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: 5.4

Hyper-spectral Remote Sensing

Due date of deliverable:	29-06-2012
Actual submission date:	12 July 2012
Version:	V1
Main Authors:	Richard Lucas (P11), Brian O'Connor (P11), Jordi Inglada (P16), Sander Mucher (P4), Laure Roupioz, Harini Nagendra (P5), Maria Adamo(P1), Cristina Tarantino (P1), Palma Blonda (P1)

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

Deliverable title	Hyper-spectral Remote Sensing
Deliverable number	D5.4
Deliverable version	V1
Previous version(s)	0
Contractual date of delivery	1 July, 2012
Actual date of delivery	12 July, 2012
Deliverable filename	D5.4
Nature of deliverable	R = Report
Dissemination level	PU
Number of pages	63
Workpackage	WP5
Partner responsible	P11
Author(s)	Richard Lucas (P11), Brian O'Connor (P11), Jordi Inglada (P16), Sander Mucher (P4), Laure Roupioz, Harini Nagendra (P5), Maria Adamo (P1), Cristina Tarantino (P1), Palma Blonda (P1)
Editor	Richard Lucas
EC Project Officer	Florence Beroud

Abstract	Evaluates the role of hyper-spectral data within the EODHaM system.
Keywords	VHR, land cover maps, habitats, semantic nets

Signatures

Editor	**Responsibility-Company**	**Date**	**Signature**
Richard Lucas	Editor (P11)	11-12-2012	
Verified by			
Laurent Durieux	WP5 Leader (P12)	11-12-2012	
Nagendra Harini (P5)		11-12-2012	
Approved by			
Palma Blonda	Coordinator (P1)	12-12-2012	
Fifamè Koudogbo	QAP (P7)	11-12-2012	

Table of Content

Executive Summary

Deliverable 5.4 provides an overview of the potential of hyper-spectral imagery for enhancing the classification of land covers and habitats for Natura 2000 sites and surrounds. The Deliverable builds on the use of single and multi-date high and very high-resolution (VHR) satellite imagery (D5.3) for progressively classifying land covers, with these defined according to the Food and Agricultural Organisation (FAO) Land Cover Classification Scheme (LCCS). In implementing the classification, hyper-spectral data were considered to be useful for providing additional spectral bands, indices and endmember fractions (e.g.. photosynthetic. non-photosynthetic and moisture) for enhancing discrimination of species or genera, either dominating communities or being indicator or rare components, and hence classification of land covers and habitats. A suite of algorithms unique to the classification of hyper-spectral data are reviewed, with reference made to their potential for better discrimination of land covers or habitats. Additionally, hyper-spectral data were also determined to provide strong potential for delineating certain important habitat features that can be used as inputs into the rule-based EODHaM classification (e.g., tree crowns). Focusing on Cors Fochno in Wales, classifications undertaken using hyper-spectral and multi-temporal multi-spectral data are compared, with results suggesting that the latter (if acquired at times of maximum senescence and leaf cover and in transition periods) can provide similar or better discrimination of land covers/habitats. Whilst hyper-spectral data are beneficial for habitat classification, these are often expensive and logistically more complex to acquire compared to spaceborne optical data and processing of data is compromised by large data volumes and atmospheric illumination and bidirectional effects. The study therefore recommends targeted use of hyper-spectral data for identifying species that are indicators of habitat condition, rare or endangered or of limited extent, or invasive or exotic. Procedures for integrating these data within the EODHaM system are conveyed.

1 Introduction

Within the EODHaM 1st, 2nd and 3rd stages outlined in the Deliverable (D5.3), the focus has been on the use of Very High Resolution (VHR) and HR spaceborne optical sensors for segmenting the landscapes within and surrounding Natura 2000 sites, classifying land covers according to the Food and Agricultural Organisation (FAO) Land Cover Classification Scheme (LCCS) and translating these to (plant and non-plant) life forms and associated General Habitat Categories (GHCs). Within D5.3, significant benefits were obtained by integrating VHR and HR data acquired on at least two dates with these timed to coincide with periods of peak flush (wet season or summer) or senescence (dry season or winter)(D4.4). The inclusion of data acquired in the transitioning periods also facilitated better discrimination of land covers and habitats dominated by individual species, genera or combinations of these.

For Deliverable 5.4 (D5.4), the potential of using hyper-spectral data within the EODHaM system has been evaluated, both through a review of literature and practical application. Specific focus has been on investigating the use of these data for:

- Generating indices and retrieving endmember fractions to assist the description and discrimination of plant communities and contained species.
- Classifying land covers and habitats using algorithms specific to hyper-spectral data.
- Improving segmentation of features (e.g.. tree crowns) that can provide important inputs into the rule-based classification for subsequent stages of EODHaM (habitat classification and monitoring of pressures).

To achieve this aim, a number of sub-objectives were defined and are listed below:

- To define indices and extract endmembers from hyper-spectral data that can be used primarily to describe life forms (e.g. in terms of biophysical characteristics, dominant species) but also non-life forms (e.g. surface materials, water quality) of the LCCS.
- To use spectral unmixing techniques to generate maps of biophysical properties or species distributions, which can be used subsequently to provide input to the rule-based classification of LCCS categories, give more detailed descriptions of the distribution of GHCs and/or identify plant species and/or communities, including those regarded as indicators.
- To practically evaluate the relative benefits of using single date hyper-spectral and multi-date VHR data within the EODHAM 1st and 2nd stages.
- To critically review the role of hyper-spectral data within the EODHaM system.

The Deliverable is structured as follows. Section 2 reviews hyper-spectral data and their use in support of the classification of land covers/habitats, including through segmentation of the landscape and extraction of features. Key indices and endmember fractions that can be used for the description and discrimination of plants and classification techniques specific to hyperspectral data are described. Section 3 then provides an overview of the airborne and also spaceborne hyper-spectral data available for the BIOSOS sites in Wales, Italy and the Netherlands, with specific details on the sensors used provided in Appendix I. Section 4 documents approaches to the classification of land covers/habitats for selected BIOSOS sites, as case studies, based on the review conducted in Section 2. Section 5 provides examples of how hyper-spectral data have been used for the classification of land covers and associated habitats in the selected case study locations, and provides a comparison of classification outputs generated using single-date hyper-spectral and both single- and multi-date multi-spectral data. Section 6 provides a critique of the use of hyper-spectral data, focusing specifically on the cost-benefits of acquiring and processing hyper-spectral over the use of multi-spectral (single-date and multi-date) data. Section 7 provides recommendations as to the inclusion of hyper-spectral data within the EODHaM system.

2 Review of hyper-spectral data

2.1 Hyper-spectral versus multi-spectral

The EODHaM system is designed primarily to use multi-spectral VHR or HR data to provide classifications of land covers (based specifically on the FAO LCCS taxonomy), which can be translated subsequently to habitat categories (i.e., GHCs). A particular advantage of using spaceborne over airborne data is that greater coverage in one data-take and consistency in the imagery across the landscape (e.g. in terms of reflectance characteristics) are provided, particularly as image strips from airborne data often need to be combined. For some habitats, such as those dominated by forbs (e.g., bracken; *Pteridium aquilinum*) or aquatic species (e.g., *Phragmites australis*) or which contain a diverse mix of species (e.g., active bogs with a mix of shrubs, forbs, graminoids and lichens), information on the distribution and/ relative amounts (e.g., in terms of cover) of particular plant species is required. For non-life forms, spatial information on surface materials (e.g., sand, mud) and water characteristics (e.g., sediment load) is needed. When single-date multi-spectral data (typically visible blue, green and red and near infrared; NIR) are used for this purpose, capacity for discrimination is hampered because of correlations between bands, the limited dynamic range (quantization levels) and the broad spectral range in the channels provided. The launch of the Worldview-2 (WV-2) with eight spectral bands (including in the coastal, yellow and red edge as well as a second (overlapping) NIR channel; Table 2.1) has provided new opportunities for discrimination of land covers/habitats, hence its preferred adoption with the EODHaM system. Nevertheless, single-date imagery is still limited for the level of classification detail required.

Table 2.1: WV-2 wavelength regions

Bands	Lower Edge (nm)	Upper Edge (nm)
Coastal	400	450
Blue	450	510
Green	510	580
Yellow edge	585	625
Red	630	690
Red edge	705	745
NIR1	710	895
NIR2	860	1040

Hyper-spectral sensors offer finer spectral measurements than multi-spectral instruments, with often hundreds of spectral bands of narrow (e.g.. 0.1 nm) width allowing a near continuous spectrum to be generated for each pixel. This presents opportunities for more precise identification of surface materials compared to when broadband multispectral sensors are used. In the spectral region, these sensors can record reflectance from the visible blue through to the shortwave infrared (SWIR) wavelength regions. For example, APEX is an advanced scientific instrument for the European remote sensing community, recording hyper-spectral data in approximately 300 bands in the wavelength range between 400 nm and 2500 nm and at a spatial ground resolution of 2 m to 5 m. Others include the HYMAP and Airborne Visible Infrared Imaging Spectrometry (AVIRIS) and the Multi-spectral Infrared and Visible Imaging Spectrometer (MIVIS), with the latter extending into the thermal infrared. However, many sensors record only in the visible and NIR (VHIR; e.g. the Compact Airborne Spectrographic Imager (CASI) or EAGLE) or SWIR wavelength regions (e.g. the HAWK).

Whilst airborne sensors have acquired the majority of hyper-spectral data, spaceborne sensors have also operated over the past decade (e.g., HYPERION and CHRIS PROBA). As an example, Christian and Krishnayya (2007) used HYPERION data for land cover studies in India. Future sensors are also being constructed for launch in the next five years. Italy's ASI space agency plans to launch Prisma, a medium-resolution hyper-spectral imaging mission, in 2012. Prisma's hyper-spectral camera will be able to acquire images in about 235 channels in the VNIR and SWIR wavelength regions. The German Aerospace Center (DLR) and the German Research Centre for Geosciences (GFZ) are planning to launch the EnMAP hyper-spectral satellite in 2014 to map the Earth's surface in over 200 narrow

wavebands. EnMAP is designed to record bio-physical, biochemical and geo-chemical variables to increase understanding of biospheric/geospheric processes (http://www.esa.int). In 2015, NASA plans to launch the HyspIRI mission, which will acquire imagery across 210 spectral bands, with focus on providing information on ecosystems, including the nutrient and water status of vegetation. Thus, spaceborne data are going to become increasingly available by the end of the decade. In an endeavour to be forward-planning, BIO_SOS is evaluating (within this deliverable) the use of hyper-spectral data for development of the EODHam system, as this is anticipated to become an important issue for the remote sensing and conservation community in the coming years.

2.2 Information content of hyper-spectral data.

2.2.1 Spectral reflectance

The primary data provided by hyper-spectral sensors are spectra, with the magnitude (primarily reflectance) and shape used commonly to describe and discriminate land surfaces. An example of a typical spectrum for photosynthetic (green) vegetation is given in Figure 2.1, but characteristic spectra relevant to land cover and habitat mapping within BIOSOS are also available for non-photosynthetic (brown) vegetation, soils, water (in liquid and frozen form), bare areas and urban surfaces.

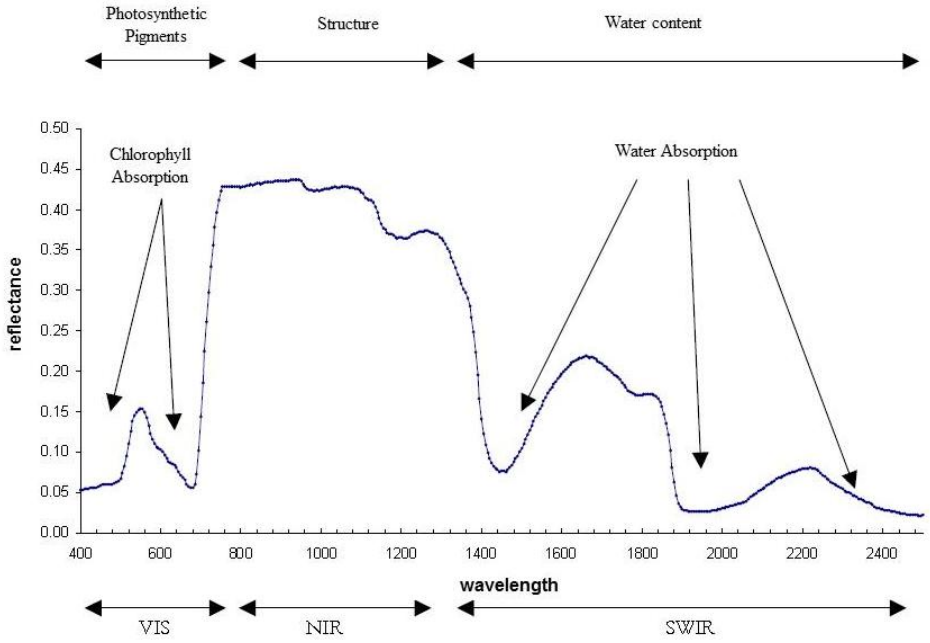


Figure 2.1. Typical spectra for vegetation highlighting the main contributors to reflectance.

Within the vegetation (photosynthetic) spectra, characteristic features include the green peak, red trough, red edge and NIR plateau with absorption features (relating to moisture content) evident in the latter and also in the SWIR wavelength regions. Reflectance in the visible regions is primarily a function of pigment concentrations in foliage whilst in the NIR and SWIR, the internal leaf structure and moisture content of the leaves respectively influence reflectance (Swain and Davis, 1978). In all cases, it should be noted that the reflectance of vegetation canopies varies from that of individual components (e.g., leaves, branches); largely because of the different contributions to the reflectance from plant materials and also the underlying surface and shadowing as a function of canopy heterogeneity, which particularly influences the NIR and SWIR wavelength regions.

The loss of pigments, cell structure and moisture content during senescence of leaves leads to the loss of most of the characteristic features of green leaves (with the exception of the water absorption features) and the transition to the spectral curve typical of non-photosynthetic vegetation. Dead wood

exhibits similar spectra to that of dead leaves and, as these components ultimately develop into soil, the spectra of the latter often contain similar spectral features when derived from vegetation (e.g., humic soils). When soils are derived from rock weathering, their reflectance contains elements of the contained minerals, which are often distinct and thereby allow the generation of reference spectral libraries. Materials derived from soils or rocks (e.g., slates, clay tiles) are used for building materials and hence can exhibit a similar reflectance to these components of the landscape, hence leading to confusion when classifying. As an example, Figure 2.2 shows reflectance spectra from a range of materials typical to urban areas.

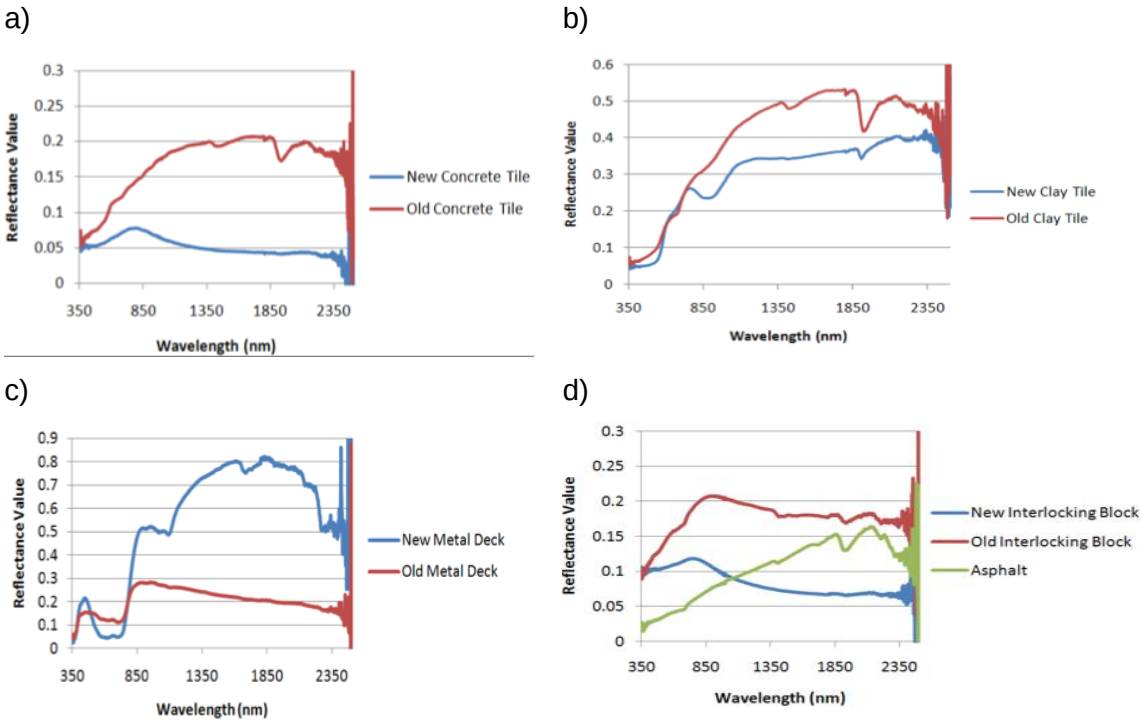


Figure 2.2. Example spectra for a) concrete and b) clay tiles, c) metal structures and d) road surfaces(Nasarudin and Shafri, 2011).

The reflectance of water in the visible regions is typically < 10-15 % depending on, for example, sediment loads and the characteristics of the bed in shallow waters, but is lower (typically close to zero) in the NIR and SWIR regions and hence the more moisture contained within materials, the lower their reflectance. Shadow exhibits a very similar reflectance to water and hence is often confused spectrally but can also be used interchangeably in some cases (e.g., in spectral unmixing). When water is in a frozen state (i.e.. as snow or ice), the reflectance is considerably higher although then varies as a function of surface contaminants in the visible regions (e.g.. dust) or grain size, with lower NIR reflectances associated with larger grain sizes because of the associated reduction in the amount of air/water interfaces. Hence, the reflectance of snow in the visible and near infrared regions can indicate condition. In the shortwave infrared regions, snow and ice have very low reflectance allowing differentiation from cloud. Through knowledge of the reflectance characteristics of surfaces, an insight into the use of spectra for the description and discrimination of surfaces can be obtained and used to assist the classification process.

A large number of studies have used spectral reflectance data to differentiate plant species and communities on the basis of differences in spectral reflectance, with this being attributed largely to differences in foliar chemistry, the internal structure of leaves, moisture content and the overall canopy structure (e.g., in terms of shadowing and relative amounts of plant components; leaves, branches). As examples, Lucas *et al.* (2008) extracted reflectance spectra (based on CASI data) from the sunlit

portions of delineated tree crowns in Australia savannas, discriminating species of *Callitris*, *Eucalyptus*, *Acacia* and *Angophora* through discriminant analysis. Lu *et. al.* (2009) used hyper-spectral date to map the distribution of two spectrally similar grasses (*Miscanthus sacchariflorus* and *Phragmites australis)* in Japan on the basis of subtle differences in canopy density, leaf and canopy structure as well as biochemical properties. The benefits of using hyper-spectral data for mapping aquatic vegetation (e.g., different species of *Spartina* in San Francisco Bay, USA; Rosso *et al..* 2005), identifying and mapping invasive species(e.g.Ustin et al., 2004, Hestir et al., 2008, Walsh et al., 2008, He et al., 2011),; and differentiating between trees of the same species that are of different ages and sizes have also been conveyed (Christian and Krishnayya, 2009).

2.2.2 Indices

As outlined in D5.3, a wide range of indices can be derived from broadband multi-spectral data in order to characterize and classify land surfaces. However, a limitation of these sensors for index calculation, and particularly those acquiring at VHR from space, is that only VNIR data are provided. Hence, the range of indices that can be derived is limited, particularly those that require observations in the SWIR wavelength regions, with these being especially critical for the mapping and discrimination of some plant communities. Furthermore, the wavebands used for calculating these indices have often been identified with reference to field, laboratory or hyper-spectral data and are often quite specific. Hence, their translation to 'similar' wavelength regions observed by multi-spectral data may compromise their use and reliability. Examples of indices that have been used in D5.3 for discriminating and mapping land covers are provided in Table 2.2. Here, the more specific wavebands used in calculations with hyper-spectral data are provided, together with their equivalents (if available) from Quickbird, WV-2 and Landsat (as an examples of a sensor that observes in the SWIR wavelength regions).

The narrower and greater number of wavebands provided by hyper-spectral sensors allows calculation of a wider range of indices that can potentially facilitate better discrimination of land covers but also precision measurements for retrieval of biophysical properties and discrimination of surface materials or plant species. In the latter case, plant species or individuals of the same species in different stages of growth can be discriminated because the biophysical properties (e.g., pigment type and content, internal structure of leaves, moisture content) differ by an amount sufficient to change the reflectance characteristics, and hence derived indices. It should be noted that discrimination of individual plants or groups of individuals or species depends critically on the spatial scale of the images with respect to the size of the units (plants or communities) being discriminated. Where pixels resolutions are larger than the size of the communities being observed, spectral contributions from a wide range of surfaces can occur leading to mixed pixels, spectral confusion and limited separability.

In general, indices can be grouped into three broad categories depending upon their use and formulation, with examples given in Table 2.3.

a) Indices that correlate with structural variables measured in the field and describing, for example, vegetation canopies (e.g., leaf area index (LAI) and canopy cover) or water (e.g. snow) surfaces.
b) Indices that are relate to the chemical composition or structure of components of materials (e.g., leaves or branches in the case of vegetation or sediments or minerals in water and soils respectively). For example, specific foliar chemicals that can be retrieved include chlorophyll a and b and carotenoids (Gitelson *et al.*, 1996).
c) Indices based on the shape of the reflectance spectra. For example, the Red Edge Position (REP); (Curran *et al.*, 1995, Clevers and Buker, 1991, Clevers and Kooistra, 2012)) considers the point of maximum slope in the reflectance spectra of leaves.

Whilst in the case of a), such relationships can and have been established with multi-spectral data, b) and particularly c) require the use of hyper-spectral data for their calculation. In each case, atmospheric effects (scattering and absorption) as well as background (in the case of vegetation; e.g., olive trees over a soil background) can compromise the values obtained (Huete *et al*., 2002).

Table 2.2 Examples of indices used to characterise a range of surfaces and their utility with a range of sensors.

Index	Hyper-spectral		VHR (RGB) [1]		VHR (WV-2) [2]		Multispectral [3]
Normalised Difference Vegetation Index (NDVI)	$\frac{\rho_{800}-\rho_{670}}{\rho_{800}+\rho_{670}}$	----->	$\frac{NIR-Red}{NIR+Red}$	----->	$\frac{NIR1-Red}{NIR1+Red}$	----->	$\frac{NIR-Red}{NIR+Red}$
Plant Senesence Reflectance Index (PSRI)	$\frac{\rho_{680}-\rho_{500}}{\rho_{750}}$		----->		$\frac{Red-Blue}{Red\ Edge}$		
Water Band Index (WBI)	$\frac{\rho_{900}}{\rho_{970}}$		----->		$\frac{NIR1}{NIR2}$		
Normalised Difference Water Index (NDWI)	$\frac{\rho_{858}-\rho_{1240}}{\rho_{858}+\rho_{1240}}$			----->			$\frac{NIR-SWIR}{NIR+SWIR}$
Normalised Difference Snow Index (NDSI)	$\frac{\rho_{550}-\rho_{1650}}{\rho_{550}+\rho_{1650}}$			----->			$\frac{Green-SWIR}{Green+SWIR}$

[1]Quickbird (Blue (450-520 nm), Green (520-600 nm), Red (630-690 nm), NIR (760-900 nm)
[2]Worldview (Coastal (400-450 nm), Blue (450-510 nm), Green (510-580), Yellow-edge (585-625 nm), Red (630-690), Red-edge (705-745 nm), NIR (1) (770-895 nm), NIR (2) (860-1040 nm)
[3]Blue (450-515 nm), Green (525-605 nm), Red (630-690 nm), NIR (775-900 nm), SWIR1 (1550-1750 nm), SWIR2 (2090-2350 nm).

A useful concept to consider is explained in Equation 1 where:

$$NDWI_{corr} = NDWI\left(1 - \frac{\rho_{SWIR} - \rho_{SWIR_{min}}}{\rho_{SWIR_{max}} - \rho_{SWIR_{min}}}\right) \qquad \text{Equation (1).}$$

The NDWI is the Normalized Difference Water Index and $\rho SWIR$ is the reflectance recorded in a spectral band of the SWIR region (e.g. centred at 2226 nm) for a considered pixel. $\rho SWIR_{min}$ is the minimum reflectance in the SWIR associated with completely closed canopies and $\rho SWIR_{max}$ the reflectance associated with open canopies. Using this measure, the contribution of vegetation beneath the canopy on the signal is reduced so the index value relates almost exclusively to the spectral response of the tree cover. Hence, background effects can be compensated for in some cases.

The absence of the SWIR channel from many airborne hyperspectral and all currently operating VHR multi-spectral data is currently a major limitation for land cover and habitat mapping. The inclusion of a SWIR channel of the forthcoming WV-3 sensor is, however, anticipated to partly fill this gap. The primary benefits of the SWIR channel are:

a) Better provision of information on the moisture content of surfaces. The SWIR channels with greatest sensitivity to surface moisture content are the 1450, 1950 and, to a lesser extent, the 2200 nm bands, with these corresponding to primary water absorption features (Carter, 1991)..

b) Better characterisation, discrimination and classification of non-photosynthetic components of vegetation. The absorption bands of dry vegetation are centred at 1720 and 1780 nm and are sensitive to chemical compounds such as cellulose, lignin and starch (Curran, 1989).
c) Enhanced discrimination of snow/ice cover and cloud, with the former exhibiting a low reflectance equivalent to open water in the SWIR channels.

In each case, the inclusion of the SWIR channels (whether from multi-spectral or hyper-spectral sensors) would improve the operations of the EODHaM system. The added benefits of a SWIR band are discussed further in Section 5.5 in relation to the MIVIS sensor. However, Table 2.3 includes some MIVIS-based vegetation indices related to water content in addition to two simple band ratio indices (i.e. Index1 and Index2) related to the peaks of absorption due to water content at 1450 and 1950 nm.

Whilst a high diversity of indices have been developed based on field, laboratory and image spectra, their adaption for use with multi-spectral data, as explained in D5.3, needs to be carefully considered, with this based on an understanding of the reasons why certain wavelength regions are used and combined and ensuring that the biophysical meaning of the index is not compromised by, for example, simple translation of the wavebands used with hyper-spectral data to the ‘nearest’ multi-spectral channel.

Whilst hyper-spectral imagery are anticipated to provide more specific information for describing surface materials, particularly if SWIR data are included, the level of spectral detail provided is likely to be unnecessary for the classification of most broader level LCCS categories, although better definition of class boundaries, whether sharp or gradual (as in the case of vegetation mosaics), and contained habitats and species, might be achieved.

Table 2.3 Examples of hyper spectral indices and their applications (including those with MIVIS data)

Index	Description	Equation	Source
Red Edge NDVI	Red edge position and senescence	$\frac{\rho_{750}-\rho_{705}}{\rho_{750}+\rho_{705}}$	(Gitelson and Merzlyak, 1994)
Red Edge Position Index	Red edge position and chlorophyll	Wavelength of steepest slope within the range $\rho_{690-740}$	Curran *et al*.. 1995
Modified Red Edge SR Index	Red edge position and chlorophyll	$\frac{\rho_{750}-\rho_{445}}{\rho_{705}-\rho_{445}}$	(Datt, 1999, Sims and Gamon, 2002)
Photochemical Reflectance Index	Light use efficiency, stress and productivity	$\frac{\rho_{531}-\rho_{570}}{\rho_{531}+\rho_{570}}$	(Gamon *et al.*, 1997, Gamon *et al*., 1992)
Vogelmann Red Edge Index 1 (Vog1)	Chlorophyll	$\frac{\rho_{740}}{\rho_{720}}$	Vogelmann *et al*. (1993)
GMI	Chlorophyll	$\frac{\rho_{750}}{\rho_{555}}$	Gitelson and Merzlyak (1994)
Carter Index (Ctr)	Chlorophyll	$\frac{\rho_{710}}{\rho_{760}}$	(Carter, 1994)
Double Difference (DD)	Chlorophyll	$(\rho_{749}-\rho_{720})-(\rho_{701}-\rho_{672})$	Le Maire *et al*. (2004)
Simple Ratio Water Index (SRWI)	Moisture content	$\frac{\rho_{858}}{\rho_{1240}}$	Zarco-Tejada *et al*. (2003)
Water Spectral Index (WSI)	Moisture content	$\frac{\rho_{820}-\rho_{1450}}{\rho_{820}+\rho_{1450}}$	
Relative Depth Index (RDI)	Water depth and turbidity	$\frac{\rho_{1116}-\rho_{1150}}{\rho_{1116}}$	(Rollin and Milton, 1998)
Moisture Stress Index (MS)	Moisture content	$\frac{\rho_{1600}}{\rho_{820}}$	$\frac{b_{21}-b_{23}}{b_{21}}$ (Hunt Jr and Rock, 1989)
Normalized Difference Water Index (NDWI)	Presence or absence of water on the surface	$\frac{\rho_{858}-\rho_{1240}}{\rho_{858}+\rho_{1240}}$	$\frac{b_{19}-b_{23}}{b_{19}+b_{23}}$ [1]
SWIR Ratio (1)		$\frac{\rho_{1270}}{\rho_{1375}}$	$\frac{b_{23}}{b_{25}}$
SWIR Ratio (2)		$\frac{\rho_{2226}}{\rho_{2000}}$	$\frac{b_{58}}{b_{31}}$

[1] MIVIS bands

2.2.3 Endmember fractions

The spectra recorded by a remote sensing instrument are typically an amalgamation of the response from multiple surfaces within the sensor's field of view (Dennison and Roberts, 2003). The mix can, however, be considered as the contribution of 'pure' spectra from these surfaces (e.g., snow and rock), which are often referred to as 'endmembers'. These endmembers are typically defined with reference to field spectroradiometer measurements but can also be derived from areas known to be 'pure' in the image. Once obtained, these can be used with spectral unmixing techniques to obtain fractional abundances of materials, which can often be related to the biophysical composition of surfaces and also used to improve the classification of images.

A wide variety of endmembers exist, but two types are particularly relevant to the BIOSOS objectives:

a) Endmembers that can be associated with the biophysical characteristics of materials, namely photosynthetic (green) vegetation (PV), non-photosynthetic (brown) vegetation (NPV); (Roberts *et al.*, 2003)), water in various states (liquid or frozen) or minerals within soils and exposed rock(Mustard, 1993).

b) Endmembers linked to the categories outlined in a) but which are associated more specifically with particular plant species or surface materials. For example, endmembers can be defined with reference to different plant species (e.g., *Sphagnum*) or, in the case of NPV, the endmember can be associated with dead leaves or wood, either standing or as litter (Guerschman *et al.,* 2009), and used subsequently to differentiate species or communities (e.g., those dominated by *Molinia caerulea*, which is senescent in winter, although highly productive in summer).

In each case, endmembers are typically associated with, and often defined with reference to, extreme values in the feature space. Endmembers for common surfaces (i.e., PV, NPV, water, soil, snow) outlined in a) can be extracted from both multi-spectral and hyper-spectral data, although this depends upon the number of wavebands available, which can be limiting. A particular advantage of hyper-spectral data though is that, as there are more wavebands, the number and type of endmember spectra can be increased providing increased capability for discrimination of particular plant species, water chemistries and minerals in soils or geologies. Assuming rigorous calibration and atmospheric correction of image data, endmember spectra extracted from field or laboratory measurements (e.g.. from ground soil samples) can be used in spectral unmixing procedures to estimate sub-pixel proportions or 'hard' classifications (e.g., using algorithms such as the Spectral Angle Mapper or SAM).

The endmember spectra for water is similar to that of shade, which represents areas of decreased illumination and shadow and is therefore not a material component of the image (Nichol and Wong, 2007). Nevertheless, its presence in an image can deter from the actual determination of other cover fractions. Hence, shade is frequently included as an endmember within spectral unmixing procedures to account for variations in illumination (Hamada *et al*., 2011) or, for example, to differentiate forest from other canopies because of the presence of more shade in the former.

For the classification of land covers, proportions of endmember fractions obtained from optical data have been used for the discrimination of regrowth stages in Amazonia (Adams *et al*., 1995) and semi-natural habitats in Wales, with the latter based primarily on differences in the photosynthetic component of vegetation in the spring and summer periods (Lucas *et al.*, 2011). Other studies have also used endmember fractions for quantifying the moisture content of leaves and landscapes (Colombo *et al*., 2008), discriminating minerals and mapping geologies as well as assessing water quality(Isidoro *et al.*, 2010). The use of endmembers in spectral unmixing procedures is also beneficial in areas with heterogeneous vegetation cover, when densities of vegetation are low and spatial patterns are scattered (He *et al*., 2011).

2.3 Hyper-spectral processing

A wide range of studies have focused on the use of hyper-spectral data for extracting sub-pixel proportions of biophysical properties of land surfaces, classifying land covers and habitats, and delineating features within the landscape, such as water bodies and tree crowns (Bunting and Lucas, 2006). Techniques that are specific for use with hyper-spectral data are often aimed at reducing the

dimensionality of the data, which is necessary because of the high degree of correlation between the large numbers of spectral bands. The following sections provide an overview of these techniques, focusing specifically on those that allow retrieval of sub-pixel proportions (e.g. PV, NPV) and classification of data, with some applied subsequently at the BIOSOS test sites.

2.3.1 Pre-processing

External factors such as land surface morphology, angle of view, geometry, lighting conditions and weather impact on the quality of airborne hyper-spectral data. A wide range of methods have been developed for correction. For example, atmospheric correction methods can be empirical (e.g., the Empirical Line Correction; using scene characteristics such as reflectance from variable or invariant objects or local meteorological data) or model-based (using atmospheric models such as ATCOR, FLAASH, HYCOR). Correction of these data is, nevertheless, a significant challenge without appropriate ground truth data to either develop procedures such as the Empirical Line Correction or to validate the output of atmospheric correction models.

Some studies conduct additional processing of data to remove white noise through, for instance, Principle Component Analyses, first removing the principal component mostly associated with uninformative data (noise), and then recombining the remaining components into a new set of composite bands which are then used as inputs for classification. Such methods are computationally intensive, and needs to be repeated for each flightline or image, but can be useful for enhancing the detection of surfaces. For example, Underwood *et al*. (2003) used the technique to detect the invasive highway ice plant (*Carpobrotus edulis*), which threatens the survival of endangered native scrub habitats in the US state of California.

2.3.2 Sub-pixel unmixing.

Sub-pixel unmixing techniques seek to extract the sub-pixel proportions of endmember spectra from optical data, including those listed in Section 2.3. Prominent amongst these are linear and non-linear spectral mixture analysis (SMA), Mixture Tuned Matched Filtering (MTFM), and multiple endmember spectral mixture analysis (MESMA). In each case, reference libraries of endmember spectra can be extracted from the image data themselves or from spectral measures in the field or laboratory.

2.3.2.1 Spectral Mixture Analysis (SMA)

Linear spectral mixture analysis is based on the assumption that the pixel spectra can be modelled from a linear combination of a number of endmember spectra of varying proportions, with multiple scattering ignored (Dennison and Roberts, 2003). As examples, SMA has been used to estimate sparse vegetation cover in arid environments within the Mojave National Preserve in south-eastern California where multiple endmembers derived from hyper-spectral data performed significantly better compared to when vegetation indices such as the Soil Adjusted Vegetation Index (SAVI) and Normalised Difference Vegetation Index (NDVI) were used (McGwire *et al*., 2000). Li *et al*. (2005) derived a nine-endmember set from AVIRIS data for a Californian saltmarsh in order to map the spatial distribution of *Salicornia virginica*, *Grindelia stricta* and *Spartina foliosa*. Kooistra *et al*. (2008)focused on the use of continuous fraction images derived from an SMA analysis of hyper-spectral AHS-160 imagery to assess the quality of habitat at a heathland site in the Netherlands and, in particular, to quantify grassland encroachment (outlined later). An accepted limitation of SMA is that the technique is compromised by variable reflectance (e.g., because of shadowing within a vegetation canopy or different viewing and illumination angles) from the same surface as only one reference spectra can be selected per endmember. Non-linear mixture models take into account multiple scattering between components in the pixel. Chen and Vierling (2006)provide a useful summary of SMA.

2.3.2.2 Matched Filtering and Mixture Tuned Matched Filtering (MTMF)

Matched Filtering (MF) focuses on maximizing and reducing the contributions of known endmembers and the background respectively, thereby providing a partial unmixing approach. Mixture Tuned Matched Filtering (MTMF) provides a score representing unfeasibility, with lower values indicating that the matched filter score is high and *vice versa*; each pixel is classified on the basis of the highest score. An example of the use of MTMF is provided by Parker Williams and Hunt Jr (2002).

2.3.2.3 Multiple Endmember Spectral Mixture Analysis (MESMA).

MESMA is an enhanced SMA which accounts for natural variation in the spectra of most materials on a sub-pixel level as it allows the number and types of endmembers to vary on a per-pixel basis (Roberts *et al*., 2007). MESMA is therefore a more powerful tool than SMA because it uses numerous mixture models of the various endmember spectra, determined iteratively for each pixel, rather than a linear combination of a few endmember spectra (Rosso *et al*., 2005).

As with SMA, MESMA is typically implemented by first generating a spectral library from which optimal endmembers are selected, with candidates derived either from the image itself using techniques such as the pixel purity index (PPI) or a library of known materials (reference endmembers), derived from field, laboratory or image spectra. Within VIPER tools, which were used to perform the MESMA for Cors Fochno in this study, spectral diversity is preserved by extracting the spectra from each pixel in a selected region of interest(Halligan, 2011). This allows for an assessment of the variability within a given ROI and informs on which outlier spectra to remove. The selection of optimal endmembers from the spectral library is guided by four methods:

a) The Constrained Reference Endmember Selection (CRES); (Roberts *et al*., 1993)
b) The Count-based Endmember Selection (COB); (Roberts *et al*., 2003).
c) The Endmember Average RMSE (EAR); (Dennison and Roberts, 2003)
d) The Minimum Average Spectral Angle (MASA); (Dennison *et al.*, 2004)

Following selection of the optimal endmembers per group (e.g. vegetation), the sub-pixel mixture modelling is constrained (e.g. minimum fit, fraction and residuals) . Each pixel within the image, whether hyper- or multi-spectral, is then unmixed using all possible combinations of endmembers, with multiple models tested within the parameter constraints. Through this approach, the number of endmembers does not limit the number of mixture models generated and hence there is greater opportunity to map a greater number of these across the image. As an example, Roberts *et al*. (2007)applied MESMA to AVIRIS data and, for landscapes in California, endmembers representing photosynthetic (green) vegetation (PV; 4), non-photosynthetic vegetation (NPV; 4) and other urban surfaces (38) were used to generate 608 models and a 4 endmember model of PV, NPV, other and shade.

2.3.2.4 OTB Toolbox algorithms for estimating endmembers

The OTB toolbox offers a number of algorithms for estimating the number of endmembers within an image, including the Eigenvalue Likelihood Maximisation algorithm(Bin and Chanussot, 2009), which uses the covariance and correlation matrices of the input data to estimate endmember numbers and a corresponding log-likelihood of occurrence, and the Virtual Dimensionality algorithm (Chein and Qian, 2004, Jing and Chein, 2006). Three families of unmixing algorithms are also available. The first (Family 1) requires, as a minimum, that one pure pixel is associated with each endmember, with these estimated for each pixel location using the Vertical Component Algorithm (VCA);(Nascimento and Dias, 2005). If non-pure endmembers occur, an estimate of the error associated with the spectra is generated. Family 2 is comprised of algorithms that seek a simplex of minimum volume circumscribing the data. A numerical optimization scheme then minimizes a function to estimate the endmembers, with this based on fully constrained, unconstrained and sparse approaches. The ISRA unmixing filter (Heinz et al., 1999)performs a fully least squares optimization on each pixel of the image, with the constraints being positivity (each component is positive) and additively (all components sum to 1). The unconstrained least square image filter solves the classical linear mixing model as a linear system of equations without constraints. The sparse approach is the most advanced and detects linear dependencies from N wavelength decompositions, performing a linear unmixing from a sparse representation of N signals (through multi-scale wavelet transforms) by means of line detection in the scatterplot. Family 3 algorithms are based on Non-negative Matrix Factorization (NMF) which factors a non-negative matrix (X) into a product of non-negative matrices (AS) by minimising a distance between X and AS with an adapted regularization to perform the degeneration in a manner adapted to the physical problems associated with unmixing. OTB proposes the MDMDNMF image filter, which implements unmixing based on MNF. This algorithm simultaneously finds the end members and abundance matrix that most closely represents the observed data(Theis et al., 2005). The algorithms of Families 1 and 2 estimate only the spectra of endmembers and the abundance maps are held *a posteri* and require application of

an algorithm such as the Fully Constrained Least Square (FCLS). Initialization of these algorithms is important as unmixing is a general non-convex problem.

2.3.3 Classification methods

Techniques for direct classification of hyper-spectral (e.g. into land cover and habitat categories or species types) include Support Vector Machines (SVM) and Spectral Angle Mapping (SAM).

2.3.3.1 Support Vector Machines (SVM)

SVM is a supervised non-parametric statistical learning technique (Mountrakis *et al*., 2011), that has provided, in many cases, higher levels of classification accuracy compared to more traditional techniques such as Maximum Likelihood Classification (Melgani and Bruzzone., 2004). SVMs make no assumption about the distribution of the input data and uses geometrical criteria to separate classes in hyper-dimensional feature space (Melgani and Bruzzone, 2004). As an example, these authors applied SVM to AVIRIS data acquired in the summer (June) over the Indian Pines area of Indiana, USA, concluding that it performed better than non-parametric classifiers (in this case, a kernel-based neural network and k-means algorithm) in terms of accuracy, computational time and robustness to parameter settings. In the Netherlands, land cover classifications based on SVM tree-structured (TS) Markov Random Field (MRF) have been performed for the BIOSOS test site (Figure 3.1). In the TS-MRF, a binary tree is constructed that exploits the hierarchical structure exhibited in the image, with the leaves of the tree corresponding to the end classes. With *k* end classes, the *k*-ary field reduces to a sequence of binary MRFs that are all configurable separately. Consequently, this technique allows many more degrees of freedom compared to conventional models that only use a single set of parameters estimated on the whole image irrespective of the difference in statistics encountered in different regions. The tree structure is selected because of its relatively simplicity and low computational burden. More complex MRF models are also allowed, as are diverse models relevant to specific nodes of the tree.

2.3.3.2 Spectral Angle Mapper (SAM)

The SAM technique uses a pixel spectrum, composed of unknown reflectance targets, represents these as separate vectors in *n*-dimensional feature space and compares them to a set of reference spectra. The angle between the reference vectors and the pixels vector is resolved before a threshold for the angle is defined. Spectra that exceed the threshold are not considered to match. The process is repeated until the constituent spectra of the pixel can be associated with a reference spectrum. A number of studies have applied SAM to hyper-spectral data. For example, Niemann *et al*. (2011) used SAM to detect insect infestations in conifer forests stands through comparison with spectra representing healthy trees. Overall accuracies of 70 % were achieved. Mucher and Kooistra (2011) also used SAM to identify different plant life forms at the pixel level and GHCs at the broader level. The accuracy of classification of life forms varied from 64 % to 78 %. A study in the Netherlands focused on evaluating the use of airborne HYMAP data for classifying GHCs based on the different life forms (Figure 3.2). For this purpose, a SAM classifier was applied to four sites of different complexity in different peat bogs (Haaksbergerveen and Korenburgerveen) and floodplains (Millingerwaard and Wageningen). The SAM classification was applied as it dealt with the entire spectral information delivered by the HYMAP (visible to SWIR) and was not affected by the albedo. The endmember spectra used for the classification were selected with reference to aerial photography and vegetation maps. The classification was assessed using standard error matrices, with this including the kappa-statistic per classified image. Whilst the accuracy in the classification of the life form generally exceeded 64 % and was typically > 70 % (Table 3.1), that of the GHCs was typically lower. This partly was a consequence of the scale of the classification, with the accuracy assessments for the SAM classification performed at the site level, whilst those of the GHC were applicable also to smaller patches. A better classification of the GHCs was obtained for larger patches, although this varied between life forms. The study concluded that the mapping success decreased with increases in biological complexity. This supports the conclusions of Andrew and Ustin(2008), who found an inverse relationship between spectral complexity and mapping success.

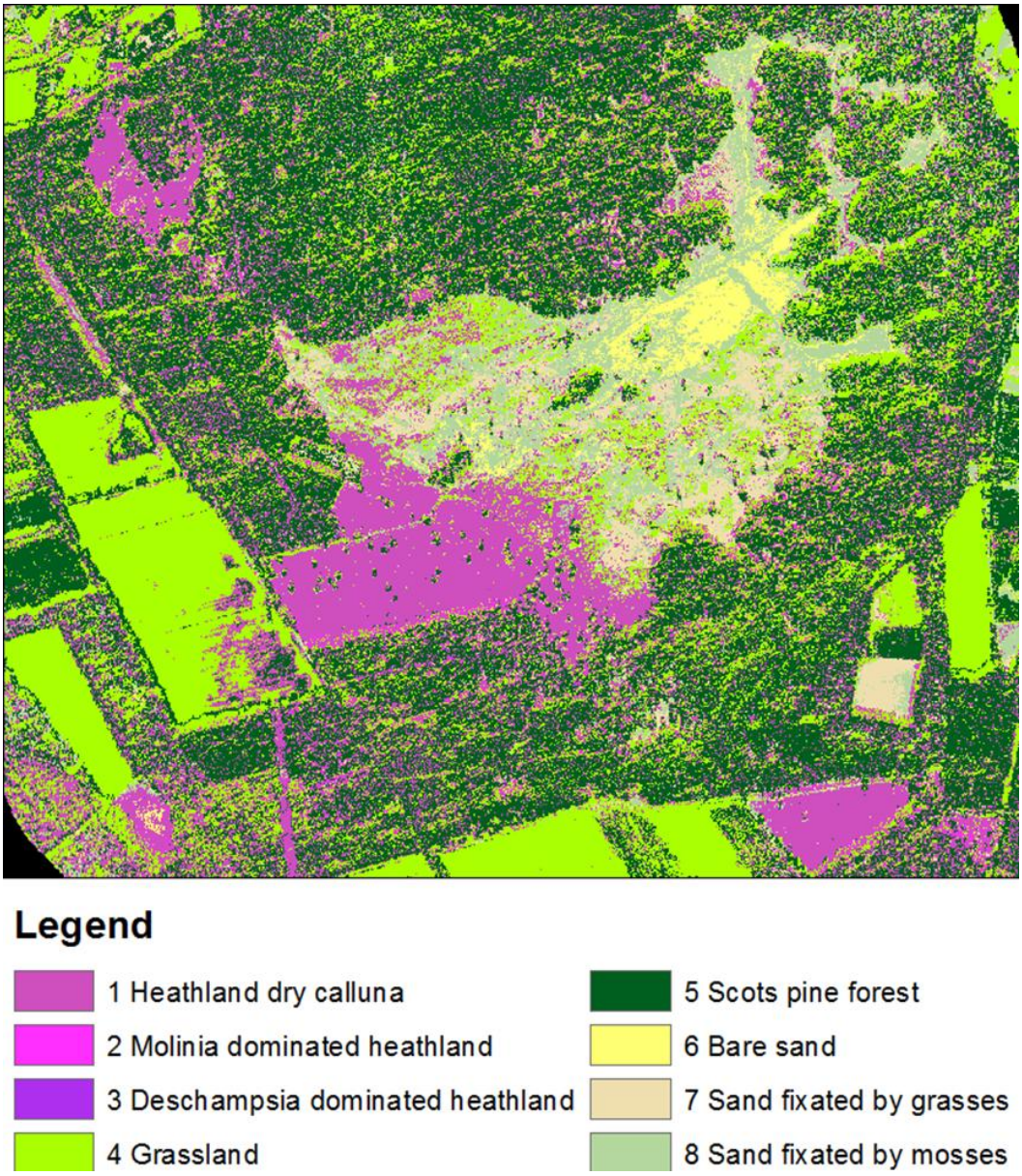


Figure 3.1 Classification of life forms and non-life forms using a SVM tree-structured (TS) Markov Random Field (MRF), the Dutch study area (source: Guy Thoonen, UA)

Table 3.1. Accuracies in the classification of life forms and GHCs, the Netherlands, from HYMAP data.

	Floodplains		Peat bogs	
Accuracy	**Wageningen**	**Millingerwaard**	**Korenburgerveen**	**Haaksbergerveen**
Lifeform (%)	64.0	75.0	77.8	71.2
GHC (%)	39.9	64.7	50.9	40.4

2.3.4 Feature extraction

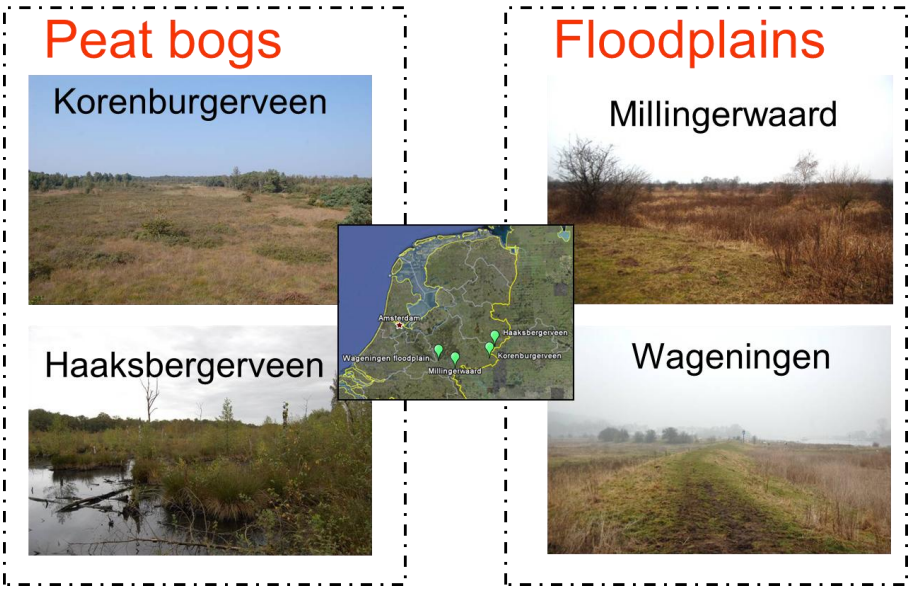
Whilst hyper-spectral data have largely been used for sub-pixel analysis and classification, a number of studies have used the enhanced spectral discrimination of surface features from hyper-spectral sensors to identify features within the landscape. For example, airborne CASI-2 hyper-spectral data acquired over a forested landscape near Injune, central east Queensland, Australia, were used to delineate tree crowns with an accuracy of 70 % (Bunting and Lucas, 2006). The tree crowns were also differentiated to species using linear discriminant analysis (Lucas *et al*., 2008). The characteristics of the tree crowns themselves have been resolved from CASI-2 data in south-eastern Australia, where Coops *et al*. (2004) assessed tree crown condition by measuring leaf and crown damage and discolouration characteristics using spectral indices in the red-green and red-edge region, with these derived from both leaf reflectance measurements and airborne imagery.

Morphological approaches that utilize spatial information in combination with spectral data can also be very useful for classification. For instance, Dell'Acqua *et al*. (2004) studied an urban setting in Pavia, Italy using 2.6 m airborne Digital Airborne Imaging Spectrometer (DAIS) imagery, with 80 data channels in the VNIR region. Classification was undertaken separately using a set of different spatial and spectral classifiers, after which the four best individual classifications were combined using a linear opinion pool, resulting in improvements in accuracy of 4.0-5.5 %. Plaza *et al*. (2005)utilize a more complicated approach, using concepts from mathematical morphology theory to characterize extended morphological transformations, based on simultaneous consideration of the spectral and spatial information available in airborne hyper-spectral imagery collected by the NASA JPL Airborne Visible-Infrared Imaging Spectrometer, and DAIS imagery. This approach was found to be especially reliable in landscapes with features that display regular differences in shape as well as orientation (e.g., urban landscapes), and thus may be useful for certain BIO_SOS landscapes such as those in in Italy where, for instance, olive groves have a specific pattern of orientation that can guide their identification (D 5.3). Such approaches have a high potential for exploitation, yet most classification techniques continue to base themselves solely on analysis of spectral information, without taking any account the two dimensional, spatial representation of data afforded by hyper-spectral images. There is, therefore, a need for further research in this direction (Schaepman *et al.*, 2009).

2.3.5 Anomaly detection

An anomaly in a scene is defined as an element that one does not expect to find. The unusual element is likely to be different from its environment and present as a minority component of a scene. Anomalous pixels are typically rare but sufficiently distinct, spectrally and/or spatially. Examples include a vehicle in a natural environment, a rock in a field or a wooden hut in a forest. Using hyper-spectral imagery, algorithms are available that allow the spectral response of the anomaly to be discriminated from that of the "background clutter". The algorithms are based on spectral and spatial hypotheses and described as either those based on target detection or anomaly detection. Target detection is where the spectral response of the target, which is known *a priori*, is used as input to the detection algorithm with examples being the Adaptive Matched Filter (AMF) or the Adaptive Cosine/Coherence Estimator (ACE). However, obtaining such *a priori* information is often difficult. Anomaly detection is where images are provided as input. An adaptive thresholding procedure is then used to generate a mask that is then compared to a map or ground truth data. The OTB using an anomaly detection approach (RX), which is based on models, statistics, hypotheses testing and moving windows (within a pre-defined local neighbourhood) and compares pixels (or sets of neighbouring pixels) against a model of background pixels. The anomaly detection algorithms can be used to assist mapping of features in the landscape (e.g., caravans) or identify land covers, habitats or species that are rare (of limited extent) within a scene.

a)



b)

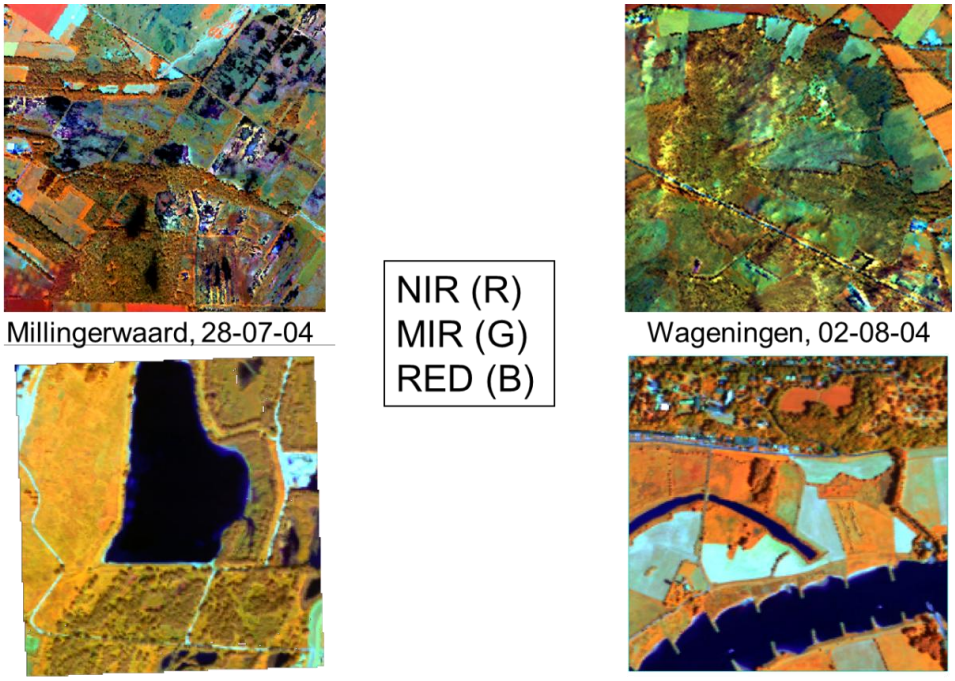


Figure 3.2. a) Locations and characteristics of four sites in the Netherlands. with these representing wetlands and peat bogs, b) and their manifestation within HYMAP data.

2.3.6 Direct correlations with plant structure and species composition

A small but increasing number of studies are beginning to explore the direct association of hyperspectral bands and derived indices with species composition and plant structure. For example, Oldeland *et al*. (2012)used HyMap and CHRIS-PROBA data to assess plant traits and species composition in a dwarf shrub savannah habitat in Namibia. Plant functional traits, such as growth form, height, leaf area and seed length, were recorded from field plots, and related with vegetation indices derived from the VNIR and SWIR region using RLQ analysis. This approach relates a matrix of environmental (in this case, spectral) variables by site to a matrix of species by traits, using a species-by-sites matrix as a link. They also evaluated another approach, which used the fourth-corner statistic to quantify the relationship between spectral indices and species traits. This approach enabled the association of a matrix of plant

functional traits with habitat characteristics (in this case, spectral indices) of the sites where sampling was conducted, given the distribution (here, abundance) of species across different sampling sites (Dray and Legendre, 2008). Spectral indices derived from HyMap were more related with vegetation traits linked to vegetation structure, while indices derived from CHRIS PROBA were more related with vegetation traits linked to canopy cover. Oldeland *et al*. (2012) attributed this to the finer spatial resolution of HyMap.

2.4 Overview

Within the BIOSOS project, the algorithms reviewed in the previous sections can assist the characterization and classification of land covers/habitats and, in some cases, segmentation of the landscape. Within the EODHaM 1st and 2nd stage, a range of hyper-spectral techniques reviewed can be exploited, as long as they result in reliable differentiation of the LCCS categories within key levels of the hierarchy. Within LCCS Levels 1-2 and 3, the use of indices and endmember fractions (e.g., derived through SMA or MESMA) potentially allow better differentiation of primary categories, with these including photosynthetic, non-photosynthetic and aquatic vegetation (submerged, emergent or overlying a waterlogged surface), materials within bare areas, forming natural or artificial surfaces, and even sediment types and loads in water. Beyond LCCS Level 3, targeted use of hyper-spectral data (e.g., based on SAM or SVM) can facilitate the classification of broad categories relating to natural/semi-natural and cultivated life forms (e.g., different types of trees, shrubs, forbs and graminoids), surface materials (sand, mud, rock) and water. Feature extraction modules can also be exploited to assist the classification of agricultural areas (e.g., through delineation of tree crowns in orchards or olive groves). For the more detailed classification of plant species and communities undertaken in the EODHaM 3rd stage, and at the habitat level, techniques such as MESMA can be used, with these allowing endmembers representing plant species to be derived as inputs to classifications, including those at the sub-pixel level. Therefore, in summary, the techniques exploited depend upon the requirement for information at each stage within the EODHaM system and the types of land covers and habitats occurring but also the spatial, spectral and temporal characteristics of the available hyper-spectral data. No single algorithm can be adopted but instead, the EODHaM system needs to take the outputs from targeted classifications of hyper-spectral data (e.g., life form maps) and use these to generate the LCCS categories, as outlined in D5.3. Whilst the EODHaM system is capable of using hyper-spectral data, temporal data are likely to be needed despite the increased spectral bands and resolution. This additional requirement is likely to reduce uptake using airborne data alone because of the costs involved, but significantly benefit the system if spaceborne systems become operational.

3 Hyper-spectral data for BIOSOS test sites.

For three of the BIOSOS test sites, hyper-spectral data have been acquired, primarily from airborne sensors, details of which are outlined in Table 4.1. The corresponding dates of the VHR spaceborne multi-spectral images are also provided.

3.1 Italy, Le Cesine

Airborne Multi-spectral Infrared and Visible Imaging Spectrometer (MIVIS) were acquired over Le Cesine on the 24th May 2009 (08:06 UTC) at a flying height of 1.5 km, providing a spatial resolution of 3 m at nadir. Data were acquired simultaneously in 91 channels in the VNIR and SWIR and also the thermal infrared (TIR). Further details can be found in Appendix 1. The MIVIS data are illustrated in Figure 3.1.

3.2 Cors Fochno. Wales

Airborne hyperspectral ASIA Eagle data (Specim Ltd., Oulu, Finland) were acquired over Cors Fochno (Borth Bog) by the Natural Environment Research Council's (NERC) Airborne Research and Survey Facility (ARSF) on June 1st 2009 under clear cloud free conditions. ASIA Eagle is a 12-bit pushbroom sensor with 252 spectral bands covering the VNIR regions of the electromagentic spectrum and has a 2.9 nm spectral resolution. The sensor scene consisted of 8 lines of imagery with a spatial resolution of between 0.5 and 0.6 m at nadir. The composite of all flight lines covered the entire length and breadth of

Cors Fochno (Figure 3.2). Colour aerial photographs and LiDAR data were also collected concurrently with the imagery. Prior to data acquisition, a LiDAR calibration flight flown at 1350 m altitude determined the mean elevation error to be 4.3 cm (SD of 4.8 cm).

Table 4.1: Summary of the available hyper spectral and multi-spectral data for the three test sites

Site	**Multispectral**	**Hyper-spectral**	**LiDAR**	**Date**	**Season**
Cors Fochno[1]	WV-2			27th July 2011	Summer
				15th Nov 2011	Winter
				23rd March 2012	Spring
		EAGLE		1st June 2009	Summer
NP Veluwe	WV-2			June. 2011	Summer
	WV-2			September. 2011	Spring
		AHS-160		7th Oct 2007	Autumn
		CHRIS-Proba		Oct. 2007	Autumn
Le Cesine	WV-2			October 2010	Autumn
	Quickbird			June. 2009	Summer
		MIVIS		24th May. 2009	Summer

[1]LiDAR Canopy Height Model (CHM) also acquired in 1st June, 2009.

Figure 3.1. MVIS image of the Le Cesine site acquired on 24th May. 2009.

Figure 3.2. A composite of the EAGLE dataset for Cors Fochno acquired on the 1st June, 2009, with a true colour aerial photograph as a backdrop.

The airborne imagery were delivered in radiometrically calibrated Hierarchical Data Format (HDF). The data were subsequently geometrically corrected for pitch, roll and yaw effects using the NERC-ARSF software azgcorr programme (Azimuth-Systems, 2011). The software, employed aircraft position and attitude data, along with a combined LiDAR and NextMap Digital Elevation Model (DEM) to correct the data to the UK national grid with a pixel size of 1 m. The output data were tested for geometric error using a combination of vector overlays from the UK Ordnance Survey (OS) standard map products and a high-resolution orthorectified aerial photograph. It was found that the accuracy of the initial correction method was insufficient locally to precisely match the OS map products, thus an additional localised geometric correction was applied based on a series of empirical ground control points. Nearest neighbour resampling was used to retain actual pixel values. The subsequent geometrically corrected data had an overall RMSE of approximately 1 m.

The AISA Eagle at-sensor radiances were atmospherically corrected to surface reflectance using the Fast Line-of-Sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) module in RSI ENVI (v.4.8). FLAASH is a physics-based atmospheric correction model based upon the widely-used and well-validated MODerate resolution atmospheric TRANsmission (MODTRAN4) radiative transfer code developed by Spectral Sciences Incorporated (Adler-Golden et al., 1999). FLAASH has been used successfully in previous research for the atmospheric correction of hyperspectral data over both land (Cooley et al., 2002) and water ((Hunter et al., 2010, Kutser, 2004). The FLAASH inputs comprised of the altitude of the sensor, pixel size, flight time and date. The mid-latitude summer MODTRAN4 atmospheric module was chosen and the concentration of aerosols was retrieved using a rural model. Water vapour was retrieved on a pixel-by-pixel basis using the water absorption feature located at 820 nm. Water vapour estimated from FLAASH was within 0.5 cm of that recorded on a MICOTOPS® sunphotometer in the field. The accuracy of the atmospheric correction was checked for a single flight line (Line 3) for which concurrently collected *in situ* data were. After atmospheric correction, some cross-track variation in reflectance remained evident across the multiple flight lines, primarily as a consequence of differences in atmospheric path length and surface bidirectional reflectance distribution (BRDF) effects, which can introduce radiometric effects (intensity and spectral)(Asmat et al., 2011). Thus, to create an almost seamless mosaic of imagery, an empirical flight line correction was subsequently employed to normalize the individual flight lines to flight line 3; the single selected "key" flight line, following the methods of Canty *et al*. (2004). The Eagle data were subsequently smoothed using a Savitsky-Golay filter and bad bands removed, leaving a total of 224 from the original 259 bands, available for further processing.

A colour infrared aerial photograph was acquired from an unmanned Airborne Vehicle (UAV) in March 2012 over the active raised bog area. The exeGesIS company acquired the four-band orthomosaiced image for the Countryside Council for Wales (CCW). A DEM was used to project every image pixel in order to calculate their geographical locations for a georeferenced orthomosaic. Indicative accuracies of the orthomosaic are 10 cm relative accuracy and an absolute accuracy of 3 metres although this strongly depends on the flight height, lighting conditions, availability of textures, image quality, overlap, and type of terrain (ExeGIS, 2012).

3.3 The Netherlands

Hyper-spectral data for the Dutch study area were acquired with the AHS-160 sensor. Specifications of the AHS-160 sensor can be found in Appendix 1. The AHS-160 images were acquired in two overpasses on 7 October 2007 by INTA (Spain), using a CASA 212-200 aeroplane (Figure 3.3). Additional preprocessing of the two images from DN values to radiance and surface reflectance was carried out in the processing and archiving facility of VITO (Belgium). The PARGE model was used for geometric correction of the data whilst the ATCOR model was used for atmospheric correction. Further details for the sensor and an assessment of image quality as well as information on weather conditions can be found in Kooistra *et al.* (2008). For the Ginkelse and Ederheide sites in the Netherlands, Compact High Resolution Imaging Spectrometer (CHRIS) PROBA data were acquired on October 22, 2007 as part of HABISTAT, although was not used in this study. Further details can be found in Appendix 1.

3.4 Data summary

The main spectral and spatial characteristics of the different hyper spectral instruments are summarised in Table 4.2. While the EAGLE sensor has the highest spatial resolution (1 m), it has the most limited spectral range (VIS to NIR) owing to the absence of the SWIR region provided by the HAWK sensor. The AHS-160 and MIVIS sensors have comparable spectral ranges but the MIVIS data is higher in spatial resolution. The CHRSI-Proba instrument is more limited in terms of spatial resolution and spectral range.

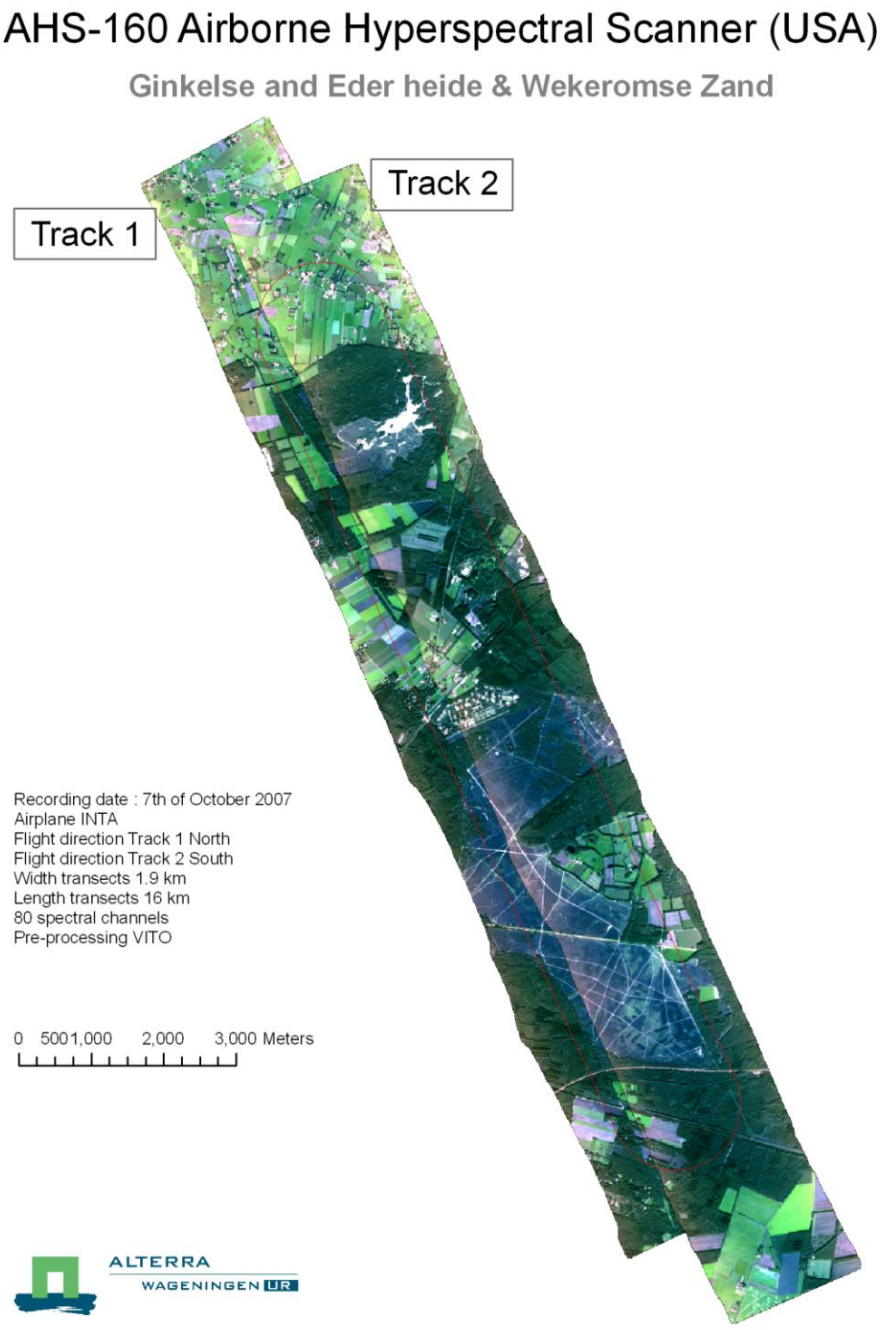


Figure 3.3. AHS-160 hyper-spectral image of the the Dutch study area.

Table 4.2: Summary of the hyper-spectral sensors providing data for the BIOSOS test sites

Sensor	Site	Digitisation	Sensor type	Swath width	Spatial Resolution (m)	Spectral Resolution	Number of bands	Spectral range (nm)	Date	Provider
AISA EAGLE	Cors Fochno	12 bit	Pushbroom	1000 pixels	1	2.9nm	252	400 to 970	01/06/2009	NERC
AHS-160	Veluwe	12 bit	Whiskbroom	1.5 km	2.5	VIS and NIR: 30 nm SWIR 1: 200 nm SWIR 2: 13 nm MIR: 300 nm LWIR: 400 nm	80	430 to 12.700	07/10/2007	INTA (Spain)
CHRIS -Proba	600	12 bits	Pushbroom	14 km in VNIR and SWIR	18	1.3 nm @ 410nm to 12 nm @ 1050nm	82	410 to 1.050		Other project (HABISTAT)
MIVIS	1.5	12 bit	Whisk-broom	Unknown	3	411 to 819 (20 nm) 1.145 to 1.540 (50 nm) 1.992 to 2.474 (9 nm) 8.340 to 1.242 (360 nm)	102	411 to 12.420	24/05/2009	Other project (TIZIANO)

4 Methods:

For the test sites of Cors Fochno and Veluwe, an investigation into options for using hyper-spectral data within the EODHAM system were undertaken, with particularly focus on the interchange of indices and endmember fractions and estimation of sub-pixel cover fractions (e.g., through spectral unmixing) for detailed mapping of habitats, including those assigned Annex 1 status. The use of these data for classifying land covers according to the LCCS scheme but also habitats through discrimination of dominant genera or species, including those that are considered to be indicators (e.g., *Sphagnum* species on active bogs), is also explored. The following sections provide an overview of the approaches considered.

4.1 Image subset selection

For the BIOSOS test sites, subsets were taken of the hyper-spectral imagery because of the large number of channels and data volumes. For Cors Fochno, a single EAGLE flight line was selected (Figure 5.1) with this encompassing a range of LCCS categories as well as proportions of water, PV and NPV. As the data were acquired in June 2009, much of the vegetation was green (particularly in improved fields dominated by the grass species *Lolium perenne*) with the exception of two large patches of non-green cotton grass, with both *Eriophorum angustifolium* and *E. vaginatum* dominating. The active raised bog supported a diversity of plant life forms, including woody shrubs (*Calluna vulgaris*, *Myrica gale* and *Erica tetralix*), mosses (e.g., *Sphagnum* species) and lichens (e.g., *Cladonia* species). The area contained a diversity of species associated with both Annex I habitats, the active raised bog and the modified raised bog, and served as an area suitable for comparing the use of the hyper-spectral and WV-2 single and multi-date data for more detailed classifications of habitats. For Veluwe, the main areas associated with the Annex I habitats at Ginkelse and Ederheide were selected in the image subsets. MVIS data were acquired over the entire area of Le Cesine and these data were used specifically to indicate the benefits of including SWIR data.

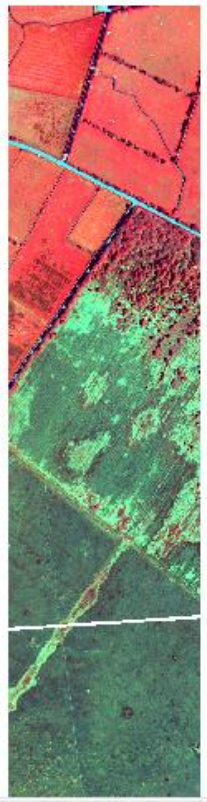


Figure 5.1. A subset of the EAGLE data acquired over Cors Fochno. with these including the Annex I habitats of the active raised bog and modified raised bog.

4.2 Ground-based measurements

For two of the test sites (Cors Fochno and Veluwe), ASD field spectroradiometer measurements and ground observations were taken to support the pre-processing and interpretation of the hyper-spectral and also multi-spectral data.

4.2.1 Cors Fochno

For Cors Fochno, field data had been acquired previously (in September, 2009) but additional data were collected in 2011 and 2012. In 2009, the vegetation percentage cover by dominant species within 1 x 1 m plots was recorded for the active raised bog in 2009, with data collected at the same time as the Eagle data flight overpass. In 2011, ground data relating to GHCs and associated species were collected to support the development of the LCCS classification and subsequent translation to habitat categories. In the summer-autumn period of 2011, two field campaigns recorded species data within the active bog and surrounds based on the GHC taxonomy and with a plot size of 2 x 2 m. The data from all plots (Figure 5.2) were combined subsequently into a single database.

Two subsequent visits (in January and May, 2012) were made to the active raised bog and the modified raised bog with the specific purpose of guiding the labelling of endmember pixels within the EAGLE hyper-spectral data, as selected by the PPI. Points corresponding to the centres of these pixels were overlain onto colour infrared (CIR) aerial photographs acquired in 2006. Approximately 20 were visited and dominant plant species recorded so that a physically meaningful label (i.e., a species or plant functional type) could be associated with the endmember.

In addition to the ground-based estimates of species coverage, field spectra were collected over a relatively small extent using an ASD field spectroradiometer and from areas < ~50 cm diameter dominated by a single species or from mosaics where several occurred. In each case, average spectra were recorded. Target species included *Molinia caerulea*, *Eriophorum angustifoliun*, *Myrica gale*, *Calluna vulgaris*, *Erica tetralix* and *Sphagnum* species. For some species (e.g., *M. caerulea*), litter was observed and hence, from all species combined, spectra representing PV and NPV were collected. Spectra were taken from vegetation associated with features of interest including hummocks (dominated by woody shrubs such as *E. tetralix* and *C. vulgaris*), hollows in-filled by sphagnum pools, and peat excavation sites that formed linear ridges and troughs across the surface of the modified active bog. Several zones of transitions between covers associated with identified endmember spectra were identified such that gradients of change observed within the imagery could be better understood. The field spectra were subsequently used to inform on the endmember selection in spectral unmixing procedures.

4.2.2 Veluwe:

To support the required atmospheric correction of the hyper-spectral data acquired over the Veluwe site and characterise the distribution of habitats present (according to the BioHab methodology), field data had been collected over the same period as the overpass of the AHS-160 data (October, 2007). For atmospheric correction, reference targets considered included bare sand, gravel covered areas, water (small ponds) and asphalt roads. The reference plots for survey were 10 x 10 m (Figure 5.3). The reflectance of selected vegetation types (based on the life forms present) was also measured at several locations over the growing season (Bunce *et al*., 2008). Within each life form, the dominant species was recorded together with their percentage cover.

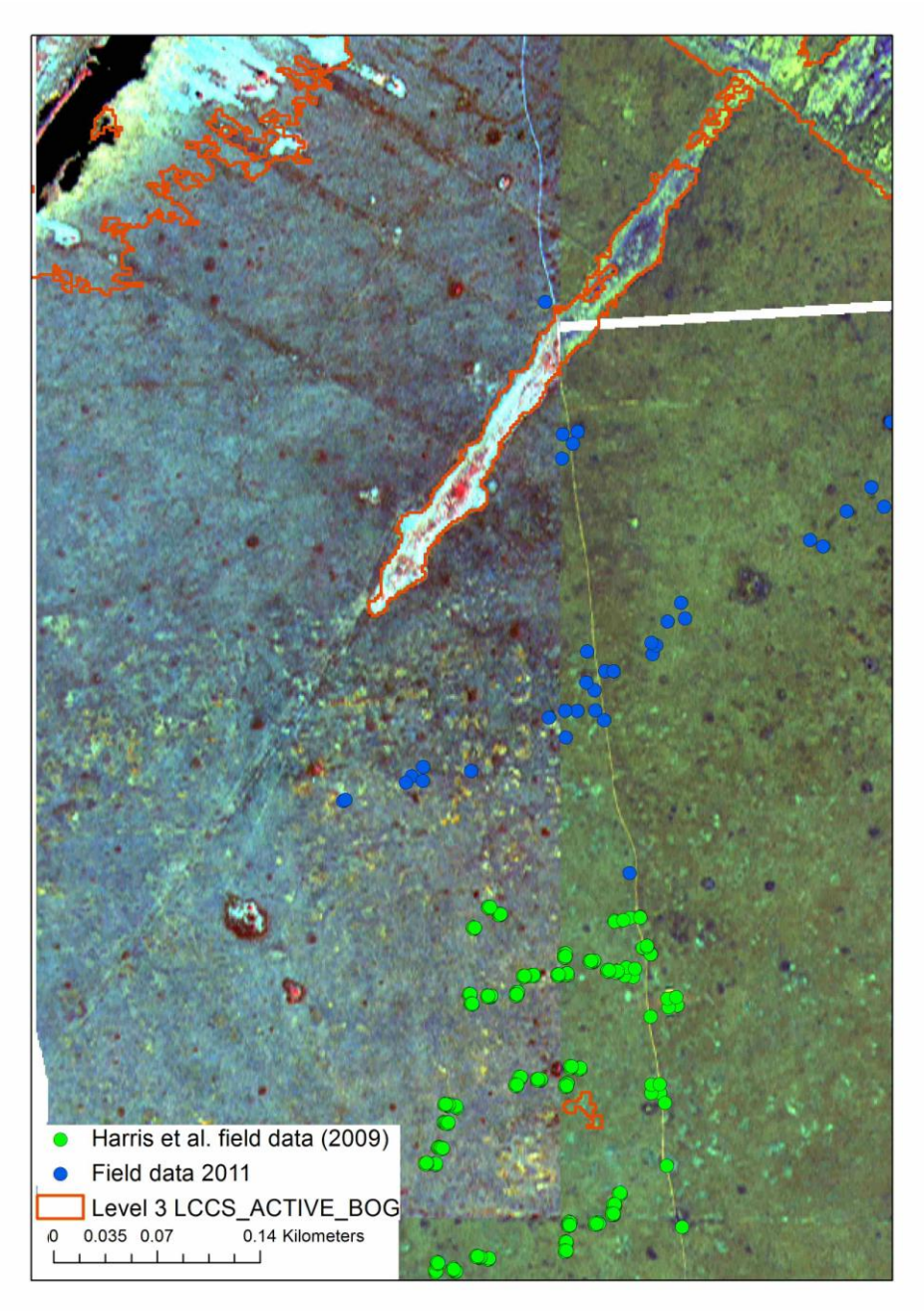


Figure 5.2. Field data from the Cors Fochno field campaign in 2011 as well as that collected in 2009 overlain on a false colour composite of EAGLE flight lines (R:b212.G:b99.B:b10). The red line separates the active bog from closed perennial grassland dominated by *Molinia Caerulea*.

Plots were located within a relatively small area so that they could be measured in one day so as to minimise atmospheric conditions. Reflectance measurements were also taken during the day of the overflight. To characterise the variability within the plot, five reflectance measurements per location were performed along a transect of 10 m (at 0.2,5, 7.5 and 10 m) in an east-west direction. Spectral sampling was conducted by holding the spectroradiometer probe in a nadir position (zenith and azimuth =0°) from a distance of ~ 1 m above the ground (GFOV ~ 44 cm). A digital photograph was taken of each area and geographic coordinates were noted.

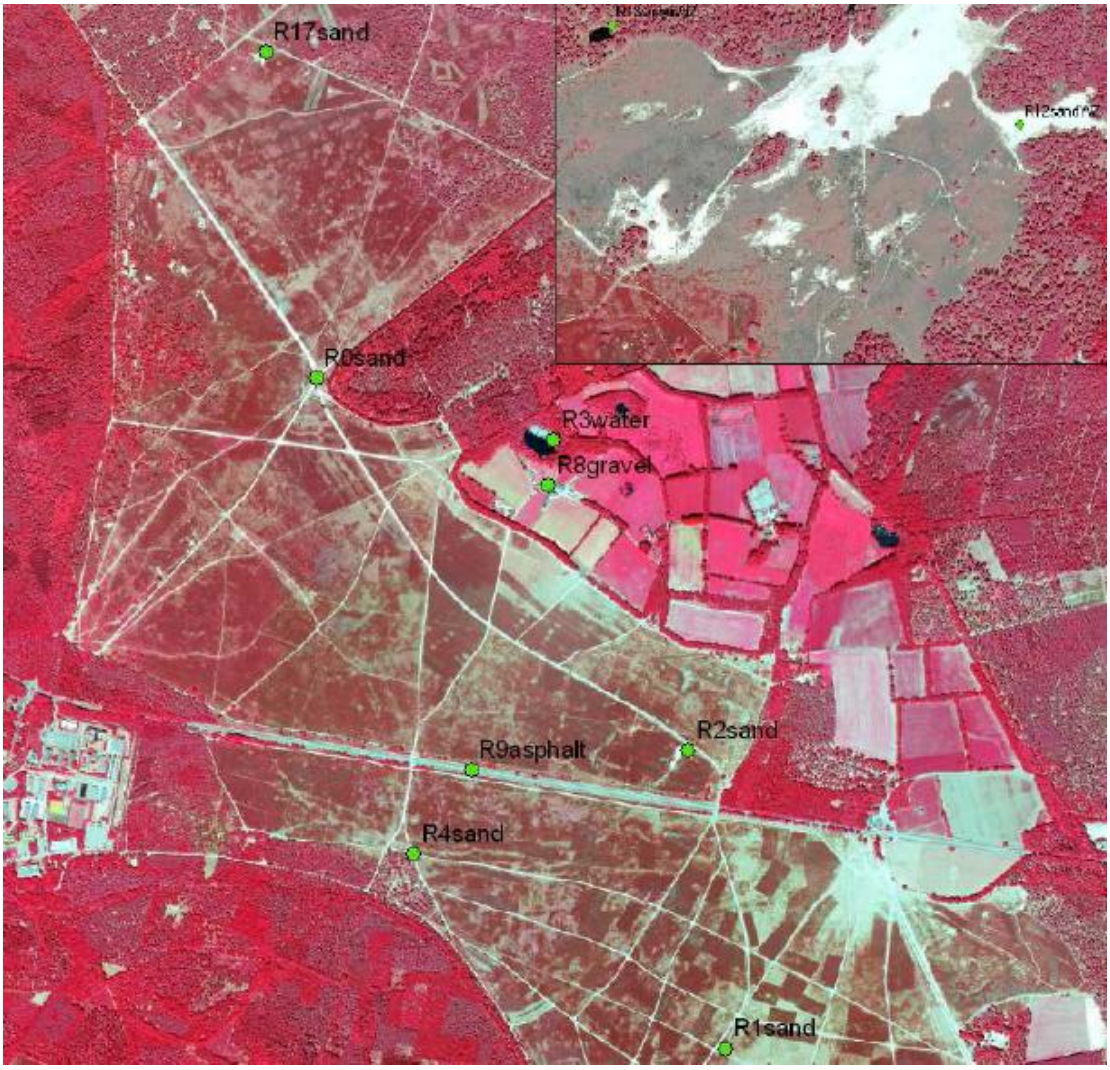


Figure 5.3. Location of radiometric reference measurements in Ginkelse & Ederheide. The inset shows the 2 locations in the Wekeromse Zand

4.3 Relationships between indices and endmember fractions.

A key component of the EODHaM system is the use of indices for identifying boundaries between land cover categories, with these including the NDVI (for differentiating photosynthetic vegetation), the PSRI (non-photosynthetic vegetation) and the NDWI (water). These indices were developed with reference to continuous reflectance spectra and are therefore directly applicable to hyper-spectral data acquired in the appropriate wavelength regions. However, when using broadband multi-spectral data, the indices are necessarily defined using the channels that most closely correspond to the narrow wavebands used in the original formulations.

Endmember fractions present an alternative data suite to indices but the information obtained can be similar. In particular, the endmember fractional images of PV (green), NPV (brown) and moisture/shade (SM, including water) should correspond to that of the NDVI, PSRI and NDWI respectively, as indicated in Table 5.1. Hence, in the absence of some wavebands required for calculation of indices (e.g., the red edge band used for the PSRI is not available within Quickbird or IKONOS images), endmember fractions might be used. A focus of the work was therefore to establish the extent to which the key indices and endmember fractions were interchangeable, with consideration given first to hyper-spectral and subsequently multi-spectral data.

Table 5.1. Endmember fractions and corresponding spectral indices

Hyper-spectral Endmembers	WV-2 biophysical indices	Descriptor
PV	Normalised Difference Vegetation Index (NDVI)	Vegetation productivity or photosynthetic activity
NPV	Plant Senescence Reflectance Index (PSRI)	Dead or senescent vegetation
Water	Normalised Difference Water Index (NDWI) and Water Band Index (WBI)	Moisture content of vegetation, water.
Shade	NDWI or texture	Canopy shadows

Based on the Wales imagery, the endmember fractions associated with PV, NPV and SM were derived separately from the hyper-spectral and multi-spectral data and then compared against the NDVI, PSRI, soil indices and the WBI to establish the relationships between these. These measures were difficult to compare between sensors because of differences in the times (months and years) of acquisition and the spectral wavelength regions considered, although relative differences and similarities were considered.

4.4 Classification of habitats and dominant species/genera.

The approach to the classification of habitats and their constituent plant life forms was adapted from Hamada *et al*. (2011), which considered a two-tier methodology in the estimation of fractional cover of Californian Sage Scrub communities (i.e., the life forms of true shrub, sub-shrub, herb and bare ground) in California's Mediterranean-type ecosystems. In this study, two different approaches were adopted based on contrasting classification models, the H-resolution and L-resolution models, whose characteristics are summarised in Table 5.2. In the H-resolution model, where the target object is larger than the pixel resolution, per pixel classifiers and object-based image analysis (OBIA) were used. In contrast, for the L-resolution model, where the target object is smaller than the pixel resolution, spectral mixture analysis (SMA) was used to retrieve the sub-pixel cover fractions of the target material. Hamada *et al*. (2011) concluded that an object-based classification of a pan-sharpened Quickbird multispectral image (0.6m) was the most cost-effective long-term monitoring method of sage shrub communities overall and that the H-model approaches (i.e., OBIA) generally performed better than the L-model (i.e.,SMA), with this based on reference to ground data. More generally, bare ground and true shrubs were classified with higher accuracy than sub-shrub and herb, regardless of the imagery used.

Similarly, for the purpose of generating classifications of LCCS and GHC habitat categories in the context of BIOSOS from hyper-spectral and multi-spectral imagery, two approaches were considered.

a) Classification within an object-orientated rule-based system, within this including the use of endmember fraction images, derived from spectral unmixing, and indices (top down; i.e., H-model).
b) Classification based on endmembers for dominant species, with these derived through the use of MESMA or SMA (bottom up; i.e., L-model). Comparisons with relative proportions generated through fuzzy classification were also undertaken.

For Cors Fochno, both approaches were considered using WV-2 and EAGLE hyper-spectral data respectively, with focus on classifying life forms/GHCs and dominant species associated with Annex 1 habitats of the active raised bog and modified bog (Table 5.35).

Table 5.2. Characteristics of the L-resolution and H-resolution models for estimating fractional cover in Californian Mediterranean ecosystems using high resolution (SPOT-5) and VHR Quickbird multispectral imagery (Hamada *et al*., 2011).

	H-resolution model	**L-resolution model**
Description	Image object > pixel resolution	Image object < pixel resolution
Classification approach	Per pixel classifiers (MXL.ANN[2]) or OBIA	SMA
Algorithms used	Segmentation followed by Parametric or non-parametric classifiers	SMA, MESMA
Typical application	VHR imagery (≤2m) and per-pixel analysis	Fraction retrieval of PV, NPV, soil and shade etc. on high resolution imagery (≥10m)
Most accurate in fractional cover calculation	ANN with a grid cell size of 50m	MESMA on SPOT-5 data with a grid cell size of 50m

Table 5.3. Lifeforms and associated descriptions based on land covers (LCCS), habitat (GHCs) and species/genera Cors Fochno

Lifeform	**Dominant GHC**	**Species/genera**
Graminoids	CHE	*Eriophorum angustifolium* *Eriophorum vaginatum* *Molinia caerulea* *Juncus effusus* *Mixed improved grassland species*
Forbs	LHE	*Narthecium ossifragum*
Woody Shrubs	LPH/ MPH	*Myrica gale* *Erica tetralix* *Calluna* vulgaris
Lichens	LIC	*Cladonia* species
Bryophytes	BRY	*Sphagnum* species

5.4.1 Classification within an object-orientated rule-base

As described in D5.3, the object-based approach to classification (i.e., akin to the H model) was applied to multi-date WV-2 data acquired over Cors Fochno (July and November, 2011 and March, 2012), with the area corresponding to that of the selected hyper-spectral flight line. Initially, the WV-2 data were segmented using the algorithm of Shepherd *et al*. (In prep) and the hierarchical scheme for the classification of LCCS Levels 1-2 (EODHaM 1st Stage) and Level 3 was then implemented, with categories including vegetation and non-vegetation, aquatic and terrestrial and artificial/cultivated and managed and semi-natural natural land covers (Figure 5.4a-d). A more detailed LCCS classification was then applied with the main categories associated with the active raised bog, woodlands and cultivated grasslands (Figure 5.4e).

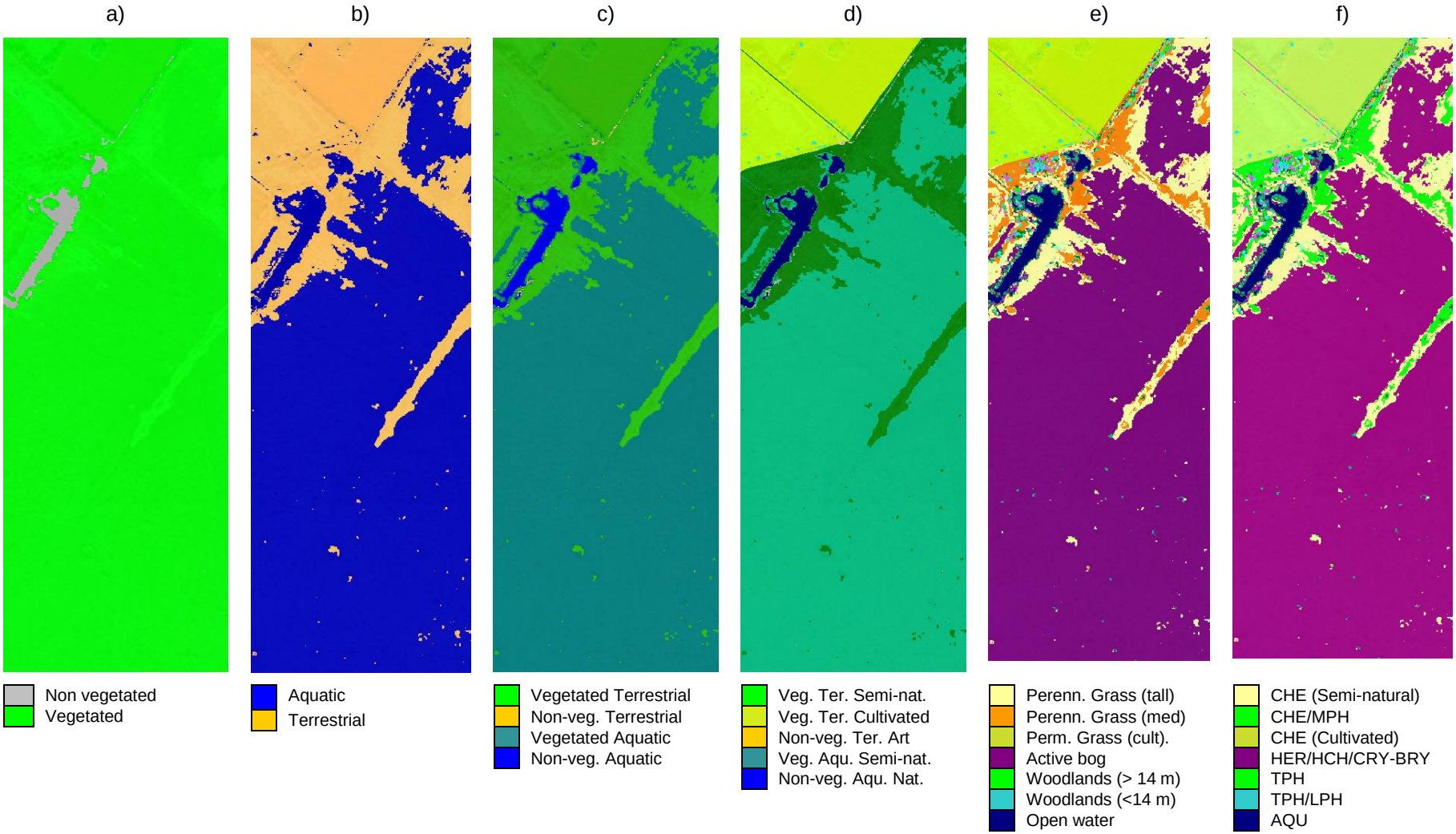


Figure 5.4. Classification of a) vegetated versus non-vegetated and b) aquatic versus terrestrial surfaces. The Level 2 and 3 classifications are shown in c) and d) respectively. The LCCS categories are shown in e) with these translated to GHCs f).

Following translation to GHCs (Figure 5.4f), a finer spatial resolution segmentation was applied to allow a more focused classification of species and communities within the Annex 1 habitats of the active raised bog and the modified bog as well as immediate surrounds. The classification was undertaken using a rule-based approach, with consideration then given to sub-pixel classification through application of fuzzy-membership functions within eCognition.

5.4.2 Classification based on endmember estimation

Classification based on endmember estimation was conducted by applying MESMA to the EAGLE data acquired over Cors Fochno. Prior to unmixing, pre-processing steps were followed to reduce the dimensionality of the hyperspectral datasets and produce fewer uncorrelated and noise-free bands. For this purpose, the Minimum Noise Fraction (MNF) was applied within ENVI to the EAGLE data. The MNF is designed to address a shortcoming of Principal Component Analysis (PCA), which is a lack of a definite trend of increasing noise with component number. In the MNF transformation, the principal components are re-ordered so that the higher the component number, the greater the noise fraction (Green *et al*., 1988). To select spectral endmember candidates automatically from the MNF bands, the Pixel Purity Index (PPI) was used. The PPI is an iterative process that gives the frequency that a pixel is recorded as extreme in *n*-dimensional feature space (Boardman *et al*., 1995). The values of the PPI pixels represent the number of times that a pixel is counted as extreme and a user-defined threshold is used to accept those that are above a certain value, with these considered to be suitable endmember candidates.

To retrieve sub-pixel cover fractions for Cors Fochno, a MESMA was applied to the subset of the EAGLE data shown in Figure 5.1a. A method of MESMA library construction was adapted from Powell and Roberts (2008). First, the endmembers were derived from pixels selected by the PPI and labelled into groups, with these including graminoids, shrub and moss categories. The assignment of spectra to an endmember class was undertaken with reference to the field data collected, as outlined in Section 5.2.1. A default photometric shade endmember was also considered to account for variations in illumination across the scene. Within the spectral library generated, 980 endmembers were contained. Sorting of candidate endmembers per group was carried out to reduce spectral confusion within each category and to find the optimal spectra. First, endmember spectra that could not be well modelled by a combination of the other spectra from within the group were eliminated. While the more spectra per group, the more effective the elimination process, a compromise between having fewer spectra per group and allowing effective modelling was necessary. Once the groups were internally sorted, 2, 3 and 4-endmember models were generated as well as a default endmember for shade.

A series of constraints (fraction values between −0.05 and 1.05 and a RMSE threshold of 0.02 between the 2 and 3 endmember models and 0.008 between the 3 and 4 endmember models) were applied for candidate models. The RMSE thresholds represent cut-off values for model complexity. If exceeded, the pixel is modelled with more endmembers in the combination, which leads to a more complex model. Conversely, in increasing the threshold, a simpler model is used. Images of the different endmember combinations were generated to ascertain which combinations were optimal for modelling the components of a given pixel. Whilst selection of this RMSE threshold was subjective, the analysis was guided by knowledge of how many endmembers would be expected for modelling a pixel. Several endmember combinations were viewed before the optimal model combination for each pixel was selected. On this basis, and using the optimal model, an image of endmember fractions was generated. The results from the MESMA analysis were used directly to classify the communities occurring within the active bog, with these based on the sub-pixel proportions estimated.

5.4.3 Evaluating classifications

For Cors Fochno, a range of data were available for interpreting and validating the classifications generated from both the WV-2 and hyper-spectral data, with these including field-based measurements but also Unmanned Airborne Vehicle (UAV) data acquired in March, 2012, over the active bog. These data can be summarised as:

a) Polygons (derived from ROIs) representing 'pure' communities as interpreted from the UAV and covering the main species types

b) Polygons drawn from the UAV but aligned with the four corner points of the 2009 plot data and assuming minimal change between 2009 and 2012.

c) Polygons with a 1 m buffer around GPS points sampled in the 2011 survey

The latter two datasets had species information attached, as recorded in the field survey, while the ROIs were interpreted based on their spectral distinction and homogeneity in the UAV image. The process of deriving the validation dataset is summarised in Figure 5.5 and a summary of the number of ROIs and ground points per species recorded is presented in Table 5.4. Additional field campaigns are being conducted during the summer of 2012 to allow a more comprehensive validation of both the hyper-spectral and WV-2 classifications, with this feeding into the EODHaM protocols, and hence only a preliminary comparison is reported.

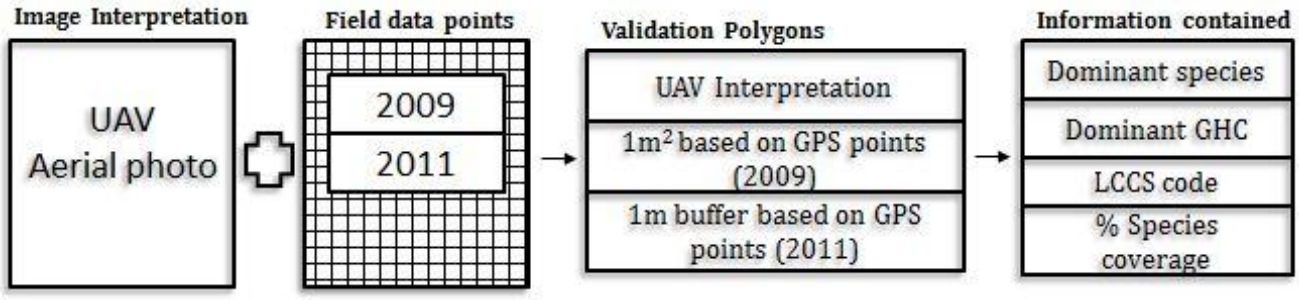


Figure 5.5: A graphical summary of the method used to derive the validation dataset based on aerial photo interpretation, location of field points using GPS readings and the types of species data extracted from the field points

Table 5.6. The number of points and ROIs per species group, as derived from the combined 2009/2012 Cors Fochno dataset as a function of GHC

GHC	No	Field		UAV	GHC	No	Field		UAV
		2011	2009	2012			2011	2009	2012
BRY	23	2	2	33	LHE	3			
BRY/CHE	1				LIC				14
BRY/HEL	1				LIC/SCH	1			
BRY/MPH	1				LPH	6	4		9
BRY/SCH	5	2			LPH/BRY	1			
CHE	49			69	LPH/CHE	1			
CHE/BRY	1				M	1			
CHE/LPH	1				MPH	2			30
CHE/SCH	6		2		SCH	34	2	6	
EAR	1				SCH/	1			
EHY	7				SCH/BRY	4	2	1	
GEO	3				SCH/CHE	8		2	
HEL	15	1	1		SCH/HEL	11	3		
HEL/BRY	3				SCH/LIC	2	1		
HEL/CHE	4				WAT	1			
HEL/SCH	4								
					TOTAL	**201**	**17**	**15**	**208**

4.5 SWIR spectral indices for aquatic vegetation (A24) discrimination.

As discussed in Section 3.2.2, a limitation of using VHR optical data within the EODHaM system is that the currently operating sensors do not support a SWIR channel. These data are, however, considered to be useful for detecting aquatic vegetation and also discriminating a number of plant communities based on the moisture content of the plant foliage (e.g., forests and grasslands). The potential benefits of using these data and derived indices was therefore investigated using the MIVIS visible to SWIR data for Le Cesine with particular focus on discriminating aquatic and terrestrial vegetation categories (i.e., A11 and A12 from A23 and A24). Wavebands with noise (belonging to the 3rd sensor) were excluded.

Within the MIVIS image, three primary classes were recognisable (both as LCCS and GHCs), with their type and extent determined through a previous Interreg III project through supporting field campaigns.

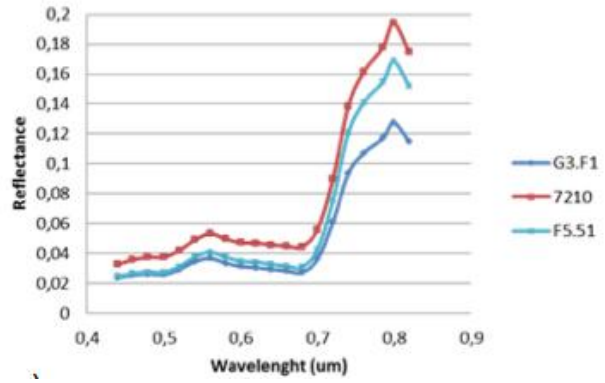
a) Aquatic regularly flooded vegetation dominated by herbaceous (gramminoid) perennial vegetation (A24/A2.A6.E6+R1.O3.M233.N12-HS; *Cladium mariscus*; Annex 1: 7210).

b) Cultivated and managed coniferous plantations, needle-leaved and evergreen (A11/A1.A8.A9_W7; EUNIS G3.F1)

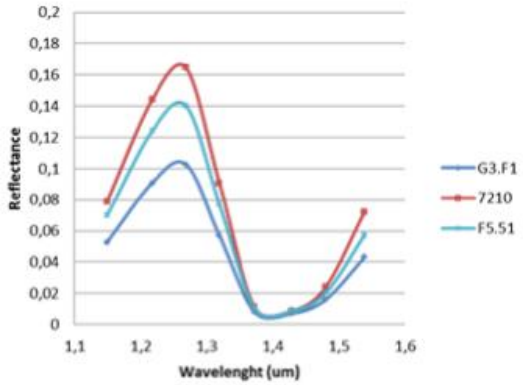
c) Woody broad-leaved deciduous shrubs (A12/A11.A4.D1.E2+O3.M233.N.12; LP+*Rubus* species; EUNIS F5.51).

Typical reflectance spectra for the three habitats are given in Figure 4.6.

a)



b)



c)

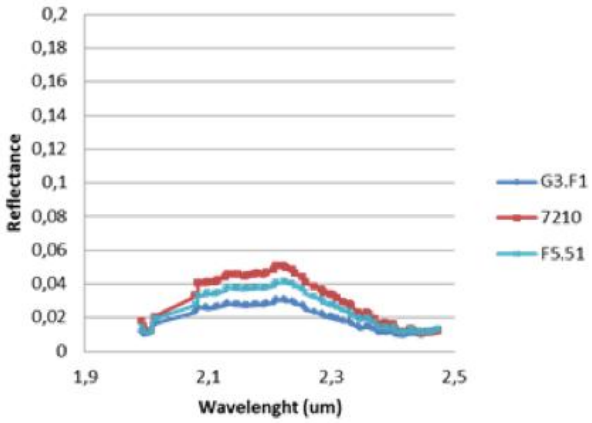


Figure 4.6. Reflectance curves for managed coniferous forests (EUNIS G3.F1), broadleaved deciduous shrubs (EUNIS F5.51) and aquatic graminoid vegetation (Annex 1 7210) extracted from MIVIS imagery

The classes represent examples of aquatic vegetation and both needle-leaved and broad-leaved terrestrial vegetation, which were difficult to distinguish using VNIR data alone. All three were in their active growth phase in May (i.e., on the date of image acquisition), according to the phenological class description provided in D5.2 by botanists. The analysis focused on establishing whether the accuracies

in the mapping of the aquatic category (which occurs in EODHAM Stage 1) could be improved through inclusion of the SWIR data, with comparisons made against classifications undertaken using multi-spectral VNIR data from Quickbird and WV-2 using a similar set of ground truth data.

5 Results

5.1 Relationships between indices and endmember fractions.

Using MESMA and the EAGLE hyper-spectral data as input, fractional images relating to grass (primarily green vegetation), both *Eriophorum* species (photosynthetic but not green) and *Calluna vulgaris* (mixed communities with non-photosynthetic grass and woody material as well as some green vegetation of low productivity) were generated. The comparison with the NDVI (Figure 6.1a) indicated a general increase for the grass (PV) and *Eriophorum* fractions but a decrease with the *Calluna* fraction. These comparisons suggest that the NDVI and PV fraction are broadly interchangeable.

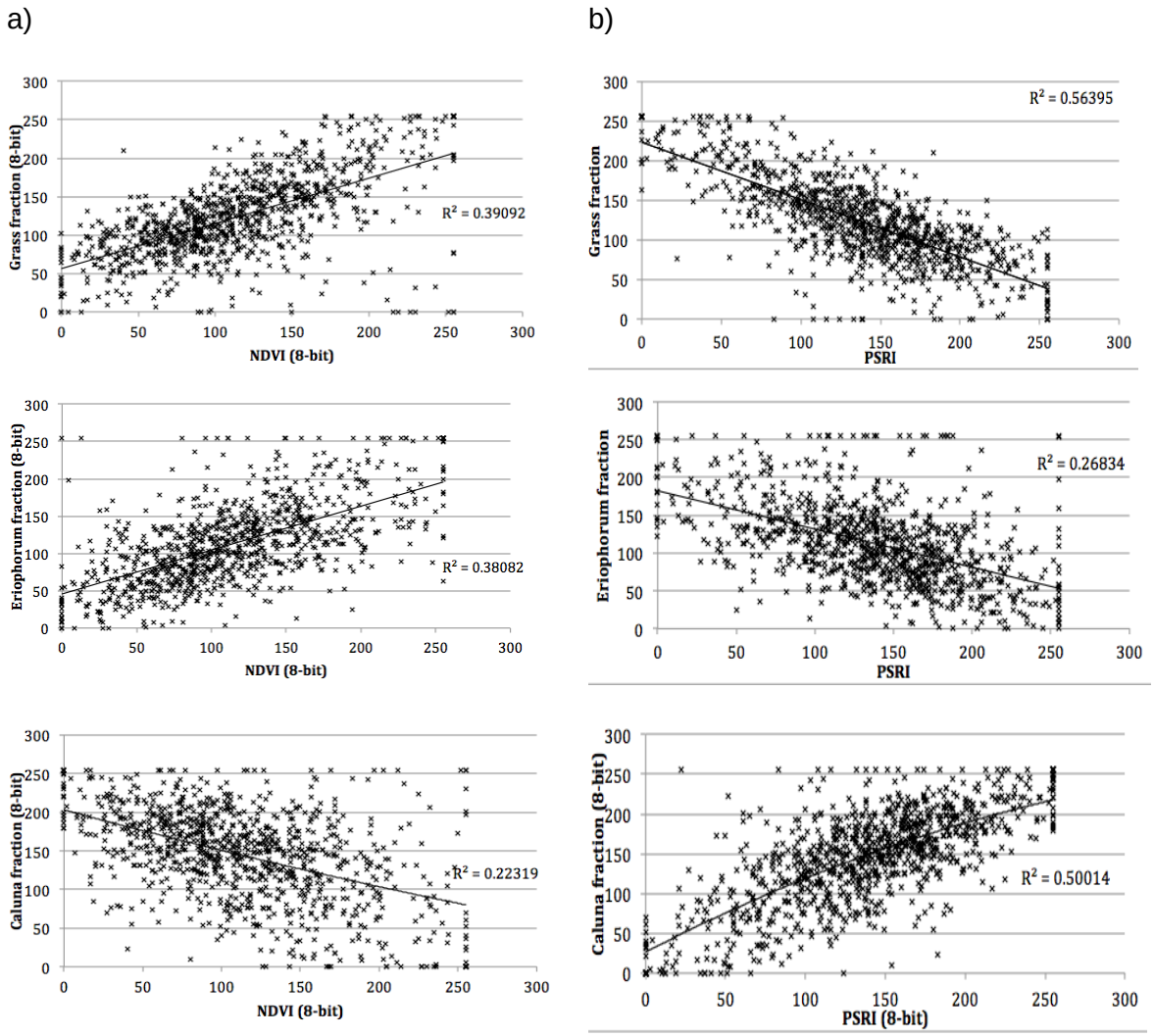


Figure 6.1. Scatterplots of grass, Eriophorum and Calluna fractions for a) the NDVI and b) the PSRI. In each case, both the index and the respective fractions were scaled to the 0-255 range.

The grass fraction and, to a lesser extent, the *Eriophorum* fraction decreased with increasing PSRI (Figure 6.1b) but an increase in the *Calluna* fraction was observed. Whilst *Eriophorum* species give the appearance of being non-photosynthetic, because of the red colouration of the blades, these are photosynthetically active and hence an increase with the NDVI was observed. The analysis suggests that areas dominated by these species cannot be used to define the NPV endmembers. Defining the

endmember fraction for the NPV in the summer months is therefore problematic as few areas of grass or forb litter occur. By contrast, in the pre-flush and senescent periods, extensive grass and forb litter occur within these, in Wales, being comprised of plant material from *Molinia caerulea* and *Pteridium aquilinum* (or bracken) respectively. Nevertheless, a correspondence between the *Calluna* fraction and NPV was observed indicating the large amount of woody material and other dead components that are present. These observations suggest that research is needed to establish the reflectance characteristics of a range of woody and senescent plant materials in order for the NPV endmembers to be better defined and fractions retrieved from the imagery.

In the case of the WV-2 comparisons, a close correspondence was again observed between the PV fraction and the NDVI and also between the water fraction and WBI, with these fractions defined through linear SMA (Figure 6.2). The NDVI is a correlate of vegetation productivity (Cohen *et al*., 2003) and therefore should be an appropriate surrogate for the PV fraction and *vice versa*. However, a weaker correspondence between the NPV fraction and the PSRI was observed, with the endmember used extracted from areas dominated by *Molinia caerulea*, which is largely senescent in March when the WV-2 image was acquired.

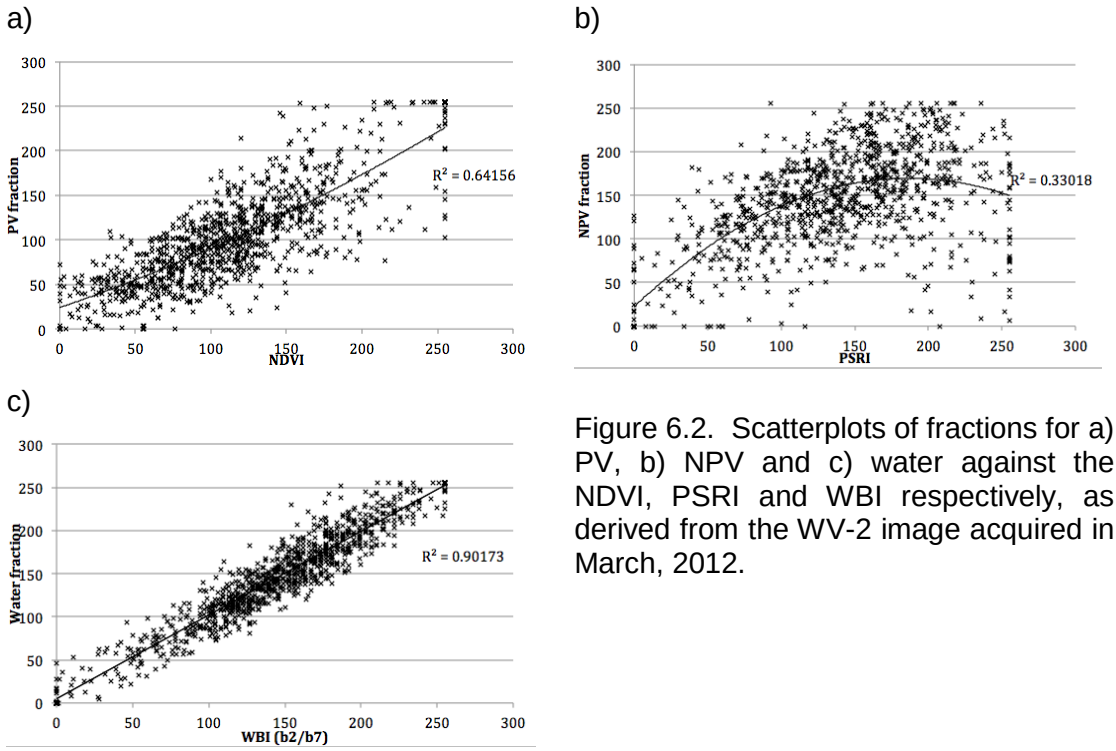


Figure 6.2. Scatterplots of fractions for a) PV, b) NPV and c) water against the NDVI, PSRI and WBI respectively, as derived from the WV-2 image acquired in March, 2012.

5.2 Rule-based classifications.

The rule-based classification of the main categories of vegetation on the active bog based on multi-date (March, July and November) WV-2 data is given in Figure 6.3. The classification utilised a combination of vegetation indices, endmember fractions and reflectance data. The classification confines areas of productive grasslands and also *Juncus effusus* primarily to the cultivated areas in the north-east (with these defined through the LCCS classification at Level 3). However, both grassland types also occur within the semi-natural areas surrounding the active bog. The remainder of the cultivated area as well as the margins surrounding the bog are dominated by *Molinia caerulea*, although *Phragmites australis* is also present but difficult to distinguish. Across the bog, woody shrubs also occur, with these primarily being *M. gale*, *C. vulgaris* and, to a lesser extent, *Erica tetralix*. The majority of the surface is dominated by grass species, with these being primarily *Eriophorum* species interspersed with *M. caerulea*. The bog pools are clearly identified, with these dominated by *Sphagnum* species. Overall, the classification was considered to represent the broad distribution of dominant species within the bog and the surrounds.

a)

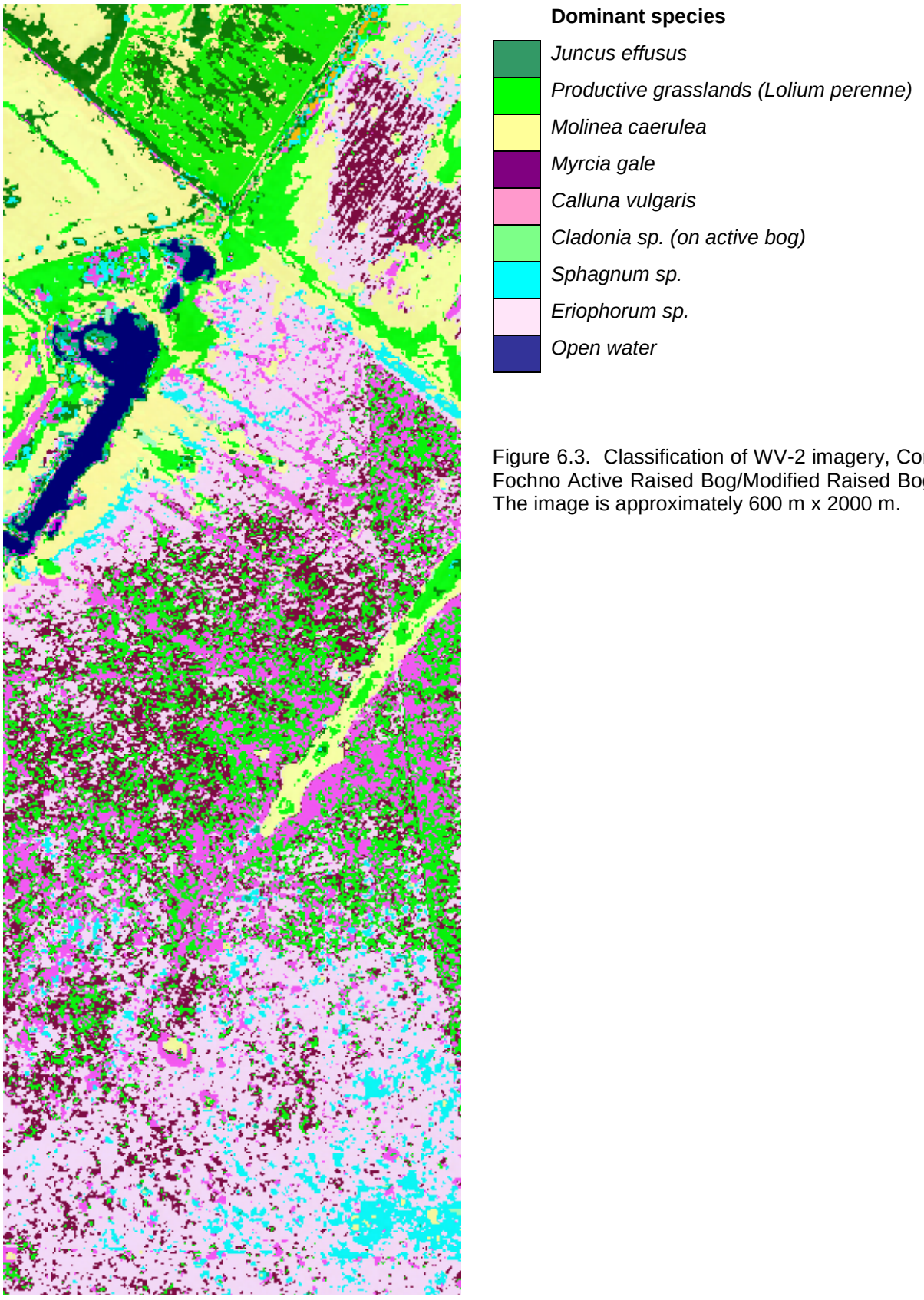


Figure 6.3. Classification of WV-2 imagery, Cors Fochno Active Raised Bog/Modified Raised Bog. The image is approximately 600 m x 2000 m.

Whilst a comprehensive accuracy assessment has yet to be undertaken, with subsequent improvements to the classification anticipated, comparison with the UAV imagery indicated a close correspondence

between the mapped and observed distribution of the dominant species, with examples given in Figure 6.4. In particular, *Sphagnum* species occurring within the depressions (pools) within the active bog as well as *Calluna* are indicated, with the former representing a key indicator of the relative wetness of the bog and of the pool and hummock sequence. Most of the species classified are typical of the Annex I habitats represented by the active raised bog and modified raised bog. The approach therefore illustrates the potential of the EODHaM system to achieve detailed classifications starting from the upper levels of the LCCS.

a) b)

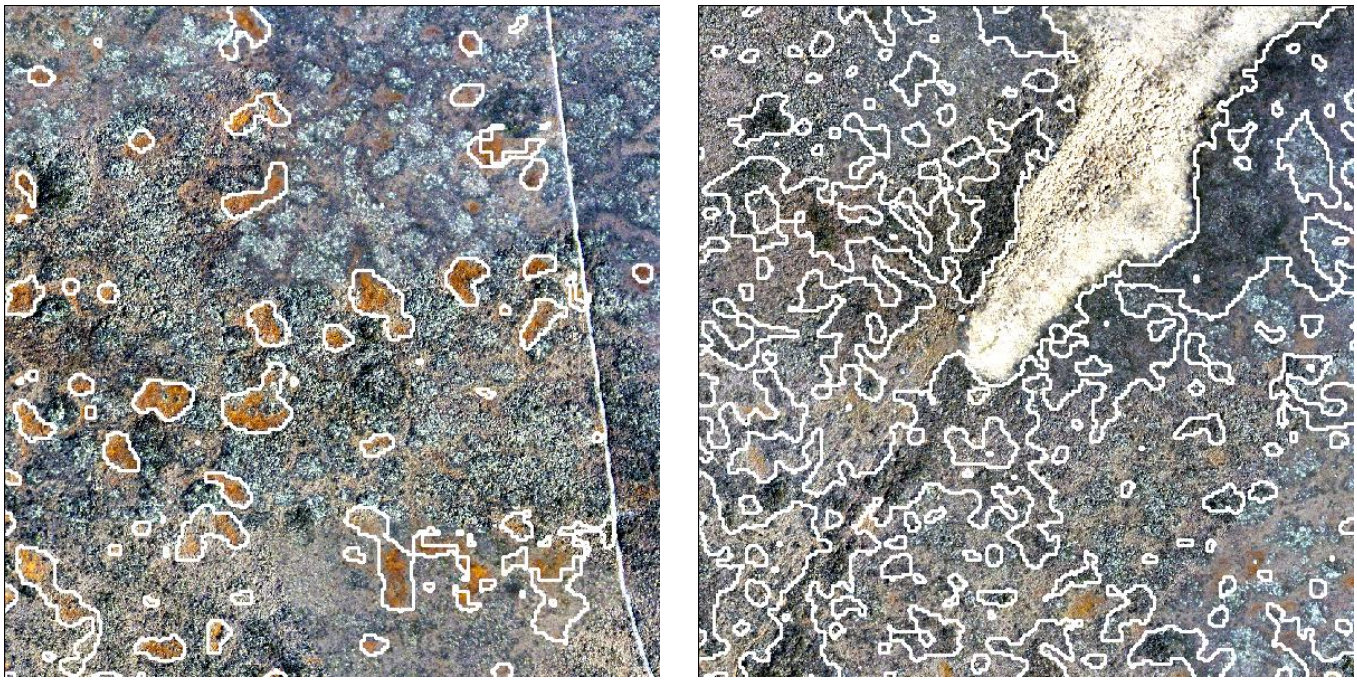


Figure 6.4. Outlines of polygons representing a) *Sphagnum* pools and b) *Calluna*-dominanted active bog (darker shade), Cors Fochno, as classified using WV-2 multi-date imagery, with these overlain over a UAV image acquired in 2012.

5.3 Mapping through spectral mixture analysis.

5.3.1 Cors Fochno

The optimal endmember combination image (Figure 6.7) is from a first iteration of the MESMA analysis of EAGLE data, with this generated using the spectral library sorting procedure described in Section 5.4.2. The water areas are removed from the image using a mask based on the shade fraction. All of endmember fractions were shade-normalised. The species groups are mapped using the optimal combinations of 2- (e.g. *Myrica*), 3- (e.g., *Myrica* and grass) and 4- (e.g., *Myrica*, *Molinia* and Grass) endmember models, with shade representing the additional endmember. Within the 3-4 endmember models, shade was a component but was removed after normalisation. As with the WV-2 classification, grasslands within the cultivated areas were classified primarily as those with combinations of *Juncus effusus* and meadow grass (dominated primarily by *L. perenne*). A close correspondence in the distribution of vegetation with a dominant cover of *Calluna* and also *Molinia* was also observed. The area of the active bog with *Sphagnum* species as a component was, however, overestimated compared to the classification using the WV-2 data, with the latter providing a closer correspondence with the distribution observed within the UAV imagery. Within this area are lawns of *Sphagnum* species such as *S. capillifolium*, with *S. pulchrum, S. papillosum* and *S. magellanicum* while S. *cuspidatum* is found in wetter hollows (Crowther and Groome, 2004). The main bog surface consisted primarily of mixtures of *Eriophorum*, *Calluna* and *Molinia*.

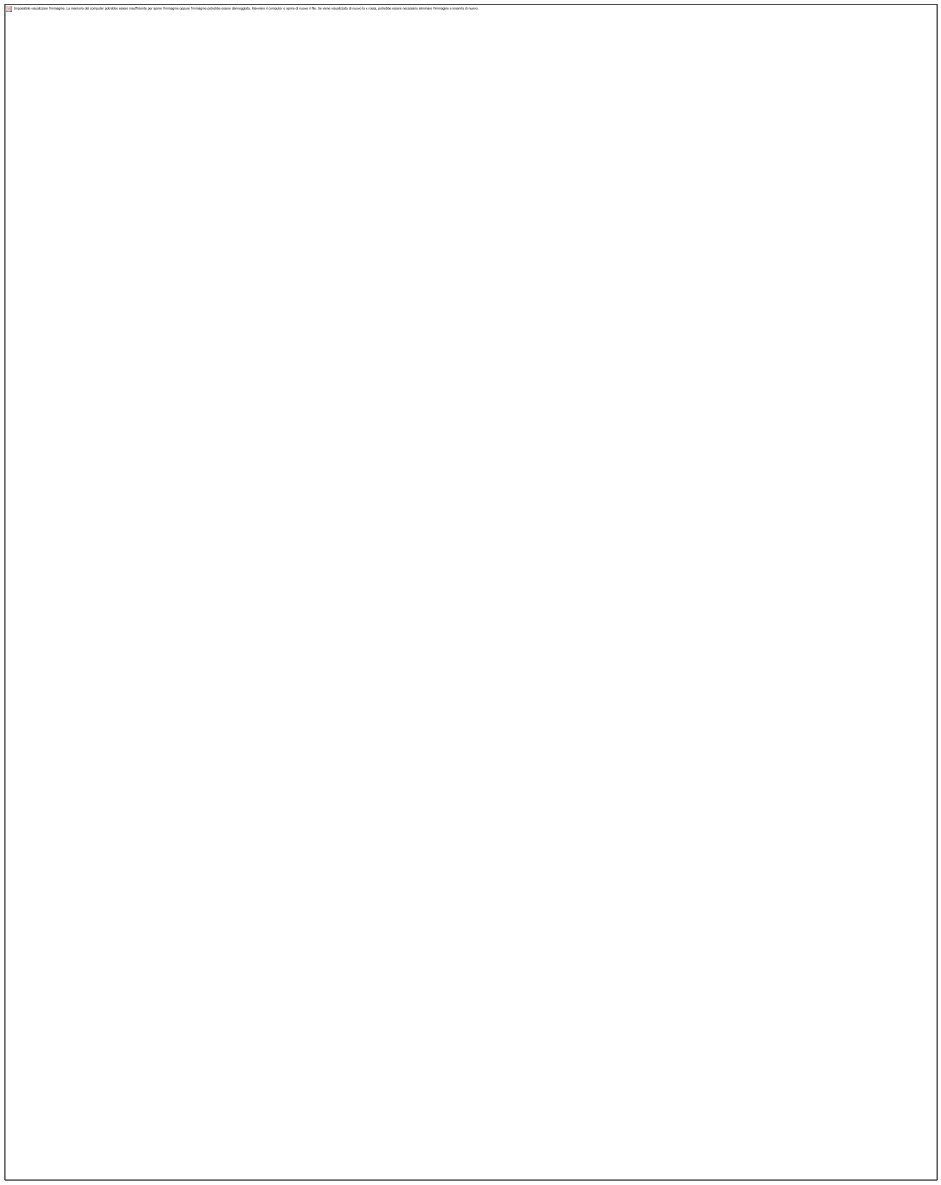


Figure 6.5: The optimal MESMA endmember combinations of the EAGLE subset using RMSE thresholds of 0.2% between the 2 and 0.08% between the 3- and 4-endmember models

Selected endmember fractions are highlighted in Figure 5.6a-c, with the corresponding fuzzy membership functions generated from selected indices of the multi-date WV-2 images also shown (d-f), noting that these were from different seasons and different years In both cases, the image brightness is indicative of the sub-pixel cover fraction for each, with brighter values corresponding to a higher fraction. For some categories (e.g., *Calluna vulgaris*, *Juncus effusus*), the fractional and membership images were broadly similar, although extensive areas of *J. effusus* were excluded from the bog surface using the WV-2 data by the *a priori* classification of this species within the cultivated land cover category at LCCS Level 3. However, in others (as exemplified by *Sphagnum* species), the endmember fractional image tended to overestimate the distribution compared to the WV-2. Whilst the classifications of both the hyper-spectral and WV-2 data are being refined and further validation (e.g., of mixtures) undertaken, the classifications highlight the potential of using both datasets for detailed classifications of habitats, including indicator species, but also the potential to substitute one dataset with the other. Whilst the level of detail able to be provided by the hyper-spectral data, the use of multi-date WV-2 data within a rule-based approach is simpler and yields similar if not improved results.

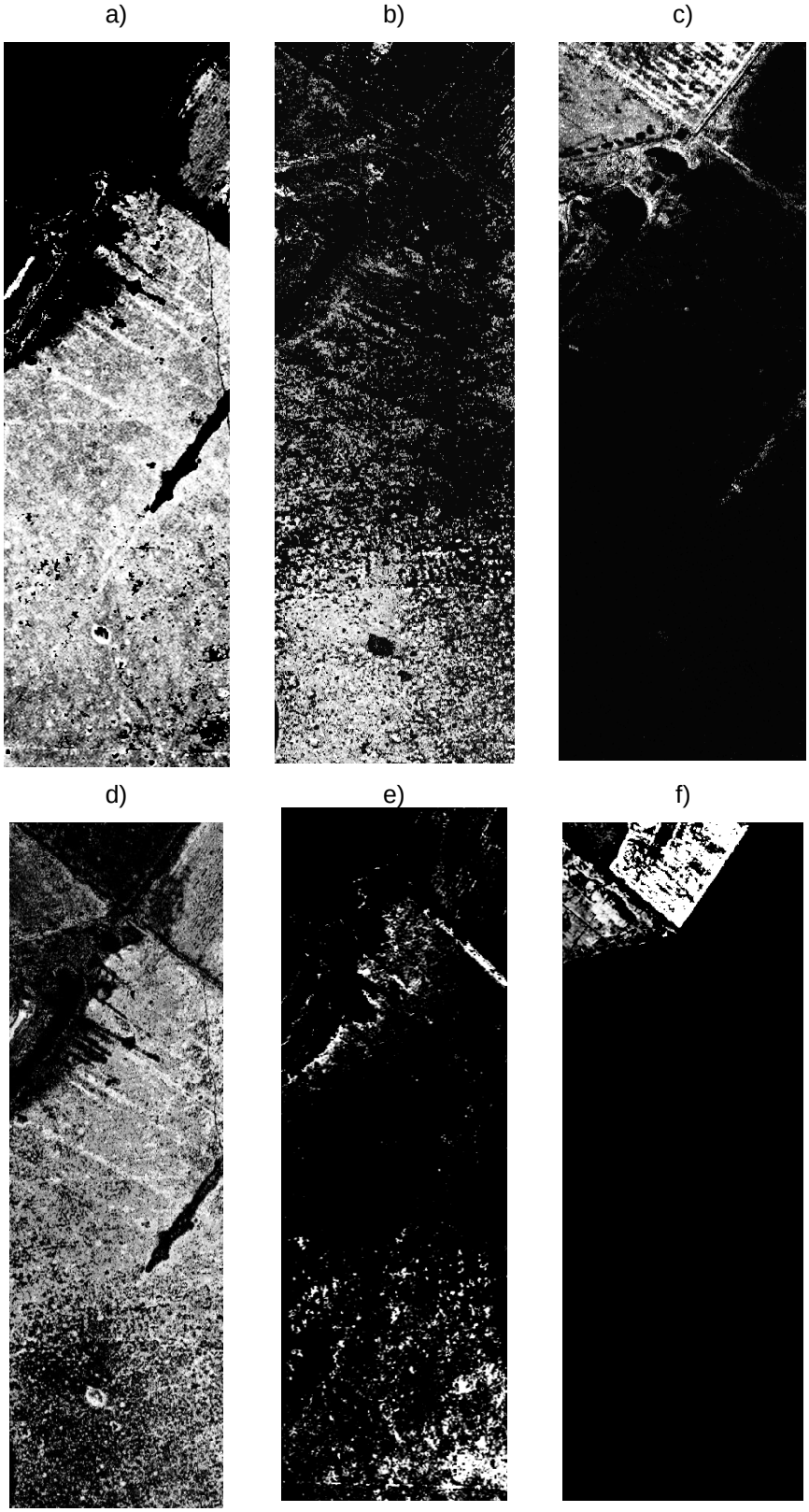


Figure 6.6. Endmember fractions for a) *Calluna vulgaris*, b) *Sphagnum* species and c) productive grasslands. The corresponding fuzzy membership images are provided in d), e) and f).

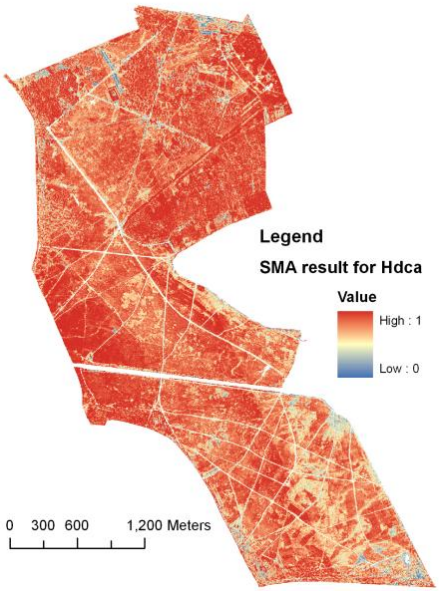
As described in Section 5.4, the RMSE was used in the MESMA as a constraint to find the best model combination per pixel. The selection of the optimal endmember models was found to be a critical step in MESMA and must be considered carefully before unmixing a pixel and producing the image fractions. Different RMSE thresholds were tested in order to accept or reject the endmember combinations; However, the 0.02 and 0.008 cut-off values were chosen after extensive testing of several endmember combinations in their ability to characterise the spatial distribution of plant species. The endmember combination is highly sensitive to the RMSE threshold. For example, accepting a more complex model does not always produce the best scenario for a given pixel. While a complex model might suit a pixel in one part of the image, it may over-complicate modelling in another part. This was particularly relevant in the subset of the EAGLE flight line where there were areas of homogenous vegetation as well as complex species mosaics which required a balance between model complexity and RMSE threshold adjustment.

5.3.2 The Netherlands.

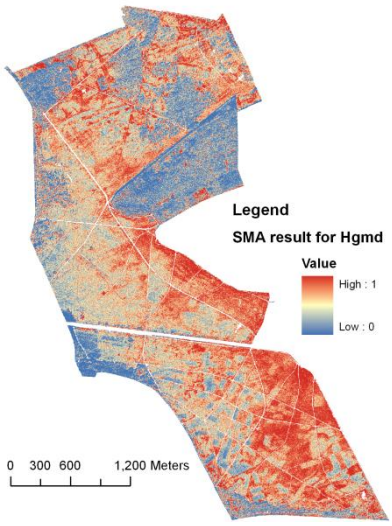
Using SMA, seven fractional maps, with each representing one of the main vegetation classes within Ederheide and Ginkelse Heidem were generated, together with an RMSE image (Figure 5.6). These classes were *Calluna*-dominated heathland (Hdca), *Molinia*-dominated heathland (Hgmd) and coniferous forest dominated by Scots pine (Fcps; *Pinus sylvestris*).

The SMA fraction estimation results correlated reasonably well with the field estimates for both the *Calluna* (Hdca) and the overall grass encroachment (Hgmd and Gpnd), with R^2 values of 0.41 and 0.48 respectively. Overall, RMSE values were relatively low, but some higher values occurred along boundaries between sandy roads and vegetation. This is likely to be caused by shadow and/or scattering effects, but the addition of a shadow endmember did not improve the results. Due to the nature of the mathematical model used for the SMA, the fractional abundance images were allowed to have a negative value, with this set to zero. The *Calluna* heath (Hdc) was abundant over the entire area (Figure 6.5). However, differences in the abundance of *Calluna* were observed, which was attributed, in part, to management practices. The spatial distribution of grass encroachment was observed from the Hgmd fractional image (Figure 6.5b). Two intensively managed areas with short grasses were readily identified from the Gpnd fraction image (not shown). All fraction images conveyed, to a certain extent, the effect of mosaicking within the two flight transects with abrupt fraction changes occurring along a central NW-SE oriented line (e.g. Figure 6.5a). This was mainly attributed to across track illumination differences between both flight lines, which resulted in, for example, relatively high Hdc fractions in the western compared to the eastern flight line. This effect may also partly be responsible for the higher abundance estimations of some of the image-based *Calluna* fractions compared to the actual field-observed coverage (Figure 5.5). As the SMA model was not fully constrained (only unit sum was constrained), some fractions were smaller than 0 and larger than 1. Although the most representative endmembers were selected with care for each vegetation class, some within-class spectral variation still remained, which has not been accounted for in this analysis. A possible alternative approach would be to use Multiple Endmember SMA (MESMA), as with Cors Fochno, which evaluates a list of candidate endmembers and selects the best fit for each pixel, thus avoiding overfitting due to too many endmembers (Rosso *et al*.. 2005).

a)



b)



c)

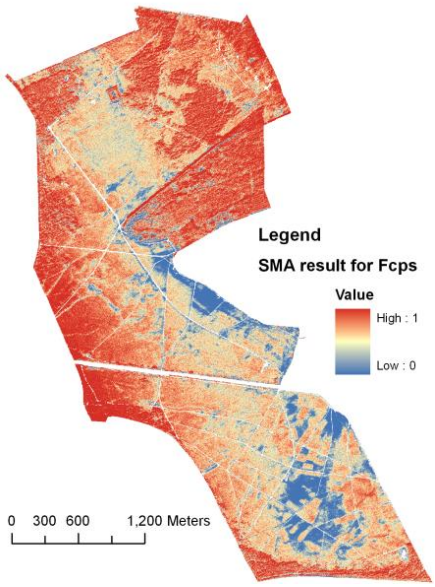


Figure 6.7 Thematic fraction images for the endmembers of a) *Calluna*-dominated heatland (Hdca), b) *Molinia*-dominated heathland (Hgmd) and c) coniferous forest, mainly Scots pine (Fcps), obtained by applying a spectral mixture analysis (SMA) to a AHS hyper-spectral image acquired in October 2007

5.4 Segmentation and feature extraction from hyperspectral data.

For segmentation prior to classification of hyper-spectral imagery, the algorithm of Shepherd *et al*. (2012) was used. Using multi-spectral WV-2 data, the segmentation was based initially on reflectance bands but improvements were found for Cors Fochno using primary indices including the NDVI and NDWI (all dates) and the PSRI (March and November, when vegetation was senescent). A trial segmentation was also conducted on the MNF bands of the EAGLE dataset rather than the full range of bands, with the number of seeds ranging from 30 to 120 (Figure 6.8a-d). For the active bog area, the finer detail provided by the 30-seed classification was considered to best represent the diversity of habitats occurring whilst a coarser segmentation was considered to be more appropriate for the more homogeneous areas (e.g., cultivated fields). Within the EODHAM system, the recommendation is that segmentation of hyper-spectral data should be conducted to allow classification of the broader LCCS

categories (in EODHAM 1st and 2nd stages) and then a more detailed segmentation (i.e., to the pixel or small object level) undertaken to allow spectral unmixing procedures to be performed.

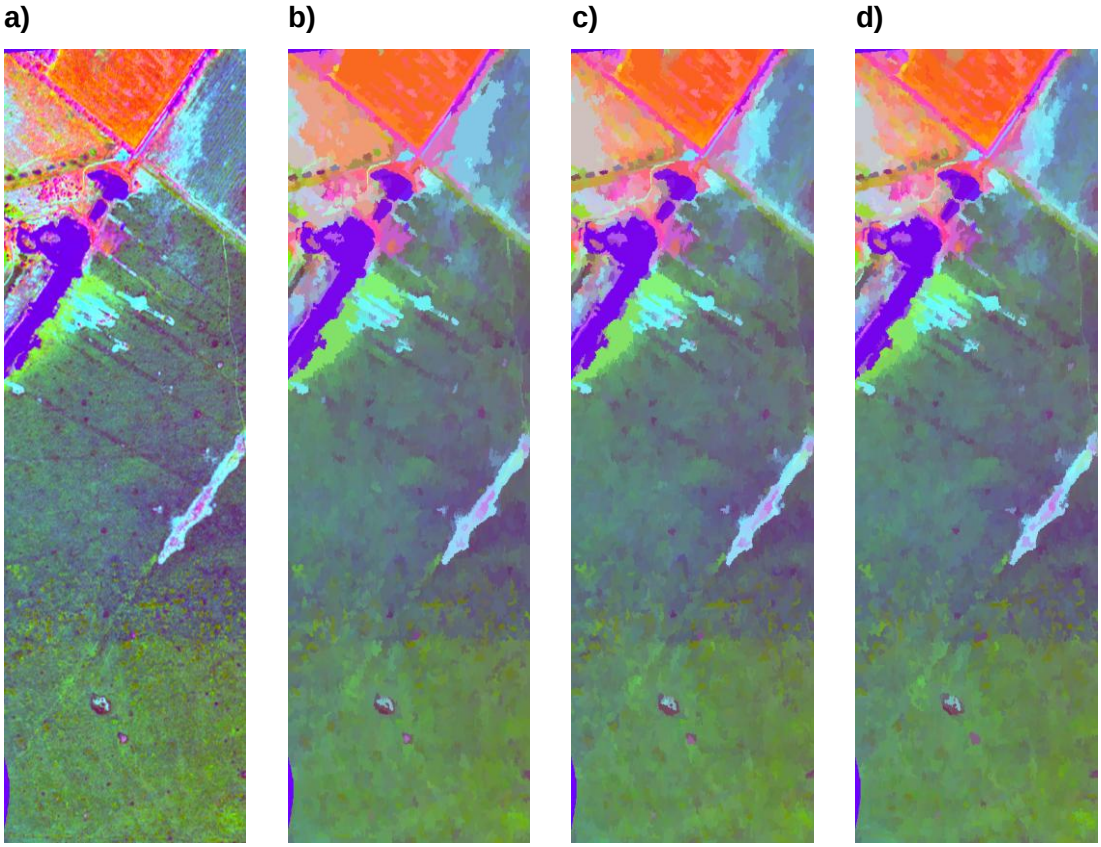


Figure 6.8a) A composite of the Minimum Noise Fraction (MNF) bands derived from the EAGLE hyper-spectral dataset segmented with the algorithm of Shepherd *et al*. (2012) and eliminated to a) 30 pixels (i.e. minimum unit size) with b) 50, c) 60 and d) 120 seeds.

The use of hyperspectral data for extracting key features of the landscape is ongoing as the exercise is being undertaken in parallel with that of the VHR data (to be reported in D5.5). Key areas where hyper-spectral data are likely to benefit the segmentation of the image include the delineation of tree crowns (Bunting and Lucas, 2006; based on ratios of the red edge and red channels) and in defining urban features (e.g., buildings with different roofing materials).

5.5 Analysis of SWIR data, Le Cesine.

For Le Cesine, classifications of LCCS categories (to Level 2) were generated using both the WV-2 and Quickbird images, without and with the MIVIS data (Figure 6.9), with particular focus on evaluating the use of the SWIR for identifying aquatic (emergent) vegetation. According to the EODHaM 1st stage DFD proposed in D5.3, 1st order texture along with specific indices were used to identify the intermediate layer of emergent aquatic vegetation. This produced an underestimate of emergent vegetation, as evidenced in the confusion matrix of Table 3.6 in D5.3. The hyper-spectral Index 1, as described in Table 2.3, was therefore introduced in a combined analysis of MIVIS and multi-spectral data following co-registration at a spatial resolution of 3 m. Index 1 was used with the spectral indices and the 1st order texture measure used in D5.3 for Le Cesine but the threshold of the 1st order texture was also changed in order to capture a greater number of objects considered to represent emergent aquatic vegetation.

The selection of Index 1 from MIVIS data was based on an evaluation of the contribution of all other MIVIS indices given in Table 2.3 in combination with the 1st order texture measure and spectral indices from QuickBird and WV-2 images and the rules already used in D5.3. The finding was that Index1 performed better than other MIVIS indices for discriminating A24 categories of aquatic vegetation.

However, the selection of thresholds was critical, as demonstrated by Table 6.1a-b, with these reporting the classification accuracy and error tolerance obtained for the Le Cesine site at LCCS level 2. The number of reference samples was lower than the ones used in D5.3 for validation purposes (see Table 3.6 in D5.3) since the MIVIS image covered a smaller area than the QuickBird and WorldView2 images. The comparison indicated that whilst the overall classification accuracy was lower, the classification of aquatic vegetation was more consistent.

a) b)

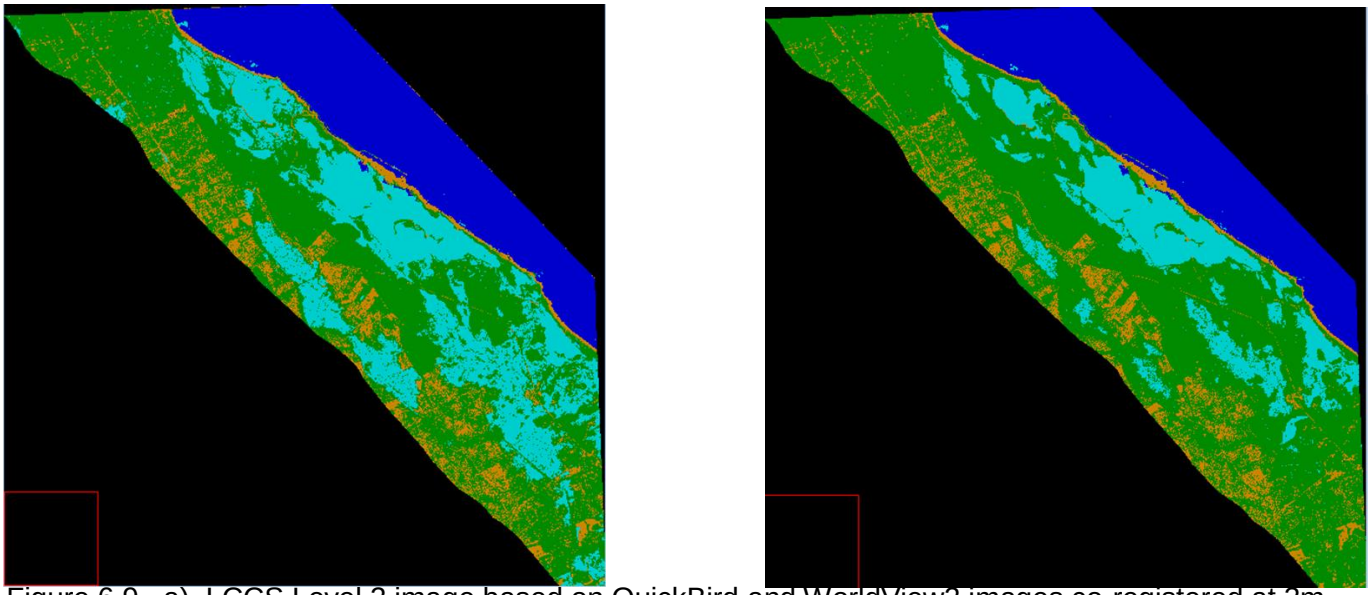


Figure 6.9. a) LCCS Level 2 image based on QuickBird and WorldView2 images co-registered at 3m spatial resolution. b) LCCS level 2 image based on the combined use of indices from MIVIS, QuickBird and WorldView2 co-registered images.

Table 6.1.a Validation by reference random polygons for LCCS Level 2 categories based on a) QuickBird June 2009 and Worldview2 October 2010 images and b) including MIVIS.

a)

Ground Truth Fields

Satellite-based Classification	A1 TERR VEG	A2 AQU VEG	B1 TERR NON-VEG	B2 AQU NON-VEG	Number	User's Accuracy (%)
A1 TERR VEG	144	22	4	0	170	84.70
A2 AQU VEG	0	59	0	0	59	100
B1 TERR NON-VEG	0	0	34	0	34	100
B2 AQU NON-VEG	0	0	0	37	37	100
Number	144	81	38	37		
Producer's Accuracy (%)	100	72.83	89.47	100		
				Overall (%)	300	91.33
				Delta		3.18

b)

	Ground Truth Fields					
Satellite-based Classification	**A1 TERR VEG**	**A2 AQU VEG**	**B1 TERR NON-VEG**	**B2 AQU NON-VEG**	**Number**	**User's Accuracy (%)**
A1 TERR VEG	122	12	4	0	138	88.40
A2 AQU VEG	22	69	0	0	91	75.82
B1 TERR NON-VEG	0	0	34	0	34	100
B2 AQU NON-VEG	0	0	0	37	37	100
Number	144	81	38	37		
Producer's Accuracy (%)	84.72	85.18	89.47	100		
				Overall	300	87.33
				Delta		3.76

6 Discussion

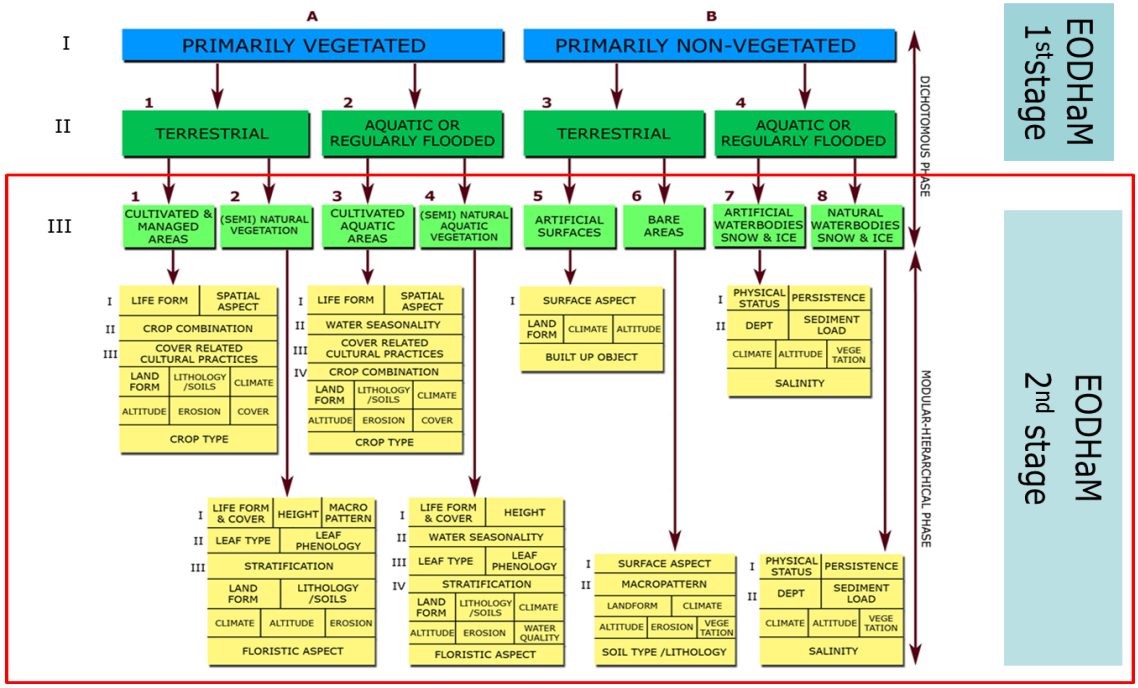
The EODHaM system has been designed to ingest optical data for the classification of land covers according to the FAO LCCS scheme. The following sections discuss the benefits of integrating hyper-spectral data for land cover and habitat classification within the EODHaM system. The practicalities of using hyper-spectral data as well as quality indicators are conveyed. The relative benefits of hyper-spectral data over multi-spectral data are also outlined, with consideration given to the information gains through inclusion of multi-date imagery.

6.1 Inclusion of hyper-spectral data within EODHaM

Hyper-spectral data can be used to generate classifications of LCCS categories within the EODHaM 1st and 2nd stages (Figure 7.1) in much the same way as multi-spectral data (see D5.3). More specifically:

a) For discrimination of vegetated from non-vegetated and aquatic from terrestrial categories in the EODHaM 1st stage, hyper-spectral data provide a wider range of indices and allow more targeted generation of endmember fractions (e.g., representing photosynthetic and non-photosynthetic components of vegetation) for use within rule-based classifications.
b) In the EODHaM 2nd stage (LCCS Level 3), contextual rules applied to multi-spectral data can be similarly used with hyper-spectral data, although some features may be better delineated (e.g., tree crowns, using ratios of specific narrow wavebands within the red edge and red channels; Bunting and Lucas, 2006).
c) In the EODHaM 2nd stage (Beyond LCCS Level 3), single-date hyper-spectral data provide more options for discrimination of categories within the layers associated with plant life forms (namely leaf type, phenology, and cover), and non-plant life forms, such as bare categories (materials), water (sediment loads, subsurface material and depth) and sediment loads.
d) Multi-temporal hyper-spectral, as with multi-spectral, data, can be used to map water duration and dynamics and crop seasonality patterns although multi-temporal optical and SAR spaceborne data are more useful given the greater frequency of observation.

a)



b)

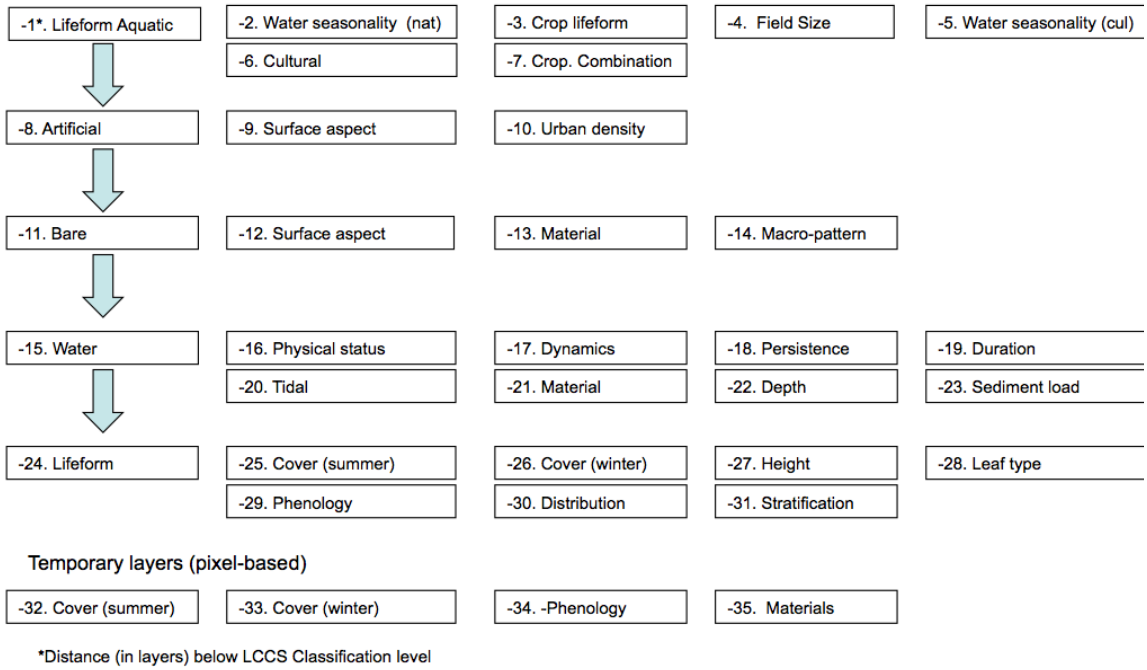


Figure 7.1a) The LCCS classification scheme and its context within the EODHaM 1st and 2nd stage. b) Layers generated in the EODHAM 2nd stage for classification of LCCS categories.

Whilst the benefits of hyper-spectral over multi-spectral data are relatively marginal for land cover classifications within the FAO LCCS, they can be used to provide more information on the distribution of dominant species or genera comprising habitats, including those that are rare, endangered or of limited extent,. For example, in Cors Fochno, several *Sphagnum* species could be discriminated from other plant species within the active bog with these being important indicators of habitat condition. However, if species are limited in extent or occur in the understorey, then discrimination is likely to be difficult if not unlikely without inference based on site preferences. Furthermore, these data can allow discrimination and mapping of invasive or exotic species, with these considered a threat to most of the BIOSOS sites,

and particularly in Portugal where Australian *Acacia* and *Eucalyptus* species are displacing other habitats, as well as in the BIO_SOS Indian sites where *Lantana camara* and *Chromolaena odorata* represent invasive species of major conservation significance. However, the use of hyper-spectral data is likely to be very site specific and focused on only a few species. In some cases, the ability to discriminate species can be assessed *a priori* on the basis of field spectra acquired at specific times and through application of separability techniques (e.g., linear discriminant analysis).

6.2 Practicalities in the use of hyper-spectral data

Experience with these data within the BIOSOS project has highlighted a number of limitations, particularly when acquired by airborne sensors (as in the case of Cors Fochno, Eder and Ginkelseheide and Le Cesine), with these listed below:

a) Variable levels of geometric accuracies, with these often inconsistent within each flightline and often reducing towards the outer edges.
b) Difficulty in atmospheric correction in the absence of ground targets and associated reflectance measurements.
c) Requirements for spectral smoothing and channel removal algorithms to reduce noise in spectra.
d) The cost of acquisition over larger areas and in several points in time.
e) Limb brightening and other variations/gradients in illumination, within images and differences between images when combined into a mosaic.
f) Lower image reflectance compared to ground-based measurements, with this compromising the use of the latter spectra in atmospheric correction and spectral retrieval.
g) Different sensors (although often on the same platform) used to acquire data in the visible/NIR and SWIR wavelength regions, with the latter often compromised by noise and artefacts.
h) Bidirectional effects across and between images when combined into a mosaic.
i) Difficulties in atmospheric correction using techniques such as the Empirical Line or atmospheric models (e.g., ATCOR, HYCORR).
j) Complex algorithms for retrieval of biophysical properties and spectra, particularly vegetation, and subsequent classification of imagery.
k) Requirements for validation at the time of the overpass.
l) Permission from authorities to acquire data or because of proximity of sites to airports.

Many of these limitations are specific to the fact that most hyper-spectral imagery is currently available through airborne platforms, and the increasing availability of spaceborne hypersectral data in years to come may reduce many of these.

Specific limitations within the BIOSOS project for Cors Fochno and the Veluwe in the Netherlands were that the hyper-spectral data had been acquired on only one occasion in 2009 (during the summer, June) and in 2007 (October) respectively. Atmospheric correction of the datasets resulted in image spectra that varied depending on the algorithm used (e.g., Empirical Line, ATCOR) and the geometric accuracy varied inconsistently with the swath. Whilst field data had been collected at the time of the overflights, these were limited only to a small area of the active bog, and were largely collected for different *Sphagnum* species (in the case of Cors Fochno). The ground dataset was therefore also limited for assessing the accuracy of classification for other plant groups and non-vegetation. Whilst ground spectra were collected, these were specific to the time and period of the overflights. Despite the difficulty in obtaining sufficient and reliable ground truth, the UAV data provide a highly valuable source of information on the distribution of plant species across the active raised bog at Cors Fochno and the future use of this technology is advocated.

On this basis, several recommendations are made when considering the acquisition and use of hyper-spectral data as input to the EODHaM system. In terms of flight planning, these include:

a) A direction flight into and away from the sun to avoid bidirectional effects and facilitate generation of seamless mosaics in subsequent processing. For multiple acquisitions, the aircraft should ideally fly in the same direction to reduce deviations associated with wind effect.

b) Flights occurring near to the solar noon and where the sun elevation is sufficient to avoid shadow within the image, from terrain but also tall vegetation.
c) An overlap between flightlines of at least 20-30%.
d) Sufficiently slow flight speed to allow sufficient dwell time for energy to be recorded.
e) Acquisitions taken under clear sky (normally late morning or around noon) or, to reduce shadowing and variations in illumination, full cloud.
f) Provision of appropriate ground and navigational control to ensure a high level of geometric correction, particularly when ground spectra are collected from small targets.

In addition, the following is recommended:

a) Acquisition of airborne LiDAR at the same time as that of hyper-spectral data acquisition to assist with geometric correction.
b) Multi-temporal acquisitions to capture the seasonal variation (in relation to phenology)
c) Collection of ground reflectance spectra from targets (bright, dark and intermediate) located in the area of the swath and at the same time as the overpass.
d) Collection of sufficient ground data relating to vegetation and other surface materials, with these targeted to the classes of interest in subsequent classifications, retrieval of biophysical attributes (e.g., proportion of non-photosynthetic material) or identification of species (e.g., indicators, as required by the BIOSOS project).
e) Inclusion of UAV data at very high resolution to allow interpretation of hyper-spectral and also multi-spectral images and validation of output products.

A number of Quality Indicators required for inclusion of hyper-spectral data within the EODHaM system are:

a) Robust atmospheric correction of data (to within a few percent of ground measurements) using either techniques such as the Empirical Line or atmospheric models.
b) A high level of geometric accuracy (< 1 pixel) throughout the swath to facilitate reliable retrieval of spectra, particularly in complex environments and when comparing against ground based spectra or measurements (e.g., of plant community composition).
c) Minimisation of bidirectional reflectance across the scan within minimal differences across the swath for homogenous surfaces.
d) Provision of ground calibration targets which are of sufficient size (at least greater than 2x the pixel resolution) and appropriate material (i.e., uniform dark and bright reflectance across the spectral range) to support or improve atmospheric correction.
e) Acquisition at a spatial resolution that is appropriate to the size of targets being considered (land cover classes, habitats, vegetation communities, tree crowns, etc). Thus, if airborne platforms are used to collect data, the spatial resolution should be maintained as high as possible and in accordance with the spatial scale of critical objects being classified in the landscape.

6.3 Comparisons with VHR optical date: single date and time-series

Whilst the benefits of hyper-spectral data have been conveyed, D5.3 and D5.4 have collectively indicated that equivalent or better classifications can be obtained using multi-temporal WV-2 data acquired during periods of peak senescence and biomass and ideally and additionally, during transition periods (i.e., spring and autumn). In terms of single-date acquisitions, the hyper-spectral data would provide advantages over multi-spectral date if appropriately corrected. However, multi-temporal acquisitions of multi-spectral VHR data are considered to provide better discrimination as long as the timing of acquisitions occurs during periods of maximum separability between target communities. In terms of retrieval of endmember fractions (e.g., PV and NPV), these broadly correspond to indices generated from either hyper-spectral or multi-spectral data. Comparisons between similar measures (fractions, indices) derived from multi-spectral and hyper-spectral data were difficult because of differences in the years and dates of observations.

6.4 Procedures inclusion of hyper-spectral data within the EODHaM system

A summary of procedures to integrate hyper-spectral data within the EODHaM system is given in Figure 7.2. Airborne sensors will have acquired the majority of hyper-spectral data, although the potential exists in the future for provision of spaceborne hyper-spectral data. In each case, modules within the EODHaM system need to include atmospheric and geometric correction procedures, although often these are site or sensor specific. For this reason, the provision of data that is corrected prior to ingestion by the system is desirable, with the EODHaM focusing on ensuring correction and calibration based on comparison with reference spectra for known materials. For this purpose, the system will require reference to standard spectra or those collected in the field for known targets (e.g., black or white tarpaulins) or surfaces (e.g., sand, asphalt).

A key component is the extraction of endmembers (NPV, PV etc.) and calculation of key indices (e.g., NDVI, PSRI or NDWI) for classification of vegetated and aquatic surfaces in the EODHaM 1st stage. A correspondence between specific indices and endmember fractions has been observed, suggesting that either might be used within the EODHaM system. However, this may not always be the case. For example, the SMA fraction for live green vegetation was found to correlate better than NDVI with ground-measured percentage cover in a Californian semi-arid environment (Elmore *et al*., 2000). A greater diversity of endmembers and indices are also likely to be needed for specific descriptions of life forms, materials and sediment load. Within the EODHaM 3rd stage, and following translation of LCCS categories to life forms and associated GHCs, a finer segmentation of the landscape is advocated with this aimed at achieving the following:

a) Delineation of features, such as tree crowns and marginal aquatic vegetation and subsequent application of discriminant procedures (e.g., linear discriminant analysis) for differentiating species/genera.
b) Spectral unmixing to determine relative proportions of, for example, NPV or PV which can be used subsequently in a rule-based approach to classifying GHCs and the species comprising these or different species directly.

In the final stage, the accuracy of the classifications needs to be assessed through independent field samples, with the requirement that these are representative (in terms of number and area) of the species, genera or habitats of interest. The accuracy of classifications will also depend on spatial scale, with higher accuracy potentially achieved at the level of discrimination of land cover classes and/or GHCs, and greater challenges posed in the discrimination of finer spatial features such as species assemblages. Therefore, the spatial resolution of the image datasets, and the spatial scale of the objects of interest in the classification, will play a major role in impacting the accuracy of classification outcomes. The use of UAV data provided considerable benefits for interpreting and validating the hyper-spectral and multi-spectral classifications of Cors Fochno and its wider use is advocated, particularly if acquisitions are on the same or similar dates.

6.5 Overview

Deliverable 5.4 has provided a realistic assessment of the role of hyper-spectral data within the EODHaM system, recognising the limitations associated with limited data acquisitions (both spatially and temporally) and the complexity of pre-processing, interpreting, retrieving useful information and classifying these data. Comparison with multi-temporal WV-2 data for Cors Fochno in Wales suggests that multi-temporal hyperspatial data can provide equivalent if not better information for discriminating habitats than hyper-spectral sensors (based on the study from Cors Fochno), unless temporal hyper-spectral data can be acquired. However, the usefulness of hyper-spectral data can be increased by acquiring data at times of year known to be optimal for discrimination of target species, although this can be challenging when there are multiple species of interest in a landscape, as is frequently the case in most landscapes of conservation interest.

For the mapping of LCCS categories in Levels 1-2 (EODHaM 1st stage) and 3 and beyond (EODHaM 2nd stage) can be achieved using hyper-spectral data and, in some cases, to a better level of accuracy compared to multi-spectral data, these data are best used for the detection of specific species, genera or

habitat which are difficult to identify because of similarities in the reflectance of other habitats or because of their relatively localised extent. A key role seen for these data is the identification of invasive or exotic species, which often exhibit spectral characteristics that differ from that of native species, and which present for instance particular problems for conservation in BIO_SOS sites in Portugal and India. The endmember fractional images and vegetation indices derived from hyper-spectral sensors, and exploited in EODHaM are similar to those able to be generated from multi-spectral data although more spectral (e.g., absorption) features specific to species or conditions of habitats can be exploited.

A number of tools are available for the processing of hyper-spectral data. Unmixing methods that are more likely to be adopted are expected to include MESMA and SMA because of their availability in commercial software packages, such as ENVI, where there are freely available mixture modelling extensions such as VIPER tools. However, routines available within the ORFEO Toolbox are open source and their integration within the EODHAM system is advocated.

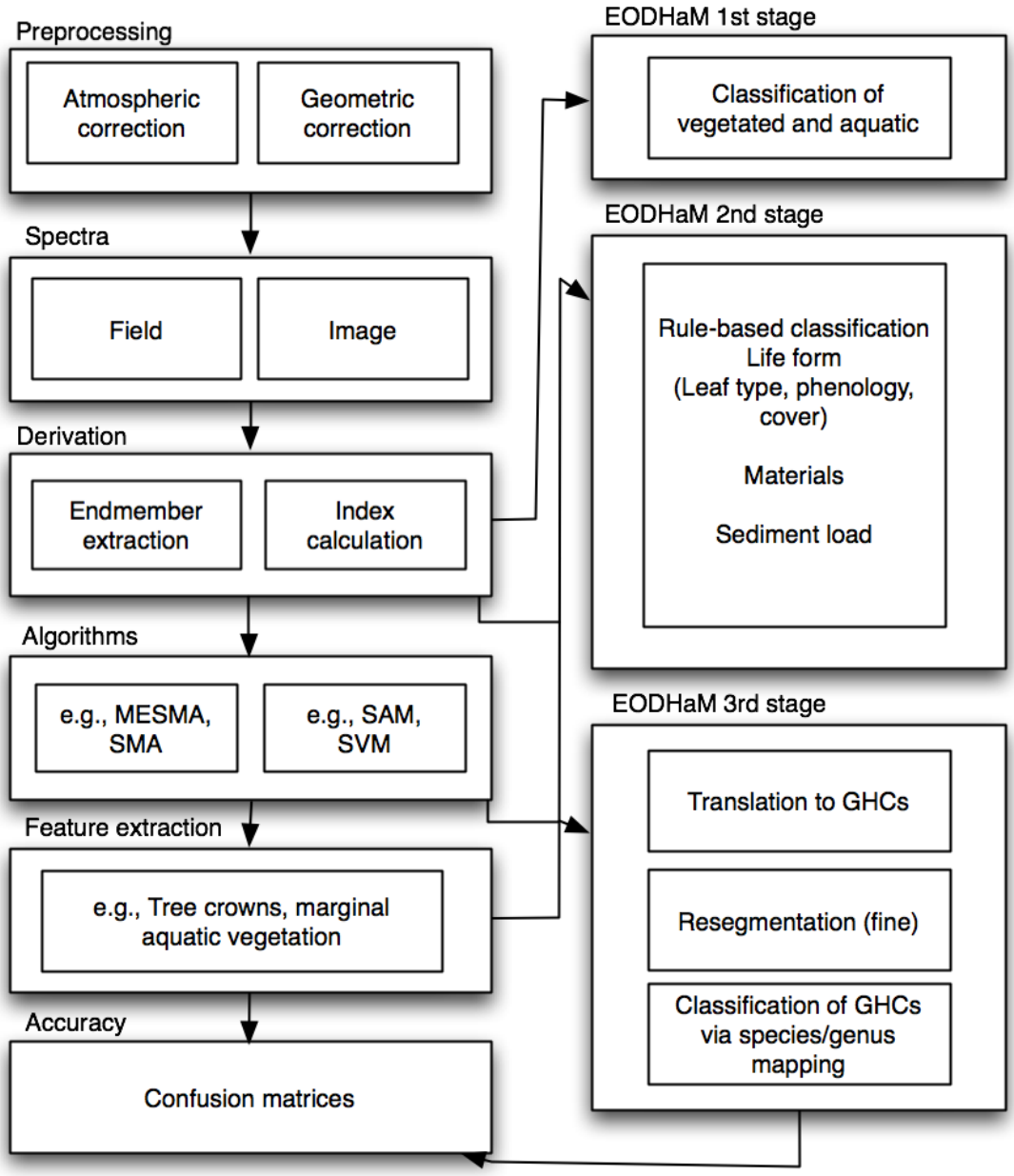


Figure 4.12. Approach to integration of hyper-spectral data within the EODHaM system

Based on the analyses conducted through D5.4, the following recommendations can be made.

a) Whilst hyper-spectral data can be used to map the land covers associated with FAO LCCS, this role can equally or perhaps be even better performed using VHR data, particularly if time-series are acquired. This is mainly because the cost and technical challenges of acquiring, processing and potentially automating the use of hyper-spectral data are substantial and considerably higher than that associated with spaceborne VHR datasets.

b) The analysis of hyper-spectral data often requires reference to field data acquired at the time of the sensor overflight (including spectral reflectance measurements of calibration targets) and hence introduces additional costs and time constraints that bring significant challenges to the concept of pre-operationalization of automated expert-guided classification approaches, as envisaged by BIO_SOS. Thus, whilst hyperspectral approaches may be useful in certain contexts, we do not recommend their automatic incorporation, and instead suggest that these can be used in specific locations where the discrimination of certain land cover classes and/or habitats make the additional time, cost and technical challenges associated with utilization of these images worthwhile in terms of outcomes.

7 References

ADAMS, J. B., SABOL, D. E., KAPOS, V., ALMEIDA FILHO, R., ROBERTS, D. A., SMITH, M. O. & GILLESPIE, A. R. 1995. Classification of multispectral images based on fractions of endmembers: Application to land-cover change in the Brazilian Amazon. *Remote Sensing of Environment,* 52**,** 137-154.

ADLER-GOLDEN, S. M., MATTHEW, M. W., BERNSTEIN, L. S., LEVINE, R. Y. A. B., RICHTSMEIER, S. C., ACHARYA, P. K., ANDERSON, G. P., FELDE, G., GARDNER, J., HOKE, M., JEONG, L. S., PUKALL, B., RATKOWSKI, A. & BURKE, H.-H. Atmospheric Correction for Short-wave Spectral Imagery Based on MODTRAN4. SPIE Proc. Imaging Spectrometry, 1999. 61-69.

ANDREW, M. E. & USTIN, S. L. 2008. The role of environmental context in mapping invasive plants with hyperspectral image data. *Remote Sensing of Environment,* 112**,** 4301-4317.

BEGIEBING, S. & BACH, H. 2004. Analyses of hyperspectral and directional CHRIS data for agricultural monitoring using a canopy reflectance model. *In:* EUROPEAN SPACE AGENCY (SPECIAL PUBLICATION) ESA SP. 578 ED. (ed.).

BIN, L. & CHANUSSOT, J. Unsupervised classification of hyperspectral images by using linear unmixing algorithm. Image Processing (ICIP), 2009 16th IEEE International Conference on, 7-10 Nov. 2009 2009. 2877-2880.

BOARDMAN, J. W., KRUSE, F. A. & GREEN, R. O. 1995. Mapping target signatures via partial unmixing of AVIRIS data. *Summaries of the Fifth Annual JPL Airborne Geoscience Workshop.* Jet Propulsion Laboratory, Pasadena, CA.

BUNCE, R., METZGER, M., JONGMAN, R., BRANDT, J., DE BLUST, G., ELENA-ROSSELLO, R., GROOM, G., HALADA, L., HOFER, G., HOWARD, D., KOVÁŘ, P., MÜCHER, C., PADOA-SCHIOPPA, E., PAELINX, D., PALO, A., PEREZ-SOBA, M., RAMOS, I., ROCHE, P., SKÅNES, H. & WRBKA, T. 2008. A standardized procedure for surveillance and monitoring European habitats and provision of spatial data. *Landscape Ecology,* 23**,** 11-25.

BUNTING, P. & LUCAS, R. 2006. The delineation of tree crowns in Australian mixed species forests using hyperspectral Compact Airborne Spectrographic Imager (CASI) data. *Remote Sensing of Environment,* 101**,** 230-248.

CANTY, M. J., NIELSEN, A. A. & SCHMIDT, M. 2004. Automatic radiometric normalization of multitemporal satellite imagery. *Remote Sensing of Environment,* 91**,** 441-451.

CARTER, G. A. 1991. Primary and Secondary Effects of Water Content on the Spectral Reflectance of Leaves. *American Journal of Botany,* 78**,** 916-924.

CARTER, G. A. 1994. Ratios of leaf reflectances in narrow wavebands as indicators of plant stress. *International Journal of Remote Sensing,* 15**,** 697-703.

CHEIN, I. C. & QIAN, D. 2004. Estimation of number of spectrally distinct signal sources in hyperspectral imagery. *Geoscience and Remote Sensing, IEEE Transactions on,* 42**,** 608-619.

CHEN, X. & VIERLING, L. 2006. Spectral mixture analyses of hyperspectral data acquired using a tethered balloon. *Remote Sensing of Environment,* 103**,** 338-350.

CHRISTIAN, B. & KRISHNAYYA, N. S. R. 2009. Classification of tropical trees growing in a sanctuary using Hyperion (EO-1) and SAM algorithm. *Current Science,* 96**,** 1601-1607.

CLEVERS, J. G. P. W. & BUKER, C. Feasibility of the red edge index for the detection of nitrogen deficiency. Physical measurements and signatures in remote sensing. Proc. 5th international colloquium, 1991 Courchevel. 165-168.

CLEVERS, J. G. P. W. & KOOISTRA, L. 2012. Using Hyperspectral Remote Sensing Data for Retrieving Canopy Chlorophyll and Nitrogen Content. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 5**,** 574-583.

COHEN, W. B., MAIERSPERGER, T. K., GOWER, S. T. & TURNER, D. P. 2003. An improved strategy for regression of biophysical variables and Landsat ETM+ data. *Remote Sensing of Environment,* 84**,** 561-571.

COLOMBO, R., MERONI, M., MARCHESI, A., BUSETTO, L., ROSSINI, M., GIARDINO, C. & PANIGADA, C. 2008. Estimation of leaf and canopy water content in poplar plantations by means of hyperspectral indices and inverse modeling. *Remote Sensing of Environment,* 112**,** 1820-1834.

COOLEY, T., ANDERSON, G. P., FELDE, G. W., HOKE, M. L., RATKOWSKI, A. J., CHETWYND, J. H., GARDNER, J. A., ADLER-GOLDEN, S. M., MATTHEW, M. W., BERK, A., BERNSTEIN, L. S., ACHARYA, P. K., MILLER, D. & LEWIS, P. FLAASH, a MODTRAN4-based atmospheric correction algorithm, its application and validation. Geoscience and Remote Sensing Symposium, 2002. IGARSS '02. 2002 IEEE International, 2002 2002. 1414-1418 vol.3.

COOPS, N., STONE, C., CULVENOR, D. S. & CHISHOLM, L. 2004. Assessment of Crown Condition in Eucalypt Vegetation by Remotely Sensed Optical Indices. *Journal of Environmental Quality,* 33**,** 956–964.

CROWTHER, K. & GROOME, G. 2004.Monitoring of Raised Mire Vegetation at Cors Fochno CSAC.636b The Countryside Council for Wales.

CURRAN, P. J. 1989. Remote sensing of foliar chemistry. *Remote Sensing of Environment,* 30**,** 271-278.

CURRAN, P. J., WINDHAM, W. R. & GHOLZ, H. L. 1995. Exploring the relationship between reflectance red edge and chlorophyll concentration in slash pine leaves. *Tree Physiology,* 15**,** 203-206.

DATT, B. 1999. A New Reflectance Index for Remote Sensing of Chlorophyll Content in Higher Plants: Tests using Eucalyptus Leaves. *Journal of Plant Physiology,* 154**,** 30-36.

DELL'ACQUA, F., GAMBA, P., FERRARI, A., PALMASON, J. A., BENEDIKTSSON, J. A. & ARNASON, K. 2004. Exploiting spectral and spatial information in hyperspectral urban data with high resolution. *Geoscience and Remote Sensing Letters, IEEE,* 1**,** 322-326.

DENNISON, P. E., HALLIGAN, K. Q. & ROBERTS, D. A. 2004. A comparison of error metrics and constraints for multiple endmember spectral mixture analysis and spectral angle mapper. *Remote Sensing of Environment,* 93**,** 359-367.

DENNISON, P. E. & ROBERTS, D. A. 2003. Endmember selection for multiple endmember spectral mixture analysis using endmember average RMSE. *Remote Sensing of Environment,* 87**,** 123-135.

DRAY, S. & LEGENDRE, P. 2008. TESTING THE SPECIES TRAITS–ENVIRONMENT RELATIONSHIPS: THE FOURTH-CORNER PROBLEM REVISITED. *Ecology,* 89**,** 3400-3412.

ELMORE, A. J., MUSTARD, J. F., MANNING, S. J. & LOBELL, D. B. 2000. Quantifying Vegetation Change in Semiarid Environments: Precision and Accuracy of Spectral Mixture Analysis and the Normalized Difference Vegetation Index. *Remote Sensing of Environment,* 73**,** 87-102.

EXEGIS. 2012. *Air Photo Service - Technical details* [Online]. U.K. Available: http://www.esdm.co.uk/aps-technical-details.asp [Accessed 6th July 2012].

GAMON, J. A., PEÑUELAS, J. & FIELD, C. B. 1992. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sensing of Environment,* 41**,** 35-44.

GAMON, J. A., SERRANO, L. & SURFUS, J. S. 1997. The photochemical reflectance index: an optical indicator of photosynthetic radiation use efficiency across species, functional types, and nutrient levels. *Oecologia,* 112**,** 492-501.

GITELSON, A. & MERZLYAK, M. N. 1994. Spectral Reflectance Changes Associated with Autumn Senescence of Aesculus hippocastanum L. and Acer platanoides L. Leaves. Spectral Features and Relation to Chlorophyll Estimation. *Journal of Plant Physiology,* 143**,** 286-292.

GITELSON, A. A., KAUFMAN, Y. J. & MERZLYAK, M. N. 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment,* 58**,** 289-298.

GREEN, A. A., BERMAN, M., SWITZER, P. & CRAIG, M. D. 1988. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *Geoscience and Remote Sensing, IEEE Transactions on,* 26**,** 65-74.

GUERSCHMAN, J. P., HILL, M. J., RENZULLO, L. J., BARRETT, D. J., MARKS, A. S. & BOTHA, E. J. 2009. Estimating fractional cover of photosynthetic vegetation, non-photosynthetic vegetation and bare soil in the Australian tropical savanna region upscaling the EO-1 Hyperion and MODIS sensors. *Remote Sensing of Environment,* 113**,** 928-945.

HALLIGAN, K. 2011. *VIPER Tools, Open Source Spectral Mixture Analysis software* [Online]. Available: http://www.vipertools.org/?q=content/about-viper-tools [Accessed 13th March 2012].

HAMADA, Y., STOW, D. A. & ROBERTS, D. A. 2011. Estimating life-form cover fractions in California sage scrub communities using multispectral remote sensing. *Remote Sensing of Environment,* 115**,** 3056-3068.

HE, K. S., ROCCHINI, D., NETELER, M. & NAGENDRA, H. 2011. Benefits of hyperspectral remote sensing for tracking plant invasions. *Diversity and Distributions,* 17**,** 381-392.

HEINZ, D., CHANG, C.-I. & ALTHOUSE, M. L. G. Fully constrained least-squares based linear unmixing [hyperspectral image classification]. IEEE International Geoscience and Remote Sensing Symposium, 1999. 1401-1403.

HESTIR, E. L., KHANNA, S., ANDREW, M. E., SANTOS, M. J., VIERS, J. H., GREENBERG, J. A., RAJAPAKSE, S. S. & USTIN, S. L. 2008. Identification of invasive vegetation using hyperspectral remote sensing in the California Delta ecosystem. *Remote Sensing of Environment,* 112**,** 4034-4047.

HUETE, A., DIDAN, K., MIURA, T., RODRIGUEZ, E. P., GAO, X. & FERREIRA, L. G. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment,* 83**,** 195-213.

HUNT JR, E. R. & ROCK, B. N. 1989. Detection of changes in leaf water content using Near- and Middle-Infrared reflectances. *Remote Sensing of Environment,* 30**,** 43-54.

HUNTER, P. D., GILVEAR, D. J., TYLER, A. N., WILLBY, N. J. & KELLY, A. 2010. Mapping macrophytic vegetation in shallow lakes using the Compact Airborne Spectrographic Imager (CASI). *Aquatic Conservation-Marine and Freshwater Ecosystems,* 20**,** 717-727.

ISIDORO, D., QUÍLEZ, D. & ARAGÜÉS, R. 2010. Drainage water quality and end-member identification in La Violada irrigation district (Spain). *Journal of Hydrology,* 382**,** 154-162.

JING, W. & CHEIN, I. C. 2006. Applications of Independent Component Analysis in Endmember Extraction and Abundance Quantification for Hyperspectral Imagery. *Geoscience and Remote Sensing, IEEE Transactions on,* 44**,** 2601-2616.

KOOISTRA, L., MÜCHER, C. A. & NIEWIADOMSKA, A. 2008.Monitoring of Natura 2000 sites using hyperspectral Remote Sensing.CGI Report: CGI-08-001 Wageningen University, the Netherlands.

KUTSER, T. 2004. Quantitative detection of chlorophyll in cyanobacterial blooms by satellite remote sensing. *Limnology and Oceanography,* 49**,** 2179-2189.

LE MAIRE, G., FRANÇOIS, C. & DUFRÊNE, E. 2004. Towards universal broad leaf chlorophyll indices using PROSPECT simulated database and hyperspectral reflectance measurements. *Remote Sensing of Environment,* 89**,** 1-28.

LI, L., USTIN, S. L. & LAY, M. 2005. Application of multiple endmember spectral mixture analysis (MESMA) to AVIRIS imagery for coastal salt marsh mapping: a case study in China Camp, CA, USA. *International Journal of Remote Sensing,* 26**,** 5193-5207.

LU, S., SHIMIZU, Y., ISHII, J., FUNAKOSHI, S., WASHITANI, I. & OMASA, K. 2009. Estimation of abundance and distribution of two moist tall grasses in the Watarase wetland, Japan, using hyperspectral imagery. *Isprs Journal of Photogrammetry and Remote Sensing,* 64**,** 674-682.

LUCAS, R., BUNTING, P., PATERSON, M. & CHISHOLM, L. 2008. Classification of Australian forest communities using aerial photography, CASI and HyMap data. *Remote Sensing of Environment,* 112**,** 2088-2103.

LUCAS, R., MEDCALF, K., BROWN, A., BUNTING, P., BREYER, J., CLEWLEY, D., KEYWORTH, S. & BLACKMORE, P. 2011. Updating the Phase 1 habitat map of Wales, UK, using satellite sensor data. *Isprs Journal of Photogrammetry and Remote Sensing,* 66**,** 81-102.

MCGWIRE, K., MINOR, T. & FENSTERMAKER, L. 2000. Hyperspectral Mixture Modeling for Quantifying Sparse Vegetation Cover in Arid Environments. *Remote Sensing of Environment,* 72**,** 360-374.

MELGANI, F. & BRUZZONE, L. 2004. Classification of hyperspectral remote sensing images with support vector machines. *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING,* 42.

MOUNTRAKIS, G., IM, J. & OGOLE, C. 2011. Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing,* 66**,** 247-259.

MUCHER, S. & KOOISTRA, L. 2011. Relationship between remote sensing classification success and site complexity. *7th EARSeL Workshop of the Special Interest Group in Imaging Spectroscopy.* Edinburgh.

MUSTARD, J. F. 1993. Relationships of soil, grass, and bedrock over the kaweah serpentinite melange through spectral mixture analysis of AVIRIS data. *Remote Sensing of Environment,* 44**,** 293-308.

NASARUDIN, N. E. M. & SHAFRI, H. Z. M. 2011. Development and Utilization of Urban Spectral Library for Remote Sensing of Urban Environment. *Journal of Urban and Environmental Engineering,* 5**,** 44-56.

NASCIMENTO, J. M. P. & DIAS, J. M. B. 2005. Vertex component analysis: a fast algorithm to unmix hyperspectral data. *Geoscience and Remote Sensing, IEEE Transactions on,* 43**,** 898-910.

NICHOL, J. & WONG, M. S. 2007. Remote sensing of urban vegetation life form by spectral mixture analysis of high-resolution IKONOS satellite images. *International Journal of Remote Sensing,* 28**,** 985-1000.

NIEMANN, K. O., LOOS, R., QUINN, G. S. & VISINTINI, F. Previsual detection of insect infestation using airborne imaging spectroscopy and LiDAR data 7th EARSeL Workshop of the Special Interest Group in Imaging Spectroscopy, 11-13th April 2011 Edinburgh.

OLDELAND, J., WESULS, D. & JÜRGENS, N. 2012. RLQ and fourth-corner analysis of plant species traits and spectral indices derived from HyMap and CHRIS-PROBA imagery. *International Journal of Remote Sensing,* 33**,** 6459-6479.

PARKER WILLIAMS, A. & HUNT JR, E. R. 2002. Estimation of leafy spurge cover from hyperspectral imagery using mixture tuned matched filtering. *Remote Sensing of Environment,* 82**,** 446-456.

PLAZA, A., MARTINEZ, P., PLAZA, J. & PEREZ, R. 2005. Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations. *Geoscience and Remote Sensing, IEEE Transactions on,* 43**,** 466-479.

POWELL, R. L. & ROBERTS, D. A. 2008. Characterizing Variability of the Urban Physical Environment for a Suite of Cities in Rondônia, Brazil. *Earth Interactions,* 12**,** 1-32.

ROBERTS, D., HALLIGAN, K. & DENNISON, P. 2007.VIPER Tools User Manual.Version 1.5

ROBERTS, D. A., DENNISON, P. E., GARDNER, M. E., HETZEL, Y., USTIN, S. L. & LEE, C. T. 2003. Evaluation of the potential of Hyperion for fire danger assessment by comparison to the Airborne Visible/Infrared Imaging Spectrometer. *Ieee Transactions on Geoscience and Remote Sensing,* 41**,** 1297-1310.

ROBERTS, D. A., SMITH, M. O. & ADAMS, J. B. 1993. Green vegetation, nonphotosynthetic vegetation, and soils in AVIRIS data. *Remote Sensing of Environment,* 44**,** 255-269.

ROLLIN, E. M. & MILTON, E. J. 1998. Processing of High Spectral Resolution Reflectance Data for the Retrieval of Canopy Water Content Information. *Remote Sensing of Environment,* 65**,** 86-92.

ROSSO, P. H., USTIN, S. L. & HASTINGS, A. 2005. Mapping marshland vegetation of San Francisco Bay, California, using hyperspectral data. *International Journal of Remote Sensing,* 26**,** 5169-5191.

SCHAEPMAN, M. E., USTIN, S. L., PLAZA, A. J., PAINTER, T. H., VERRELST, J. & LIANG, S. 2009. Earth system science related imaging spectroscopy—An assessment. *Remote Sensing of Environment,* 113, Supplement 1**,** S123-S137.

SIMS, D. A. & GAMON, J. A. 2002. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment,* 81**,** 337-354.

SWAIN, P. N. & DAVIS, S. M. 1978.Remote Sensing. The Quantitative Approach.

THEIS, F. J., STADLTHANNER, K. & TANAKA, T. First results on uniqueness of sparse non-negative matrix factorization. European Sig. Proc. Conf. (EUSIPCO), 2005.

UNDERWOOD, E., USTIN, S. & DIPIETRO, D. 2003. Mapping nonnative plants using hyperspectral imagery. *Remote Sensing of Environment,* 86**,** 150-161.

USTIN, S. L., ROBERTS, D. A., GAMON, J. A., ASNER, G. P. & GREEN, R. O. 2004. Using Imaging Spectroscopy to Study Ecosystem Processes and Properties. *BioScience,* 54.

VOGELMANN, J. E., ROCK, B. N. & MOSS, D. M. 1993. Red edge spectral measurements from sugar maple leaves. *International Journal of Remote Sensing,* 14**,** 1563-1575.

WALSH, S. J., MCCLEARY, A. L., MENA, C. F., SHAO, Y., TUTTLE, J. P., GONZALEZ, A. & ATKINSON, R. 2008. QuickBird and Hyperion data analysis of an invasive plant species in the. Galapagos Islands of Ecuador: Implications for control and land use management. *Remote Sensing of Environment,* 112**,** 1927-1941.

ZARCO-TEJADA, P. J., PUSHNIK, J. C., DOBROWSKI, S. & USTIN, S. L. 2003. Steady-state chlorophyll a fluorescence detection from canopy derivative reflectance and double-peak red-edge effects. *Remote Sensing of Environment,* 84**,** 283-294.

8 Appendix 1: Characteristics of hyper-spectral sensors acquired over the BIOSOS sites.

The MIVIS hyper-spectral sensor is a whisk-broom scanner with an axe head double mirror. It is a passive scanning and imaging instrument consisting of 4 spectrometers that simultaneously acquire radiation in the VNIR, SWIR and thermal infrared (TIR) (Table A.1). All channels were acquired and provided in units of radiance (μW/cm^2/sr/nm. for bands 1 to 91; VNIR to SWIR) or brightness temperature (°C. for bands 92 to 102; TIR). Data were converted to surface reflectance (%) or brightness temperature (^{o}C) and reprojected from Gauss-Boaga (Zone 2; Rome 1940) to UTM coordinates (Zone 33. WGS-84). The first 20 bands of the sensor were within the same wavelength region as the WV-2.

Table A.1. Spectral and radiometric characteristics of MVIS.

Channels (no)	Range (μm)	FWHM (nm)	IFOV (mrad)	Quantization accuracy
20	0.411-0.8019	20		
8	1.145 to 1.54	50	2.0	12 bit
64	1.992 to 2.474	9		
10	8.34 to 12.42	360		

102 spectral bands; IFOV of 2 mrad and digitized FOV of 71.1°; quantization (ADC) of 12 bits; scan rotational speed of 25. 16.7. 12.5. 8.3 and 6.25 scan s^{-1}; 2 reference black bodies selectable between 15°C and 45°C; a Position Attitude System (PAS) composed of a GPS receiver for measuring aircraft's position (accuracy 15-20 m) and speed (accuracy 0.05-0.20 m/sec) and a gyroscope for determining aircraft's roll and pitch (accuracy ±0.2°) with a roll correction in real time between ±10°; flux gate compass for finding aircraft's variation around the yaw axis (accuracy ±0.56°) and a computer-aided data quality check of all channels in real time.

Table A.2: MIVIS spectral ranges. The correspondence with the WV-2 bands is also indicated.

MVIS		
Bands	**Centre (nm)**	**FWHM (nm)**
1	441	20
2	460	20
3	480	20
4	500	20
5	521	20
6	541	20
7	561	20
8	581	20
9	601	20
10	621	20
11	641	20
12	661	20
13	681	20
14	701	20
15	721	20
16	740	20
17	760	20
18	779	20
19	798	20
20	819	20

Table A.3: Specifications of the AHS-160 sensor as operated by INTA

Item	**Characteristic**
Field of View	90˚
Instantaneous Field of View	2.5 mrad
No. of channels	80
No. of samples/lines	750
Scan principle	Whiskbroom
Scan frequency	Variable
Ground resolution	2.5 – 10 m

Radiometric resolution	12 bit

Table A.4: Band specifications of the AHS-160 sensor as operated by INTA

Optical port	Number of bands	Spectral region (nm)	Band width (nm)
1 - Visible and Near Infrared	20	430 – 1030	30
2A - Short wave Infrared	1	1550 – 1750	200
2B - Short wave Infrared	42	1994 – 2540	13
3 - Mid Infrared	7	3.3 – 5.4 µm	300
4 - Long Wave Infrared	10	8.2 – 12.7 µm	400

Table A.5: Geometric characteristics of AHS data of Eder Heide (NL)

Item	Characteristic
Sampling	
Line rate (lines per second)	35 rps
Pixel size	2.40294 x 2.402716 m
Resampling	Nearest Neighbour
Map projection	UTM. Zone 31 N
Geodetic Datum	WGS-84

CHRIS-PROBA (The Netherlands)

The CHRIS PROBA is a physically compact payload (weighing under 15 kg) and operates in the 'push-broom' mode3. The telescope in the CHRIS instrument is nadir pointing. CHRIS has the additional advantage of being relatively cheap and easy to manufacture since it has no moving parts. Its main applications are in environmental monitoring, forestry inventory and precision farming (Table A.6). The first version of the instrument was flown on the PROBA (Project for On Board Autonomy) platform, which carried the instrument in a sun-synchronous elliptical polar orbit. at a mean altitude of about 600 km. All parts of the Earth's surface were imaged when PROBA's across track pointing ability was used. From a 600 km orbit, CHRIS imaged the Earth in a 14 km swath with a spatial resolution of 18 m (this is somewhat variable as the altitude varies around the orbit). Using PROBA's agile steering capabilities in along and across track directions, observation of selectable targets well outside the nominal field of view of 1.3^{o} was achieved. Images were generally acquired in sets of 5, with these being taken at along track angles of +/- 55^{o}. +/- 36^{o} and as close to nadir as possible. CHRIS operated over the visible/near infrared band from 400 nm to 1050 nm, in 63 spectral bands at a spatial resolution of 36m and 18 bands

[3] http://www.chris-proba.org.uk/mission/instrument_characteristics.html and http://earth.esa.int/missions/thirdpartymission/proba.html

at full spatial resolution. The spectral sampling varied from 2-3 nm at the blue end of the spectrum. to about 12 nm at 1050 nm. Sampling was about 7 nm near the red edge (~690-740 nm). The instrument was very flexible and different sets of bands could be used for different applications (Table 5.7 and Figure 5.3). (http://eo.belspo.be/Directory/ProjectDetail.aspx?projID=824).

The CHRIS instrument was programmable. so that a variety of band selections were possible. For the mission, the number of different configurations was minimised. The principal configurations, based on a pre-launch wavelength calibration, are set out in Table A.6. The locations of the bands compared to a typical spectrum of green vegetation and to the spectrum of atmospheric absorption are shown in Figure A.1.

Table A.6: Specifications of the CHRIS-PROBA sensor

Item	**Characteristic**
Spatial sampling interval	18m on ground at nadir
Image area	14 km X 14 km (748 X 748 pixels)
Number of images	5 downloads at different view angles
Data per image (for a 14 X 14 km2)	131 Mbits
Spectral range	410nm to 1050 nm
Number of spectral bands	19 bands at a spatial resolution of 18m, 63 at 36m
Spectral resolution	1.3 nm @ 410nm to 12 nm @ 1050nm (i.e it varies across the spectrum)
Programmable operation	
Across track pixel size	18m or 36m
Along track pixel siz	Finest resolution is 18m but can be made coarser by changing the integration time
Spectral	Variable bandwidth and location
Digitisation	12 bits
Signal-to-noise ratio	200 @ a target albedo of 0.2

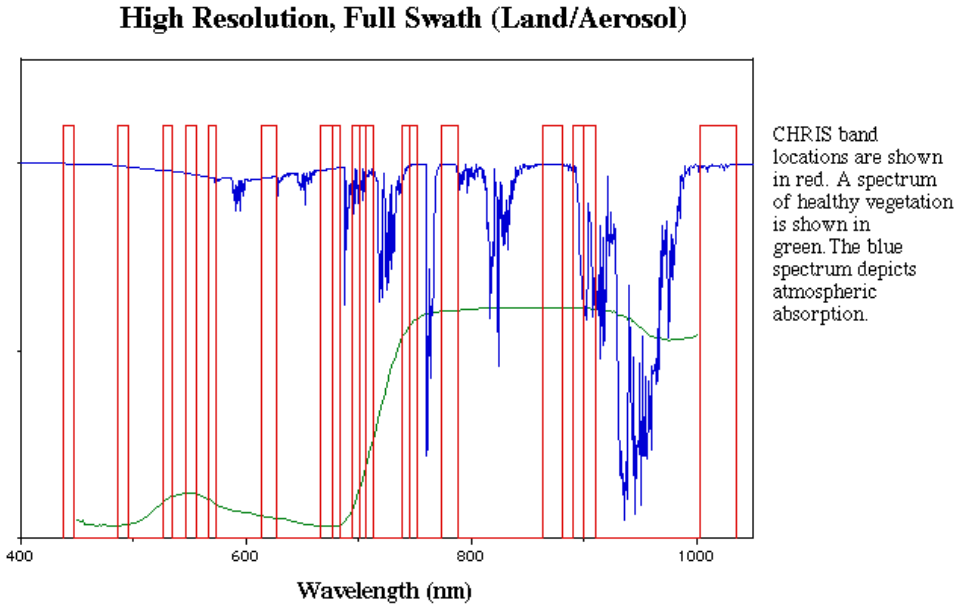


Figure A.1: Band positions of the CHRIS-PROBA sensor for MODE 3 compared to a vegetation (green) and atmospheric (blue) spectrum.

Table A.7: CHRIS PROBA bands For Full Swath.

High Resolution Land/Aerosols (MODE 3)

Band	Detectors	From	To	lmid	Width
L1	41-45	438.0	446.8	442.4	8.8
L2	65-68	485.6	494.8	490.2	9.2
L3	81-83	525.6	534.2	529.9	8.6
L4	88-90	546.4	556.1	551.2	9.7
L5	94-95	566.3	573.4	569.8	7.7
L6	109-110	626.6	636.0	631.2	9.4
L7	115-116	655.8	666.3	661.0	10.5
L8	117-118	666.3	677.2	671.7	10.9
L9	122	694.3	700.1	697.2	5.9
L10	123	700.2	706.2	703.2	6.0
L11	124	706.2	712.4	709.3	6.2
L12	129	738.2	745.0	741.6	6.8
L13	130	745.0	751.9	748.4	6.9
L14	134-135	773.3	788.4	780.8	15.0
L15	145-146	863.1	881.3	872.1	18.3
L16	148	890.7	900.2	895.4	9.5
L17	149	900.2	909.8	905.0	9.7
L18	159-161	1002	1035	1019.0	32.9

9 Acronyms

COB	Count-based Endmember Selection
CRES	Constrained Reference Endmember Selection
EAR	Endmember Average RMSE
ETM	Landsat 7 Enhanced Thematic Mapper
FCLS	Fully Constrained Least Square
GHCs	General Habitat Categories
GIS	Geographic Information System
GMES	Global Monitoring for the Environment and Security
HYMAP	
HR	High Resolution
ISRA	Image Space Reconstruction Algorithm
LC	Land Cover
LCC	Land Cover Change
MASA	Minimum Average Spectral Angle
MESMA	Multiple end-member spectral mixture analysis
MR	Medium Resolution
MRF	Markov Random Field
MS	Multi-Spectral
MF	Matched Filtering
MTMF	Mixture Tuned Matched Filtering
NDVI	Normalised DifferenceVegetation Index
NPV	Non Photosynthetic Vegetation
OBIA	Object Based Image Analysis
OTB	ORFEO Toolbox
PAR	Photosynthetically Active Radiation
PV	Photosynthetic Vegetation
RS imagery	Remote Sensed imagery
RS-IUS	Remote Sensing Image Understanding System

SAM	Spectral Angle Mapping
SAR	Synthetic Aperture Radar
SCI	Site of Community Importance
SDD	Service Design Document
SMA	Spectral Mixture Analysis
SPA	Special Protection Area
SPOT	Satellite Pour l'Observation de la Terre
SR	Spatial Resolution
SVM	Support Vector Machine
TM	Thematic Mapper
TOA	Top of Atmosphere
TOARF	Top of Atmosphere ReFlectance
TS-MRF	Tree-Structured Markov Random Field
VCA	Vertical Component Algorithm
VHR	Very High Resolution
VIPER	The Video Performance Evaluation Toolkit (or Video Performance Evaluation Toolkit)