6. a) Minimum curvature surface indicating alternations in elevated banks and gullys associated with peat cuttings and b) a WV-2 false colour (R:NIR, G: Red, B:Blue) composite of the banks and gully terrain at Cors Fochno raised bog. Note that the modified raised bog occurs in the centre of the image whilst the active (undisturbed) bog is towards the top left.

5.1.6 The requirement for contextual information

Following on from LCCS Level 2, Level 3 separates areas of cultivated and/or managed vegetation from semi-natural vegetation and also artificial from natural bare surfaces and waterbodies. For the Welsh and Dutch sites, existing thematic layers representing cadastral information and infrastructure were used to separate semi-natural and natural surfaces from those that are more influenced by humans. However, the BIOSOS project seeks to formulate methods for automated extraction from spectral information and, for this reason, methods are being implemented to reduce dependency on these (see above). A number of options are available but, in many cases, contextual information as well as expert knowledge class description is needed, although this can also make use of spectral data and derived measures. For example, hedgerows can be identified as contrast with the reflectance of the fields they envelope. Region growing rules can also be used to grow ‘seeds’ located within sections of the hedgerow to encompass the entire length. Large objects surrounded by hedgerows can then be associated with cultivation (which includes permanent pastures). Artificial water bodies can also be identified by considering infrastructure surrounding areas of open water.

Within EODHAM, appropriate segmentation is critical. The preference is for the selective segmentation of particular features within the landscape followed by a segmentation of remaining areas using, for example, the algorithm of Shepherd *et al*., (2012). The spectral bands and indices for segmentation

then need to be specific to the features being considered. For example, indices such as the Water Band Index; WBI) can be used to segment areas of water whilst those best representing non-vegetated surfaces (e.g., the NDVI) can be used to segment urban infrastructure. In each case, the segmentation needs to focus only on identifying and delineating those features that are described by and make up the LCCS categories (e.g., high density industrial buildings).

5.1.7 Evaluation of spectral indices.

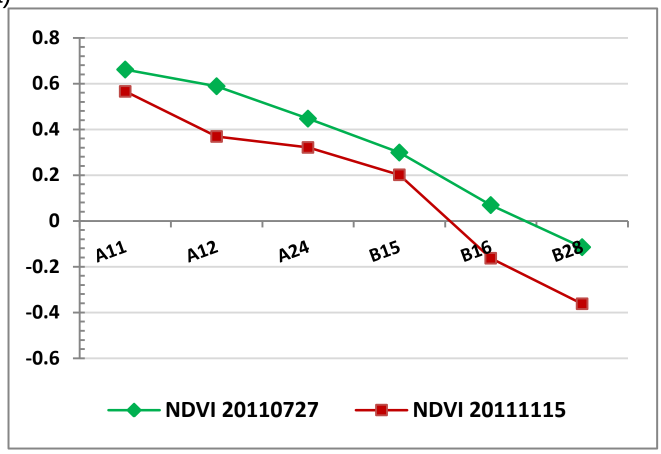
The 1st Stage of the EODHAM system relies heavily on spectral indices and the choice of these is critical. In order to ensure that the proposed workflow is robust in relation to differences between different years and also to different sites with different landscape characteristics, a detailed analysis of a range of spectral indices available in the literature can be undertaken to assess their usefulness for the classification of LCCS categories in Level 3 and beyond. The following sections outline the approach to identifying optimal indices for mapping classes associated with the LCCS classification.

For key indices, an analysis of separability of key categories was undertaken in two stages:

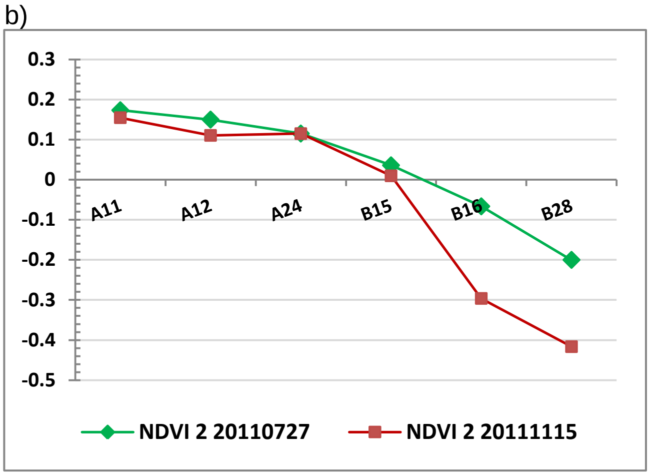
a) Several indices were computed (as in Tables 2.2 and 2.3) and compared in order to determine the degree of separability of the classes associated with LCCS Level 3 as well as the stability between dates. In the analysis, it was recognized that some indices require the availability of specific spectral bands, which may not be imaged by some sensors. This point is of major interest for the generalization of the approach.

b) The pertinence of the features used to classification LCCS categories to Level 3 was determined and ranked accordingly so that the impact of a missing feature could be estimated. This mapping was done using decision tree learning, which allowed generation of classification rules in an IF/THEN format, similar to the ones used in SIAM.

a)



b)



b)

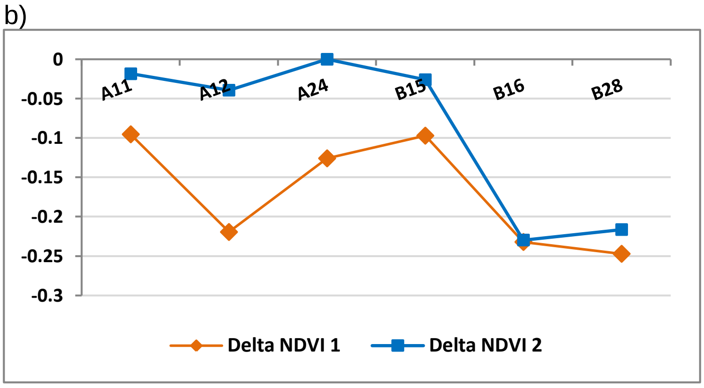


Figure 5.7. Comparison of vegetation indices for cultivated agricultural land (A11), terrestrial and aquatic natural and semi-natural vegetation (A12 and A24 respectively), artificial areas (B15), bare areas (natural; B16) and natural waterbodies (B28) based on a) the NDVI(1), b) NDVI(2) and c) the difference between the NDVI(1) and NDVI(2) for the two dates (July and November, 2011).

In the first case, some examples are presented based on WV-2 data. For each of the LCCS Level 3 classes, the mean values for each band on two acquisition dates (July and November, 2011) were

calculated and differences in these for these two dates and for each class were highlighted. Some examples are presented as illustration of the approach based on the NDVI(1) and NDVI(2), which are sensitive to the presence of photosynthetic and non-photosynthetic/senescent vegetation respectively. The comparison (Figure 5.7a) indicated that the NDVI(2) provided better separation for vegetated categories, with the NDVI(2) providing best separation for bare and water classes (Figure 5.7b). The difference between the values for each of the two dates was less in the NDVI(2) compared to the NDVI(1) (Figure 5.7c). As a second example, a comparison of the WBI(1) and WBI(2) indicated that the latter provided better discrimination (Figure 5.8). Similar analyses are being carried out currently for indices related to bare surfaces and built-up areas.

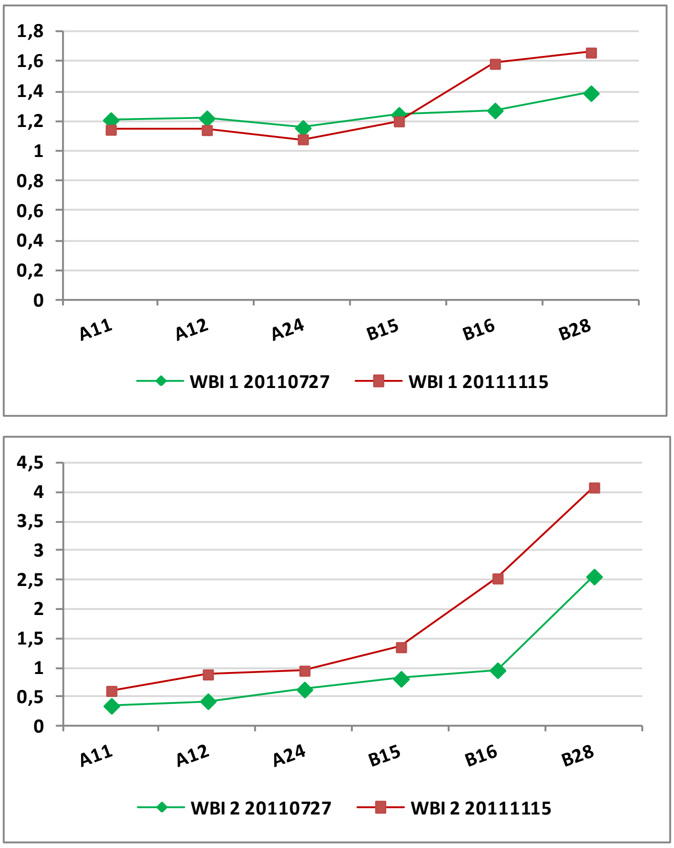


Figure 5.8. Comparison of the separability of open water.

5.1.8 Decision tree learning

As stated above, the goal of the feature analysis using decision trees is to assess quantitatively the pertinence of the different spectral features available in the result obtained using the workflow implemented in eCognition. The main motivation for this assessment was the estimation of the impact of a lacking feature, for example in the case where no WV data is available and QuickBird (QB) imagery is used instead. In this case, since some spectral bands are missing, some indices cannot be computed. The same approach can also be used to assess the information content of other features (texture, LIDAR data, ancillary data) and this study will be reported later. Once this feature selection (ranking) and analysis is streamlined, the results between different test sites can be compared.

The approach adopted for the feature selection is to use machine learning algorithms to find decision rules to map spectral features into LCCS level 3 classes. For this purpose, the land cover classification produced using the workflow implemented in eCognition was used as target for the mapping. The features which were used were the 8 spectral bands of one single WorldView date (July) and 3 spectral indices, namely NDVI(1), PSRI and WBI(1), since these were the primary indices used for the determination of the dichotomic LCCS workflow described above.

The machine-learning algorithm used was the decision tree due to the fact that, in contrast to other approaches such as neural networks, decision trees represent rules. Rules can readily be expressed so that humans can understand them and their implementation is straightforward. The main disadvantage of decision trees is that they may be less accurate than other approaches, since the thresholds define rectangular regions of the solution space, but this can be overcome by the appropriate choice of non-linear combinations of features (and thus the importance of exploring many different spectral indices). An example of a sub-tree from a decision tree obtained by supervised learning using the inputs and outputs described above is outlined below:

```
ndvi <= 0.39: A12 (29/6.2)
ndvi > 0.39:
:...psri <= -0.34: A24 (637/0)
psri > -0.34:
:...psri <= -0.34: A12 (34.4/2.5)
psri > -0.34: A24 (522.3/56.8)
```

The values presented as (X/Y) for each leaf of the tree are the number of examples correctly (X) or incorrectly (Y) classified by the rule.

Decision trees presented like this can be difficult to read when the number of classes is large. However, it is easy to re-write a decision tree as a set of rules as for instance:

```
IF    ndvi > 0.52   AND
      psri > -0.09  AND
      wbi > 1.17    AND
      wbi <= 1.120
      ->  class A12
      confidence [0.925]
```

It is worth noting that several rules may lead to the same class, but since they are applied in decreasing confidence order, this is not a problem.

The estimation criterion in the decision tree learning algorithm is the selection of an attribute, at each decision node in the tree, that is most useful for classifying examples. A good quantitative measure of the worth of an attribute is a statistical property called information gain that measures how well a given attribute separates the training examples according to their target classification. This measure is used to select among the candidate attributes at each step while growing the tree. This very same measure can be used to rank the set of features with respect to their pertinence. In Tables 5.1 to 5.3, three examples of feature ranking for the same data set used above are presented together with the associated classification accuracy. A further property of decision trees is that they can be learnt and used for classification even when some features are missing. This property has not yet been investigated at this stage, but will become very useful for the generalization of the approach.

The use of decision trees for this task needs further analysis, since the number of rules obtained is high and, for generalization purposes, it would be interesting to generate compact rule sets. This can be achieved by using several available image acquisitions from different dates, but more generally, the use of a richer set of features needs to be investigated. These features could include other vegetation indices, texture and statistical features. Finally, trying to learn the decision trees jointly over different test sites is anticipated to give a good insight on the degree of generalization within the EODHAM System.

Table 5.1. Feature ranking based on reflectance values only

Rank	Feature	Information content (normalized)
1	B8	1.00
2	B6	0.96
3	B3	0.92
4	B7	0.75
5	B1	0.70
6	B2	0.69
7	B5	0.58
8	B4	0.38
Classification accuracy		**80.3%**

Table 5.2. Feature ranking based on indices only.

Rank	Feature	Information content (normalized)
1	NDVI	1.00
2	PSRI	1.00
3	WBI	0.84
Classification accuracy		**77.6%**

Table 5.3. Feature ranking based on reflectance values and spectral indices

Rank	Feature	Information content (normalized)
1	B8 (NIR2)	1.00
2	B4 (Yellow)	1.00
3	B3 Green)	1.00
4	B7 (NIR1)	0.97
5	B6 (Red Edge)	0.84
6	WBI	0.80
7	B1	0.77
8	B2	0.73
9	PSRI	0.73
10	NDVI	0.70
11	B5	0.51
Classification accuracy		**80.5%**

5.1.9 Classification incorporating the Dempster-Shafer theory

As previously stated, differences in the availability of data from site to site is a common case, especially when studying diverse regions and landscapes. As an example, the optical data might not come from the same sensors for all sites while LiDAR information available for one site might be unavailable for another. Also, uncertainty both in the input data, due to quantization errors or noise, and in the expert rules inserted to the classification process, needs to be taken into account in a robust classification scheme.

The Dempster-Shafer theory, providing a framework able to handle uncertainty and missing information, is expected to add robustness to the BIO_SOS classification scheme and improve its generalization performance, both up to Level 2 of LCCS classification and levels 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, 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 case all missing information were in favour of this event. This formalism allows a classification scheme based on this theory to express naturally the uncertainty in the different events, providing the lowest and highest degree of confidence that they hold.

The framework allows the assignment of belief intervals to multiple events in addition to single ones. In our case, this property is particularly useful in handling missing information, under the lack of which the discrimination of two or more potential classes for a specific landscape patch is not possible. 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 either to the one of the other class and the classification process continues. In the appearance of appropriate evidence, the classification is refined through the assignment of belief intervals to the individual classes.

The Dempster-Shafer theory can be used either for the design of an individual classifier or for the fusion of two or more classification results produced from other classifiers. In the first case, the system is based on rules provided by experts, including field ecologists and botanists, or rules extracted from experimentation, simulations and machine learning algorithms (e.g., through eCognition or decision trees) as discussed above. The validity of the conditions dictated by each rule are examined and propagated and a final belief interval is given to each patch. A particular effort is given on fuzzification approaches to be included in the algorithm, in order to counteract errors in data and inaccurate crisp rules provided by the experts. The framework offers great flexibility in incorporating various fuzzification schemes, thus making the algorithm more robust to changes of thresholds in rule conditions, small changes in reflectance values of the data, etc.

The theory can be used also as a way to fuse classification results produced from individual classifiers. In our case, such classifiers may involve different rule sets applied in eCognition on the same site, different decision trees, or other employed classification approaches. The results of all classifiers are merged based on the Dempster rule of combination and, as previously, belief intervals are assigned to each patch for its potential classes.

The final classification result for each patch may arise from different approaches. The final class for a specific patch may be decided based on the maximum belief value, compared with the other potential classes, the maximum plausibility, or a combination of the two. In summary, the Dempster-Shafer theory offers a great flexibility in rule-based classification and is expected to be particularly effective 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.

5.2 Beyond Level 3

5.2.1 Overview

In the subsequent classification of LCCS categories, reference is made to additional layers (e.g., life form, spatial aspects, surface aspect, physical status and persistence; Figure 5.9). Within these, a range of subcategories exists, which are formally listed or can be described (e.g., in terms of surface material).

To replicate the dichotomous sequence of classification within the LCCS, a series of layers was created with these then classified separately in subsequent stages. The layers were created following the generation of LCCS categories 1 to 3, as indicated in Figure 5.1, with the layer representing the LCCS Level 3 categories (including the classification and segmentation) copied multiple (31) times to levels below. Each of these levels was created to represent key descriptors in the LCCS classification. Four additional layers were generated below the lowest level, with these segmented to a pixel level and used primarily for subsequent estimation of life Form, canopy cover, vegetation phenology and surface materials. The construction of the layers is illustrated in Figure 5.9 with descriptors identified in Figure 5.10.

The following sections describe the reasoning behind the generation of the layers for the classification of LCCS categories and indicate reflectance bands and derived measures (e.g., indices, contextual information, ancillary layers) that allow discrimination of key features. The benefit of the approach is that the potential exists to generate the layer classifications from a diversity of remote sensing data without reliance on a single sensor.

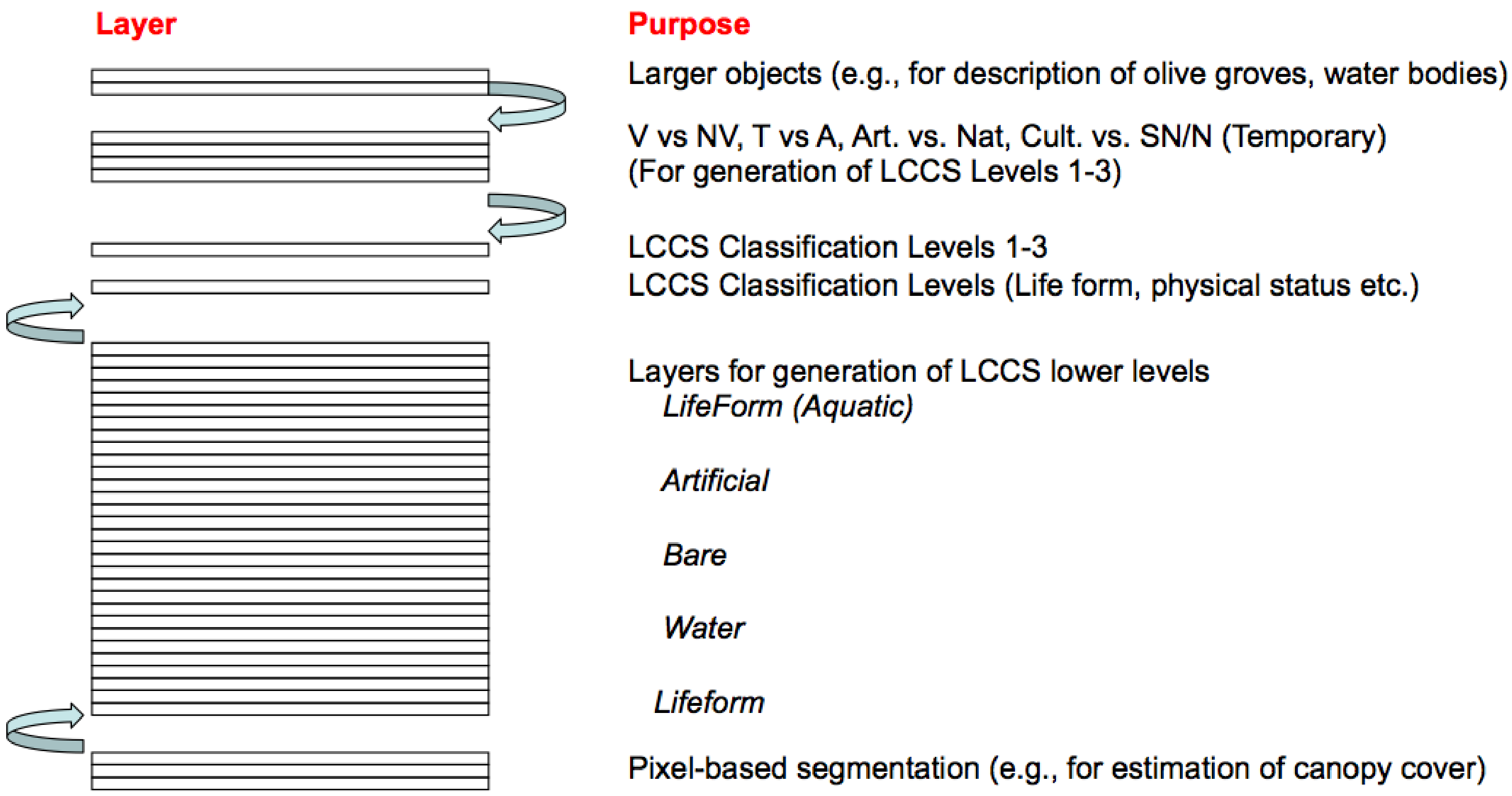


Figure 5.9. Sequence of layers generated to support the classification of LCCS categories.

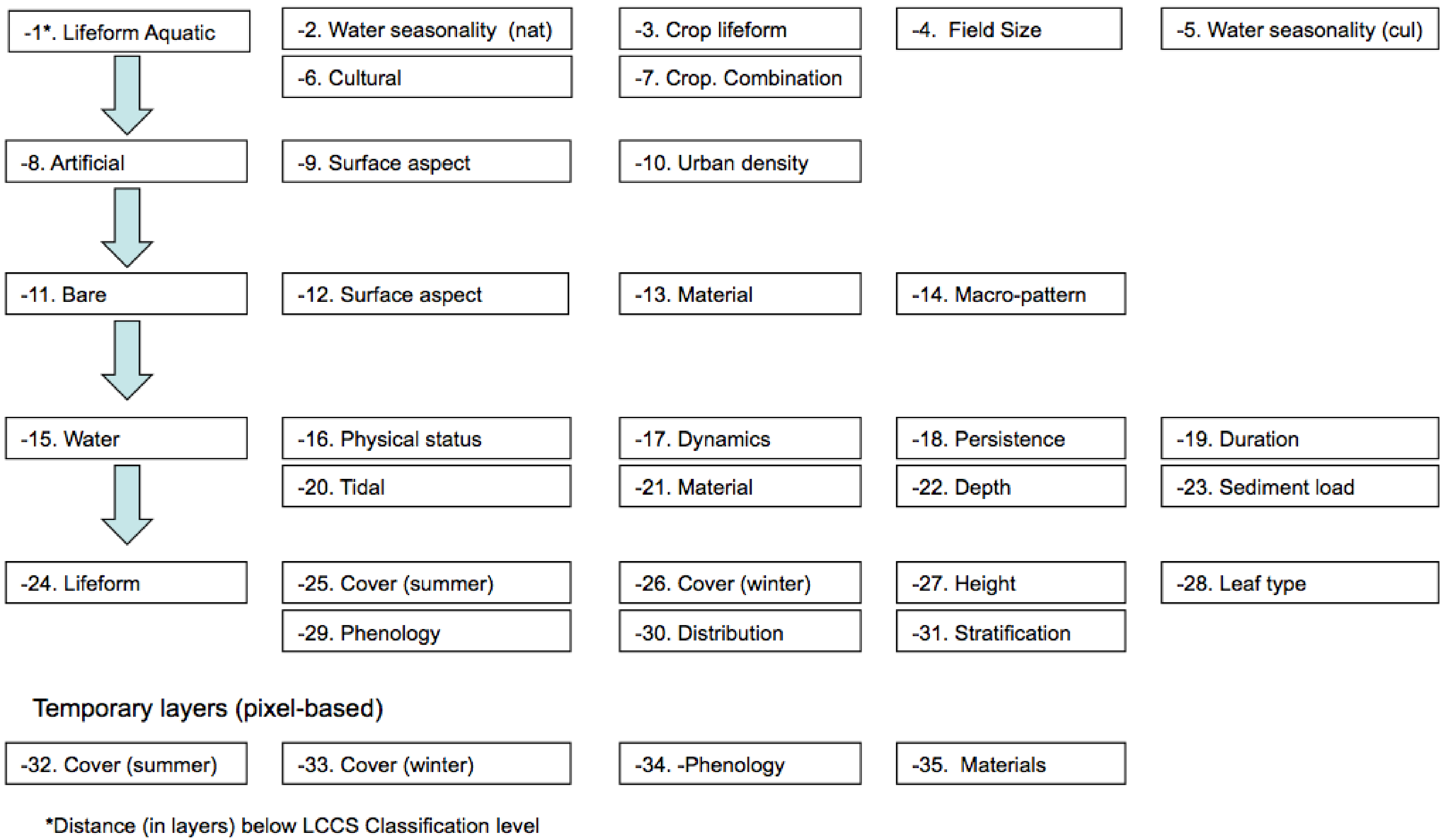


Figure 5.10. Description of layers and their location below the level of the LCCS classification.

5.2.2 Aquatic vegetated

Lifeform and water seasonality: The EODHAM approach to the classification of life forms (vegetation structure) and water seasonality is outlined in later sections dealing with the classification of vegetation and aquatic systems.

5.2.3 Cultivated/managed areas

Cultivated and managed areas are identified within both terrestrial and aquatic environments and can be described on the basis of crop life form, field size, water seasonality, land use activities and crop combinations, as outlined below.

Crop life form: The LCCS describes the life form on the basis of the main crop, with this defined as being trees, shrubs or herbaceous (graminoid or non-graminoid). Differentiation of these (e.g., as a function of phenology) is described in the later section on life form. However, a key additional component is the identification of plantations and tree crops (e.g., orchards, olive groves) and urban vegetated areas (Figure 5.11).

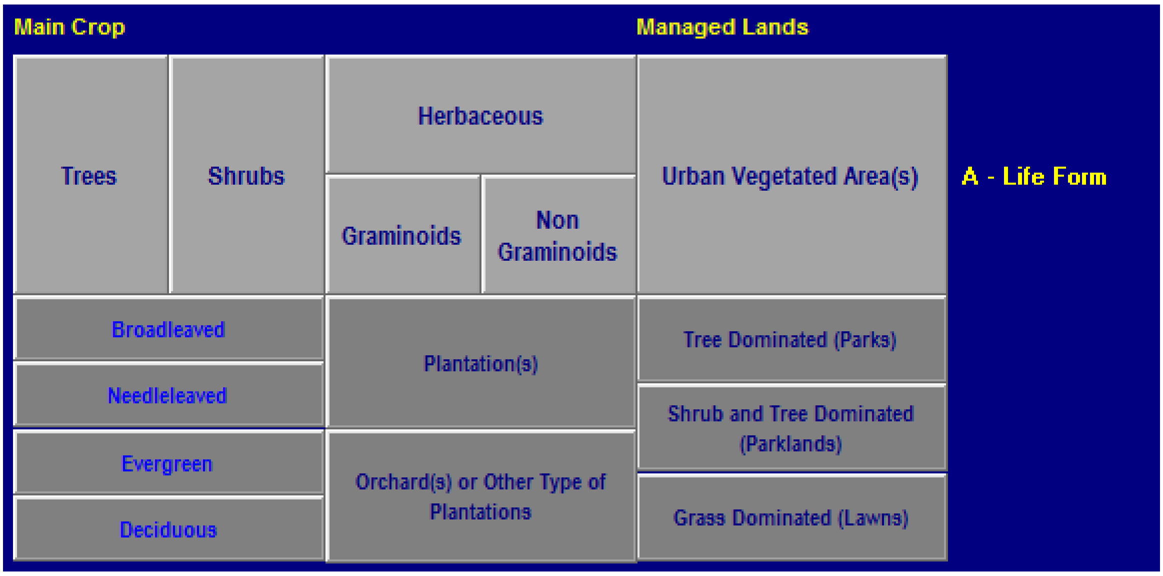


Figure 5.11. Main life forms and associated descriptors used in the LCCS classification of cultivated/managed areas.

In the case of plantations, measures that focus on the homogeneity of forest canopies are required, with textural measures of reflectance bands or indices (e.g., the WV-2 coastal-red) being particularly useful. For the mapping of tree crops in Wales, the system exploits larger objects associated with updated cadastral units, with counts of tree crowns and descriptors of row orientation contributing to the classification of LCCS categories. For other sites (e.g., Le Cesine), geometric features and additional contextual features based on class description (ontologies) are being used. Managed plantations can also be identified as these often occur as fragments within landscapes dominated by agriculture although care needs to be taken in their classification.

Field size: The LCCS system defines three sizes of fields; small (< 2 ha), medium (2-5 ha) and large (> 5 ha; Figure 5.12). In Wales, field sizes are quantified in the upper ('large objects') level in the hierarchy where each object represents the cadastral unit, defined either through reference to existing thematic information (if available) or that generated through segmentation of the image and subsequent feature extraction. The field size classification is then transferred to smaller objects in lower levels of the hierarchy (i.e., at the level of the LCCS classification). For optimal identification and classification of small fields (< 2ha), HR or VHR satellite images are needed.

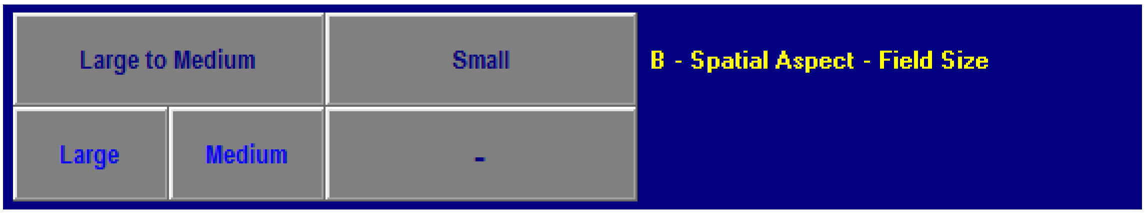
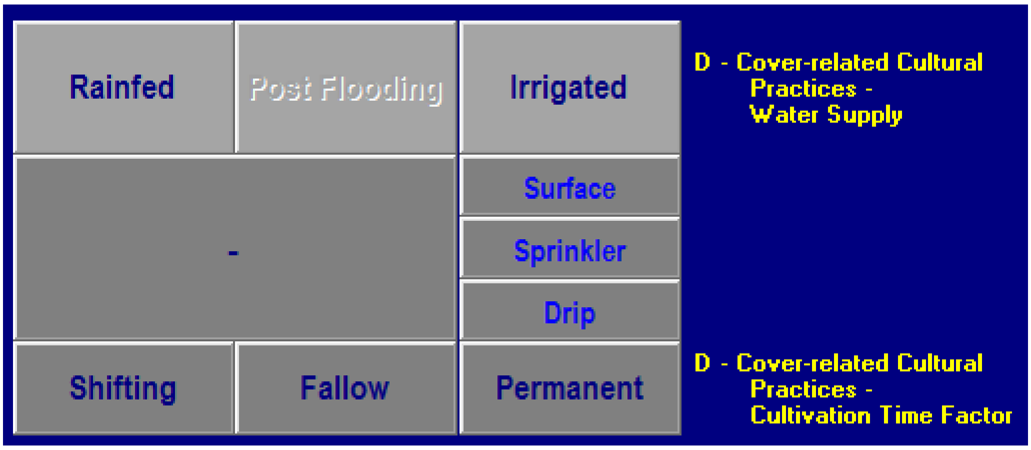


Figure 5.12. LCCS descriptors of field size.

Cultural: Within both terrestrial and aquatic cultivated/managed environments, a range of cultural practices occurs. For example, water for cultivation in terrestrial landscapes is either through rainfall (e.g., as in many northern European countries) or is provided through irrigation (as in many Mediterranean countries; Figure 5.13a). In the latter case, the productivity (as reflected in indices such as the NDVI) is artificially high, and distinct patterns are often associated with surface, sprinkler and drip methods. Within rainfed systems, the reflectance characteristics vary temporally and depend upon the rainfall amounts and variability. In some cases (e.g., in India), the natural flooding cycles facilitate irrigation and production of aquatic crops (e.g., rice).

a)



b)

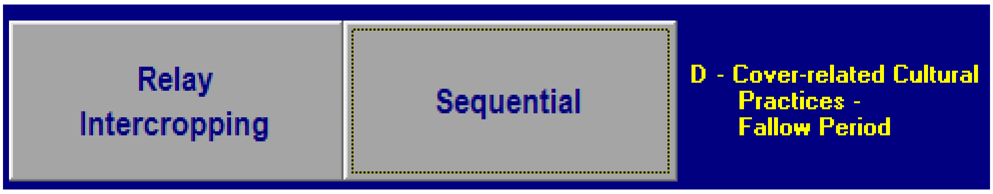


Figure 5.13. LCCS descriptors of cultural practices associated with the cultivation of a) terrestrial and b) aquatic cultivated/managed environments.

The timing of cropping is also important. In many northern European countries, the cover of cultivated vegetation is permanent but a rotation is evident in all regions, with the introduction of fallow periods commonplace and occurring typically on an annual or, less often, a sub-annual basis. In many tropical countries, shifting cultivation is more prevalent whereby the forest is cleared for agriculture, used for a number of years (e.g., 3-4) and then abandoned to regenerating forests for periods often exceeding several decades. The use of time-series observations using EO data is therefore essential; not only to distinguish all major agricultural crops and other vegetation types, but also to derive temporal classifications of agricultural land and clearances from which frequencies of use and turnover can be determined. As an example, a time-series classification of the age of regrowth forests, the periods of active land use prior to abandonment and the reclearance frequency can be used to identify agricultural land, regrowth and mature forest at the BIOSOS Tapajos test site south of Santarem in the Brazilian Amazon (Figure 5.14). Within aquatic systems, the fallow period may be associated with relay intercropping or may be sequential.

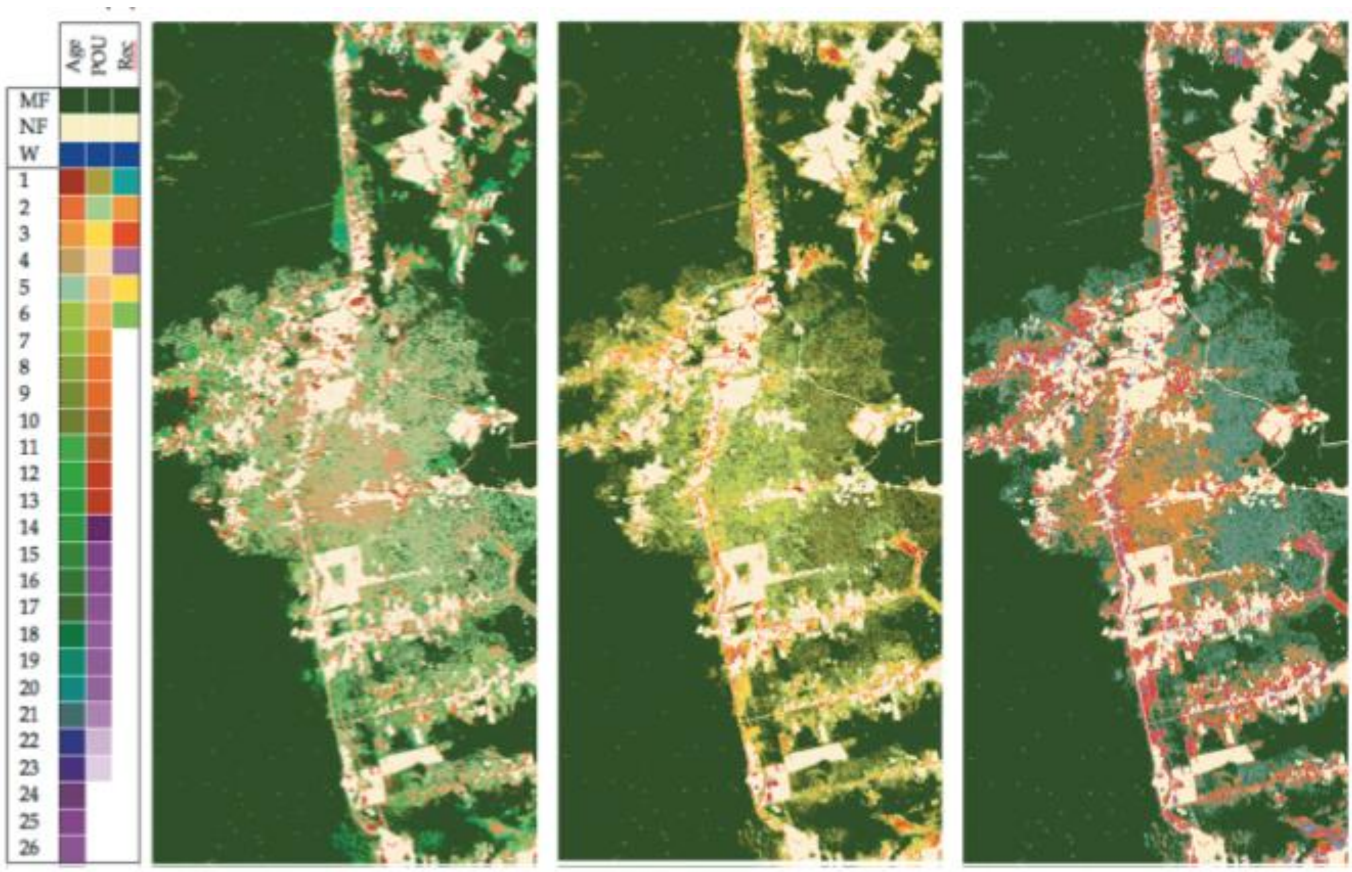


Figure 5.14. Classifications of the age of regenerating forest, the periods of active land use (POU) prior to agricultural abandonment and the frequency of reclearance (Rec) of vegetation for an area south of Santerem in the Brazilian Legal Amazon.

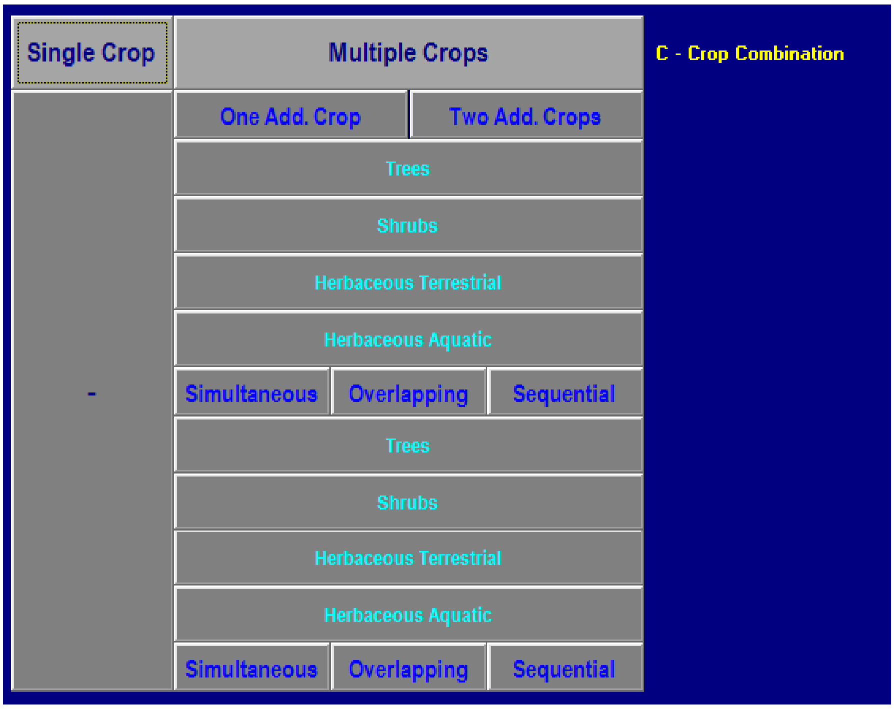
Crop combinations: Within both terrestrial and aquatic systems, the EODHAM system requires, in most cases, time-series of satellite sensor observations in order to identify different crop combinations, particularly if multiple crops are considered (Figure 5.15a and b). The timing of the observations is also important depending upon the type of crop and the agricultural practices followed. The selection of dates depends on an expert knowledge-based class-specific description, and it is important to identify the minimum number of images needed for a reliable classification and for reducing the costs of multiple-date acquisition. In some areas of Northern Europe, thematic information on agricultural crops, yields and cycles is associated with cadastral information as part of reporting requirements and can be exploited in order to enhance the reliability of the classification. In some cases, however, the cropping patterns and sequences can be inferred through reference to spectral signatures and also the SAR backscatter (e.g., as provided by Terra-SAR-X) associated with specific crops.

5.2.4 Bare areas

Areas that are naturally bare (e.g., mudflats, sand dunes) can be described on the basis of surface aspect (including material) and macropattern (Figure 5.16).

Surface aspect: Material is either considered to be consolidated or unconsolidated, with the former associated with bare rock a/o coarse fragments or hardpans and the latter with bare soil (and other unconsolidated material) or loose and shifting sands. The classification of these categories is undertaken only for objects classified in Levels 3 as bare (natural).

a)



b)

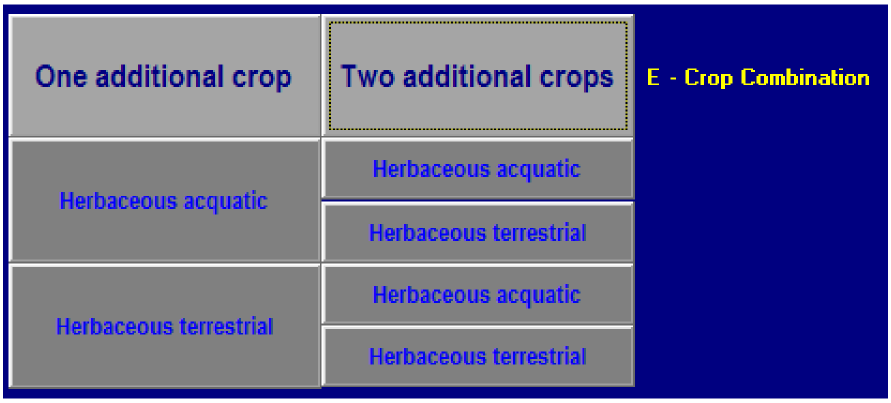


Figure 5.15. LCCS descriptions of crop combinations for a) terrestrial and b) aquatic cultivated areas.

Consolidated material is typically (although not exclusively; e.g., as in the case of limestone pavements) associated with rocky outcrops with some slope. Unconsolidated material can also be found on steep slopes (as in the case of high mountain scree) but can often be differentiated on the basis of textural measures or context (i.e., below and adjacent to a cliff). In many coastal areas, unconsolidated surfaces include mudflats, sandflats and dunes. These areas are typically confined to low elevations of low slope. Sand is often the most spectrally distinct, particularly in the visible wavelengths where the comparatively high reflectance contrasts with that of other bare areas and vegetation. Spectrally, however, these surfaces can be highly variable, particularly if observed during low tide (i.e., when the surface material is wet). At the other extreme, unconsolidated material can occur in high mountain areas (e.g., debris associated with glaciers and hydrological flows) or inland sand dunes, as in the Netherlands.

Table 5.4. Reflectance channels and indices used to unconsolidated from consolidated material.

Index/fraction	Measure
Visible (green)	Increased reflectance from brighter surfaces, particularly sand
Texture (green)	Entropy
Elevation	Range (DTM_{max} - DTM_{min})
Slope	Rate of change of elevation, calculated in degrees or percentage

Materials: Although related to surface aspect, a separate classification of surface materials benefits the discrimination of LCCS classes relating to bare ground. In particular, textural measures can be used to separate materials such as gravels, stone and boulders, including those generated using SAR data. Estimates of the percentage cover for each within larger objects can be obtained using the same approach adopted for estimating the canopy cover of vegetation (see Section 5.2.7). The differentiation of hardpans (i.e., ironpan/laterite, petrocalcic and petrogypsic) is, however, difficult without the use of ancillary information.

Macropattern: The macropatterns relates to the horizontal spatial distribution of elements within the landscape and, with the LCCS, dunes, saltflats, gilgai and termite mounds are defined. The dune systems are divided further into barchans, parabolic dunes and longitudinal dunes, all of which are associated with unconsolidated material. In the description of dunes, which are common to the BIOSOS sites, these can be differentiated from other loose and unconsolidated sands (e.g., beaches) through reference to slope and other topographical features.

a) b)

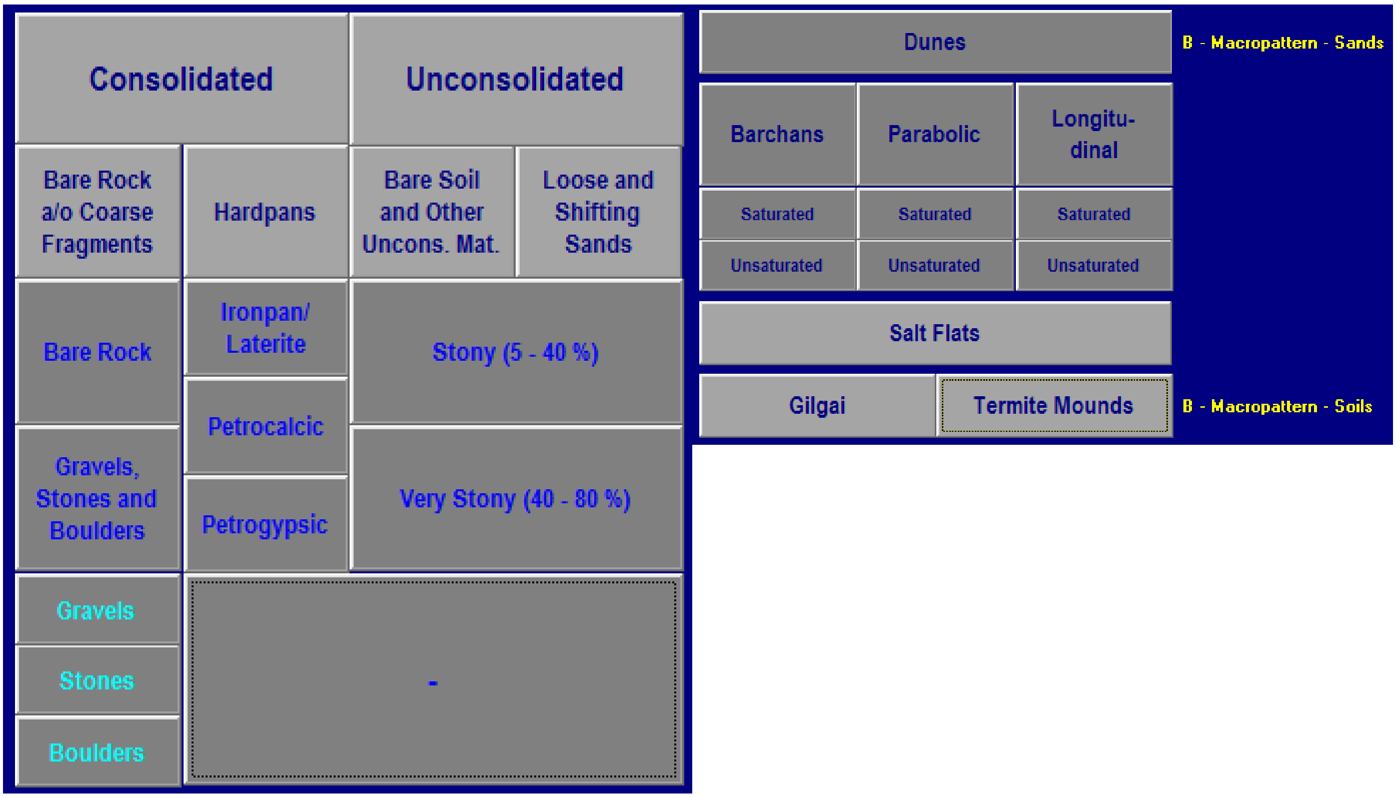


Figure 5.16. LCCS categories for bare areas.

5.2.5 Water bodies and aquatic environments

Water bodies are diverse in their extent and, in the LCCS, are described on the basis of physical status, persistence, depth and sediment load. However, additional layers are also needed to establish whether water of ice is moving or stationary, the duration of cover, the presence of tidal influence and the substrate (Figure 5.17).

Physical status: Differentiation of water, snow and ice is achieved primarily through reference to spectral indices, as listed in Table 5.5.

a) b)

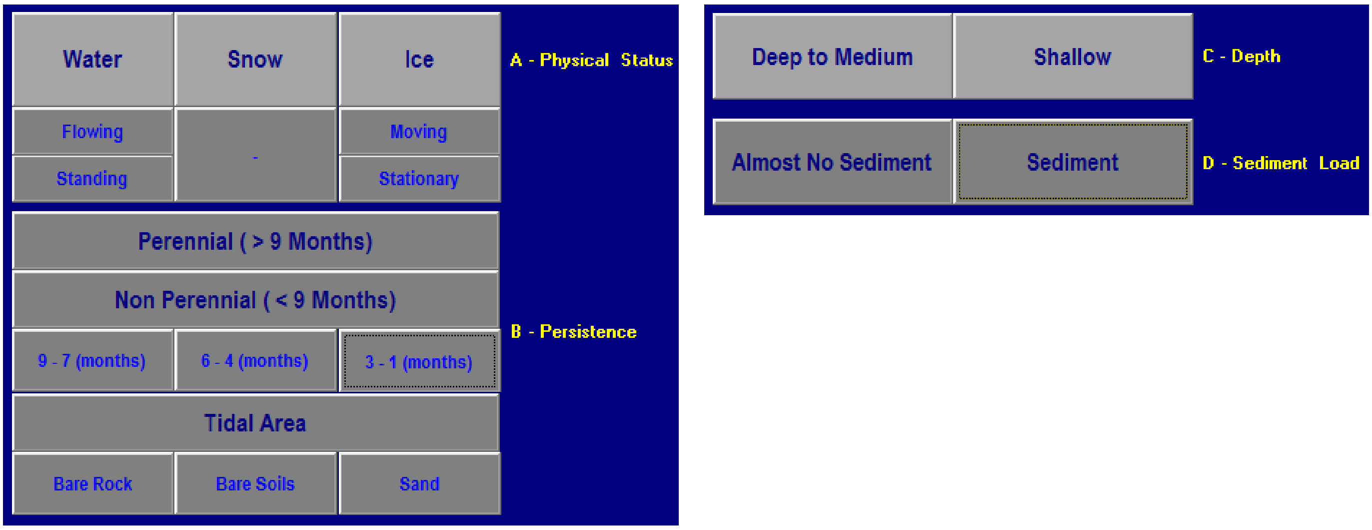


Figure 5.17. LCCS options for the description of waterbodies.

Table 5.5. Indices used for the detection of water types

Index/fraction	Formula	State
WBI(1)/WBI(2)	As in table 2.3b	Water from other surfaces
NDWI	As in table 2.3b	Snow from ice
Normalised Difference Glacier Index (NDGI)	$\frac{Green - Red}{Green + Red}$	Glacial ice
Normalised Difference Snow Index (NDSI)	$\frac{Green - SWIR}{Green + SWIR}$	Snow from other surfaces
Slope and aspect	Lack of a slope and aspect for water areas	DEM

Dynamics: Within the LCCS, water is described as flowing or standing whilst ice can be either moving or stationary. Where water is located within river channels, estuaries or the sea it is typically moving and these larger objects can generally be recognised through reference to tidal and river flow information. Even where the river is flowing under vegetation, the associated slope surfaces indicate movement of water. By contrast, most areas of standing water are surrounded by terrestrial surfaces. Hence, by merging all segments associated with water and 'other, differentiation of moving and standing water could be achieved based on partial or complete enclosure by a non-water surface respectively. Distance rules were also used to allow areas enclosed by two bridges to be mapped as water. In the

case of ice (e.g., on glaciers), movement can be determined through reference to, for example, SAR interferometry or offset tracking based on radar or optical data. From moderate to VHR data, observations of glacial features (e.g., cravasses, flow lines) can be used to establish dynamics (Table 5.6).

Table 5.6. Measures used to quantify the dynamics of aquatic surfaces

Method	Measure
Enclosure rules	Standing of flowing water
Interferometry	Glacial movement
Offset tracking	Glacial movement
Glacial features	Counts indicate movement
Enclosure rules	Standing and flowing water.

Persistence and duration: To establish whether open water is perennial (present for more than 9 months) or otherwise, time-series of EO data are required particularly as the LCCS divides the duration of non-perennial aquatic surfaces into 1-3 months, 4-6 months and 7-9 months. In each case, the extent of the aquatic surface can be established using the indices and measures described in Table 2.3 but change detection is needed to associate each with an LCCS category. As well as artificial and natural waterbodies, spatial data on the temporal dynamics of water within the landscape are needed for the description of aquatic vegetation, whether cultivated/managed (Figure 5.18a,b) or natural/semi-natural (Figure 5.18c). In these latter cases, the duration of inundation differs from within daily (e.g., for tidal areas) to more than 9 months. Hence, the mapping of water inundation using temporal EO data or through reference to hydrological models is essential.

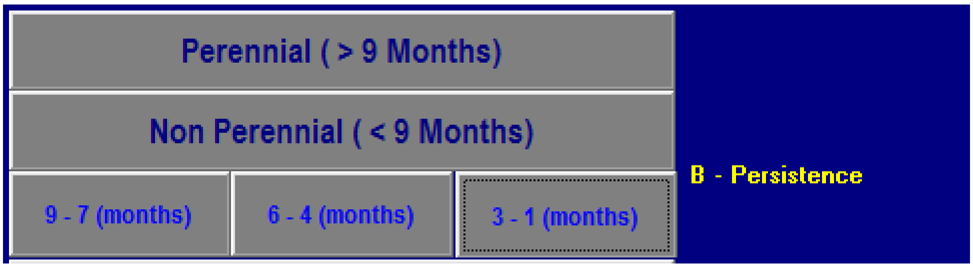


Figure 5.18a LCCS descriptors of water within the landscape (artificial and natural)

Figure 5.18b LCCS descriptors of water seasonality (for aquatic cultivated).

Figure 5.18c LCCS descriptors of water seasonality
(for aquatic natural and semi-natural environments).

Tidal area: The tidal area is typically associated with flowing water with a connection to the sea. Within the tidal area, the subsurface material can be approximated from classifications of the material at low tide.

Depth and sediment load: The depth of water is described on the basis of shallow and medium deep to deep (Figure 5.19). The depth of water can be approximated on the basis of a) coverage of water in one image but not the next allowing an approximation of depth based on known tidal ranges, b) the use of the WV-2 coastal or visible channels (all optical sensors) in combination with the NIR reflectance, which is typically lower also for deeper water or c) exploitation of the Normalied Difference Depth Index (NDDI), with this based solely on WV-2 data (Table 5.7). The LCCS identifies water with almost no sediment and differentiates from water with sediment. The presence of sediment can be detected through reference to the visible wavebands or derived indices such as the Normalised Difference Turbidity Index (NDTI; Table 5.7).

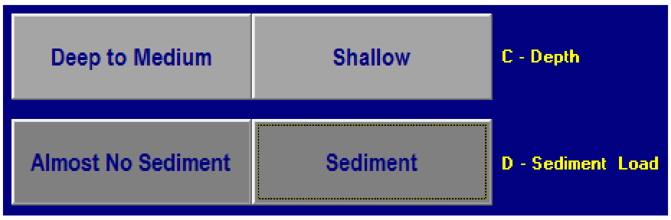


Figure 5.19 LCCS options for the description of waterbodies.

Table 5.7. Indices used for the detection of water types

Index/fraction	**Formula**	**State**
Normalised Difference Turbidity Index (NDTI)	$\frac{Red - Green}{Red + Green}$	Turbid from non-turbid water
Normalised Difference Depth Index (NDDI)	$\frac{Coastal - Green}{Coastal + Green}$	Shallow and deep water

5.2.6 Artificial surfaces

For the description of artificial surfaces, the LCCS describes these on the basis of a surface aspect (Figure 5.20). However, with EODHAM, a further layer is generated to describe the density of urban areas. The classification of artificial surfaces is important within EODHAM as, in many cases; the expansion of settlements, industry and associated infrastructure has led to the loss of biodiversity as have increases in the density of urban areas. In many regions, existing maps of urban areas exist and form a baseline against which to quantify change. However, identification of artificial surfaces from

moderate or VHR optical imagery is needed, particularly in areas where these spatial layers do not exist or where rapid changes are occurring. The area associated with these surfaces is known beforehand because of classification within LCCS Level 3.

Within the area associated with artificial surfaces (e.g., identified through appropriate segmentation and use of spectral indices such as the NDVI), non-built up urban areas are divided into waste dump deposits and extraction sites which can often be identified through changes in topography (e.g., elevation, slope, aspect and shape). Using a DEM, extraction sites (quarries) can also be identified through hydrological network analysis as these open represent areas of internal drainage (i.e., sinks). However, confusion with exposed rock and other bare surfaces can occur and the arrangement of infrastructure around the sites needs to be considered.

Within the built up area, linear features are identified by considering a range of descriptors (primarily length/width ratios; Table 5.8). Within this category, railways can be distinguished from roads as these are typically of low slope, generally have gentle changes in angle and are also often either raised above or are below the surrounding landscape. The nature of the landscape surrounding railways can be discerned from DTMs. When roads are paved, these tend to be of constant width and have a visible reflectance that typically differs from unpaved roads. Communication lines/pipelines are more difficult to distinguish because these are generally narrow although are often associated with linear clearances or reductions in height of vegetation, with these discernible within optical and CHM data.

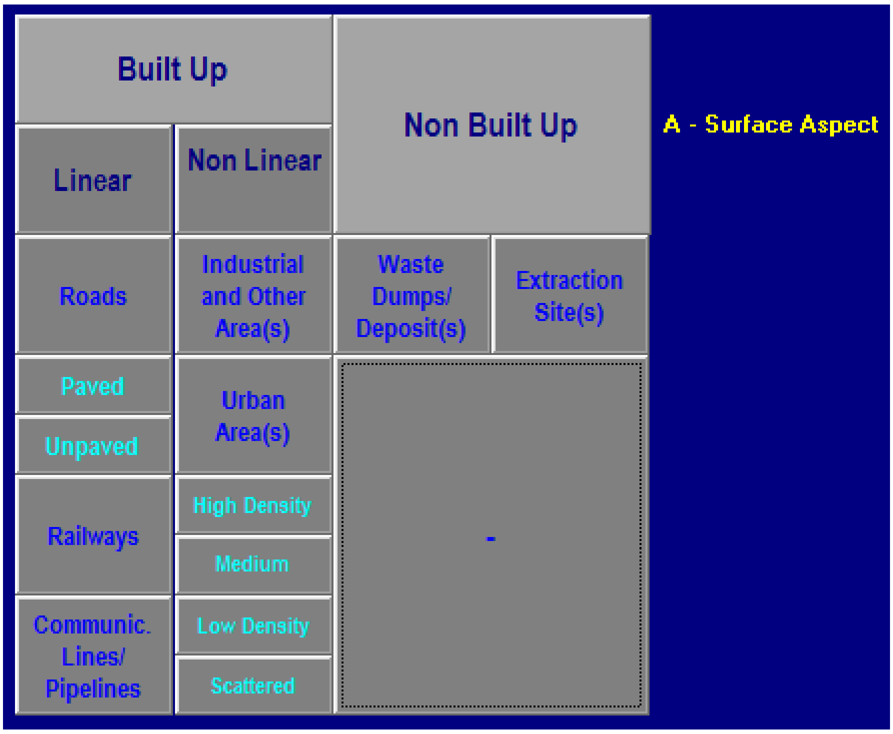


Figure 5.20. LCCS options for the description of urban areas

Industrial and other areas are typically separated from other urban areas as the buildings (e.g., warehouses, storage yards) are generally larger and more homogeneous compared to residential areas. The density of residential area can be determined by counting the number of buildings within larger objects representing the extent of urban infrastructure.

Table 5.8. Key features facilitating differentiation of LCCS artificial categories

Index/fraction	Measure
Linear	Length/width ratio
Paved/unpaved	Soil indices, PV/Bare endmember fractions
Railways/roads	Slope, angle and local linear topography of surrounding objects
Extraction sites	Sinks within DEMs
Waste dumps	Changes in topography

5.2.7 Vegetation

Vegetation can be natural and semi-natural and is found within cultivated/managed environments, whether terrestrial or aquatic. Descriptors of vegetation life forms are found in both aquatic and terrestrial vegetated categories, with some variations. For example, in addition to the woody trees and shrubs, herbaceous forbs and graminoids and lichens/mosses associated with terrestrial vegetation, rooted and free floating forbs are noted in aquatic natural/semi-natural environments (Figure 5.21a). Within cultivated landscapes, urban vegetated areas (tree dominated parks, shrubs and tree dominated parklands and grass-dominated lawns) are found (Figure 5.21b) whilst within aquatic environment, the moss/lichen lifeform is not included and woody vegetation does not distinguish between trees and shrubs (Figure 5.21c). Recognising the commonality in the descriptions of life forms and also the exceptions (additions, omissions), the EODHAM system classifies life forms before assigning these to a cultivated/managed (A11/A23) or natural/semi-natural (A12/A24) category. Once the life form has been defined, all vegetation is subsequently classified on the basis of cover, height, leaf type, phenology, distribution and stratification, as outlined in the following sections.

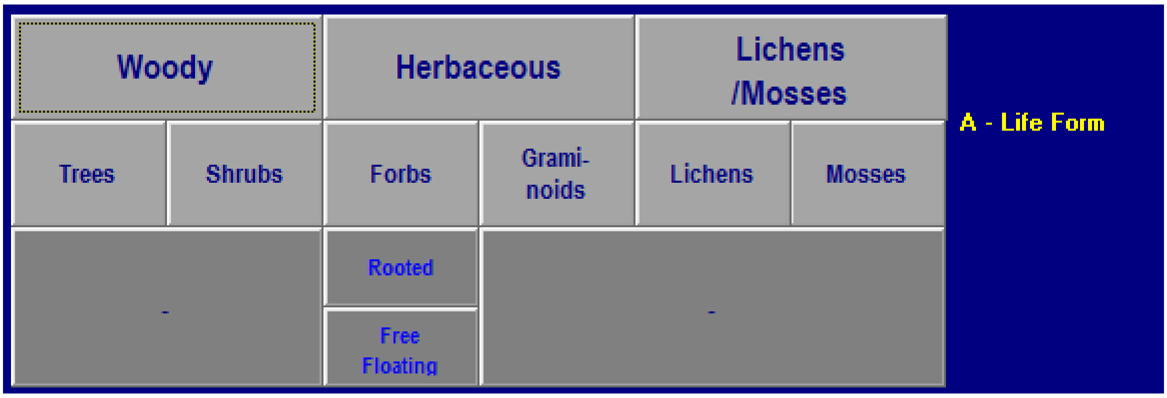


Figure 5.21a LCCS descriptors of life form for semi-natural and natural vegetation. Note that rooted and free floating forbs are differentiated only in aquatic environments.

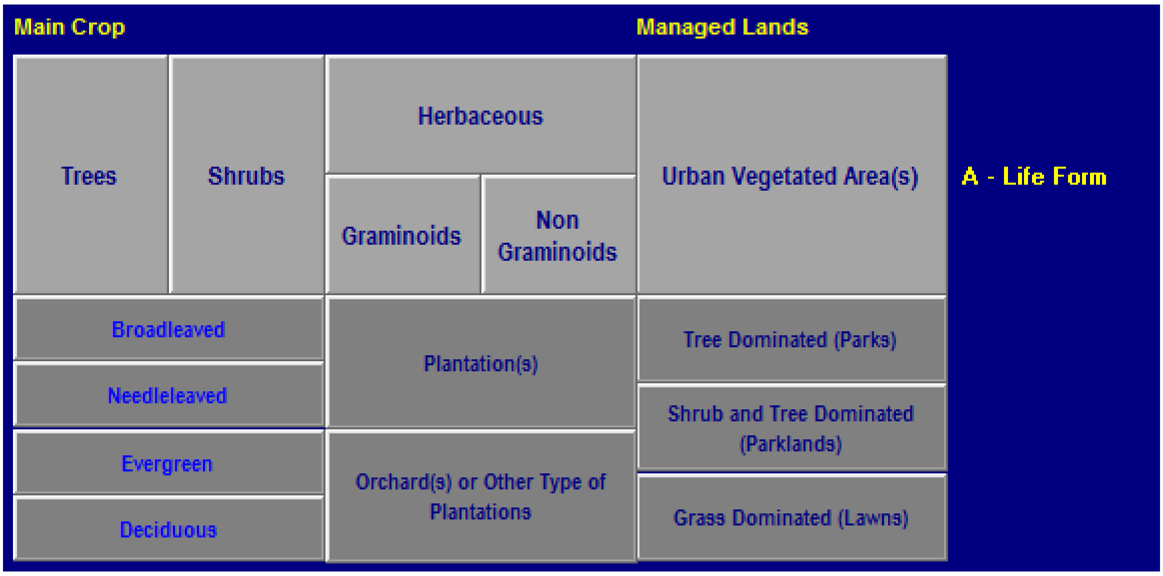


Figure 5.21b LCCS descriptors of life form for cultivated terrestrial environments.

Figure 5.21c LCCS descriptors of life form for cultivated aquatic vegetation.

Life form: When all life forms are considered, the primary categories relate to woody trees/shrubs, herbaceous vegetation (forbs and graminoids) and lichens/mosses. The optimal approach to the separation of woody from non-woody vegetation is to utilise a canopy height model (CHM), with that generated using LiDAR considered to be the most reliable, although other sources of information are available (e.g., interferometric SAR, stereo imagery; Table 5.9). Where optical data are not available, the alternative is to use surrogates for height with textural measures based on entropy (green layer) considered to be optimal. Lower frequency (e.g., L-band SAR), such as the Advanced Land Observing Satellite (ALOS) Phased Arrayed L-band SAR (PALSAR) also allow separation of woody and non-woody vegetation as a higher backscatter is observed from woody vegetation, with this attributed to the increased amount of plant material (and hence biomass). Better differentiation of woody and non-woody vegetation is achieved at HV polarisations.

Table 5.9. Methods for retrieval of height information

Index/fraction	Measure/equation
LiDAR	
Stereo imagery	
Interferometry	
Surrogates	Texture (green)
	Shade fraction
Indices	[1]$Coastal - Green$

[1]WV-2 only.

Canopy cover: The estimation of canopy cover has been achieved in a wide range of studies and is often based on relationships established between ground measurements and optical imagery. As an example, Foliage Projected Cover (FPC) is derived routinely from Landsat sensor data for the state of Queensland, Australia, based on relationships with spectral data and climate variables. However, the estimation of cover is based primarily on the amount of green vegetation whereas in many environments,

other vegetated states contribute to the total cover (e.g., non-photosynthetic and submerged vegetation). Hence, estimates of cover need to include area estimates from the full range of vegetated states. The amount of green vegetation can be defined using, for example, the VDI or the NDVI (Table 2.2). Other vegetated states can be defined using the indices described in Section 2.5.1.

An approach to the estimation of canopy cover is to sum the number of pixels associated with vegetation and express as a percentage of the total for the object (Figure 5.22). Whilst the use of total cover provides an overall estimate, estimates of the amount of vegetation in different vegetative states (i.e., photosynthetic or otherwise) are also useful to understand, for example, seasonal changes in green cover. Once canopy cover is estimated, each object identified as woody, herbaceous and lichens/mosses can be assigned to one of the LCCS categories shown in Figure 5.23. The LCCS classification assigns several vegetation cover estimates for woody, herbaceous and lichens/mosses. The same scheme can therefore be applied to objects based on the estimates of cover.

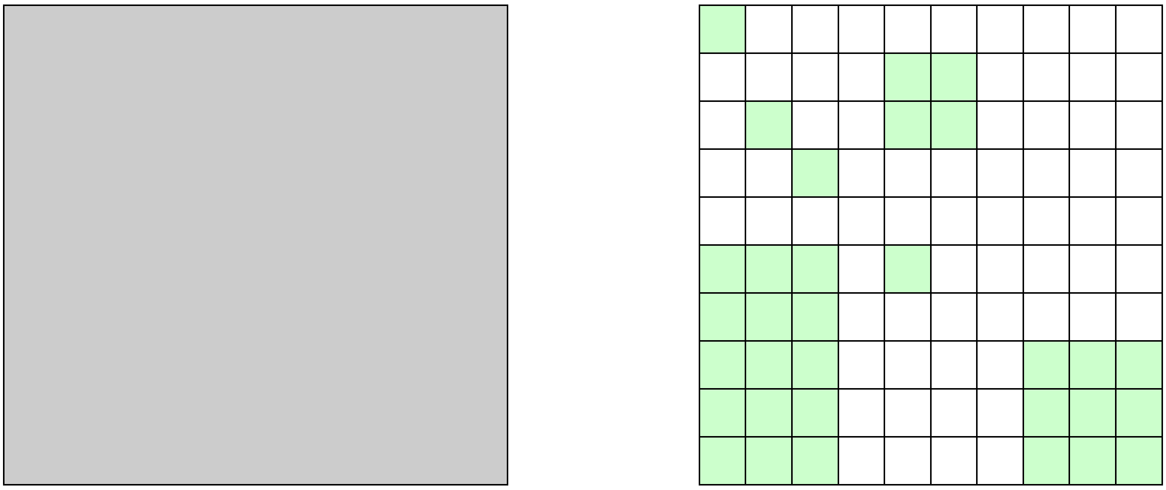


Figure 5.22. A large object consisting of 100pixels allocated to photosynthetic vegetation, non-photosynthetic and submerged vegetation. The percentage cover for each vegetated state is associated with the larger object, with this being 23 % in the case illustrated.

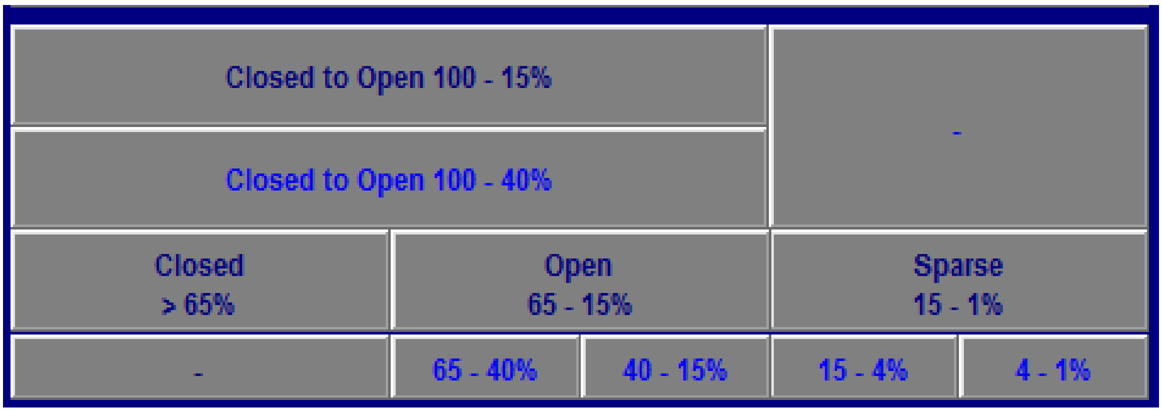


Figure 5.23 LCCS cover categories for woody and herbaceous vegetation and also lichens/mosses.

Height: The height of plant canopies, which is required for LCCS description (Figure 5.24) is best determined using a CHM generated from LiDAR, although other datasets (listed in Table 5.9) can be used. The assignment of an object to the height categories is based on simple thresholding using pre-defined values (Figure 5.24), with the reliability of classification depending upon the accuracy of the CHM. The main limitation of the LCCS classification is that those categories below approximately 0.5 m are difficult to separate as the noise in the LiDAR data can exceed this value.

> 30 - 3 m	5 - 0. 3 m		3 - 0.03 m	
> 30 - 14 m	5 - 0.5 m		3 - 0.3 m	
14 - 7 m	5 - 3 m	3 - 0.5 m	3 - 0.8 m	0.8 - 0.3 m
7 - 3 m	< 0.5 m		0.3 - 0.03 m	

Figure 5.24 LCCS height categories for woody and herbaceous vegetation and lichens/mosses.

Leaf type: The LCCS considers three leaf types; broadleaved, needle-leaved and aphyllous, with these categories applied primarily to woody vegetation. Discrimination between these leaf types is best achieved through reference to differences in the NIR reflectance and, to a lesser extent, the short wave infrared (SWIR) reflectance during periods of peak flush. These differences are primarily a function of differences in cell size and arrangement within the leaves. However, from optical imagery, the distinction is compromised by shadowing within canopies and hence taking the meanlit (or sunlit) spectra from delineated crowns, a better indication of the reflectance of vegetation can be established. The procedure is similar to that outlined in Figure 5.22, but the mean of only those values that are considered non-shaded (as determined using, for example, the WBI) is used.

Phenology: In the LCCS, woody vegetation is assigned to evergreen, deciduous, semi-deciduous (if containing > 75 % of plants within the community) or mixed and to establish this phenology, changes in the cover of photosynthetic (green) vegetation during observations acquired over pre-flush and full flush periods can be quantified. Such comparisons can be undertaken at the pixel level but are best achieved by comparing pixels and/or objects (in VHR imagery) representing delineated crowns or stands of trees. In the case of herbaceous vegetation, the LCCS assigns a perennial or annual descriptor, which again can be described through reference to layers representing green vegetation during pre- and peak-flush periods. By defining leaf type and then phenology in different layers, woody vegetation can be characterised, with examples given in Table 5.10. Similarly, herbaceous vegetation can also be described and discriminated (Table 5.11).

Table 5.10. Classes used to describe woody vegetation, with examples

	Phenology	
Leaf type	**Evergreen**	**Deciduous**
Needle-leaved	Scots Pine *(Pinus sylvestris)*	European Larch *(Larex decidua)*
Broad-leaved	Rhododendron (*Rhododendron ponticum*)	Oak (Quercus robur)
Aphyllous	Gorse (*Ulex europeas*)	

Table 5.11. Classes used to describe herbaceous vegetation

Lifeform	**Perennial**	**Annual**
Graminoids	Purple moor grass (*Molinea caerulea*)	Annual Meadow Grass (*Poa annua*)
Forbs	Bracken (*Pteridium aquilinum*)	Dandelion *(Taraxacum officinale)*

Stratification: The LCCS scheme considers the stratification of woody (primarily forests) into one or multiple layers and can only be achieved realistically using the LiDAR data, which can be analysed to retrieve the number and depth of canopy layers. Assumptions on stratification can be made based on the height of vegetation but is difficult to achieve from optical remote sensing data.

Distribution: The classification of vegetation can consider whether land cover categories are continuous or fragmented, with this differentiation requiring finalisation of the LCCS map or *a priori*.

5.3 Approaches to contextual classification

Whilst the classification of LCCS categories described is based largely on spectral bands and derived indices, contextual information allows improved and, in some cases, a more consistent classification of some LCCS categories. In this regard, elements of the landscape can be classified using textural, morphological, geometric features and both topological and non-topological spatial relationships, as outlined in the following sections.

5.3.1 Textural features

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 *et al*., 2006). First-order texture indices are not related to contextual information since such measures are derived from the histogram of pixels intensities in a moving window but ignore the spatial relationships of pixels. Second order statistics are calculated from the grey-level co-occurrence matrix, which indicates the probability that each pair of pixels values co-occur in a given direction and distance (Haralick et. al 1973, St-Luis et al. 2006). First-order entropy does not consider such probability in direction and distance. A limitation of textural measures is that these are often inconsistent when applied to optical imagery because of the variability in reflectance over time and as a function of wavelength and kernel size. Nevertheless, textural measures can facilitate discrimination of some LCCS categories (e.g., managed and unmanaged forests).

5.3.2 Morphological features

Morphological features can be used to describe, for example, bright objects over dark backgrounds or *vice versa*. Examples include olive groves (in the visible channels, dark trees over bright soil) or windfarms (bright wind turbines over vegetation). Delineation of objects (e.g., tree crowns) in the foreground is required for characterisation, with the background assuming the characteristics of the area remaining. In terms of the background, consideration needs to be given to the variability primarily in the reflectance of soils, rock surfaces and photosynthetic and non-photosynthetic vegetation.

5.3.3 Geometric features

A number of geometric features can be used including area, length/width, perimeter length, angle, compactness, rectangularity, elongatedness and straightness of boundaries. Within the LCCS, examples including the definition of field sizes (small, large and medium based on area in ha), differentiation of roads from buildings (based on length/width ratios), ponds (compactness), hedgerows and ditches (straightness of boundaries) and railway tracks (curvature and straightness).

5.3.4 Spatial non-topological features

Non-topological relationships are based on qualitative and cardinal measures (relative differences), including distance (e.g., close, far) and direction (e.g., north, southwest). Examples include aquatic vegetation occurring in proximity to water bodies or agricultural terraces where relative differences in elevation can assist their discrimination.

5.3.5 Spatial topological relationships

Spatial topological relationships describe relative position such as below, above, front-of and behind and adjacency. As an example, gardens occur in front of or behind buildings, rivers occur below bridges or bridges occur above rivers. Another example is the case of flowing and standing water, where the latter is typically surrounded by terrestrial/aquatic vegetated or non-vegetated land covers.

5.3.6 Temporal relationships

Temporal relationships are important to the discrimination of a number of LCCS categories, particularly those associated with water inundation, agricultural crop types and productivities and phenology.

5.3.7 Overview

The use of contextual information is undertaken in the second stage of the EODHaM system and is particularly useful for the discrimination of artificial surfaces (e.g., buildings, roads, railways, reservoirs).

A key requirement, however, is appropriate and reliable segmentation of the landscape, as discussed in Section 5.1.1.

5.4 Classification of LCCS categories beyond Level 3

Following generation of the information within each of the layers, a subsequent process was established whereby each LCCS category was generated (at Level 0) through reference to the classification in each of the sub-layers (where occurring). To achieve the classification, Level 1-3 categories not relevant to the subsequent classifications in the sub-layer were removed (see Figure 5.9). For example, if water areas were to be classified, all vegetated and non-vegetated bare or artificial classes were removed. The remaining area of water was then classified further by referencing the appropriate class in the layers physical status, dynamics, persistence, duration, tidal, material, depth and sediment load. Examples of application are provided in subsequent sections.

5.5 EODHaM 2nd Stage (LCCS Level 3): Implementation

The EODHaM 2nd Stage classification to Level 3 focuses on separating natural and semi-natural surfaces from those that have been altered by humans through agriculture (cultivation), forestry (management) and construction of artificial surfaces (e.g., buildings, roads, water reservoirs and artificial channels). Whilst this stage can bring in existing thematic layers (e.g., field boundaries, roads and buildings), the generation of these from the image data themselves is desirable and, in many cases, essential. The following sections outline how the EODHaM 2nd Stage classification to Level 3 has been implemented for sites in Italy, Wales and the Netherlands.

5.5.1 Italy (Le Cesine)

For the Le Cesine site, regularly updated thematic layers representing cadastral boundaries and urban infrastructure are not available as well as a detailed DEM. Whilst some field data have been collected through a range of projects, these data have not been always been collated centrally. Le Cesine therefore provides an example of a Natura 2000 site where supportive information for classification is not available. The features used are presented in Table 5.12

Artificial: Most of the artificial surfaces are associated with urban areas as well as reservoirs and dykes. Buildings are typically small and are sparsely distributed throughout the site whilst artificial waterbodies are associated with the golf course to the east of the main reserve. To differentiate built up areas, the Brightness Index was used as values were typically higher than barren land. However, additional geometric indices (e.g., length, rectangularity, number of vertices) also facilitated detection. Artificial water bodies were detected using the geometric Border Index, which measures the jaggedness of the border. The index is used following merging of adjacent segments labelled as aquatic such that the border of the combined area was considered

Cultivated areas: The agricultural areas surrounding the Natura 2000 site is comprised mainly of tree crops, with olive trees dominating. The coniferous tree plantations are also considered to be managed. The intermediate layers of *Bare Soil*, *Sand* and *Sea* were used to implement adjacency rules as tree crops can be identified through context-sensitive features mainly based on spatial topological and non-topological and geometric features (i.e. roundness) and (first-order) texture measurements. In particular, the condition of adjacency of the tree segments to soil, the roundness condition of the tree crowns and topological distances between trees were used for detecting this category. The two intermediate strata corresponding to tree crops and managed coniferous were merged into the LCCS level 3 *Cultivated Terrestrial* category.

The DFD for EODHaM 2nd stage, which provides as output the LCCS classes within Level 3 is conveyed diagrammatically in Figure 5.25.a to evidence the use of class specific contextual features. The same approach was also adopted for the other sites. The DFD emphasises the focus of the EODHaM 2nd stage on context-sensitive features. LCCS categories at Level 3 are shown in Figure 5.26.

Table 5.12 . Features used for the detection of LCCS level 3 categories.
$\rho_x = \text{TOA Reflectance at spectral band centred on wavelength (x)}$

Symbols in DFD	Index	Acronym	Formula	Application
I_1	$\frac{1}{\text{Water Band Index}}$	$\frac{1}{\text{WBI}(WV2)}$	$\frac{\rho MS_7}{\rho MS_2}$	Non-submerged aquatic vegetation and open water discrimination
I_2	Greenness Index	GI(WV2)	$\frac{\rho MS_5}{\rho MS_2}$	Vegetation discrimination
I_3	$\frac{1}{\text{Water Band Index}}$	$\frac{1}{\text{WBI}(QB)}$	$\frac{\rho QB\_MS_4}{\rho QB\_MS_1}$	Non-submerged aquatic vegetation and open water
I_4	Greenness Index	GI(QB)	$\frac{\rho QB\_MS_3}{\rho QB\_MS_1}$	Vegetation detection
I_5	Normalised Difference Depth Index	NDDI(WV2)	$\frac{(\rho MS_1 - \rho MS_3)}{(\rho MS_1 + \rho MS_3)}$	Deep and shallow water discrimination
I_6	Brightness	BR	$\frac{\sum \text{All spectral bands}}{\text{Total number of spectral bands}}$	Bright Soils detection
I_7	Water Index	WI(QB)	$\frac{\rho QB\_MS_4}{\rho QB\_MS_3}$	Water detection
T_1	Entropy	Texture Occurrence (WV2)	Green band with a 3x3 window	Grassland and trees discrimination
T_2	Entropy	Texture Occurrence(QB)	Green band with a 3x3 window	Grassland and trees discrimination
G_1	Border Index	BI	$\frac{b_v}{2(l_v+w_v)}$ b_v=image object border length l_v=length of an image object v w_v=width of an image object v "Ratio between the border length of the image object and the smallest enclosing rectangle (v)"	How jagged an image object is
G_2	Roundness	R	ϵ_v^{Max} - ϵ_v^{Min} ϵ_v^{Max}=radius of smallest enclosing ellipse ϵ_v^{Min}=radius of largest enclosed ellipse "Difference of the enclosing ellipse and the enclosed ellipse"	How similar an object is to an ellipse
S_1	Adiacency	Ad	$\boxed{\frac{\sum_{u \in N_v(d,m)} b(v,u)}{b_v}}$ b(v,u) : topological relation border length N_v(d) : neighbors to an image object v at a distance d b_v: image object border length "Ratio of the shared border length of an image object with a neighboring image object assigned to a defined class to the total border length"	Adiacency to an image object

S_2	Distance to	Dist	$\boxed{\min_{u \in V_i(m)} d(v,u)}$ d(v,u) : distance between v and u V_i (m) : image object level of a class m "The distance (in pixels) of the image object's center concerned to the closest image object's center assigned to a defined class"	Distance to an image object

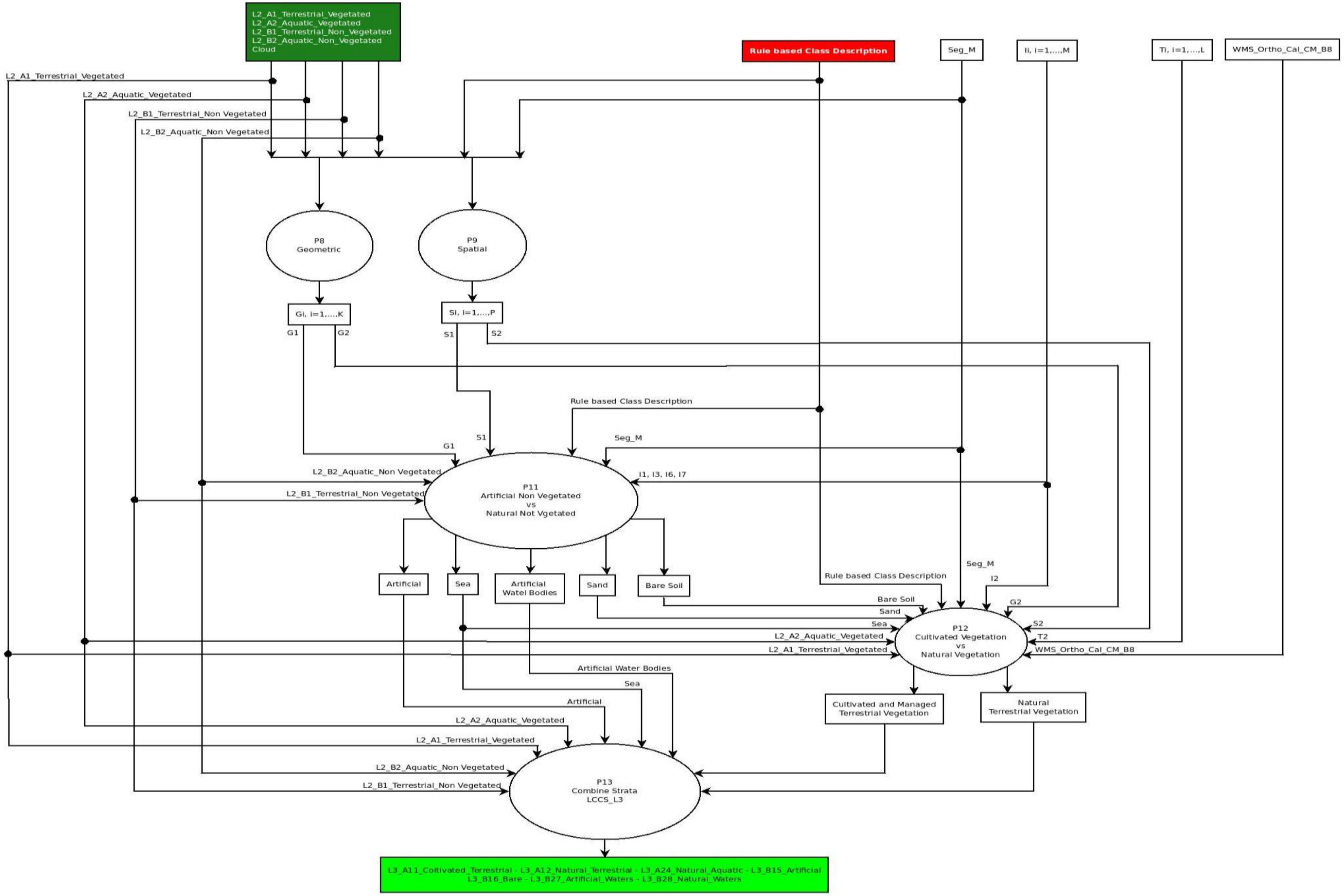


Figure 5. 25 EODHaM 2nd stage. Data Flow Diagram (DFD) outlining the approach to the classification of LCCS 3. An instantiation of LCCS Level 2 to LCCS Level 3 classification for the Le Cesine site. The legend is provided in Tables 2.1. and Table 5.12

(a)

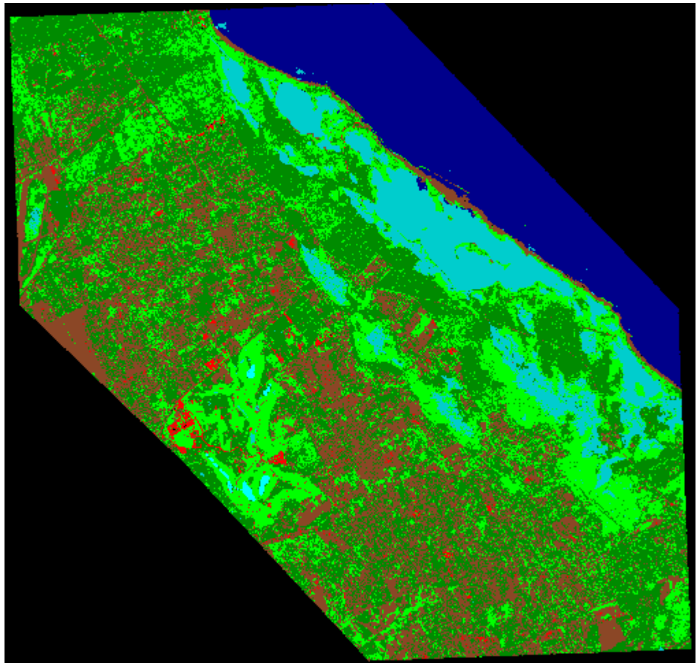


Figure 5.26. Classification of a) the Le Cesine site to Level 3 and a subset of b) the Worldview2 October image, c) the classified map and d) the Quickbird image

L3_A11_CULTIVATED_TERRESTRIAL

L3_A12_NATURAL_TERRESTRIAL_VEGETATED

L3_A23_CULTIVATED_AQUATIC (Not present)

L3_A24_NATURAL_AQUATIC

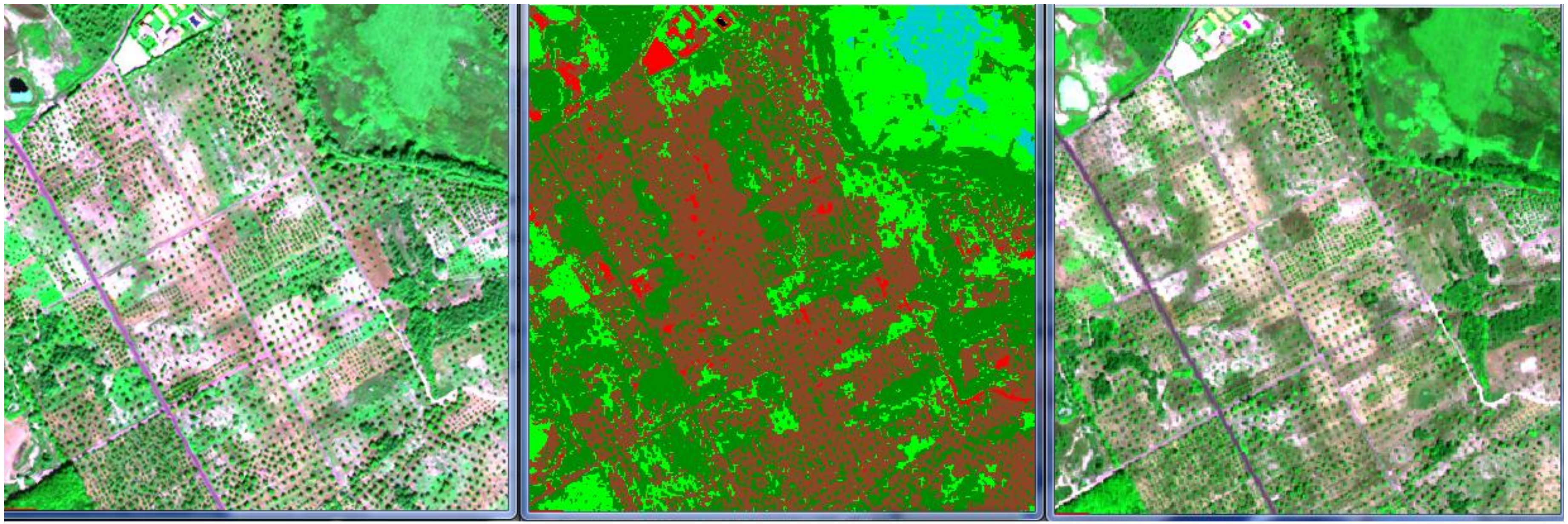
L3_B16_BARE

L3_B15_ARTIFICIAL

L3_B27_ARTIFICIAL_WATER

L3_B28_NATURAL_WATER

b) c) d)



5.5.2 Wales (Cors Fochno).

For Cors Fochno, the classification was based primarily on the use of WV-2 data acquired in July and November, 2011.

Artificial surfaces: These surfaces are typically associated with urban areas as well as waterbodies, including reservoirs and dykes. In many cases, these can be identified by the diversity of infrastructure surrounding the water. All remaining surfaces are assigned to a ***natural*** category.

Cultivated areas: For the agricultural areas surrounding the Natura 2000 site, vectors representing field boundaries are generated routinely, although this did not include managed tree plantations (primarily) coniferous). However, cultivated areas could be identified through feature extraction. All remaining areas were assigned to ***natural or semi-natural vegetation***.

The Level 3 classification for Cors Fochno was achieved by combining the classes from each level (Figure 5.27) and reflects the broad distribution of land covers in the region.

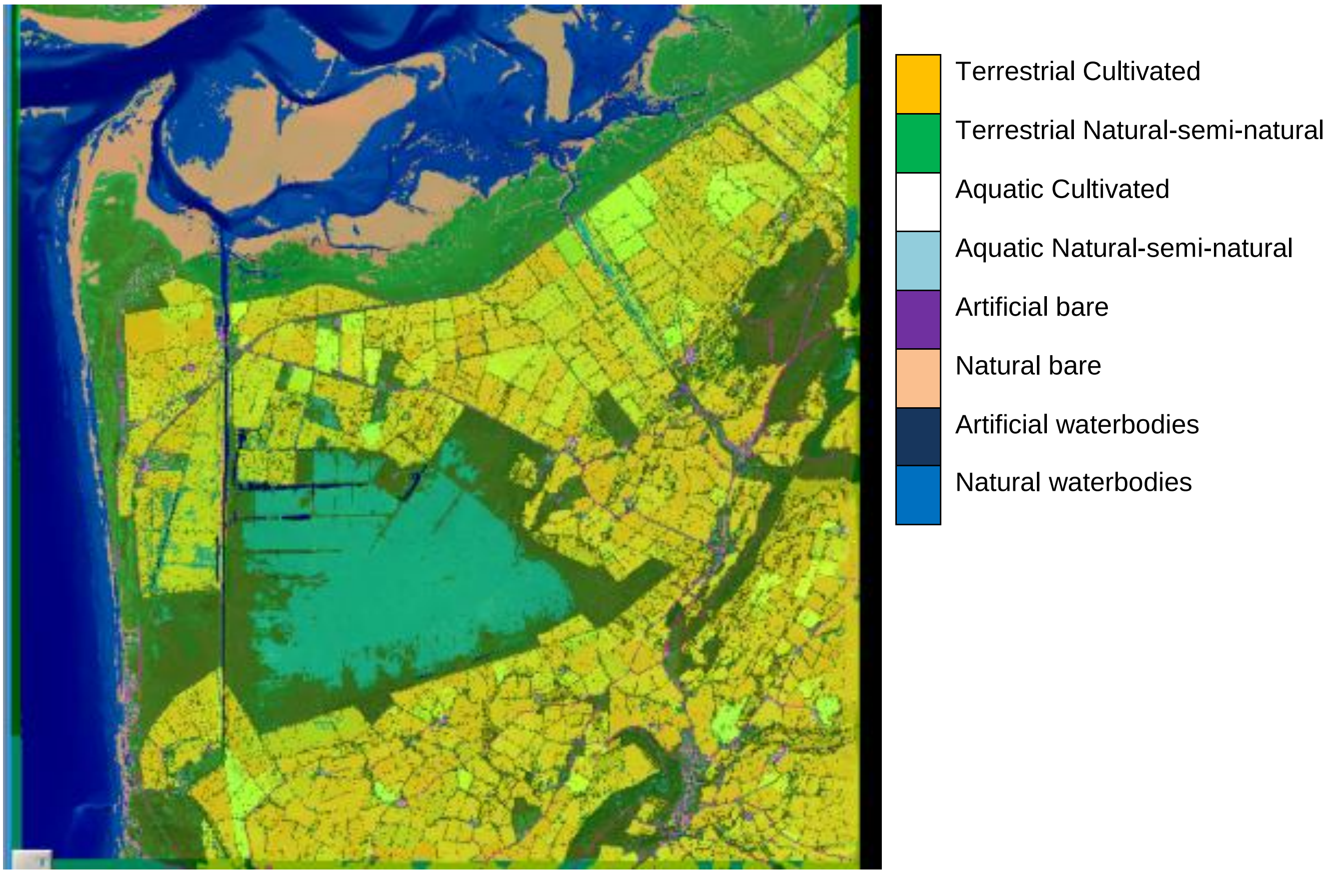


Figure 5.27. Classification of the Cors Fochno site to Level 3

5.5.3 The Netherlands

For the Dutch site, the thematic layers such as the topographic maps 1:10.000 are regularly updated, around once every four years. The Dutch LPIS information system focused completely on agricultural areas and therefore there is a preference to use the Top10-vector as a basis, as this has more complete thematic coverage. The Dutch DEM is regularly updated as new airborne LiDAR campaigns are completed.

Artificial: Most of the artificial surfaces are associated with urban areas. The largest urban area that falls partly within the area is Ede with approximately 120.000 inhabitants. Within the study area there are also some small villages as Wekerom, and are characterized by scattered buildings. The classification utilised existing data layers but approaches to direct retrieval from the imagery are being developed.

Cultivated areas: The agricultural areas surrounding the Natura 2000 site is comprised mainly of pastures and arable land with crops such as maize, wheat, potatoes and sugar beets. The agriculture in the surrounding area is very intensive with high amounts of fertilizers. Together with the intensive pig and dairy farming are placing significant pressures on the Natura 2000 site because of nitrogen deposition. Areas of cultivation were identified primarily with reference to existing cadastral layers, although approaches to automated mapping are being developed.

For the Netherlands, the classification to LCCS Level 3 (Figure 5.28) corresponded to the known patterns within the landscape. Confusion was evident between cloud shadow and natural water and also bare areas and cloud although cloud-screening algorithms have been implemented to remove affected areas prior to segmentation.

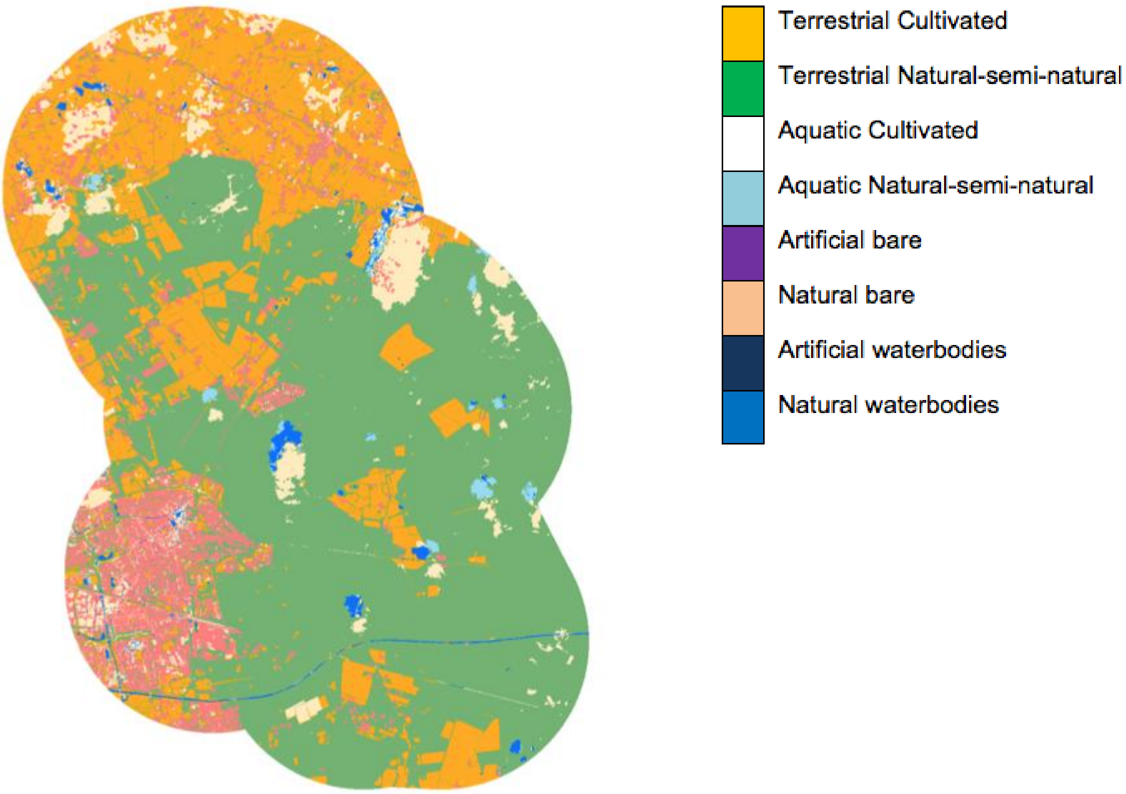


Figure 5.28. Classification of the Veluwe site, the Netherlands, to LCCS Level 3.

5.5.4 Indian sites

For the Indian site, the classification of the GeoEye-1 image to LCCS Levels 1-3 has not been explored using EODHaM, but the approach has been evaluated for feasibility as discussed below. The classification to LCCS Levels 1-3 involves separation of semi-natural and modified terrestrial and aquatic vegetation, and semi-natural and artificial bare and aquatic surfaces.

Vegetation: Diverse types of semi-natural vegetation occur, of which the main types are shola-grassland forests (a complex of montane forests and natural grasslands that is specific to this region), evergreen, semi-evergreen, moist deciduous and dry deciduous/scrub forests. Modified vegetation consists of plantations and cultivated areas. Plantations are largely of *Eucalyptus*, coffee (mostly *Coffea robusta*), and silver oak (*Grevillea robusta)* - however, coffee is mostly shade-grown, below a canopy of silver oak and hence difficult to separate from plantations of silver oak. As the patches of silver oak plantations (devoid of coffee) are very few in number, we will not attempt this separation. Cultivation is largely of ragi (*Eleusine coracana*), a local millet, with some fields in cultivation at the time of the image, and other fields fallow at the time of survey.

Natural/Semi-Natural Vegetation: It is relatively easy to differentiate vegetated from non-vegetated surfaces in IN-1. It should also be possible to use a rule-based classification approach to discriminate between semi-nnatural vegetation types based on canopy cover, and the use of multi-season imagery to discriminate between different phenologies, as can be seen in Figure 5.29.

Modified Vegetation: Modified vegetation categories include plantations of *Eucalyptus*, coffee, and fallow and seasonally planted fields of Ragi. Figure 5.29 indicates that it should be possible to differentiate plantations of *Eucalyptus* from other vegetated areas, based on spectral differences as well as variations in texture. Spectrally differentiating coffee plantations from other locations will be challenging because of their shaded canopy of silver oak, although texture could perhaps be exploited for this purpose, as shown in Figure 5.26. Data from previous years on the location of coffee plantations is available and can be used to help in identifying many plantations, although this will be insufficient in locations where the areas of coffee have expanded, or where new plantations have been initiated in areas where they were previously not grown. An additional challenge is that coffee is planted in all types of forest areas, making it difficult to define simple rules to separate such patches from their surroundings. Distance to roads can be used to assist in classification – in addition, there may be temperature differences in areas of shade coffee vs semi-natural forest, and the use of thermal bands can be explored to help differentiate this. Fallow and ragi fields can be easily differentiated spectrally especially if multi-season data are used, and fallow fields are quite different spectrally and in terms of the regularity of their spatial boundaries from natural grasslands in the shola-grassland complexes. In addition, these can also be easily separated by virtue of the fact that shola-grasslands occur at high elevations.

There are very few areas of aquatic vegetation, mostly with the growth of lotus and other aquatic plants on fresh water lakes. It should be possible to differentiate these using context based rules (surrounded by water), as can be seen from the Figure 5.30.

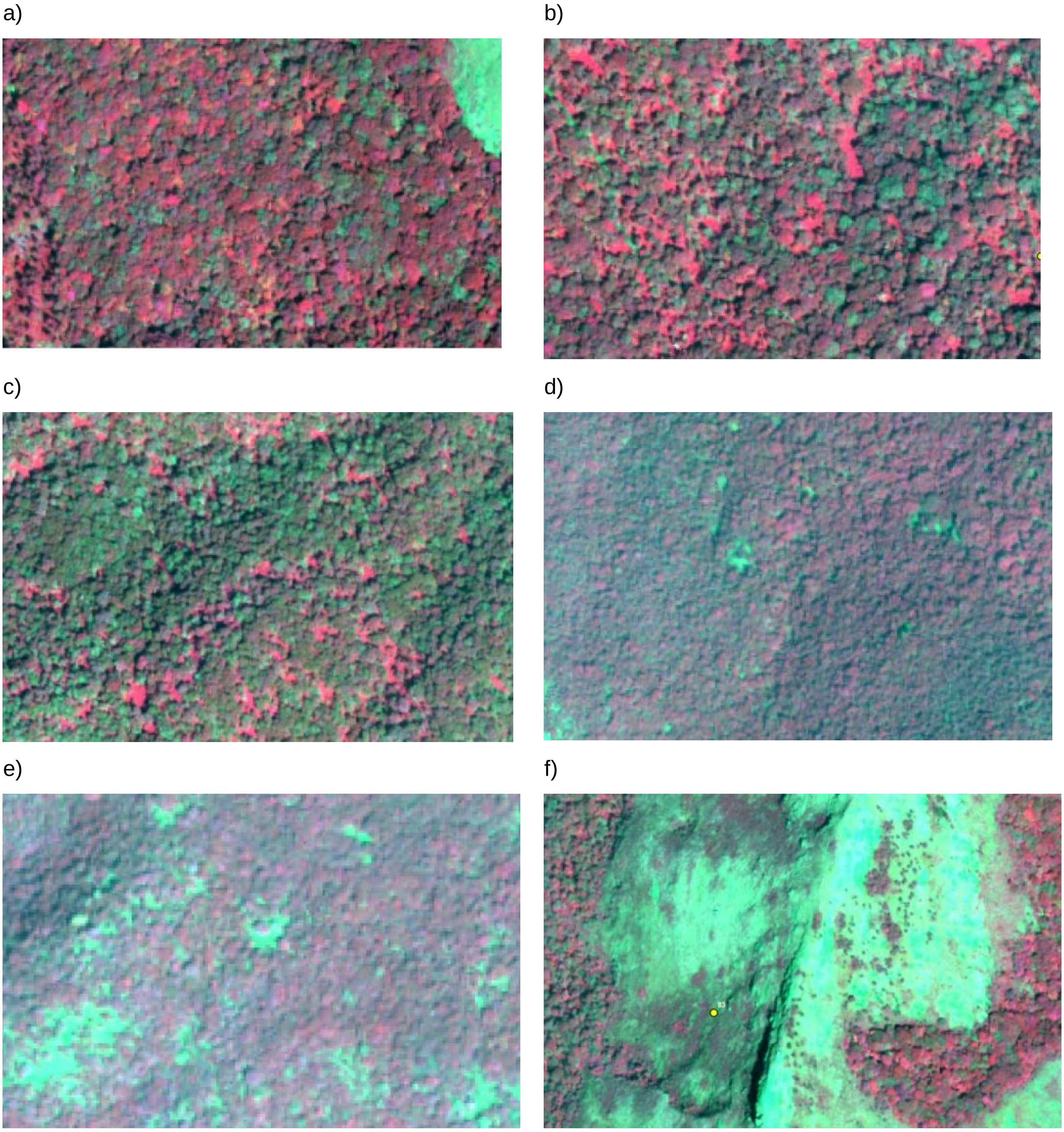


Figure 5.29. Appearance of a range of forest types within Geoeye-1 imagery acquired in the pre-monsoon period. a) Evergreen forest, b) semi-evergreen forest, c) moist deciduous forest, d) dry deciduous forest, e) scrub/dry deciduous forest and f) shola-grassland complex.

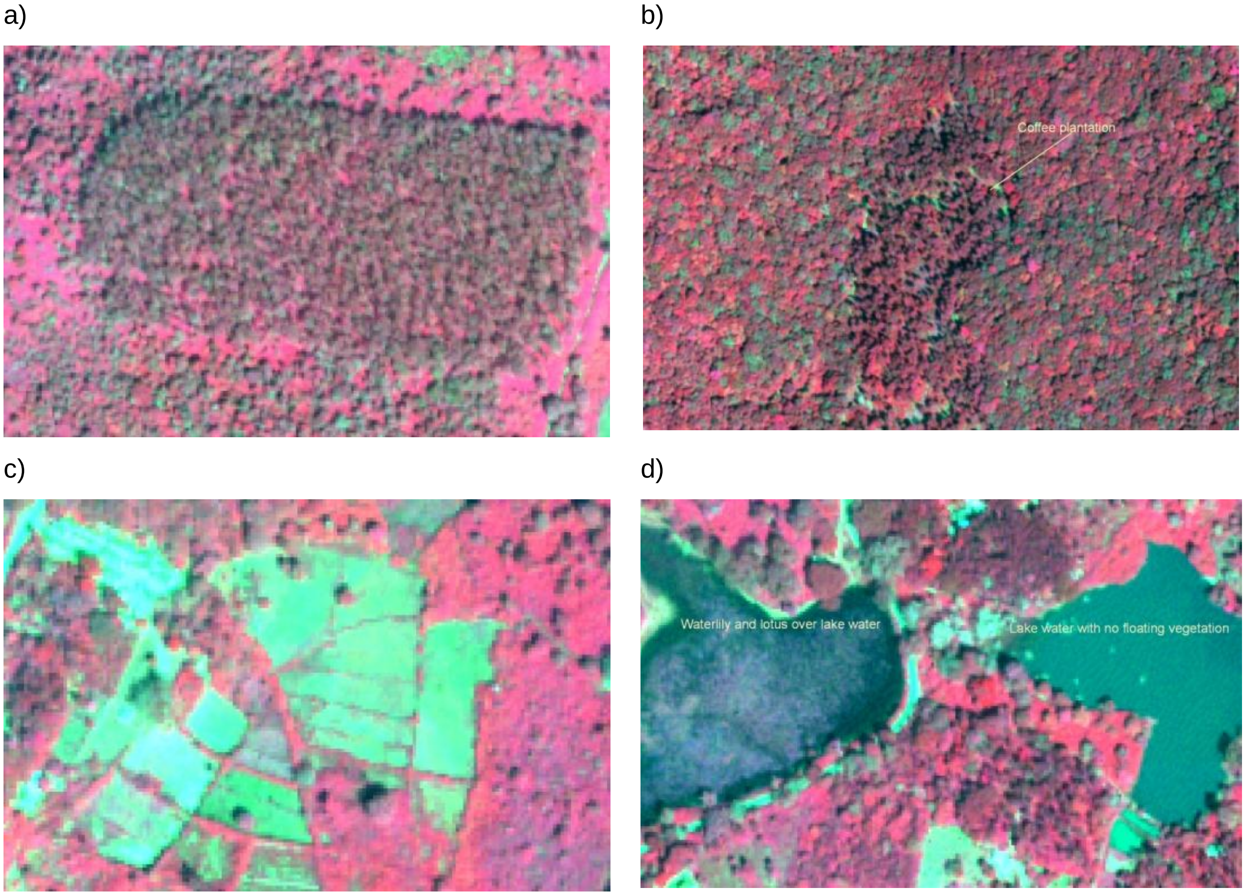


Figure 5.30. Appearance of a range of culrivated/managed and aquatic vegetation within Geoeye-1 imagery acquired in the pre-monsoon period. a) Eucalpytus plantation, b) Coffee plantation, c) Fallow fields and d) Aquatic vegetation on lake surface (left).

Non-vegetated:

Natural: Natural non-vegetated areas are of two major types – bare exposed rock (terrestrial) and areas of open water (aquatic). As Figure 5.30 indicates, bare exposed rock should be easy to differentiate based on spectral differences as well as on texture and shape. Aquatic areas are of two types –fresh water streams and lakes. Streams are often covered with riparian vegetation and may be difficult to identify spectrally, but indicators of shape (long and winding) can be used to classify these areas. In addition, stream feature layers are available from old maps and should be able to predict the current coverage of many streams, although some (especially in mountain areas) may have changed course, or dried up in recent years. Fresh water lakes are semi-natural, with water collected in depressions but in some areas strengthened and maintained by bunds, with road networks around the boundary of the lake in some instances. These can be identified spectrally as well as based on information collected from old maps.

Artificial

Artificial non-vegetated areas are of two major types in this location – roads and settlements. Roads may be difficult to discriminate spectrally, as they are narrow in width and covered by trees (Figure 5.31), but they could be discriminated based on shape. It may be challenging to clearly separate streams from roads (Figure 5.32), and this problem will require further attention. Built settlements are largely found in a single location with a temple, resorts and guesthouses, and small houses – this area is easy to discriminate. In addition, there are clusters of huts scattered in different locations, but these are thatched and surrounded by vegetation, and difficult to identify due to their size and nature of building material.

a) b)

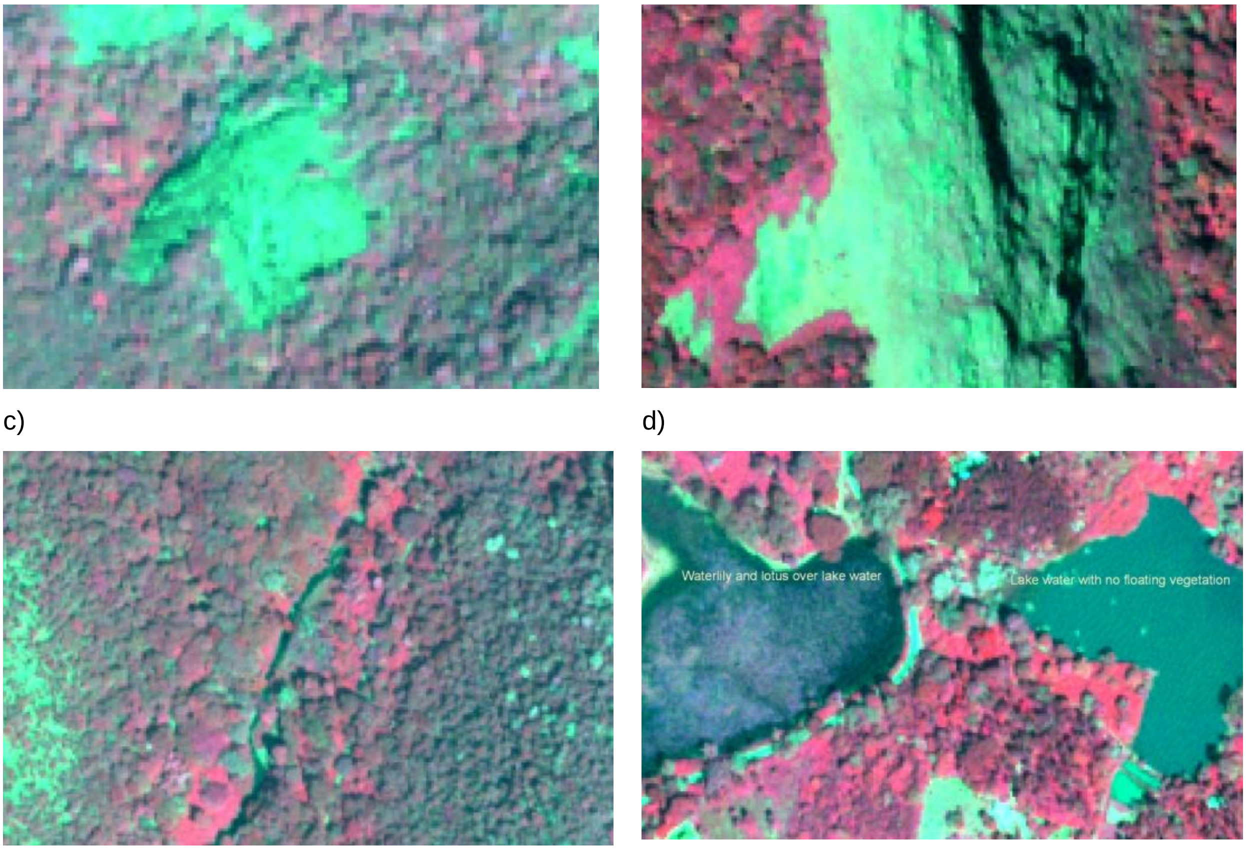


Figure 5.31. Appearance of a) and b) bare exposed rock, c) a stream covered with riparian vegetation and d) a lake with vegetation on the right.

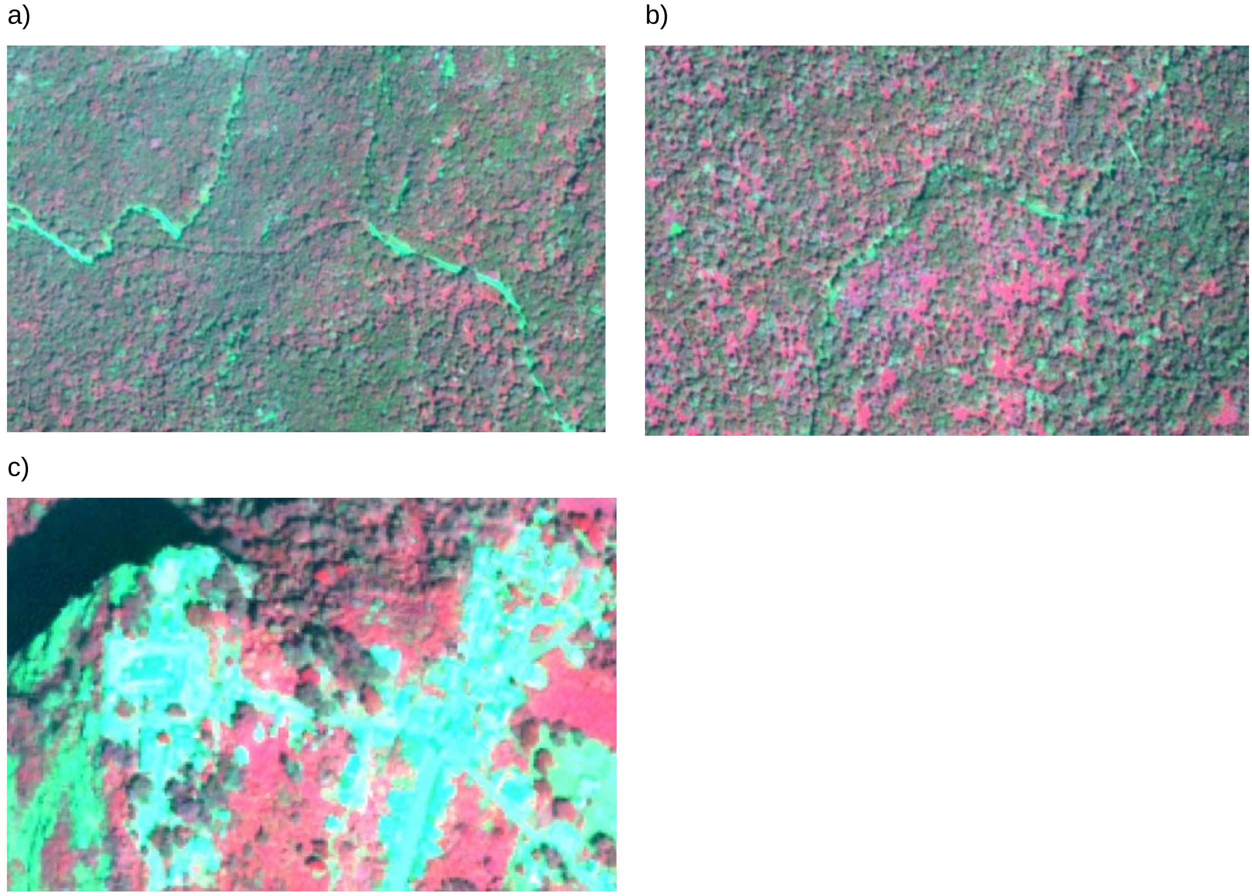


Figure 5.32. The appearance of a) and b) a road through forest, partially obscured by the tree canopy and c) a temple and settlements.

In conclusion, it appears that the approaches developed for BIO_SOS using EODHAM as described in D5.3 and applied to sites in Italy, Netherlands and Wales are conceptually capable of application to India. The greater heterogeneity in land cover and land use, as well as lack of supporting ancillary datasets and of EO data such as LiDAR make it much more challenging to develop approaches for automated classification in this context, however. In addition, preliminary explorations suggest that it may be difficult to develop one-to-one relationships between LCCS classes and GHCs in this area, and some land cover categories (such as shola-grassland complexes) may be especially difficult to differentiate in the LCCS scheme as well. The India site thus offers an interesting case to explore the transferability of the EODHAM approaches to the tropical context, and will require additional investigation

5.5.5 Summary

For the study areas, VHR optical data have been acquired with these providing significant opportunities for land cover change and habitat mapping. The WV-2 data are considered optimal, largely because the 8 spectral channels include those in wavelength regions that benefit discrimination of vegetation communities and non-vegetated surfaces. The inclusion of panchromatic data also allows finer segmentation of spectral features.

5.6 EODHaM 2nd Stage (Beyond Level 3).

The classification of land covers beyond Level 3 in the LCCS has been applied to Cors Fochno and surrounds in Wales and is currently being implemented for other sites. The following sections provide an overview of the approach, which focused on the classification of cultivated and managed areas, semi-natural/natural environments, artificial surfaces, bare ground and natural waterbodies. Cultivated aquatic systems do not occur within the area. Pools excavated in the area at the margins of the active raised bog and straightened rivers sections were considered to be natural because the former were designed to support conservation and restoration of the active raised bog whilst the latter had existed for over 100 years. A list of the LCCS categories for Cors Fochno is provided in Deliverable 5.2 and the layers used to generate these are also outlined below.

5.6.1 Cultivated and managed areas

The majority of the agricultural landscape around Cors Fochno is occupied by improved or semi-improved grasslands enclosed by hedgerows or ditches. Many forests are either needle-leaved or broad-leaved plantations, with the former including both evergreen (e.g., *Pinus sylvestris*) and deciduous species (e.g., *Larix europeas*).

In the EODHAM Scheme, layers relating to crop lifeform, field size, distribution, cultural practices and crop combinations were generated. Within the crop lifeform, woody vegetation was first identified using either LiDAR data or the coastal-green reflectance index (Table 5.9). All remaining areas were assumed to be herbaceous, with photosynthetic grasslands identified as the near infrared reflectance above a threshold in both months. The PSRI was also used to capture graminoid crops, which had been cut and were hence largely non-photosynthetic (e.g., because of haying off during the drier summer months). The phenology of the woody vegetation was determined through reference to varying amounts of photosynthetic vegetation during the July and November acquisition periods.

In Wales, field size was determined initially through reference to the size of agricultural units, as determined from updated thematic cadastral layers but subsequently to the distribution of objects classified as such (see section of segmentation). Time-series datasets were insufficient to establish whether a single or multiple-crop occurred. However, the majority of the fields are permanent grasslands and hence this category was simply assigned. In the future, multi-temporal datasets (including SAR) might be used to detect cultivation of single or multiple crops. All systems are rainfed and are permanent. Tree crops are entirely plantations and hence were assigned to this category. Nevertheless, options are available for identifying individual trees and describing their distribution (e.g., spacing) and characteristics (e.g., crown size), thereby enabling assignment to an orchard.

Table 5.13. Layers referred to in the classification of cultivated/managed terrestrial vegetation. Note that the symbols (A, B, C and D) are used to describe the land covers occurring and assign LCCS codes (see Table 5.10).

Code	A	A	A	A	A	A	A	B	B	B	B	C	C	D	D	W
	Trees	Herbaceous	Graminoids	Broad-leaved	Needle-leaved	Evergreen	Deciduous	Large to medium	Large	Medium	Continuous	Single crop	Multiple crop	Rainfed	Permanent	Plantation
Number	1	3	4	7	8*	9*	10	1	3*	4*	5	1	2	1	9*	7*
Crop Lifeform	BCD	A	A	D	CB	B	CD									BD
Field size								CA		BCD						
Distribution											ABCD					
Cultural												ABCD				
Crop combination														ABCD	ABCD	

*Modifier

The LCCS categories identified in Deliverable 5.2 are listed in Table 5.14, with these derived through reference to the layers identified in Table 5.14. In Table 5.14 and those following, the symbols (i.e, A, B, C and D etc. below the grey bar) are used to describe the land covers occurring and assign LCCS codes. For example, in Table 5.13, the LCCS code (A) would consider A3, A4, B1, B5, C1, D1 and D9 with the letter and number derived from the upper and lower grey bars respectively. The asterisk (*) indicates a modifier. For LCCS code (B), A1, A8, A9, B4, B5, C1, D1m, D9 and W7 would be considered and so on. These codes are then combined to produce the final LCCS class. The equivalent GHC category is given in Table 5.14 and similarly for all remaining tables that describe the LCCS categories.

Table 5.14. LCCS Categories defined for Cors Fochno and surrounds (cultivated).

?	Cat.	LCCS Code Modifer	Description	GHC
A	A11	A3.A4.B1.B5.C1.D1.D9_B4	Permanently cropped area: Graminoid crops	CUL/CRO or URB/GRA
B	A11	A1.B1.B5.C1.D1.D9-W7_A8.A9.B4.	Permanently cropped area with rainfed needle-leaved coniferous tree crops (plantations).	CUL/WOC or URB/TRE or TRS/TPH/(EVR, DEC)/CON or TRS/FPH/(EVR,DEC)/CON
C	A11	A1.B1.B5.C1.D1.D9-W7_A8.A10.B4.	Permanently cropped area with rainfed needle-leaved deciduous tree crops (plantations).	CUL/WOC or URB/TRE or TRS/TPH/(EVR, DEC)/CON or TRS/FPH/(EVR,DEC)/CON
D	A11	A1.B1.B5.C1.D1.D9_A7.A10.B4	Permanently cropped area with rainfed broad-leaved deciduous tree crops (plantations).	CUL/WOC or URB/TRE or TRS/TPH/DEC or TRS/FPH/DEC

5.6.2 Natural and semi-natural terrestrial vegetation

A diverse range of semi-natural and natural vegetation occurs, ranging from broadleaved deciduous forests and shrublands to grasslands and forbs, including bracken (*Pteridium aquilinum*). To classify these according to the LCCS, layers relating to life form, cover, height, leaf type and phenology as well as spatial distribution were generated (Table 5.14). As with cultivated and managed landscapes, woody vegetation was identified first, all remaining areas were assigned to a herbaceous category; graminoids were then defined using the NIR reflectance and PSRI. All remaining herbaceous vegetation was assigned to forbs. Cover was defined using the peak-flush imagery (July) on the basis of the VDI, PSRI and NIR difference at the pixel level, with these reflecting the distribution of photosynthetic, non-photosynthetic and non-submerged aquatic vegetation respectively. The LCCS categories generated through reference to Table 5.15 are listed in Table 5.16. The input layers are illustrated in Figure 5.33.

Table 5.15. Layers referred to in the classification of natural and semi-natural vegetation. The symbols (A, B etc.) are used to describe the land covers occurring and assign LCCS codes (see Table 5.16).

Code	A	A	A	A	A	A	A	B	C	C	D	E	E	E	E
	Woody	Trees	Shrubs	Herbaceous	Forbs	Graminoids	Cover categories	Height categories	Continuous	Fragmented	Broadleaved	Evergreen	Deciduous	Mixed	Perennial
Number	1	3	4	2	5	6	10-16	2- 13	1	2	1	1	2	5	6*
Lifeform	ABC	A	BC	DEFGH	GH	DEF									
Cover							ABC DEFG								
Height								ABC DEFG							
Distribution									D	ABCF					
Leaf type											ABC				
Phenology												C	AB	DEF	DE

*Modifier

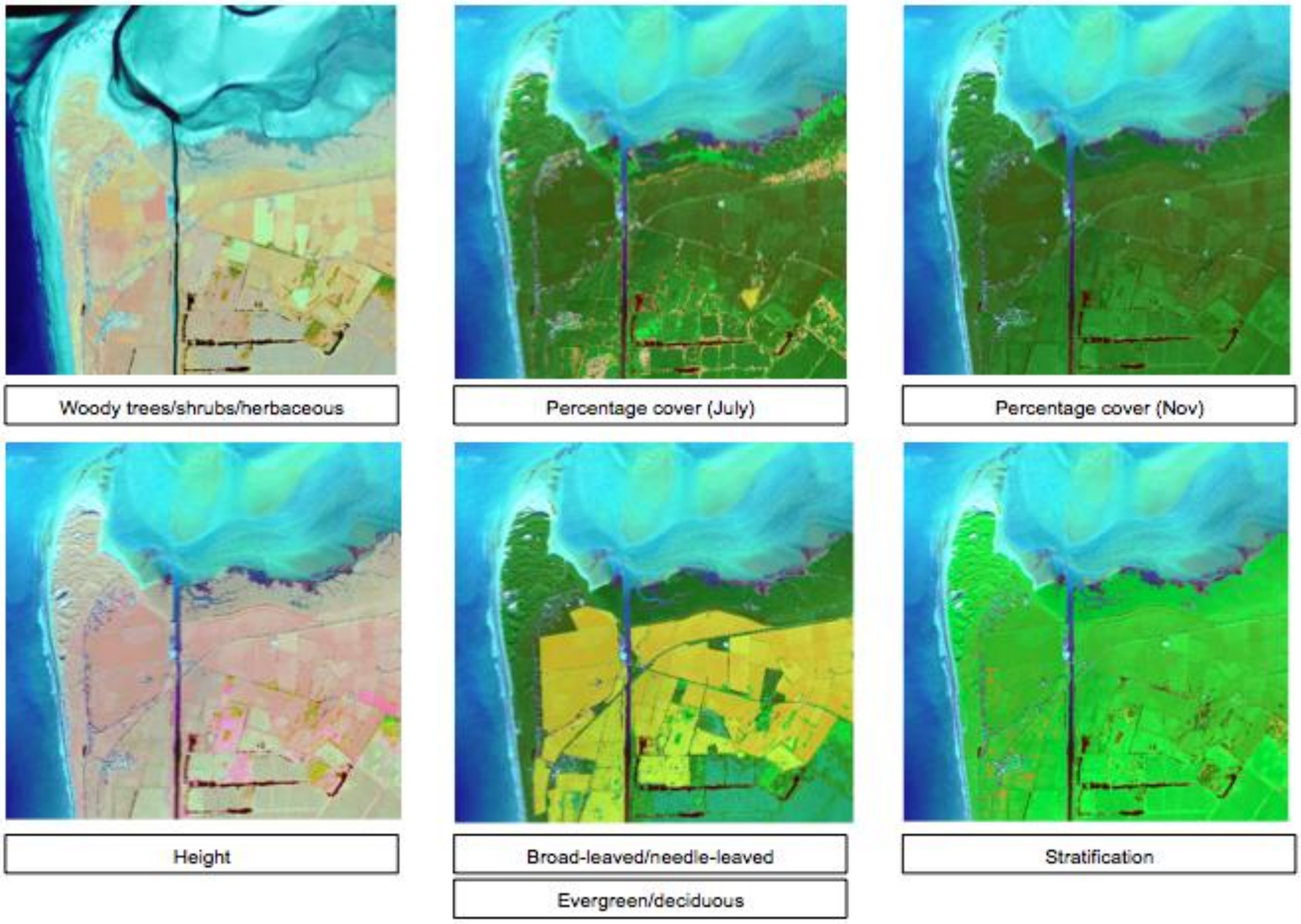


Figure 5.33. Layers used to describe semi-natural terrestrial and also aquatic vegetation.

As highlighted in Section 3, broad-leaves were separated from needle-leaves on the basis of the NIR reflectance. Phenology was determined by comparing the distribution of photosynthetic vegetation between the two image dates (July and November), noting that the majority of herbaceous vegetation was mixed in composition and also perennial.

Table 5.16. LCCS Categories defined for Cors Fochno and surrounds (natural/semi-natural)

	Cat.	LCCS Code Modifer	Description	GHC
A	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
B	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
C	A12	A1.A4.A11.B3.C2.D1.E1	Broad-leaved Evergreen Fragmented Shrubland single layer.Heathland (uplands)	TRS/SCH/EVR
D	A12	A2.A6.A10.B4.C1.E5_B12.E6	Closed Perennial Medium Tall Grassland (e.g., Molinia/Juncus)	HER/CHE
E	A12	A2.A6.A11.B4.XX.E5_A12.B12.E6	Open ((70-60)-40 %) Perennial Medium Tall Grassland (e.g., Eriophorum)	HER/CHE
F	A12	A2.A6.A10.B4.C2.E5_B13	Closed short grassland	HER/CHE
G	A12	A2.A5.A10.B4_B11	Closed medium tall forbs (3.0-0.8 m)	HER/LHE
H	A12	A2.A5.A10.B4_B12	Closed medium tall forbs (0.8-0.3 m)	HER/LHE

5.6.3 Natural and semi-natural aquatic vegetation

The primary habitats protected by the Cors Fochno Natura 2000 site are the active and modified raised bog; perennial closed tall grasslands (e.g., dominated by *Phragmites australis*) and saltmarshes are also prevalent on the margins of the bog and around the estuary respectively.

Table 5.17. Layers referred to in the classification of aquatic natural and semi-natural terrestrial vegetation. Note that the symbols (A, B, C and D) are used to describe the land covers occurring and assign LCCS codes (see Table 5.18).

Code	A		A	A	A	A	A	A	B	C	C	C	D	D	E	E	E	E	F	F
	Woody	Herbaceous	Trees	Shrubs	Forbs	Graminoids	Lichens/mosses	Closure	Height	Water seasonality	Persistent whole day	Within daily variations	Broadleaved	Needlet-leaved	Evergreen	Deciduous	Mixed	Perennial	No more layers	Herbaceous
Number	1	2	3	4	5	6	7	13-20	2-4	1	4*	5*	1	2	1	2	5	6*	1	4
Aquatic life form	A	BC		A		BC														
Cover								ABC												
Height									ABC											
Water seasonality										ABC	AB*	C								
Leaf type													A	A						
Leaf phenology															A		B	B	A	A

*Modifier

The extent of natural and semi-natural aquatic vegetation had been defined within Level 3 and the only additional layer required to classify the active raised bog was water seasonality. From ecological knowledge, all land covers were waterlogged with the main separation between them being that water was persistent for the whole day within the active bog and perennial closed tall grasslands (primarily phragmites), but varied daily over the saltmarshes. In the latter case, the classification was based on knowledge of the tidal regime and the extent of cover as a function of elevation. However, differences in the WBI between the July and November acquisitions also indicated, to a certain extent, the level of tidal inundation but this varies daily depending upon the state and relative height of the tide. The LCCS categories generated through reference to Table 5.17 are listed in Table 5.18.

Table 5.18. LCCS Categories defined for Cors Fochno and surrounds (aquatic natural/semi-natural)

	Cat.	LCCS Code Modifer	Description	GHC
A	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)
B	A24	A2.A6.A12.B4.C1.E5_B11.C4.E6	Perennial closed tall grassland on permanently flooded land (persistent)	HER/HEL
C	A24	A2.A6.A13.B4.C1_B13.C5	Open short grassland on permanently flooded land (with daily variations) (Unmanaged Saltmarsh)	HER/HEL

5.6.4 Artificial surfaces

Whilst the area of urban development is not large, a wide range of infrastructures occurs. Buildings are primarily residential with some industrial development (primarily boat yards) and agricultural constructions (e.g., cattle sheds associated with farmhouses). The main transport infrastructure is the roads, the majority of which are paved, and the railway. To detect urban areas, appropriate segmentation is required. Once delineated, these were identified primarily because of a lack of a recognisable vegetation spectrum, allowing separation based on a combination of differences in the red and green reflectance (lack of a red trough) and also the red and near infrared (lack of a red edge). Linear features were separated on the basis of length/width ratios of segments following merging and all remaining areas were assigned to buildings. The density of buildings can be determined by first isolating individuals and creating larger objects representing, for example, caravan parks. The number of individual buildings counted within the larger objects can then be used to assign a density estimate. The LCCS categories for the urban areas surrounding Cors Fochno are defined with reference to Table 5.19 and are listed in Table 5.20.

Table 5.19. Layers referred to in the classification of artificial surfaces. The symbols (A, B etc.) are used to describe the land covers occurring and assign LCCS codes

Code	A	A	A	A	A	A		A	A
	Built up	Linear	Non-linear	Roads (paved)	Railways	Industrial and other areas	Urban areas	Urban density (medium)	Urban density (scattered)
Number	1	3	4	8	10*	12*	13*	15*	17*
Surface aspect		AB	CDE	A	B	E	CD		
Urban density								C	D

*Modifier

Table 5.20. LCCS Categories defined for Cors Fochno and surrounds (natural/semi-natural)

	Cat.	**LCCS Code Modifer**	**Description**	**GHC**
A	B15	A3_A8	Paved road(s)	URB/ART
B	B15	A3_A10	Railway(s)	URB/ART
C	B15	A4_A13.15	Urban areas (medium density)	URB/ART/NON
D	B15	A4_A13.17	Urban areas (scattered)	URB/ART/NON
E	B15	A4_A12	Industrial and/or other areas	URB/ART?NON

5.6.5 Bare areas

The main naturally bare areas occur in the estuarine complex and are associated primarily with the sand dunes and flats. Three layers are used to describe the surfaces; surface aspect, material and macropattern (Table 5.21). The main areas of consolidated material are coarse fragments of gravels, stones and boulders that are located in some parts of the estuary. Unconsolidated material includes the bare soil (mud) and loose and shifting sands, which are located towards the landward margins or as isolated islands (in lower tidal regimes) within the estuary. The sand dunes are barchans and are considered to be unsaturated.

Table 5.21. Layers referred to in the classification of bare areas (natural). The symbols (A, B etc.) are used to describe the land covers occurring and assign LCCS codes (see Table 5.22).

Code	A	A	A	A	A	A	A	A	A	B	B	B	B
	Consolidated	Unconsolidated	Bare rock and or coarse fragments	Gravels, stones and boulders	Bare soil and other unconsolidated material	Loose and shifting sands	Bare rock	Stony	Very stony bare soil	Dunes	Barchans	Unsaturated	Saturated
Number	1	2	3	8	5	6	7	12	13	1	2	5	6
Surface aspect			A		B								
Material						CD	A	D	E				
Macropattern										C	C		C

*Modifier

Within the area mapped previously as bare (B16), unconsolidated material is associated with land which is within the tidal area and which is of low slope and elevation. The tidal area between the July and November image acquisitions was defined using the difference in the WBI. In addition, dry sand above the tidal area was mapped using the green reflectance bands. Areas of consolidated material were mapped based on homogeneity of reflectance in the green channels, with reflectance being comparatively lower than the surrounding muds and sands. Dunes were identified as these supported a higher slope compared to the estuarine sands and were differentiated into barchans (saturated) based on knowledge of the sand dune complex. The LCCS categories mapped are listed in Table 5.22.

Table 5.22. LCCS Categories defined for Cors Fochno and surrounds (natural/semi-natural)

	Cat.	LCCS Code Modifer	Description	GHC
A	B16	A3_A7	Bare rock	SPV/ROC
B	B16	A5	Bare soil and other unconsolidated material	SPV/EAR
C	B16	A6.B2	Shifting Sands.Saturated Barchans	SPV/SAN
D	B16	A6_A12	Stony loose and shifting sands	SPV/STO
E	B16	A5_A13	Very stony bare soil and unconsolidated material(s)	SPV/STO/GRV

5.6.6 Artificial and natural waterbodies

A diverse range of water bodies occurred within the Cors Fochno area and surrounds, including the open sea, estuarine (tidal) water and pools. Whilst a number of the waterbodies could be classified as artificial, the majority had been created to support the long-term water holding capacity of the active bog. Furthermore, the rivers had been straightened over 100 years previously and hence could be regarded as natural. A large number of layers were used to describe water bodies including physical status (water, ice or snow), dynamics (standing or flowing), persistence, tidal regimes, depth and sediment load (Table 5.23).

Table 5.23. Layers referred to in the classification of bare areas (natural). The symbols (A, B etc.) are used to describe the land covers occurring and assign LCCS codes (see Table 5.18).

Code	A	A	A	B	B	C	C	D	D
	Water	Flowing	Standing	Perennial (> 9 months)	Tidal area	Shallow	Deep to medium	Almost no sediment	Sediment
Number	1	4	5	1	3	1	6	2	5
Physical status	ABCDEF	BEF*	ACD*						
Dynamics				ACD	E				
Persistence									
Duration									
Tidal area									
Depth						AD	BCE		
Sediment load								A	

*Modifier

To define the extent of water, the WBI was used. Flowing and standing water were differentiated on the basis of enclosure, as outlined previously. All water bodies within the study area are perennial but time-series of satellite sensor data could be used to establish the persistence of water. The tidal area was connected to the sea, with the upper level based on elevation. The depth of the water was identified within reference to the near infrared wavebands whilst the sediment load made use of the coastal bands. The LCCS categories for both artificial and natural categories are listed in Tables 5.24 and 5.25. The layers used are illustrated in Figure 5.34.

Table 5.24. LCCS Categories defined for Cors Fochno and surrounds (natural/semi-natural)

	Cat.	LCCS Code Modifer	Description	GHC
A	B27	A1.B1.C2.D2.A5	Clear shallow artificial waterbody (Standing)	SPV/AQU
B	B27	A1.B1.C6_A4	Turbid Deep to Medium Deep Artificial Perennial waterbodies (Flowing)	SPV/AQU
C	B27	A1.B1.C1_A5	Deep to Medium Perennial Artificial Waterbodies (Standing)	SPV/AQU

Table 5.25. LCCS Categories defined for Cors Fochno and surrounds (natural/semi-natural)

	Cat.	LCCS Code Modifer	Description	GHC
D	B28	A1.B1.C1_A5	Deep to Medium Perennial Natural Waterbodies (Standing)	SPV/AQU
E	B28	A1.B3_A4.B6	Tidal Area (Flowing); Surface Aspect (sand)	SPV/AQU(TID)
F	B28	A1_A4	Natural waterbodies, flowing (ocean/sea)	SPV/AQU/SEA

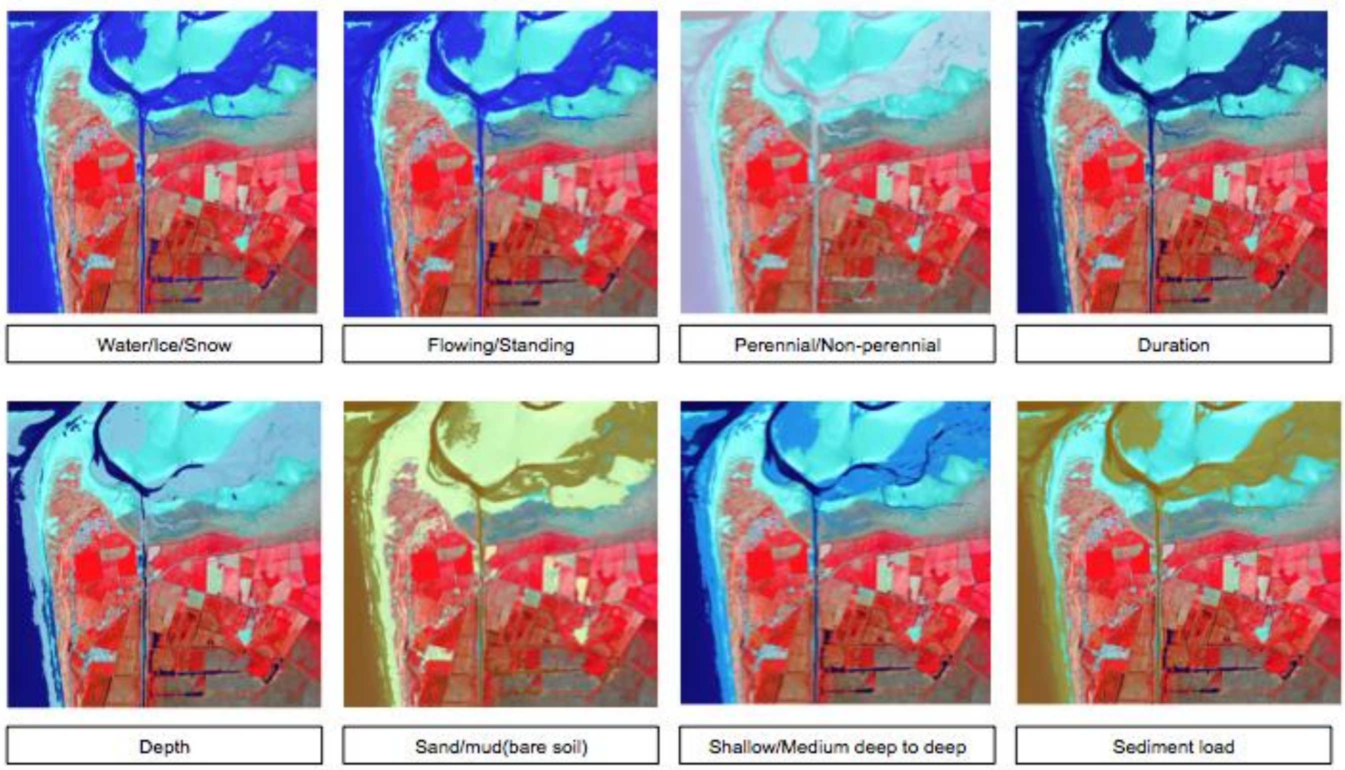


Figure 5.34 Layers used to describe aquatic vegetation

5.6.7 The LCCS Classification Beyond Level 3.

By referencing the different layers generated, a classification of all objects into LCCS categories was achieved for Cors Fochno and surrounds (Figure 5.35). The main natural and semi-natural categories are associated with the aquatic habitats of the active raised bog and saltmarshes and terrestrial vegetation on the sand dunes as well as the areas of semi-natural broadleaved woodland, shrub (e.g., gorse) and herbaceous grasslands (e.g., dominated by *Molinea caerulea*) and forbs (e.g., bracken; *P. aquilinum*). The cultivated area consists primarily of permanent grasslands used for grazing sheep and cattle. A diverse range of naturally bare categories occur, with these associated largely with the estuarine environment and sand dunes. Water bodies include those that are flowing through the sea and estuary (tidal) and in smaller rivers and streams (non-tidal) and stationary such as in the ponds in the active bog. The main artificial bare categories are associated with the roads, railways and residential developments, with the latter including a large number of caravan parks. The distribution of land covers reflects the patterns observed within the landscape, with some (e.g., woodlands) described in more detail because of the inclusion of height and cover metrics and also phenological information.

To assess the accuracy in the classification of LCCS categories, field data were collected during 2011 and are currently being collected during the summer of 2012. Hence, the assessment of accuracy is not fully reported in this deliverable. The use of the different layers within the LCCS, does however, raise some issues in relation to assessing classification accuracy as this can be assessed in two ways:

a) On the basis of the final LCCS classification (and subsequently its translation to habitat categories)
b) By considering the accuracies in the classification within the different layers (e.g., height, cover, leaf type), with the errors then accumulated.

In this latter approach, a greater knowledge of the causes of misclassification is provided, allowing then the user to focus on the layers that are contributing to the greatest errors in the final classification. This new perspective on the assessment of classification is being considered and will also be reported in Deliverable 5.5.

Whilst the classification provides good detail, this is still insufficient for many conservation managers who require detailed information on habitats and their associated flora and fauna. For this reason, the EODHAM 3rd stage is needed which provides the translation of LCCS categories to habitat categories (i.e., the GHC) and a more detailed classification of the latter, as well as changes in the state of the habitat for monitoring purposes. An overview of the EODHAM 3rd stage as applied to Cors Fochno, is outlined in Section 5.7, and will be reported in full in Deliverable D5.5 for this and other sites.

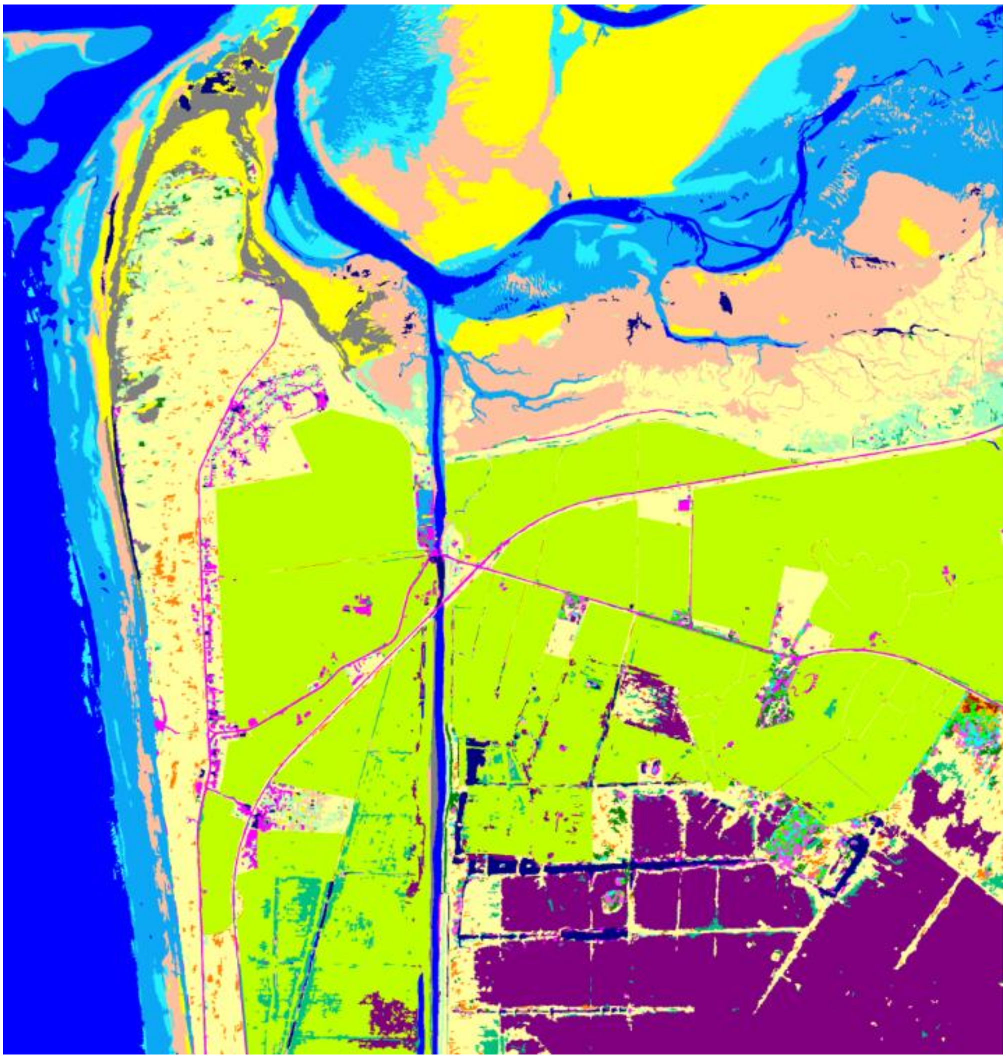


Figure 5.35. Classification of LCCS categories, Cors Fochno and surrounds. The classification reflects the distribution of land covers including the active bog (purple), the cultivated areas (light green), forests (multiple colours because of height and cover categories), unconsolidated sands and muds (yellow and orange), gravels (grey) and urban (magenta).

5.7 EODHAM3rd stage

5.7.1 Translation of LCCS to GHCs.

As indicated in Deliverable 5.2, each of the LCCS categories defined needs to be translated to a GHC. For Cors Fochno, the list of GHCs associated with each of the LCCS categories is listed in Table 5.26 a and b. In each case, a one-to-one translation can be achieved. Once the translation to the GHCs has been performed, a more detailed classification of habitats within the framework of the GHCs can be provided. As an example, Figure 5.36 conveys the translation from the LCCS category of active raised bog (Closed to Open Broad-leaved Evergreen Shrubs with Herbaceous Vegetation on Permanently Flooded Land; Persistent) to the equivalent GHC (TRS/DCH/EVR or TRS/SCH/EVR or TRS/LPH/EVR,

or TRS/MPH-EVR HER/EHY/HEL/SHY-FLO/LEA). Here, a large area is mapped to one class but much more detail is required within the area of the active bog to support on-going conservation management but also monitoring. The active bog is a particularly good example of this problem because the LCCS does not allow a complete description with the vegetation consisting of many components that are discrete in the taxonomy (i.e., woody shrubs, herbaceous forbs and graminoids and lichens and mosses).

Table 5.26a. 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 T all 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

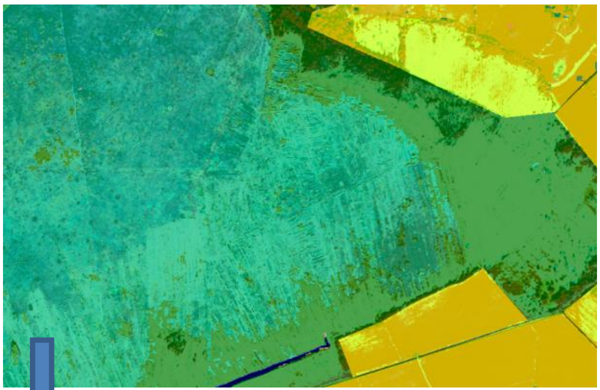
Table 5.26b. LCCS categories and General Habitat Categories, Cors Fochno, mid Wales.

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

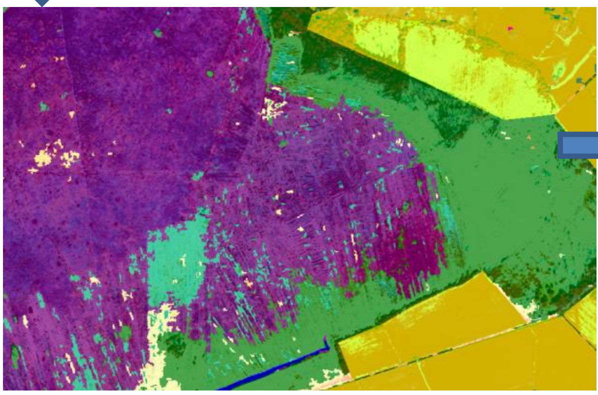
5.7.2 Detailed classification of GHCs

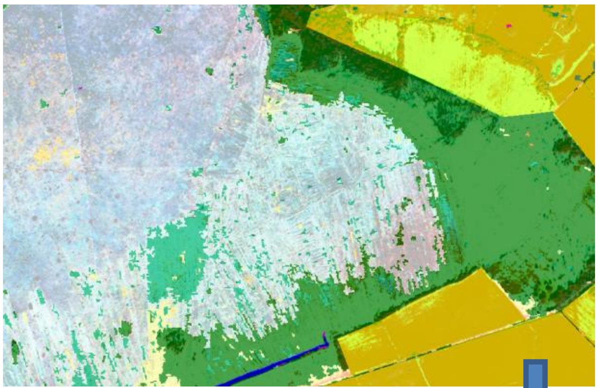
To obtain the detail required for conservation management and monitoring, a more involved classification of the GHCs needs to be implemented to capture detail relating to the dominant life forms within vegetated habitats. For most non-vegetated categories, a one-to-one translation from the LCCS to GHCs is sufficient. However, for many vegetated LCCS classes identified and translated to a GHC, further segmentation is required with this typically to the size of a pixel or object consisting of several pixels. Each of these objects is then run through a second GHC classification (currently implemented with eCognition) to provide a more detailed overview of the life forms occurring within the active bog. A fuzzy-based classification is also performed based on the spectral characteristics of different life forms and/or dominant species, which allows the generation of the GHC classification that takes into account percentage cover and relative dominance. The scheme is illustrated in Figure 5.36, which conveys the sequence of processing needed to perform the classifications.

LCCS to GHC



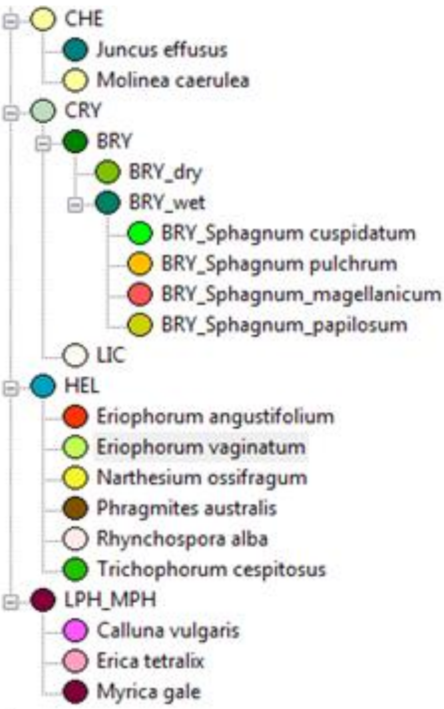
LCCS Level 3

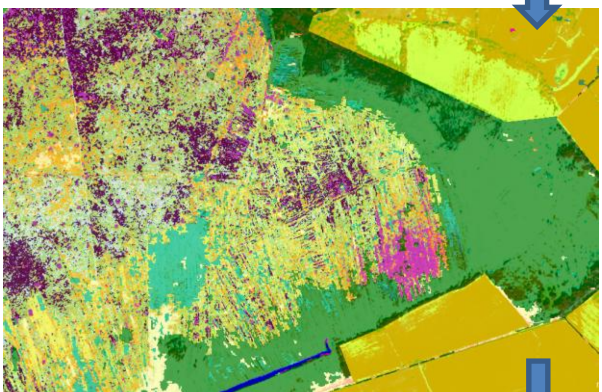




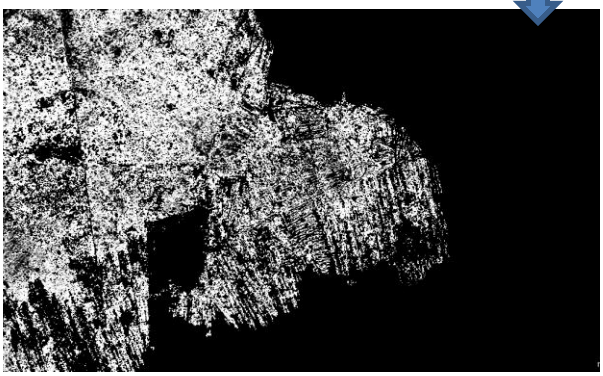
LCCS Classification (active bog)

GHC classification (active bog)





GHC Classification (detailed)



Subpixel classification (*Eriophorum vaginatum*)

Figure 5.36. Approach to the detailed classification of GHCs, Active raised bog, Cors Fochno.

5.8 Using ontologies

5.8.1 Introduction

In BIO_SOS, a Geographic Object-Based Image Analysis (GEOBIA) is performed for land cover and habitat mapping. This technique emerged in the early 2000s and represents an important paradigm shift for processing remote sensing images (Hay et Castilla 2008). GEOBIA is defined as a sub discipline of Geographic Information Science (GIScience) devoted to developing automated methods to partition remote sensing imagery into meaningful image-objects, and assessing their characteristics through spatial, spectral and temporal scales, so as to generate new geographic information in GIS-ready format (Hay et Castilla 2008).The object-oriented classification approach uses semantics based on descriptive assessment and knowledge, i.e. it incorporates the knowledge of the user (Blaschke et Strobl 2001). The user guides the image processing operation based on expert knowledge in order to produce more reliable maps. However, whereas GEOBIA is currently widely adopted, there are issues associated with its application which have been summarized in detail Comber, Fisher, et Wadsworth 2005).

First, each data producer conceptualizes the reality he intends to represent in the image. Indeed, this is more of a computer-aided photo-interpretation process, where two experts using the same data can potentially produce two different results. Therefore, GEOBIA does not support the automation of image interpretation. Secondly, the processing chain to achieve a classification is neither entirely controlled nor well documented. For example, segmentation algorithms implemented through GEOBIA can sometimes be ill-posed (Hay et Castilla 2008). as the choice of segmentation parameters can be subjective and thus hinder the quality assessment of the classification. The classification step itself is based on features of interest selected by the user and the determination of thresholds to identify these features, based on a “trial-and-error’ analysis. Thirdly, as a consequence, GEOBIA leads to non-robust methods that are non-transferrable. Although it is possible to exchange process trees performed on a defined dataset acquired for a given study area, its reuse will only be valid for a comparable dataset (i.e. comparable sensor images, study areas and acquisition dates).

Thus, although GEOBIA is a promising technique for the efficient processing of remote sensing images, one may wonder how the information on data conceptualization and feature/threshold selection is to be communicated to the remote sensing community, i.e. to the users and to the scientists interested in the geographic information produced. Since neither the method nor the cognitive process that led to the production of the geographic information, are well documented, we argue that GEOBIA needs to be accompanied by a method that would allow the knowledge of GEOBIA specialists to be represented in order to aggregate and share it with the remote sensing community.

Since the future of GEOBIA appears to be an issue of knowledge management, it is likely that knowledge representation techniques can play a pivotal role. Ontologies are widely used in various scientific areas such as in biological sciences (Bard et Rhee 2004; Renear et Palmer 2009). Thus, various authors have highlighted the necessity for developing ontologies in Geographic Information Science (GIS) (Mark et al. 2005; Visser et al. 2002; Fonseca et al. 2002; Agarwal 2005; Buccella, Cechich, et Fillottrani 2009; Couclelis 2010). In this deliverable, after a brief introduction to ontologies, we identify the requirements for building ontologies in order to represent the expert knowledge used in the classification of LCCS categories as part of the EODHaM 1st and 2nd stages. We then illustrate how ontologies can help in describing expert knowledge and in image analysis.

5.8.2 Introduction to ontologies

5.8.3 Definition

An ontology is an explicit expression of a concept, which is an abstract, simplified view of the world that we wish to represent for some purpose (Gruber 1995). In the conceptualization phase, we identify the concepts and their relationships within a scientific domain such as ecology, geography or earth observation (J. S. Madin et al. 2008). When the knowledge of a domain is formally represented, it is first necessary to define the set of objects that can be represented in what is called the universe of discourse (Gruber 1995). Therefore, a more concise definition of an ontology is a formal depiction of reality through explicit definition of concepts (terms), representing real (individual) domain entities, relations between these entities, and the rules and constraints which govern the behavior between them (Tripathi et Babaie

2008). Here a concept denotes a collection of “instances” that share common characteristics (J. S. Madin et al. 2008). In ontologies, the main relations between objects consist of rules based on specialization/generalization (is-a) and aggregation/composition (part-of). However, the user is allowed to define other types of semantic relations, which can then include spatial and/or temporal rules. Finally, rules and constraints on concepts can be included in the ontologies through description logics.

5.8.4 Building ontologies

The first step in the construction of ontologies consists of a good conceptualization of the domain to be represented. A good conceptualization must be based on clearly defined concepts and relations. Broad concepts mean that the set of objects represented by the extension of a given term is very large (Essler, 1982 (Heink et Kowarik 2010). For instance, Amazonia is a broad concept as it can refer to the Amazon basin, the Amazonian forest or the Legal Amazon.

Once the domain of interest has been conceptualized, it is then necessary to build the ontology as close as possible to the conceptualization. A good ontology must be neutral, computationally tractable, acceptable to and reusable by those working in the domain that the ontology describes (Smith et al., 2007).therefore, Gruber 1995 proposed a set of design criteria for assessing if an ontology is accurate enough for representing a domain of interest:

- Clarity: An ontology should represent objective complete definitions of terms. These definitions must be documented with natural language.
- Coherence: An ontology should identify inferences that are inconsistent. Specific software designed for ontologies such as Protégé can help in verifying consistency.
- Extendibility: An ontology should be designed to anticipate the uses of the shared vocabulary. One should be able to define new terms for special uses based on the existing vocabulary, in a way that does not require the revision of the existing definitions.
- Minimal ontological commitment: Ontological commitments are agreements to use the shared vocabulary in a coherent and consistent manner. In that sense, an ontology should require the minimal ontological commitment sufficient to support the intended knowledge sharing activities.

Since ontologies are useful for representing and sharing domain knowledge, it should allow any user to include its own knowledge in an ontology. In order to allow someone to extend an ontology, it is necessary to structure the way ontologies are built. Thus, it has been suggested to develop different ontological levels whose architecture is described by (Guarino 1998; Fonseca et al. 2002) (Figure 7.1):

- Top-level ontologies describe very general concepts like space, time, matter, object, events and action independently of a specific domain or problem.
- Domain ontologies describe the vocabulary related to a specific domain (e.g. ecology, geography, remote sensing) by specializing the terms introduced in the top-level ontology.
- Task ontologies, similar to domain ontologies, describe the vocabulary related to a generic task or activity (e.g. image processing).
- Application ontologies describe concepts depending both on a particular domain and task, which are often specializations of both the related ontologies.

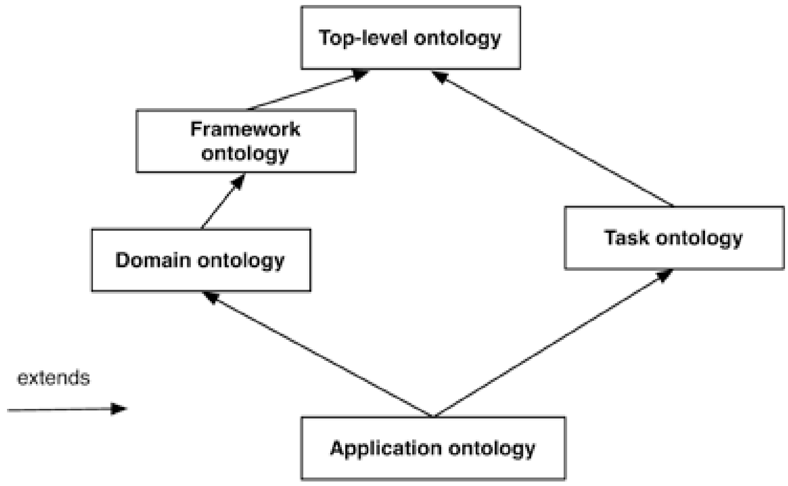


Figure 5.37. Architecture of ontological levels

5.9 Requirements for the construction of an Earth Observation ontology

Since ontologies are used for creating, aggregating and sharing knowledge, any potential user should be able to extend ontologies for their purposes. Therefore, ontologies are structured in various ontological levels, including top-level, domain, task and application ontologies (Figure 5.37) (Guarino 1998; Fonseca et al. 2002). It is first necessary to describe these ontological levels for the future construction of Earth Observation ontology.

5.9.1 Top-level ontology

Top-level ontologies provide definitions for general-purpose terms and act as a foundation for interconnecting more specific domain ontologies (Niles and Pease, 2001; Madin et al., 2008). For example, an upper ontology would allow connecting an ontology of Earth Observation with other domain ontologies about ecology (Madin et al. 2008), hydrogeology (Tripathi et Babaie 2008) or biology (Bard et Rhee 2004). Various top-level ontologies exist, including BFO, Cyc, DOLCE, GFO, PROTON, Sowa's ontology and SUMO (Mascardi, Cordì, et Rosso 2007). In the context of Earth Observation, relevant framework ontologies should include general concepts regarding **Observation** for describing the **Earth Environment**.

The Extensible Observation Ontology (OBOE) introduced by Madin *et al.* (2007) serves as a formal conceptual framework for describing the semantics of observational data sets. The ontology asserts that an **Observation** is of some **Entity** (e.g. a forest observed on a Landsat image or in the field) and is carried out through the **Measurement** of one (or more) **Characteristic** of this Entity (e.g. the forest's spectral response in the Near-Infrared Band) based on a **Measurement Standard** (e.g. reflectance in percent). Finally, the measurement has a value and, eventually, an associated precision. It is essential to teat both the observation and the measurement with equal importance. First, the observation has a **Context** that refers to the relationships the entity maintains with other entities at the time of the observation. Second, the measurements are taken by a **Recorder** using a **Protocol** at a particular time and place (e.g. the forest of interest can be observed at different locations and at different instants in the year) (Madin et al., 2007).

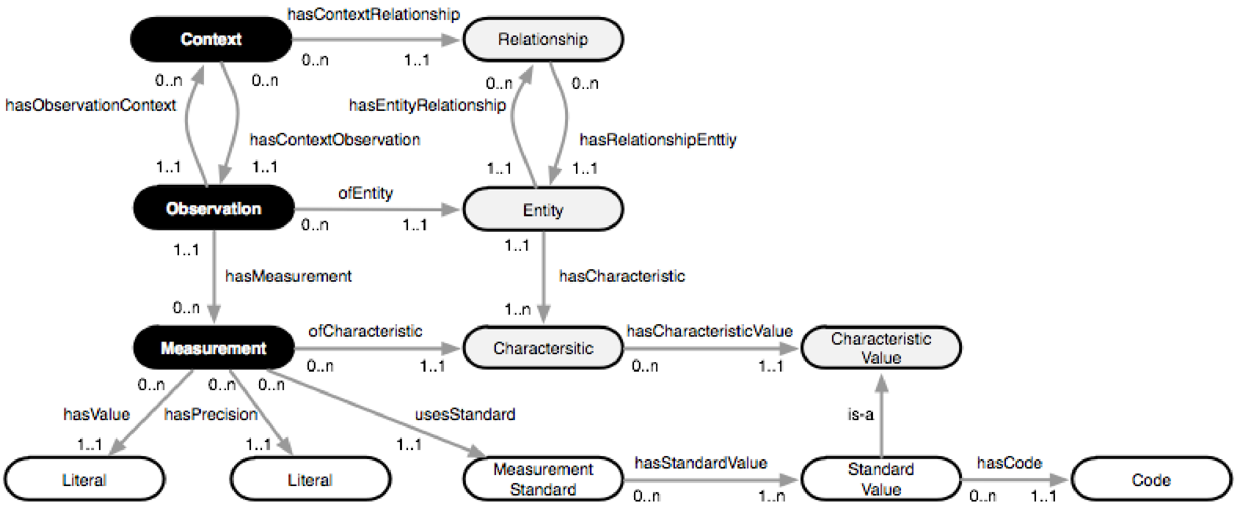


Figure 5.38. OBOE

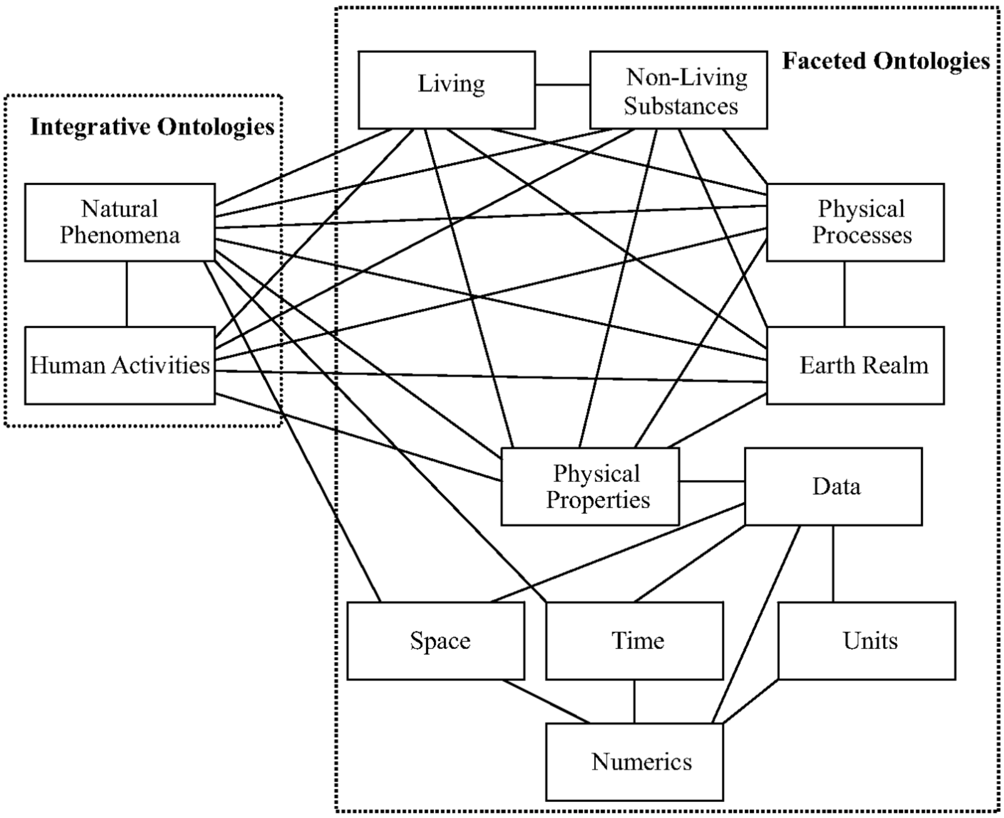


Figure 5.39: SWEET ontologies and their interrelationships

Regarding the Earth Environment, the Semantic Web for Earth and Environmental Terminology (SWEET) proposed by NASA is an appropriate ontology (Raskin et Pan 2005) (http://sweet.jpl.nasa.gov/). This ontology is publically available and is composed of nine top-level ontologies (Figure 5.39) written in the OWL language and containing around 6000 concepts. The nine ontologies are organized in “faceted” and “integrative” ontologies, including orthogonal concepts and integrative science knowledge concepts, respectively (Madin et al. 2008). The “faceted” ontologies refer to **Representation**, **Process**, **Matter**, **Realm**, **Property**, **State** and **Relation** while Integrative ontologies refer to **Phenomena** and **Human Activities**. These ontologies are highly modular since, potentially; any compound concept related to environment can be described based on concepts introduced in orthogonal

ontologies. As an example, the compound concept *air temperature* can be described by the physical property *temperature* (coming from the ontology on Properties), which applies to the substance *air* (coming from the ontology on Matter) (Tripathi et Babaie 2008). Then, it is expected that complex compound concepts of interest for domain applications (e.g. land cover mapping) can be described by faceted concepts selected in the SWEET “faceted” ontologies.

The strength of OBOE and SWEET ontologies lies in their extension capabilities that can be used to connect various ontologies. Thus, OBOE and SWEET can be interconnected and can be used to develop relevant domain and task ontologies.

5.9.2 Domain ontologies

Domain ontologies describe the vocabulary related to a specific domain (e.g. ecology, geography, remote sensing) by specializing the terms introduced in the top-level / framework ontologies. Based on the OBOE and SWEET ontologies, it is possible to identify main domain ontologies that have to be built for GEOBIA applications. We identified four relevant domain ontologies.

5.9.3 Land cover ontology

In the context of Earth Observation, the Entity concept should be specified as a Geographic Entity, which corresponds to any entity that occupies a position in space (Mark 1993), i.e. Lake Victoria, Amazon River, Amazonian Forest, Paris, etc. The observation of such entities is based on the measurement of a characteristic which refers to land cover. The land cover class that is assigned to a geographic entity is then extracted from a specific measurement standard, which refers to a classification domain, i.e. a land cover nomenclature. Such nomenclature must then be translated into a land cover ontology. There has been significant effort towards building land cover ontologies (http://harmonisa.uni-klu.ac.at/content/land-use-land-cover-ontologies), especially, the Land Cover Meta Language (LCML), also named Land Cover Classification System (LCCS), produced by the Food and Agriculture Organization (FAO) which is an attempt to classify the geographic entities in very simple groups of characteristics, arranged in different ways. These groups act as building blocks to describe the more complex land cover classes (http://www.glcn.org/ont_2_en.jsp). However, the LCCS has two limitations for remote sensing-based land cover mapping. Firstly, the context of the observation is not considered. Secondly, the LCCS is used for classifying "real world" geographic entities whereas the remote sensing expert classifies representations of these entities in images as revealed through image processing techniques. These two points highlight the need for the building of new domain ontologies.

5.9.4 Image ontology

An image ontology should contain all information on concepts usually used by remote sensing experts to describe representations of geographic entities in satellite images. The representation of a geographic entity in a digital image is called a geographic object. Whereas the geographic entity can be described by its real characteristics (vegetation height, leaf type, etc), the geographic object is described by its characteristics in the image. In a GEOBIA approach, the image characteristics are organized in four levels: Level 0 features refers to singe pixel values, Level 1 features refer to single object characteristics, including its spectral response, texture, geometry; Level 2 features refer to spatial relations between two objects; Level 3 features refer to the way two or more objects are arranged together (Liu, Guo, et Kelly 2008). This refers in photo-interpretation to what is called Structure and is linked to the context. In the SWEET, the Property ontology mainly refers to physical properties so that it should be extended to include image properties with important concepts such as texture, shape as well as spectral indices.

5.9.5 Spatiotemporal ontologies

The context of the observation refers to the relationships the entity maintains with other entities at the time of the observation. These relationships are either spatial or temporal. For example, a forest can border a river or can be a secondary forest that regenerated after a clearing. For example, it has been

defined that spatial relations can refer to topology, distance or orientation (distance and orientation being refinements of topological relations). Topological relations have mainly been described in two models : the 9-intersection Matrix (Egenhofer et Herring 1991) and the RCC8 calculus. If these relations are suitable for GIS applications including multi-valued space (like vectors); its application to single-valued space (raster) is limited. Indeed, after an image segmentation process, the only topological relations that can exists between two image objects are “disjoint” or “touch”. It is then necessary to develop additional metrics to describe the way two objects are spatially related in images (Liu, Guo, et Kelly 2008; Alboody, Sedes, et Inglada 2009). Such relations should then be included into spatial-temporal ontologies. The SPAN/SNAP ontology is a framework ontology dedicated to such relations. The SWEET ontologies also include important concepts on spatial and temporal relations that could be used to express such relations between geographic entities.

5.9.6 Sensor ontology

In OBOE, the **Recorder** of a measurement can be human or non-human (J. Madin et al. 2007). In remote sensing, non-human recorders especially refer to **Sensors** onboard satellites (e.g. the Thematic Mapper onboard Landsat 5 that measures reflectance values in different spectral bands). It would then be necessary to develop a specific domain ontology for remote sensors. Various initiatives in this area will be analyzed: OntoSensor (Russomanno, Kothari, et Thomas 2005), SensorML (http://www.opengeospatial.org/standards/sensorml) and the Semantic Sensor Networks (Compton et al. 2009).

5.9.7 Task ontology

A task ontology describes the reasoning concepts and their relationships occurring within a certain domain and for a specific task (Timpf 2002). It is organized into three different levels, i.e. the lexical level, the conceptual level and the symbol level (Ikeda et al. 1997). The lexical level deals with the syntactic aspect of the problem solving description. The conceptual level captures conceptual meaning of the description and the symbol level corresponds to a runnable program and specifies the computational semantics of the problem solving (Ikeda et al. 1997). For example, at the lexical level, a user might claim that vegetation indices are efficient to monitor vegetation in remote sensing images. Examples of vegetation indices (NDVI, EVI, SAVI, etc) and their corresponding mathematic equations would then be specified at the conceptual level. Finally, at the symbol level, the vegetation indices should be linked with operators from an image processing software that enables the vegetation index to be computed (e.g. otb::Functor::NDVI, from the Orfeo Tool Box image processing software that allows NDVI images to be computed).

It is noteworthy that the lexical level actually refers to knowledge contained in application ontologies (see next section) where links between vegetation and vegetation indices would be described between the land cover ontology and the image ontology, respectively. Consequently, the task ontology must focus on the conceptual and symbol levels, i.e. on (i) the methods used for extracting information from the remote sensing images, and on (ii) the tools used to carry out the method. It has thus been proposed to create method ontologies and software tool ontologies (Benjamin et al. 2005), where the method ontology describes the generic tasks used in remote sensing applications (data acquisition, data processing, data analysis, error assessment, final presentation) and the software tools ontology includes specific software operators that can be used to perform a defined task. The method and tools ontologies should then be connected to upper-level ontologies such as SWEET which already include a large amount of concepts linked to physical and mathematical processes. Ideally, the task ontologies should be built to extend the SWEET ontologies to build an image processing ontology. It is expected that the task ontologies, including knowledge on methods and associated operators from different software, can improve the interoperability of processes and data, enhancing the capability to design optimal scientific workflows.

5.9.8 Application ontologies

Ontologies are composed of taxonomies and axioms. Taxonomy is a hierarchical system of concepts and axioms are rules, principles, or constraints among the concepts (Ikeda et al. 1997). Whereas the taxonomy is introduced in the top-level, domain and task ontologies, the axioms that describe the concepts are defined in the application ontology. Thus, application ontologies describe concepts depending both on a particular domain and task, which are often specializations of both of these related ontologies.

5.10 Example of application: from knowledge to image

5.10.1 Including knowledge on geographic entities

The building of ontologies has yet to be carried out. The first task will consist of building the domain ontology of land cover and habitat based on the LCCS and GHC nomenclatures, respectively. The ontology will be built in OWL on Protégé software. A brief example based on a GHC class (FPH/CON for Coniferous Forest Phanerophytes) illustrates the approach (Figure 5.40). First, the main concepts are listed. In the case of FPH/CON, these include concepts such as Geographic Entity, Property, and Measurement Standard. These concepts can be easily connected to OBOE and SWEET ontologies. Secondly, the main relations such as has Leaf, has Tree, has Property are defined. Third, the main data properties are included to express class properties. For example, the Height concept can have a float value. Finally, concepts are defined based on the upper concepts, relations and data properties identified. For example, the FPH/CON concept is described as a sub-class of the GHC class FPH, which is a sub-class of the concept forest. In the example of FPH/CON, an area is defined as having at least 2 tall trees (a tall tree is a tree whose height is higher than 5 meters), 2 coniferous trees (a coniferous tree is a tree whose leaves are needleleaves) and whose vegetation cover is higher than 30%.

To check for consistency, we defined virtual instances of geographic entities (Figure 5.41). For instance, an individual named ForestFPH/CON has been defined as a geographic entity whose vegetation cover is 30% and who contains one coniferous tall tree and one coniferous medium tree (higher than 5 meters). A reasoner algorithm was then used to infer to which class this individual belongs. In Figure 5.4, it appears that this individual has been correctly classified as a member of the concept FPH/CON. We then expect that any entity that would be described as an individual could be classified based on either GHC or LCCS nomenclatures.

5.10.2 Including knowledge on geographic objects for classifying images

Once the entities are described in the ontology, their description in the remote sensing image must be added. The next task will then consist of representing all the descriptions contained in this deliverable in one ontology. As a short example, we proposed that the concept of vegetation can be observed in a remote sensing image using vegetation indices, such as the NDVI, whose values for vegetated pixels range from 0.3 to 0.8. The concept can then be defined as follows:

$$\begin{gathered}\mathrm{SegmentVegetal} \equiv \mathrm{Segment} \ \sqcap \ \exists \mathrm{hasFeature}(\mathrm{SegmentFeature} \\ \sqcap \ \exists \mathrm{hasProcessing}.\{\mathrm{ndvi}\} \ \sqcap \ \exists \mathrm{Mean}\{\geq 0,3\} \ \sqcap \ \exists \mathrm{Mean}\{\leq 0,8\})\end{gathered}$$

Where the right hand side of the equivalent axiom is only built with concepts and properties imported from the image ontology. This equation was then translated into Protégé and used to classify a remote sensing image.

First, we used the Soprano interface to extract a semantic description of remote sensing images thanks to a methodology based on the semantic web vocabularies in order to extract and serialize the concepts from the image in the RDF/XML syntax. For example, the description of one segment (an individual) is represented below: I

. . .

<img:Segment rdf : about=" ht tp: // cartam . s a t /uneImage#1">
<img:segmentOf
        r d f : r e s o u r c e=" ht tp: // cartam . s a t /uneImage#segmentat ion "/>
</ img:Segment>
. . .
<img:PseudoSpect ralSegmentFeature
        rdf : about=" ht tp: // cartam . s a t /uneImage#ndvi 1 ">
<img:Kur tos i s rdf :d a t a t yp e=" ht tp: //www.w3 . org /2001/XMLSchema#double ">
        -0.98629782506</ img:Kur tos i s>
<img:Maximum rdf : da t a t y pe=" ht tp: //www.w3 . org /2001/XMLSchema#double ">
        -0.055116762289</img:Maximum>
<img:Mean r df :da t a t ype=" ht tp: //www.w3 . org /2001/XMLSchema#double ">
        -0.12566546593</img:Mean>
<img:Median rdf :d a t a t yp e=" ht tp: //www.w3 . org /2001/XMLSchema#double ">
        -0.12635899372</ img:Median>
<img:Minimum rd f :da t a t ype=" ht tp: //www.w3 . org /2001/XMLSchema#double ">
        -0.21288248665</img:Minimum>
<img:Var ianc e rd f :da t a t ype=" ht tp: //www.w3 . org /2001/XMLSchema#double ">
0.0021005467679</ img:Var ianc e>
. . .

</ img:PseudoSpect ralSegmentFeature>
. . .

The descriptions of each segment then populate the image ontology so that, at this point, both the image and the remote sensing expert knowlegde are expressed based on the same formalization. It then becomes possible to infer new knowledge thanks to the reasoning capabilities of ontologies. For this purpose, we used a reasoner to classify segments in the image through inference between the semantic description of features and the remote sensing ontology.

We tested our approach on a calibrated (in reflectance and temperature) Landsat 5-TM image of the city of Santarem (in the Brazilian Amazon) from 07/12/2009 where we intended to detect vegetated segments. The segmentation and feature extraction steps were performed separately to prepare the data that serves as input for our ontological approach used at the classification step. The ontological engineering was used to perform the classification thanks to the Fact++ reasoner. We obtained the image shown in Figure 5.42a where white segments represent vegetation. We validated this experimental result by comparing it with the image obtained by using a traditional operator in image processing software (Figure 5.42b). Both images are consistent if we focus on the vegetated zone (in white in the pictures) and validate the possibility of using ontologies to interpret satellite images.

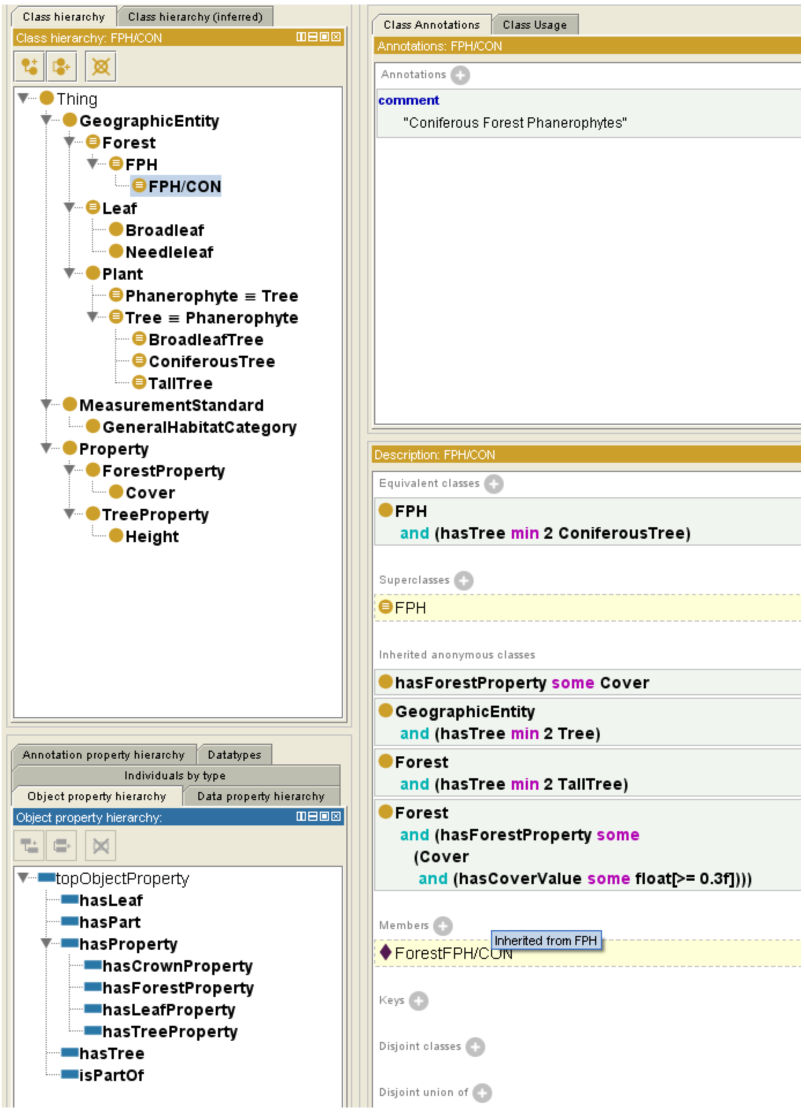


Figure 5.40. Use of Protégé software

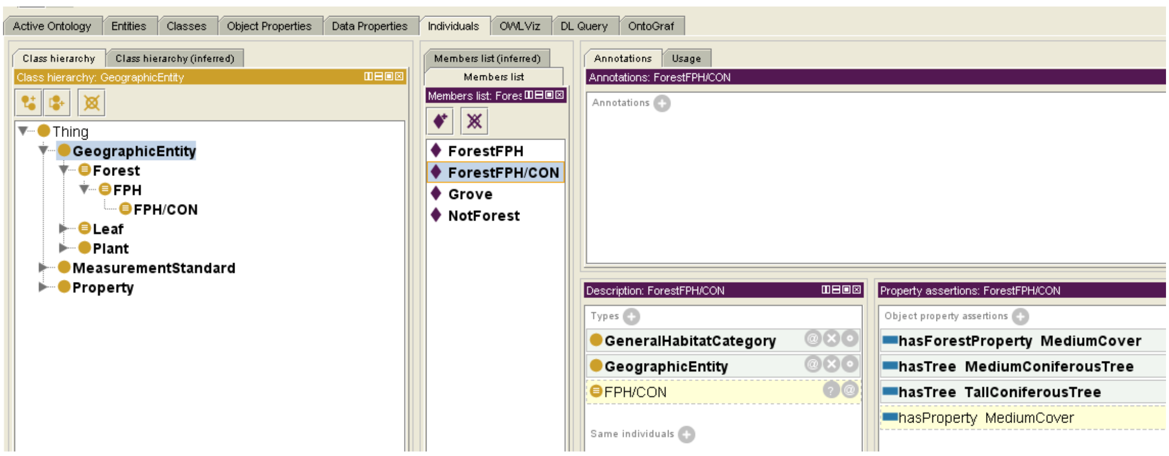


Figure 5.41. Use of Protégé software

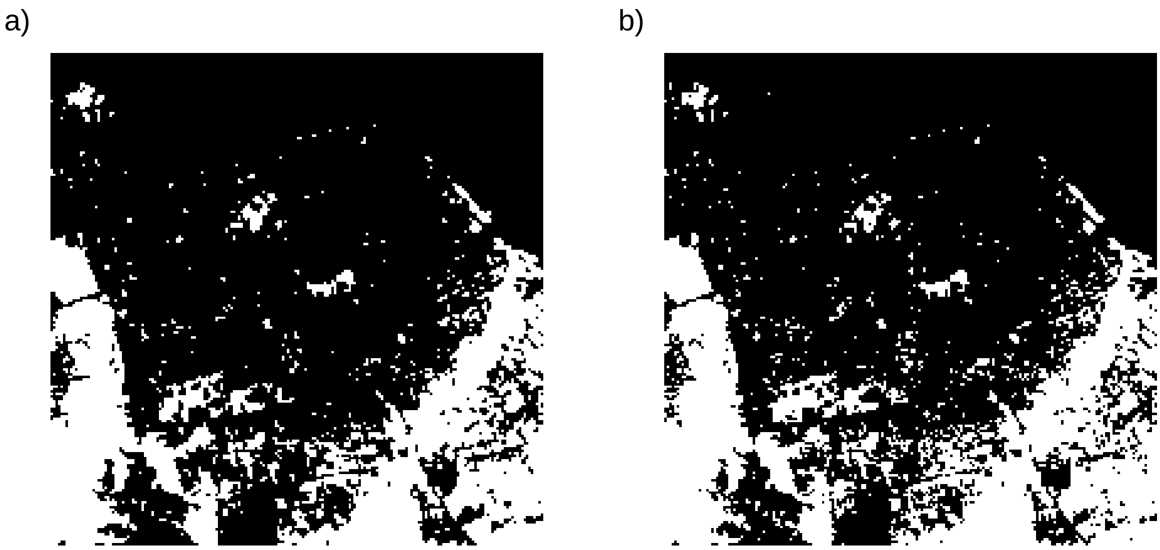


Figure 5.42. Classifications based on ontological engineering

6 Annex 2.

6.1 Le Cesine (Italy)

6.1.1 Le Cesine (Italy): spectral rules adopted for LCCS Level 1 to 2 classification

The decision tables with spectral rules adopted in P5, P6 and P7 sub-components of DFD in Figure 5.25 are reported hereafter

Table 6.3. Decision Table of P5 procedure, LCCS_L1_Vegetated_vs_NotVegetated

Predicates	Truth Values			
(I3>1 and I4<=1) or (I1>1 and I2<=1)	Y	N	Y	N
I1<=1 and I3<=1 and I5>=0.2	Y	N	N	Y
Resulting Class				
L1_Non Vegetated		X		
L1_Vegetated	X		X	X
Photosynthetic Vegetation (PV)	X		X	
Submerged Vegetation (SV)	X			X

Table 6.4. Decision Table of P6 procedure, Intermediate strata Terrestrial_vs_Aquatic

Predicates	Truth Values			
I1<=1 and I3<=1	Y	N	Y	N
I3>1 and I4<=1 and I1<=2.5 and T1<=0.65	Y	N	N	Y
Resulting Class				
Water or Shadow	X		X	
Emerged Vegetation (EV)	X			X
Terrestrial		X		
Aquatic	X		X	X

Table 6.5. Decision Table of P7 procedure, Combine_Strata_LCCS_L2

Predicates	Truth Values			
is L1_Vegetated	Y	Y	N	N
is L1_Non_Vegetated	N	N	Y	Y
is Terrestrial	Y	N	Y	N
is Aquatic	N	Y	N	Y
Resulting Class				
L2_A1_Terrestrial_Vegetated	X			
L2_A2_Aquatic_Vegetated		X		
L2_B1_Terrestrial_Non Vegetated			X	
L2_B2_Aquatic_Non Vegetated				X

6.1.2 Le Cesine (Italy): rules adopted for LCCS Level 2 to 3 classification

The pseudo-code representing the classification procedure adopted in P11 and P12 sub-components of DFD in Figure 5.25 is reported hereafter, with symbols explained in Table 6.6

Table 6.6 Symbols and nomenclature used in the following pseudo-code procedures:

- Ci, Z_i: generic segment taken from one of the segmented map C (or Z)
- C, Z: whole segmented maps
- class_*something* : class derived from another process named *something*
- merge(*C*, *class*): merge adjacent segments in a segmented map *C* if they belong to the same class *class*
- unmerge(*C*, *class*): unmerge adjacent segments in a segmented map *S* if they belong to the same class *class*
- Enclosed_in(*Ci*,*class*): returns true if generic segment *Ci* is enclosed in *class*
- class(*Ci*): class of the generic segment *Ci*
- All class values are represented in red colour
-

6.1.2.1 *Natural_Non_Vegetated_vs_Ariticial_Non_Vegetated*

proc Natural_Non_Vegetated_vs_Artificial_Non_Vegetated(C):

```
for each Ci:
        class(Ci) <- Unclassified

// Aquatic

for each Ci:
        if I1(Ci)<=1 and I3(Ci)<=1:
                class(Ci) <- Water
for each Ci:
                if class(Ci)=Water and class_Combined_Strata_LCCS_L2(Ci)=L2_B2_Aquatico_Not_Vegetated:
                class(Ci) <- Sea
merge(C,Sea)
for each Ci:
        if class(Ci)=Water and Enclosed_in(Ci,Sea):
                class(Ci) <- Sea
merge(C,Water)
for each Ci:
        if class(Ci)=Water and G1(Ci)<2 and I1(Ci)<0.9:
                class(Ci) <- Artificial_Water_Bodies
for each Ci:
        if area(Ci)<=80px and class(Ci)=Sea:
                class(Ci) <- Water

// Terrestrial
```

```
for each C_i:
    if I6(C_i)>=0.5 and class_Combined_Strata_LCCS_L2(C_i)=L2_B1_Terrestrial_Not_Vegetated and not
adjacent(C_i,Sea):
        class(C_i) <- Artificial
    loop on C:
        for each C_j with class(C_j) != Artificial:
            if I6(C_j)>=0.45 and
class_Combined_Strata_LCCS_L2(C_j)=L2_B1_Terrestrial_Not_Vegetated and adjacent(C_j,Artificial):
                class(C_j) <- Artificial
            else break loop
    if class_Combined_Strata_LCCS_L2(C_i)=L2_B1_Terrestrial_Not_Vegetated and I6(C_i)<0.5:
        class(C_i) <- Bare_soil

merge(C,Bare_soil)

for each C_i:
    if class(C_i)=Bare_soil and adjacent(C_i,Sea):
        class(C_i) <- Sand
for each C_i:
    if class(C_i)=Artificial and ajacent(C_i,Sand):
        class(C_i) <- Sand

return C
```

endproc

6.1.2.2 *Cultivated_Vegetation_versus_Natural_Vegetation*

proc Cultivated_Vegetation_vs_Natural_Vegetation(C)

```
Z = Natural_not_vegetated_vs_artoficial_not_vegetated(C)

for each C_i:
    if class_Combined_Strata_LCCS_L2(C_i)=L2_A2_Aquatic_Vegetated:
        class(C_i) <- Natural_acquatic_veg
    if class_Combined_Strata_LCCS_L2(C_i)=L2_A1_Terrestrial_Vegetated and T2(C_i)>=0.85:
        class(C_i) <- High_textured_veg
    if class_Combined_Strata_LCCS_L2(C_i)=L2_A1_Terrestrial_Vegetated and T2(C_i)<0.85:
        class(C_i) <- terrestrial_vegetation

for each C_i:
    if class(C_i)=High_textured_veg ...
        and adjacent(C_i,class(Z_i)=Bare_soil)  and G2(C_i)<=0.8 and T2(C_i)>1.3:
            class(Ci) <- Single_trees

merge(C,Single_trees)

for each C_i with class(C_i)=Single_trees:
    if distance(C_i,Single_trees)<=0.8px and distance(C_i,Single_trees)>=2px:
```

class(C_i) <- Cultivated_with_trees_core

unmerge(C,Single_trees)

for each C_i with class(C_i)=Single_trees:
class(C_i) <- Terrestrial_vegetation

for each C_i with class(C_i)=High_textured_veg:
class(C_i) <- Terrestrial_vegetation

for each C_i with class(C_i)=Terrestrial_vegetation:
if I4(C_i)<0.95 and T2(C_i)<=0.85 and Wv2_nir2(C_i)<0.15:
class(C_i) <- Managed_vegetation

merge(C,Terrestrial_vegetation)
merge(Z,Sand)

for each C_i with class(C_i)=Managed_vegetation and adjacent(Z_i,Sand):
class(C_i) <- Terrestrial_vegetation

for each C_i with class(C_i)=Managed_vegetation and adjacent(Z_i,Water):
class(C_i) <- Terrestrial_vegetation

for each C_i:
if class(C_i)=Cultivated_with_trees_core or class(C_i)=Managed_vegetation:
if class_Combined_Strata_LCCS_L2(C_i) = L2_A1_Terrestrial_Vegetated:
class(C_i) <- Cultivated_and_managed_terrestrial_vegetation

for each C_i with class(C_i)=Terrestrial_vegetation:
if class_Combined_Strata_LCCS_L2(C_i) = L2_A1_Terrestrial_Vegetated:
class(C_i) <- Natural_terrestrial_vegetation

return C

endproc

6.1.2.3 Decision Table for P13 sub-component

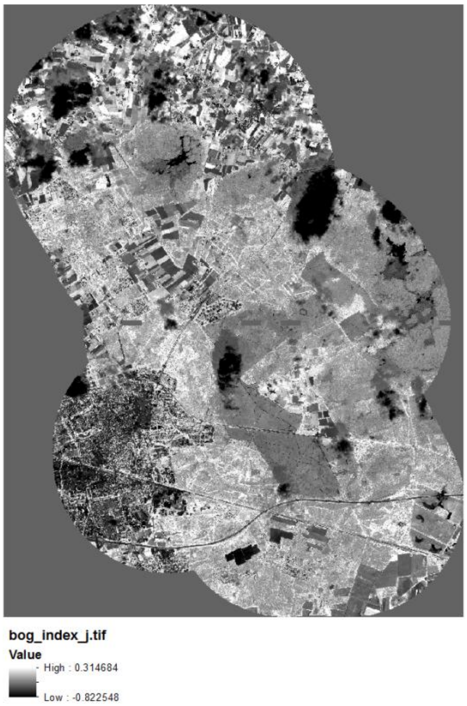
Table 6.7. Decision Table of P13 procedure, Combine_Strata_LCCS_L3

Predicates	Truth Values						
is Cultivated_managed_terrestrial	Y	N	N	N	N	N	N
is Artificial	N	N	N	Y	N	N	N
is Artificial Water Bodies	N	N	N	N	N	N	Y
is Sea	N	N	N	N	N	Y	Y
is L2_A2_Aquatic_Vegetation	N	N	Y	N	N	N	N
is L2_A1_Terrestrial_Vegetation	Y	Y	N	N	N	N	N
is L2_B2_Aquatic_Not Vegetation	N	N	N	N	N	Y	Y
is L2_B1_Terrestrial_Not Vegetation	N	N	N	Y	Y	N	N

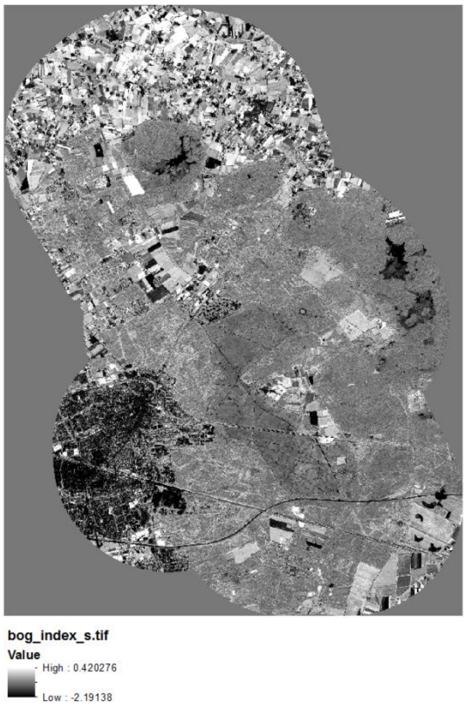
Resulting Class							
L3_A11_Cultivated_Terrestrial	X						
L3_B15_Artificial				X			
L3_A12_Natural_Terrestrial		X					
L3_A24_Natural_Aquatic			X				
L3_B16_Bare					X		
L3_B28_Natural_Water						X	
L3_B27_Artificial_Water							X

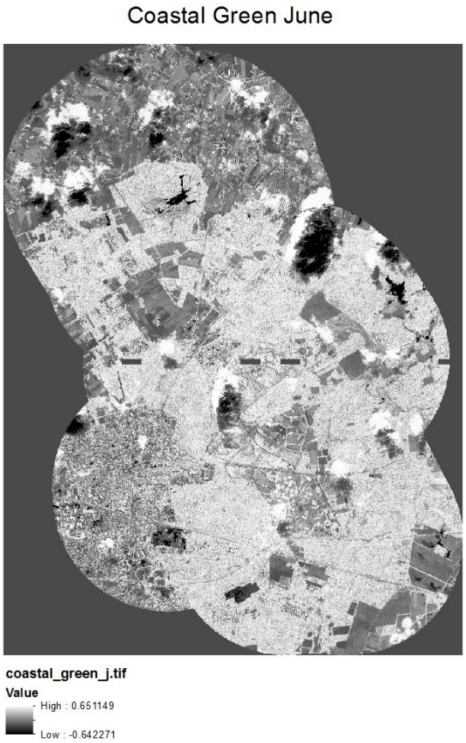
6.2 Examples of indices generated from Worldview data for the Dutch sites

Bog Index June



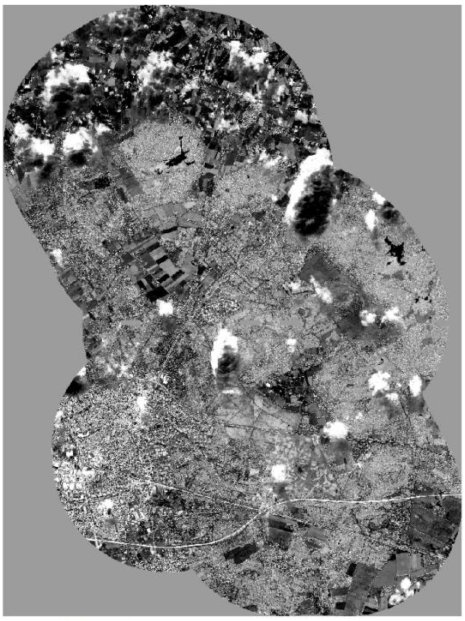
Bog Index September




Coastal Green June
coastal_green_j.tif
Value
High : 0.651149
Low : -0.642271

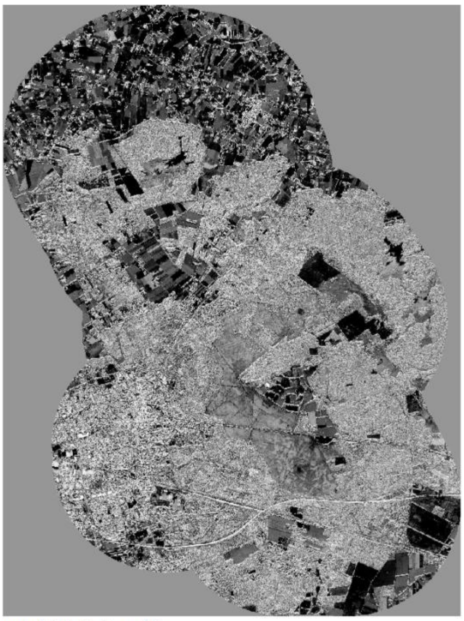

Coastal Green September
coastal_green_s.tif
Value
High : 1.99017
Low : -0.730365

Coastal Red Edge June



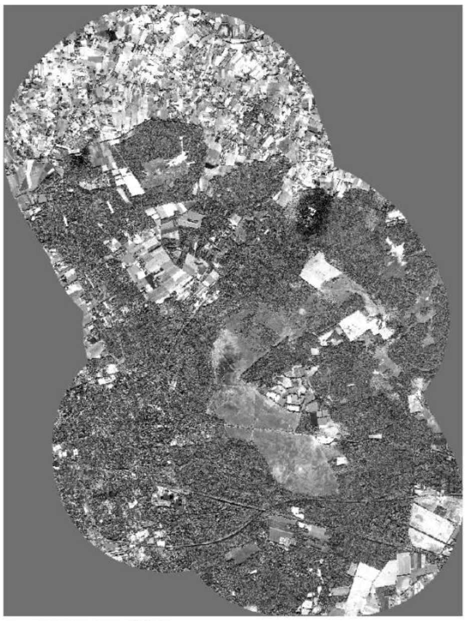
coastal_Red-edge_j.tif
Value
High : 0.352527
Low : -0.57996

Coastal Red Edge September



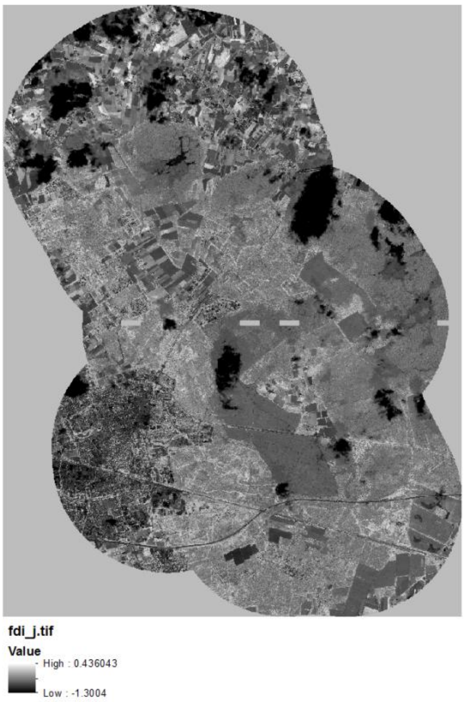
coastal_Red-edge_s.tif
Value
High : 0.712871
Low : -1.97243

Coastal Red Edge Diff

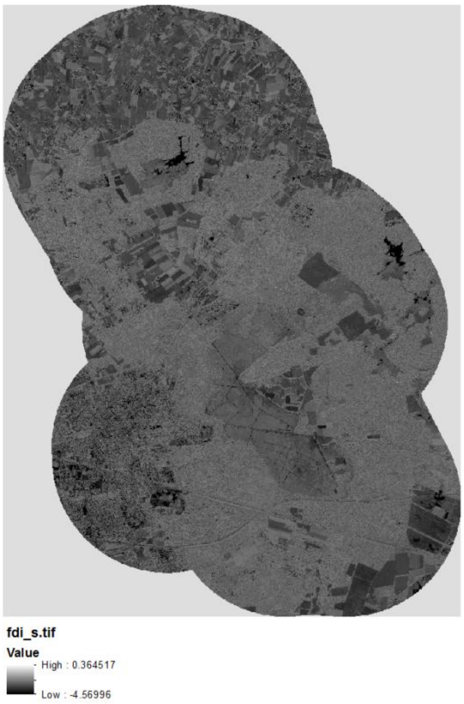


coastal_red-edge_diff.tif
Value
High : 1.9673
Low : -0.690116

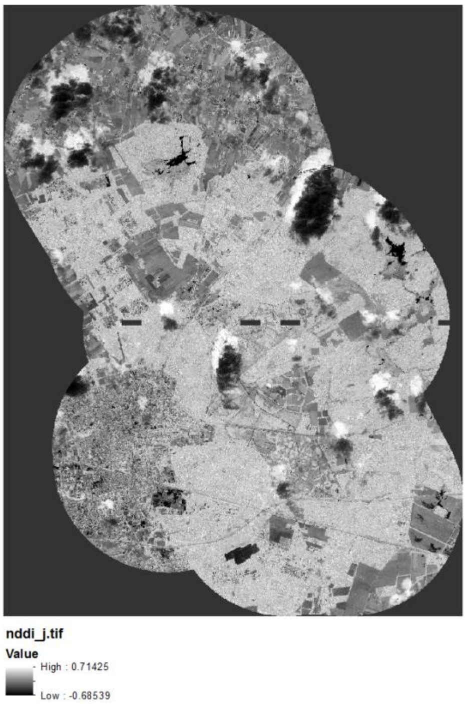
FDI June



FDI September

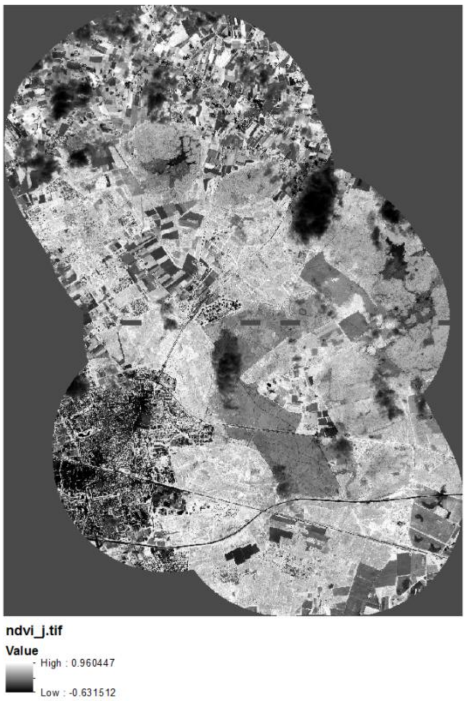


NDDI June



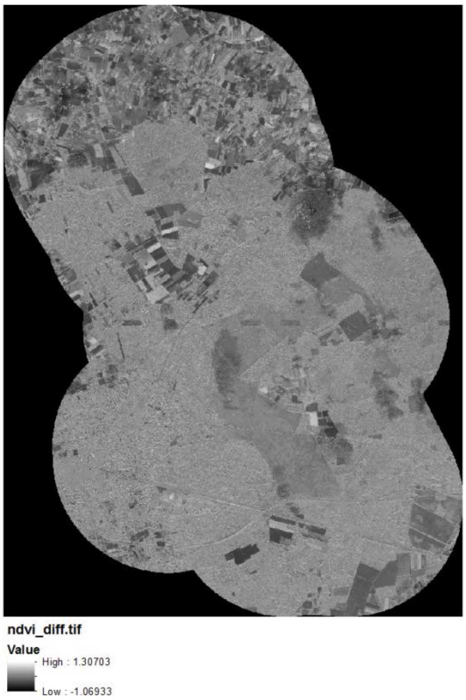
NDDI September

NDVI June

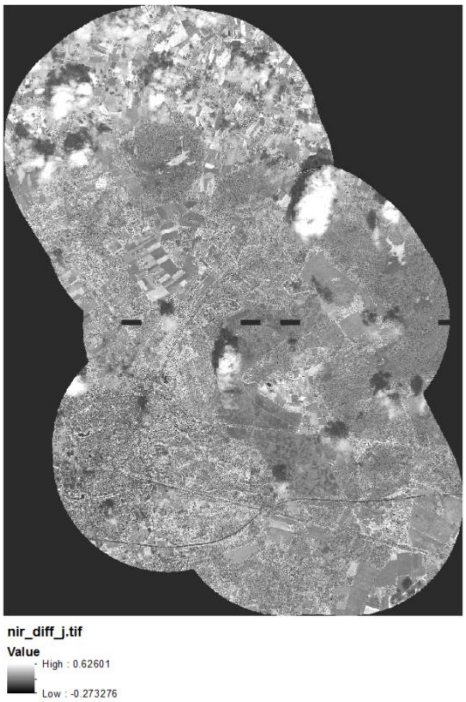


NDVI September

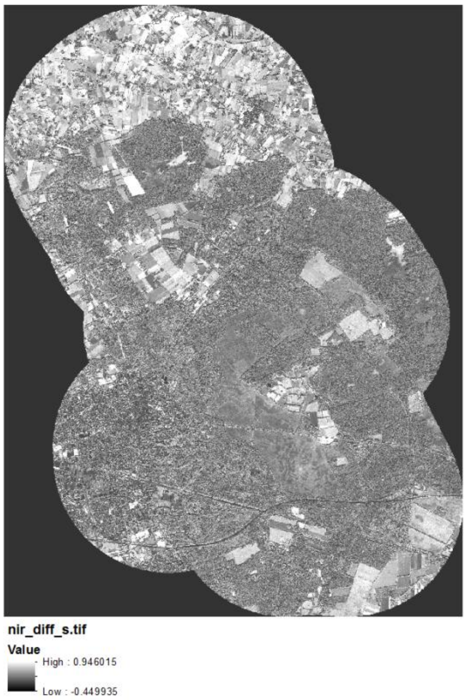
NDVI Diff



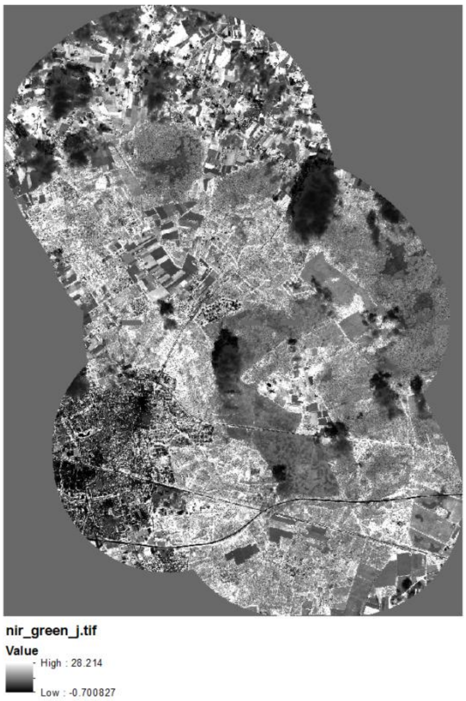
NIR diff June



NIR diff September

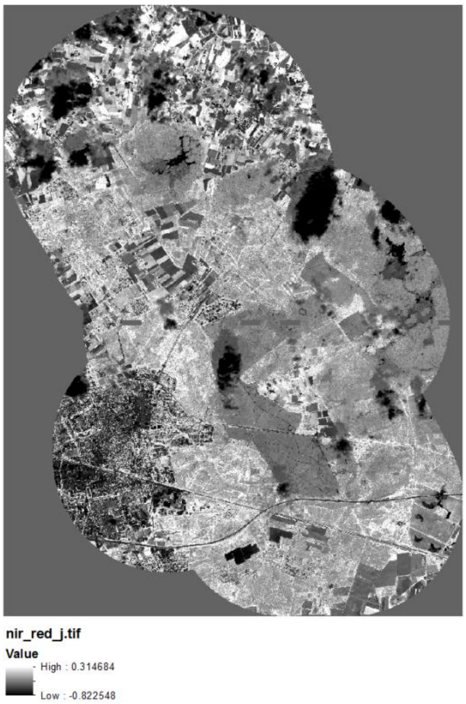


NIR Green June

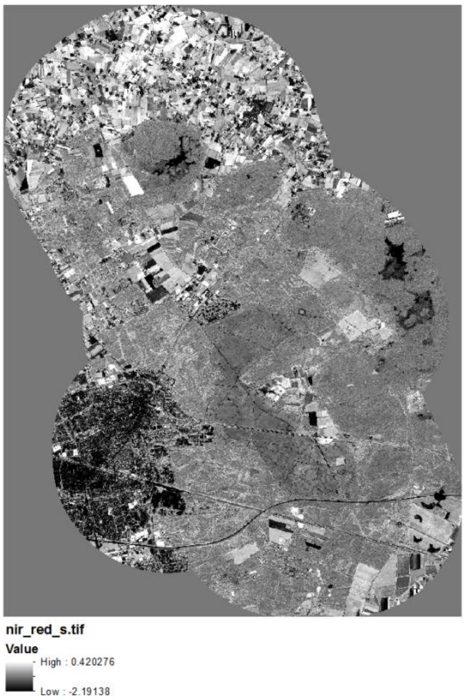


NIR Green September

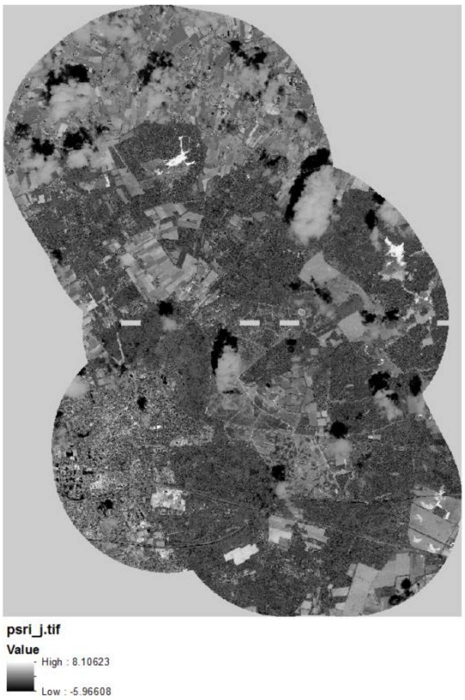
NIR Red June



NIR Red June

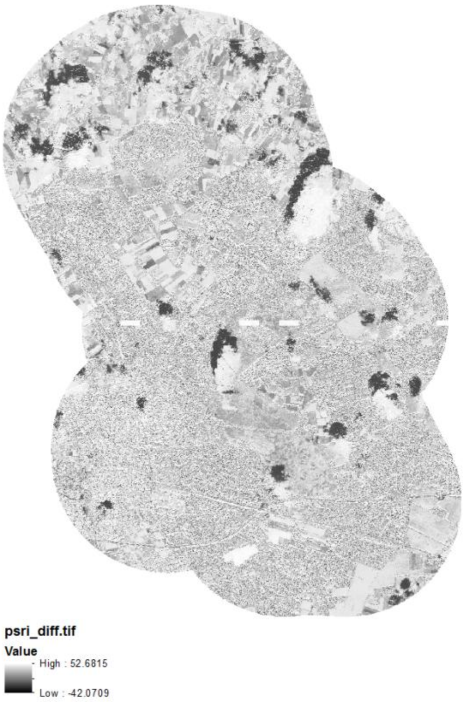


PSRI June

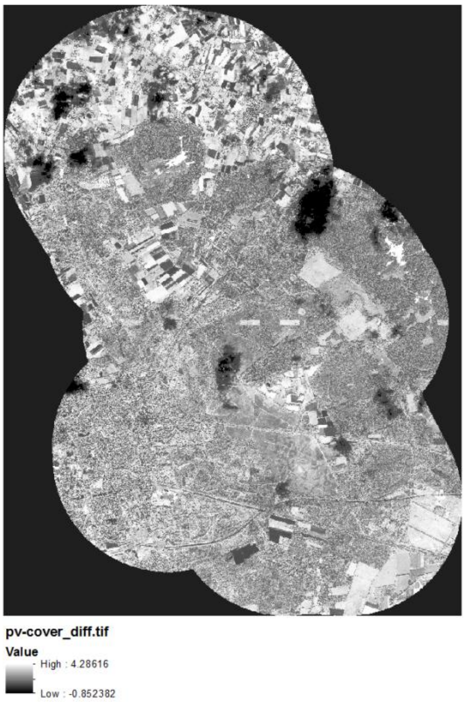


PSRI September

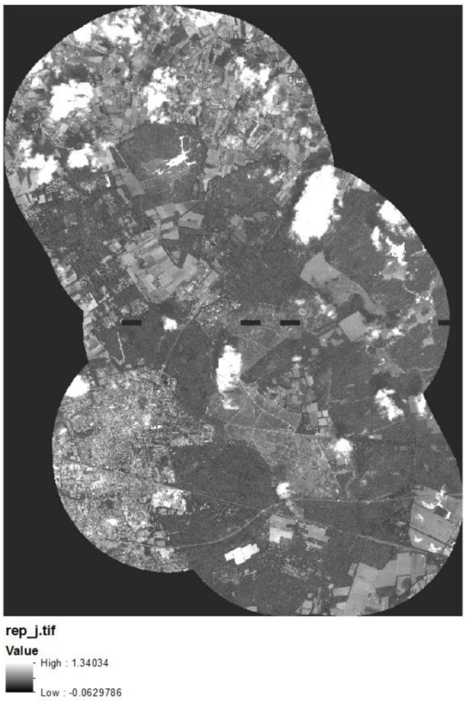
PSRI Diff



PV Cover Diff

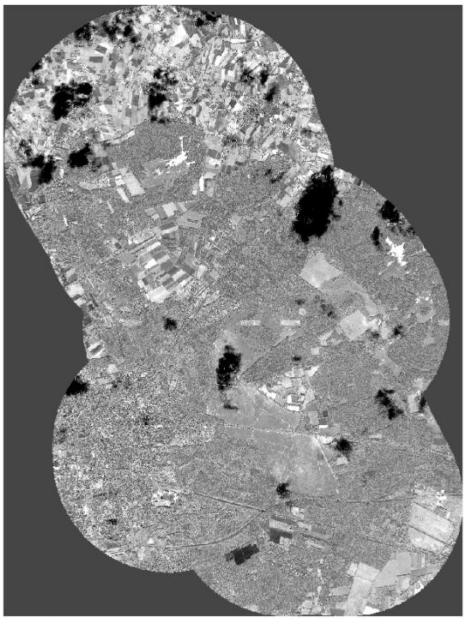


REP June



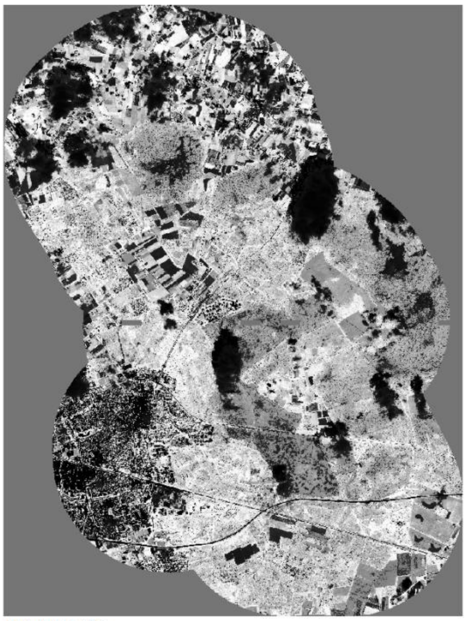
REP September

REP Diff



rep_diff.tif
Value
High : 3.9668
Low : -1.10589

Soil Index June



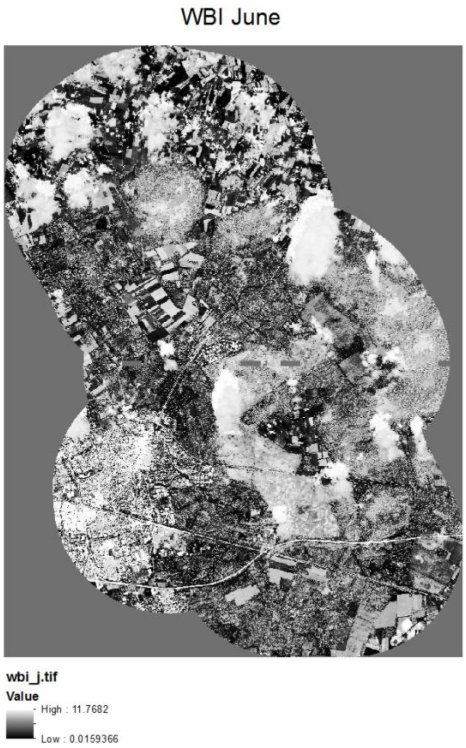
soil_index_j.tif
Value
High : 0.966669
Low : -0.782949

Soil Index September

soil_index_s.tif
Value
High : 0.694809
Low : -0.972396


WBI Green June
wbi_green_j.tif
Value
High : 434.681
Low : 0.911391


WBI Green September
wbi_green_s.tif
Value
High : 1253.16
Low : 1.13824


WBI June
wbi_j.tif
Value
High : 11.7682
Low : 0.0159366


WBI September
wbi_s.tif
Value
High : 120.289
Low : 0.145096

WVVI June

WVVI September

7 References

Agarwal, P. 2005. Ontological considerations in GIScience. International Journal of Geographical Information Science 19 (5) (mai): 501-536. doi:10.1080/13658810500032321.

Alboody, Ahed, Florence Sedes, et Jordi Inglada. 2009. Multi-level Topological Relations of the Spatial Reasoning System RCC-8 . In , 13-21. IEEE. doi:10.1109/DBKDA.2009.13. http://ieeexplore.ieee.

Bard, Jonathan B. L., et Seung Y. Rhee. 2004. Ontologies in biology: design, applications and future challenges . Nature Reviews Genetics 5 (3) (mars): 213-222. doi:10.1038/nrg1295.

Benjamin, P., K.V. Akella, K. Malek, et R. Fernandes. 2005. An ontology-driven framework for process-oriented applications . In Simulation Conference, 2005 Proceedings of the Winter, 9 pp. doi:10.1109/WSC.2005.1574526.

Blaschke, Thomas, et Josef Strobl. 2001. What's wrong with pixels? Some recent developments interfacing remote sensing and GIS . Interfacing Remote Sensing and GIS 6: 12-17.

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

Buccella, Agustina, Alejandra Cechich, et Pablo Fillottrani. 2009. Ontology-driven geographic information integration: A survey of current approaches . Computers & Geosciences 35 (4) (avril): 710-723. doi:10.1016/j.cageo.2008.02.033.

Christophe, E. and Inglada, J., 2007. Robust Road Extraction for High Resolution Satellite Images. In Conference on Image Processing, IEEE International Conference - ICIP , vol. 5, pp. 437-440. DOI: 10.1109/ICIP.2007.4379859

Comber, Alexis, Peter Fisher, et Richard Wadsworth. 2005. Comparing statistical and semantic approaches for identifying change from land cover datasets . Journal of Environmental Management 77 (1) (octobre): 47-55. doi:10.1016/j.jenvman.2005.02.009.

Congalton, R.G., Green K..2009. Accuracy Assessment. Taylor & Francis Group. Second Edition ISBN 978-1-4200-5512-2

Couclelis, Helen. 2010. Ontologies of geographic information . International Journal of Geographical Information Science 24 (12) (novembre 26): 1785-1809. doi:10.1080/13658816.2010.484392.

Compton, Michael, Corey Henson, Holger Neuhaus, Laurent Lefort, et Amit Sheth. 2009. A Survey of the Semantic Specification of Sensors . In 2nd International Workshop on Semantic Sensor Networks, at 8th International Semantic Web Conference.

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

Dennison, P. E. and D. A. Roberts (2003). Endmember selection for multiple endmember spectral mixture analysis using endmember average RMSE. Remote Sensing of Environment **87**(2-3): 123-135.

Egenhofer, Max J., et John R. Herring. 1991. Categorizing Binary Topological Relations Between Regions, Lines, and Points in Geographic Databases. Technical Report. Tech. Report, Department of Surveying Engineering. Orono, ME: University of Maine. http://www.spatial.maine.edu/%7Emax/9intReport.pdf.

Fonseca, Frederico T., Max J. Egenhofer, Peggy Agouris, et Gilberto Camara. 2002. Using Ontologies for Integrated Geographic Information Systems . Transactions in GIS 6 (3) (juin): 231-257. doi:10.1111/1467-9671.00109.

Gruber, Thomas R. 1995. Toward principles for the design of ontologies used for knowledge sharing? International Journal of Human-Computer Studies 43 (5-6) (novembre): 907-928. doi:10.1006/ijhc.1995.1081.

Guarino, N. 1998. Formal Ontology in Information Systems: Proceedings of the 1st International Conference June 6-8, 1998, Trento, Italy. 1st éd. Amsterdam, The Netherlands, The Netherlands: IOS Press.

Hay, G. J., et G. Castilla. 2008. Geographic Object-Based Image Analysis (GEOBIA): A new name for a new discipline . In Object-Based Image Analysis, éd par. Thomas Blaschke, Stefan Lang, et Geoffrey J. Hay, 75-89. Lecture Notes in Geoinformation and Cartography. Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-540-77058-9_4.

Haralick R.M., Shanmugan K., Dinstein I. (1973), textural features for image classification, **SMC-3** (6), pp. 610-621

Heink, Ulrich, et Ingo Kowarik. 2010. What are indicators? On the definition of indicators in ecology and environmental planning . Ecological Indicators 10 (3) (mai): 584-593. doi:10.1016/j.ecolind.2009.09.009.

Huete, A., K. Didan, et al. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment **83**(1–2): 195-213.

Ikeda, Mitsuru, Mitsuru Ikeda, Kazuhisa Seta, Kazuhisa Seta, Riichiro Mizoguchi, et Riichiro Mizoguchi. 1997. Task Ontology Makes It Easier To Use Authoring Tools . In In Proc. of the 15th IJCAI, 342–347. Morgan Kaufmann Publishers, Inc.

Liu, Yu, Qinghua Guo, et Maggi Kelly. 2008. A framework of region-based spatial relations for non-overlapping features and its application in object based image analysis . ISPRS Journal of Photogrammetry and Remote Sensing 63 (4) (juillet): 461-475. doi:10.1016/j.isprsjprs.2008.01.007.

Madin, Joshua, Shawn Bowers, Mark Schildhauer, Sergeui Krivov, Deana Pennington, et Ferdinando Villa. 2007. An ontology for describing and synthesizing ecological observation data . Ecological Informatics 2 (3) (octobre): 279-296. doi:10.1016/j.ecoinf.2007.05.004.

Madin, Joshua S., Shawn Bowers, Mark P. Schildhauer, et Matthew B. Jones. 2008. Advancing ecological research with ontologies . Trends in Ecology & Evolution 23 (3) (mars): 159-168. doi:10.1016/j.tree.2007.11.007.

Mark, David M. 1993. Toward a theoretical framework for geographic entity types . In Spatial Information Theory A Theoretical Basis for GIS, éd par. Andrew U. Frank et Irene Campari, 716:270-283. Berlin, Heidelberg: Springer Berlin Heidelberg. http://www.springerlink.com/index/10.1007/3-540-57207-4_18.

Mark, David M., Barry Smith, Max J. Egenhofer, et Stephen C. Hirtle. 2005. Ontological Foundations for Geographic Information Science . In A research agenda for geographic information science, éd par. Robert Brainerd McMaster et E. Lynn Usery. Boca Raton, Fla.: CRC Press.

Mascardi, Viviana, Valentina Cordì, et Paolo Rosso. 2007. A comparison of upper ontologies. In Dagli Oggetti agli Agenti, Agenti e Industria: Applicazioni tecnologiche degli agenti software. Edited by Matteo Baldoni, Antonio Boccalatte, Flavio De Paoli,Maurizio Martelli, Viviana Mascardi, WOA 2007, Genova, Italy, 24th to 25th September 2007

Merzlyak, M. N., A. A. Gitelson, et al. (1999). Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. Physiologia Plantarum **106**(1): 135-141.

Niles, I. and Pease, A. Toward a Standard Upper Ontology. In Formal Ontology in Information Systems. Proceedings of the 2nd International Conference (FOIS-2001), Christopher Welty and Barry Smith (eds.), 2–9. New York: ACM Press.

Penuelas, J., I. Filella, et al. (1993). The reflectance at the 950-970 nm region as an indicator of plant water status. International Journal of Remote Sensing **14**(10): 1887-1905.

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

Petrou, Z.I., C. Tarantino, M. Adamo, P. Blonda and M. Petrou, (2012). Estimation of Vegetation Height through Satellite Image Texture Analysis, 22nd Cong. of the Int. Soc. of Photogrammetry and Remote Sens. (ISPRS), Melbourne, Accepted.

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

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

Raskin, Robert G., et Michael J. Pan. 2005. Knowledge representation in the semantic web for Earth and environmental terminology (SWEET). Computers & Geosciences 31 (9) (novembre): 1119-1125. doi:10.1016/j.cageo.2004.12.004.

Renear, A. H., et C. L. Palmer. 2009. Strategic Reading, Ontologies, and the Future of Scientific Publishing . Science 325 (5942) (août 13): 828-832. doi:10.1126/science.1157784.

Russomanno, David J., Cartik R. Kothari, et Omoju A. Thomas. 2005. Building a Sensor Ontology: A Practical Approach Leveraging . In ISO and OGC Models, The 2005 International Conference on Artificial Intelligence, Las Vegas, NV, 2005, 637–643. Press.

Smith, Barry, Michael Ashburner, Cornelius Rosse, Jonathan Bard, William Bug, Werner Ceusters, Louis J Goldberg, et al. 2007. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration . Nature Biotechnology 25 (11) (novembre): 1251-1255. doi:10.1038/nbt1346.

Siegel M.R. 1961. Statistics, New York: McGrow-Hill.

St Luis, V., Pidgeon A. M., Radeloff V.C., Hawbaker T.J., Clayton M.K.. (2006). High resolution image texture as a predictor of bird species richness Remote Sensing of Environment 105, 299-312

Tarantino C., Blonda P., Pasquariello G., (2007) Remote sensed data for automatic detection of land use changes due to human activity in support to landslide studies, Natural Hazards, vol. 41 pp. 245-267.

Timpf, Sabine. 2002. The need for task ontologies in interoperable GIS. Zurich: University of Zurich, Department of Geography. The need for task ontologies in interoperable GIS.

Tripathi, Ajay, et Hassan A. Babaie. 2008. Developing a modular hydrogeology ontology by extending the SWEET upper-level ontologies . Computers & Geosciences 34 (9) (septembre): 1022-1033. doi:10.1016/j.cageo.2007.08.009.

Vanegas, M.C., Bloch, I. & Inglada, J. 2009, "Fuzzy spatial relations for high resolution remote sensing image analysis: The case of "to go across"", *International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. IV773.

Visser, U., H. Stuckenschmidt, G. Schuster, et T. Vögele. 2002. Ontologies for geographic information processing . Computers & Geosciences 28 (1) (février): 103-117. doi:10.1016/S0098-3004(01)00019-X

8 Abbreviations and Acronyms

ETM	Landsat 7 Enhanced Thematic Mapper
GHCs	General Habitat Categories
GIS	Geographic Information System
GMES	Global Monitoring for the Environment and Security
HR	High Resolution
IDL	Interactive Data Language
LC	Land Cover
LCC	Land Cover Change
MR	Medium Resolution
MS	Multi-Spectral
OBIA	Object Based Image Analysis
PAR	Photosynthetically Active Radiation
RS imagery	Remote Sensed imagery
RS-IUS	Remote Sensing Image Understanding System
SAR	Synthetic Aperture Radar
SCI	Site of Community Importance
SDD	Service Design Document
SPA	Special Protection Area
SPOT	Satellite Pour l'Observation de la Terre
SR	Spatial Resolution
TM	Thematic Mapper
TOA	Top of Atmosphere
TOARF	Top of Atmosphere REFlectance
VHR	Very High Resolution