The Earth Observation Data for Habitat Monitoring (EODHaM) system

Richard Lucas a, Palma Blonda b,*, Peter Bunting c, Gwawr Jones c, Jordi Inglada d, Marcela Arias d, Vasiliki Kosmidou e, Zisis I. Petrou e, Ioannis Manakos e, Maria Adamo b, Rebecca Charnock c, Cristina Tarantino b, Caspar A. Mücher f, Rob H.G. Jongman 1, Henk Kramer 1, Damien Arvor 1, João Pradinho Honrado h, Paola Mairota 1

a Centre for Ecosystem Sciences, School of Biological, Earth and Environmental Science, The University of New South Wales, High Street, Kensington, NSW 2052, Australia
b National Research Council – Institute of Intelligent Systems for Automation (CNR-ISSIA), Via G. Amendola 122, 70126 Bari, Italy
c Institute of Geography and Earth Sciences, Aberystwyth University, Aberystwyth, Ceredigion SY23 3DB, United Kingdom
d CESBIO (CNES/CNRS/UPS/IRD), 18, Avenue Edouard Belin, 31401 Toulouse Cedex 9, France
e Information Technologies Institute (ITI), Centre for Research & Technology Hellas, 6th km Harilaou – Thermi, 57001 Thessaloniki, Greece
f Alterra, Wageningen UR, Droevendaalsesteeg 3, 6708 PB Wageningen, The Netherlands
h IRD-UMR 228 ESPACE-DEV, MTD-Montpellier, 500 rue Jean-François Breton, 34093 Montpellier Cedex. France
i InBIO/CIBIO & Faculdade de Ciências da Universidade do Porto, Edifício FC4 (Biologia), Rua do Campo Alegre, s/n, 4169-007 Porto, Portugal
j Department of Agro-Environmental and Terrestrial Sciences, University of Bari, Aldo Moro, Via Orbona 4, 70126 Bari, Italy

A R T I C L E   I N F O

Article history:
Available online 18 November 2014

Keywords:
Habitat
Land cover
Classification
Monitoring
Remote sensing

A B S T R A C T

To support decisions relating to the use and conservation of protected areas and surrounds, the EU-funded BIODiversity multi-SOunce monitoring System: from Space TO Species (BIO-SOS) project has developed the Earth Observation Data for HABitat Monitoring (EODHaM) system for consistent mapping and monitoring of biodiversity. The EODHaM approach has adopted the Food and Agriculture Organization Land Cover Classification System (LCCS) taxonomy and translates mapped classes to General Habitat Categories (GHCs) from which Annex I habitats (EU Habitats Directive) can be defined. The EODHaM system uses a combination of pixel and object-based procedures. The 1st and 2nd stages use earth observation (EO) data alone with expert knowledge to generate classes according to the LCCS taxonomy (Levels 1 to 3 and beyond). The 3rd stage translates the final LCCS classes into GHCs from which Annex I habitat type maps are derived. An additional module quantifies changes in the LCCS classes and their components, indices derived from earth observation, object sizes and dimensions and the translated habitat maps (i.e., GHCs or Annex I). Examples are provided of the application of EODHaM system elements to protected sites and their surrounds in Italy, Wales (UK), the Netherlands, Greece, Portugal and India.

© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

Introduction

Land use remains a significant driver of habitat degradation and removal and associated losses in biodiversity. In Europe, Natura 2000 sites were established to halt such losses (Mücher et al., 2006; Mücher, 2009; EC, 2011) but increasingly pressures from within but particularly around these sites are leading to deterioration of the habitats they were designed to protect (Lomba et al., 2013; Mairota et al., 2013; Vicente et al., 2013). Quantitative assessments of the changing extent and quality of habitats and the threats and pressures affecting these are therefore urgently needed.

As many protected sites and their surrounds are changing rapidly, the use of Earth Observation (EO) data combined with local knowledge of the sites has been advocated for monitoring. The benefit of EO data is that these are acquired in different modes (e.g., optical, radar and LiDAR) and often routinely at various spatial and temporal scales. However, experts (e.g., ecologists, reserve wardens, vegetation surveyors) with good knowledge of the sites being observed are often needed to ensure that the interpretation of the EO data is accurate and that classification outputs are appropriate for conservation purposes. Common obstacles to achieving this link have often included partial knowledge of habitats and of the needs of users by EO scientists, and skepticism amongst the potential users on the ability of EO data to deliver the information

http://dx.doi.org/10.1016/j.jag.2014.10.011
needed for practical conservation and management. Therefore, the knowledge of both parties needs to be tapped in order to optimize maps and maximize benefit for practical applications (Lucas et al., 2006, 2011; Blonda et al., 2013). A further obstacle is that there has been no systematic approach to the classification of habitats from EO data that is applicable to all sites and available as a standard. Indeed, many areas are mapped and monitored using a range of different data sources, types and classification schemes. Furthermore, the schemes used have often generated classes of limited value for conservation purposes.

To address these issues, the FP7-funded BIO_SOS project sought to develop the Earth Observation Data for Habitat Monitoring (EODHAM) system, with this providing a standardized framework for consistent land cover and habitat mapping and monitoring inside and around protected areas with particular focus on European Natura 2000 sites and their surroundings. A key component of the system is the inclusion of decision rules within a hierarchical classification structure with these generated from expert knowledge from both ecologists and remote sensing scientists. This paper conveys the framework of the EODHAM system and provides examples of elements as applied to selected sites.

The EODHAM system overview

Overall structure

The EODHAM system adopts the Food and Agriculture Organization’s (FAO) Land Cover Classification System (LCCS) (di Gregorio and Jansen, 2005; Fig. 1), which shows the closest correspondence of any common classification scheme (Tomaselli et al., 2013) to the habitat taxonomy of General Habitat Categories (GHCs) (Bunce et al., 2008). This has been tested previously in the context of habitat and biodiversity monitoring (Bunce et al., 2013a) and provides a useful framework for EO and in situ data integration for Annex I mapping. By using GHCs in combination with information on environmental variables (e.g., biogeographical regions, surface moisture) and on dominant or indicator species, Annex I categories (Bunce et al., 2013b) can be delineated although end-user interaction is often a requirement at this stage.

The EO component of the EODHAM system (Fig. 2) is based on geographic object-based image analysis (GEOBIA; Hay and Castilla, 2008; Blaschke and Strobl, 2001) and also incorporates a pixel-based analysis. The system is comprised of (a) data input involving preparation and pre-processing (orthorectification, radiometric, atmospheric and/or topographic correction), (b) spectral feature extraction, segmentation and classification to LCCS Level 2 (1st stage), (c) classification to Level 3 and beyond (2nd stage), with this involving expert knowledge, and (d) translation of these classes to GHCs and Annex I Classes (3rd stage) of conservation importance (EU Habitats Directive). An additional module focuses on change detection and validation of outputs, which include maps of land cover, habitats and changes in these. The output products feed subsequently into modules that perform ecological modeling at the landscape level, biodiversity indicators extraction, and biodiversity indicators change detection.

The processing within the EODHAM is automated, with the exception of threshold value determination, and is undertaken primarily using the Remote Sensing and GIS Library (RSGLib) software (Bunting et al., 2014), the Geospatial Data Abstraction Library (GDAL), and the ORFEO Toolbox (Inglanda and Christophe, 2009), with XML and PYTHON scripting. The classification system also makes use of the KEA image file format (Bunting and Gillingham, 2013), which allows for processing within a raster attribute table (RAT). Within the RAT, which has been developed such that large datasets can be efficiently analyzed (Clewley et al., 2014), all pixels of the same object share the same ID. This table is first populated with image data and derived products (e.g., vegetation indices) and class codes are added progressively as the classification proceeds, with the final attribution being the LCCS, GHCs and Annex I classes for different periods in time. This allows changes to be detected. All attributes can be readily queried to fine tune any step in the classification process.

EO data requirements

The EODHAM system was developed for use with very high resolution (VHR) optical (including hyperspectral) data, but has the benefit of being able to ingest data from any sensor (including radar and LIDAR) and at any spatial scale provided that information extracted (e.g., on plant leaf phenology or type or water inundation extent) is relevant to the classification process and accurate within acceptable limits. At VHR, the Worldview-2 is currently the sensor of choice because the eight bands of spectral data allow calculation of a wider range of spectral indices and images acquisitions can be targeted to periods that are phenologically optimal for the discrimination of land cover classes and the detection of change. When using optical data, the standard processing provides data expressed in units of top of atmosphere (TOA) reflectance but the preference is to remove the effect of the atmosphere so data are processed to surface reflectance (e.g., using the 6S code; Vermote et al., 2003) and corrected for topographic and/or bidirectional effects (e.g., using the algorithm of Shepherd and Dymond (2003).

The system has the advantage of being able to bring in knowledge from the users (e.g., whether water bodies are persistent.
Fig. 2. Overview of the EO components within the EODHaM system. The system is hierarchical and top down and commences with the classification of Levels 1 and 2, with this based on spectral processing alone, and continues to LCCS Level 3 and beyond. The final LCCS categories are then translated to GHCs and subsequently to Annex I. Change analyses are then performed.

or otherwise) and also from other data sources, including digital terrain models (DTMs), canopy height models (CHMs), estimates of the number of vegetation strata, and outputs from hydrological models (e.g., inundation extent). Thematic information relating to the distribution of urban infrastructure, production agriculture and forestry, and water bodies (e.g., reservoirs) can also be included. In the latter case, a high level of geometric correction accuracy is essential.

For detailed classification, information on the seasonal phenological behavior of vegetation (Fisher et al., 2006; Bhandari et al., 2012; Zhu and Woodcock, 2012) is a pre-requisite to distinguish many of the required land cover and habitat types. However, pheno-logy differs depending on the environment being considered. For example, in the humid tropics (e.g., Brazilian Amazonia), only one image is generally needed because of the generally low variation in spectral reflectance over time. However, for other areas, what is termed as a ‘pre’ or a ‘post’ flush and a ‘peak’ flush image are required with these relating to the periods of lowest and highest productivity of vegetation, respectively. In northern Europe, the pre and peak flush images would be acquired in the early spring and mid summer respectively, whereas in the Mediterranean, the periods of acquisition would be associated with the spring (peak flush) and summer (post flush). This would also be the case in season al forests (e.g., in India) where there is a distinct dry (pre or post flush) season and a wet season (peak flush). In some regions (e.g., southern Italy, Wales, India) with seasonal environments, more acquisitions would be needed to define the different land cover categories and associated habitats with these acquired in what is termed here as ‘transition’ periods. These often present the best opportunities for discriminating specific habitats, particularly where these are only distinguishable spectrally for a very short period of time. Multi-date imagery is also required to capture hydrological cycles and particularly the periodicity of inundation, or to deal with the seasonal management of some land use classes (e.g., ploughing and other practices of annual farming cycles). In the LCCS classification, for example, water seasonality and persistence are considered. Water seasonality is defined on the basis of sites being inundated for 2–4 months or more than 4 months and waterlogged areas are also distinguished. For natural and artificial waterbodies, water persistence is defined on the basis of these being non-perennial (1–3 months, 4–6 months and 7–9 months) or perennial (>9 months). Tidal areas (with diurnal variations) and inundation within cultivated areas are further considered. Such information can be obtained from temporal imagery, by referencing hydrological models or from local knowledge.

The EODHaM 1st stage

The EODHaM 1st stage uses only spectral information for extracting objects, segmenting the imagery, and classifying the landscape to LCCS Level 2 (Fig. 2). The three essential steps in the EODHaM 1st stage are (a) initial extraction of distinct and identifiable features of varying size and dimension within the landscape followed by (b) segmentation of the remaining areas to divide the landscape into spectrally homogeneous units and, once completed, (c) a classification of objects within the imaged scene.

Object extraction

Whilst many studies have focused on segmenting an entire image (e.g., Lucas et al., 2006), a limitation is that the algorithms used perform well in delineating some landscape objects but rarely all objects of interest. Instead, many segmentation methods are not adapted to detect the variety of geographical entities comprising a complex scene (Marceau et al., 1994). Blaschke and Strobl (2001) also consider that segments in an image will never represent meaningful objects at all scales and address all applications, and recommend a multi-scale segmentation approach.

For this reason, the EODHaM system first automatically extracts recognizable objects from the image prior to segmentation through an extraction procedure that utilizes specific spectral bands or derived indices (Arias et al., 2013). These objects, observed in VHR data, include individual trees, hedgerows, roads and ponds. Their detection is typically validated through reference to existing thematic layers (e.g., buildings extent or ground observations; e.g., of tree crowns). For example, in Wales, over 76% of mobile and static caravans were detected using this approach. However, the success in extraction depends upon factors such as the ground surface topography and solar illumination as well as the contrast of the features with others in the scene. The reliability of extraction will also decrease for composite objects (e.g., farmyards). As the resolution decreases, most of these objects become less distinct (Marceau et al., 1994). Hence, the number of objects of different type that can be extracted from, for example, 10 m SPOT HRG or 30 m Landsat sensor data decrease with the spatial resolution.

Within the EODHaM system, and prior to segmentation of VHR imagery, algorithms for automatically extracting objects corresponding to individual tree crowns and clusters of crowns, buildings (including caravans) and hedgerows have been developed within the ORFEO Toolbox. These utilize individual or combinations of bands, derived indices (e.g., entropy) or context-sensitive features such as geometric (e.g., area, compactness, elongatedness), morphological, topological (e.g., adjacency) and non-topological (e.g., distance between objects) attributes (Arias et al., 2013). For extracting hedgerows, for example, thresholds of Haralick texture measures and binary morphological operators (e.g., dilation, erosion and closure) are used; connections are then made between segments such that hedgerows are formed. These thresholds have been defined with reference to VHR imagery from a range of European sites but can be varied by the user. An example of delineated tree crowns and clusters of crowns within both a natural and managed setting is provided in Fig. 3. Depending upon
their spectral uniformity, larger field units can also be delineated. However, this is compromised where boundaries are indistinct or a large number of objects with a high spectral diversity occur within each unit.

Segmentation

Following automated extraction of landscape objects, two algorithms are available within the EODHaM 1st stage for segmenting the remaining areas, with these based primarily on spectral information. However, LIDAR data can be integrated as this often leads to improvements in the segmentation of forest areas. Within RSGISLib, and for image segmentation, a clustering and small object elimination algorithm is applied whilst the ORFEO Toolbox uses the algorithm of Comaniciu and Meer (2002).

At this stage, the concept of ‘small’ and ‘large’ objects is introduced. ‘Small’ objects are typically those that can be discerned within the imagery and include (in the case of the VHR data) previously extracted objects (e.g., buildings, trees, hedgerows). These ‘small objects’ are extracted using the procedures outlined in the object extraction section. However, small objects can also be obtained from within the remainder of the image by parameterizing the segmentation algorithms such that the object size is commensurate with many of the finer but often less distinct or recognizable elements of the landscape (e.g., patches of shrub or marshy grasslands in otherwise dry fields). Existing thematic layers (e.g., representing buildings or field boundaries) can also be integrated within the segmentation process, which often results in splitting of objects. However, the benefit is that the resulting objects align with units (e.g., roads) that have already been mapped. A separate segmentation is then performed for the entire image to generate ‘large’ objects, which can be associated with well-defined landscape and management units, including fields, forestry plantations, urban infrastructure (e.g., airport runways, large industrial buildings) and reservoirs. These larger objects are generated through parameterization of the segmentation algorithm and delineation of specific features (e.g., field or forest boundaries) can often be validated through reference to existing thematic layers. The landscape objects already extracted (e.g., buildings, individual trees) are ignored in this segmentation process such that these larger units are captured in their entirety. However, reference can be made interchangeably between the large and small objects through the RAT of both the large and small object layers. The small and large segmentation generated using the RSGISLib code is illustrated in Fig. 4. In this segmentation, which is based on an unsupervised K-means clustering of the imagery, objects below a certain size are eliminated following each iteration and associated pixels are reassigned to new clusters in subsequent iterations. The number of clusters, maximum number of iterations and degrees of spectral change required by the algorithm can be altered from default values as can the minimum size of objects to be eliminated. The algorithm gives the same result with the same input parameters (Clewley et al., 2014).

Linking small and large objects

Large objects contain smaller objects, which are often relatively homogeneous (spectrally) and spatially related (Coulcelis, 2010). For example, olive groves (the large objects) contain olive trees, with each individual showing different characteristics (e.g., crown dimensions, height, species type) or sharing attributes with other similar individuals (e.g., orientation, distance to). In some cases, however, the larger object may not capture all of the components of a composite feature (e.g., an airport or caravan park) in a landscape as it is not sufficiently well delineated. This particularly the case where the composite feature is comprised of a large number of smaller objects with differing spectral characteristics. In this case, a large object can be constructed from the smaller simpler objects (e.g., trees of different species type or buildings) (Coulcelis, 2010). As an example, a caravan park might consist of individual caravans but also roads and grass verges and lawns. Hence, by first classifying and then combining these objects (if they can be extracted) based on, for example, proximity by object type and connectivity, a larger object can be formed and described. Another example would be an airport with runways, grass areas, terminal buildings and road infrastructure. Through these two approaches, large objects are defined and described.

Classification in the EODHaM 1st stage

The EODHaM 1st stage involves classification at the pixel and small object level and, in its current implementation, utilizes a sequence of decision rules, with these minimized to include only a narrow range of spectral indices that allow discrimination of LCCS Level 1 and Level 2 categories (i.e. vegetated and not vegetated terrestrial and aquatic). Whilst the sequence of rules and data layers used is set, the thresholds applied are subject to change depending upon the users’ a priori knowledge and interpretation of the scene although they are generally similar within and often between environments (e.g., Temperate or Mediterranean). At the pixel level,
simple binary masks representing, for example, the extent of vegetation cover are generated. However, at the small object level, the raster attribute table (RAT) associated with the KEA format is accessed for classification (Bunting and Gillingham, 2013). Within the RAT, each small object has a unique ID with all pixels associated with each object having the same characteristics. Each object within the RAT is then progressively attributed with the means of reflectance data and derived indices calculated from the pixels contained, binary mask information (e.g., water, not water) and ultimately the class assigned (e.g., vegetated, non-vegetated). The large object layer used later also uses the RAT.

Pixel level

Classification at the pixel level is undertaken to generate binary masks of the extent of different vegetative states (Table 1) and open water, based on indices such as the Normalized Difference Vegetation Index (NDVI), Plant Senescence Reflectance Index (PSRI) and the Water Band Index (WBI) (Sims and Gamon, 2002; Fig. 5). These are used to discriminate green (photosynthetic) and brown (non-photosynthetic; dead or senescent) vegetation but options are also available for mapping the extent of other states, including non-submerged, submerged (e.g., algae) or burnt. Once mapped, all vegetated states are merged into a vegetated category and remaining pixels associated with a non-vegetated category. Within the vegetation category, woody vegetation is differentiated using the ratio of the blue and green reflectance or the LIDAR canopy height model (CHM), in preparation for subsequent classification at Level 4. Where dual or multi-season images are available, the pixel-level classification of the different vegetative states at each time step is used to determine the extent of evergreen, deciduous and also low productivity vegetation (e.g., in aquatic environments) and hence indicate phenology, which is a component of the LCCS classification.

Small object level

At the small object level, non-vegetated areas are associated with (a) open water and urban infrastructure, (extracted from the image either a priori (e.g., buildings; see object extraction section) or classified using the indices given in Table 2), and (b) remaining objects not classified as vegetation using the indices given in Table 1. This latter case avoids separate classification of the wide range of non-vegetated surfaces that are common to many scenes. A second component then identifies aquatic surfaces, with these including the open water areas used for the classification of non-vegetation but also submerged and non-submerged aquatic vegetation. These latter categories are often difficult to differentiate because of environmental variability (e.g., extent of inundation, type (water, ice or snow) and turbidity) and may warrant the use of more specific spectral indices, digital elevation models or ancillary information (e.g., the output from hydrological models). Once defined as aquatic, all remaining objects are assigned as terrestrial using an inverse rule. The classification at LCCS Level 2 (i.e., vegetated or non-vegetated terrestrial or aquatic) is then achieved through cross tabulation of the areas of vegetation and non-vegetation with those that are aquatic and terrestrial.

Table 1

Indices used in the rule-based implementation of the EODHaM System for identifying vegetated states.

<table>
<thead>
<tr>
<th>Green vegetation</th>
<th>Formula</th>
<th>Other vegetative states</th>
<th>Formula</th>
<th>Woody vegetation</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDVI</td>
<td>$\frac{\rho_{\text{red}} - \rho_{\text{NIR1}}}{\rho_{\text{NIR1}} + \rho_{\text{NIR2}}}$</td>
<td>PSRI&lt;sup&gt;a&lt;/sup&gt;</td>
<td>$\frac{\rho_{\text{red}} - \rho_{\text{NIR1}}}{\rho_{\text{NIR2}}}$</td>
<td>BG</td>
<td>$\frac{\rho_{\text{NIR2}}}{\rho_{\text{NIR1}}}$</td>
</tr>
<tr>
<td>Greenness&lt;sup&gt;b&lt;/sup&gt;</td>
<td>$\frac{\rho_{\text{NIR2}} - \rho_{\text{red}}}{\rho_{\text{NIR2}}}$</td>
<td>REP&lt;sup&gt;c&lt;/sup&gt;</td>
<td>$\rho_{\text{NIR2}} - (\rho_{\text{NIR2}} - \rho_{\text{NIR1}})$</td>
<td>CHM&lt;sup&gt;d&lt;/sup&gt;</td>
<td>HEIGHT</td>
</tr>
<tr>
<td>FDI</td>
<td>$\rho_{\text{red}} - (\rho_{\text{NIR1}} + \rho_{\text{NIR2}})$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Non-photosynthetic/senescent (brown) vegetation.

<sup>b</sup> Non-submerged aquatic vegetation.

<sup>c</sup> Submerged aquatic vegetation (used in combination with the WBI; see Table 2).

<sup>d</sup> Canopy Height Model (e.g., derived from LIDAR data).

<sup>e</sup> Forest Discrimination Index.
Fig. 5. (a) the NDVI and (b) the PSRI representing areas of green and brown vegetation respectively and (c) the WBI, Le Cesine, Italy. The scale bar indicates values ranging from low (black) to high (white).

The EODHaM 2nd stage

The EODHaM 2nd stage has two main components focusing on classification to LCCS Level 3 and, subsequently, Level 4. In its current implementation, the classification utilizes user-defined thresholds of reflectance data or derived indices but other data can be incorporated including Digital Elevation Models (DEMs) and derived measures (such as DTMs and CHMs obtained from.

Table 2
Indices used in the rule-based implementation of the EODHaM system for identifying non-vegetated areas.

<table>
<thead>
<tr>
<th>Water</th>
<th>Formula</th>
<th>Soil</th>
<th>Formula</th>
<th>Urban</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDWI</td>
<td>( \frac{R_{\text{NIR}1} - R_{\text{R}}}{R_{\text{NIR}1} + R_{\text{R}}} )</td>
<td>NDSI</td>
<td>( \frac{R_{\text{G}} - R_{\text{R}}}{R_{\text{G}} + R_{\text{R}}} )</td>
<td>NHFD</td>
<td>( \frac{R_{\text{C}} - R_{\text{R}}}{R_{\text{C}} + R_{\text{R}}} )</td>
</tr>
<tr>
<td>WBI</td>
<td>( \frac{R_{\text{B}}}{R_{\text{NIR}1}} )</td>
<td>Brightness</td>
<td>DSM(^a)</td>
<td>Height</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) Digital Surface Model (for buildings).
Table 3
Primary and secondary descriptors of the land covers associated with each LCCS class at Level 3.

<table>
<thead>
<tr>
<th>No</th>
<th>LCCS Component</th>
<th>Level 3 category</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>A11</td>
</tr>
<tr>
<td>1</td>
<td>Lifeform$^{1,2,3}$</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Cover$^{1,2}$</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Height$^{1,2}$</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Spatial distribution$^{1,2,3}$</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Leaf type$^{1}$</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Phenology$^{1}$</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Life cycle$^{1}$</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Stratification$^{2}$</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Spatial aspects$^{1,2,3}$</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Crop combination$^{1,3}$</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Cover-related cultural practices</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Water seasonality$^{3}$</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Surface aspect$^{1,2}$</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Macropattern$^{1,2}$</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Physical status$^{3}$</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Persistence$^{1,2}$</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Depth$^{1,2}$</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Sediment Load$^{1}$</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Landform$^{1,2}$</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Climate</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Altitude$^{1}$</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Lithology/Soils</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Erosion$^{1,2}$</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Crop cover/density$^{1}$</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Water quality$^{1}$</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Surface aspect$^{1,2}$</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Vegetation$^{1,2}$</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Crop type$^{1,3}$</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>Floristic aspect$^{1}$</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Built up objects$^{1,2,3}$</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>Salinity$^{3}$</td>
<td></td>
</tr>
</tbody>
</table>

LCCS components are derived from $^1$optical and/or $^2$LIDAR data. Some codes can be allocated using $^3$polarimetric or interferometric radar data, either singularly or in combination with optical or LIDAR data. All information can be obtained or augmented by ground surveys or in situ data/knowledge. Not all layers require population for a classification to be generated.

for example, LIDAR), and thematic information (e.g., presence of buildings). Such information is integrated into the RAT. Thematic information may be included as a numeric value (e.g., indicating overlap or otherwise) or a string. This provides a wide range of information that can be exploited by the users for land cover and subsequently habitat description through the EODHaM process. For the classification at Level 3 and particularly Level 4, expert knowledge of the information content of remote sensing data in relation to the land covers being considered can be incorporated into the rules (e.g., in defining threshold bands and values; Adamo et al., 2014).

Classification to Level 3

The LCCS Level 3 classification requires that the landscape be differentiated according to elements that are cultivated, managed or artificial or natural or semi-natural. The system focuses on classifying the landscape through reference to the extracted context-sensitive objects (e.g., large objects containing individual trees in rows) or existing thematic (e.g., cadastral, infrastructure) layers.

Initially, objects in the small object layer are associated with the class they represent using the thematic label applied when these were extracted (e.g., trees, buildings). In the RAT for the large object layer, the proportion of the area represented by these small objects is calculated, with this potentially expandable to include the number of objects, their orientation and so on. Some measures are calculated to allow for description of objects or disambiguation. For example, roundness is a measure of how different a shape is from a circle and this measure can be used to identify and describe tree crowns (e.g., olive trees). Such information can then be used to classify forest plantations (based on the proportion of pixels representing woody vegetation), orchards (based on the size and density of trees) and urban areas (based on the number, size and density of buildings) and hence managed and cultivated areas. Each large object is also associated with a field size class as the LCCS differentiates between small ($<8$ ha), medium ($\geq 8$ and $<20$ ha) and large ($\geq 20$ ha) areas. A key criterion here is to ensure appropriate delineation of the larger enclosing objects, which can be problematic (e.g., when dealing with olive groves) and may necessitate the use of existing ancillary information (e.g., relating to field boundaries).

Once areas are associated with these labels (i.e., ‘terrestrial’ or ‘aquatic’, ‘vegetated’ or ‘not vegetated’, ‘cultivated, managed or artificial’, or ‘natural or semi-natural’), these are cross-tabulated to generate a classification of Level 3 categories (e.g., terrestrial vegetated cultivated/managed; see Fig. 1). The accuracy in the classification of these categories needs to be high as separate classifications of each are performed subsequently when classifying beyond Level 3.

Classification beyond Level 3

When classifying beyond Level 3, thirty-two separate layers (columns in the RAT) are generated, which are populated subsequently with class codes defined within the LCCS (Table 3). These codes are then combined within the RAT to generate the final classification label. As an example, eleven codes are used for vegetation. In the first instance, woody (A1) and herbaceous (A2) lifeforms
and cryptograms (A7) are differentiated. In the woody category, trees (A3) and shrubs (A4) are separated whilst in the herbaceous category, forbs (A5) and graminoids (A6) are distinguished. Cryptograms are divided into lichens (A8 code if terrestrial vegetated (A2) and A10 if aquatic vegetated (A24)) and mosses (A9 if A12; A11 if A24). Lifeforms are then categorized according to their cover (A12–A20 codes, for different percentage covers), height (B1–B10, for different vegetation heights), spatial distribution (continuous (C1) or fragmented (C2)), leaf type (broad-leaved (D1), needle-leaved (D2) or aphyllous (D3)), seasonality/leaf strategy (evergreen (E1), deciduous (E2), semi-evergreen, E3) and/or lifecycle (annual (E5) or perennial (E6)) and stratification (one, two or more layers; F1, F2 etc.). The assigned codes are then combined and translated to a meaningful description; for example, A12.A3.A11.B5.C1.D1.E1 translates to semi-natural open continuous high broad-leaved evergreen forest. An example of the classification for the Veluwe in

Fig. 6. Classifications of vegetation as a function of (a) height and (b) cover based on LIDAR and VHR optical data respectively and (c) the resulting LCCS classification for Veluwe, the Netherlands.
employing Dempster–Shafer theory principles and fuzzy logic has been implemented, with this complementing the basic LCCS to GHCs translation framework, as proposed by Petrou et al. (2014b). The rationale is to handle uncertainty in the outcome of expert rules and missing information and counteract both the potential noise afflication of the data and inaccurate rule thresholds provided by the experts. Using a linear membership function, each potential GHC class event of a certain object is given a basic probability assignment value. In case inadequate information is available to discriminate among certain potential GHC classes, events consist of multiple GHC classes. Belief and plausibility values are then assigned to each event. The final event is selected as the one with the smallest number of classes under the requirement of minimum belief value of 0.75 or 0.94, for events with single or multiple classes respectively (Petrou et al., 2014b).

The EODHaM approach has been extended to include the extraction of Annex I maps, with rulesets defined for the translation of LCCS to GHC and to Annex I classes (using GHC qualifiers) (Tomasselli et al., 2013). The LCCS attributes and GHC qualifiers refer to additional layers of information (e.g., lithology, moisture, soil aspect, acidity, elevation, climate), which help to resolve ambiguities in the classification to Annex I. Whenever such data layers are unavailable, multiple Annex I classes may result when translated from some GHCs classes.

Several approaches to assess the accuracy of the land cover and habitat classifications are available, with these based upon the LCCC class components, the combined LCCS class codes, the GHCs translated from these and/or the Annex I categories. Accuracy can be assessed against the data or other remote sensing data (e.g., aerial photography). As an example, for the sites considered, users and producers accuracies (Congalton, 1991) for the LCCS classes were generally lowest (typically over 97%) for homogeneous classes (e.g., coniferous plantations, open water, active raised bog in temperate environments) but least (sometimes as low as ~30%; close to chance) for more complex and heterogeneous land covers (e.g., natural aquatic perennial graminoids in Italy and mosses on flooded land in Wales).

### Change detection modules

A wide range of techniques have been developed for detecting change using EO data (Singh, 1989; Bovolo et al., 2012). Common amongst these are image and class differencing, change vector analysis, and cross-correlation analysis (Koeln and Bissonette, 2000; Blaschke, 2010; Chen et al., 2012). Within the EODHaM system, six types of change assessment are considered, with these based on the use of the RAT matrix of the KEA file format: changes in (i) LCCS classes (or GHCs), (ii) LCCS component codes, (iii) the number of extracted objects belonging to the same category (e.g., buildings), (iv) object size and geometry (splitting or merging), (v) EO data and derived metrics (e.g., LIDAR-derived height or the NDVI), and (vi) calculated landscape indicators. Change detection is possible, even when some information on habitats or indicators is missing. Examples of changes that are detected are provided in Fig. 8, whereby individual caravans are tracked over time (delineated tree crowns could similarly be mapped over time to quantify losses) and changes in class assignment (vegetation to water) and the PSRI are indicated.

### Discussion

**Use of the LCCS and GHCs taxonomies**

When classifying land covers from EO data, a common approach has been to take training areas for pre-defined classes (e.g.,

---

**Fig. 7.** Areas of the grassland (GRA), which can be differentiated where these occur adjacent (with a border) to urban buildings (URB).

The Netherlands is provided in Fig. 6. In this case, the classification benefits from the inclusion of LIDAR data as the LCCS describes vegetated areas on the basis of their height (with divisions at 0.03, 0.3, 0.5, 0.8, 3, 5, 7, 14, and 30 m) and cover (with divisions at 1, 4, 15, 40 and 65%). The LIDAR point cloud data can also be binned on a per unit area basis to differentiate vegetation with one or two or more layers, with this describing their stratification (Miura and Jones, 2010; Bunting et al., 2013). Water can similarly be described in relation to its state (water, ice or snow), persistence, depth, sediment load and salinity. In the EODHaM System, not all layers need to be populated to generate the final classification. For example, information on leaf type (D1, D2 or D3) may be missing or woody shrubs are unable to be distinguished from trees and hence remain as woody (A1). Hence, land cover maps can still be generated even though some LCCS codes are not available.

### EODHaM 3rd stage: translation to GHCs

In the 3rd stage of EODHaM, GHCs and Annex I habitat maps are automatically produced based on a translation from the LCCS classes extracted in the previous stages. Mapping relationships and discrepancies in definitions between the LCCS and GHCs taxonomies are outlined in Kosmidou et al. (2014). Several one-to-

many relationships from LCCS to GHCs classes have been observed. To resolve such ambiguities, additional site-specific expert rules, using ancillary data, morphological and topological features and contextual information (e.g., proximity, adjacency), have been identified and reported by Adamo et al. (2014), with these integrated within the EODHaM system. For example, herbaceous graminoids (CHE) adjacent to buildings (URB), regardless of density, are described as Urban/Herbaceous (GRA; Fig. 7), as discussed by Bunce et al. (2011). LIDAR data can also be used to determine the GHCs relating to low, medium, tall or forest phanaeroophytes (i.e., LPH, MPH, TPH and FPH respectively; Bunce et al., 2008, 2011). In case LIDAR data are not available, texture analysis measures (i.e., local entropy) are used as surrogates to discriminate vegetation height categories (Petrou et al., 2012, 2014a; Adamo et al., 2014).

To increase the robustness and transferability of the framework to landscapes in different geographical locations, a scheme
broadleaved deciduous forest) and use the statistics from these to provide a classification result (e.g., Muchoney et al., 2000; Xie et al., 2008). The EODHaM system takes a different approach by classifying the components of each LCCS class (e.g., broadleaved, deciduous and forest) based on a range of input data (airborne, spaceborne, thematic) and then combining these to provide a meaningful description (i.e., broadleaved deciduous forest). The main benefit is that the classes, which can be numerous, are relevant to any location worldwide (by virtue of using the LCCS and GHC taxonomies) and the classification can be applied at any spatial scale provided that features within the landscape can be adequately resolved and appropriate spectral data (in terms of wavelength regions) are available. Hence, sensors such as the Worldview-2, Quickbird and/or Landsat data can be used to classify sites ranging in size from several to hundreds of km². Where available, LiDAR data have also been integrated primarily to classify vegetation height and cover. Where only coarser spatial resolution data are available, such as provided by the Landsat sensors, the classification of some elements of the landscape (e.g., buildings, hedgerows) may not be achieved. In general, the sequence of rules used for classification remains the same but the threshold values may vary depending upon the image data used, prevailing environmental conditions and local knowledge. Thresholds are generally transferable to other regions when using data from the same sensors, with preference given to imagery acquired prior to and during the peak vegetation flush period.

Detection of change

Most change detection techniques typically focus on just one element of the change process, whether it be a change in class (e.g., to indicate processes of deforestation) or reflectance values

Fig. 8. Classifications of change (circled) in Cors Fochno, Wales, showing (a) an increase in the number of extracted objects (in this case, mobile caravans), (b) water inundation (blue) from flooding and tides, with this determined as a change in the LCCS class label and sub-code, and (c) dieback of stands of P. australis (whiter shades) between Time 1 and Time 2, as indicated by an increase in the PSRI. These change observations give land managers evidence of the changes occurring within and surrounding the Natura 2000 site.
and derived measures. The change detection method is also often applied to the entire image, and the change process is rarely placed in context. Within the EODHaM system, the change process considers differences in the LCCS class codes (including their components), the numbers, sizes and dimensions of objects in the landscape, and differences in image data values and derived measures. As such, the EODHaM system provides users with a more complete insight into the changes that are occurring within protected areas and their surrounds and also results that are easier to interpret. Rather than stating only that a change has taken place, the change itself is described in detail.

Limitations of approach

The greatest difficulty in the mapping of the LCCS classes is the differentiation of aquatic vegetation as well as cultivated and managed areas. This can be partly overcome by developing context-sensitive rules or referring to existing ancillary data layers or hydrological models. In the case of ancillary layers, cadastral information can be used to identify vegetation occurring, for example, in rice or cotton fields. Whilst the classification is currently rule-based, outputs from other classification procedures (e.g., supervised, support vector machines) can be used. For example, components of these classifications (e.g., maps of forest extent, water or urban areas) can be used as direct input to the EODHaM system. A further limitation is that subpixel proportions (e.g., of plant functional types) are not yet integrated within the classification. These can, however, be determined a priori based on, for example, spectral end-member unmixing or fuzzy classification and added to the RAT. Each object would then be classified according to the LCCS taxonomy based on the dominant class (e.g., woody shrubs, graminoids, forbs) or by genus or species, with information on the relative proportions of the remaining life forms, genera or species occurring within the object also retained for further analysis.

Summary and conclusions

The EODHaM system (which is outlined in Fig. 2) is comprised of software that allows for classification within the framework of the FAO LCCS and subsequent translation of the LCCS classes (Level 3 and beyond) to GHGs and Annex I habitats (taking into account ambiguities). These taxonomies are recognized for their global application. The EODHaM system was developed using VHR data but data from moderate resolution optical sensors can also be used as the same LCCS taxonomy is applied. However, there is a scale dependency on the number and types of classes that can be discriminated. Classifications can also be undertaken based on knowledge of the user without the need for extensive ground truth datasets. Change detection modules are available with these based on changes in LCCS codes and classes, spectral data and indices and automated or semi-automated extraction (e.g., trees, buildings). The system is largely automated, with user interaction required primarily to fine-tune the thresholds used in the rule-based system. The system benefits from interactions by the end users in refining the rules used for classification and selecting ancillary data (e.g., tidal levels) relevant to the discrimination of land cover types. Overall, the EODHaM system provides a framework and potentially operational approach to the monitoring of protected areas and their surrounds from EO data.

Based on EO data availability, the methods have been successfully applied to Natura 2000 sites and their surrounds in Wales, the Netherlands and Italy and have been evaluated at other locations in Greece, Portugal and India, thereby covering a range of environmental and biogeographic contexts. In each case, a consistent classification of the landscape has been achieved confirming the transferability of the system. The EODHaM system is currently implemented within open source software and uses a diversity of EO and also ancillary datasets. The system can be readily adopted by managers of protected sites and their surrounds for consistent land cover, habitat and biodiversity monitoring within and between sites and improving the effectiveness of policy and management whilst coping with national and international reporting obligations.

References


