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: D 6.10

Software for habitat maps production from LC

Due date of deliverable:	31/05/11
Actual submission date:	31/05/11
Version:	4th version of D 6.10
Main Authors:	Vasiliki Kosmidou (CERTH), Zisis Petrou (CERTH), and contribution from: Richard Lucas (ABERY), Valeria Tomaselli (CNR-IGV), Maria Petrou (CERTH), Robert Bunce (ALTERRA), Marion Bogers (ALTERRA), Alexander Mucher (ALTERRA), Cristina Tarantino (CNR-ISSIA), Palma Blonda (CNR-ISSIA), Andrea Baraldi (BACRES)

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

Deliverable title	Software for habitat maps production from LC
Deliverable number	D6.10
Deliverable version	v4
Previous version(s)	
Contractual date of delivery	31 May 2011
Actual date of delivery	31 May 2011
Deliverable filename	BIO_SOS_D6.10_software_for_habitat_maps_from_LC_maps_v4.odt
Nature of deliverable	R
Dissemination level	PU = Public
Number of pages	
Workpackage	WP 6 Task 6.6
Partner responsible	CERTH
Author(s)	Vasiliki Kosmidou (CERTH), Zisis Petrou (CERTH), and contribution from: Richard Lucas (ABERY), Valeria Tomaselli (CNR-IGV), Maria Petrou (CERTH), Robert Bunce (ALTERRA), Marion Bogers (ALTERRA), Alexander Mucher (ALTERRA), Cristina Tarantino (CNR-ISSIA), Palma Blonda (CNR-ISSIA), Andrea Baraldi (BACRES)
Editor	Maria Petrou
EC Project Officer	Florence Beroud

Abstract	This report describes the design of the algorithm for the production of habitat maps from land cover maps based on a dictionary of expert rules. It also discusses the future refinement of the preliminary algorithm with the introduction of ancillary data and contextual information to resolve the ambiguities arising from the one-to-many mapping of land cover classes to habitat classes, and, in a subsequent stage, the introduction of fusion of information methods that allow the uncertainty in the data to be taken into consideration.

Keywords	Habitat mapping, LCCS, GHC

Signatures

Written by	**Responsibility- Company**	**Date**	**Signature**
Vasiliki Kosmidou	CERTH	30/05/11	
Verified by			
Maria Petrou	Responsible for D6.10, CERTH	30/05/11	
Harini Nagendra	WP6 Leader, ATREE		
Approved by			
Maria Petrou	Quality Manager, CERTH	30/05/11	
Palma Blonda	Project Coordinator, CNR		

Table of Contents

1. Executive summary......6

2. Introduction......8

3. Land Cover Classification Systems and General Habitat Categories......10

  3.1 Land Cover Classification Systems......10

  3.2 General Habitat Categories 12

  3.3 Relation between the two taxonomies......14

    3.3.1 Height......14

    3.3.2 One-to-many relationships......15

    3.3.3 Scale......16

    3.3.4 Linking with landscape features......16

4. An algorithm for habitat production maps......17

  4.1 Expert rules for mapping the LCCS to the GHC classes......17

    4.1.1 Top level categorical mapping......17

    4.1.2 Mapping the Life-Form classifiers......20

  4.2 Algorithm design for habitat map production......24

  4.3 Algorithm refinement 26

5. Implementation of the algorithm 29

  5.1 Description of the implementation procedure......29

  5.2 Application of the algorithm and results for Le Cesine site......30

  5.3 Application of the algorithm and results for the Wales site......34

  5.4 Discussion on the results - Future considerations......39

6. Conclusion......40

7. Appendices......41

8. References......49

1. Executive summary

In the Framework of the WP6, with the title *"EO Data for Habitat Monitoring (EODHaM) modelling module development"*, Task 6.6 refers to the *Development of an algorithm for conversion of land cover maps into habitat maps*. Deliverable D6.10 is the output of WP6, Task 6.6, and aims at the design and implementation of the algorithm for producing habitat maps given land cover maps.

Common approaches for habitat monitoring of changes require definitions and rules that are harmonised continentally and globally. Habitat is a widely used term, but, the content of the concept "habitat" remains diverse, ambiguous, and difficult to be used consistently in monitoring. The term "habitat" as used in the EBONE Manual [1] comes as an ecological refinement of land cover categorisation as developed by the FAO-Land Cover Classification Systems (LCCS) [2]. To this end, D6.10 deliverable focuses on the production of General Habitat Categories (GHC) maps [3] from LCCS maps as a base for Annex 1 .

The produced maps will be used further towards the production of Annex I habitat maps, using the rule based hierarchical Key developed within the EBONE [4] project. The Key is available as EBONE Deliverable 4.2 through www.ebone.wur.nl, while the Annex I of the EU Habitats and Species Directive is available at http://ec.europa.eu/environment/nature/legislation/habitats directive/index_en.htm.

In the Mediterranean Natura 2000 sites selected for the BIO_SOS project several habitat types which are of great ecological importance are not considered by Annex I of the EU Directive [see D6.1], even if functionally linked to habitats included in Annex I. For IT4 and IT3 Italian sites such habitats are listed in Table IT4_1 and IT3_1 of D6.1 and evidenced by an X code in the Annex I column. As an example, the communities belonging to "eutraphent reed and sedge beds", are not included in Annex I, even though these types are important habitats in coastal wetlands for bird reproduction and the presence of rare species [see D6.1]. These habitats are strongly threatened by combined anthropic pressures and are exposed to reduction and high fragmentation processes. For such reason the Users who signed the SLA are particularly interested in their conservation.

In EBONE Deliverable 4.2, the Key for the mapping of these habitat types is obviously not available. Within BIO_SOS, focusing mainly on Mediterranean Natura 2000 sites, a specific set of expert rules will be defined for the direct mapping of LCCS classes to both GHC and habitat types as expressed by other habitats classification taxonomy, such as CORINE Biotopes or EUNIS.

In Task 6.6, the dictionary provided in Task 6.1 was taken into consideration along with additional expert rules towards the conversion of LCCS maps into GHCs. Both expert mapping rules, that harmonize the LCCS and the GHC taxonomies, and elaborated classification techniques are employed in the design of the algorithm. Due to the discrepancies between the two classifications, high level of uncertainty is imposed in the classification scheme; thus, the use of a module that takes into consideration the reliability of the process is vital. Such a module can be realised based on Bayesian techniques, which facilitate the use of prior information into the algorithm and account for the uncertainty in the input.

The algorithm in its present form serves as the basic scheme that will evolve into the final project deliverable. It has been implemented and applied in two test sites, i.e. the Le Cesine in Italyand Cors Fochno in Wales. The results presented in D6.10 are indicative of the feasibility of producing GHC maps from land cover maps. Identification of mapping relationships between LCCS classes into GHCs is accomplished here for the first time at the 3rd-stage of the EODHaM system, giving added value to the BIO_SOS project towards the development of a cost effective biodiversity monitoring system.

In general, the identified LCCS-to-GHC mapping relationships are many-to-many, which include relationships one-to-many, many-to-one and one-to-one as special cases. Eligible for further improvements, these initial relationships are presented in D6.10

The refinement of the proposed algorithm within the EODHaM 3rd stage will be a constant procedure to concur with parallel activities in BIO_SOS throughout the duration of the project. The design of the algorithm is such, allowing the future insertion of additional modules and data inputs for the elaboration of the produced habitat maps and the enhancement of the classification scheme.

In particular, to increase the classification accuracy of GHCs generated from LCCS classes, the initial many-to-many relationships of LCCS classes onto GHCs must be better constrained (modeled) to obtain ideal one-to-one relationships or at least one-to-few relationships. To better condition the EODHaM 3rd-stage mapping of LCCS classes into GHCs, two future strategies can be pursued in parallel.

- Employ as input additional sources of ancillary information (where ancillary information is defined as non-pictorial information which cannot be detected in RS imagery), if any.
- Adopt semantic nets, consisting of nodes, equivalent to classes of objects, and arcs (relationships) between nodes. It is noteworthy that semantic nets are also adopted by the EODHaM 2nd stage to generate as output LCCS classes from input spectral categories automatically detected by the 1st-stage SIAM™ preliminary classifier. This means that a synergistic use of semantic nets is expected by both the EODHaM 2nd and 3rd stages.

2. Introduction

A key challenge that has to be addressed in the BIO_SOS project is the cost effective and timely monitoring of changes in the land cover within and along the borders of protected areas, to judge the effectiveness of protecting and conserving the regions from human impacts. Habitat maps, which are at the base of indicator extraction, can be obtained by interpreting land cover maps of sufficient detail with ancillary data, other EO derived products and by re-labelling and, where appropriate, by merging similar land cover classes, according to the 92/43 EEC Directive and to General Habitat Categories (GHC) based on life forms as defined in the previous BioHab project [3].

Many studies have been carried out with the aim to find effective procedures for monitoring EU habitats. The development of consistent recording procedures is essential, especially for long-term monitoring [3], [5]-[7]. Based on the recommendations and conclusions extracted from deliverable D6.1, the adoption of the Land Cover Categorisation System [2],[10] seems to be the most reasonable choice for the transition from LC maps to habitats, due to the highest correlation between this and the targeted GHC in comparison with other LC categorizations. In addition, LCCS provides greater detail mainly for natural and semi-natural class types discrimination if compared to other commonly used LC taxonomies, such as CORINE and IGBP. FAO-LCCS uses life forms and has been recently adopted by the Global Land Cover 2000 (GLC2000) project. As recognized in the EBONE Handbook [1], habitats can be considered as an ecological refinement of the LCCS categorisation as developed by FAO [3] and are the basis for surveillance and monitoring from remotely sensed data. The output habitat maps will be in either the Annex I of the European Directive or the GHC framework, which are also related to each other.

The FAO-LCCS is based on the use of a set of independent diagnostic criteria rather than on establishing a pre-defined land cover class set. It is both hierarchical and well posed. It has proven to be a valid tool in the detection of changes and serves the goals of the EODHaM system. In fact, a land cover change may occur in two ways, as a conversion from one land cover class to another or as a modification within the same class. Conversion implies an evident change and it can be easily represented on a map (changing the LC class), whereas modifications are less apparent and their representation in a map is not always possible, depending on the detail and flexibility of the LC classification used. With the LCCS approach, land cover change detection becomes possible both at the level of conversion of a class, and modification within a certain class type. In this last case, the change becomes immediately identifiable by a difference in the output of a classifier, or through the use of additional qualifiers, although maintaining the same major class type [8],[9].

In parallel, a GHC is specifically designed to be recorded consistently especially for detection and mapping of changes. Furthermore, this system applies stringent criteria to ensure that real change is recorded and not results that are distorted by differences in definitions, between observers or in recording techniques. One of the key elements of this approach is its potential for the detection and evaluation of flows between habitats [3]. The GHC classification system is a promising tool for the detection of changes, not only those involving a change from a habitat type to another, but also those involving a modification within the same habitat type. Such a change can be represented by adding or modifying environmental qualifiers.

Moreover, the framework for the transition of GHC codes to Annex I codes has been fully described in the EBONE Handbook [4]. A rule based key to Annex I habitats has been produced using GHCs to provide a hierarchical key. The Key is available as EBONE Deliverable 4.2 through www.ebone.wur.nl. Annex I of the EU Habitats and Species Directive is available at http://ec.europa.eu/environment/nature/legislation/habitatsdirective/index_en.htm. Therefore, the focus of WP6.6 is the design and implementation of an algorithm to translate LCCS land cover maps into GHC maps which are at the base of the mapping both of the habitat types included in Annex I of the EU Directive and other habitat types of ecological interest for Natura 2000 sites in the Mediterranean areas.

To create the framework for the transition from LC maps to habitat maps, the definitions of two classification schemes, along with the reliability of the procedure needs to be taken into consideration. Therefore, the relation between the two taxonomies, which was initially indicated in D6.1, is investigated here thoroughly to produce the mapping rules by experts and examine the degree of ambiguity induced to the translation system, as outlined in the following Sections of D6.10. These expert rules are the basis of the algorithm for the production of habitat maps from land cover maps.

3. Land Cover Classification Systems and General Habitat Categories

3.1 Land Cover Classification Systems

The Land Cover Classification System (LCCS) is the result of an initiative to take a first step towards an internationally agreed reference base for land cover [2],[10]. The objectives of the Africover Programme of the Environment and Natural Resources Service (SDRN), FAO, are to develop an approach for conceptualizing, defining and classifying land cover that coincides with the UNEP/FAO initiative on harmonization of land cover and land use classifications.

The LCCS aims at capturing any land cover all over the world, independently of specific applications and/or geographical areas. It is intended to overcome problems due to the interpretation of different land cover class definitions, because, rather than establishing other land cover classes based on nomenclature, it defines a set of independent diagnostic criteria strictly based on vegetation physiognomy and structure, leading to criteria-based land cover classes, compatible with any definition, and allowing for relation with existing classifications and labels [10].

One of the basic principles adopted in the LCCS approach is that a given land cover class is defined by the combination of a set of independent diagnostic attributes, the so-called *classifiers*. This approach allows the use of the most appropriate classifiers and reduces the total number of impractical combinations of classifiers. The increase of detail in the description of a land cover feature is linked to the increase in the number of classifiers used. In other words, the more classifiers added, the more detailed the class will be [2]. The class boundary is then defined either by the different amount of classifiers, or by the presence of one or more different types of classifiers. Thus, emphasis is no longer on the class name, but on the set of classifiers used to define this class.

The classification according to LCCS has two main phases (see Fig. I-1, in Appendix I):

1) The *Dichotomous phase*, where a dichotomous key is used to define eight major land cover types. In the Dichotomous phase of LCCS, three classifiers are used: *Presence of Vegetation*, *Edaphic Condition* and *Artificiality of Cover* to identify the top level LCCS categories. These eight top land cover categories are:

    - *A11: Cultivated and Managed Terrestrial Areas*, which refers to primarily vegetated, terrestrial areas, where the vegetation is artificial and requires human activities to maintain;
    - *A12: Natural and Semi-Natural Terrestrial Vegetation,* which refers to primarily vegetated, terrestrial areas, with natural vegetation, that is not planted, but could be influenced by human actions;
    - *A23: Cultivated Aquatic or Regularly Flooded Areas,* which refers to artificial vegetation, cultivated in aquatic or regularly flooded areas;
    - *A24: Natural and Semi-Natural Aquatic or Regularly Flooded Vegetation,* which refers to natural primarily vegetated, aquatic or regularly flooded areas;
    - *B15: Artificial Surfaces and Associated Areas,* which refers to primarily non-vegetated, terrestrial areas with artificial cover;
    - *B16: Bare Areas,* which refers to primarily non-vegetated, terrestrial areas with no artificial cover;
    - *B27: Artificial Waterbodies, Snow and Ice,* which refers to primarily non-vegetated areas, covered by water due to the construction of artifacts; and
    - *B28: Natural Waterbodies, Snow and Ice,* which refers to primarily non-vegetated areas,that are naturally covered by water.

    The definitions of these LCCS top categories are given in detail in Table I-1 (see Appendix I).

2) The *Modular-Hierarchical phase*, in which land cover classes are created by a combination of sets of pre-defined classifiers, different for each one of the top level LCCS categories, facilitating the definition of more detailed land cover classes and reducing the likelihood of impractical combinations of classifiers. The set of classifiers changes from one class to another. In each set, the classifiers are divided in three groups:
    - *pure land cover* classifiers, which for the primarily vegetated areas (A11, A12, A23 and A24) may include a subset of the *life form, cover, height, spatial aspects, crop combination, leaf type, leaf phenology, stratification* and *water seasonality*, whereas for the primarily non-vegetated areas (B15, B16, B27 and B28) may include one or more of *surface aspect, physical status, persistence, depth, macropattern,* and *sediment load,* as depicted in detail in Fig. I-1 (see Appendix I)*;*
    - *environmental* attributes (e.g., *climate, landform, altitude, soils, lithology, water quality* and *erosion*), which influence land cover, but are not inherent features of it and should not be confused with the pure land cover classifiers; and
    - *specific technical* attributes, which refer to the technical discipline and include the *floristic aspect* for (Semi-)Natural Vegetation, the *crop type* for Cultivated Areas, and the *soil type* for Bare Soil.

In the LCCS classification system the user is not obliged to follow and check the whole series of classifiers and attributes, when no more information is available, or if no more details are required. The major LCCS categories are included in the Table I-2 reported in the Appendix I. The more information (classifiers and attributes) is obtained, the more detailed the land cover class derived will be.

Before starting to use the classifiers, the user has to take into account some basic rules governing the concepts of classification of (semi-) natural vegetation, namely:

- the definition of Life Form; and
- the definition of dominance.

Life Form of a plant is defined by its physiognomic aspect. This is the case when Woody plants, subdivided into Trees and Shrubs, are distinguished from Herbaceous plants, subdivided into Forbs and Graminoids, and Lichens/Mosses.

A condition of Height is applied to separate Trees from Shrubs: woody plants higher than 5m are classified as Trees. In contrast, woody plants lower than 5m are classified as Shrubs. This general rule is subject to the following exception: a woody plant with a clear physiognomic aspect of trees can be classified as Trees even if the Height is lower than 5m, but more than 3m. Plants essentially herbaceous, but with a woody appearance (e.g., bamboos and ferns) are classified as Trees, if the height is more than 5m, and as Shrubs if the height is less than 5m.

Concerning the concept of dominance, the main criterion is the **uppermost canopy layer.** This means that the dominant layer goes from Tree canopy to Shrub to Herbaceous/Forbs/Graminoids. This general rule is subject to a sub-condition of cover. In particular, it is only valid if the dominant Life Form has a cover, either Closed (over 65%) or Open (15 to 65%). If the Life Form is sparse (4-15% cover), the dominance goes to another Life Form that has a Closed or Open cover.

Additional vegetation layers can be added in the *stratification* classifier to determine combination of vegetation classes. An example is the "Tree Savannah", which is clearly defined by two main elements: a Herbaceous vegetation layer and a Sparse Trees layer. Thus, the Stratification of the two elements Herbaceous and Tree layer is crucial for the definition of this class. Some limitations in the use of the classifier Stratification have been introduced in order to avoid class combinations that are irrelevant from the structural point of view. Further details on the Stratification/Layering classifier can be found in [2].

3.2 General Habitat Categories

The General Habitat Categories (GHC) classification system [3] resulted from previous projects, such as BIOHAB (www.edinburgh.ceh.ac.uk/biota/biohab_page.htm), EBONE (http://www.ebone.wur.nl), BIOBIO (http://www.biobio-indicator.org) and GEOBON (http://www.earthobservations .org/cop_bi_geobon.shtml), dealing with surveillance and monitoring of biodiversity, and from the necessity in getting changes in habitat cover and composition. The objectives of GHC are to provide the lowest common denominator for European-wide habitat statistics. It has been developed as the primary structure for recording ecosystems or habitats and is specifically designed for detection and mapping of changes.

This system arises from the need of harmonizing different habitat characterization processes at a continental or global level. Furthermore, GHC applies stringent criteria to ensure that real change is recorded and not results distorted by differences in definitions, between observers or in recording techniques [1],[3].

The GHC approach can be seen as a habitat framework for biodiversity monitoring, while linking national and other higher level, continental classifications. The GHCs are mainly based on Life Forms [1],[11] with added qualifiers on environment, site, management and species composition. The term was developed to include Life Forms (LF). Non-Life Form Habitats (NLF) have been added to complete the system, such as urban categories and sparsely vegetated land. Rock and various categories and combinations of bare ground are considered as habitats in their own right and are especially important in deserts and mountains.

When reporting the GHCs, the map is separated into elements (i.e. areal, linear and point elements) according to their size, as defined by the following rules:

- The Minimum Mappable Element (MME) for an areal element is 400m^2 with minimum dimensions of 5 x 80m or 20 x 20m.
- If the element is narrower than 5m, it is recorded as a linear element with a Minimum Mappable Length (MML) of 30m.
- Elements that do not pass the MME or MML criteria for either areal or linear elements can be mapped and recorded as point elements or as proportions of a larger element.

The key to the GHCs can be applied to any of the three elements, whereas some additional qualifiers for linear elements do exist.

Determination of the GHC is based upon a sequence of five dichotomous divisions (decision tree) related to a set of six super-categories [1], which determine the series of LFs and NLFs used to identify the appropriate GHC (see Fig. II-1, in Appendix II):

- Urban (URB), which refers to 'urban' or 'built-up' land, within the boundary of the land functionally related to buildings, but also refers to parks and recreation areas. It is recognised that the term is not based on Life Forms, but is a land-use division.
- Cultivated (CUL), which refers to habitat currently occupied by herbaceous or woody crops, or bare land with less than 30% cover and evidence of cultivation (recorded with appropriate qualifiers). If the colonizing vegetation has smothered the crop stems (usually 3-7 years), then it should be recorded as life forms only within a qualifier that there was evidence of former cropping.
- Sparsely Vegetated (SPV), which refers to bare areas and is recorded as such only if the vegetation cover is less than 30%. The Lichens and Bryophytes are not assumed as vegetation, but instead they would be included as a life form qualifier.
- Trees and Shrubs (TRS), which refers to LFs and includes the woody habitats, trees and shrubs. It also includes the habitat of species, which do not have secondary ligneous woody thickening in

strict botanical sense, but have a shrubby form and perennating buds above ground level. Height is therefore the only consistent arbiter.

- Herbaceous wetland (HER), which refers to habitat of plants that grow in aquatic or waterlogged conditions, and
- Other Herbaceous (Other HER), which refer to all the other herbaceous species, like forbs and graminoids.

The division of each super-category in LFs and NLFs as well as its subdivisions and life form qualifiers are presented in Fig. II-2 (see Appendix II). The GHC codes and all possible combinations within each of the super-categories can be given in any mapping element, as described in detail in the EBONE Handbook [1]. No other combinations are possible than those reported in the Handbook, i.e. there is a maximum number of 140 GHCs.

After the identification of the GHC super-category using the rules described in the dichotomous phase (Fig. II-1, Appendix II), the determination of the GHCs subcategories is achieved based on two percentage rules: over 70% for single GHCs or 40-60% for GHCs that are combinations of two habitats, only within the possible combinations. In particular, an element with >70% cover of a single LF or NLF category is a GHC with a single code (e.g. ART or HEL) or a double code if the GHC belongs to the TRS super-category (e.g. FPH/CON and FPH/DEC), where CON and DEC are denominators for the TRS category. On the other hand, elements with 40-60% cover of two life forms or two non-life form categories belonging to the same super-category are also GHCs, but with a double code (e.g. ROC/GVR or SHY/EHY) or with a triple code, if belonging to the TRS super-category (e.g. mixed Deciduous/Conifer Forest, FPH/DEC/CON). If there are equal proportions of LFs, then precedence rules are provided. The precedence will be given in the order of the GHCs as listed in Fig.II-2 (see Appendix II), or in case of trees to the tallest category over 30% cover (e.g. if an element has a coverage of TPH/DEC 30, TPH/EVR 30, MPH/CON 30, LPH/CON 10, then precedence is given to the ranking above, i.e. TPH/DEC/EVR). Further details on the abbreviations are provided in the Handbook [1].

As analytically explained in the Handbook, the GHCs contain in themselves information about life form, leaf type, phenology and height. The primary sources for the Life Forms have been various floras. The height categories have been designed to fit in with previous work, especially in the Mediterranean literature. Other qualifiers can be also added to express variations between elements that may have the same GHC, to identify the habitat type. Such qualifiers are

- *additional life form qualifiers*, which refer to species that develop on the ground below or between other Life Forms important for the description of the habitat;
- *environmental qualifiers*, which apply only on the TRS, HER and Other HER super-categories and refer to soil moisture, soil salinity, acidity and eutrophy;
- *global codes,* which refer to the setting of an element (height or scattered trees) or reference previous data;
- *site qualifiers,* which characterize the geomorphology, archaeology and life form complexity of elements in order to express variations between elements with the same primary code; and
- *management qualifiers,* which are organised in several levels, the first level being the time of the management, the second level the general categories where management is taking place (e.g. forest or urban), and the third level is a more specific management activity.

The complete list of the qualifiers is provided in the Handbook [1] and is used to further refine the habitat recording. The GHCs are specifically designed to be recorded consistently especially for detection and mapping of changes. Furthermore, this system applies stringent criteria to ensure that real change is recorded and not results distorted by differences in definitions, between observers or recording techniques. One of the key elements of this approach is its potential for the detection and evaluation of flows between habitats [3].

3.3 Relation between the two taxonomies

To create a framework for the translation of LCCS maps into GHC maps and secure minimum error propagation in the final system, the correspondence between the two taxonomies needs to be established. However, there are some difficulties in the translation between the two systems because of differences in definitions and interpretations of the landscape, even though both have been designed to meet the fundamental attributes for global vegetation classification [12]. In particular, LCCS focuses on describing land cover, whereas GHC also employs the land use. However, even at the land cover level, the definition of LFs according to FAO-LCCS is based on a classification of trees, shrubs and herbs (graminoids and forbs), the definitions of which differ from that adopted in the GHC framework [3],[11]. The main differences between the two classifications can be found in the variation in the height of the vegetation and the rules relating to the percentage of vegetation cover. There are issues also with the existence of one-to-many or many-to-one relationship, scale and the need to define recognisable features in the landscape as outlined below.

3.3.1 Height

The height of vegetation in the definition of LCCS and GHC classes is listed in Table 3-1. As an example of the discrepancies between the two systems, the height ranges defined by the GHC for chamaephytes and phanerophytes do not correspond exactly with those defined by LCCS for trees and shrubs. Moreover, the LCCS defines different ranges of height for herbaceous types, whereas these ranges are not provided in GHC. These discrepancies most likely arise because of the differences in end-user requirements. For example, the GHCs were set-up by ecologists from an ecological perspective and have a basis in the UK countryside survey, while the LCCS was set up by land cover, forest and remote sensing specialists with different backgrounds compared to ecologists. These discrepancies and different perspectives on the landscape must be taken into consideration, when designing the algorithm for the transition from LCCS to GHC maps.**Cover**

The differences in definition of cover between the two schemes are listed in Table 3-2. In terms of vegetation cover, the LCCS taxonomy requires that vegetation cover exceeds 4% whilst the GHC requires over 30% cover. These definitions lead to very different results in the final classification when these two schemes are applied. Both taxonomies make use of both the uppermost canopy layer and the percentage of vegetation cover, but at a more detailed level many discrepancies do exist. A representative example of this could be a natural area consisting of 84% herbaceous-forbs and 16% cover of dispersed trees. According to the *uppermost canopy layer* rule in the LCCS, this area would be recorded as A12 *Natural and semi-natural terrestrial vegetation,* A3 *Woody Trees*_A11 *Open (15-65%)* and the forbs would be recorded within the *stratification/layering* qualifier. On the contrary, according to the *uppermost canopy layer* rule in the GHC (see Fig. II-1, in Appendix II), the percentage of Trees is not adequate enough to be recorded as the primary vegetation, so precedence would be given to the lower vegetation. In particular, it would be recorded within the *Other HER* super-category and the trees would be included in the *OPE* qualifier. Moreover, if instead of trees, the 16% was shrubs these would be ignored in the GHC.

Table 3-1: Height definitions in the LCCS and GHC categories

LCCS classification	GHC classification
	>40m Mega Forest Phanerophytes
14->30m B5 Trees	5-40m Forest Phanerophytes
7-14m B6 Trees	
3-7m B7 Trees	2-5m Tall Phanerophytes
3-5m B8 Shrub	
0.5-3m B9 Shrub	0.6-2m Mid Phanerophytes
0.3-0.5m B10 Shrub	0.3-0.6m Low Phanerophytes
	0.05-0.3 Shrubby Chamaerophytes
	<0.05cm Dwarf Chamaephytes
0.8-3m B11Herbaceous	
0.3-0.8m B12 Herbaceous	
0.03-0.3m B13 Herbaceous	

Table 3-2: Cover conditions in the LCCS and GHC categories

Species category	LCCS classification	GHC classification
Trees	>4%, as the primary class >15%	>1%, as the primary class >30%
Shrub	>4%, as the primary class >15%	
Herbaceous	>4%, as the primary class >15%	>30%
Lichens	>25%	only as a qualifier

3.3.2 One-to-many relationships

When only land cover classifiers are used to define the LCCS classes, in a number of instances, several GHC classes can be contained within one LCCS class and therefore, it is difficult to relate the two classifications, even at the life form level. Therefore, in order to facilitate translation, the LCCS map must contain as much detail as possible. For instance, if the stratification qualifier is not recorded in the previous example, the transition to the correct GHC would be impossible and the system would probably fail. Thus, the relation between the two classifications is susceptible to the scale of information available. Moreover, some classes (e.g., water) can also be considered to be both artificial and natural. For example, straightened channels (e.g., for freshwater or brackish water) are artificial, but these often link to the natural system (e.g., estuaries or unstraightened channels up or downstream). Artificial pools may be constructed to recreate, restore and eventually become part of a 'natural' system (e.g., active raised bog). Assignment to a specific LCCS class is sometimes difficult. When translating existing land cover maps to LCCS classes, difficulties arise if the original mapping is relatively coarse as the accuracy of delineation and classification is compromised; hence, in reality, defined polygons may contain several LCCS classes. Therefore, new classifications may have to be developed with these focusing on direct assignment to an LCCS class, which can then be translated subsequently to GHC classes.

3.3.3 Scale

The relationships between the two classification schemes are dependent upon the scale and hence the spatial resolution of satellite or airborne systems. In particular, the linear elements, which are especially important to biodiversity in managed landscapes [3] and would be recorded with additional qualifiers in the GHC, might be considered as a part of the adjacent classes in the LCCS classification. In many cases, the spatial resolution of the observing sensor needs to be high to very high in order to allow discrimination of LCCS classes and information on vegetation height (e.g., as derived from airborne LiDAR) may become essential. In other environments (e.g., the active raised bog of Cors Fochno in mid Wales), a complex mosaic of woody shrubs, graminoids and lichens occurs and several LCCS classes might be assigned to this habitat depending on the spatial resolution of the observing sensor. Sensor resolution may limit class size and specification of what a class actually consists of. In such cases, and particularly where the resolution is not high, mixed pixels will be commonplace, and the reliability of the selection of the LCCS code is questionable. If an individual pixel happens to lie completely within, or fortuitously coincides with, the boundaries of a given class, then the multiband spectral properties of the dominant material(s) in the enclosed class will determine the multiband digital numbers for that pixel. In many cases though, the pixel will straddle or cut across several class or feature boundaries. The resulting spectral content will then be a function of the spectral responses from each internal class. Recognition of each feature or class becomes difficult, since there are two primary unknowns to account for - the identity of the classes and their relative proportions in the mix.

Mathematical methods are available to solve for these unknowns, but there always remains some statistical uncertainty. As a general rule, the level of accuracy obtainable in a remote sensing classification depends on diverse factors, such as the suitability of the size, shape, distribution, and frequency of occurrence of individual areas assigned to each class, which together determine the degree to which pixels are mixed, the sensor performance and resolution, and the methods involved in classification. Mapping of LCCS classes is expected to be more reliable, where data acquired at very high spatial resolution are used. In many cases, the assignment of an LCCS class is difficult where the spatial resolution of the observing sensor is insufficient. Hence, a system whereby LCCS classes that can be differentiated at different (and typical) spatial resolutions might have to be developed to facilitate the extraction of more detailed habitat maps.

3.3.4 Linking with landscape features

In many cases, particularly in human-managed landscapes, distinct boundaries are evident (e.g., those associated with field boundaries and urban areas). A key component is, therefore, to define these boundaries, either within the image itself or through reference to other datasets (e.g., land parcel boundaries or building locations). By identifying such features, differentiation between artificial and natural land covers can be better achieved. The inclusion of such information is strongly recommended to resolve ambiguity. A satellite-based classification that concurs with the spatial layout of the landscape (with hedgerows etc.) is needed and information on easily identifiable boundaries (e.g. fields) needs to be taken into consideration.

4. An algorithm for habitat production maps

The contribution of the experts is crucial for the determination of the most appropriate mapping between the two taxonomies (LCCS and GHC) on the terminology level. The basis of the algorithm for the production of GHCsmaps from LC maps consists of expert rules, which at the top level link the LCCS classes, defined by using only pure land cover classifiers, with all the possible corresponding habitat types, based not only on their definitions, but their potential for discrimination using RS data as well. Additionally, the inspection of all the available information deduced by the first two stages of the EODHaM system and the detailed description of the 3rd stage module requirements can contribute to the design of an optimum algorithm.

In the following section (4.1), the expert rules for mapping between the LCCS and GHC classes, based strictly on the definitions, are considered. The production of the habitat maps is completed in three stages. At the first stage, the algorithm makes a mapping from the top classes of LCCS to the top classes of GHC (Section 4.1.1) to exclude all the mismatches between the two taxonomies. At the second stage, the mapping expands to the more analytical categories contained within the LCCS and GHC, according to the expert rules (Section 4.1.2) and within a fuzzy scheme as described by the architecture design (Section 4.2). This is further enriched at the third stage, where the algorithm is refined with the incorporation of ancillary data and elaborated classification techniques (Section 4.3).

4.1 Expert rules for mapping the LCCS to the GHC classes

As previously described in detail, the two taxonomies differ in their definitions to such an extent that the need of expert rules and harmonization between LCCS and GHC appears to be urgent. A major contradiction occurs in the percentage of vegetation cover, which in order to be mapped in the LCCS taxonomy should be more than 4%, whereas in the GHC case should be over 30%. The measurement of the percentage of vegetation cover for determining the LCCS classes as detected by the RS images is per se a very difficult task, with high error variance, posing fuzziness in the selection of the correct class. This can be perceived as the reliability of the measurement. Taking this into consideration, it is agreed to ignore this discrepancy between the two taxonomies, when the percentage of vegetation cover is around 20% and to use the SCA and OPE global qualifiers codes of the GHC taxonomy [1] in the case of percentages of tree cover lower than 10%. In the latter, if the *stratification* qualifier points out the land cover, this would be used to identify the GHC code. Alternatively, the output of the 1st stage should be considered to identify the surface content and specify the land cover. Then the corresponding GHC code is extracted along with the SCA/OPE qualifier using the expert rules.

The expert knowledge on botanical definitions is used to provide the project with the framework for the mapping of LC classes into habitat classes. The following sections describe the reasoning on which all the possible relations between the LCCS and GHC codes can be found.

4.1.1 Top level categorical mapping

The top level classes of the LCCS are mapped into the GHC super-categories according to the expert knowledge as tabulated in Table 4-1. On the left column, the 8 top LCCS categories are listed, while the right column includes the potential GHC super-categories. The acronyms of the six GHC super-categories are bolded. In the parentheses, although not necessary at the first stage of the software, possible sub-categories are presented, whose full names can be found in Appendix II. Where there is no parenthesis after a GHC super-category acronym, it is implied that all its possible sub-categories could be present.

A general reasoning of the mapping is provided below.

- *Cultivated and managed terrestrial areas (A11)*. This category may include either cultivated areas (*CUL*), urban/constructed areas (*URB*) with the sub-categories that include some form of vegetation, or trees (*TRS*) with lower probability. All other GHC super-categories are excluded, since they are either (semi-)naturally vegetated or non-vegetated.
- *Natural and semi-natural terrestrial vegetation (A12)*. Non-aquatic vegetated GHC super-categories correspond to *A12* LCCS class, i.e. trees and shrubs (*TRS*) or non-wetland herbaceous (*Other HER*). Additionally, GHC may characterize an area as sparsely vegetated (*SPV*) in the case of lichens which are identified as a qualifier in this taxonomy, while according to the LCCS framework an area is characterized as vegetated if having more than 25% lichens. In the GHC category the *URB* should be also considered due to the assignment of semi-natural fields to parklands, which may not be identified as such in the LCCS derived from a RS image.
- *Cultivated aquatic or regularly flooded areas (A23)*. This category refers to aquatic plantations such as rice and is therefore directly linked to the GHC *CUL* class. Other non-harvested aquatic plantations are considered impossible or with extremely low probability in the GHC category.
- *Natural and semi-natural aquatic or regularly flooded vegetation (A24)*. Wetland herbaceous (*HER*) habitats belong in this category. Trees and shrubs (*TRS*) and other herbaceous (*Other HER*) growing in such aquatic conditions, if existing, may belong in this LCCS category as well as deduced in D6.1, but should be identified as less possible by the algorithm. The *URB* should also be included here considering the land use, as described in the *A12 Natural and semi-natural terrestrial vegetation* above, but the probability of such mapping will be much lower than before.
- *Artificial surfaces and associated areas (B15)*. The non-vegetated parts of the *URB* super-category are included here, since this category refers exclusively to the artificial areas.
- *Bare areas (B16)*. Terrestrial sparsely vegetated (*SPV*) areas correspond in this LCCS category.
- *Artificial waterbodies, snow and ice (B27)*. Waterbodies will most frequently correspond to the SPV category. Where waterbodies with concrete base occur, these will fall into the *URB (ART)* GHC class. This is due to the fact that in the GHC classification only the areal elements of more than 400m^2 are considered, which is rare in natural fields, and the waterbodies within an urban area are not recorded in GHC. Artificial areas of snow and ice are considered solely within the *SPV* GHC category.
- *Natural waterbodies, snow and ice (B28)*. The respective non-vegetated non-artificial areas fall also into the *SPV* GHC super-category.

As deduced from the inspection of Table 4-1, there are one-to-one relations between the two taxonomies for the main categories A23, B15, B16 and B28 and one-to-many relations for the A11, A12, A24 and B27. This one-to-many relations are induced mainly because of the introduction of the land use in the categorization of the GHC along with the land cover, which is the basis of the LCCS. For instance, a field of planted fruit trees belongs to the LCCS *A11 Cultivated and managed terrestrial areas* class, but it could correspond either to the GHC *CUL* if it is an agricultural plantation, or the GHC *URB* if it is near a building and is intended for personal use only. Additionally, a coniferous plantation that falls also under the LCCS *A11 Cultivated and managed terrestrial areas* class, though managed would be characterized as the GHC *TRS,* since it is not harvested.

Table 4-1: LCCS to GHC preliminary mapping

LCCS main categories	GHC super-categories
A11 Cultivated and managed terrestrial areas	**CUL** / **URB**-(VEG/GRA/TRE) / **TRS / Other HER**
A12 Natural and semi-natural terrestrial vegetation	**SPV** (LIC) / **TRS** / **Other HER** / **URB**-(VEG/GRA/TRE)
A23 Cultivated aquatic or regularly flooded areas	**CUL**-(CRO)
A24 Natural and semi-natural aquatic or regularly flooded vegetation	**HER** / **TRS** / **Other HER** / **URB**-(VEG/GRA/TRE)
B15 Artificial surfaces and associated areas	**URB**-(ART/NON)
B16 Bare areas	**SPV**-(ROC/BOU/STO/GRV/SAN/EAR)
B27 Artificial waterbodies, snow and ice	**SPV**-(AQU/ICE) / **URB**-(NON/ART)
B28 Natural waterbodies, snow and ice	**SPV**-(SEA/AQU/ICE)

4.1.2 Mapping the Life-Form classifiers

The approach discussed above constitutes the preliminary mapping rules from the 8 top LCCS categories to the 6 GHC super-categories, necessary for the development of the first stage of the algorithm for the conversion of LC maps to habitat maps. In the same notion, the second stage focuses on the mapping at the level of life forms with the incorporation of further expert rules and ancillary data. The crude expert rules, which are tabulated in Table 4-2, form the basis of the algorithm, whereas further details concerning the life form qualifiers, i.e., height, leaf type, leaf phenology etc, are also included in the classification module.

As observed in Table 4-2, the mapping at the level of life forms also includes one-to-many matching, mainly due to the discrepancies in the botanical definitions between the two taxonomies, as described in Section 3.3. A representative example is the case of a garden with fruit trees, such as black currents and gooseberries, less than 3 meter tall. This will be assigned with the label *A11.A2 cultivated shrubland* in the LCCS taxonomy and consindering the land use as *CUL(WOC)* or *URB-TRE* (if more than 0.6m tall) or *URB-VEG* (if less than 0.6m height) if land use is considered. So, one single LCCS class could correspond to up to three GHC classes as noted in the D6.1 deliverable, where the LCCS *A12.A4.A10.B3.B9 medium/high shrubland* (B9 refers to height between 0.5-3m as shown in Table I-2 in Appendix I) could be either *TRS-LPH*(0.3-0.6m) or *MPH*(0.6-2m) or *TPH*(2-5m) (see Fig. II-2 in Appendix II for further details), with *TRS-MPH* being the most probable. This reveals the urgent need to take into consideration the probability of appearance and the reliability of measurements of phenological and structural details, when designing the algorithm to produce the GHC habitat maps from LCCS maps.

The one-to-many relations appear even in the case of the NLFs, where there is also inconsistency induced by definitions, e.g. the *gravel* is defined as coarse fragments having a size less than 6 cm within the LCCS taxonomy, whereas in the GHC as having diameter less than 5cm. Such details cannot be resolved using remote sensing data and only a coarse assumption of the LC class is possible. Strictly based on the definitions and not taking into consideration the capability of a RS system, the *B16.A1_A14 Gravel* will correspond to the *SPV-GRV* or the *SPV-STO,* assigning a very low probability (p=0.01) to the latter. Taking into consideration the uncertainty in the measurement of the size, the *B16.A1_A14 Gravel* in the GHC would only correspond to *SPV-GRV.*

Moreover, classes such as the hardpans recorded in the LCCS under *B26 Bare areas_A1 Consolidated _A4 Hardpans* would be recorded in the GHC only based on the composition of the ground, most probably as *SPV-EAR.* The *_A9 Ironpan/Laterite* would be specified only through additional qualifiers in the GHC taxonomy. The determination of these qualifiers is not a trivial issue, and the contribution and feedback of experts in the area is further needed in order to validate or correct and improve this mapping effort.

Finally, it should be noted that in the GHC mapping the codes may appear both distinctly or as one of the valid combinations as described in detail in the EBONE Handbook [1]. These combinations are also included in the expert rules and considered in the design of the algorithm. This increases the ambiguity in the output classes, but provides a more detailed habitat map, enhancing the monitoring of the biodiversity.

Table 4-2. LCCS to GHC mapping on the Life Form level

LCCS to GHC mapping	
LCCS class	**GHC possible codes**
A11 Cultivated and managed terrestrial areas	CUL / URB-(VEG/GRA/TRE) / TRS-(FPH/TPH) / Other HER
A1 Trees	CUL-(WOC) / URB-(TRE) / TRS-(FPH/TPH)
A7 Trees-Broadleaved	CUL-(WOC) / URB-(TRE) / TRS-(FPH/TPH)
A8 Trees-Needleleaved	CUL-(WOC) / URB-(TRE) / TRS-(FPH/TPH)
A9 Trees-Needleleaved-Evergreen	CUL-(WOC) / URB-(TRE) / TRS-(FPH/TPH)
A10 Trees-Needleleaved-Deciduous	CUL-(WOC) / URB-(TRE) / TRS-(FPH/TPH)
A2 Shrubs	CUL-(WOC) / URB-(VEG/TRE)
A7 Shrubs-Broadleaved	CUL-(WOC) / URB-(VEG/TRE)
A8 Shrubs-Needleleaved	CUL-(WOC) / URB-(VEG/TRE)
A9 Shrubs-Needleleaved-Evergreen	CUL-(WOC) / URB-(VEG/TRE)
A10 Shrubs-Needleleaved-Deciduous	CUL-(WOC) / URB-(VEG/TRE)
A3 Herbaceous	CUL-(CRO) / URB-(VEG/GRA) / Other HER
A4 Herbaceous-Graminoids	CUL-(CRO) / URB-(GRA) / Other HER(CHE)
A5 Herbaceous-Non Graminoids	CUL-(CRO) / URB-(VEG/GRA) / Other HER (except CHE)
A6 Urban vegetated areas	URB-(GRA/TRE)
A11 Urban-parks	URB-(GRA/TRE)
A12 Urban-parklands	URB-(GRA/TRE)
A13 Urban-lawns	URB-(GRA/TRE)
A12 Natural and semi-natural terrestrial vegetation	SPV-(LIC) / other HER / TRS / URB-(VEG/GRA/TRE)
A1 Woody	TRS / URB-(VEG/TRE)
A3 Woody-Trees	TRS-(TPH/FPH/GPH) / URB-(TRE)
A4 Woody-Shrubs	TRS-(DCH/SCHLPH/MPH/TPH) / URB-(VEG/TRE)
A2 Herbaceous	(weak)TRS-(DCH/SCH) / other HER-(LHE/CHE/THE/HCH/GEO) / URB-(VEG/GRA)
A5 Herbaceous-Forbs	(weak)TRS-(DCH/SCH) / other HER-(LHE/CHE/THE/HCH/GEO) / URB-(VEG/GRA)
A6 Herbaceous-Graminoids	other HER-(CHE) / URB-(GRA)
A7 Lichens/Mosses	SPV-(LIC) / other HER-(CRY)
A8 Lichens	SPV-(LIC) / other HER-(CRY-LIC)
A9 Mosses	SPV-(LIC) / other HER-(CRY-BRY)
A23 Cultivated aquatic or regularly flooded areas	CUL-(CRO)
A1 Graminoid Crop	CUL-(CRO)
A2 Non-Graminoid Crop	CUL-(CRO)

Continued on the next page...

Table 4-2. - Continued

LCCS to GHC mapping	
LCCS class	**GHC possible codes**
A24 Natural and semi-natural aquatic or regularly flooded vegetation	HER / Other HER-(CRY) / TRS / URB-(VEG/GRA/TRE)
A1 Woody	TRS / URB-(VEG/TRE)
A3 Woody-Trees	TRS-(TPH/FPH/GPH) / URB-(TRE)
A4 Woody-Shrubs	TRS-(DCH/SCH/LPH/MPH/TPH) / URB-(VEG/TRE)
A2 Herbaceous	HER
A5 Herbaceous-Forbs	HER
A8 Herbaceous-Forbs-Rooted	HER-(SHY/EHY/HEL/-LEA)
A9 Herbaceous-Forbs-Free floating	HER-(EHY-FLO)
A6 Herbaceous-Graminoids	HER
A7 Lichens/Mosses	Other HER-(CRY)
A10 Lichens	Other HER-(CRY-LIC)
A11 Mosses	Other HER-(CRY-BRY)
B15 Artificial surfaces and associated areas	URB-(ART/NON)
A1 Built-up	URB-(ART/NON)
A3 Linear	URB-(ART/NON)
A7 Linear-Roads	URB-(ART/NON)-(ROA/TRA)
A8 Linear-Roads-Paved	URB-(ART)-(ROA)
A9 Linear-Roads-Unpaved	URB-(NON)-(TRA)
A10 Linear-Railways	URB-(ART)
A11 Linear-Comm.Lines/Pipelines	URB-(ART/NON)
A4 Non-Linear	URB-(ART/NON)
A12 Non-Linear-Industrial a/o other	URB-(ART/NON)
A14 …High density	URB-(ART)
A15 …Medium density	URB-(ART)
A16 …Low density	URB-(ART/NON)
A13 Non-Linear-Urban areas	URB-(ART/NON)
A14 …High density	URB-(ART)
A15 …Medium density	URB-(ART)
A16 …Low density	URB-(ART/NON)
A2 Non Built-up	URB-(NON)
A5 Non Built-up -waste dump deposits	URB-(NON)
A6 Non Built-up -extraction Sites	URB-(NON)

Continued on the next page...

Table 4-2: Continued

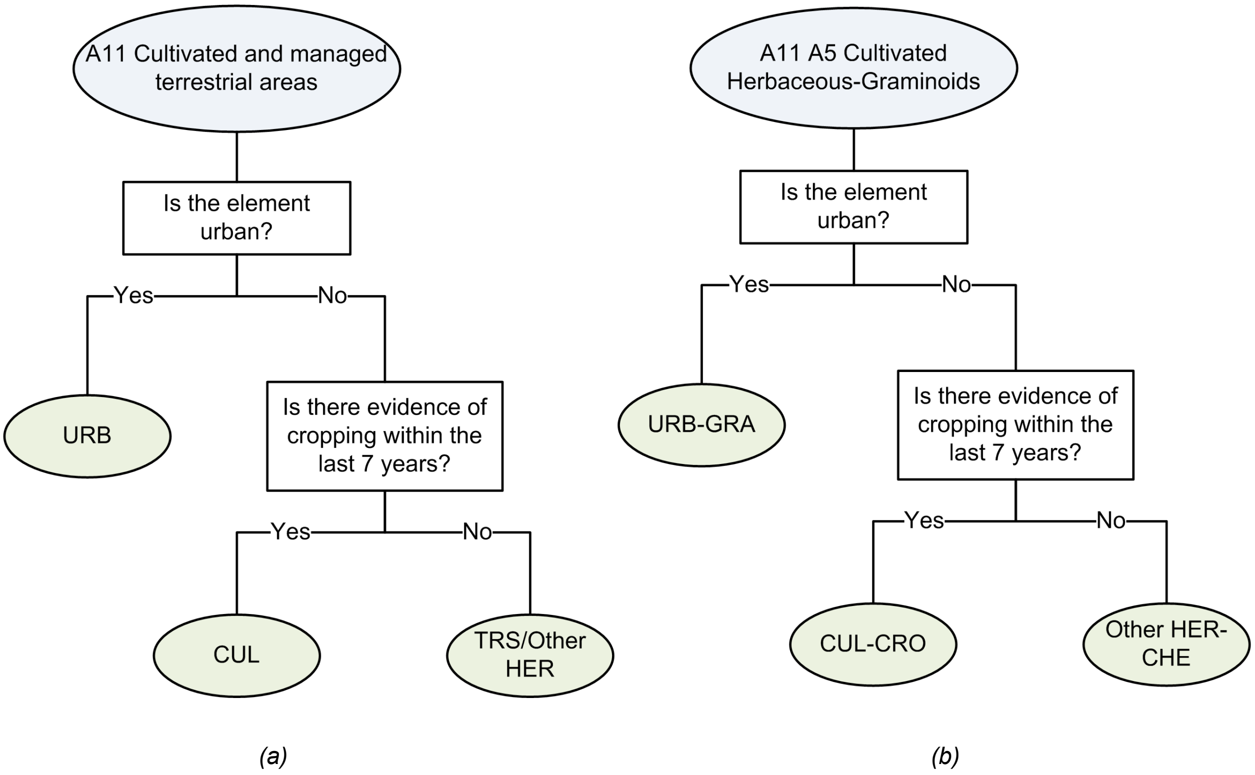
LCCS to GHC mapping	
LCCS class	**GHC possible codes**
B16 Bare areas	SPV-(ROC/BOU/STO/GRV/SAN/EAR)
A1 Consolidated	SPV-(ROC/BOU/STO/GRV/EAR)
A3 Consolidated-Bare rock a/o coarse fragm.	SPV-(ROC/BOU/STO/GRV)
A7 Consolidated-Bare rock a/o fragm-Bare rock	SPV-(ROC)
A8 Consolidated-Bare rock a/o fragm-Gravel/Stones/Boulders	SPV-(BOU/STO/GRV)
A14 …Gravel	SPV-(GRV)
A15 …Stones	SPV-(STO)
A16 …Boulders	SPV-(BOU)
A4 Consolidated-Hardpans	SPV-(EAR)
A9 Consolidated-Hardpans-Ironpan/Laterite	SPV-(EAR)
A10 Consolidated-Hardpans-Petrocalcic	SPV-(EAR)
A11 Consolidated-Hardpans-Petrogypsic	SPV-(EAR)
A2 Unconsolidated	SPV-(ROC/BOU/STO/GRV/SAN/EAR)
A5 Unconsolidated-Bare Soil a/o other Uncon. Mat.	SPV-(STO/GRV/EAR)
A12 …Stony (5-40%)	SPV-(STO/GRV/EAR)
A13 …Very Stony (40-80%)	SPV-(STO/GRV)
A6 Unconsolidated-Loose and Shifting Sands	SPV-(STO/SAN)
A12 …Stony (5-40%)	SPV-(STO/SAN)
A13 …Very Stony (40-80%)	SPV-(STO/SAN)
B27 Artificial waterbodies, snow and ice	(weak)URB-(ART/NON) / SPV(AQU/ICE)
A1 Artificial Waterbodies	(weak)URB-(ART/NON) / SPV(AQU)
A4 Artificial Water-Flowing	(weak)URB-(ART/NON) / SPV(AQU)
A5 Artificial Water-Standing	(weak)URB-(ART/NON) / SPV(AQU)
A2 Artificial Snow	SPV-(ICE)
A3 Artificial Ice	SPV-(ICE)
A6 Artificial Ice-Moving	SPV-(ICE)
A7 Artificial Ice-Stationary	SPV-(ICE)
B28 Natural waterbodies, snow and ice	SPV-(SEA/AQU/ICE)
A1. Natural Waterbodies	SPV-(SEA/AQU)
A4 Water-Flowing	SPV-(SEA/AQU)
A5 Water-Standing	SPV-(AQU)
A2 Snow	SPV-(ICE)
A3 Ice	SPV-(ICE)
A6 Ice-Moving	SPV-(ICE)
A7 Ice-Stationary	SPV-(ICE)

4.2 Algorithm design for habitat map production

The main inputs of the algorithm are LC maps, following the FAO-LCCS scheme. The map is converted to habitat maps, through a one-to-one and one-to-many mapping, i.e. each LC area is classified to a specific habitat, or a combination of potential habitats, according to the expert rules described above.

The algorithm is designed as a rule-based system, which also incorporates input uncertainty (reliability of the 2nd stage classifier) to work out the classification of the type of habitat. Thus, decision trees are employed to built the expert rules within a tree-like structure that predict the target variable in the classification module. Decision trees are intuitive and easy to understand, but also flexible to modifications and collaboration with other methods. Each LC class is the root of a decision tree, whose leaves are all the possible habitat classes that correspond to the specific LC map. An indicative example of the crisp decision trees translating the expert rules is depicted in Fig. 4-1(a),(b), where the one-to-many relations between the two classification schemes are evident.

Moreover, decision trees can be used to analyse complex statistical and probabilistic situations. As proposed by the expert rules, a single LCCS class can be mapped to various GHC classes, which do not necessarily occur with the same probability. To account for the heterogeneity in the output, the branches of the decision trees are determined to be probabilistic. A common problem when using prior probabilities is that they can bias the posterior probability of the output classes. This bias is often induced due to errors and/or uncertainty in the information used to prescribe these prior probabilities. The uncertainty of classification can be reliably estimated within a Bayesian averaging technique [13]. Therefore, Bayesian methods are intended to be used to address this problem, i.e. decrease the effect of prior probabilities on the classification results and resolve the ambiguity in the output. An important aspect of this method is that knowledge regarding the quality of ancillary information can be incorporated



(a) *(b)*

Figure 4-1 Decision tree example for the (a) the subroutine (1) and (b) the subroutine (4).

into Bayes' rule, providing a refinement of the algorithm and facilitating more accurate mapping from LCCS to GHC classes.

Accordingly, the expert rules include the mapping rules and the probability of assessment in the input (e.g. errors in the identification of vegetation cover and height of plantation). Additionally, the reliability of the process by which it was estimated will be taken into consideration, as depicted in Fig. 4-2. The reliability of measurement refers to the uncertainty in the output (e.g. the uncertainty of the appearance of a specific habitat in the landscape). The software design representing the functionality of the algorithm in an abstract way is simple, comprises well defined, independent components and is easily extensible to future added features.

Four subroutines have been designed according to the level of detail of the input and output data.

1. The first procedure takes as input LC maps containing information on solely the 8 top LCCS classes arisen from the dichotomous phase. The output of the procedure is a habitat map comprising the potential GHC super-category, or combinations of them, that may result for each LCCS class. A decision tree representing this scheme is shown in Fig. 4-1(a).
2. In the second procedure, more detailed output at the level of GHC life-forms is attempted, having as starting point the 8 top LCCS categories. The error deviation in this case is prohibitive, but in case of absent data, this module would be used to specify the habitat class.
3. The uncertainty of classification is reduced in the third procedure, where the input includes apart from the 8 top LCCS classes, additional life form classifiers. The classifiers have the ability to specify with a greater accuracy the potential GHC super-categories that may result from a specific LCCS class, thus reducing the total number of ambiguous classes in the habitat level.
4. The fourth procedure takes as input an LC map of the same level of detail as before, but expresses the conversion result not only at the level of the GHC super-categories, but also at the level of potential GHC life and non-life form categories. A representative example is depicted in Fig. 4-1(b). The potential output can be weighted, according to the expert rules, facilitating an integrated solution to the mapping problem.

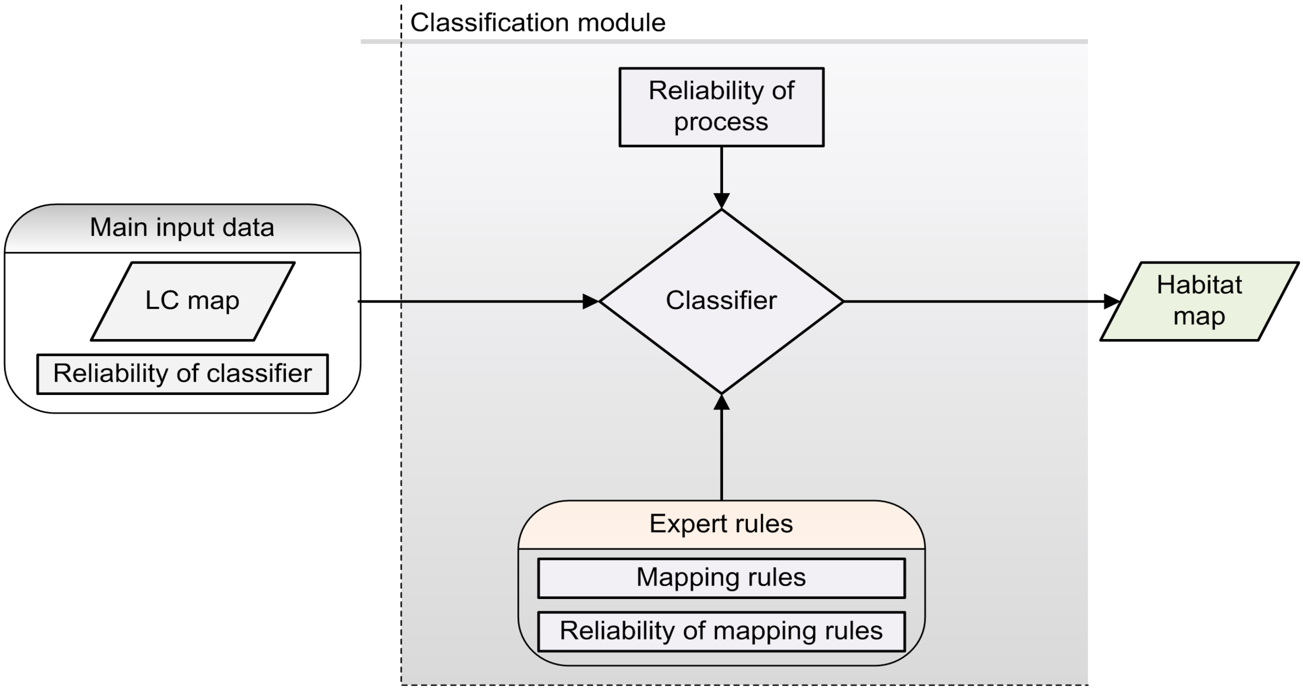


Figure 4-2 Classification module for the production of habitat maps from LC maps

4.3 Algorithm refinement

The algorithm for the conversion of LCCS land cover maps into GHC habitat maps is a fundamental part of the EODHaM system architecture. The first version of the algorithm, as described in this report, is already operational including the core structural elements and constitutes the skeleton upon which future versions will build. The algorithm will keep on evolving together with supplementary activities in BIO_SOS throughout the duration of the project. As new information is derived from interacting modules of the EODHaM system, new features will be incorporated in the algorithm to elaborate the produced habitat maps and minimize the uncertainty in the classification.

To a certain degree, the refinement of the algorithm will proceed in parallel with the evolution and refinement of the 2nd stage of the EODHaM, where the algorithm will take its main input from. The more semantic the information, including contextual and temporal, generated from the 2nd stage, the more additional data will be available for the refinement of the algorithm. In addition, the system will be further trained to include additional information from external sources and incorporate supplementary mapping and classifying rules.I In a further level the accuracy of both input data and expert rules will be questioned and fuzzy classification schemes will be adopted to improve the resulting habitat maps.

The upcoming ameliorations of the algorithm can be organized into three feature categories shown below, not necessarily chronologically implemented in the order presented here. The final architecture of the LC to habitat conversion system is depicted in Fig. 4-3.

1. Incorporation of ancillary data. Apart from the core LCCS map extracted from the 2nd stage of the EODHaM system, additional available data will be exploited in order to improve classification accuracy and minimize uncertainty. Ancillary data include:

   - Data from external sources. Any additional information concerning the landscape to be translated into habitats, mainly in the form of additional GIS layers, may be incorporated in this framework. As an example, cadastral maps might be used to discriminate between managed and natural or semi-natural habitat categories. Digital elevation models (DEM), when available, might decrease classification uncertainty, by excluding habitats not occurring at specific elevations, slopes or aspects. Similar information can be extracted from data on soil or lithology structure of the landscape, climatic conditions of the area, dominant species, etc.

   - Semantic information from the EODHaM 2nd stage. The information from the semantic net and the ontologies generated at the 2nd stage of the EODHam system will provide further information as far as spatial or temporal relations between LC classes, or inherited properties from parental classes, is concerned. The more meaningful the semantic net becomes, the more informative the input of the algorithm and its capabilities in intraclass and interclass discrimination.

   - Data from previous stages of EODHaM. One of the structural features and advantages of the EODHaM system is the preservation of intermediate results, which might be useful in further processes. This mainly includes the output from the 1st stage, where spectral pixel-based information on the landscape is generated and can be exploited in certain cases as additional data during the conversion from LC classes to habitats.

   The available ancillary data will be considered by the system through additional rules provided by the experts. Various thresholds and decision rules will be applied to refine the distinction between potential habitat classes.

2. Refinement of expert rules. As more data become available, existing mapping rules might be modified, while new rules will be added to the algorithm. A semantic net will be deployed in the habitat level linking the different GHC super-categories, life and non-life forms (nodes of the

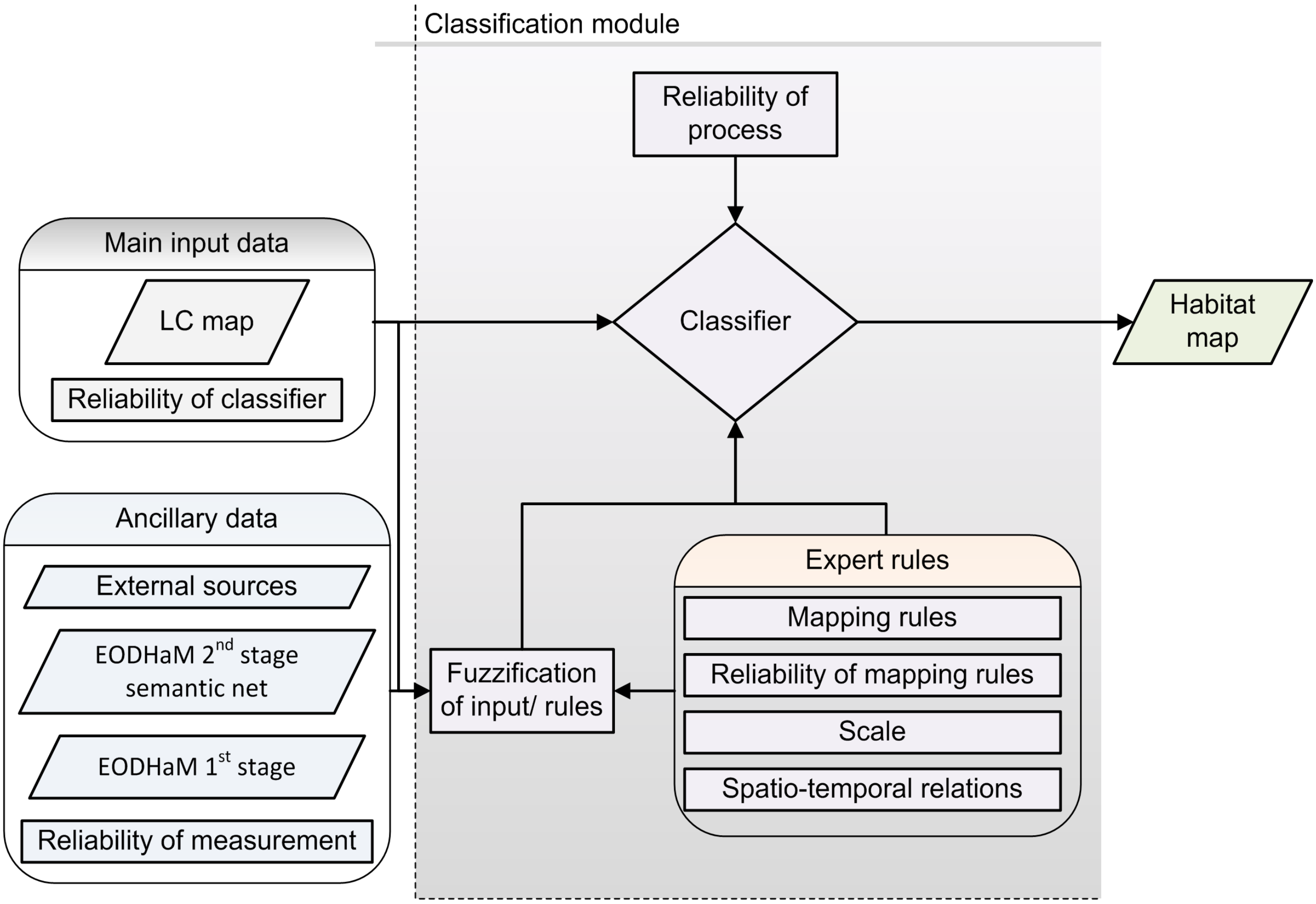


Figure 4-3: Complete architecture of the LC to habitat conversion system.

Figure 4-3: Complete architecture of the LC to habitat conversion system.

semantic net) with meaningful expressions and relationships (arcs of the net). The different relationships might include:

- spatial correlations, such as adjacency (e.g. two habitats might not be possible to be adjacent, or two specific adjacent LCCS classes might result in a certain combination of habitats);
- temporal correlations (e.g. the sequence of LCCS classes of an area throughout specific time instances of the year may exclude some potential habitats and favor the existence of others);
- scale issues. The difference in scale between the LC and the habitat mapping elements will be taken into consideration. As an example, the scale of the individual mapping elements in an LCCS map might be meaningless for the exact translation into GHC classes, element by element. Such a mapping might be realistic for the translation into life and non-life form categories, but not for entire GHCs. For instance, a GHC might be characterized as such because of the dominant life form, while there might exist a secondary one in a percentage of up to 30%; in such a case, the area might be represented or occupied by two or more different LCCS categories, resulting in one and the same GHC. Concepts from fields such as plausibility theory might need to be incorporated to counterbalance scale discrepancies.

3. Fuzzification of input data and expert rules. At a further stage of the development of the algorithm the uncertainty of both input data and expert rules will be taken into account. As far as input parameters, including the ancillary data, is concerned, uncertainty may arise when parameters that naturally represent continuous variables are grossly quantized, such as salinity of the water

or slope of the terrain. In addition, another source of uncertainty may result from the way some attributes are measured or calculated, which is subject to errors due to noise or intraclass variability. Various crisp thresholds adopted as expert rules may also perform better when fuzzified. The adoption of approaches such as fuzzy neural networks might be incorporated in the algorithm to reduce uncertainty and improve the classification accuracy from LCCS to GHC classes.

5. Implementation of the algorithm

5.1 Description of the implementation procedure

In order to test the validity of the theoretical aspects, the algorithm was implemented in a software format. The largest part of the land cover to habitat map conversion algorithm has been implemented in the MATLAB® computing language and software environment [14].

The software receives as input a geospatial data file in a vector format, in particular in Esri shapefile format (.shp extention). The shapefile and its accompanying files split the area of interest in polygons, assigning to each one certain attributes. Each polygon is characterized of an LCCS top class and, when available, the respective life form classifiers.

The mapping rules that will be used for the translation of the land cover classes into habitats, as well as the definitions, descriptions and abbreviations of the LCCS and GHC classes, are inserted in the algorithm as text files appropriately read by the software. The mapping rules, as already discussed in 4.1, map the LCCS classes into the potential GHC super-categories. In a further level of detail, the LCCS classes are mapped also to the possible GHC life and non-life form categories.

The software reads the shapefile, which contains information on the location of the vertices of each polygon and the LCCS class it represents. The respective LCCS life form classifiers are read also if existing. Depending on the level of detail of the available input data and the mapping rules, four possible mapping combinations exist, as mentioned in 4.2:

1. Only LCCS top classes are available as input which are mapped in the level of GHC super-categories.
2. Only LCCS top classes are available, but further rules map them to the possible GHC life and non-life forms, depending on the relevant super-category.
3. Both LCCS top classes and life form classifiers are available as input; each class is mapped to the potential GHC super-categories.
4. Both LCCS top classes and life form classifiers are available, while the translation rules include the mapping to the potential GHC life and non-life form categories.

The more information is added concerning the level of detail of either the input data or the mapping rules, the less the uncertainty that arises in the final classification results into habitats. Specific examples of the aforementioned combinations are discussed in 5.2 and 5.3.

For each polygon of the area of interest, the software maps the LCCS class, according to the mapping rules, to the potential habitat classes. The translation, in certain cases, constitutes one-to-one mapping while in others one-to-many. When life form classifiers and more detailed rules are provided to the system, the classification results become more specific. In case the LCCS life form classifiers provided are mixed with additional data, such as environmental or technical attributes, the algorithm extracts the useful, for the moment, life form classifiers; the additional data might prove useful depending on further expert rules provided to the system at an upcoming stage.

New attributes are generated and added to each polygon, describing the habitats that may exist. The attributes are added to the shapefile with the form of new fields in its attribute table. The new fields introduced are the following:

i) GHC_code. It contains the abbreviations of the possible GHC super-categories that each specific polygon might belong to. The three letter code names of the GHC super-categories are drawn in Appendix II. As an example, '*URB* or *CUL* or *TRS*' denotes that the specific polygon might indicate either urban or cultivated or tree and shrub-covered habitat.

- ii) GHC_id. It contains the same information with the aforementioned GHC_code, but instead of the abbreviations, it includes the indices of the GHC super-categories; the convention followed is '1' represents urban areas, '2' cultivated, '3' sparsely vegetated, '4' herbaceous wetland, '5' other herbaceous and '6' trees and shrubs. The respective value of the GHC_id field for the example mentioned for GHC_code above would be '*1-2-6*'. This field is introduced mainly to be used by further processing from the algorithm.
- iii) GHC_description. It acts as an explanatory field of the abbreviations and index names adopted above. The description field for the discussed example would be '*Urban/constructed areas or cultivated areas or trees and shrubs*'.
- iv) GHC_LNL_forms. Based on the elaborated version of the mapping rules that refer to the potential GHC life and non-life form categories, this field contains information on both GHC super-category level and the possible GHC life and non-life form categories. The previous example could be '*URB (TRE)* or *CUL (CRO)* or *TRS (LPH/MPH/TPH)*', which indicates that the specific habitat might be either urban woody area or cultivated woody crops or low, mid or tall phanerophytes.
- v) GHC_cat. Each combination of GHC super-categories of the area is represented with a category number. The field is particularly useful in quickly inspecting the amount of different GHC super-category combinations existing and in enhancing the rasterization of the vector file, when necessary.

The output of the software is a shapefile containing all information existing in the input shapefile together with the additional fields added, as described above. All geospatial and attribute information is preserved and the shapefile is enriched with the new habitat layers, which further facilitates the comparison of the two types of maps, namely land cover and habitat.

When inspecting the resulting habitat classes stored in the shapefile with any visualization software, each combination of super-categories is assigned a separate color. For visualization purposes, and for clearer inspection of the individual super-categories included in each combination, a further, optional step was added to the algorithm. The shapefile is converted into raster format depending on the values of the GHC_cat field using the GRASS software [15]. Each pixel of the created grid is assigned the underlying GHC_cat value. The file is saved in ENVI raster geospatial data format. With the use of MATLAB, pixels in each category are changed according to the included GHC super-categories in such a way that different textures are created revealing the composition of each category.

The implementation procedure and the different algorithmic steps are further explained below with the use of examples from two areas, Le Cesine in Italy and Cors Fochno in Wales. The respective results and the constructed maps are drawn.

5.2 Application of the algorithm and results for Le Cesine site

Le Cesine is one of the test sites of BIO_SOS project. It is located on the south east side of Puglia region, Italy. Four out of eight LCCS top classes are present in the site, namely

- a) cultivated and managed terrestrial areas (A11),
- b) natural and semi-natural terrestrial vegetation (A12),
- c) natural and semi-natural aquatic or regularly flooded vegetation (A24),
- d) artificial surfaces and associated areas (B15).

The location and extent of the aforementioned top LCCS classes are drawn in Figure 5-1.

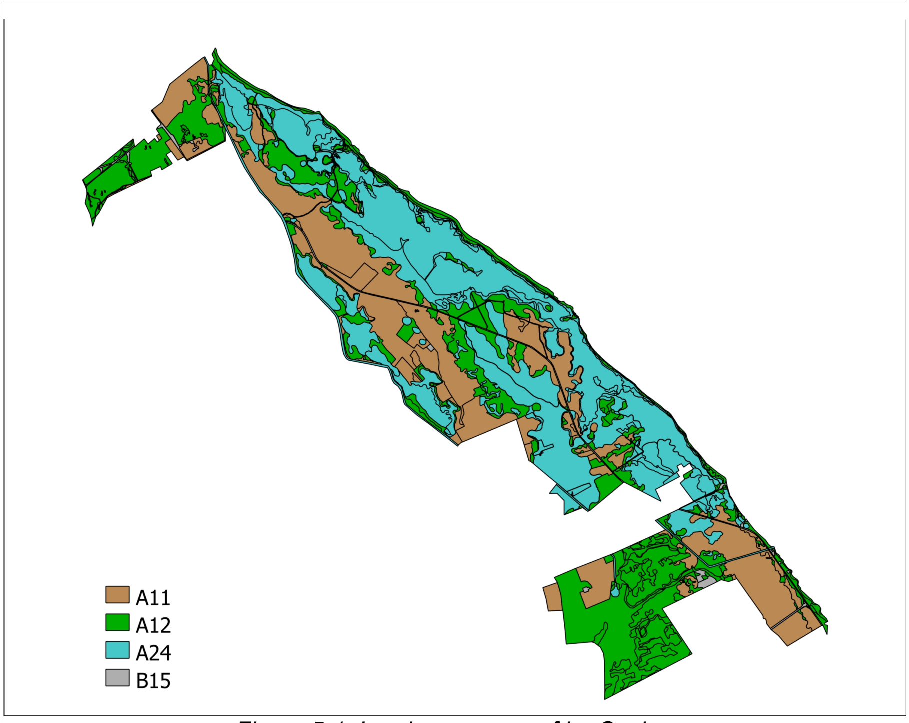


Figure 5-1: Land cover map of Le Cesine

Depending on the level of detail of the available input and the level of detail of the expert rules incorporated in the system, different GHC classes may arise, as explicitly mentioned in 5.1. At the beginning, having as input only the LCCS top classes, a translation was attempted into the potential, for each LCCS class, GHC super-categories. TFour different combinations of possible GHC super-categories were created as a result, namely

i) urban or cultivated or other herbaceous or tree and shrub areas,

ii) herbaceous wetlands or other herbaceous vegetation or trees and shrubs or urban areas,

iii) urban or sparsely vegetated areas or other herbaceous vegetation or trees and shrubs, and

iv) urban areas.

When additional mapping rules were considered for the mapping of the LCCS top classes into possible GHC life and non-life form categories, the number of possible combinations of super-categories remained the same, as expected. However the possible life and non-life form categories for each combination were further specified. Figure 5-2 shows the resulting habitat map including information on the potential GHC super-categories and, for each super-category, the potential life and non-life form category.

When, accompanying life form classifiers are introduced to the system, beside the top LCCS classes, the ambiguity between certain classes is, in general, reduced, in proportion with the level of detail of the adopted mapping rules. Figure 5-3 depicts the potential GHC super-categories when both the LCCS top classes and life form classifiers are used as input. More combinations appear in the latter case, denoting that certain areas are specified in a greater detail, when the additional life form classifiers were introduced. The most obvious example is the conversion of almost the entire area characterized as 'herbaceous wetland or other herbaceous vegetation or trees and shrubs or urban' area in Figure 5-2 into simply 'herbaceous wetland vegetation' in Figure 5-3. In addition, the part of the south west area of the site that was characterized as 'either urban or sparsely vegetated areas or other herbaceous vegetation or trees and shrubs' in Figure 5-2, it is split now into two different areas, namely 'urban or other herbaceous or trees and shrubs' and 'urban or trees and shrubs'; the life form classifiers helped eliminate the possibility of the habitat to be sparsely vegetated. Similarly, on the eastern side of the area, the ambiguous strip of the same combination in Figure 5-2, is now split in three combinations, i.e. the ones described above and the 'urban or other herbaceous'. Finally, the 'other herbaceous vegetation' super-category is excluded from most of the areas previously characterized as 'urban or cultivated or other herbaceous areas or trees and shrubs'; in the rest of these areas, the possibility of 'trees and shrubs' is eliminated.

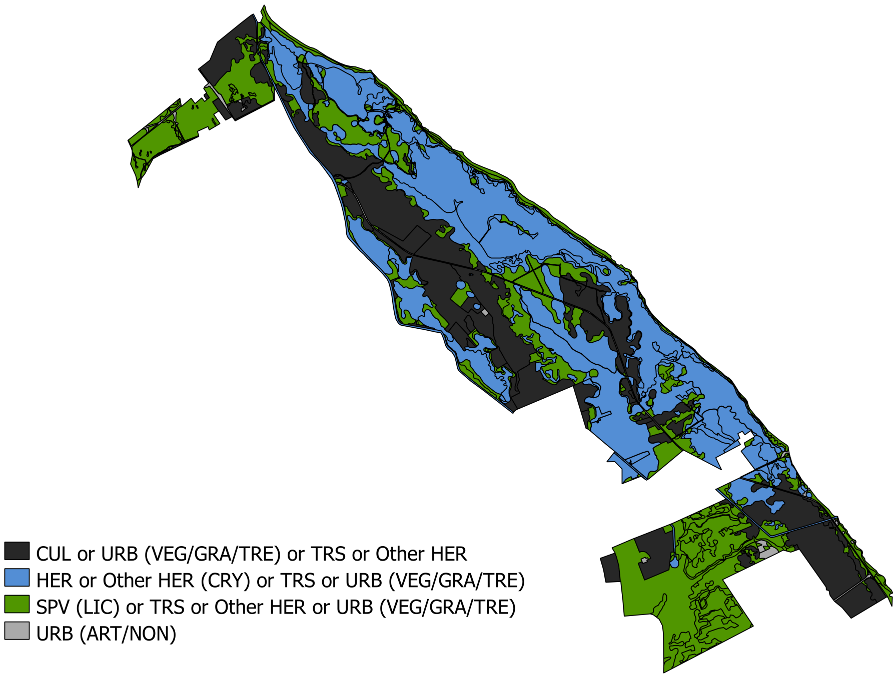


Figure 5-2: GHC super-categories and life forms from LCCS top classes for Le Cesine.

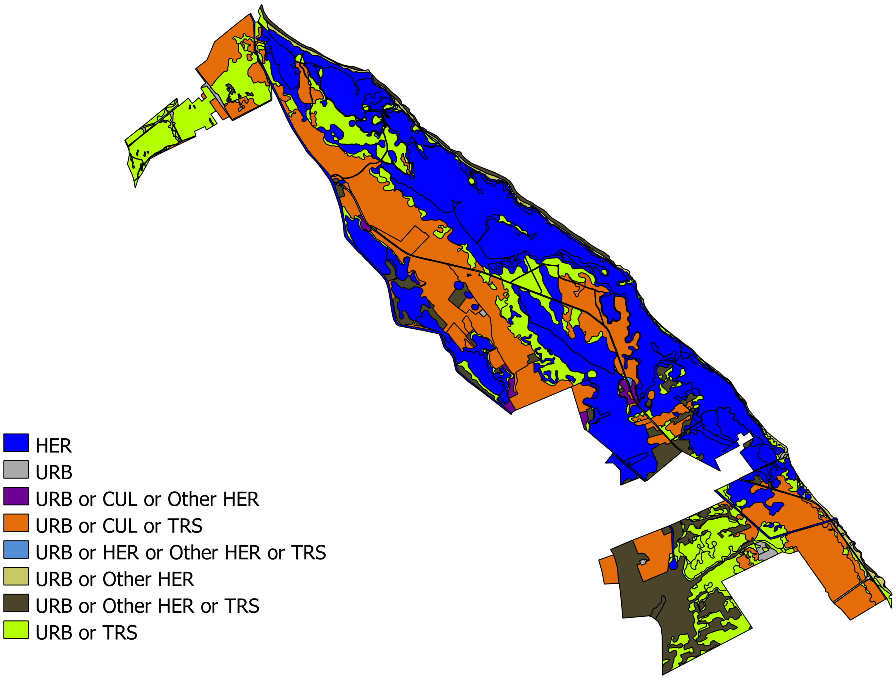


Figure 5-3: GHC super-categories from LCCS top classes and life form classifiers for Le Cesine.

In case the mapping rules provided by the experts extend to the specification of the potential GHC life and non-life form categories, the classification is further refined. The number of the potential combinations in the level of super-categories remains the same, but certain classes may split into two or more new combinations, depending on the possible life and non-life form categories. Such a case is presented in Figure 5-4, where two instantiations of the 'urban' class can be found, one including only artificial areas and the other artificial or non-vegetated. The 'urban or other herbaceous vegetation or trees and shrubs' combination is also split in two new combinations.

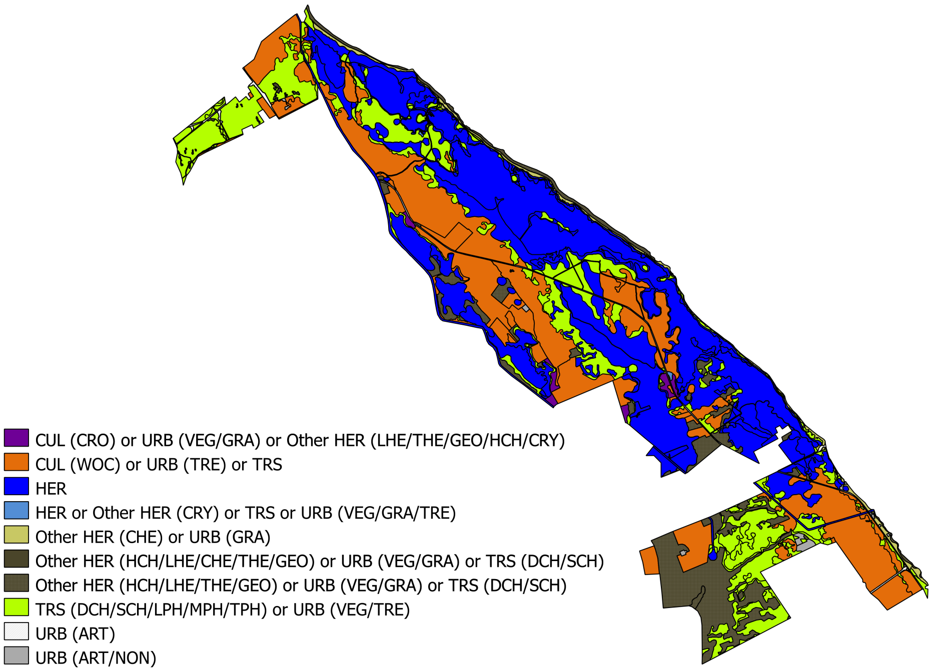


Figure 5-4: GHC super-categories and potential life and non-life form categories extracted from LCCS top classes and life form classifiers for Le Cesine.

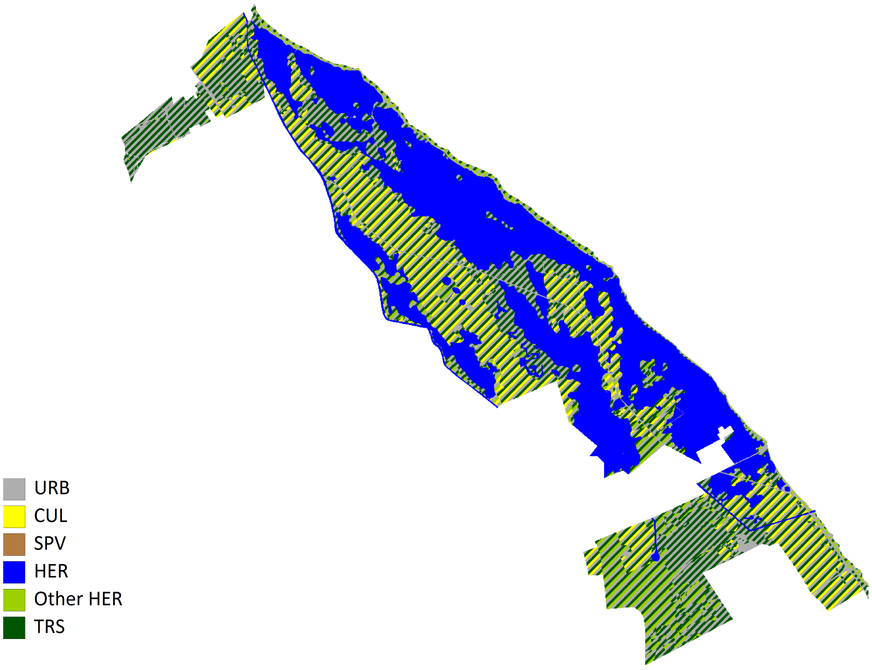


Figure 5-5: Visualization of GHC super-categories included in each combination for Le Cesine.

For visualization purposes, Figure 5-5, presents the different GHC combinations, as they appear in Figures 5-3and 5-4, as combinations of the colors of the included GHC super-categories. The mixed combinations are visually distinguished from the pure GHC super-categories, as the ones with multiple colors (strips).

In Figures 5-3, 5-4 and 5-5, the GHC super-category URB seems to have a potentially excessive extension compared to the actual arrangement of the site. As an example, along the coastal dunes there are no traces of URB. The inclusion of additional information in successive processing phases will be used to better discriminate GHCs and eliminate these ambiguities.

5.3 Application of the algorithm and results for the Wales site

Cors Fochno is also one of the test sites examined by BIO_SOS. It is an estuarine mire complex located with the Dyfi catchment in Wales, containing the largest uncut area of lowland raised bog in the UK. Six top LCCS land cover classes coexist in the site, namely

a) cultivated and managed terrestrial areas (A11),
b) natural and semi-natural terrestrial vegetation (A12),
c) natural and semi-natural aquatic or regularly flooded vegetation (A24),
d) artificial surfaces and associated areas (B15),
e) bare areas B(16),
f) natural waterbodies, snow and ice (B28).

The location and extent of the aforementioned top LCCS classes are drawn in Figure 5-6, with this based on the existing Land Cover map (LCM2000).

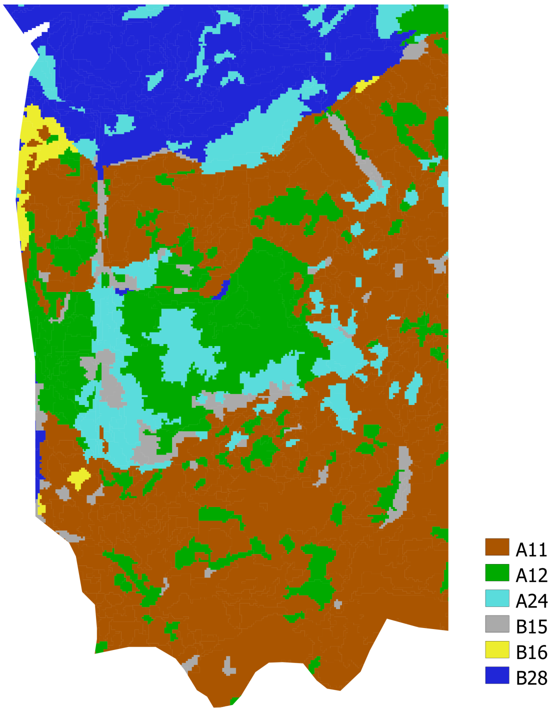


Figure 5-6: Land cover map of Cors Fochno.

As in the example examined in 5.2 for Le Cesine site, different habitat class combinations arise depending on the level of detail of the available input and the level of detail of the expert rules incorporated in the system. Following a similar procedure, at the beginning, only the LCCS top classes were considered as input and a translation was attemptedbetween each land cover class and the GHC super-categories. For Cors Fochno, the resultingd combinations of possible GHC super-categories are

i) sparsely vegetated areas,

ii) urban areas,

iii) urban or cultivated or or other herbaceous or tree and shrub areas,

iv) urban areas or areas with either herbaceous wetland or other herbaceous vegetation or trees and shrubs, and

v) urban or sparsely vegetated areas or other herbaceous vegetation or trees and shrubs.

The resulting combinations of GHC super-categories are drawn in Figure 5-7. LCCS classes B16 and B28 are both classified as sparsely vegetated areas.

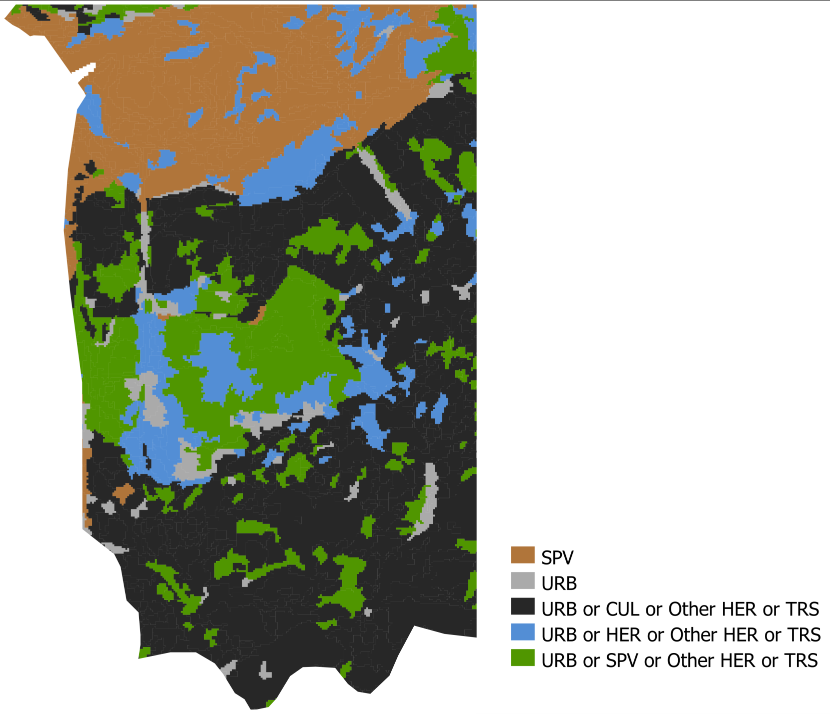


Figure 5-7: GHC super-categories from LCCS top classes for Cors Fochno.

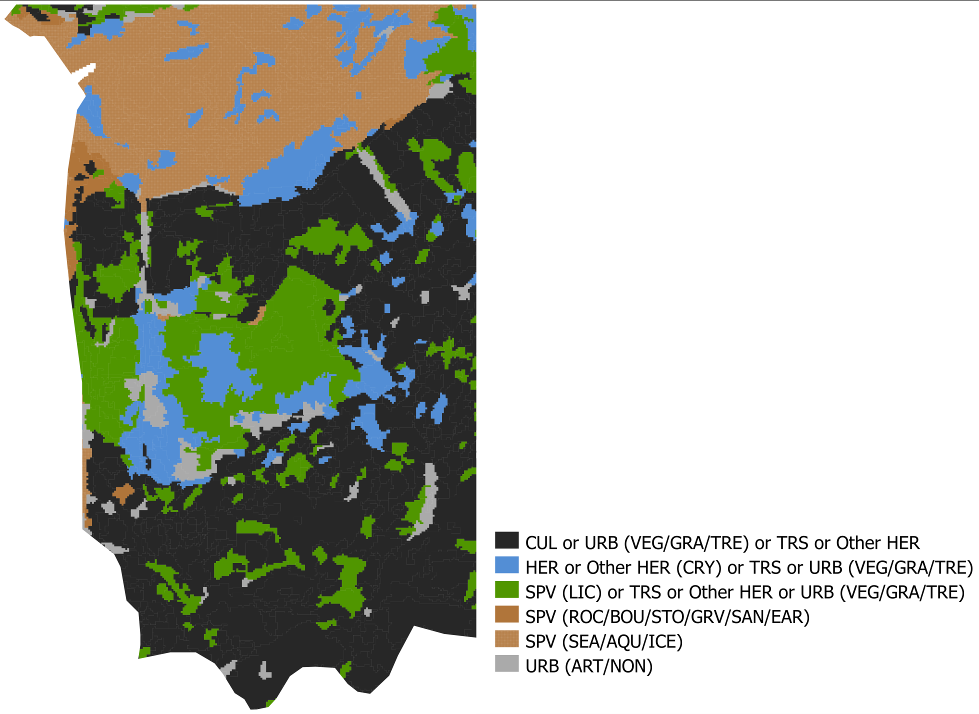


Figure 5-8: GHC super-categories and life forms from LCCS top classes for Cors Fochno.

When additional mapping rules were considered for the mapping of the LCCS top classes into possible GHC life and non-life form categories, the number of possible combinations of super-categories remained, as expected, the same. However the possible life and non-life form categories for each combination were further specified. That resulted in the separation of the 'sparsely vegetated areas' category into two new ones, the first including the terrestrial non-life form categories, resulted from the B16 LCCS class, while the second the aquatic, resulted from the B28. Figure 5-8 illustrates the resulting habitat map.

When, life form classifiers are added to the LCCS classes, the ambiguity between certain classes is reduced, in proportion with the level of detail of the adopted mapping rules. Figure 5-9 depicts the potential GHC super-categories, when both LCCS top classes and life form classifiers are used as input. More combinations appear, compared to the ones in Figure 5-7, denoting that certain areas are specified in a greater detail, when the additional life form classifiers were introduced. The most obvious example is the exclusion of the 'trees and shrubs' category from the largest part of the area previously characterized as 'urban or cultivated areas or other herbaceous or trees and shrubs'; from the rest of such areas, the possibility of presence of 'other herbaceous' vegetation is eliminated. In addition, most areas previously classified as 'herbaceous wetland or other herbaceous vegetation or trees and shrubs or urban area' in Figure 5-7 are now classified simply as 'herbaceous wetland vegetation' in Figure 5-9, while the rest as 'urban area or trees and shrubs'. The vague combination expressed as 'either urban or sparsely vegetated areas or other herbaceous vegetation or trees and shrubs' in Figure 5-7, is split now in different smaller combinations, each one excluding either one or two GHC super-categories.

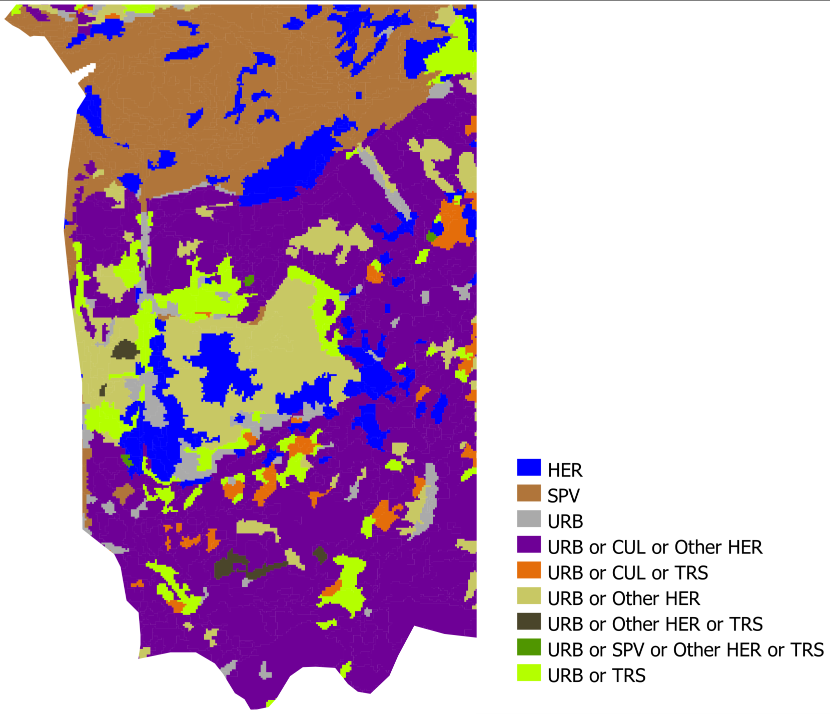


Figure 5-9: GHC super-categories extracted from LCCS top classes and life form classifiers for Cors Fochno.

In case the provided mapping rules from the experts extend to the specification of the potential GHC life and non-life form categories, the classification is further refined, as was observed also in the Le Cesine site. The number of the potential combinations in the level of super-categories remains the same, but certain classes split into two or more new combinations, depending on the possible life and non-life form categories. This case is illustrated in Figure 5-10. The sparsely vegetated areas are split into four different combinations of non-life form categories, such as terrestrial, aquatic or ice elements. The areas with wetland herbaceous vegetation are separated into two groups, one of which includes all possible herbaceous vegetation, while the other is restricted to certain of them (e.g., SHY-LEA). Similar differentiations are observed in the areas classified as 'urban or cultivated or other herbaceous' and 'urban or trees and shrubs'.

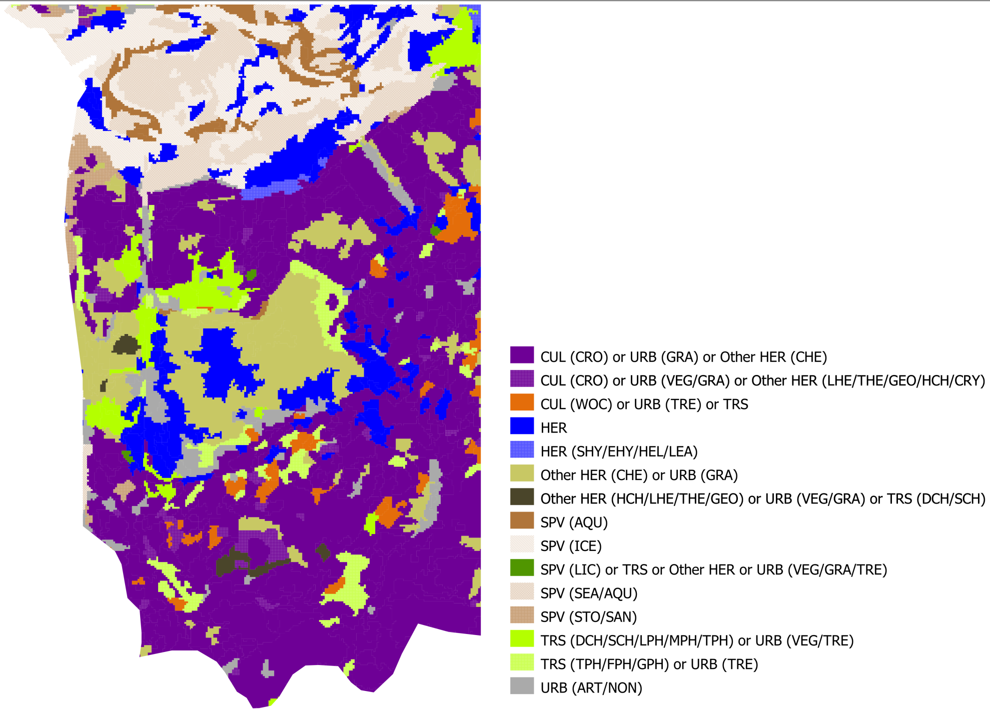


Figure 5-10: GHC super-categories and potential life and non-life form categories extracted from LCCS top classes and life form classifiers for Cors Fochno.

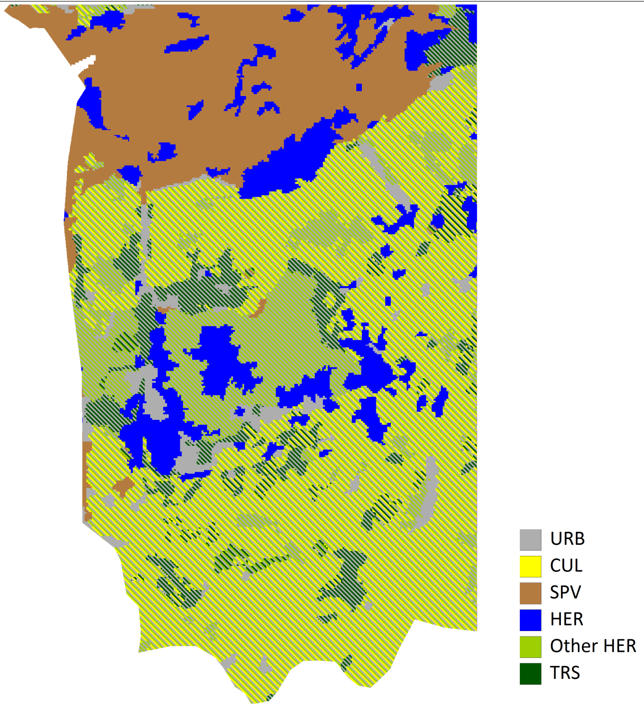


Figure 5-11: Visualization of GHC super-categories included in each combination for Cors Fochno.

For visualization purposes, Figure 5-11 presents the different GHC combinations, as they appear in Figures 5-9 and 5-10, as combinations of the included GHC super-categories. The mixed combinations are visually distinguished from the pure GHC super-categories, as the ones with multiple colors.

5.4 Discussion on the results - Future considerations

After the presentation of the structure of the software for the conversion of land cover maps into habitat maps, in Section 5.1, the functionality of the software was demonstrated through examples on two of the BIO_SOS test sites, Le Cesine in Italy and Cors Fochno in Wales. For each site, land cover maps according to the FAO-LCCS framework were fed to the software, together with expert rules for the translation of the land cover into habitat classes. Habitat maps were produced showing the feasibility of the algorithm.

Apart from the LCCS eight top classes, additional life form classifiers were used as input to the software, when available. The life form classifiers reduced the uncertainty of the classification and the number of possible GHC super-categories in certain combinations. In the case of both Le Cesine and Cors Fochno, the life form classifiers introduced managed to decrease the ambiguity of certain areas and better specify the potential habitats; in Le Cesine site, with the use of LCCS life form classifiers, the four initial vague GHC super-category combinations were increased to eight smaller ones, while for Cors Fochno they increased form five to nine.

On the other hand, the level of detail of the expert rules used plays a crucial role in the improvement of the classification. In a basic level, expert rules consider the mapping from the LCCS classes into the six GHC super-categories. At a second level, the rules were refined in such a way to handle and extract information on life and non-life form categories. Before the adoption of the refined mapping procedure, the B16 and B28 LCCS classes were both classified as 'sparsely vegetated areas' for Cors Fochno. After the refinement of the rules to include the discrimination at a level of life and non-life form categories, the areas belonging in the class B16 were classified as 'sparsely vegetated areas' with terrestrial non-life form categories, while the areas belonging in B28 were classified as 'sparsely vegetated areas' with aquatic non-life form categories.

In parallel, the software was further trained to evaluate available information on LCCS life form classifiers; both these improvements resulted in a more refined result, as shown in Figures 5-4 and 5-10, by considering both the LCCS top classes and their life form classifiers and expressing the habitat maps in terms of not only potential GHC super-categories, but of possible life and non-life form categories as well.

The software, as implemented up to the moment, constitutes the skeleton upon which the 3rd stage of the EODHaM module will build. The results obtained are encouraging, although they include a large degree of classification uncertainty. As the EODHaM modules are being developed, the software will be refined with additional and more sophisticated expert rules and with useful ancillary data and contextual and temporal information, in order to improve the classification accuracy and reduce any ambiguities.

As far as the pure implementation and interface level is concerned, the software has to be compatible with the rest of the EODHaM modules, not only at the level of input and output data specifications, but also on its means of implementation. Although initially implemented in MATLAB computing language, the software can trivially be translated into other computing languages or environments complying with the EODHam architecture.

6. Conclusion

Deliverable D6.10 focuses on the design and implementation of an algorithm for the conversion of Land Cover (LC) maps into Habitat maps for Biodiversity monitoring by the integrated use of remote sensing data with *in-situ* and ancillary data. In particular, an algorithm for the production of GHC maps from LCCS land cover maps has been designed and implemented. The resulting GHC maps will be used further towards the extraction both of Annex I habitats, by employing the Key introduced in the EBONE Deliverable 4.2, and other habitat types not included in Annex I, but of great ecological value for the Natura 2000 sites seleted in the Mediterranean areas.

The framework for the linkage between the FAO-LCCS and the GHC taxonomies was created and expert-rules were built upon this. Decision trees were employed for the representation of the expert knowledge, implementing a first stage crisp-classifier.

Due to the discrepancies between the two taxonomies, high level of uncertainty is imposed in the classification scheme; thus, the use of a module that takes into consideration the reliability of the process is vital. Such a model can be realised based on Bayesian regression model, which facilitates the use of prior information into the algorithm and account for the uncertainty in the input.

The algorithm is implemented and applied in two test sites, i.e. the Le Cesine in Italy and Cors Fochno in Wales. The output is produced at different levels of detail, given as input the top LCCS categories and, when available, respective life form classifiers. The ambiguity of the results is imposed by the differences between the two classification schemes, but can be resolved using sophisticated methods.

The results presented in D6.10 are indicative of the feasibility to convert the land cover maps into habitat maps, enhancing the innovative approach introduced by the BIO_SOS project. The results come as a proof of concept that gives added value to the BIO_SOS project towards the development of a cost effective biodiversity monitoring system.

The algorithm can be further refined by the use of ancillary data and semantic nets in order to increase the classification accuracy in the output. The refinement of the proposed algorithm will be a constant procedure to concur with parallel activities in BIO_SOS throughout the duration of the project. The design of the algorithm is such that allows the future insertion of additional modules and data input towards the elaboration of the classification result. As new information is derived from interacting modules of the EODHaM system, new features will be incorporated in the algorithm to elaborate the produced habitat maps and minimize the classifying uncertainty.

7. Appendices

Appendix I. Different stages and definitions of the dichotomous and the hierarchical phase in LCCS classification

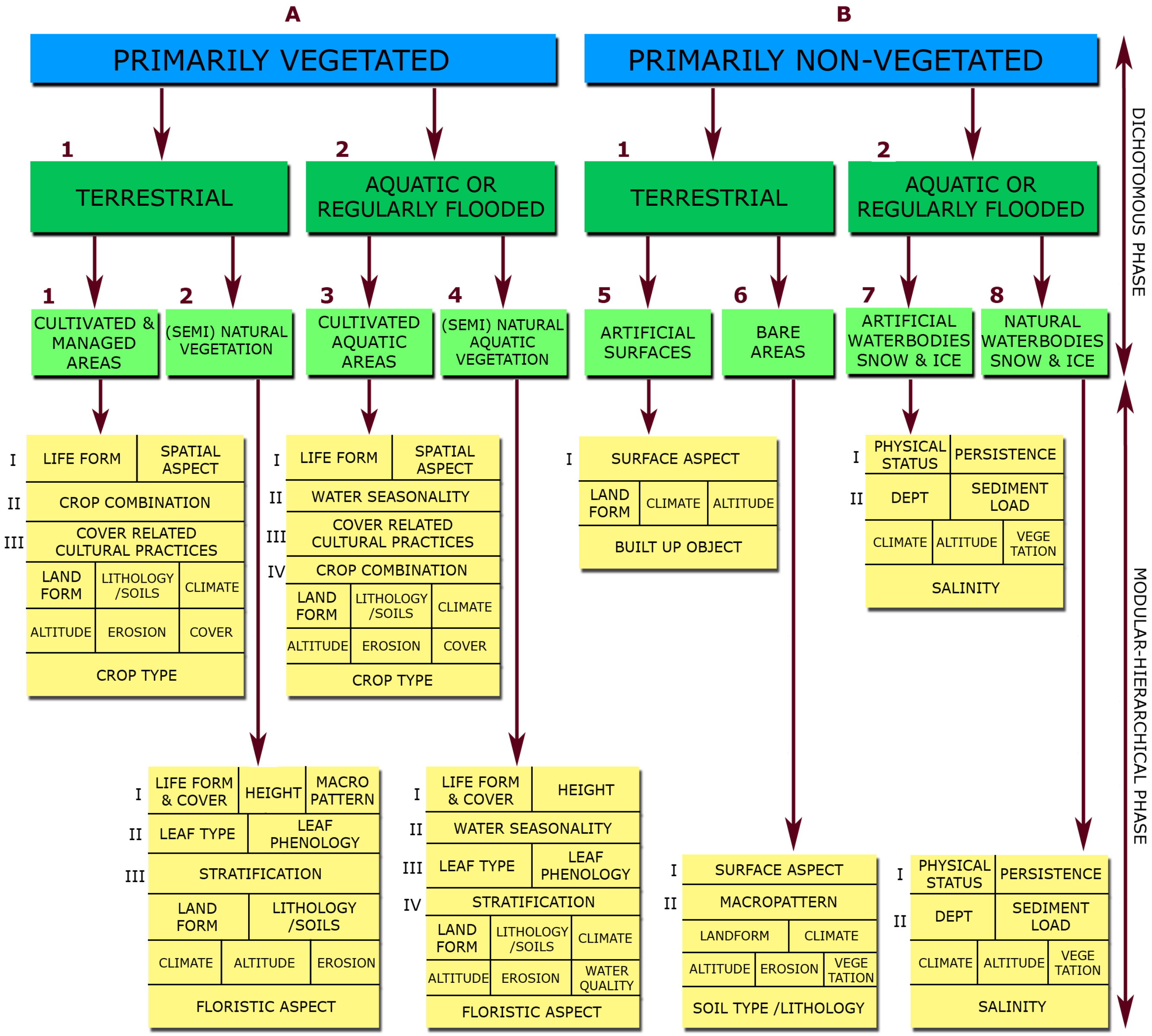


Figure I-1. Two phases and classifiers of LCCS (De Gregorio & Jansen, 1998)

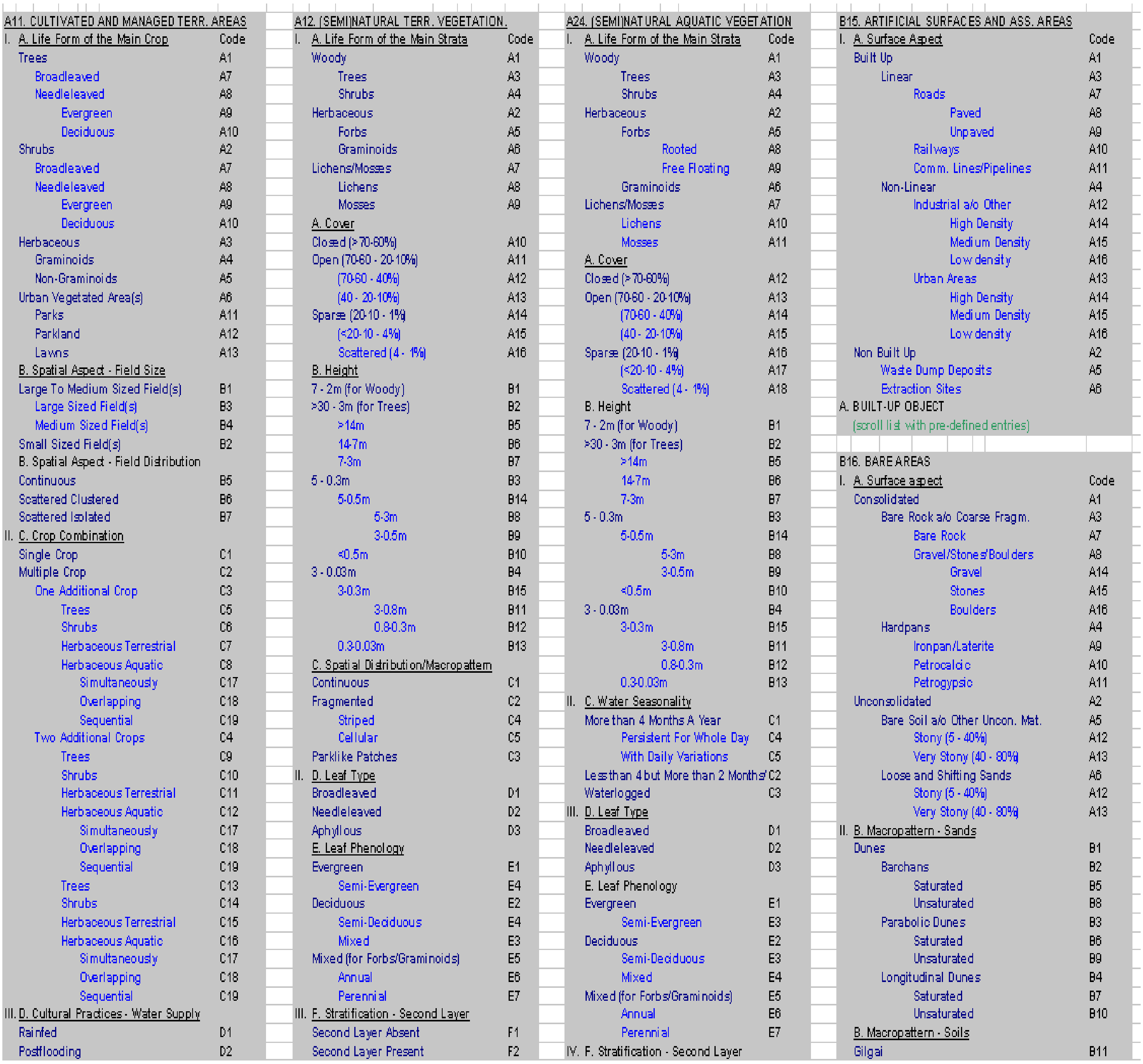
Table I-1 Definitions of the different stages of the dichotomous phase in the LCCS classification

Classifiers used	**Land Cover Class Name and Description**
DICHOTOMOUS PHASE: INITIAL-LEVEL DISTINCTION	
Presence/Absence of Vegetation: *Primarily vegetated*	A. Primarily Vegetated Areas: This class applies to areas that have a vegetative cover of at least 4% for at least two months of the year. This cover may consist of the life forms *Woody* (Trees, Shrubs), *Herbaceous* (Forbs, Graminoids) or a combination of them, or consist of Lichens/Mosses (only when other life forms are absent). A separate cover condition exists for Lichens/Mosses that can be only applied if this life form contributes at least 25% to the total vegetative cover (see Appendix A).
Presence/Absence of Vegetation: *Primarily non- vegetated*	B. Primarily Non-Vegetated Areas: This class includes areas that have a total vegetative cover of less than 4% for more than 10 months of the year, or in the absence of Woody or Herbaceous life forms less than 25% cover of Lichens/Mosses
DICHOTOMOUS PHASE: SECOND-LEVEL DISTINCTION	
Primarily vegetated Edaphic Condition: *Terrestrial*	A1. Terrestrial Primarily Vegetated Areas: The vegetation is influenced by the edaphic substratum.
Primarily non-vegetated Edaphic Condition: *Terrestrial*	B1. Terrestrial Primarily Non-Vegetated Areas: The cover is influenced by the edaphic substratum.
Primarily vegetated Edaphic Condition: *Aquatic or regularly flooded*	A2. Aquatic or Regularly Flooded Primarily Vegetated Areas: The environment is significantly influenced by the presence of water over extensive periods of time. The water is the dominant factor determining natural soil development and the type of plant communities living on its surface. Includes marshes, swamps, bogs and all areas where water is present for a substantial period regularly every year. This class includes floating vegetation.
Primarily non-vegetated Edaphic Condition: *Aquatic or regularly flooded*	B2. Aquatic or Regularly Flooded Primarily Non-Vegetated Areas: The environment is significantly influenced by the presence of water over an extensive period of time each year.

Classifiers used	Land Cover Class Name and Description
DICHOTOMOUS PHASE: TERTIARY-LEVEL DISTINCTION	
Primarily vegetated Terrestrial Artificiality of Cover: *Artificial/managed*	A11. Cultivated and Managed Terrestrial Areas: This class refers to areas where the natural vegetation has been removed or modified and replaced by other types of vegetative cover of anthropogenic origin. This vegetation is artificial and requires human activities to maintain it in the long term. In between the human activities, or before starting crop cultivation, the surface can be temporarily without vegetative cover. Its seasonal phenological appearance can be regularly modified by humans (e.g., tillage, harvest, and irrigation). All vegetation that is planted or cultivated with an intent to harvest is included in this class (e.g., wheat fields, orchards, rubber and teak plantations).
Primarily vegetated Terrestrial Artificiality of Cover: *(Semi-) natural*	A12. Natural and Semi-Natural Vegetation: Natural vegetated areas are defined as areas where the vegetative cover is in balance with the abiotic and biotic forces of its biotope. Semi-natural vegetation is defined as vegetation not planted by humans but influenced by human actions. These may result from grazing, possibly overgrazing the natural phytocenoses, or else from practices such as selective logging in a natural forest whereby the floristic composition has been changed. Previously cultivated areas which have been abandoned and where vegetation is regenerating are also included. The secondary vegetation developing during the fallow period of shifting cultivation is a further example. The human disturbance may be deliberate or inadvertent. Hence semi-natural vegetation includes vegetation due to human influences but which has recovered to such an extent that species composition and environmental and ecological processes are indistinguishable from, or in a process of achieving, its undisturbed state. The vegetative cover is not artificial, in contrast to classes A11 and A24, and it does not require human activities to be maintained in the long term.
Primarily vegetated Aquatic or Regularly Flooded Artificiality of Cover: *Artificial/managed*	A23. Cultivated Aquatic or Regularly Flooded Areas: This class includes areas where an aquatic crop is purposely planted, cultivated and harvested, and which is standing in water over extensive periods during its cultivation period (e.g., paddy rice, tidal rice and deepwater rice). In general, it is the emerging part of the plant that is fully or partly harvested. Other plants (e.g., for purification of water) are free-floating. They are not harvested but they are maintained. This class excludes irrigated cultivated areas.
Primarily vegetated Aquatic or Regularly Flooded Artificiality of Cover: *(Semi-) natural*	A24. Natural and Semi-Natural Aquatic or Regularly Flooded Vegetation: This class describes areas which are transitional between pure terrestrial and aquatic systems and where the water table is usually at or near the surface, or the land is covered by shallow water. The predominant vegetation, at least periodically, comprises hydrophytes. Marshes, swamps, bogs or flats where drastic fluctuations in water level or high concentration of salts may prevent the growth of hydrophytes are all part of this class. The vegetative cover is significantly influenced by water and dependent on flooding (e.g., mangroves, marshes, swamps and aquatic beds). Occasionally-flooded vegetation within a

	terrestrial environment is not included in this class. *Natural Vegetated Aquatic* habitats are defined as biotopes where the vegetative cover is in balance with the influence of biotic and abiotic forces. *Semi-Natural Aquatic* vegetation is defined as vegetation that is not planted by humans but which is influenced directly by human activities that are undertaken for other, unrelated purposes. Human activities (e.g., urbanization, mining and agriculture) may influence abiotic factors (e.g., water quality), affecting species composition. Furthermore, this class includes vegetation that developed due to human activities but which has recovered to such an extent that it is indistinguishable from its former state, or which has built up a new biotope which is in balance with the present environmental conditions. A distinction between Natural and Semi-Natural Aquatic Vegetation is not always possible because human activities distant to the habitat may create chain reactions which ultimately disturb the aquatic vegetative cover. Human activities may also take place deliberately to compensate for effects as noted above with the aim of keeping a "natural" state.
Primarily non-vegetated Terrestrial Artificiality of Cover: *Artificial/managed*	B15. Artificial Surfaces and Associated Areas: This class describes areas that have an artificial cover as a result of human activities such as construction (cities, towns, transportation), extraction (open mines and quarries) or waste disposal.
Primarily non-vegetated Terrestrial Artificiality of Cover: *(Semi-) natural*	B16. Bare Areas: This class describes areas that do not have an artificial cover as a result of human activities. These areas include areas with less than 4% vegetative cover. Included are bare rock areas, sands and deserts.
Primarily non-vegetated Aquatic or Regularly Flooded Artificiality of Cover: *Artificial/managed*	B27. Artificial Waterbodies, Snow and Ice: This class applies to areas that are covered by water due to the construction of artefacts such as reservoirs, canals, artificial lakes, etc. Without these the area would not be covered by water, snow or ice.
Primarily non-vegetated Aquatic or Regularly Flooded Artificiality of Cover: *(Semi-) natural*	B28. Natural Waterbodies, Snow and Ice: This class refers to areas that are naturally covered by water, such as lakes, rivers, snow or ice. In the case of rivers, the lack of vegetation cover is often due to high flow rates and/or steep banks. In the case of lakes, their geological origin affects the life conditions for aquatic vegetation. The following circumstances might cause water surfaces to be without vegetation cover: depth, rocky basins, rocky and/or steep shorelines, infertile washed-in material, hard and coarse substrates.

Table I-2 Major Land Cover type and their stuctural domain



A11. CULTIVATED AND MANAGED TERR. AREAS		A12. (SEMI)NATURAL TERR. VEGETATION.		A24. (SEMI)NATURAL AQUATIC VEGETATION		B15. ARTIFICIAL SURFACES AND ASS. AREAS	
I. A. Life Form of the Main Crop	Code	I. A. Life Form of the Main Strata	Code	I. A. Life Form of the Main Strata	Code	I. A. Surface Aspect	Code
Trees	A1	Woody	A1	Woody	A1	Built Up	A1
Broadleaved	A7	Trees	A3	Trees	A3	Linear	A3
Needleleaved	A8	Shrubs	A4	Shrubs	A4	Roads	A7
Evergreen	A9	Herbaceous	A2	Herbaceous	A2	Paved	A8
Deciduous	A10	Forbs	A5	Forbs	A5	Unpaved	A9
Shrubs	A2	Graminoids	A6	Rooted	A8	Railways	A10
Broadleaved	A7	Lichens/Mosses	A7	Free Floating	A9	Comm. Lines/Pipelines	A11
Needleleaved	A8	Lichens	A8	Graminoids	A6	Non-Linear	A4
Evergreen	A9	Mosses	A9	Lichens/Mosses	A7	Industrial a/o Other	A12
Deciduous	A10	A. Cover		Lichens	A10	High Density	A14
Herbaceous	A3	Closed (>70-60%)	A10	Mosses	A11	Medium Density	A15
Graminoids	A4	Open (70-60 - 20-10%)	A11	A. Cover		Low density	A16
Non-Graminoids	A5	(70-60 - 40%)	A12	Closed (>70-60%)	A12	Urban Areas	A13
Urban Vegetated Area(s)	A6	(40 - 20-10%)	A13	Open (70-60 - 20-10%)	A13	High Density	A14
Parks	A11	Sparse (20-10 - 1%)	A14	(70-60 - 40%)	A14	Medium Density	A15
Parkland	A12	(<20-10 - 4%)	A15	(40 - 20-10%)	A15	Low density	A16
Lawns	A13	Scattered (4 - 1%)	A16	Sparse (20-10 - 1%)	A16	Non Built Up	A2
B. Spatial Aspect - Field Size		B. Height		(<20-10 - 4%)	A17	Waste Dump Deposits	A5
Large To Medium Sized Field(s)	B1	7 - 2m (for Woody)	B1	Scattered (4 - 1%)	A18	Extraction Sites	A6
Large Sized Field(s)	B3	>30 - 3m (for Trees)	B2	B. Height		A. BUILT-UP OBJECT	
Medium Sized Field(s)	B4	>14m	B5	7 - 2m (for Woody)	B1	(scroll list with pre-defined entries)	
Small Sized Field(s)	B2	14-7m	B6	>30 - 3m (for Trees)	B2		
B. Spatial Aspect - Field Distribution		7-3m	B7	>14m	B5	B16. BARE AREAS	
Continuous	B5	5 - 0.3m	B3	14-7m	B6	I. A. Surface aspect	Code
Scattered Clustered	B6	5-0.5m	B14	7-3m	B7	Consolidated	A1
Scattered Isolated	B7	5-3m	B8	5 - 0.3m	B3	Bare Rock a/o Coarse Fragm.	A3
II. C. Crop Combination		3-0.5m	B9	5-0.5m	B14	Bare Rock	A7
Single Crop	C1	<0.5m	B10	5-3m	B8	Gravel/Stones/Boulders	A8
Multiple Crop	C2	3 - 0.03m	B4	3-0.5m	B9	Gravel	A14
One Additional Crop	C3	3-0.3m	B15	<0.5m	B10	Stones	A15
Trees	C5	3-0.8m	B11	3 - 0.03m	B4	Boulders	A16
Shrubs	C6	0.8-0.3m	B12	3-0.3m	B15	Hardpans	A4
Herbaceous Terrestrial	C7	0.3-0.03m	B13	3-0.8m	B11	Ironpan/Laterite	A9
Herbaceous Aquatic	C8	C. Spatial Distribution/Macropattern		0.8-0.3m	B12	Petrocalcic	A10
Simultaneously	C17	Continuous	C1	0.3-0.03m	B13	Petrogypsic	A11
Overlapping	C18	Fragmented	C2	II. C. Water Seasonality		Unconsolidated	A2
Sequential	C19	Striped	C4	More than 4 Months A Year	C1	Bare Soil a/o Other Uncon. Mat.	A5
Two Additional Crops	C4	Cellular	C5	Persistent For Whole Day	C4	Stony (5 - 40%)	A12
Trees	C9	Parklike Patches	C3	With Daily Variations	C5	Very Stony (40 - 80%)	A13
Shrubs	C10	II. D. Leaf Type		Less than 4 but More than 2 Months	C2	Loose and Shifting Sands	A6
Herbaceous Terrestrial	C11	Broadleaved	D1	Waterlogged	C3	Stony (5 - 40%)	A12
Herbaceous Aquatic	C12	Needleleaved	D2	III. D. Leaf Type		Very Stony (40 - 80%)	A13
Simultaneously	C17	Aphyllous	D3	Broadleaved	D1	II. B. Macropattern - Sands	
Overlapping	C18	E. Leaf Phenology		Needleleaved	D2	Dunes	B1
Sequential	C19	Evergreen	E1	Aphyllous	D3	Barchans	B2
Trees	C13	Semi-Evergreen	E4	E. Leaf Phenology		Saturated	B5
Shrubs	C14	Deciduous	E2	Evergreen	E1	Unsaturated	B8
Herbaceous Terrestrial	C15	Semi-Deciduous	E4	Semi-Evergreen	E3	Parabolic Dunes	B3
Herbaceous Aquatic	C16	Mixed	E3	Deciduous	E2	Saturated	B6
Simultaneously	C17	Mixed (for Forbs/Graminoids)	E5	Semi-Deciduous	E3	Unsaturated	B9
Overlapping	C18	Annual	E6	Mixed	E4	Longitudinal Dunes	B4
Sequential	C19	Perennial	E7	Mixed (for Forbs/Graminoids)	E5	Saturated	B7
III. D. Cultural Practices - Water Supply		III. F. Stratification - Second Layer		Annual	E6	Unsaturated	B10
Rainfed	D1	Second Layer Absent	F1	Perennial	E7	B. Macropattern - Soils	
Postflooding	D2	Second Layer Present	F2	IV. F. Stratification - Second Layer		Gilgai	B11

Appendix II. The GHC decision tree and the (Non/) Life Forms

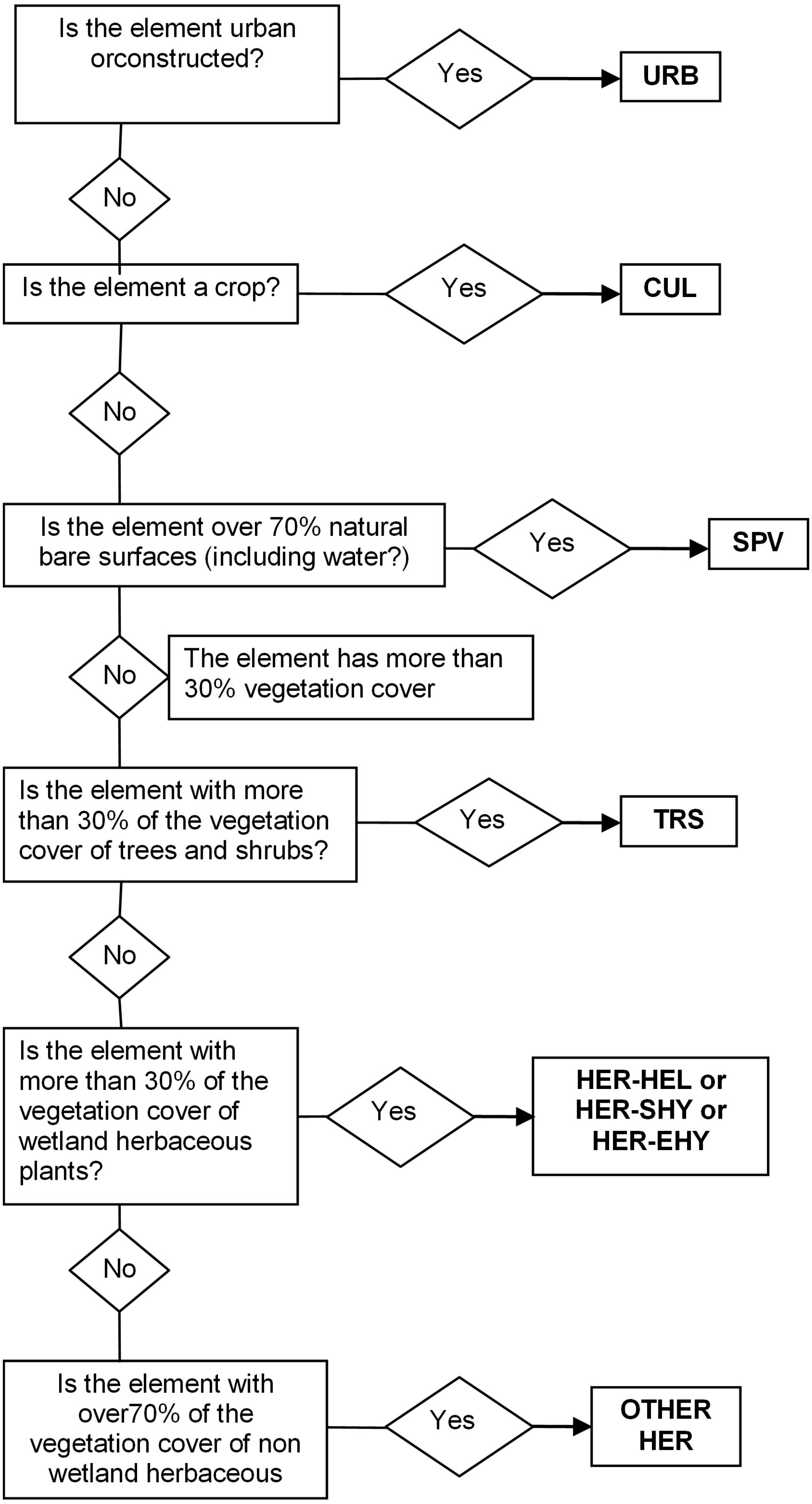


Figure II-1 Decision tree for super categories (in Ebone Handbook [1], fig. 4, p 28)

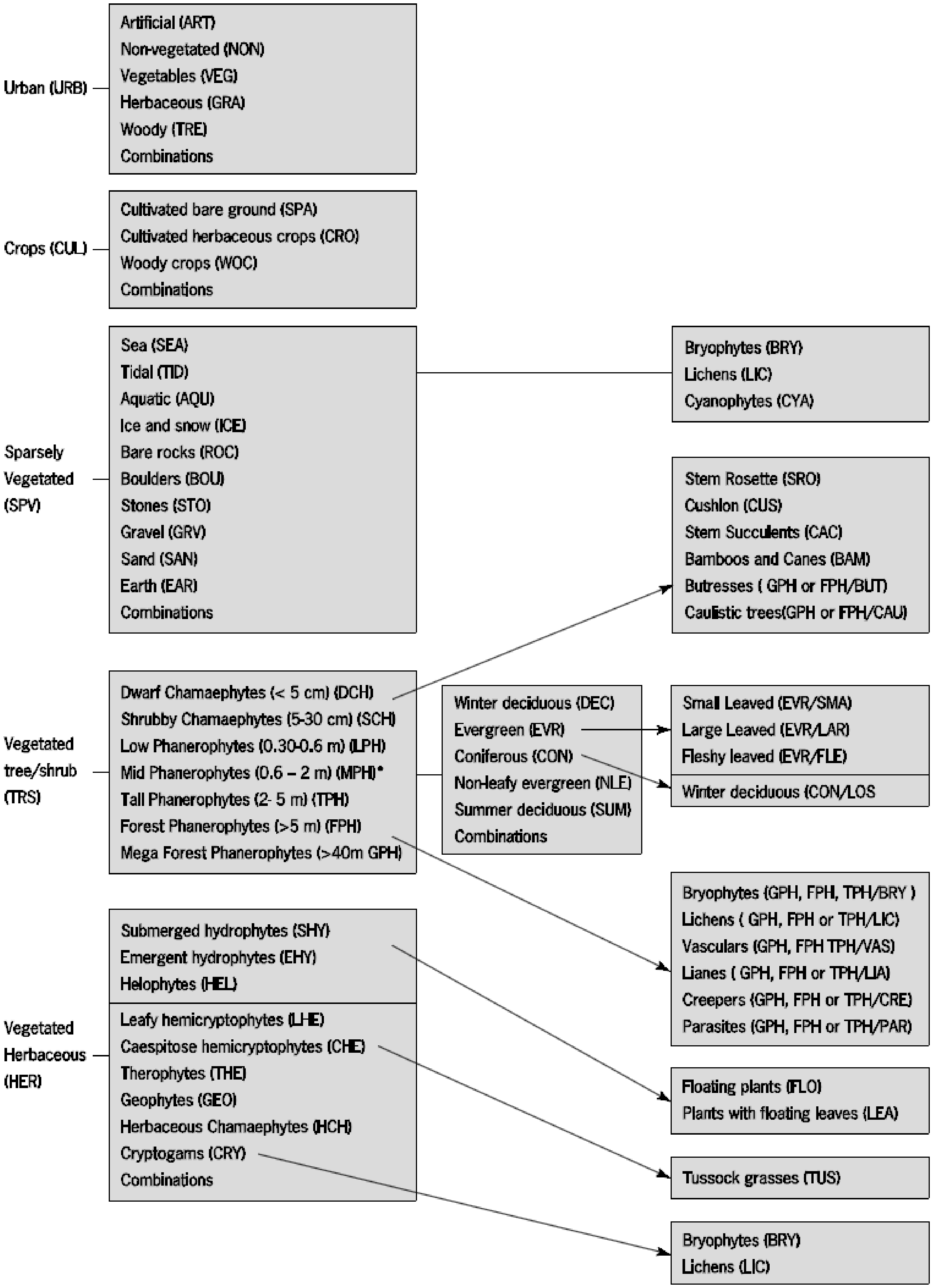


Figure II-2 Diagrammatic representation of the GHC Life Forms, Non-Life Forms and the basic qualifiers (in Ebone Handbook [1], fig. 5, p 31).

Appedix III. Acronym List

ABERY	**University of Aberystwyth – Inst. of Geography And Earth Sciences**
ANSI	**American National Standards Institute**
ATREE	**Ashoka Trust for Research in Ecology and the Environment – India**
BIOHab	**Acronym description: a framework for the coordination of BIOdiversity and Habitats**
BIO_SOS	**Biodiversity Multi-Source MOnitoring System: From Space To Species**
CBD	**Convention of Biological Diversity**
CERTH	**Informatics And Telematics Institute Of The Centre For Research And Technology – Greece**
CNR	**Consiglio Nazionale delle Ricerche**
EEA	**European Environmental Agency**
EEC	**European Economic Community**
EBONE	**European Biodiversity Observation Network**
EC	**European Community**
EO	**Earth Observation**
EODHaM	**EO Data for Habitat Monitoring**
EU	**European Union**
EUNIS	**European Nature Information System**
FAO	**Food and Agriculture Organization**
FAO-LCCS	**FAO - Land Cover Classification System**
FP7	**Seventh Framework Program**
GHC	**General Habitat Categories**
GIS	**Geographic Information System**
GLC	**Global Land Cover**
LC	**Land Cover**
LCC	**Land Cover Change**
LCCS	**Land Cover Classification System**
RS-IUS	**Remote Sensing Image Understanding System**
WP	**Work Package**

8. References

[1] Bunce R.G.H., Bogers M.M. B., Roche P., Walczak M., Geijzendorffer I.R. and Jongman R.H.G., 2011. Manual for Habitat and Vegetation Surveillance and Monitoring: Temperate, Mediterranean and Desert Biomes. First edition. Wageningen, Alterra report 2154.

[2] Di Gregorio, A. & Jansen, L.J.M. 1998. Land Cover Classification System (LCCS): Classification Concepts and User Manual. For software version 1.0. GCP/RAF/287/ITA Africover - East Africa Project in cooperation with AGLS and SDRN. Nairobi, Rome.

[3] Bunce R.G.H., Metzger M.J., Jongman R.H.G., Brandt J., de Blust G., Elena-Rossello R., Groom G.B., Halada L., Hofer G., Howard D.C., Kovàř P., Mücher C.A., Padoa Schioppa E., Paelinx D., Palo A., Perez Soba M., Ramos I.L., Roche P., Skånes H., Wrbka T., 2008. A standardized procedure for surveillance and monitoring European habitats and provision of spatial data. Landscape Ecol., 23: 11-25.

[4] Bunce R.G.H., Bogers M.M. B., Evansn D., and Jongman R.H.G., 2011. D4.2: Rule based system for Annex I habitats, Wageningen, Alterra EBONE-D4.2-2.6.

[5] Haines Young R.H., Barr C.J., Black H.I.J., et al., 2000. Accounting for nature: assessing habitats in the UK countryside. DETR, London.

[6] Dimopoulos P., Bergmeier E., Fisher P., 2005. Monitoring and conservation status assessment of habitat types in Greece: fundamentals and exemplary cases. Ann Bot, V: 7-20.

[7] Boteva Dimitrina, Griffiths G., Dimopoulos P., 2004. Evaluation and mapping of the conservation significance of habitats using GIS: an example from Crete, Greece. Journal for Nature Conservation, 12: 237-250.

[8] Jansen L. J. M., Di Gregorio A., 2002a. Parametric land cover and land-use classifications as tools for environmental change detection. Agriculture, Ecosystems and Environment, 91: 89-100.

[9] Jansen L. J. M., Di Gregorio A., 2002b. Land-use data collection using the ‘‘land cover classification system’’: results froma case study in Kenya. Land Use Policy 20: 131–148.

[10] Di Gregorio, A. & Jansen, L.J.M., 2005. Land Cover Classification System (LCCS): classification concepts and user manual. Food and Agriculture Organization of the United Nations, Rome.

[11] Raunkiaer C, 1934. The life forms of plants and statistical plant geography, being the collected papers of C Raunkiaer. Clarendon, Oxford

[12] Running, S.W., Loveland, T.R., Pierce, L.L., Nemani, R.R., Hunt, E.R., 1995. A remote sensing based vegetation classification logic for global land cover analysis. Remote Sensing of Environment, 51 (1), pp. 39-48.

[13] Chen, M.-H., Ibrahim, J. G., & Yianoutsos, C., 1999. Prior elicitation, variable selection, and Bayesian computation for logistics regression models. Journal of the Royal Statistical Society B, 61, 223– 242.

[14] MATLAB (version R2010a) [Software]. (2010). Natick, MA: The Mathworks Inc.

[15] GRASS Development Team, 2010. Geographic Resources Analysis Support System (GRASS) (version 6.4.0) [Software]. Open Source Geospatial Foundation. Retrieved from http://grass.osgeo.org.