Project Title:	**BIO_SOS Biodiversity Multisource Monitoring System: from Space TO Species**
Contract No:	FP7-SPA-2010-1-263435
Instrument:	
Thematic Priority:	SPA.2010.1.1-04
Start of project:	1 December 2010
Duration:	36 months

Deliverable No: 4.3

Protocols for new on-site campaigns

Due date of deliverable:	31-07-2011
Actual submission date:	19-08-2011
Version:	3rd version of D4.3
Main Authors:	PART1 Andrea Baraldi (P15) PART 2 Richard Lucas (P11), Palma Blonda , Valeria Tomaselli, Cristina Tarantino, Francesco Lovergine (P1) PART 3 Paola Mairota (P8), Richard Lucas (P11), Sander Mucher (P4), Joao Honrado (P9), Marion Bogers (P4), Valeria Tomaselli (P1), Panayotis Dimopoulos (P2), Dino Torri (P1), Rocco Labadessa (P8), Harini Nagendra (P5), Madhura Niphadkar (P5)

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

Deliverable title	Protocols for new on-site campaigns
Deliverable number	D4.3
Deliverable version	v3
Previous version(s)	v1 and v2
Contractual date of delivery	31 July 2011
Actual date of delivery	19 august 2011
Deliverable filename	BIO_SOS standard deliverable
Nature of deliverable	R
Dissemination level	PU = Public,
Number of pages	105
Workpackage	WP4
Partner responsible	UNIBA (P8)
Author(s)	PART 1: Andrea Baraldi (P15) PART 2: Richard Lucas (P11), Palma Blonda , Valeria Tomaselli, Cristina Tarantino, Francesco Lovergine (P1) PART 3: Paola Mairota (P8), Richard Lucas (P11), Sander Mucher (P4), Joao Honrado (P9), Marion Bogers (P4), Valeria Tomaselli (P1), Panayotis Dimopoulos (P2), Dino Torri (P1), Rocco Labadessa (P8), Harini Nagendra (P5), Madhura Niphadkar (P5)
Editor	Paola Mairota (P8)
EC Project Officer	Florence Beroud

Abstract	D4.3 anticipates the description of the protocol for accuracy assessment of the thematic maps generated from the EODHaM system (PART 1), provides examples for the instantiation of such a protocol for two training sites (PART 2), specializes the protocol to in-field collection of data relating to General Habitat Categories (GHCs) as well as flora, fauna and soil data (PART 3). These are required to support validation of the different stages outputs, of the EODHaM

	system.
Keywords	EODHaM System validation, LCCS, GHC, map validation habitat, biodiversity and pressure monitoring, field recording and ecosystem assessment

Signatures

Editor	Responsibility- Company	Date	Signature
Paola Mairota	UNIBA		
Verified by			
Panayotis Dimopulos	WP4 Leader UOI		
Approved by			
Palma Blonda	Coordinator, CNR		
Fifamè Koudogbo	Quality Manager, ALTAMIRA		

TABLE OF CONTENTS

1. Executive summary

D4.3 is a Deliverable of WP4, Task 4.4. The activity of Task 4.4, focusing on *field campaigns for system validation,* includes*:*

1) Planning of new in-field campaigns for system validation with definition of: vegetation, soil and fauna data/information to be considered
2) Protocols definition for updated on-site *in-situ* data collection
3) On site data collection for soil and vegetation monitoring

D4.3 describes the protocols for updated on-site *in-situ* (in field) campaigns to be carried out in BIO_SOS training sites. In addition, D4.3 anticipates the description of the protocol for accuracy assessment of a thematic map generated from earth observation (EO) data. Such a protocol should be described in D5.5 which is due too late (i.e.,. month 12). Since the preliminary SIAM™ maps have been already provided and the second stage land cover/use maps (LC/LU) have to be produced in a shorter time frame, the protocol has been included in D4.3following discussions with P1, P15 and P8

In the BIO_SOS project, on-site *in-situ* (in field) campaigns are required to support validation of the proposed EODHaM system. The outputs obtained by the different stages of the system, as described in D3.1, include:

a) Land cover/use maps
b) General Habitat Category (GHCs) and Annex 1 Habitat maps
c) Change maps
d) Biodiversity indicators (belonging to the focal areas: Status and trends of the components of biological diversity and Ecosystem integrity and ecosystem goods and services) and vegetation dynamics (D2.1).
e) Biodiversity indicator trends for Biodiversity pressure scenario analysis.

The outputs will be provided within WP5 and WP6 by integrating high resolution (HR) and very high resolution (VHR) EO-derived products with “on site” data through ecological modelling. “On site” data include ancillary data/information which, by definition, is any data/information which cannot be inferred from appearance (visual, pictorial) properties of 3-D objects in a (3-D) scene depicted in a (2-D) EO image domain. The “on site” campaigns will also include “in field” campaigns, and consequently will be focused on the collection of data (flora, vegetation, fauna, soil) for both biodiversity and pressures/threats indicators extraction, as well as for GHCs identification.

D4.3 is organized into three parts.

PART 1 introduces concepts for probability sampling design, response design for accuracy assessment and analysis of the thematic maps derived from EO data in WP5 (Task 5.2, Task 5.3 and Task 5.4) and WP6 (Task 6.6). The first part is based on the contribution of P15.

PART 2 includes the instantiation of the protocol described in PART 1 for two training sites, in Italy and Wales and is based on the contributions of P11 and P1

PART 3 specializes the protocol to in-field collection of data relating to GHCs as well as flora, fauna and soil data, which are required specifically for WP4 (Task 4.4) and WP6 (Tasks 6.2, 6.3, 6.4, 6.7). The third part is based on contribution from and discussions with many partners: P8, P11, P4, P9, P1, P2, P5.

2. PART 1

2.1 Introduction

Two complementary types of variables exist in data/information processing systems (Liang,2004).

- Objective quantitative varying **sub-symbolic (non-semantic) sensory *continuous variables*.**
- Subjective qualitative (fuzzy, vague, but stable) **symbolic (semantic) *categorical variables*** belonging to a prior knowledge-based community-agreed ***taxonomy*** or ***ontology*** equivalent to **a finite and discrete set (e.g., hierarchy) of *percepts*, *concepts* or classes of 3-D objects in the (3-D) real world, also called 3-D object-models** (Matsuyama et al. 1990).

By definition, ***classification*** **is the mapping of continuous sensory variables into a finite and discrete set of thematic (symbolic, semantic, categorical) variables**.

The aforementioned duality in the taxonomy of variables is one-to-one related with the well-known information gap existing between (sub-symbolic, sensory, instantaneous, numerical, quantitative, absolute, non-semantic) ***sensations*** and (symbolic, linguistic, qualitative, vague, discrete and semantic, persistent, stable) ***percepts*** which has been thoroughly investigated in both philosophy and psychophysical studies of perception. In practice, "we are always seeing objects we have never seen before at the sensation level, while we perceive familiar objects everywhere at the perception level" (Matsuyama et al. 1990).

By definition, ***geospatial variables*****, either continuous or categorical, are geographically explicit quantitative sub-symbolic sensory data (e.g., surface reflectance values) or qualitative symbolic concepts (e.g., land cover classes) provided with a geographic attribute (e.g., a lat-long coordinate position on the Earth surface).** It is noteworthy that investigation in the geographic domain of geospatial variables through remote sensing (RS) images of the Earth differentiates RS image analysis from, say, computer vision and biomedical imaging.

Geospatial variables, either continuous or categorical, may be sampled in the field, inferred from RS images of the Earth or collected from pre-existing information sources (e.g., pre-existing thematic maps). They may be described in terms of:

- **single-number (scalar) summary measures** (Stehman, 1997) or **summary statistics** (e.g., areal extent), eligible for use when the geospatial distribution and pattern (related to the so-called landscape connectivity/fragmentation) of the target geospatial variable is considered negligible.
- **cartographic maps (**e.g., thematic maps, where the geospatial distribution and pattern of the target geospatial variable are modeled explicitly).

The present PART 1 focuses its attention on **cartographic maps of categorical variables**, namely, ***thematic maps*** **whose legend belongs to a finite and discrete set of concepts (categorical variables)**.

On the one hand, to generate thematic maps at regional, national, continental or global scale, an operational satellite-based measurement system is required to accomplish a synergistic use of RS image understanding techniques, such as inductive (bottom-up, data-driven) or deductive (top-down, prior knowledge-based) classifiers, with reference to thematic information (semantic labeled data) required for training and/or testing the RS image mapper (Gutman et al. 2004). Typical advantages of generating thematic maps from RS imagery are the following.

(i) Assess the geospatial distribution of thematic areas in place of their summary statistics.
(ii) Assess geospatial trends in thematic area change through time.
(iii) Reduce significantly the volume and cost of reference sampling measurements needed to be undertaken in the field (Maniates et al. 2010).

On the other hand, thematic map accuracy assessment and validation quantifies the quality of geospatial (mapped) categorical variables so that map users may evaluate the utility of a thematic map for their intended applications. The symbolic labels from the thematic map are compared to the reference classifications, and the extent to which these two classifications agree is defined as thematic map accuracy.

As a consequence, when a thematic map is, first, generated from RS imagery and, second, assessed and validated in terms of labeling and spatial accuracy (Baraldi et al. 2005), some confusion may raise among RS researchers and practitioners about the meaning and purpose of reference thematic information.

In this specific context, the use of reference thematic information is three-fold, as summarized below.

1. Training the RS image classification system capable of generating, as output, a thematic map. A reference (semantic labeled) training data set is required if (and only if) the RS image classifier is capable of inductive learning-from-data. For example, no training data are required by a deductive prior knowledge-based classifier (e.g., a decision-tree non-adaptive to data such as the Satellite Image Automatic Mapper™, SIAM™ (Baraldi et al. 2006; Baraldi, 2009)). The training data set, if any, is generally collected based on a non-probability sampling approach where reference locations are selected by purposeful, convenient, or haphazard procedures (see section 2.2.1).
2. Testing the RS image classification system whose output is a thematic map. As with the training data set, if any, the test data set is generally collected based on a non-probability sampling approach (refer to Point 1. above). If the training data set is not required because the classifier employs a deductive inference approach, then the whole reference data set is employed for testing the mapping system (Baraldi, 2005).
3. Validating thematic maps, whether or not generated from RS imagery, is based on a statistically rigorous accuracy assessment to permit generalization from the discrete and finite reference sampled data set to the accuracy of the full population. A probability sampling design is a key element of a statistically rigorous thematic map accuracy assessment. Thematic map accuracy assessment is usually conducted by selecting a statistically meaningful discrete and finite set of sampling units, and comparing the classifications at these reference sample locations to the classifications provided by the thematic map at hand (Stehman, 1998). In RS common practice, one-third to one-half of the budget of a RS image classification project can be allocated for a statistically rigorous high-level product validation.

Possible non-alternative, but complementary sources of reference thematic information (semantic labeled data) for training and/or testing a RS image classifier and/or thematic map validation purposes can be the following.

- Field campaigns in geographic areas to be considered accessible. The costs of field measurements are typically large.
- High-quality RS images (e.g., fine spatial resolution aerial photography or videography) of the reference sampling units to be photo-interpreted by domain experts, scientists and analysts having a clear understanding of the thematic classification taxonomy (ontology) being used in the mapping project. This approach encompasses surface areas considered inaccessible and typically reduces significantly the volume and cost of reference sampling measurements needed to be undertaken in the field (Maniates et al. 2010).
- Prior thematic knowledge stemming from pre-existing thematic information sources, (e.g., pre-existing thematic maps).

The primary objective of this work is to elucidate a sorted set of basic steps (called a protocol equivalent to a practice guideline document) of a statistically rigorous accuracy assessment of a thematic map, whether or not generated from RS imagery.

The proposed protocol is fully compatible with and capable of integrating recommendations provided by the following protocols found in existing literature.

- Reducing emissions from Deforestation and Forest Degradation (REDD) mechanism endorsed by the of the United Nations Framework Convention on Climate Change (UNFCCC), refer to the summary paper by Maniates and Mollicone (Maniates et al. 2010) and to the source book REDD+ version 2010 COP16-1.
- Stehman's protocol for "Design and Analysis for Thematic Map Accuracy Assessment", refer to (Stehman, 1998).

The general basic steps of any thematic map accuracy assessment protocol according to the aforementioned literature (Maniates, 2010; Stehman, 1998) are summarized in Section 2.2.

The second goal of this contribution is to discuss the accuracy assessment of thematic maps, generated from EO data in combination with ancillary information, if available. In general, thematic map legends consist of a geospatial thematic (e.g., land cover (LC)/land use (LU)) hierarchical taxonomy (ontology) featuring different levels of semantic granularities (e.g., fine, intermediate, coarse, etc.) and increasing levels of abstraction (e.g., from LC/LU to GHCs (Bunce et al. 2011)). The issue of thematic map accuracy assessment will be dealt with in more detail in D5.5, as output of Task 5.2 and Task 5.3

In this context, by definition, *ancillary data/information* is any data/information that cannot be inferred from appearance (visual/pictorial) properties of 3-D objects in a (3-D) scene depicted in a (2-D) RS image domain.

The rest of this work is organized as follows. In Section 2.2 the protocol basic steps are described. In section 2.3 a probability sampling design protocol for accuracy assessment of a thematic map from RS imagery is illustrated in more details. Two Appendices are related to this PART1.

2.2 Strategies to operationalise and successfully implement a thematic map accuracy assessment

The thematic classifications from the thematic map are compared to the reference classifications, and the extent to which these two classifications agree is defined as thematic map accuracy.

A thematic map accuracy assessment protocol, irrespective of the origin of the thematic map at hand (e.g., generated from RS imagery), can be broken down into four stages (Stehman, 1998; Stehman, 1997) .

1. A thematic map accuracy assessment begins with the **identification and clear understanding of the thematic map legend (taxonomy, ontology)**, including spatial (e.g., inclusion) and non-spatial (e.g., subset-of) relationships between geospatial thematic classes, and the selection of the target geospatial population(s) or class(es), equivalent to information layer(s) or stratum (strata), with these located in the area or region represented by the thematic map.
2. The ***sampling design protocol*** to select geospatial reference sampling units across the thematic map and the depicted surface of the Earth. There are two basic ways to approach sampling.
    i. Non-probability sampling methods provide estimates of population parameters, but the uncertainty (error tolerance) of those estimates cannot be assessed (Maniates, 2010).
    ii. Probability sampling methods provide estimates for a population based on rigorous laws of probability that allow evaluation of the uncertainty of the estimates (Maniates, 2010).
3. The ***response design protocol*** for determining the thematic label(s) of the reference geospatial sampling units, (e.g., through photo-interpretation of high-quality RS imagery, field campaigns or a combination of these two information sources). It includes:
    i. Procedures to collect information pertaining to the reference thematic determination, referred to as the ***evaluation protocol***.
    ii. Rules for assigning one or more reference classifications to each sampling unit; the ***labeling protocol***.

4. The ***estimation and analysis protocol*** for a statistically rigorous quality assessment of a thematic map in comparison with reference geospatial sampling units based on a selected set of mutually uncorrelated QIs provided with uncertainty measures.

2.2.1 Non-probability sampling of reference sampling units

In non-probability sampling methods, estimates can be provided for population parameters, but the accuracy (error tolerance) of those estimates cannot be assessed (Maniates et al. 2010).

Unfortunately, examples of nonprobability sampling are common in thematic map accuracy assessment applications. Selecting reference locations by purposeful, convenient, or haphazard procedures does not provide the structure to determine the inclusion probabilities for each sampling unit. Such designs, therefore, are not probability samples. Purposefully selecting training data for a supervised classification is a good example of a nonprobability sample. Such samples are acceptable for developing the thematic classification map (e.g., by means of a supervised data learning classifier), but often have limited use for accuracy assessment because the necessary probability foundation to permit generalization from the sample data to accuracy of the full population is lacking. Selecting the reference sample from conveniently accessible sites or available aerial photography suffers from the same problem. It is virtually impossible to assert with any confidence that these convenient sources of data have the same attributes as the entire region. We may assume this to be the case, but this assumption cannot be scientifically defended (Stehman, 1998).

To conclude, nonprobability sampling resulting from purposeful selection of sampling units within subregions of the target area of interest cannot be classified as a probability sample of the full region. Although it is possible to obtain useful information (e.g., training samples for inductive classifiers of RS imagery to generate the thematic classification map) from nonprobability samples, the statistical limitations of such data should be recognized in the map accuracy assessment design selection (Stehman, 1998).

2.2.2 Probability sampling of reference sampling units

The probability sampling design is the protocol by which sampling units are selected into the sample (target population) according to the laws of statistics. This approach provides estimates for a population, but it is based on laws of probability that allow evaluation of the uncertainty of the estimates (Maniates et al. 2010).

Thematic map accuracy assessments typically have multiple users and objectives leading to interest in a variety of accuracy parameters and subregions of the mapped area. The need to satisfy multiple objectives motivates selecting a simple, general purpose sampling design. Simplicity is a key criterion because simple sampling designs are easier to implement properly in the field and to analyze, and they are more likely to provide adequate information for a broad variety of objectives. Simple designs are also easier to understand, so the accuracy assessment data are more likely to be used correctly, even by future users who may not be familiar with the planning and details of the design. A disadvantage of a broadly adequate, simple design is that it will be less effective for any single objective relative to a design tailored specifically for that objective. For example, if a rare class is critical to the success of a mapping project, a specialized, separate design can be added to augment the sample size in the rare class which will likely not be well represented in a simple, general sampling design (Stehman, 1998).

Practical limitations (e.g., geographic accessibility, budget constraints) determine what realistically can be expected of statistical methods, and this should focus accuracy assessment planning on the priority objectives (e.g., budget) of the mapping project. If all objectives cannot be addressed well (e.g., sample set size as a function of the target accuracy and confidence interval versus budget), the sampling strategy must be constructed so that critical issues are addressed adequately. Secondary objectives may, by necessity, not receive adequate sampling resources. A practical accuracy assessment sampling strategy often represents a compromise, with the overall design goal being adequate for all critical objectives but not optimal for any single objective (Stehman, 1998).

Although practical considerations (e.g., geographic accessibility, budget constraints) play a prominent role in thematic map accuracy assessment planning, these considerations should never lead to use of inefficient or incorrect sampling designs and analyses. On the contrary, the use of efficient and correct sampling strategies should be considered a necessary condition to satisfy practical considerations, cost reduction *in primis*. For example, if a thematic map is generated from a RS image, but a poor probability sampling design is implemented for thematic map validation, then additional reference samples may be required to reduce the error tolerance of statistical estimates. This additional reference sample collection may become prohibitively expensive and sometimes impossible if too much time has passed since the RS imagery was obtained. In other words, the high cost of obtaining reference data motivates an attempt to reduce costs by employing efficient and statistically sound sampling design strategies and structures (e.g., accessibility strata) (Stehman, 1998).

Finally it is worth mentioning that while probability sampling augmentation is impractical and expensive, reanalyzing RS imagery, even long after the reference sample data set has been collected, is relatively inexpensive. In other words, if another thematic map is generated from the same RS imagery provided with a reference sample data set, the novel thematic map accuracy assessment would be relatively inexpensive.

2.3 A probability geospatial sampling design protocol for accuracy assessment of a thematic map generated from RS imagery

This section specialize the probability geospatial sampling design protocol for the accuracy assessment of a thematic map generated from RS imagery in agreement with the four steps listed in previous Section 2.2.

1. ***Identification of the 3-D thematic map legend, imaging sensor model and 2-D thematic map legend***

    1.1. Identification and clear understanding of the 3-D thematic class set legend (taxonomy, ontology) including spatial (e.g., inclusion) and non-spatial (e.g., subset-of) relationships between geospatial thematic classes, and the selection of the target geospatial population(s) or class(es), equivalent to information layer(s) or stratum (strata), located in the area or region represented by the thematic map. In other words, the following list of information must be instantiated.

    1.2. The target surface areas of potential interest within the region represented by the thematic map must be defined

    1.3. Identify the discrete and finite set (taxonomy) of geospatial thematic classes in the 3-D world, also called 3-D object-models, hereafter identified as **3-D geospatial thematic classes** (e.g., LC/LU classes, GHCs, see Section 1). This is a **(3-D) scene domain prior knowledge, also called *world model* (Matsuyama, 1990), which deals exclusively with the visible (appearance, pictorial) properties of 3-D objects, but remains independent of the satellite imaging sensor at hand (Growe, 1999).** This prior geospatial knowledge of the world is typically represented as a **semantic network.** A semantic network description is reported in D5.1, which will be delivered in parallel to D4.3. The information primitives for a semantic network representation of the 3-D world model is reported in Appendix 1.

    1.4. Identification of an imaging sensor whose characteristics are listed in Appendix 2.

    1.5. Based on the selected imaging sensor model characteristics (e.g., spatial, spectral, radiometric and temporal resolution), a so-called Specialized 3-D Object Model Selection Expert (SOMSE) transforms the 3-D appearance properties of the specialized 3-D object-models belonging to the *semantic network* representation of the *world model* into a selected set of 2-D appearance properties of nodes and links [a2a], [a5a]. As output, this transformation provides an imaging sensor-dependent discrete and finite set of target thematic classes (categorical variables) distinguishable in the (2-D) RS image domain. In the rest of this work, target thematic classes of 2-D objects distinguishable in the (2-D) RS image domain based on their appearance properties are identified as **2-D thematic classes**. Thus, **a target 2-D thematic class set is always spaceborne imaging sensor-dependent**. It is noteworthy that **a taxonomy of 2-D thematic classes is equal to or a subset of the sensor-independent 3-D *world model*.** In practice, the target 2-D thematic class set is the adopted thematic map legend when the thematic map is generated from sensor-specific RS imagery. This prior geospatial knowledge about the 2-D thematic class models is typically represented as a **semantic network** instantiated according to a semantic network vocabulary (see Appendix 1).

2. Sampling design

2.1 Definition of the target thematic map class-specific accuracies and confidence intervals, in line with the QA4EO guidelines. For example, according to the U.S. Geological Survey (USGS), the target one-class overall accuracy probability (p_{OA}) $\in$ [0, 1] ± error tolerance (δ) is fixed at 0.85 ± 2%. The per class classification accuracy, $p_{OA,c} \in [0, 1] \pm \delta_c$, $c = 1, \ldots, C$, where C is the total number of LC classs, should be about equal and never below 70%, whereas a reasonable reference standard for δ_c is about 5% (Baraldi et al. 2010). More details in subsection 2.3.1

2.2 Estimation of the sample set cardinality as a function of the target 2-D thematic class-specific accuracy, confidence interval and budget constraints according to an optimum allocation strategy. More details in sub-section 2.3.1.

2.3 Sampling strategy by which geospatial sampling units are selected into the target geospatial population. More details in sub-section 2.3.2, 2.3.3 and 2.3.4

2.4 Geospatial sampling unit selection. The sampling unit is the fundamental unit on which the accuracy assessment is based. The sampling unit can be defined without specifying what will be observed on that unit; thus no assumption about homogeneity of thematic classes for the sampling unit is necessary. More details in sub-section 2.3.5

3. Response design

The response design protocol assigns the thematic label(s) of the reference geospatial sampling units (e.g., by means of photo-interpretation of high-quality RS imagery, field campaigns or a combination of these two information sources). It includes the ***evaluation*** and ***labeling*** protocol. More details in sub-section 2.3.6

4. Estimation and Analysis

The estimation and analysis protocol for a statistically rigorous quality assessment of a thematic map in comparison with reference geospatial sampling units based on a selected set of mutually uncorrelated QIs provided with uncertainty measures. More details in sub-section 2.3.7

Some of the abovementioned steps are further discussed hereafter.

2.3.1 Sample set cardinality required to meet the thematic map project requirements of class-specific accuracy, confidence interval and budget

When using stratification, there are several ways to allocate reference sampling units to different strata (e.g., proportional allocation, optimum allocation and Neyman allocation (Maniates et al. 2010)). For example, in optimum allocation, reference sampling unit selection should cost the least for a target confidence interval and a target accuracy estimate when costs of sampling are available (see Section 5.2.3).

In order to estimate the minimum number of reference sampling units to be selected, allocated and labeled for each thematic class of the thematic map to be accuracy assessed, Biging, Colby and

Congalton (Biging et al., 1999) propose a statistical criterion independent of the costs of sampling. This criterion is described below.

2.3.1.1 The criteria for reference sample set size (SSS) estimation independent of the costs of sampling

It is well known that any classification overall accuracy (OA) probability estimate, $p_{OA} \in [0, 1]$, is a random variable (sample statistic) with a confidence interval (error tolerance), $\pm \delta$, associated with it, where $0 < \delta < p_{OA} \leq 1$. In other words $p_{OA} \pm \delta$ is a function of the specific testing data set used for its estimation and *vice versa* (Baraldi et al. 2010). For example, for a given reference sample set size (*SSS*) comprising independent and identically distributed (i.i.d.) reference samples (in practice, this hypothesis is violated in image mapping problems due to spatial autocorrelation between neighboring pixels) and an estimated classification accuracy probability p_{OA}, it is possible to prove that the half width δ of the error tolerance $\pm\delta$ at a desired confidence level (e.g., if confidence level = 95 % then the critical value is 1.96) can be computed as follows (Biging et al., 1999):

$$\delta = \sqrt{\frac{(1.96)^2 \cdot p_{OA} \cdot (1 - p_{OA})}{SSS}}. \qquad (1)$$

Vice versa, minimum SSS = f(target p_{OA}, target δ) can be computed as follows:

$$SSS = \frac{(1.96)^2 \cdot p_{OA} \cdot (1 - p_{OA})}{\delta^2}. \qquad (2)$$

For each *c*-th class simultaneously involved in the classification process, with c=*1,…, C,* where C is the total number of classes, with $C \geq 2$ (at least, the total number of classes C comprises a target land cover class and class 'outliers'; It is noteworthy that the definition of a rejection rate is a well-known objective of any RS image classification system (e.g., refer to (Swain, 1978), p. 185), it is possible to prove that (Biging et al., 1999)):

$$\delta_c = \sqrt{\frac{\chi^2_{(1,1-\alpha/C)} \cdot p_{OA,c} \cdot (1 - p_{OA,c})}{SSS_c}}, \ c=1,\dots,C, \qquad (3)$$

where α is the desired level of significance (i.e., the risk that the actual error is larger than δ_c (e.g., α = 0.05), 1 – α/C is the level of confidence (e.g., if α = 0.05 and C = 5, then 1 – 0.05/ 5 = 0.99), and $\chi^2_{(1,1-\alpha/C)}$ is the upper *(1– (α/C)) * 100*th percentile of the chi-square distribution with one degree of freedom (e.g., if the level of confidence is (1 – 0.05/ 5) = 0.99, then $\chi^2_{(1,0.99)} = 6.63$).

Vice versa, the minimum SSS_c = f(target $p_{OA,c}$, target δ_c), c = 1, ..., C, can be computed as follows:

$$SSS_c = \frac{\chi^2_{(1,1-\alpha/C)} \cdot p_{OA,c} \cdot (1 - p_{OA,c})}{\delta_c^{\ 2}}, \ c=1,\dots,C. \qquad (4)$$

For example, target values and confidence intervals of community-agreed classification accuracy measures can be selected as follows. The target one-class pOA ∈ [0, 1] ± δ = (1) is fixed at 0.85 ± 2%, in agreement with the U.S. Geological Survey (USGS) classification system constraints. The per class classification accuracy, pOA,c ∈ [0, 1] ± δc = (3), c = 1, …, C = 5 (excluding outliers) should be about equal and never below 70%, whereas a reasonable reference standard for δc is about 5% (Baraldi et al. 2010).

For example, if desired level of significance α = 0.05 and C=5, then (1 – α/C) = 0.99 and χ2(1, 1-α/C) = 6.63. In this case, if pOA,c = 85%, with δc = ± 2%, then SSSc = (4) = 2113, c = 1, ..., 5. If pOA,c = 85%, with δc = ± 5%, then SSSc = (4) = 344, c = 1, ..., 5, and so on.

2.3.1.2 The REDD criterion for reference sample set size (SSS) estimation independent of the costs of sampling for continuous variables

To determine the SSS number of sampling plots for a geophysical continuous variable (e.g., biomass), given a certain confidence level and maximum error, one can apply the following formula:

$$SSS = \left(\frac{z^* \cdot \sigma}{e \cdot \mu} \right)^2 , \qquad (5)$$

where z* is the distribution critical value at a certain confidence level, σ is the standard deviation of the continuous variable (e.g., biomass), e is the maximum allowable error, and μ is the average of the continuous variable in the target stratum (e.g., forest). For example, for a forest where μ is 400 tons per hectar (t/ha) with σ is 65 t/ha, if you want to have an error of at most 5%, with 90% confidence level (z* = 1.645) then:

$$SSS = \left(\frac{1.645 \cdot 65}{0.05 \cdot 400} \right)^2 = 28.$$

For a 95% confidence level (z* = 1.960):

$$SSS = \left(\frac{1.960 \cdot 65}{0.05 \cdot 400} \right)^2 = 40.$$

Inversely, given a certain number of samples, the expected error can be calculated:

$$e = \frac{z^* \cdot \sigma}{\sqrt{SSS} \cdot \mu} . \qquad (6)$$

In all cases the mean μ and standard deviation σ of the continuous variable (e.g., average biomass in the forest μ and its standard deviation σ) need to be established first. This is best done by professional foresters, using generally accepted techniques for sampling. In practice this implies a minimum of 30 randomly located samples per forest stratum.

2.3.1.3 The optimal criterion for assessing SSS sampling units within the reference sampling budget

To give the most information per dollar spent (in other words, to cost the least for a given accuracy of the estimate of the target population or, for a given cost, to produce a minimum variance of the statistical estimator), the costs of reference sampling (assumed to be available), should be optimized with respect to the quality of reference samples, namely, the repeatability and even the accuracy of the reference samples (Stehman, 1998).

To reduce the overall reference sampling costs, a typical response design strategy is to label the whole or a subset of the required set of reference sampling units based on high-quality RS images of the sampling units. High-quality RS image selection, purchase and photo-interpretation by scientists and domain experts typically reduces significantly the volume and cost of reference sampling measurements that needs to be undertaken in the field (Maniates et al. 2010).

2.3.2 Positive inclusion of samples in probability sampling

Probability sampling requires that all inclusion probabilities be greater than zero, and the inclusion probabilities must be known for those sampling units selected in the sample. If some sampling units have an inclusion probability of zero, the assessment does not represent the entire target region of the map. Excluding inaccessible areas or heterogeneous edges between polygons is an example of assigning sampling units an inclusion probability of zero. Requiring the inclusion probabilities to be known is necessary so that statistically valid (i.e., consistent) estimates can be computed (Stehman, 1998).

To summarize, it is crucial to develop a geospatial sampling strategy where the probability of an element being included in an arbitrary sample of the population is known and where each element in the population has a positive inclusion probability. For some issues (e.g., rare fire events), the use of non-random sampling methods is recommended (Stehman, 1998).

Systematic sampling, random sampling, stratified random sampling and cluster sampling are all probability sampling designs. When using such designs in practice, the inclusion probabilities do not have to be computed explicitly because they are already taken into account in the standard estimation formulas. But if a new or non-standard sampling protocol is constructed, then the investigators must specify the inclusion probabilities. The inclusion probabilities determine the weight attached to each sampling unit in the estimation formulas, and if the inclusion probabilities are unknown, so are the estimation weights. A **good rule to apply when planning an accuracy assessment is that if the sampling protocol cannot be identified as a standard probability sampling design and the project planners are unable to specify the non-zero inclusion probabilities, the proposed design should be discarded** (Stehman, 1998).

2.3.3 Geospatial stratification prior to selection of reference sampling units

The geospatial stratification process consists of separating the entire geospatial surface of interest into mutually exclusive (non-overlapping) and totally exhaustive ***strata*** (***layers***) (including layer "outliers", "unknown" or "no target data" complementary to target layers, if useful). These are equivalent to relatively homogenous surface areas selected on the basis of an often subjectively chosen measure of similarity in an arbitrary feature space (e.g., dealing with landscape, ecological or thematic class features). This is based on a similarity measure chosen subjectively based on its ability to create "interesting" layers, so that the similarity (variation) within each stratum is maximized (minimized) at the expense of the similarity (variation) between the strata.

When geospatial strata are available, then samples units are taken from each stratum to obtain a more efficient estimate of the total population according to rigorous statistical criteria.

In general, geospatial stratification attributes must be available before (prior to) the geospatial statistical sampling takes place. Thus, geospatial stratification is possible if, and only if, prior geospatial knowledge is available before the statistical sampling occurs.

Common stratification attributes belong to the following two geospatial information domains (Stehman, 1998).

- Geographic information different from thematic (symbolic) classes. This includes administrative regions, ecoregions, sub-symbolic (non-semantic) layers of mountainous height generated from a digital elevation model (DEM), etc. In particular, geographic stratification can be used to distribute sampling efforts evenly among administrative regions or ecoregions or to sample accessible areas with higher probability than expensive, but low-priority, inaccessible regions.
- Mapped thematic classes (e.g., forest type) based on a pre-existing thematic (symbolic, classification) map. Stratifying by mapped thematic classes may ensure that a specified sample set cardinality is obtained in each mapped class, including those rare classes that would not be prevalent in a simple random or systematic sample without stratification. A disadvantage of stratifying by mapped thematic classes is that it locks the assessment into the map version used to form the strata. If this map is subsequently revised or the thematic classification scheme changed, the original strata are still valid, but they no longer correspond to the thematic classes of the revised map.

It is noteworthy that, before stratified statistical sampling occurs, the relationship between semantic (symbolic) or non-semantic (sub-symbolic) strata available on an *a priori* basis and the target thematic classes to be sampled within strata must be carefully investigated and fully understood by the application developer.

In general, relationships between prior geospatial strata (e.g., ecoregions) and target thematic classes (e.g. deciduous forest) must be assumed to be many-to-many, where relationships one-to-many, many-to-one and one-to-one are considered as special cases. For example, in the BIO-SOS project, when a SIAM™ preliminary classification map (Baraldi et al. 2006; Baraldi, 2009) is automatically generated from RS imagery to be employed for stratification purposes, these **semantic strata** (consisting of spectral categories equivalent to sets of LC classes) **are, in general, many-to-many related to target 2-D thematic classes**.

2.3.4 Probability sampling methods

Simple random sampling, systematic sampling, stratified random sampling and cluster sampling are all probability sampling designs (Stehman, 1998). They are all considered as reference standards because they guarantee: (a) the probability of an element being included in an arbitrary sample of the population is known and (b) each element in the population has a positive inclusion probability. As a consequence, when using such probability sampling designs in practice, the inclusion probabilities do not have to be computed explicitly because they are already taken into account in the standard estimation formulas.

The three major map sampling schemes are described below (Maniates et al. 2010).

2.3.4.1 Simple random sampling

Random sampling of a statistical distribution (population) is simple to implement, but it is not optimal as there may be a lack or absence of sampling units belonging to thematic classes that have a lower occurrence.

2.3.4.2 Systematic sampling

Systematic sampling at equal intervals of a statistical distribution is simple to implement, but it has the same limitation as random sampling. Since a spatially well-distributed sample is produced, precision is typically better relative to simple random sampling.

2.3.4.3 Stratified random sampling (SRS) and SRS within a regular grid (SRSRG)

Stratification of a thematic map entails the division of a sampling area into non-overlapping strata (layers, subsets; see Section 2.3.3). When SRS is adopted as probability sampling design, the following considerations hold.

- SRS produces estimates that are unbiased provided that each stratum value is weighted according to the proportion that the stratum forms of the entire population. The accuracy of the estimate can be assessed provided that a minimum of two sampling units occur within each stratum (Maniates et al. 2010).
- SRS achieves the positive inclusion probability of samples per stratum (see Section 2.3.2).

To combine advantages of SRS and systematic sampling while minimizing their drawbacks, the REDD requirement of the United Nations Framework Convention on Climate Change (UNFCCC) promotes an SRS strategy within each cell (sampling unit area, area equivalent unit) of a regular grid (SRSRG) (Maniates et al. 2010). SRSRG provides several advantages. First, it provides a separate estimate of the mean and the variance of the statistical estimator in each stratum. The result is that there will be a different sampling density for the different land cover types. Second, for a given sampling intensity, it yields more accurate estimates of the population parameters. Finally, it ensures better distributed coverage of the population than simple random sampling (Maniates et al. 2010). Shown in Fig. 2, the SRSRG method promoted by REDD is described below.

- In SRSRG the strata have an overlay of a systematic grid.
- Each stratum (e.g., forest type), is divided in an equal number of sampling unit areas (area equivalent units). Thus, the size of a sampling unit area is stratum-specific. The REDD guidelines propose **the division of each stratum into 25 to 30 area equivalent units**. The choice of 25 to 30 area equivalent units is considered to be statistically sufficient for guaranteeing a normal distribution of sample estimators according to the large sample theory.
- Sampling measurements would be made in each unit (e.g., 10 strata would equal 250 points).
- Sampling units would be chosen at random in each area equivalent unit as follows. (i) The first sampling unit is selected at random in each area equivalent unit; (ii) To avoid that the second sample within the area equivalent unit is too close to the first, a distance rule can be imposed

on the random sampling. This either forbids sample points below a certain distance or selects a replicate at the maximum possible distance of the previously selected sampling unit.

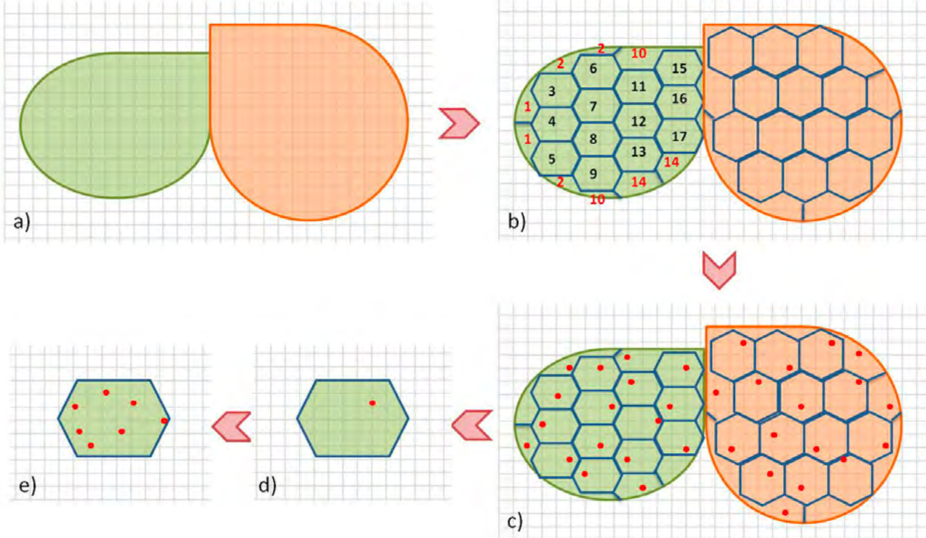

Fig. 2. Proposed sampling method, sample allocation and sample distribution for two hypothetical strata. (a) two different strata overlayed with a systematic grid; (b) the same as (a) but with each stratum divided into the same number of area equivalent units (hexagons); (c) same as (b) but with one randomly selected sample point taken in each of the area equivalent units; (d) zoom in on one area equivalent unit and (e) the distribution of seven sample points in the area equivalent unit using a distance rule for the sample location choice.

2.3.5 Design of geospatial sampling units

The sampling unit (e.g., 0.1 ha pixel, 10 ha polygon, 1000 ha circular plot) is the fundamental unit on which the accuracy assessment is based. The comparison of the thematic map and reference sample classifications is conducted on the spatial scale of a sampling unit. For example, if a (discrete, small but finite) pixel is chosen as the sampling unit, the reference sample classification is obtained for each pixel (as represented on the earth) and compared to the corresponding map pixel. If the sampling unit is a (dimensionless) point, the correspondence is between the classification provided by the map at that point, and the reference sample classification associated with the same point location on the earth. Because the sampling unit is the ultimate basis for the comparison of the thematic map and reference sample classifications, whatever sampling unit is chosen, it is essential that this choice be explicitly and clearly stated and considered acceptable to users of the thematic map (Stehman, 1998).

The sampling unit can be defined without specifying what will be observed on that unit on the ground; thus no assumption about homogeneity of thematic classes for the sampling unit is necessary. There are two types of sampling units (Stehman, 1998).

a. (Dimensionless) Point, featuring no area extent. The statistical population associated with a point sampling unit is viewed as continuous.
b. Areal unit featuring a 2-D spatial coverage. The statistical population associated with areal units is considered as partitioned into discrete spatial units such as pixels or polygons. The three primary areal sampling units are the following.
   i. Pixels, representing small areas (e.g., 30 m pixel) are related to point sampling units, but because pixels still possess some areal extent, they partition the mapped population into a finite, though large, number of sampling units, see Fig. 3.
   ii. Polygons. Polygon sampling units are usually irregular in shape and differ in size. Homogeneous thematic polygons can be selected on the thematic map. Photo-interpreted polygons are identified on Earth from RS imagery. Homogeneous thematic map polygons have an appealing convenient structure for the sampling unit and a direct correspondence to the thematic representation displayed by the

map. A disadvantage of using homogeneous thematic map polygons as sampling units is that the sampling units are now inseparably bound to a particular map. For example, if this map is updated, then the original polygon sampling units are still valid for the assessment, but they may no longer correspond to homogeneous thematic polygons of the revised map, see Fig. 4.

iii. Fixed-area plots. Fixed-area plot sampling units are usually regular in shape, and cover some predetermined areal extent. In practice, pixels and polygons are special cases of fixed-area plot sampling units. Fixed-area plots defined independently of thematic polygons, either of the map or the ground, retain their identity under map revisions and over time. The disadvantage is that these units do not correspond directly to thematic polygons, either of the map or the ground.

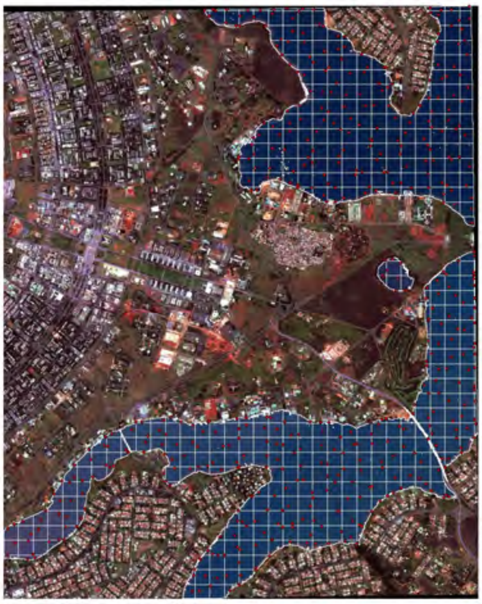

Fig. 3. Example of an SRSRG approach for reference geospatial sampling unit selection. Stratum (visible in light blue) = water or shadow (WASH) spectral category detected by SIAM™ in a 4-band WorldView-2 image of Brazilia, Brazil, radiometrically calibrated into TOARF values. This target phenomenon features no geometric property, thus, sampling unit type = (dimensionless) point, featuring no area extent. Systematic grid consisting of square cells, 1 sampling unit per cell, with the number of cells computed with Eq. (4).

Fig. 4. Example of an SRSRG approach for reference geospatial sampling unit selection. Stratum (visible in light blue) = bright bare soil or built-up (BBB) spectral category detected by SIAM™ in a 4-band WorldView-2 image of Brazilia, Brazil, radiometrically calibrated into TOARF values. This target phenomenon features geometric properties (e.g., rectangularity), thus, sampling unit type = polygon. Systematic grid consisting of square cells, 1 sampling unit per cell, with the number of cells computed with Eq. (4). Reference sampling objects, depicted in red, are selected where reference sampling units are located.

In general, some strata may have higher within-stratum variability (of a target thematic class-specific variable) and thus the sampling unit design will need to take such variability into account, resulting in a sampling unit design protocol being adopted separately for each stratum.

A thematic class-specific sample spatial type selection protocol can be the following.

Is the target population in the (2-D) image or 3-D world characterized by (non-contextual) color (either chromatic or achromatic) properties? If yes, then the sample spatial type is point,

otherwise

the target population in the (2-D) image of 3-D world is characterized by contextual (either chromatic or achromatic) properties, namely,

- texture
    - featuring foreground exclusively (e.g., forest in a VHR image).
    - featuring both foreground and background (e.g., an olive grove consisting of olive tree crowns as foreground and bare soil or grass as background in a VHR image).

OR

- Shape/morphology (e.g., buildings in a VHR image)

therefore

the sample spatial type is a polygon.

The choice of a region of interest (ROI) as a polygon must be representative of the spatial feature characterizing the target object. For example:

- If texture, then the ROI must capture the texture space period and orientation.
- If shape/morphology, then the ROI must capture the object size and shape, (e.g., rectangularity of buildings in a VHR image, elongatedness of a river, etc.).

2.3.6 The response design protocol to label the reference sampling units

The response design protocol determines the thematic label of every reference geospatial sampling unit. Response design requires input from scientists and analysts having a clear understanding of the land-cover classification scheme being used in the mapping project. Because of the interpretive (subjective) nature inherent in determining thematic classification (in fact, terms semantic and subjective are synonyms! (Baraldi, 2011)), the response design may also require a reliability or quality control component to **evaluate the repeatability and even the accuracy of the reference sampling units** themselves.

Conceptually it is useful to separate the response design into two components: the evaluation protocol and the labeling protocol which are summarized below.

2.3.6.1 The evaluation protocol

The evaluation protocol consists of the procedures used to collect information contributing to the reference sample classification determination. Information pertaining to the reference thematic determination are collected for each sampling unit from the following independent sources of evidence eligible for being used in parallel.

1. Photointerpretation of high-quality (fine spatial resolution, multi- or hyper-spectral) RS images depicting the target sampling units on the ground.
2. A ground visit.
3. Prior domain knowledge of the target sampling units on the ground by domain experts, scientists and analysts based on intuition, expertise and evidence from data observations.
4. A combination of these sources of reference information. This combination should provide the best compromise between conflicting objectives mentioned below.
    a. Be feasible in practical terms, e.g., accomplish geographic accessibility of sampling units on the ground.

b. Be consistent across time and space to reduce systematic and accidental errors in reference sampling units labeling. Systematic and accidental labeling errors should be carefully addressed and made explicit in the thematic map accuracy assessment project report, to properly deal with the propagation of errors (Baraldi et al. 2010; Bruzzone, 2009). For example, typical inconsistencies between spaceborne imagery and reference ground truth samples can be due to: (i) land cover changes occurring in the time interval (e.g., multi-annual in these experiments) between the ground truth data acquisition campaign and the RS image acquisition date, (ii) accidental problems occurred in the acquisition/ storage/ recovery of ground truth samples and/or (iii) spatial inconsistencies between the orthorectification/co-registration of reference samples with spaceborne imagery. In particular, the following thematic map assessment project requirements should be fulfilled.
   i. Be in line with the thematic map accuracy assessment requirements in terms of elapsed time between the time stamp t1 of the testing thematic map and the time stamp t_2 of reference sampling units.
   ii. Evaluate the repeatability and even the accuracy of the reference sampling units themselves. For example, in (Baraldi et al. 2010), a semantic cross-checking between the reference dataset and a SIAM™ preliminary classification map (Baraldi et al. 2006; Baraldi, 2009) automatically generated from four RS testing images revealed that the average rejection rate of reference samples across the four input images was about 5.4%. This rejection rate is statistically relevant and representative of a potential source of uncertainty of classification accuracy measurements largely underestimated in RS common practice.

c. Respect the sampling budget, e.g., by reducing costs of collecting reference information. On the one hand, it is well known that the strategic methodological option of using RS high-quality data rather than field data to assess reference samples allows a significant reduction in the volume and cost of field measurements (Maniates et al. 2010). On the other hand, in RS common practice the cost, timeliness, quality and availability of adequate reference (training/testing/validation) datasets derived from field sites, existing maps and tabular data are currently considered the most limiting factors on RS data-driven high-level product generation and validation (Gutman et al. 2004).

The evaluation protocol starts from choosing the size and shape of the **spatial support region** on which the reference sample classification evaluation will be based on the information source, either RS imagery or field campaign. In particular:

   i. **if the sampling unit is a point, the evaluation need not be limited only to what the evaluator observes at that point location.**
   ii. **If the sampling unit is an areal sampling unit, namely, pixel, polygon or fixed area, a spatial support region defined for an areal sampling unit may or may not be the areal unit itself**. For example, a 30 m pixel may be assigned a support region of 1 ha. The spatial support of a polygon sampling unit will usually just be the polygon itself.

Once the support region has been identified, numerous options are available to determine the reference sample classification. In some cases, the evaluator may visually scan the support region and record qualitative observations contributing to an eventual classification of the sampling unit. In other cases, the evaluation protocol may specify recording species composition, canopy closure, or distribution of tree sizes, or require other quantitative data needed to distinguish among land-cover classes or to characterize the land cover of the sampling unit. The evaluation protocol should conform to the users' concept of error-free classification; any compromises should be agreeable to users (Stehman, 1998).

The evaluation protocol may include subsampling within the areal unit. Using transects, quadrats, or gridded point samples are candidate response design sampling methods for estimating quantitative continuous variables that contribute to the land-cover classification of a sampling unit. The response design subsampling also provides information on within-pixel or within-polygon heterogeneity. This information may be relevant to the subsequent labeling protocol, or to characterize heterogeneity within a particular land-cover class (Stehman, 1998).

However, the primary objective of the evaluation protocol is to obtain information pertinent to identifying a reference thematic label for each sampling unit.
The evaluation protocol will be further discussed in D5.5 of WP5, Task 5.3.

2.3.6.2 The labeling protocol

The labeling protocol, which assigns a land-cover classification to the sampling unit based on the information obtained from the evaluation protocol.

1. Crisp, e.g., primary or primary and secondary thematic class. At the most basic level, the reference sampling unit is labeled as one and only one thematic class. A primary class labeling suffers from the potential problem that a sampling unit may consist of several different thematic classes, or represent a transition or mixed class not easily identified as a single cover type. Because it is not always possible or desirable to label the sampling unit as a single thematic class, the labeling protocol may specify recording both a primary and secondary thematic class (e.g., based on a defuzzification strategy).
2. Fuzzy class memberships for every class in the target 2-D thematic class taxonomy, such that the sum of fuzzy memberships ranges from 0 to T, where T is the cardinality of the target 2-D thematic class taxonomy.
3. Quantitative, e.g., area proportions for each thematic class present in an areal sampling unit or spatial support region, such that the sum of area proportions equals 1.

2.3.7 Protocol for the estimation and analysis of the degree of match between the thematic map and the reference geospatial sampling units

The protocol for the estimation and analysis of the degree of match between the thematic map and the reference geospatial sampling unit will be further discussed in D5.5 of WP5, Task 5.3.

3. PART 2

3.1 Probability geospatial sampling design and analysis for LC/LU thematic map validation in the IT4 test

In this section, the multi-step thematic map validation protocol described in Section 2.2 of PART1 of this Deliverable and based on the stratified random sampling within a regular grid (SRSRG) sampling strategy is instantiated for the validation of LCCS maps generated from RS imagery of an Italian training site. To obtain the desired levels of accuracy and error tolerance, a large number of sampling points are needed for each LCCS category. However, the costs of undertaking such sampling may be too high and hence there is a need to optimize the amount and proportion of those which are to be sampled in the field and which can be sampled through reference to thematic maps or appropriate selection of a subset (cardinality = SSS*) of the whole set of LCCS sampling units (cardinality =SSS), such that SSS* ≤ SSS are the sampling units expected to be labelled by in-field inspection and not by complementary sources of information. The sampling approach also needs to consider the protocols for GHC sampling reported in the EBONE manual (Bunce et al. 2011) (see PART3).

3.1.1 Identification of the 3D thematic map legend

3.1.1.1 Identification of the 3-D LCCS classes in the real-world: IT4 test site and its surrounding area

For the Italian test site (IT4), the list of 19 3-D LC/LU classes in LCCS taxonomy is reported in Table 3.1.1 and the corresponding GHCc are listed in D5.1. Cultivated surfaces are both within and in the neighbouring area of the IT4 Natura 2000 sites. Sea water is outside the boundary of IT4.

Table 3.1.1 - List of 3-D LC classes in LCCS taxonomy for the IT4 test site.

Class index	Secondary class index	LCCS class code		LCCS class description	ANNEX1 / EUNIS code
1		**A12** **Natural and seminatural terrestrial vegetation**	**A2.A5.**A10.B4.E5.B12.**E6**	Closed annual medium/tall forbs	X / E1.6
2		**A12**	**A1.A4.**A10.B3.**D1.E2.**B9	Broadleaved deciduous medium/high closed shrubland (thickets)	X / F5.51
3		**A12**	**A1.A4.**A10.B3.**D2.E1.**B9	Needleaved evergreen medium/high closed shrubland (thickets)	2250 / B1.63
4		**A12**	**A1.A4.**A11.B3.**D1.E1.**B10	Broadleaved evergreen open dwarf shrublands	X / F6.2C

Table 3.1.1 (continued) - List of 3-D LC classes in LCCS taxonomy for the IT4 test site.

5	5.1	**A12**	**A1.A4**.A10.B3.**D1.E1**.B9 **+** topology+other attributes	Broadleaved evergreen medium/high closed shrubland (thickets)	5330 /F5.55
	5.2				X / F5.514
6		**A12**	**A2.A5.**A11.B4.E5.B13.**E6**	Open annual short forbs	1210 / B1.1
7		**A12**	**A2.A6.**A11.B4.E5.B12.**E7**	Open perennial medium-tall grasslands	2110 / B1.31
8	8.1	**A12**	**A2.A5.**A11.B4.E5.A13.B13.**E6**	Open (40-(20-10)%) annual short herbaceous vegetation	2230 / B1.48
	8.2				6220 / E1.313
9	9.1	**A24** **Natural and seminatural aquatic or regularly flooded vegetation**	**A2.A5.**A13.B4.C2.E5.B13.**E6** **+**topology+other attributes	Open annual short herbaceous vegetation on temporarily flooded land	3170 / C3.421
	9.2				1310 / A2.51
	9.3				1310 / A2.55
10		**A24**	**A1.A4.**A12.B3.C2.**D3**.B10	Aphyllous closed dwarf shrubs on temporarily flooded land	1420 / A2.526
11	**11.1**	**A24**	**A2.A6.**A12.B4.C2.E5.B11.**E7** **+**other technical attributes (species)	Perennial closed tall (3-0.8m) grasslands on temporarily flooded land	1410 / A2.522
	11.2				7210 / D5.24
	11.3				X / D5.1
	11.4				X / D5.2
12		**A24**	**A2.A6.**A12.B4.C2.E5.B12.**E7**	Perennial closed medium-tall (0.8-0.3m) grasslands on temporarily flooded land	X / C2
13		**A24**	**A2.A5.**A16.B4.C1.E5.A15.B12.**E7**	Perennial sparse medium tall herbaceous vegetation on permanently flooded land	1150 / X03
14		**A11** **Cultivated and managed**	**A3**	Herbaceous crops	

Table 3.1.1 (continued) - List of 3-D LC classes in LCCS taxonomy for the IT4 test site.

15	**A11**	**A1.**B1.C1.D1.**W8.A7.A9.**B3	Monoculture fields of rainfed broadleaved evergreen tree crops orchards (olive groves)
16	**A11**	**A1.**B1.C1.D1.**W7.A8.A9 .B3**	Monoculture fields of rainfed needleleaved evergreen tree crops plantations
17	**B15 Artificial surfaces**	A1.A4.A12.A17	Scattered industrial or other areas
18	**B15**	A1.A3.A7.A8	Paved roads
19	**B28**	A1.B1	Perennial natural waterbodies
			Effective 19

A12=Natural and seminatural terrestrial vegetation, A11 = Cultivated and managed, Natural and seminatural aquatic or regularly flooded vegetation B15 = Artificial surfaces

3.1.1.2 Identification of a space borne imaging sensor

Scale 1:5000 or better is required by users for the LC/LU mapping of the IT4 site whose spatial extent is 3.5 km^2. This scale requirements are compatible with the spatial resolution and swath width of VHR spaceborne imagery and QUICKBIRD (QB) and WORLDWIEW-2 (WV-2) images will be considered, largely because of accessibility and their comparatively low cost.. For this test site, three seasonal images should be considered for producing LCCS, GHC and Directive 92/43/EEC Annex1 habitat maps as the seasonal variability in the spectral reflectance characteristics can be captured. The optimal periods are winter (January/February), spring (April/May and mid to late summer (July to September). Images acquired in the second and third period are particularly useful for GHC mapping since this period corresponds to the peak of the growing season of different habitats (e.g., Directive 92/43/EEC Annex 1 habitat 3170 and 1310. Both are temporarily flooded in winter but the peak of their growing period is between May (June) and August (Sept), respectively. In particular, habitat 3170 is dry in August when 1310 is most productive. A summary of selected input sensor-generated data requirements is provided below.

1. Geographic area of interest. Le Cesine, Lat_Log (see D2.2. for a more detailed description)
2. Spectral resolution. Multispectral (MS) four bands *B, G, R, NIR*, plus Panchromatic (PAN).
3. Spatial resolution. 2.4m (QB, MS) and 0.60 m (QB, PAN).
4. Radiometric resolution (typically, 1 byte, i.e. 256 gray levels): 1 byte in TOARF values (see below, thus quantization error is: (1/255) / 2 = 0.2%.
5. Observation timing (e.g., seasonal considerations). The phenological description of each class is reported in Table 2. As a minimum, three observations per year are needed.
6. Required spatial quality. Co-registration/orthorectification error: less than one pixel.
7. Geo-coding (geographic projection):UTM WGS84.
8. Required radiometric quality = Radiometric calibration = Transformation of digital numbers into TOARF values.

Required image enhancement in terms of topographic correction: no topographic correction is required because the IT4 test site corresponds to a flat surface area.

3.1.1.3 3-D LC/LU Class Description/explanation/definition in terms of (2-D) appearance properties in the 2-D RS image domain

Based on the selected sensor properties (see Section 3.1.1.2), the 3-D LC/LU class set listed in Section 3.1.1.1 maintains its cardinality equal to NC = 19 in the 2-D image domain. In particular, based on the given imaging sensor model, the target 3-D LC/LU classes are characterized by the the list of pictorial (visual, appearance) properties reported in Table 2 (Column 2 for class 15; i.e. A11/ A1.B1.C1.D1.W8.A7.A9.B3). It is noteworthy that 3-D LC/LU classes are described in the real world based on visual properties expressed in physical units of measures (e.g., the distance between trees in an olive grove ranges from 10 to 15 meters, the diameter of an olive tree crown ranges from 5 to 10 m, etc.). In the 2-D image domain, specific operators should be selected and their free parameters measured in pixel units. For the selected class, the operator output values are reported in column 3 of Table 2.

Figure 3.1.1 shows a subset of the IT4 test site depicted in the original MS (2.3 m) image (a), the pansharpened (PANSH) image (b) and the SIAM™ map generated from the PANSH image at the finest level of granularity of spectral categories (equal to 52) (c). The full Pansharpend image is shown in the Figure 1 (h). Different image subsets centred on medium aged trees and very young trees are displayed in Figure 1 (f) and (i). The SIAM™ output map is used to extract the photometric properties of the selected class reported in column 3 of Table 3.1.2. Figure 3.1.1 (g) and (l) show the SIAM™ output of the subsets in Figure 3.1.1 (f) and (i).

Morphologic operators, such as *opening–closing* transforms (Soille, 2002), are used to isolate bright (opening) and dark (closing) structures in images, where bright/dark means are brighter/darker than the surrounding features in the image (Benediktsson, 2003). As an example, consider Figure 1(i-l) where trees crowns are darker than the background soil. When morphological operator operators are used in image processing, these operators are applied with a set of structuring element (SE) of known shape (e.g., circular in the case of olive tree crowns). Most operators come as pairs. In order to isolate features with a thinner support than a given SE, a widely used technique is to take the residuals of the *opening, closing* and the original images by a morphological transformation called *top-hat* and *bottom-hat* (Soille, 2002). Once the dimension (in pixels) of the interest bright/dark structures in the image have been defined, morphological filters or other operators can be used. Table 2 (Column 2) reports the dimension of the tree crowns as dark objects in the image.

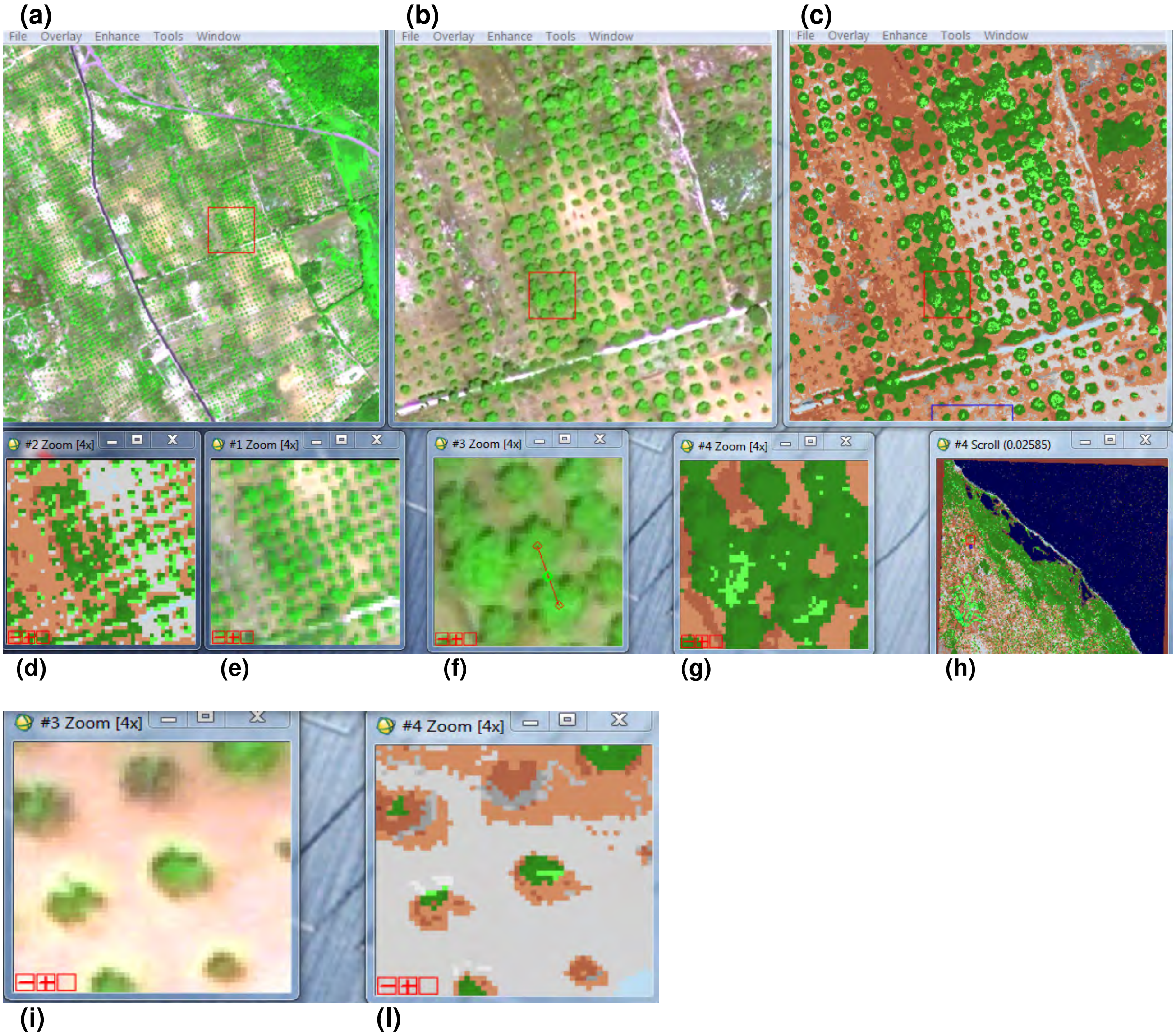

Figure 3.1.1 - (a) MS Quickbird image acquired on July 2005, RGB=3,4,1; (b) geo-linked (in ENVI) PANSH image, RGB=3,4,1; (c) PANSH SIAM™ 1st stage output map with 52 spectral categories; (d) window of the MS SIAM™ 1st stage output map; (e) window of (a), (f) window of image (b) centered on medium aged olive trees; (g) PANSH SIAM™ 1st stage output map of the window in (f); (h) PANSH SIAM™ map of the full image.

Olive groves are generally characterized by rows of tree oriented in two directions. The period (i.e., the distance between trees) and the orientation of the tree rows can be automatically detected by a texture operator. Amoruso (2009) used the Variogram Based Texture operator (Sanz, 2006) was to determine, in the 2-D image domain, the number of orientations, the orientation angle and the period of olive tree fields by analysing cumulate variance of the variogram image of a Quickbird image panchromatic band.

Table 3.1.2 - Descriptions of Class 16 (A11/A1.B1.C1.D1.W8.A7.A9.B3) in both the 3-D world domain and 2-D image domain.

Ci Index	***3-D LC/LU description/explanation/definition in terms of (3-D) appearance (visual, pictorial) properties in the 3-D world domain. Spatial units of measures: meters.*** ***Semantic net is composed of:*** • ***nodes (e.g LCCS or GCH class, object composing the class) and edges (i.e. relations between objects)*** • ***relations: IS-A, PART-OF , DATA-OF, CON-OF*** • ***attributes: photometric, geometric, morphologic, texture***	***Description/explanation/ definitions in terms of 2-D appearance (visual, pictorial) properties in the 2-D image domain of Quickbird imagery featuring: spatial resolution =0.6 m and spectral resolution= B,G,R,NIR.*** NOTE: based on these 2-D class-specific pictorial properties an expert programmer should be able to write the source code of a class-specific rule-based classifier employing prior knowledge of the 3-D real world.
15	*NODE:* Monoculture field of rainfed broadleaved evergreen tree crops, orchard (olive groves) *EDGES*: *Class* IS A cultivated area *Class Foreground:* olive tree crown (PART OF). Tree height range [1.5m, 4m] *Class Background:* soil (PART OF) and, depending on seasonality (PART OF) , shadow (PART OF) as well as grass (PART OF), the latter depending also on agricultural practices (pesticide?)). *Temporal relation: (Class Phenology)***:** perennial, evergreen (TEMPORAL RELATION) **Photometric**: colour properties: Background olive tree crown are green Background soil colour ranges from maroon to very bright green due to soil graining procedures **Geometric** (area, perimeter, compactness, straightness, elongation, rectangularity, no. of vertices) Mean perimeter (P) of foreground (i.e. tree crown): 15m and 34m for young and old trees, respectively. Mean area (A) of foreground: 18m^2 and 34 m^2 for young and old trees, respectively. Compactness of foreground (P^2/A): 12,5 Shape: Tree crowns have a circular shape	***Photometric:*** (a) chromatic properties: SIAM™ spectral categories (see Figure 1 in the text); (b) achromatic properties (range of brightness values): *Medium aged trees* o *Foreground SIAM™ spectral categories. Tree crown : 21 ASHRBR HNIR, 22 ASHRBR MNIR, 23 ASHRBR LNIR.* o *Background SIAM™ spectral categories. Soil: 32 BBB_TNCL , 33 SBBVF_LSC , 35 SBBNF_LSC, 37 ABBF , 38 ABBNF_LSC. No shadow in July image.* *Young trees* o *Foreground SIAM™ spectral categories: 22 ASHRBR MNIR, 23 ASHRBR LNIR* ***but also 34 SBBF, 35 SBBNF_LSC, 37 ABBF, 38 ABBNF_LSC.*** *Background SIAM™ spectral categories: soil 32 BBB_TNCL , 33 SBBVF_LSC , 35 SBBNF_LSC, 37 ABBF , 38 ABBNF_LSC.* ***Geometric properties:*** o *Foreground tree crowns have circular shape with diameter in ranges(in pixels) :* ▪ *[7,11] for medium aged trees* ▪ *[3, 5] for very young trees*

	Morphological attributes: (a) dark object in a bright background: tree shadows (b) bright object in a dark background: diameter (in m.) [4, 7] for medium aged trees and [1.8, 3] for young olive trees	***Morphological attributes (e.g., based on*** *opening-closing morphological filters) with SE=3*3 windows* ○ *Bright object over dark background* ▪ *Area of opening region: 106 pixels* ▪ *Area of closing pixel: 118 pixels* ▪ *Perimeter of opening region: 60 pixels* ▪ *Perimeter of closing region: 52 pixels* ▪ *Top-hat of opening: 6 pixels* ▪ *Top-hat closing: 6 pixels*
	Textural attributes: *Periodicity:* equivalent to the tree-to-tree average distance: range [10m, 20 m] *Oriented texture*: generally Y, but not always (see Figure 1) *Number of directions*: 2 generally orthogonal (but not always)	***Textural attributes*** *are based on a Variogram Based Texture (VBTA) operator (Sanz, 2006)* *Window size range for young/old (2*max period) = [36*36] pixels* *Oriented texture: VBTA orientation test positive* *Number of directions(by the analysis of cumulate variance of the variogram image): 2 with* α_1 *[*31^0*,* 53^0 *];* α_2*: [*121^0*,* 143^0 *]*
	Spatial relations *Topological (e.g adjacency, inclusion, right-of) properties*: Background shadow (in winter) is ADIACENT to foreground (tree crown) *Non-topological (e.g., close –to, distance-from, in-between- angle) topological attributes:* The distance between tree crowns for both medium aged olive trees and very young olive trees ranges in: [9m, 11m]	***Spatial properties:*** ○ *Period T ranges (in pixels):* ▪ *Medium aged as well as young trees in this area: [16,18]* ○ *Window size (spatial domain of activation) ≥ 2 * period:* ▪ *36 pixels*

3.1.2 Sampling design

3.1.2.1 Definition of the target LCCS class-specific accuracies and confidence intervals

In the IT4 test site depicted by VHR imagery, and based on RS common practice, we can assume that the target one-class overall classification probability P_{OA} is equal to the one proposed by USGS, (i.e., 0.9 (or 90 %) ± 0.2). The per class classification accuracy should be no lower than 0.85 with an error tolerance δ_c = ± 0.1, for all classes except for some natural-semi natural aquatic or regularly flooded LCCS classes (i.e. A24; for which 0.95 and δ_c = ± 0.05, is requested because of the need to correct identify the wetland classes in the IT4 test site). Project requirements will be checked with end users.

3.1.2.2 Sample set cardinality

According to Equation (4) in Part 1 (Section 2.2.2), the sample set cardinality SSSc for each class c = 1, ... NC, where NC = 19, is reported in Table 3.1.3 (Column 3). If the number of sampling units collected by the different sources of information indicated in Table 3 (Column 4) is lower than the required SSSc, then the error tolerance can be computed according to Biging, Colby and Congalton equation (3) reported in Part 1 (Section 2.3.1.1).

3.1.2.3 Stratified sampling strategy within a regular grid (SRSRG)

In a pre-existing reference thematic map available for the IT4 test case (from a previous European Program Interreg III-A Greece - Italy 2000-2006, code I3101001 project), LC/LU classes were labelled according to the CORINE taxonomy. This taxonomy was therefore converted into the LCCS taxonomy (see D6.1) for the purposes of the BIO_SOS project.

A proposed SRSRG sampling approach includes the following steps:

- For each target LCCS class *c* (with *c*= 1,…, NC=19), the corresponding class-specific stratum should be extracted from the shape file of the reference thematic map.
- Within each stratum (with *c*= 1,…, NC=19)**,** the Minimum Enclosing Rectangle (MER; *c*= 1,…, NC=19) is considered. A set of 30 areal units are overlapped with MER *c*= 1,…, NC=19.
- Within each MERc, *SSSc* points, with c= 1…19, should be randomly sampled within the set of 30 unit areas whose overlap with the stratum *c*= 1,…, NC=19 is above zero. Within each eligible areal unit, sampling units should be selected based on an above-minimum-distance criterion.
- Post-processing could provide a better selection of the sample units

The approach is illustrated for IT4 and the Wales sites, where the same but also an alternative are also proposed. For IT4 site, the distribution of two classes is shown in Figure 3.1.2. The classes are coded as 16 in Table 1 (A11/A1.B1.C1.D1.W8.A7.A9.B3) and 12 (A24/A2.A6.A12.B4.C2.E5.B11.E7) . Their shape file are evidenced in green red, respectively, when overlain onto the PANSH Quickbird image (Figure 3.1.2). These two classes are fragmented with segments located at the opposite corners of the study area and parallel to the coast respectively. For class 12, SRPRG was carried out within its MER area. The area was divided into 30 equal sampling areal units (SSUs), within which an equal number of sampling units were randomly located. According to Table 3.1.3, 148 samples for SSSc 12 are needed. The number of sample units randomly overlain for each areal sampling units is a multiple (10 times) of SSSc/30 = 5, because the final sum of the sample units should be points SSSc=148 and a large number of points are needed to cover the segments. Figure 3.1.3 illustrates the two classes considered but only the MER of class 12 is shown. Due to the peculiar distribution of class 12, only the sample units area intersecting class segments were covered by an equal number of reference sampling units. Figure 3.1.4(a) shows the samples units (i.e., 50 per each areal unit) located only on three areal sampling units intersecting the class segments and the zoomed areal unit located at line 3, column 3 of the MER grid in the upper right part of the image. Once a sufficient number of sample units are located, the sampling units should be selected for being independent and identically distributed (i.i.d).

Figure 3.1.2 - IT4. PANSH Quickbird image, RGB=3,4,1. The shape file of the LC classes is evidenced in white. In red color class 12 and green color class 16 are evidenced, respectively.

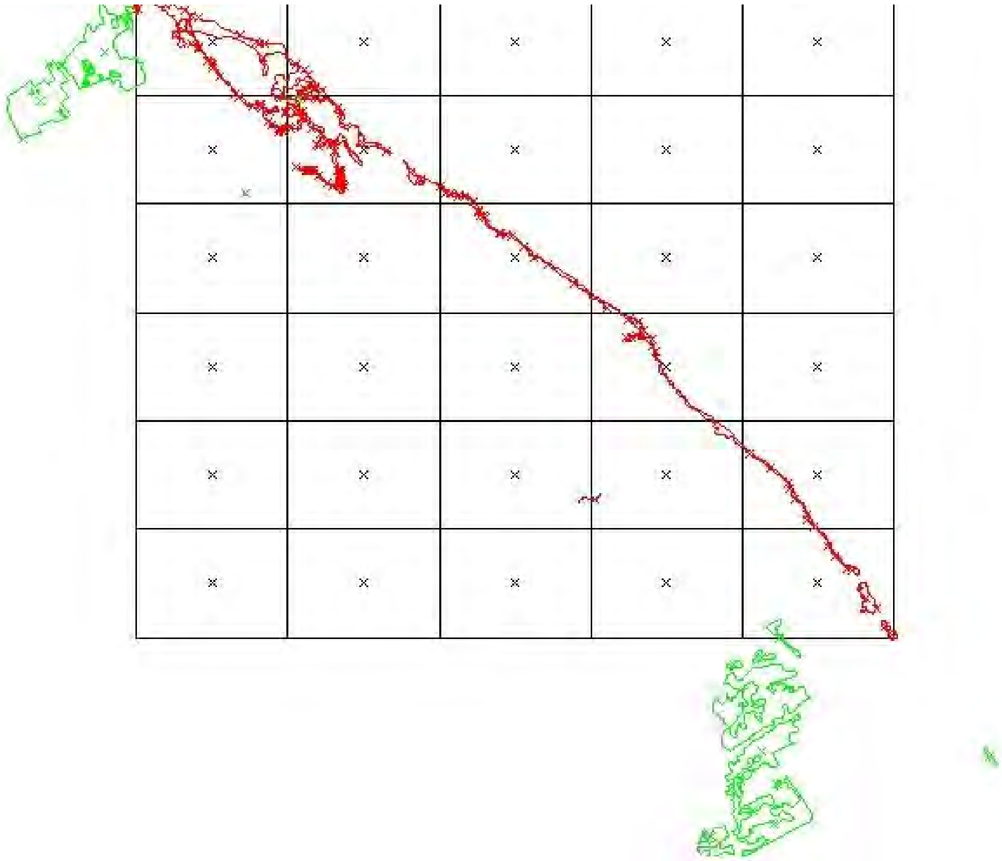

Figure 3.1.3 - The shape file of the classes 12 and 16 area shown in red and green respectively, but only the MER for class 12 is shown.

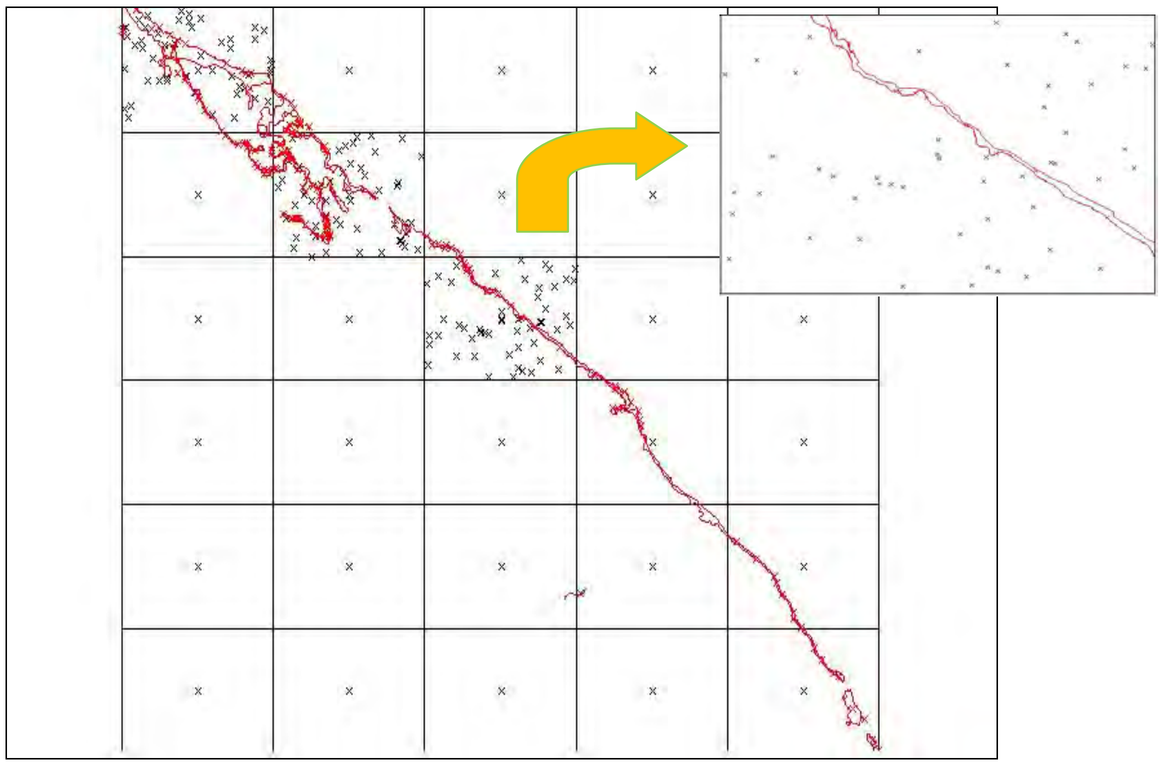

Figure 3.1.4 - Shows the samples units (i.e. 50 per each areal unit) located only on three areal sampling units intersecting the class segments and the zoomed areal unit located at line 3, column 3 of the MER grid in the upper right part of the image.

An alternative and less time-consuming approach is as follows:

- Consider the whole image and distribute a very high number of sampling units randomly based on a distance criteria. Figure 5 (a) shows the full PANSH Quickbird image and a first set of 10.000 points randomly located on the IT4 area (the white part of the figure). The points have a counter. Figure 5(b) shows in blue the sampling point of some segments of class 16
- Stratify the points per class, as done in Figure 6(a) and Figure 6(b) for class 12 and class 16.
- If the number of sample points for class c is lower than SSSc, new sample units are located randomly on the full image. If the number is higher than SSSc, the points can be selected according to their counter code (consider the points first located on the image)
- Once SSSc points are extracted and the i.d.d. condition is satisfied, then the reference point can be labelled according to Section 1.2

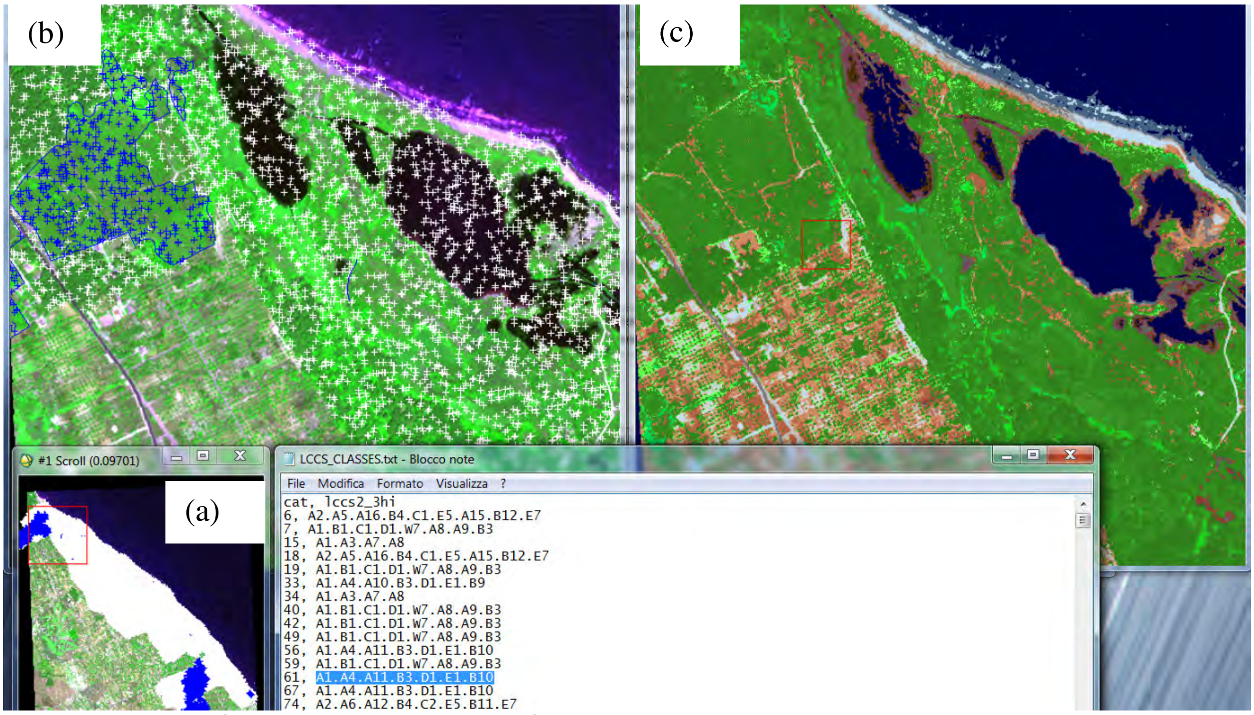

Figure 3.1.5 - (a) Quickbird 2005 image, RGB=341 with the whole set of randomly sampled points on the area of interest. In blue class 16 sample points; (b) a segment of class 16; (c)The SIAM™ 1st stage output map.

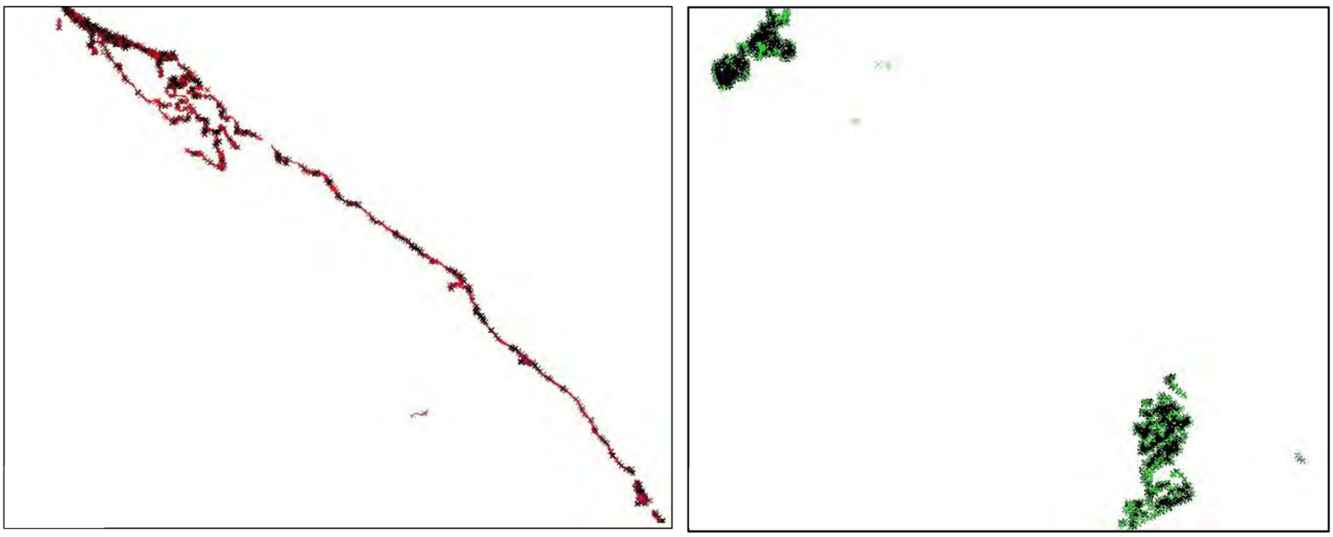

Figure 3.1.6 - (a) Reference sample points for class 12 and (b) reference sample points for class 16 to be refined according to the i.i.d. criteria and then labelled.

3.1.2.4 Stratified sampling strategy within a regular grid (SRSRG)

Sampling units for the LCCS thematic class validation will be either: *i)* a point unit (pixel) for classes not characterized by geometric/textural properties (such as water or shadow), or *ii)* an areal unit (i.e., a polygon or rectangular/square area) for classes which area characterized by geometric/textural properties. See Table 31.3 (Column 5). The criteria considered are as follows:

- For classes for which textural properties are important, a polygon ≥ 2 x texture period is needed. In the case of olive groves (LCCS= A11.A1.B1.C1.D1.W8.A7.A9.B3), the polygon size will be 36 * 36 pixels (each being 1 m in size).
- For classes for which geometric properties are important, a polygon that reproduces the shape of the target object in the (2-D) image domain (e.g., a building) is needed.

3.1.3 Response design protocol to label the reference sampling units for LC/LU validation

3.1.3.1 The evaluation protocol

To assign thematic labels to the reference sample,different independent source of information will be used. These, including LC/LU maps (e.g., regional maps) or ortho-photos, as well as prior knowledge of the target sampling units on the ground by domain experts, VHR EO images and ground visit as reported in Table 3.1.3 (Column 4). The size and shape of the spatial support regions for each strata are reported in Table 3.1.3 (Column 5).

As evidenced in PART 1 (Section 2.3.6.1), the combination of these sources of information should provide the best compromise between conflicting objectives, be feasible in practical terms and consistent in time and space, and respect the budget constraints. This combination of evidence stemming from multiple sources occurs within a class-specific spatial support region of selected size and shape with c = 1, ..., NC = 19, such that spatial support region c = 1, ..., NC = 19, is ≥ than the spatial type of the sampling units for class c = 1, ..., NC = 19.

For IT4, VHR image photo-interpretation and pre-existing information will be considered for both selecting and labelling sampling units on the VHR image to be classified. A pre-existing LC/LU map is available at a scale of 1:5000, with this developed in the framework of the cited Interreg project and with an OA of 95 % and an error tolerance of 2%. The map was produced by photo-interpretation of a 2006 ortho-photo and in-field campaigns undertaken in 2007-2008. Two Quickbird images acquired on in July 2005 and July 2009 and a WorldWiew2 image acquired in October 2010 are, at the time of writing, available for IT4. Even though the scale of the pre-existing map (i.e., 1:5,000) is compatible with the spatial resolution of the three images and no reduction of the number NC of classes is needed, there are differences in the dates of acquisition and map production. For this reason, and before assigning label to reference sample unit on the base of the pre-existing map, there is a need to verify that each strata (class) of the pre-existing map (and its label) is visible on the VHR images to be classified. For this purpose, the preliminary SIAM™ spectral output map from the 1st EODHaM stage (already produced for each of the three VHR images) should be used for selecting and excluding strata in the pre-existing map if these cannot be discriminated initially in the VHR image. As described in D3.1, the SIAM™ -derived map can provide non-overlapping semantic strata (e.g., barren land or built up, vegetation, water, etc.) that could assist in identifying strata (class) changes between the pre-existing map and the actual VHR images and/or errors. As an example, in the QUICKBIRD image acquired in 2005 some strata (e.g., rivers covered by vegetation) evident in the pre-existing map were not visible in the July image and/or were erroneously labelled when compared to with the Quickbird pre-classified SIAM™ map (e.g., classified as soil instead of vegetation). Furthermore, the areas of some habitats had changed due to fires in 2005 or flooding associated with the construction of a golf course had occurred in 2009. Quickbird. In conclusion, when photo-interpretation of VHR image and pre-existing maps are used, *no-change areas* should be used for assigning labels to reference sample unites. Such work should be undertaken by a co-operating team of EO data processing experts, domain experts and users (e.g., Management authorities of the Natura 2000 site. In-field campaigns should be undertaken to confirm the labelling of reference sample for field validation according to budget availability, time (and seasonal) constraints and site accessibility. The size and shape of the spatial support region for each strata are reported in Table 3.1.3 (Column 5).

3.1.3.2 The labelling protocol

LCCS labels assigned to sampling units selected across the IT4 test site are reported in Table 3.1.3 (Column 6).

3.1.4 Estimation and analysis protocol

The degree of match between the thematic map, in this case an LCCS map, once available and the reference sampling units will be based on an overlapping area matrix (OAMTRX). The thematic map accuracy should be estimated by a variety of Quality Indicators (QIs) (Baraldi et.al., 2005) provided with a degree of uncertainty measurements in line with QA4EO guidelines [GEO/CEOS, 2008, see References in PART1].

Table 3.1.3 - Reference sample points numbers (SSSc) to be collected for each of the LC/LU 19 classes(column 3) and their corrensponding support region on the ground (column 5). Independent source of information for the labelling o SSSc (column 4). Type of label to be used (column 6).

Ci	***LCCS Description***	***Estimated sample set cardinality SSSc with c= 1….NC=19;*** *α= 0.10; χ^2=7.9*			***Independent source of information:*** *1) prior knowledge and Photo-interpretation of:* 1.a VHR winter and summer images 2.a pre-existing LC map (Interreg Project); *2) in-field inspection*	***Size and shape of the support region on the ground ≥texture of the sampling unit (if any)***	***Labelling protocol (crisp or fuzzy label)***
		Target P_{OA}	Target δ	***SSSc***			
	A12 Natural and seminatural terrestrial vegetation						
1	**2**	**3**			**4**	**5**	**6**
1	Closed annual medium/tall forbs				1) and 2)	Areal 10*10 m	Crisp
2	Broadleaved deciduous medium/high closed shrubland (thickets)				1)	14*14 m. ~200m^2 (23*23 pixels) at 0.6 m.) or a multiple	Crisp
3	Needleaved evergreen medium/high closed shrubland (thickets)				1)	14*14 m. ~200m^2 (23*23 pixels) at 0.6 m.) or a multiple	Crisp
4	Broadleaved evergreen open dwarf shrublands				1)	7*7 m (11*11 pixels) or a multiple	Crisp
5	Broadleaved evergreen Medium/high closed shrubland (thickets)	**85%**	**10**	**99**	1) and 2)	7*7 m (11*11 pixels) or a multiple	Crisp
6	Open annual short forbs				1) and 2		Crisp
7	Open perennial medium-tall grasslands				1)	4m*4 m ~ 16m^2 (7*7 pixels) or a multiple	Crisp
8	Closed perennial tall grassland				1) and 2)	4m*4 m ~ 16m^2 (7*7 pixels) or a multiple	Crisp

Table 3.1.3 - (continued)- Reference sample points numbers (SSSc) to be collected for each of the LC/LU 19 classes(column 3) and their cornesponding support region on the ground (column 5). Independent source of information for the labelling o SSSc (column 4). Type of label to be used (column 6).

	A24 Natural and seminatural aquatic or regularly flooded vegetation				***Independent source of information:***	***Size and shape***	***Labelling protocol***
9	Open annual short herbaceous vegetation on temporarily flooded land				1) and 2)	4m*4 m ~ 16m^2 (7*7 pixels) or a multiple	Crisp
10	Aphyllous closed dwarf shrubs on temporarily flooded land				1) and 2)	4m*4 m ~ 16m^2 (7*7 pixels) or a multiple	Crisp
10	Perennial closed tall grasslands on temporarily flooded land	**95%**	**5%**	**148**	1) and 2)	4m*4 m ~ 16m^2 (7*7 pixels) or a multiple	Crisp
12	Perennial closed medium-tall grasslands on temporarily flooded land				1) and 2)	4m*4 m ~ 16m^2 (7*7 pixels) or a multiple	Crisp
13	Perennial sparse medium tall herbaceous vegetation on permanently flooded land				1)	(7*7 pixels) or a multiple	Crisp
	A11 Cultivated and Managed areas						
14	Herbaceous Crops				1)	4m*4 m ~ 16m^2 (7*7 pixels) or a multiple	Crisp label
15	Monoculture fields of rainfed broadleaved evergreen tree crops orchards (olive groves)				1)	24m*24 m (=max period of 12m*2periods) (40*40pixels)	Crisp label
16	Monoculture fields of rainfed needleleaved evergreen tree crops plantations	**85%**	**10**	**99**	1)	14*14 m. ~200m^2 (23*23 pixels) at 0.6 m.) or a multiple	Crisp label
17	Scattered industrial or other areas				1)	Areal; polygons extended to the shape of industrial areas	Crisp label
18	Paved roads				1)	Areal; polygons	Crisp label
19	See water				1)	Point features	Crisp label

3.2 Field sampling in support of GHC classification, Wales

3.2.1 Overview

The primary purpose of the field campaigns is to provide a discrete and finite ground truth dataset, collected through a probability stratified sampling strategy, to validate the LCCS maps derived from the SIAM™ first stage spectral classes and correspondingly the GHC maps translated from these. This approach allows the accuracy and uncertainty (error tolerance) in the classifications to be quantified. The ground truth data will also be supplemented using classifications and prior knowledge based on previous field campaigns, combinations of LiDAR and hyperspectral data, and aerial photography. The ground truth dataset will also be used in a statistically rigorous quality assessment of the maps generated with the target one-class overall accuracy probability (pOA) $\in$ [0, 1] ± error tolerance (δ) being at least 0.85 ± 2% and no less than 0.7 ± 5 %.

The validation needs to be undertaken in a logical sequence as follows:

a) Existing information (e.g., land cover maps) are used to initially stratify the landscape for subsequent sampling, with these converted to LCCS classes.
b) Ground data from field survey or other sources (e.g., aerial imagery) are acquired.
c) These data are used to validate LCCS maps generated independently from SIAM™ spectral categories.
d) The validated SIAM™ -derived maps are used to update the previous reference maps to generate a reference product for the site (e.g., for 2011).
e) These same maps are then used to stratify the landscape into, for example, 1 km cells which are then sampled for GHCs using standard protocols.
f) The field data and maps depicting GHCs (for selected 1 km cells) are then used to validate the maps of GHCs translated from SIAM™ -derived LCCS maps.
g) The field plots for both LCCS and GHC validation are used subsequently to validate the detection of change.

An overview of the approach is given in the following sections and above.

3.2.2 Selection of landscape area

The primary Natura 2000 site (Cors Fochno and the Dyfi Estuary) is impacted upon by processes and events occurring within the Dyfi catchment. As the system is complex and influenced significantly by hydrology, a uniform buffer zone (e.g., 3 km, as proposed for other sites) was considered inappropriate as the sole area for monitoring the impacts of human activities and natural processes. Nevertheless, changes observed in closer proximity to the borders with the Natura 2000 sites need to be highlighted as part of a monitoring system, particularly as these are subject to recreational pressures but are also being restored (e.g., by the Countryside Council for Wales; CCW).

Within the Dyfi catchment, 26 LCCS classes have been identified and rules for translating these to GHCs developed. These classes are listed in D5.1 and Table 3.2.1. Mapping of the majority of these classes from SIAM™ spectral categories associated with Landsat and SPOT sensor data has already been undertaken using a combination of locational (spatial) relationships, spectral properties, geometric and shape properties, morphological properties and links with ancillary data (e.g., elevation, slope, tidal regimes). Examples of the maps are presented in D5.1, and the LCCS map derived from Landsat sensor data is provided in Figure 3.2.1.

Table 3.2.1 - LCCS classes (Life Forms) occurring within The Natura 2000 site and the Dyfi catchment. The corresponding GHCs are listed in D5.1

Category	LCCS Code_Modifier	Description
A11	A4.B1.B5.C1.D1.D9_B4	Permanently cropped area: Graminoid crops
A11	A1.B1.B5.C1.D1.D9_A8.B4	Permanently cropped area with rainfed needleleaved tree crops (plantations).
A11	A1.B1.B5.C1.D1.D9_A7.B4	Permanently cropped area with rainfed broadleaved tree crops (plantations).
A12	A3.A10.B2.C2.D1.E2	Broadleaved deciduous fragmented high trees
A12	A4_A11_B3_C2_D1_E2__B14	Broadleaved deciduous medium to high shrubland
A12	A4.A11.B3.C2.D1.E1	Broadleaved Evergreen Fragmented Shrubland single layer.Heathland (uplands)
A12	A6.A10.B4.C1.E5_B12.E6	Closed Perennial Medium Tall Grassland (e.g., *Molinia/Juncus*)
A12	A6.A11.B4.XX.E5_A12.B12.E6	Open ((70-60)-40 %) Perennial Medium Tall Grassland (e.g., *Eriophorum*)
A12	A6.A10.B4.C2.E5_B13	Closed short grassland
A12	A5.A10.B4_B11	Closed medium tall forbs (3.0-0.8 m)
A12	A5.A10.B4_B12	Closed medium tall forbs (0.8-0.3 m)
A24	A4.A20.B3.C1.D1.E1.F2. F4.F7.G4_C4	Closed to Open Broadleaved Evergreen Shrubs with Herbaceous Vegetation on Permanently Flooded Land (Persistent) (Active Bog)
A24	A6.A12.B4.C1.E5_B11.C4.E6	Perennial closed tall grassland on permanently flooded land (persistent)
A24	A6.A13.B4.C1_B13.C5	Open short grassland on permanently flooded land (with daily variations) (*Unmanaged Saltmarsh*)
B15	A3_A8	Paved road(s)
B15	A3_A10	Railway(s)
B15	A4_A13	Urban areas
B16	A3_A7	Bare rock
B16	A6.B6	Shifting Sands.Saturated Parabolic Dunes
B16	A6_A12	Stony loose and shifting sands
B16	A5_A13	Very stony bare soil and unconsolidated material(s)
B27	A1.B1.C1_A5	Clear shallow artificial waterbody (Standing)
B27	A1.B1.C1_A4	Turbid Deep to Medium Deep Artificial Perennial waterbodies (Flowing)
B27	A1.B1.C1_A5	Deep to Medium Perennial Artificial Waterbodies (Standing)
B28	A1.B2.C1_A5	Deep to Medium Perennial Natural Waterbodies (Standing)
B28	A1.B3_A4.B6	Tidal Area (Flowing); Surface Aspect (sand)
B28	A1_A4	Natural waterbodies, flowing (ocean/sea)

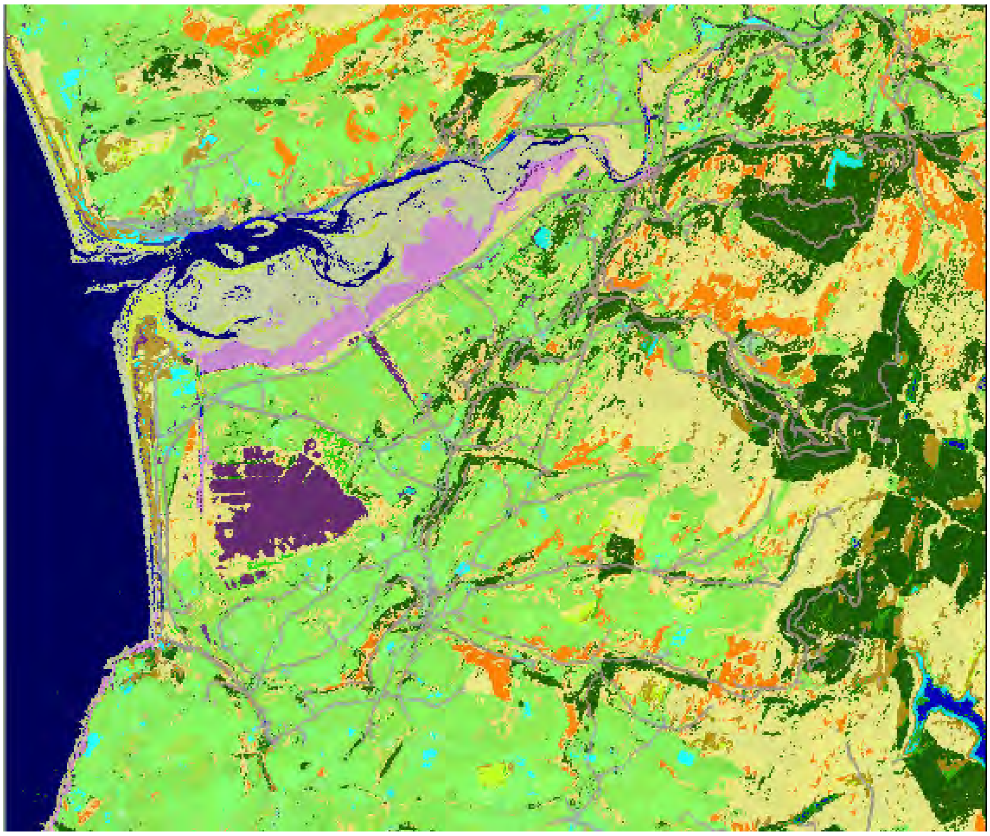

Figure 3.2.1 - Classifications of LCCS based on Landsat TM data from July, 2006. LCCS classes not able to be classified are in white in the legend below.

LCCS		LCCS Code_Modifier	Description
A11		A4.B1.B5.C1.D1.D9_B4	Permanently cropped area: Graminoid crops
A11		A1.B1.B5.C1.D1.D9_A8.B4-W7	Permanently cropped needleleaved trees
A11		A1.B1.B5.C1.D1.D9_A7.B4-W7	Permanently cropped broadleaved trees
A12		A3.A10.B2.C2.D1.E2_B5	Broadleaved deciduous trees
A12		A4.A11.B3.C2.D1.E2_B14	Broadleaved shrubland
A12		A4.A11.B3.C2.D1.E1	Broadleaved evergreen shrubland (heath)
A12		A6.A10.B4.C1.E5_B12.E6	Closed perennial medium tall grassland
A12		A6.A11.B4.XX.E5_A12.B12.E6	Open medium tall grassland
A12		A6.A10.B4.C2.E5_B13	Closed short grassland
A12		A5.A10.B4_B12/B13	Closed medium tall forbs (3.0-0.8/0.8-0.3 m)
A24		A4.A20.B3.C1.D1.E1.F2. F4.F7.G4_C4	Broadleaved evergreen shrubs flooded (bog)
A24		A6.A12.B4.C1.E5_B11.C4.E6	Perennial closed tall grassland on permanently flooded land (persistent)
A24		A6.A13.B4.C1_B13.C5	Open short grassland (saltmarsh)
B15		A3_A8	Paved road(s)
B15		A3_A10	Railway(s)
B15		A4_A13	Urban areas
B16		A3.A7	Bare rock
B16		A6.B6	Shifting Sands.Saturated Parabolic Dunes
B16		A6_A12	Stony loose and shifting sands
B16		A5_A13	Very stony and unconsolidated material(s)
B27		A1.B1.C1.D1_A5	Clear shallow artificial waterbody (standing)
B27		A1.B1.C1_A4	Turbid artificial waterbody(flowing)
B27		A1.B1.C1_A5	Deep/medium artificial waterbody (standing)
B28		A1.B2.C1_A5	Deep/medium natural waterbody (standing)
B28		A1.B3_A4.B6	Tidal area (flowing); surface aspect (sand)
B28		A1_A4	Natural waterbodies, flowing (ocean/sea)

3.2.3 Implementation of sampling in support of LCCS maps

The Landsat and SPOT-derived maps of LCCS for the Dyfi catchment are preliminary and will be updated following acquisition of very high resolution (VHR) satellite sensor data (successful acquisition of the lower Dyfi catchment on the 27th July, 2011, although at a poor viewing angle of > 25°). The accuracy of these maps is to be validated following collection of field data between mid July and early September. Furthermore, the classification is anticipated to change following the development of new rules. For these reasons, the maps were not yet considered sufficiently reliable for developing the initial stratified sampling strategy. For these reasons, and as demonstration, the initial stratification is based on the Phase 1 Habitat Survey maps which have been translated to LCCS categories and represent the most detailed level of mapping for the entire catchment. The use of an existing thematic map is not however generally advocated, particularly if historical, as changes in the landscape can lead to errors but is necessary in this instance. Therefore, in future work, changes to the Phase 1 classification determined through SIAM™ classification of VHR data will be undertaken to provide an updated reference map against which to assess change.

For the Dyfi catchment, the area of each LCCS class has been approximated from the Phase 1 Survey (Figure 3.2.2). Some classes (e.g., artificial waterbodies) were not distinguished within this Survey and hence only 20 out of the 26 present were translated.

Following the REDD sampling protocol (Maniatis and Mollicone, 2010), these LCCS categories should be divided into 25-30 area equivalent sample selection units. This is equivalent to stratified random sampling within a regular grid (SRSRG). These SSUs can be of varying shape, including grid cells, hexagons or variations on these. The size of the SSUs is determined from, for example, a) the actual area occupied by the LCCS class (e.g., as determined from previous mapping) or b) the non-orientated Minimum Enclosing Rectangle (MER). Within these SSUs, a single sample point could be located randomly, giving a total of 25-30 points for sampling. However, to achieve a target one-class overall accuracy probability (pOA) $\in$ [0, 1] ± error tolerance (δ) of at least 0.85 ± 2% (as an example), a larger number of samples are needed (e.g., 2114 based on the criteria above and with 5 classes). For this reason, random sampling within each of the 25-30 SSUs needs to be undertaken, with this utilising an underlying systematic grid (e.g., of 10, 50 or 100 m). Within this grid, sample cells can be selected randomly (based on the number desired to achieve the desired accuracy and error tolerance) and within these, points randomly located for sampling (if < 10 m). In each case, a distance threshold can be applied to avoid autocorrelation. The grid cell size can be variable depending upon the extent of the LCCS class considered.

As an example, consider the area of the active bog at Cors Fochno, which is associated with the LCCS class A4.A20.B3.C1.D1.E1.F2.F4.F7.G4_C4 (Figure 3.2.3). Here, 30 equal SSUs (of 10 ha) have been identified (a). Note that only those occupying whole 1 ha units (see grid) are illustrated but these would ordinarily include those < 1 ha (pale blue). Each of the 30 units is merged in b) and, in c), one sample point has then be located randomly in each. However, to obtain more samples as required to obtain acceptable accuracy and error tolerance (e.g., 2114 based on 5 classes), a nominal 10 m grid has been overlain (d). The number of samples required is then randomly selected from the grid (70 cells for each of the 30 area equivalent units). Within each of the selected 10 m grid cells, a point has been randomly selected using a distance threshold (in this case, 50 m to those in other cells) to avoid autocorrelation. The same would be repeated for the remaining 4 LCCS categories. For field sampling, a random selection is taken (depending on resources) and the remainder are then sampled based on existing data sources including, in the case of the Wales site, LiDAR and aerial photography.

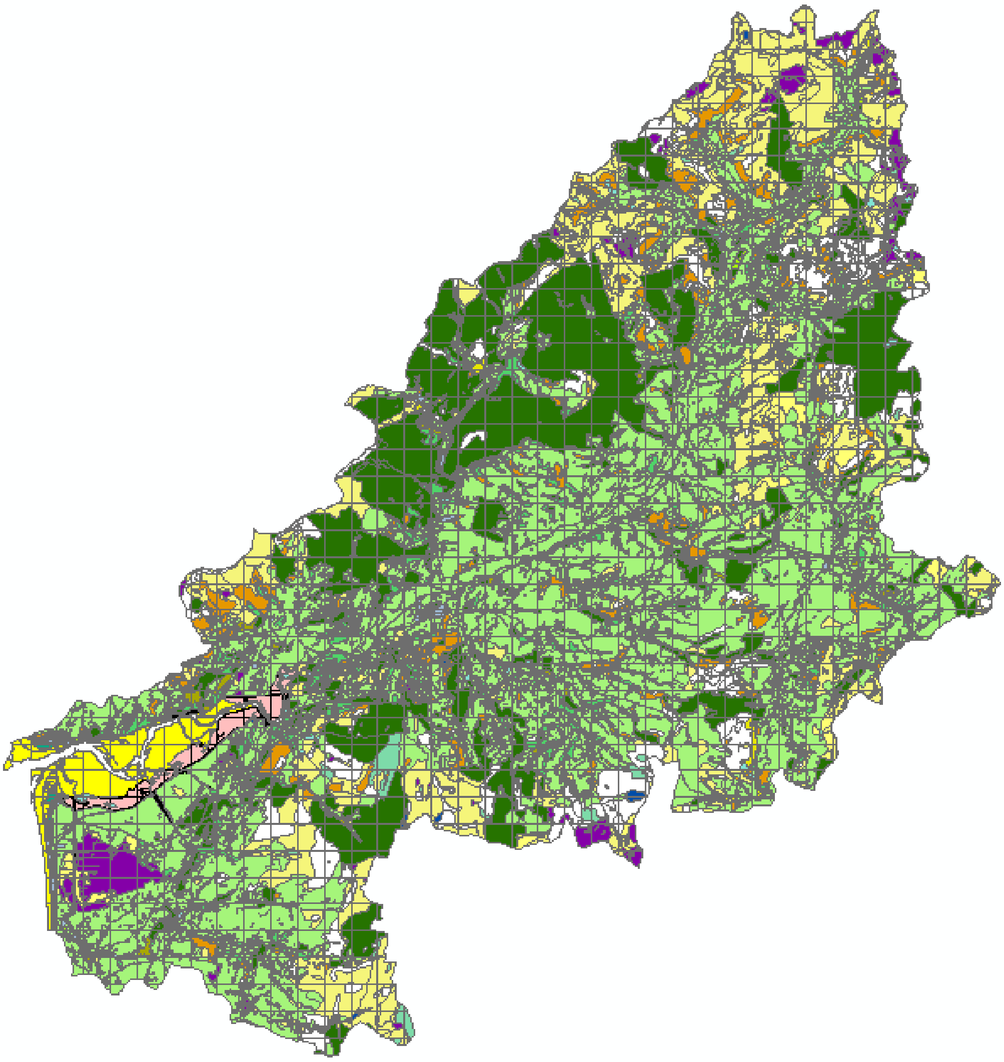

Figure 3.2.2 - LCCS classification derived from the existing Phase 1 Habitat Survey (see Figure 1 for legend).

Within the Dyfi catchment, a maximum of 20 plots per LCCS class (giving a total of 400 if 20 are considered) are deemed achievable through field sampling, noting that a proportion are non-vegetated and hence field sampling (based on observation only in the first instance) is expected to be relatively rapid. The remainder are to be sampled using other data sources. 6 of the LCCS classes were not able to be translated from the Phase 1 Survey, but an additional 80 plots would need to be located in these, giving a proposed total of 480.

An alternative and simpler approach to obtaining equivalent area sampling is to divide the MER into 25-30 SSUs and this is illustrated in Figure 3.2.4a-f. For LCCS occurring as contiguous features (e.g., tropical rainforest, as in the Brazilian site), the use of the MER is more appropriate. However, within fragmented landscapes, the MER will not be proportional to the area of coverage by each LCCS class. For example, the area of the MER will be far greater if the active bog is located in the upper and lower reaches of the catchment compared to if the bog was located only in the lower reaches. As illustration, Figure 3.2.5 shows the distribution of points derived for the active bog when using a) the Phase 1-derived LCCS map (as above), b) the MER but only including the previously mapped active bog and c) the MER encompassing the area of active bog. In each case, 2114 points are randomly located. The former approach is advocated as it can be applied regardless as to whether LCCS classes are contiguous or fragmented as the areas of the SSUs are directly proportional to the best estimate of the area occupied by the LCCS class.

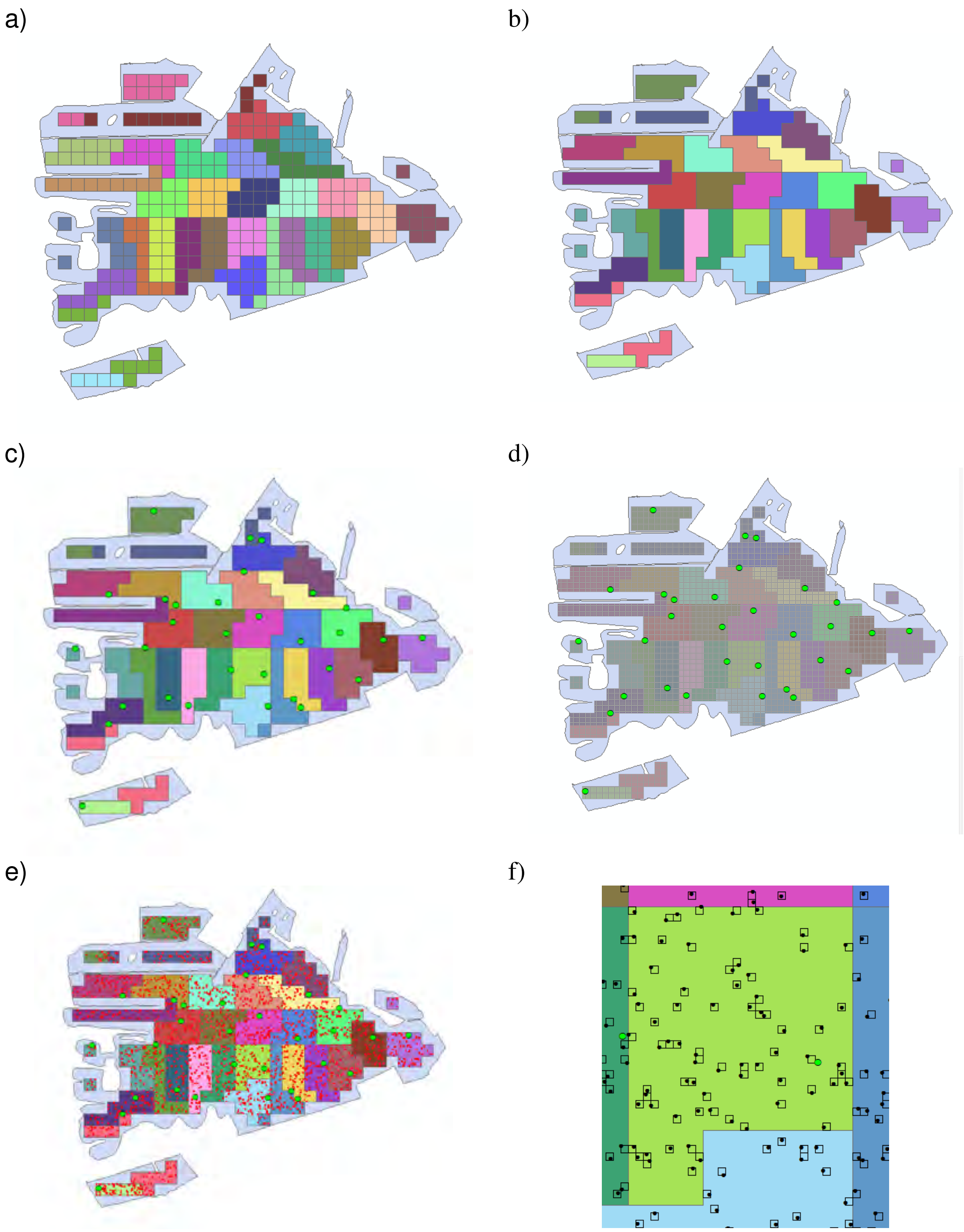

Figure 3.2.3a-f) - Proposed implementation of the REDD approach to the Wales BIOSOS site (example: active bog (A4.A20.B3.C1.D1.E1.F2.F4.F7.G4_C4). See text for explanation.

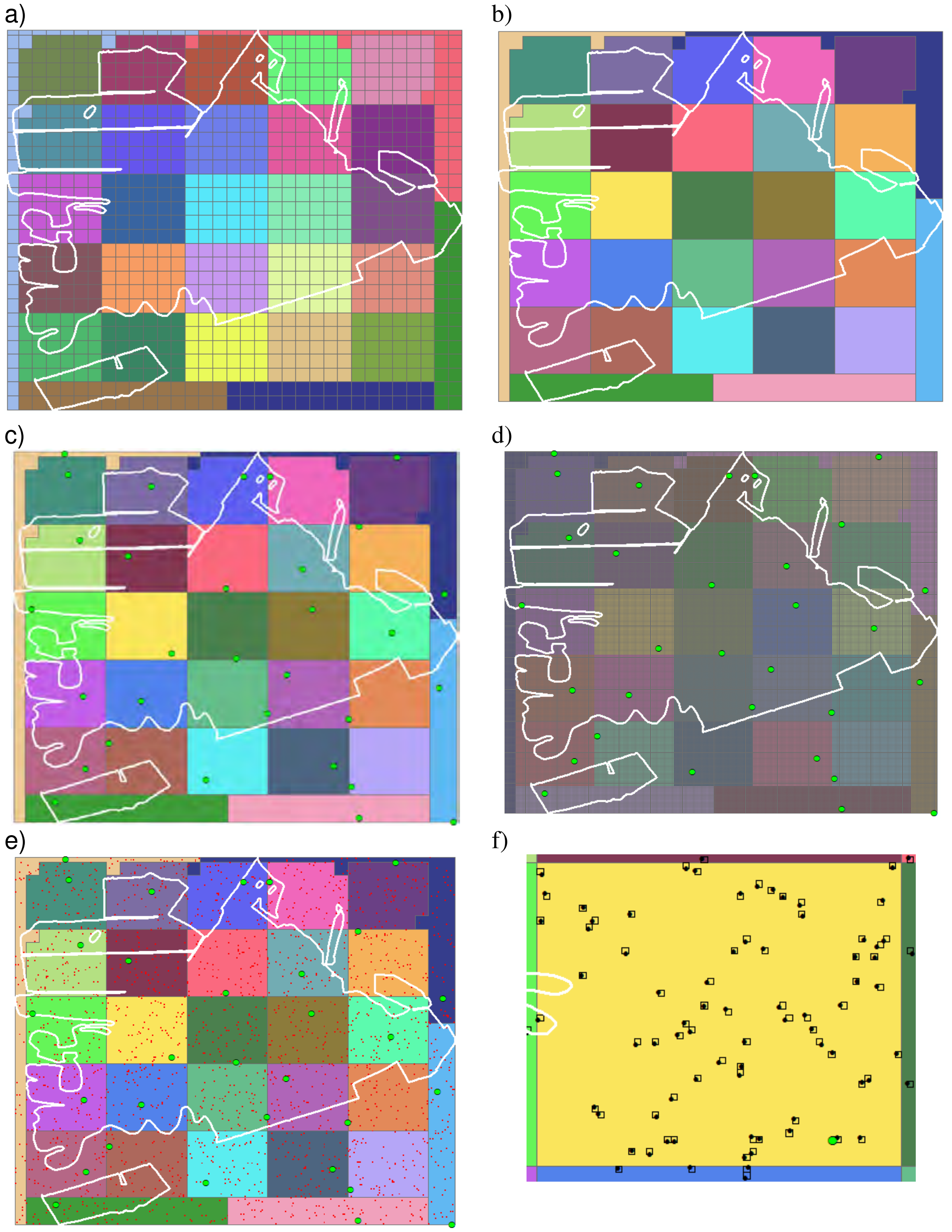

Figure 3.2.4a-f) - Proposed implementation of the REDD approach to the Wales BIOSOS site (example: active bog (A4.A20.B3.C1.D1.E1.F2.F4.F7.G4_C4). See text for explanation.

The size of the final sampling unit will also depend upon the area occupied by each LCCS class. For example, the core area of the active bog covers over 300 ha and hence cells in the underlying grid could be 37.8 x 37.8 m. However, if the active bog only covered an area of 3 ha, then the cell size could need to be 3.78 x 3.78 m to achieve allow 2100 samples to be taken. In each case, the sampled area would have to be lower to avoid autocorrelation effects. Hence, consideration needs to be given to the size of sample to the area to be sampled (i.e., the LCCS class).

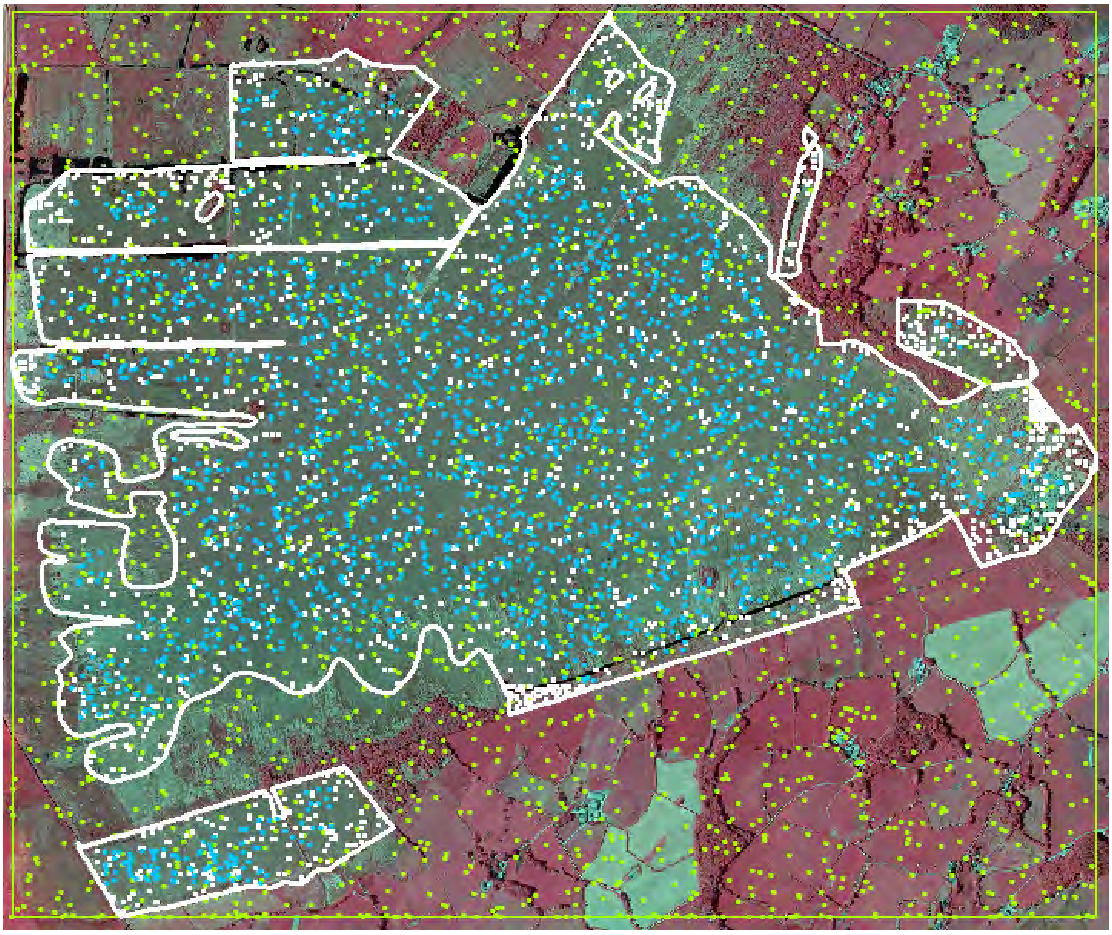

Figure 3.2 5 - Colour infrared aerial photograph of Cors Fochno active raised bog with random points generated from existing land cover mapping (blue) and the Minimum Enclosing Rectangle (MER) where the random selection of 2100 points has been taken from within the MER (green) or the raised bog only (white).

An alternative approach which is more applicable to fragmented landscapes is to establish how many samples would be required for each km^2 of the landscape occupied by an LCCS class given a desired accuracy and error tolerance. For the Dyfi catchment, 20 LCCS categories have been identified from the UK Phase 1 Habitat Survey map, with their areas varying from 0.1 km^2 to 271 km^2. A similar classification can be achieved with from the optical remote sensing data based on SIAM™ spectral classes, but needs to be undertaken following acquisition of suitable data and supportive field observations. Within each of these classes, an equal number of samples is needed, with this being 2511 based on a desired accuracy of 0.85, a half width of error tolerance of 0.02 and a level of significance of 0.1. To generate the sample areas, a 1 km grid is overlain which covers the area of the catchment. This grid is then intersected with the polygons representing each LCCS class, resulting in smaller polygons ranging in size up to 1 km^2. The 2511 samples are then distributed within these polygons, with the number in each varying as a function of their area. In Table 3.2.2, the number of samples per 1 km^2 is given, but this will be less for polygons which are smaller than this, as explained in the example below. This approach is similar to that adopted by REDD and MER but is not constrained by the requirement to initially divide the area into 25-30 equivalent area units.

The example relates to the extent of coniferous plantations (A1.B1.B5.C1.D1.D9_A8.B4-W7) within the Dyfi catchment, with this illustrated in Figure 3.1.6. The area of conifer plantation has first been divided into units not exceeding 1 km^2 in area. For each unit area, the number of points for survey has been determined and allocated. Referring to Table 3.2.2, and to conifer plantation, a maximum of 19 points can be inserted into a 1 km^2 area representing conifer forest and 8 points for a 0.5 km^2 area. However, to avoid autocorrelation, no more than one point should be placed in, for example, in an underlying grid (e.g., of 100 m x 100 m). Distance criteria can also be applied such that points within adjoining cells are sufficiently separated.

Table 3.2.2 - List of LCCS classes translated from Phase 1 and their area within the Dyfi catchment (km^2).

LCCS Class	Area (km^2)	No. samples per km^2
A1.B1.B5.C1.D1.D9_A7.B4-W7	2.3	1109
A1.B1.B5.C1.D1.D9_A8.B4-W7	130.5	19
A1.B1.C1_A4	0.8	2511
A1.B1.C1_A5	2.1	1213
A3.A10.B2.C2.D1.E2_B5	33.2	76
A3_A7	2.1	1217
A4.A11.B3.C2.D1.E1	42.9	59
A4.A20.B3.C1.D1.E1.F2.F4.F7.G4_C4	13.7	183
A4.B1.B5.C1.D1.D9_B4	270.8	9
A4.A11.B3.C1.D1.E2_.B14	1.3	1953
A4_A13	6.6	382
A5.A10.B4_B11/B12	33.5	75
A5_A13	0.1	2511
A6.A10.B4.C1.E5_B12.E6	16.7	150
A6.A10.B4.C2.E5_B13	90.0	28
A6.A11.B4.XX.E5_A12.B12.E6	0.5	2511
A6.A13.B4.C1_B13.C5	5.0	502
A6.B6	0.6	2511
A6_A12	12.0	210
A6.A12.B4.C1.E5_B11.C4.E6	1.8	1418
Unknown	6.3	399
TOTAL	672.7	

Following random allocation of the points, a next step is to randomly select 1 km^2 units for GHC survey from the set of 1 km^2 cells for each class. for GHC survey (see section 4.3). Within the 1 km^2, and following published protocol, polygons are delineated manually from aerial photography or field survey and a reconnaissance visit is made during which each mapped unit is assigned with a GHC category. Then, a 100 m grid is overlain and for each GHC category identified, only one polygon is randomly selected from the grid intersect points (e.g., PT1). The survey point is, however, located at the centre of the polygon (PT2). As an alternative, the polygon could be selected from the subset of the 2511 points distributed through the LCCS type but contained within the same 1 km^2 (PT3; coloured yellow in Figure 3.1.6c) or to the nearest intersection point (PT4) on the 100 m grid.

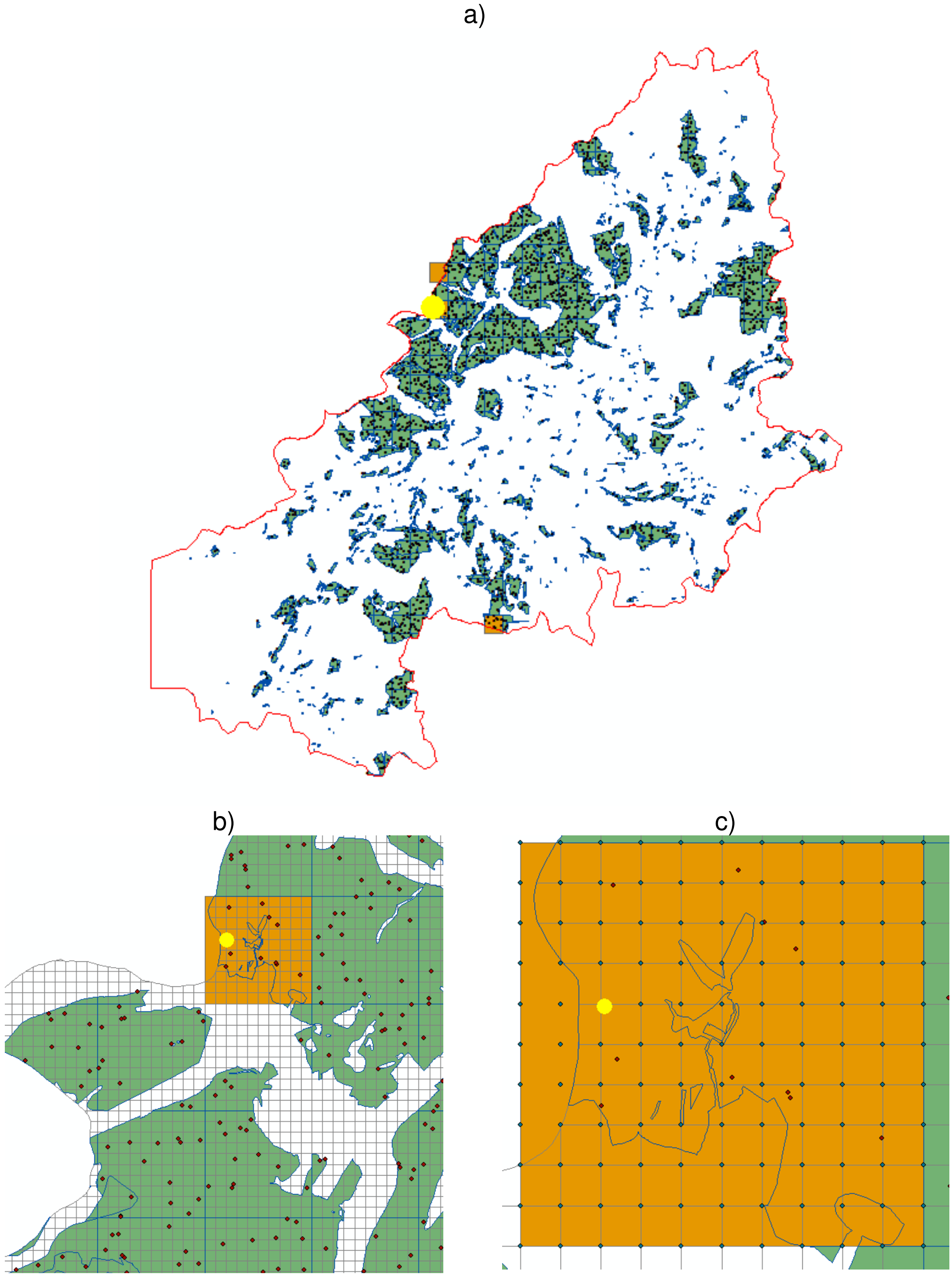

Figure 3.1.6 - The area of conifer forest within the Dyfi catchment, divided into units not exceeding 1km^2. b) Within each area unit, points are randomly located within an underlying 100 m grid (maximum of 1 per cell). c) For subsequent sampling for GHCs, one point is typically selected for field sampling within each 1 km^2 (orange) and for each GHC. Plots for field survey are located at the centroid of a polygon (defined through aerial photograph interpretation), with one polygon for each GHC selected randomly from those assigned to the same GHC type and which intersect an aligned grid of 100 m. This point could be selected as being the closest to the yellow point, determined randomly from the 2511 points or, instead, at the location of the yellow point itself.

4. PART 3

4.1 Introduction

For the purposes of WP4 and WP6, the "on site" campaigns will also include "in field" campaigns, and will be focussed on the collection of data and parameters for both biodiversity (flora, vegetation, fauna, soil) and pressures/threats indicators extraction according to D2.1. In particular, they aim at:

1. WP4 Task 4.4, WP6 Task 6.2 supplement existing data sets with data and community –both plant and animal– structure/composition as well as data on occurrence and abundance of individual species at both habitat and landscape levels ("biodiversity"), as detailed records at the community level are crucial for assessing habitat state and relative quality (Haines-Young *et al.* 2000) as well as vegetation dynamics;
2. WP6. Task 6.3 provide data for ecosystem state and functions assessment, as required to assess soil/vegetation interactions, soil physical degradation, stoniness;
3. WP6, Task 6.5 and Task 6.7 supplement existing data sets with data on fine spatio-temporal treats (e.g., soil degradation and land use and land abandonment), as required to demonstrate the adequacy of the BIO_SOS approach across a range of pressures in and around Natura 2000;
4. WP6 Task 6.2, Task 6.4 and Task 6.7 create a data set on landscape pattern at multiple scales (e.g., local, landscape) in order to explore their potential for predicting both local biodiversity attributes (e.g., species richness, abundance, diversity) according to ecological theory, and ongoing pressures and threats, as required for indicators extraction.

On site campaigns will be based on the LC/LU maps (classified according to the LCCS taxonomy, see D6.1) produced within WP5. This besides providing consistency among all sites, is required for validation purposes. Possibly non-alternative but complementary pre-existing thematic maps (section 2.1) can be used. Among these LCCS maps obtained by the reclassification of existing ancillary LC/LU maps according to the rules set in D6.1 and in D6.10. The use of such maps however, would imply solving the problem of changes in the labelling of reference sample from a previous map when compared with a new EO image (section 2.2.3.1). In order to assure internal consistency to the "on site/in field" activities which will be carried out by different research teams in different target training sites (UK, NL, IT3, IT4) as well as in all the remaining BIO_SOS study area described in D2.2, a common standard for both "sampling design" and "measurement protocols" has been agreed upon in both formal and informal discussions occurred so far among all partners of the BIO_SOS consortium, also based on the BIO_SOS Description of Work. This represents a general framework within which each team organizes its own work for the purposes of GHC field identification and collection on data on biodiversity and pressures, as well as for landscape pattern analysis.

The "sampling design" and "measurement protocol" are defined here according to EPA QA/G-5S (2002), respectively as the specification of "*the number, type, and location (spatial and/or temporal) of sampling units to be selected for measurement*", and "*a specific procedure for making observations or performing analyses to determine the characteristics of interest for each sampling unit*".

4.2 Sampling design

As far as the sampling design is concerned, while the more samples the better in order to reduce standard error (PART 1), however in most cases the number of samples is largely driven by the resources available for the survey in a given project. Therefore, particular attention has been paid to the optimization (cost/effectiveness) of the on-site/in field campaigns effort in the BIO_SOS sites, looking for a sound compromise between the efficient use of time, budgetary and human resources while ensuring the possibility of making statistical inferences and controlling for spatial autocorrelation when meeting WP4, WP5 and WP6 commitments. Moreover, the sampling design strategy adopted here is also intended to provide a basis for both surveillance and monitoring (Bunce *et al.* 2008), to integrate these activities with other data sources, and to integrate local recordings at the landscape level.

Even though site selection for BIO_SOS reflects a judgmental rather than a probabilistic criterion, so that not all the environmental zones of Europe, EnZ (Metzger *et al*. 2005; Jongman *et al.* 2006) are represented, nor are they randomly stratified, nevertheless, a probability-based sampling design will be adopted by all teams in each site to allow for statistical inferences about the sampled population when sampling for GHC identification, biodiversity and pressures.

The recommended probability-based sampling design would be a stratified (e.g. habitat/LCCS/GHC) random sampling within a cell of a regular grid (SRSRG). This is a compromise between a systematic two-dimensional sampling in space (Gilbert,1987), such as the central aligned square grid, and a random stratified sampling. This design combines the advantages of randomness with:

a) A greater precision of systematic sampling with respect to simple random sampling under circumstances when heterogeneity of sample units within the same sample population is to be expected (Cochran 1977)
b) Uniform representation of the target population in the set of samples.

The BIO_SOS project has to comply with the INSPIRE[1] (Infrastructure for Spatial Information in Europe, INSPIRE RDM PP v4-3 2002) initiative of the European Commission, which is intended to create a European spatial information infrastructure that delivers to the users integrated spatial information services. Moreover, the BIO_SOS project builds on BIOHAB (FP5) and EBONE (FP7). Within the framework of the INSPIRE initiative, a 1 km reference grid for Europe was produced which represents the spatial basis for the database holding information from all the 1 km squares in Europe produced within the EBONE project (EBONE D4.3). Such a database allowed for the identification of 13 environmental zones of Europe (EnZ) and 84 environmental strata of Europe (EnS) (Metzger *et al.* 2005; Jongman *et al.* 2006). By means of such a 1km grid associated database, the spatial distribution of any parameter available either from each km square (e.g., altitude, estimates of habitats extent) from the records made in the environmental strata and can be displayed at the continental level.

Although such grids and database integrate information from all over Europe, it has to be considered that they might not spatially coincide with national grids and databases. This means that there might be some tension in the short term by satisfying local end users and

1INSPIRE Directive:
http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_Specification_CRS_v3.1.pdf;
http://www.epsg-registry.org/)

having a common standard approach. On the other hand, such a tension might in the longer term be overcome provided that the INSPIRE initiative contributes a common basis for national grids. Therefore, it would be advisable, for the purposes of the BIO_SOS sampling activities (i.e., sampling design definition and sampling units selection) to make reference to the common standard grid (i.e., the 1km INSPIRE grid). For the purposes of further possible activities within the BIO_SOS project (e.g., up-scaling at the regional level -EnZ or EnS), a possible reference to the EBONE database might be advocated.

The area of interest for samples selection, the sample size and the size of sample units will be adequate to the specificities of each site, including their size and landscape complexity. The area of interest ought to include the protected sites themselves as well as surrounding area (i.e., the buffer zone). Land cover changes attributable to human driven harmful activities and disturbances towards the environment are likely to occur in the buffer zone (i.e., the protected/non protected fringe) leading to negative impacts on both the extent and quality of the habitats inside the protected sites. The size of the buffer zone is site specific. This can be a watershed or an area defined on the basis of a a fixed distance from site boundary. Whenever feasible in terms of available resources, paired samples (inside and outside) would be the best option as it would enable an assessment of the success of protection to be undertaken.

For smaller and/or more complex sites, sampling will be carried out with consideration given to the entire area and the number of small sample units maximised (1/4, 1/16 km^2). For larger sites or whole watersheds, sampling will be limited to a representative portion of the entire area (e.g., the extent of the EO image(s) used to produce the LCCS/GHG preliminary map). In addition, a range of sample sizes can be adopted for the different aims of the data collection having its counterpart in different (nested) sample units sizes. In particular, as indicated in the DoW (WT3-WP6), a ranked set sampling design (Gilbert,1987; Mode et al 1999, EPA QA/G-5S 2002, Wolfe 2010) can be introduced in order to improve the representativeness of the sample population while reducing sample size to compensate for the increase in sampling intensity.

The stratification will be site dependent. In the following, details for each sites are provided, where already possible including the identification of the thematic map legend (taxonomy, ontology), with spatial (e.g., inclusion) and non-spatial (e.g., subset-of) relationships between geospatial thematic classes and the target geospatial population(s) or class(es). This is equivalent to information layer(s) or stratum (strata) located in the area or region represented by the thematic map.

However, for particular activities (i.e. those within task 6.3, of WP6 relevant to soil characteristics a non probability sampling will be carried out, The following considerations are brought about: soil conditions can be favourable or adverse to species of fungi, plants, animals. In a given protected area where the habitat for a given species is present, soil characteristics are usually indifferent of favourable to the given species and to the elements that characterize the species habitat. This might not be so when anthropic action changes soil characteristics drastically. This occurs when land levelling is involved (as for urban and infrastructure works, land use change, vineyard's ex- or im-plantation: removal of ephemeral gullies and surface landslides) or when removal of rock fragments, boulders or petro-calclcic horizons is involved (as in the Murge). In these cases a characterization of the resulting anthropic soil is needed to explore its effects on the species of interests.

4.3 Stratified sampling for GHC in UK sites

4.3.1 GHC Survey

Within the Dyfi catchment, and to provide the basic ground truth dataset to validate the translation of GHCs from LCCS, a survey of GHCs is being undertaken from mid-July until early September, 2011, with this followed by a more detailed quadrat-based botanical and faunal survey in conjunction with Countryside Council for Wales (CCW) in September. The collection of these field data will support the more detailed information on the composition of flora and fauna required in WP6. Details of the proposed sampling approach are given in the following sections and the approach is outlined in Figure 4.3.1.

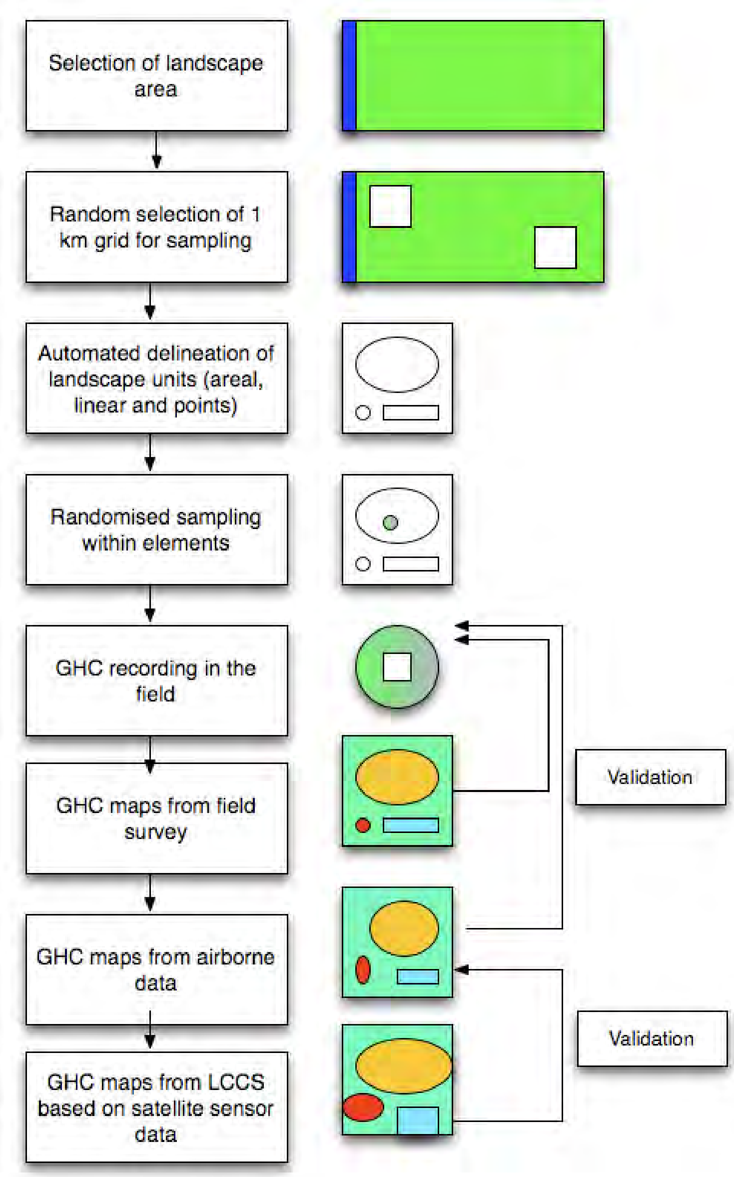

Figure 4.3.1 - Overview of the approach to GHC sampling and validation of output maps.

4.3.2 Selection of landscape area

As in Section 1, sampling of GHCs will occur throughout the Dyfi catchment, ensuring that all categories are considered.

4.3.3 Random sampling of 1 km grids within the Dyfi catchment

To direct the sampling of GHCs, selection of 1 x 1 km cells from a uniform grid has been advocated. Ideally, this grid should be linked to the existing BIOHAB stratification system. The BIOHAB project provided an environmental stratification of Europe and based on a statistical analysis of climate and topographic data, 84 environmental strata (EnS) linked to 13 zones (EnZ) have been defined. The classification was intended to provide the equivalent of 10,000 km^2 area that would be required for statistically robust survey and monitoring of GHCs. Whilst it is anticipated that the EnS and EnZ classifications will be referenced within the Dyfi catchment, a proposed 1 km grid (aligning with the British National Grid) has been generated and is presented in Figure 4.3.2.

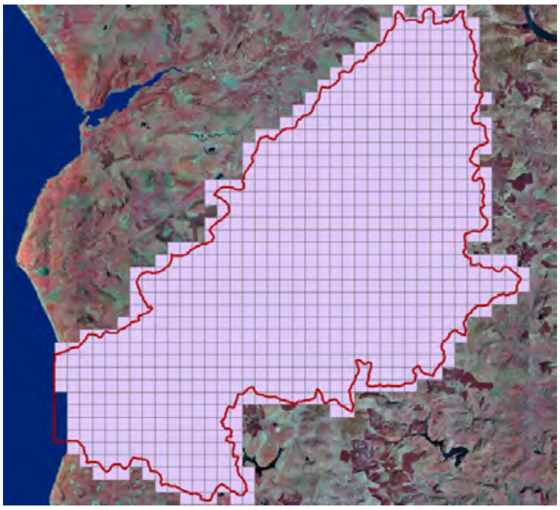

Figure 4.3.2 - 1 km grid cells contained with the Dyfi catchment, with these aligned with the British National Grid.

Simple random sampling (without stratification) within the grid is not advocated, as some thematic classes may not be represented. This is especially the case in the Dyfi catchment where, based on the UK Phase 1 Habitat Survey, grassland/marsh and woodland/shrub occupy 56.7 % and 33.9 % of the land area respectively. Whilst a diversity of other habitats occurs, these collectively only occupy 9.4 % of the catchment.

In this exercise, the extent of LCCS classes has been estimated from the Phase 1 Habitat Survey rather than the UK Land Cover Map (LCM2000) because of the greater detail provided. Ultimately, the LCCS map generated from SIAM™ spectral categories will be used to direct the sampling effort for GHCs. In a realistic approach, a target of 20 field samples per LCCS class is suggested, with the remainder obtained through reference to high-resolution datasets (i.e., aerial photography, LiDAR). This would result in 480 field-based samples as 26 GHC categories have been identified (see subsection 3,2,3).

Within GHC sampling, the standard approach is to randomly locate 1 km grid cells and, based ideally on interpretation of aerial photographs, delineate the main features in the landscape (e.g., linear roads, waterways or fields, woods). From each GHC, one is selected for more detailed sampling, as outlined earlier and in later sections. To select the 1 km grid cells, a stratified random sampling within a regular grid (SRSRG) has been proposed. To illustrate this approach, 10 % of the grid cells containing each of the LCCS classes (translated from the Phase 1 Habitat Survey) were selected randomly. As each km^2 consisted of a diversity of habitats, only those where the percentage area occupied by each was > 33 % were identified initially (Figure 4.3.3). The maximum area proposed for sampling is 30 km^2, with this area needing to encompass the 26 LCCS classes occurring. Therefore, for each of the LCCS classes, one was selected randomly for sampling. A number of cells contained a high diversity of habitats (< 33 % for all) and a random selection of these (four) was also taken as a larger number of habitats could be surveyed within a small area. An alternative option is to apply a K-means classifier to the proportions of each LCCS class within each of the 1 km grid cells and randomly select from these. The Directive 92/43/EEC Annex I habitats occurring within the catchment are included in the samples. The distribution of the 30 1 km cells within the Dyfi catchment is shown in Figure 4.3.4. Note that these are located both within and outside the Natura 2000 sites.

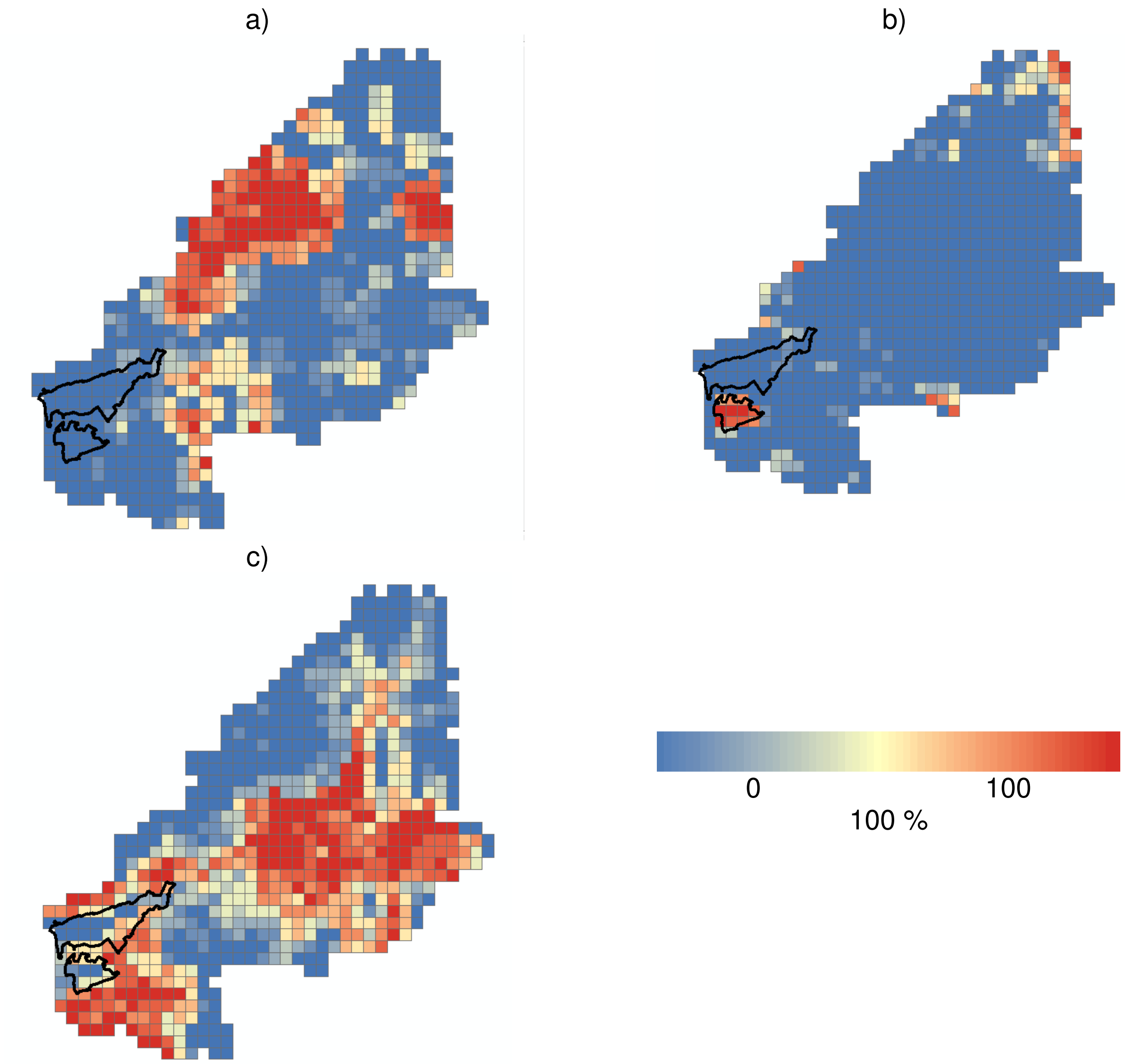

Figure 4.3.3 - The percentage of each 1 km^2 grid occupied by a) conifer forest, b) active bog and c) improved grassland. The areas covered by the Natura 2000 sites, including Cors Fochno, is outlined in black.

For the Italian sites, SRSRG within smaller cells was proposed. An example is given in Figure 4.3.5 for the Dyfi catchment whereby a ¼ km^2 grid, with each cell centred within each 1 km grid cell, has been generated. Mapping of GHCs within the 1 km^2 is to be undertaken, but priority (including field sampling) may be given (if constrained by time and resources) to those sites contained with a ¼ km^2 cell centred in each 1 km^2 selected.

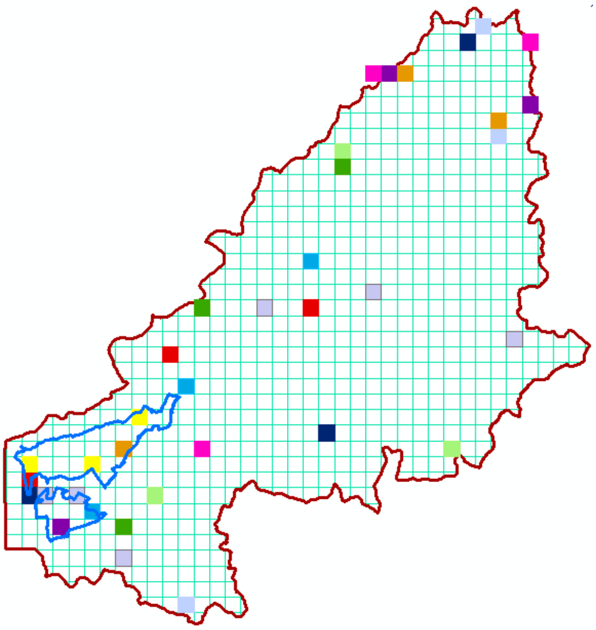

Figure 4.3.4 - The location of the 1 km sample grids by LCCS category within the Dyfi catchment (red) and in relation to the Natura 2000 area (blue).

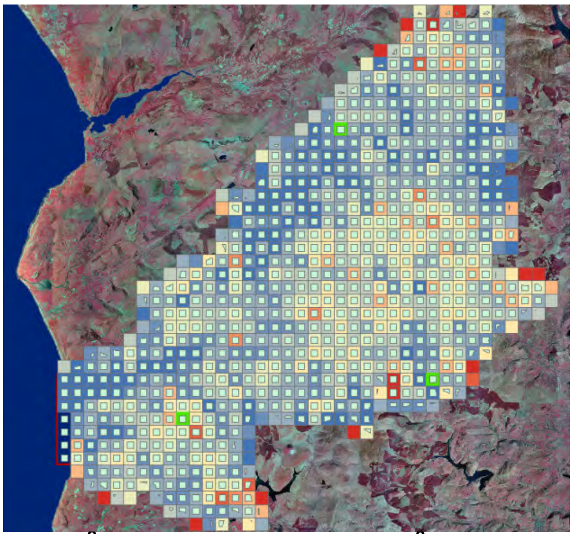

Figure 4.3.5 - ¼ km^2 cells centred on the 1km^2 cells. Whilst 1 km cells will be mapped, the survey might focus on the ¼ km cells at the centre if resources are limited.

4.3.4 Automated delineation of landscape units (areal, linear and point)

Within each 1 km^2 cell, the units within which GHC surveys are typically selected from areas delineated through manual interpretation of aerial photography. However, by using image segmentation procedures, these landscape units can be automatically delineated using very high resolution (VHR) imagery. For this purpose, eCognition image segmentation and classification software is being used.

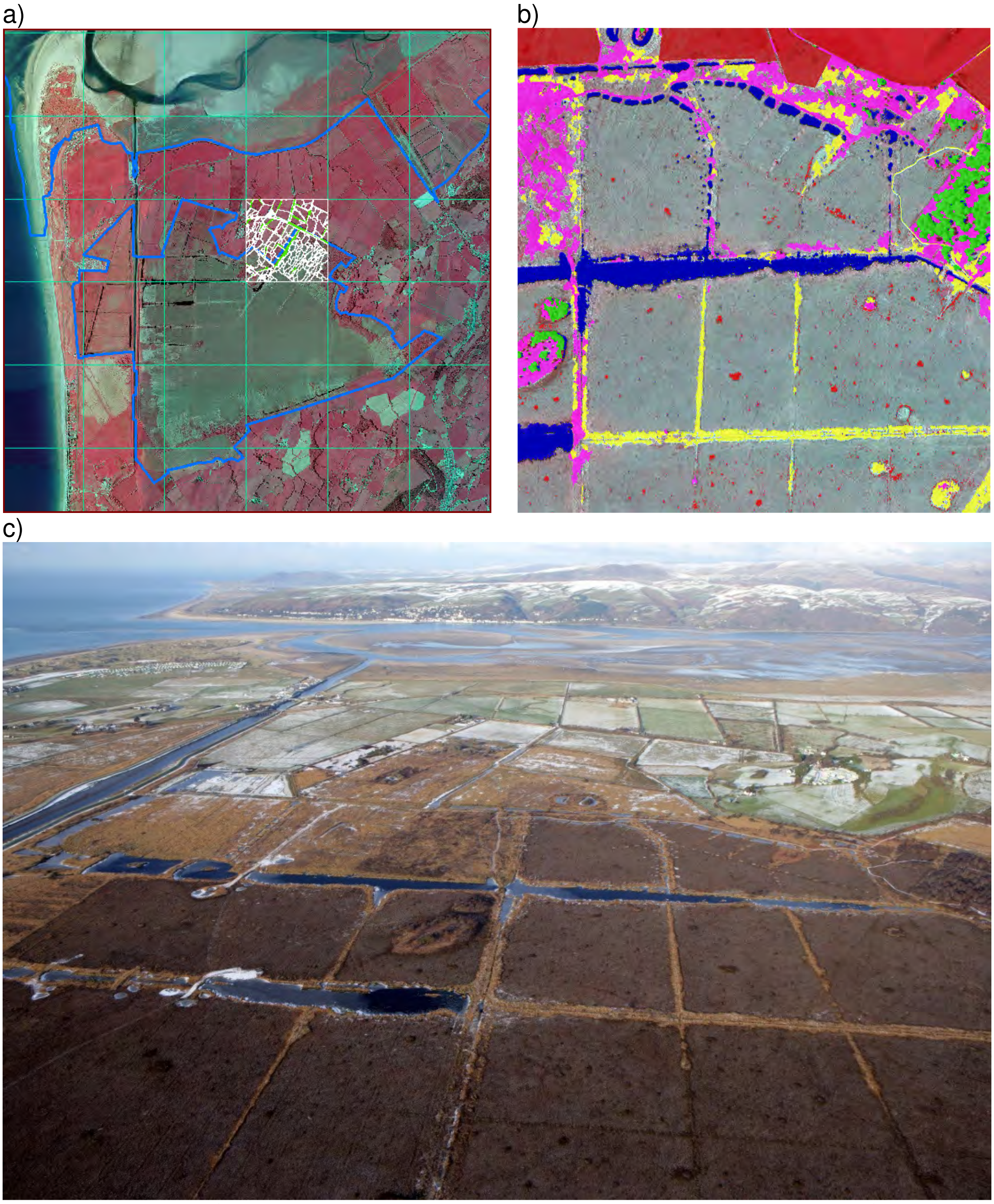

Figure 4.3.6 - a) Selected 1 km x 1 km with segments generated using eCognition overlain. b) a subset showing detailed view of segments classified into open water (blue), woodland and scrub (green), medium tall grasslands dominated by *Molinea/Phragmites* (yellow) and *Juncus* species (magenta) and clusters of shrubs (e.g., *Calluna*) (red). The active bog surface is shown in grey. Improved fields are located at the top of the image (red). The classification was undertaken within eCognition. Areal, linear and point features can be identified. c) Aerial image of Cors Fochno taken in December, 2010.

Two types of data are being used for segmentation as these are more generally available:

a) True and false colour aerial photography. These are available for all of the Dyfi catchment at < 1 m spatial resolution. Whilst the data are not calibrated, and so cannot be divided into spectral categories using SIAM™ , they can be segmented using eCognition to allow definition of landscape units. The segmentation can be improved by including existing ancillary data (e.g., thematic layers representing the urban area or field boundaries).

b) Spaceborne VHR data. VHR data were acquired on the 27th July, 2011, although at a viewing angle of > 25°, and additional tasking has been requested. An advantage of using these data is that a SIAM™ first stage classification can be generated and used subsequently to segment and classify the landscape.

Once segmented, the resulting units can be classified (Figure 4.3.6) and then translated into areal, linear and point features, with the classification refined using manual procedures.

Bunce *et al*. (2010) recommend that the mapping of habitats be carried out first, with this then used to determine the position of the vegetation plots for sampling GHCs. Typically, such mapping is undertaken in the field. However, an alternative is to use aerial photography, multispectral and/or hyperspectral data to provide a preliminary map of habitats, which can then be verified and refined during the reconnaisance field visit and used to direct the subsequent survey of GHCs. The classification also benefits significantly from the inclusion of LiDAR data. As an example, Figure 4.3.7 highlights how the landscape can be pre-classified into GHCs using a combination of LiDAR and aerial photography.

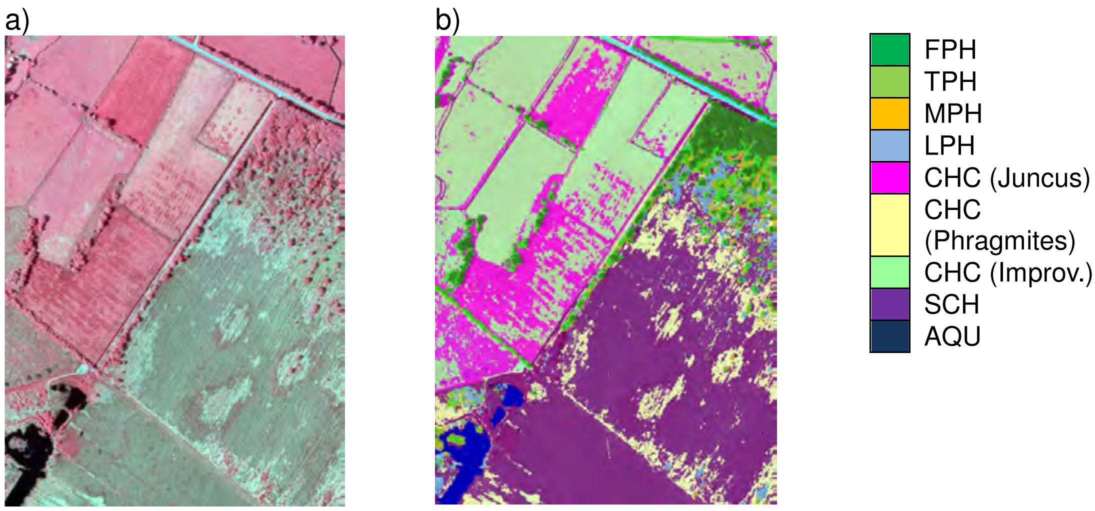

Figure 4.3.7 - a) Colour infrared aerial photography and b) Classification of GHCs using a combination of the photography and LiDAR data.

A limitation of the using airborne data to pre-segment the landscape is that too much detail is sometimes provided, as highlighted in Figure 4.3.7b. For this reason, a coarser segmentation (Figure 4.3.8) can be undertaken using the aerial photography but also available vector layers (e.g., field boundaries, urban areas) and each segment then be classified into linear or polygon features. Point features can also be identified for selected polygons using a finer segmentation. This approach is again more automated than relying entirely on manual delineation, although interactive merging of segments may be necessary prior to the field survey (e.g., where multiple segments occur within improved fields).

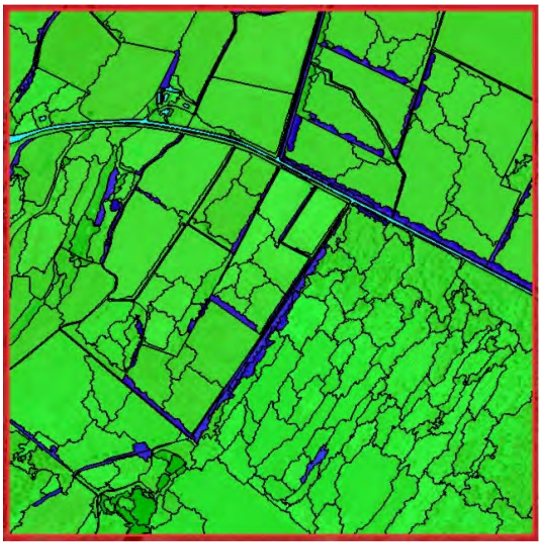

Figure 4.3.8 - Division of the landscape into larger segments and subsequent classification into linear and polygon features.

4.3.5 Randomised sampling within elements

Following generation, each segment can be associated with the dominant GHC (e.g., as a function of area), as identified within the more detailed airborne-derived classification and refined (if necessary) in the field survey. Bunce *et al*. (2010) highlight that only <u>one</u> plot needs to be located within each GHC identified, although some (e.g., grasslands) may need to be subdivided depending on the different moisture and nutrient levels occurring. Within each GHC, polygons for sampling can be selected using random points, with these generated using several procedures (refer to Figure 4.3.9):

a) A grid is overlain onto the map and points are associated with the intersection of the grid lines.
b) A grid is overlain and, within each cell, a point is randomly located.
c) The centroids of each polygon are randomly located.

Bunce *et al*. (2010) recommend the use of a). Once points have been located, one is selected randomly for each GHC category. The subsequent sample plot is then located within the centroid of a polygon, which may represent an areal, linear or point feature. Further field plots can be established following the subsequent visit to the 1 km^2 and assessment of the GHCs occurring. Plots are also not placed in urban and sparsely vegetated areas with the exception of terrestrial (e.g., beaches).

4.3.6 GHC recording in the field

Within each field plot, GHCs will be recorded using standard methods and recording sheets. A typical GHC survey requires recording within vegetation plots of variable size depending upon whether they are located within areal, linear or point elements. The plots will have the following dimensions and characteristics:

a) A Minimum Mappable Element (MME) of 400 m^2, with minimal dimensions of 5 x 80 m or 20 x 20 m.
b) Where the element is smaller than 5 m, it is recorded with a Minimum Mapping Length (MML) of 30 m.
c) All elements that do not pass the MME or MML criteria for either areal or linear elements are to be mapped as point elements or as portions of a larger element.

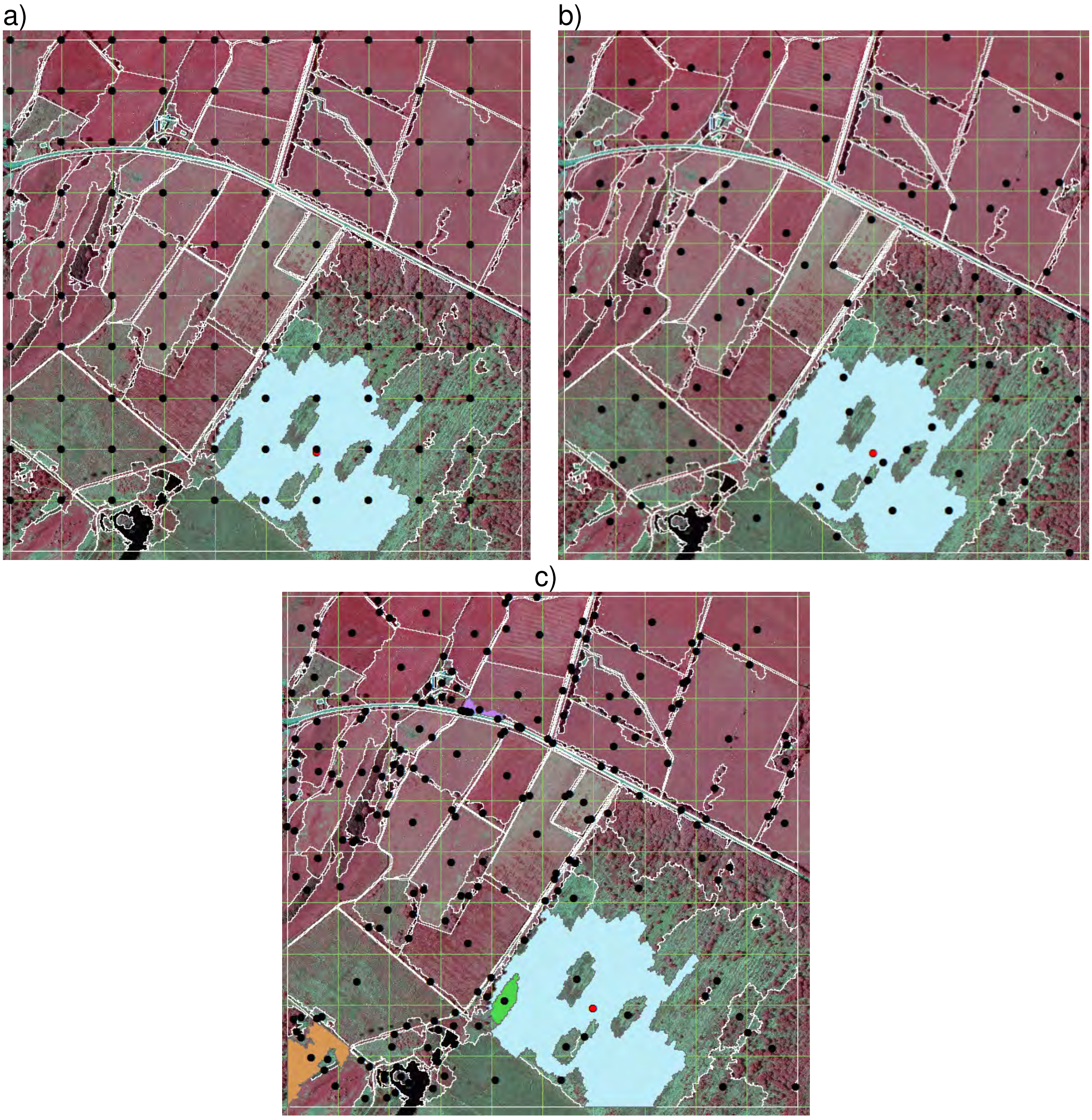

Figure 4.3.9 - The location of random points associated with a) grid intersections, b) random locations within grids and c) centroids of polygons. The selected polygon is highlighted in blue, and the centroid at which a field plot is to be located is highlighted in red. In c), polygons representing four different GHCs are highlighted as an example, with sampling to be undertaken at the location associated with the polygon centroid.

Examples are given in Figure 4.3.10. Plots should be located randomly within each defined element and be 10 x 10 m in dimension within areal elements and 10 x 1 m within linear elements. For areal elements, plots should be located at the centroid of each selected polygon. For linear elements, the centroids are placed at a random location along the length. Point elements are located within either areal or linear features (in most cases).

4.3.7 GHC maps from field survey

Following field survey, maps of GHCs can be generated for the 1 km^2 area and used to validate GHC maps generated from HR and VHR satellite sensor data. However, only 480 sample sites are proposed which is well below the amount required to achieve desired levels of accuracy and error tolerance. For this reason, additional samples are required from other sources.

a) b) c)

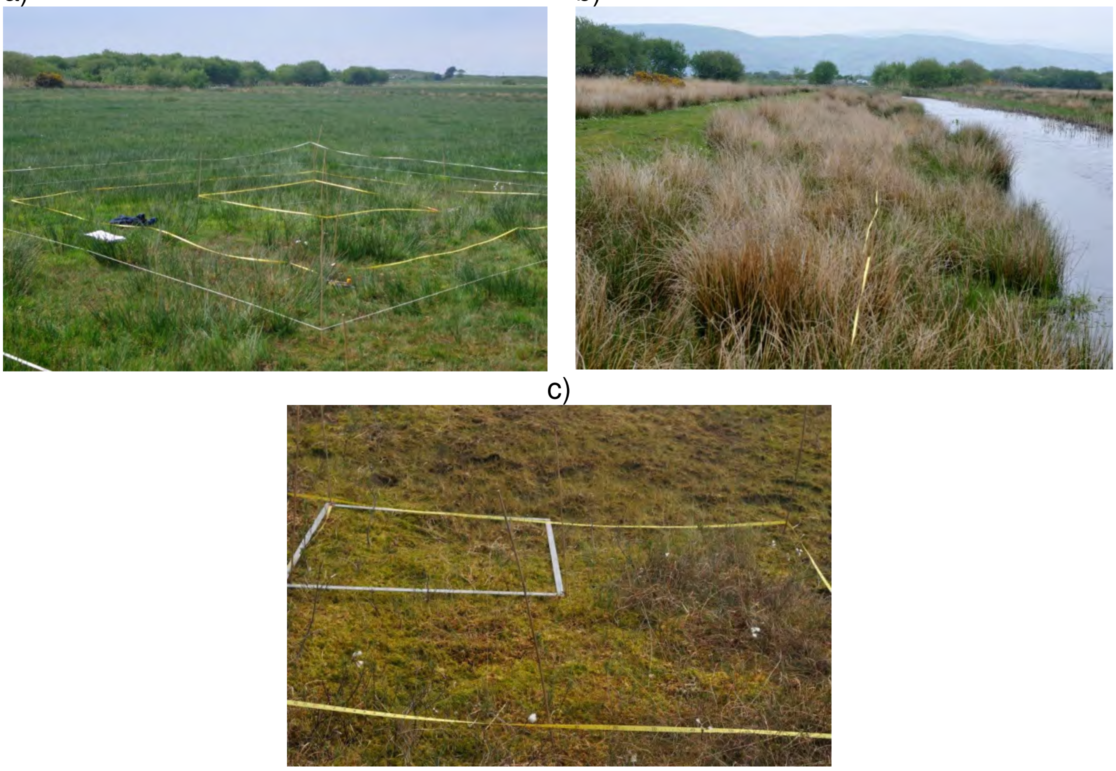

Figure 4.3.10 - Examples of sample plots located within a) areal, b) linear and c) point elements, Cors Fochno.

4.3.8 GHC maps from airborne data.

LiDAR data were acquired over Cors Fochno in 2009 at 0.5 m post-spacing. From these, a raster Digital Terrain Model (DTM), Digital Surface Model (DSM) and Canopy Height Model (CHM) have been derived. EAGLE hyperspectral data and aerial photography are also available. Based on these data, a rule-base for classifying GHCs directly is being developed and implemented to establish whether classifications can be undertaken for larger areas, thereby reducing the need for intensive survey, increasing detail in the classification and allowing a greater area to be surveyed. An example of such a classification has been given previously in Figure 4.3.7b. Accuracies in the classifications of GHCs from these airborne datasets will be assessed using field data on GHCs collected for the study area.

4.3.9 GHC maps from LCCS based on satellite sensor data.

For all BIOSOS sites, maps will have been generated by translating LCCS classes derived from the SIAM™ first stages to GHCs. Example GHC classifications based on translations from LCCS-classifications derived from Landsat and SPOT sensor SIAM™ spectral categoreis are provided in D5.1 and in Figure 4.3.10. Validation of the classification will be undertaken with reference to the following:

a) GHC maps generated for the random selection of 1 km^2 cells and based entirely on field survey.

b) GHC maps generated for the entire Natura 2000 sites and surrounds based on hyperspectral and/or LiDAR data, with the latter validated with a).

Initially, validation will be based on standard confusion matrices, recognising the limitations of these. The VHR classification will be evaluated in the first instance and, if considered sufficiently robust, will be used to validate the classification of GHCs from SPOT and Landsat sensor data.

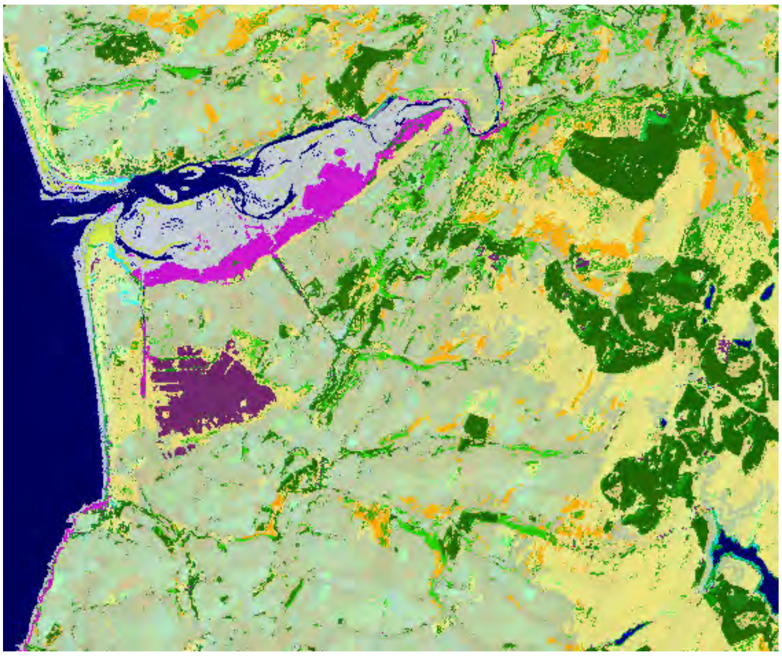

Figure 4.3.10 - Classification of GHCs translated from LCCS classes derived from SIAM™ first stage.

LCCS		LCCS Code_Modifier	Description
A11		A4.B1.B5.C1.D1.D9_B4	CRO/GRA
A11		A1.B1.B5.C1.D1.D9_A8.B4-W7	WOC/ TRE/TPH/FPH-CON(EVR/DEC)
A11		A1.B1.B5.C1.D1.D9_A7.B4-W7	WOC(DEC)
A12		A3.A10.B2.C2.D1.E2_B5	TRE/TPH/FPH-DEC
A12		A4.A11.B3.C2.D1.E2_B14	VEG/TRE/MPH/TPH
A12		A4.A11.B3.C2.D1.E1	SCH/DEC/EVR
A12		A6.A10.B4.C1.E5_B12.E6	CHE
A12		A6.A11.B4.XX.E5_A12.B12.E6	CHE
A12		A6.A10.B4.C2.E5_B13	CHE
A12		A5.A10.B4_B12/B13	LHE
A24		A4.A20.B3.C1.D1.E1.F2. F4.F7.G4_C4	TRS(DCH/SCH/LPH/MPH) HER(SHY/EHY/HEL/LEA) HER(EHY-FLO)
A24		A6.A12.B4.C1.E5_B11.C4.E6	EHY
A24		A6.A13.B4.C1_B13.C5	HEL
B15		A3_A8	ART
B15		A3_A10	ART
B15		A4_A13	ART/NON
B16		A3.A7	ROC
B16		A6.B6	SAN
B16		A6_A12	STO

B16		A5_A13	STO/GRV
B27		A1.B1.C1.D1_A5	AQU
B27		A1.B1.C1_A4	AQU
B27		A1.B1.C1_A5	AQU
B28		A1.B2.C1_A5	AQU
B28		A1.B3_A4.B6	AQU(TID)
B28		A1_A4	AQU(SEA)

4.3.10 Subsequent survey by CCW with ABERY

For recording biodiversity at the Italian sites, circular plots (100 m radius) were proposed with these being in addition to the GHC plots. Based on the GHC classification, a more detailed quadrat-based survey will be undertaken by CCW and ABERY in September, 2011. Each sample will be located within the centroids of the GHCs polygons sampled during July-August, 2011. In each quadrat, a detailed list of plant species will be recorded. Existing data on biodiversity distributions are also being collated through CCW, with these often being line transects. This information is being collated and will be made available to the BIO-SOS project and specifically for WP6. It is proposed that data on soil pH, aspect, slope and other environmental variables will be collected along previously sampled line transects. Information on animal communities will also be collected where possible, including insects and birds.

4.4 Stratified sampling for GHC in NL sites

A preliminary GHC survey is being carried out for the two NL sites Wekeroms Zan, Ederheide Ginkelse heide.

A stratified random sampling within a cell of a 250 m (1/16 km^2) regular grid built based on the National land cover database LGN-6 (reference year 2010, 25 m spatial resolution). Such a choice was due to several constraints. However, the opportunity of using the INSPIRE as a standard grid is under consideration. The same LGN-6 database was also used to determine strata, with this being this the most accurate and up-to-date land cover information available. Moreover, a very strong relationship between EODHaM (SIAM™) LCCS LC/LU maps and LGN6 is assumed, especially for major categories. Relationships will be studied in August 2011.

The sampling is being carried out considering the whole sites embedded in a common 3 km wide buffer zone, comprised of 1823 250m cells (114 km^2 in approx.) (Figure 4.4.1). Even though a 1 km wide buffer zone would have been sufficient according to most experts, it was chosen to widen it in order to include a part of a large town (Ede), since the use of protected sites for recreational purposes represents an important pressure.

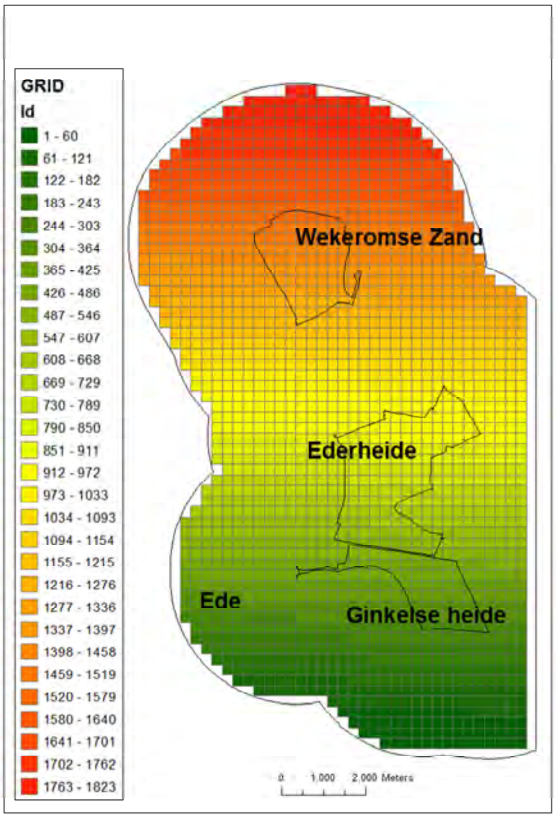

Figure 4.4.1 - 250 m fixed grid used for stratified random sampling over NL protected sites and their surroundings.

Land cover statistics were calculated separately for major categories inside and outside protected sites (within the buffer zone). The number of samples was weighted by the area of each major land cover class, resulting in 6 samples within the protected sites (4 should in

semi-natural areas, and 2 in forested areas) and in 6 samples within the buffer zone (3 in agricultural land and 3 in forested areas), for a total of 12 samples (Table 4.4.1 and Figures 4.4.2 and 4.4.3).
The GHC mapping of all samples was performed in the period June-July 2011 at the maximum development of the vegetation. Within each GHC category in each sample a vegetation plot will be recorded according to the protocols of section 5.1.6.

Table 4.4.1 - Land cover statistics for major classes in protected sites.

Major Land Cover Class	Area_ha	Area_%	Samples_n
inside protected areas			
Urban	6.9	0.5	
Agriculture	34.3	2.3	
Forest	483.9	32.6	2
Semi-natural area	959.6	64.6	3.9
Water	0.3	0	
inside totals	1484.9	100	6
outside protected areas			
Urban	2259.6	21.1	
Agriculture	3514.1	32.8	3
Forest	4395.6	41	3
Water	26.7	0.2	
Nature	515.4	4.8	
outside totals	10711.3	100	6
Totals	12196.2	-	12

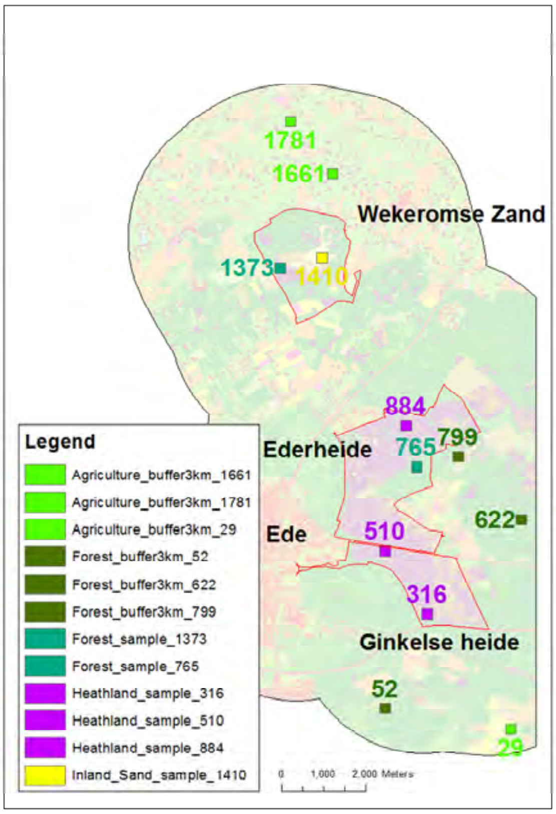

Figure 4.4.2 - Overview of 12 random selected samples with land cover as background.

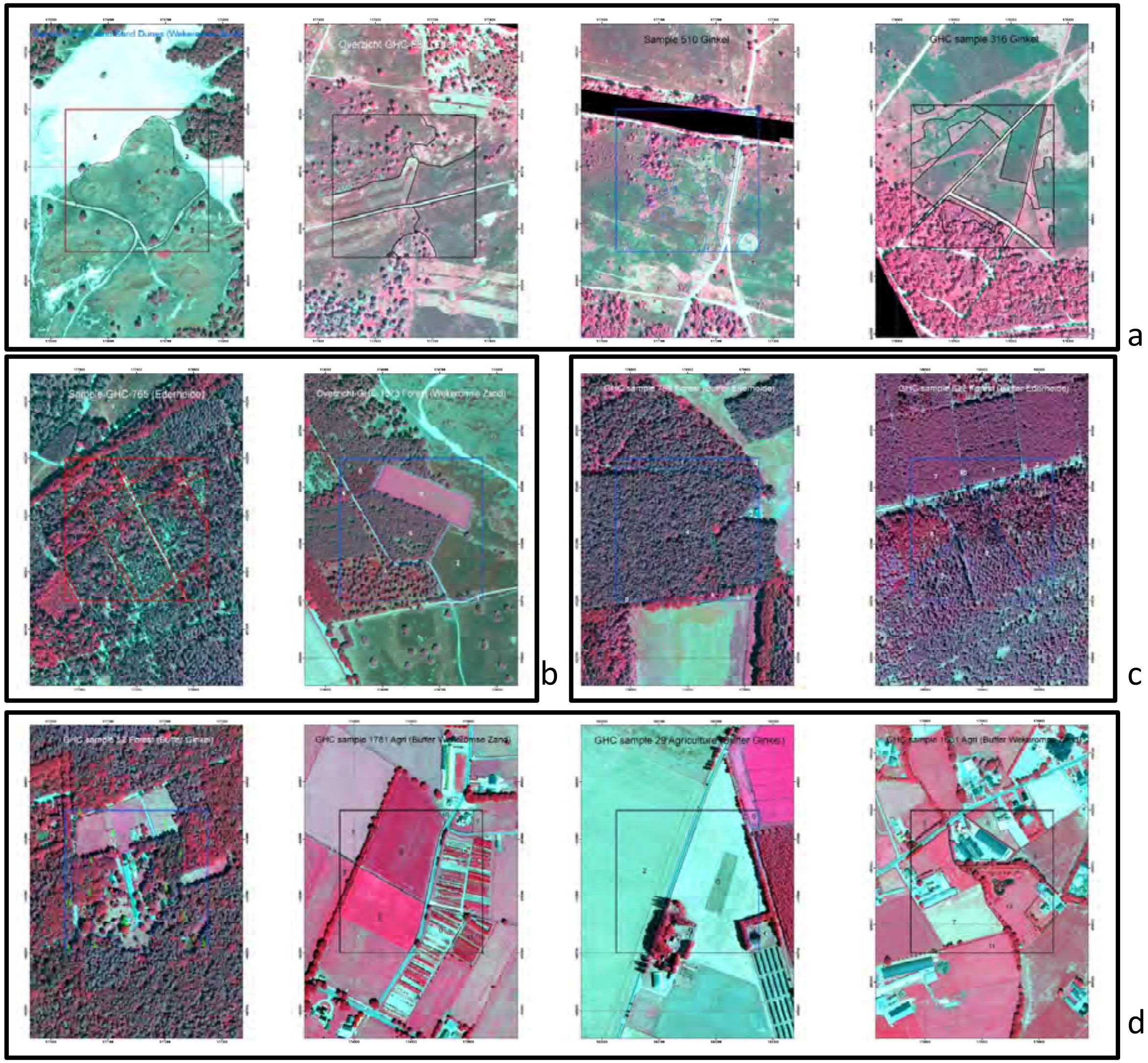

Figure 4.4.3 - Detail of samples on an aerial photographs background with basic patches prior to field survey. a inland and heathland core samples; b Forest core samples; c Forest samples in buffer zone; d Forest and agricultural samples in buffer zone.

4.5 Stratified sampling for GHC, biodiversity and soil in IT3

For this site, the sampling will be limited to the frame of the EO VHR (e.g., IKONOS; Quickbird, Worldview2) image used to produce the base LCCS map (100 km^2 in approx.). The sampling design will be a stratified random sampling within a cell of a 1 km standard regular grid (INSPIRE). An example of a possible sampling layout is given in Figure 4.5.1.

The sampling extent will coincide with the core of the frame of the EO VHR image used (WP5) to produce the base LCCS map (100 km^2 in approx.) in order to avoid incomplete 1 km^2 when selecting samples. The sampling extent will cover parts of the N2K site and the national park as well as areas outside both. Thus, three different areas will be identified according to a decreasing protection regime: N2K-NP, NP and non- N2K-NP. Two different types of buffer zones will be identified, one in which official bindings of the N2K apply but the actual legal provisions for the national park are not in force, one with neither a formal nor a legal protection.

As for the sample size, 1/3 of the entire extent will be sampled by randomly selecting 1 km^2 units among those containing portions of the target LCCS classes, with a distribution weighted according to N2K-NP, NP and non- N2K-NP relative total area and habitat amount.

Such a sampling strategy will that both the habitat fragmentation and landscape heterogeneity gradients are considered, while providing sampling for all the major LCCS classes (Figure 4.5.2).

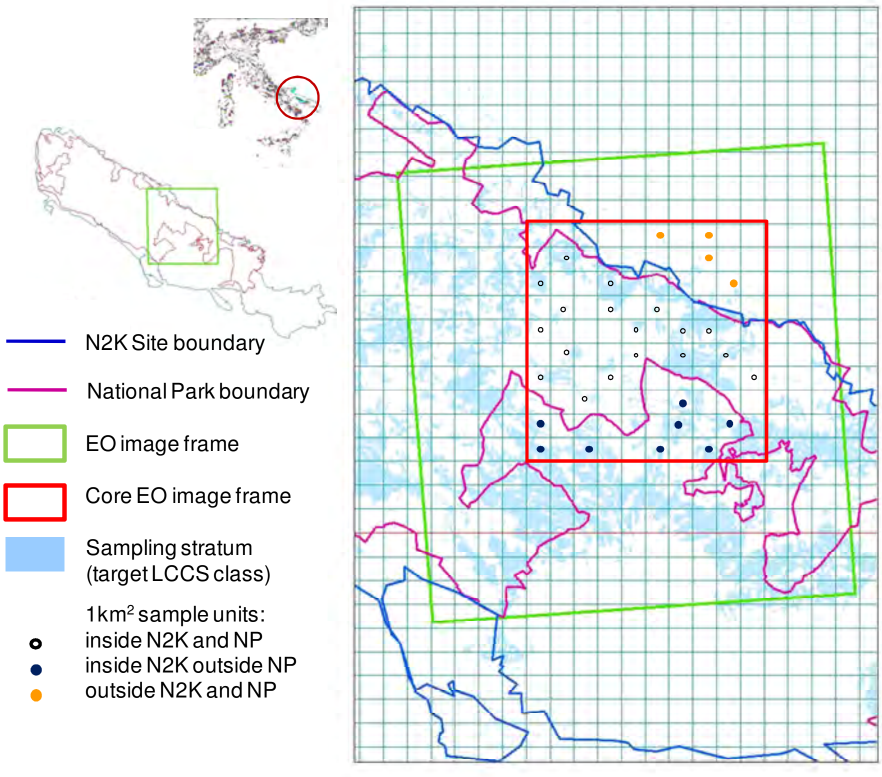

Figure 4.5.1 - Example of a possible 1km^2 sampling units for a target LCCS class in IT3.

The 1 km^2 thus selected will serve as the sample set for the purposes of landscape pattern analysis (LPA, WP6) (Section 5.3.4).

For the purposes of GHC identification and sampling (according to the standard protocol outlined in section 5.1), a sub-set of such samples will be selected by means of ranked set sampling using the expert judgment capability acquired by means of LPA and 1/4 km^2 (or 1/16 km^2) sample units will be adopted. Such samples will be nested within the previous ones.

For the purposes of data recording for biodiversity, in the same sample sub-set as for GHCs, circular units (100 m radius) will be lain. Within these, in addition to the 100 m^2 X vegetation plots within each (relatively homogeneous) GHC (section 5.1.6) 50-100 m line-transects for line intercept sampling (Eberhardt, 1978; DeVries, 1979) will be randomly lain on a possibly heterogeneous LU/LC complex (section 5.2.2). Finally, the centre or the GHC sample unit (Figure 4.5.2) will be identified for point count sampling in order for such sample group to be spatially independent.

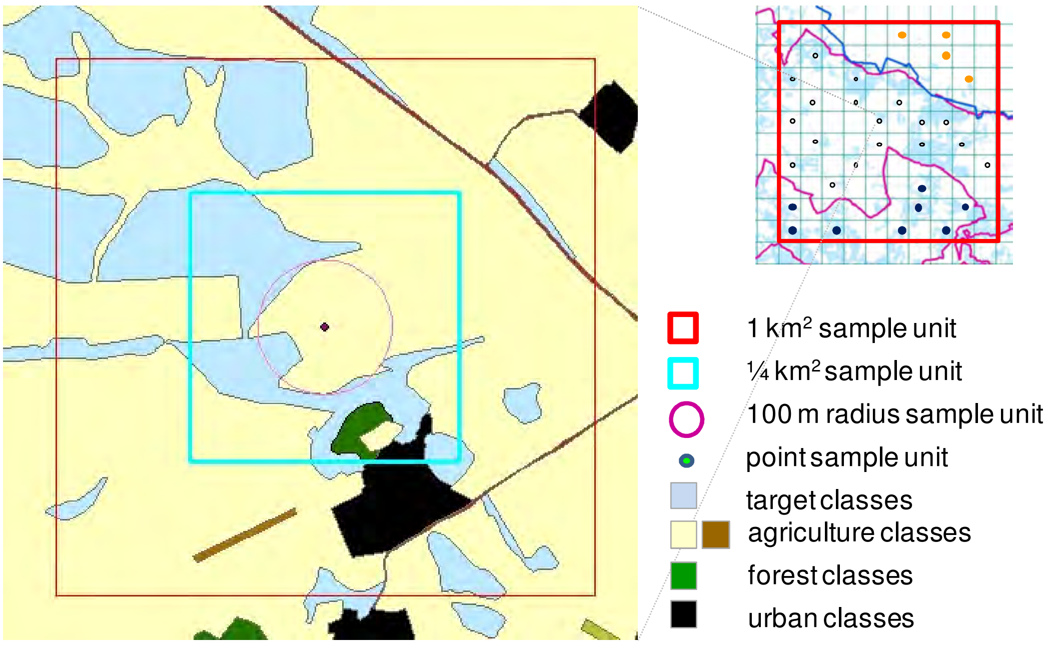

Figure 4.5.2 – Example of sample units for LPA, GHC and biodiversity data collection in IT3.

Soil will be sampled mainly in locations affected by intense anthropic activities and hence clearly different from the original soil still present in the Murge site. The sampling will be preceded by a field reconnaissance survey devoted to define whether elements of the protected habitats are affected negatively. If the survey gives positive results, the identified areas will be subdivided in supposedly uniform sub-areas and then sampled following a spatially random distribution. This unit will be chosen as much as possible in agreement with the previously described sampling technique, most probably in a subset of the aforementioned 1 km^2 units.

4.6 Stratified sampling for GHC and biodiversity in IT4

For this site, the sampling will be limited to the frame of the EO VHR (IKONOS, Quickbird, Worldview2) image used to ptoduce the base LCCS map (100 Km^2 in approx.).

A stratified random sampling will be designed within a cell of 1 km^2 based on the 1 km standard regular grid (INSPIRE).

The sampling will be carried out considering the whole site is embedded in a 1 km wide buffer zone (except for the coastal side of the site), in order to include part of the agricultural areas surrounding the site (since agricultural practices can potentially have an important impact on these natural habitats).

For the purpose of GHC and biodiversity identification and sampling, 1/16 km^2 sample units will be adopted. The choice of this small sample unit is due to the small size and large heterogeneity in terms of land covers and habitats within this site.

4.7 Stratified sampling for GHC, biodiversity and pressures in PT1 PT2 sites

The sampling strategy for GHCs in PT1 ad PT2 will basically consist of a site-specific adaptation of the general protocol described in Part I of the report. Focal areas spanning across the border of the two test sites will be surveyed to collect calibration and validation data on GHCs for image classification, under a two-stage stratified design with a random selection of sampling units within strata. Previous land cover maps will be used for a first-level stratification of the focal area (Figure 4.7.1a.) Patches within each LC class will then be classified according to their area (i.e. extent) (Figure 4.7.1b), since this attribute is expected to influence the number of distinct GHCs that can be found within the patch. When surveying each patch, one point per GHC will be recorded and characterized. Point data will be favoured over patch/polygon data in order to allow comparison among different LC/habitat classifications produced with distinct legends and at several spatial scales.

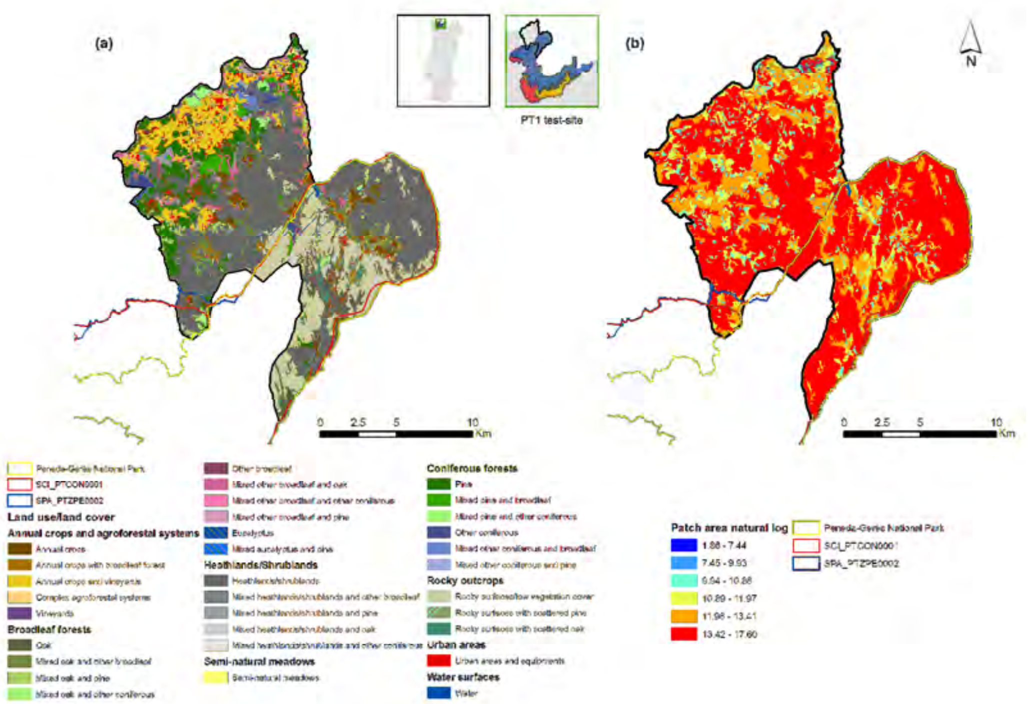

Figure 4.7.1 - Examples of (pre-existing) stratifying layers due to support the sampling design for the collection of *in situ* data on GHCs, biodiversity and pressures inside at outside the PT2 test site: (a) land cover map (scale 1:25.000; year 1990); (b) patch area map, based on the land cover map in (a).

Moreover, a set of ¼ km^2 samples will be placed at selected locations inside and outside test sites for a more detailed and spatially continuous survey of GHCs as described in section 4.1. These samples will support calibration and validation of habitat maps for rural landscape mosaics and will also be the subject of biodiversity surveys (see below). The location of ¼ km^2 samples will be stratified random, using pre-existing land cover data and related datasets as sources of stratifying layers.

The sampling strategy for biodiversity will be aimed at collecting data on (i) the distribution of target rare plant species, and (ii) the distribution of taxonomic and functional diversity across habitat types in landscape mosaics. For rare species, a model-based sampling (Figure 4.7.2) will be used to improve the efficiency of field surveys in detecting unknown populations across the focal areas in both test sites (Guisan et al. 2006; Lomba et al. 2010). Moreover, in all ¼ km^2 GHC samples described above, abundances will be estimated based on an LC-stratified random selection of habitat patches. Taxonomic and functional diversity will also be surveyed (as described in section 4.2.1) in a subset of the ¼ km^2 GHC samples. These will be selected under a ranked set sampling framework (Wolfe 2010), with land cover and GHC maps as sources of ranking variables, and plant diversity in each sample will be surveyed using a GHC-stratified random selection of habitat patches.

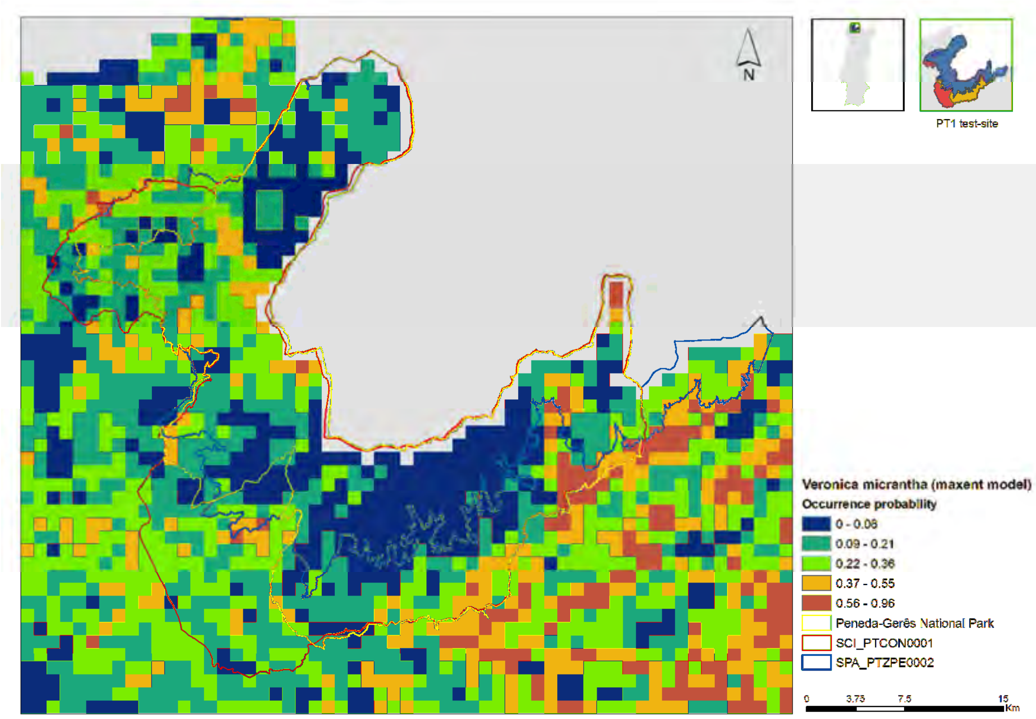

Figure 4.7.2 - Example of a (pre-existing) spatial layer due to support the sampling design for the collection of in situ data on biodiversity in Portuguese test sites: predicted regional distribution of the rare plant species Veronica micrantha in the North of Portugal (inside and outside of the PT2 site), based on predictive species distribution models (e.g., MaxEnt).

The sampling strategy for pressures in PT-1 and PT-2 will be site-specific and pressure-specific. For land use related pressures, preliminary analyses of pre-existing LC maps (1:25.000 scale; cf. Figure 4.7.1a) will provide a coarse assessment of recent land use changes, demographic and socio-economic dynamics, and wildfire occurrence across the focal areas in the two test sites. This will provide stratifying layers for subsequent sampling. For alien invasive species, a similar procedure to the one described for rare species will be followed, including the use of models for individual species (Vicente et al. 2011) and for landscape invasibility by multiple species (Vicente et al. 2010) (Figure 4.7.3). Targeted imagery classification will also provide additional distribution data on dominant woody invaders (e.g. *Acacia dealbata*). A ranked set selection based on extent of invasion will be used to select a subset of ¼ km^2 GHC samples to evaluate the impacts of invasion on habitats and on native biodiversity.

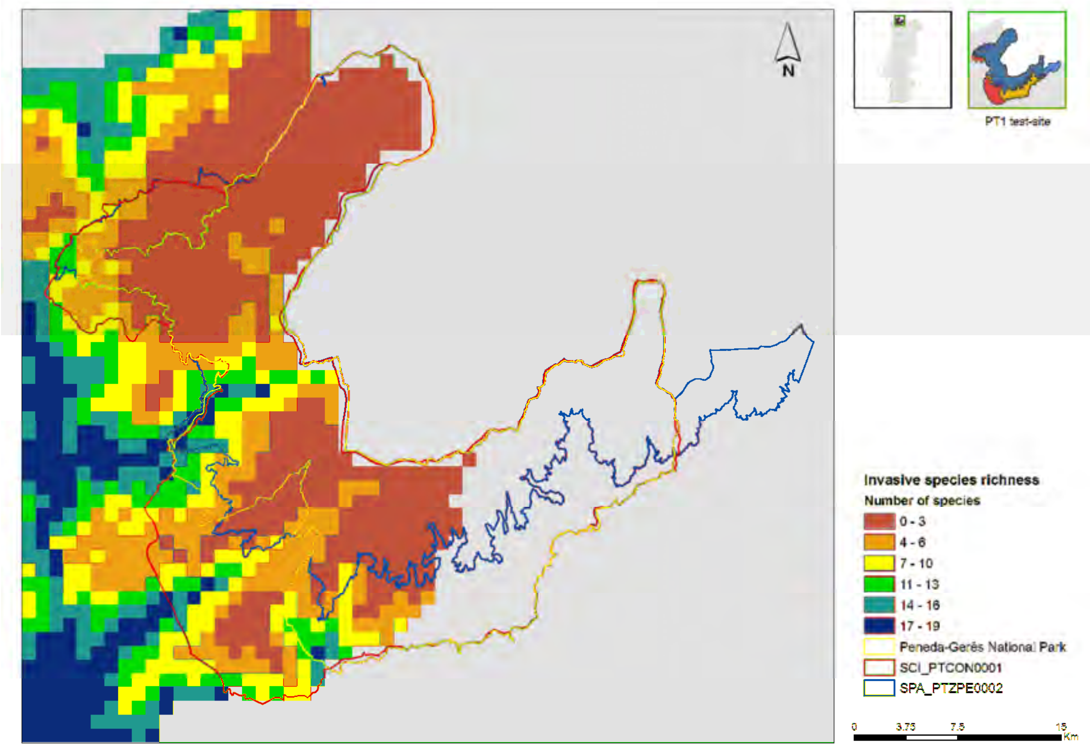

Figure 4.7.3 - Example of a (pre-existing) spatial layer due to support the sampling design for the collection of *in situ* data on biodiversity in Portuguese test sites: predicted regional distribution of species richness for alien invasive plants in the Northwest of Portugal (inside and outside the PT2 site), based on predictive species distribution models (<u>Generalized Linear Models for alien species richness</u>).

4.8 Stratified sampling for GHC, biodiversity in GR1 GR2 GR3 sites

In the BIO_SOS project, we are studying three Natura 2000 sites. For these sites, the sampling design will be a stratified random sampling within a cell of a 1 km standard regular grid (INSPIRE) (Figure 4.8.1). The sampling extent will coincide with the core of the frame of the EO VHR image used (WP5) to produce the base LCCS map (100 km^2 in approx.). The sampling extent will cover parts of the N2K sites, as well as areas outside both.

Site GR1 (GR2120001) is the Ekvoles Kalama N2K site covering an area of 8481 ha. The dominant land cover class is agricultural areas followed by forests and semi natural areas, while a considerable part of the area is covered by water. In this area, we will take five (5) samples of surface area of 250x250 m^2 (1/16 km^2) for GHC mapping. Two (2) of these samples will be taken in agricultural areas, two (2) in forests and semi-natural areas and one (1) in wetlands and water bodies. Furthermore, four (4) more samples will be taken from outside the N2K site. These samples will be located in similar land cover classes (LCCS), as the samples in the protected areas (i.e. agricultural land, forests and semi natural areas, and water bodies).

Site GR2 (GR2120002) is the Elos Kalodiki N2K site covering an area of 845 ha. Forests and semi natural areas cover approx. 43% of the site, while agricultural land covers approx. 30% of the site; the remaining proportion is wetland. Only one (1) sample will be taken from inside this area and one (1) more sample will be taken outside the area. Both samples for GHC mapping will cover a surface area of 250x250 m^2 (1/16 km^2).

Site GR3 (GR2120004) is the Stena Kalama N2k site covering an area of 1867 ha. The dominant land cover class is forests and semi-natural areas covering more than 90 % of the site. In this site, one (1) sample will be taken from inside the N2K site (from forested areas) and one (1) more sample will be taken outside the area. Both samples for GHC mapping will cover a surface area of 250x250 m^2 (1/16 km^2).

Regarding biodiversity, sampling will be restricted to sample plots of smaller size, because in an area of 62500 m^2, exhaustive recording of all plant species is prohibitive given the time and cost limitations of the BIO_SOS project. Therefore, the relevés will be carried out in a standardized way as concerns a) the plot size with four (4) plot sizes for the BIO_SOS vegetation (phytosociological) relevés, following Chytrý & Otýpková (2003) and b) a template for common data standards when effecting phytosociological relevés according to Mucina et al. (2000), discussed in Section 4.2.1. These sample plots will be used to estimate plant diversity and to help identify, in detail, the plant communities present in the study sites in correspondence with the habitat types of the Annex I of the Dir. 92/43/EEC.

Such relevés will be carried out in sample plots taken inside the large samples used for GHC mapping, but also outside them. Whenever possible, this sampling will try to identify the locations and thus to replicate a similar sampling scheme performed in the 1999-2000 period, when the habitat types mapping of the Natura 2000 sites had taken place, so as to provide some insight on the biodiversity changes occurring in the meantime. Furthermore, these samples will be focused on areas of natural vegetation, while artificial surfaces and cultivated fields will be avoided. Finally, these samples will focus on habitat types of conservation priority (according to the Habitats Directive 92/43/EEC) inside the Natura 2000 protected area.

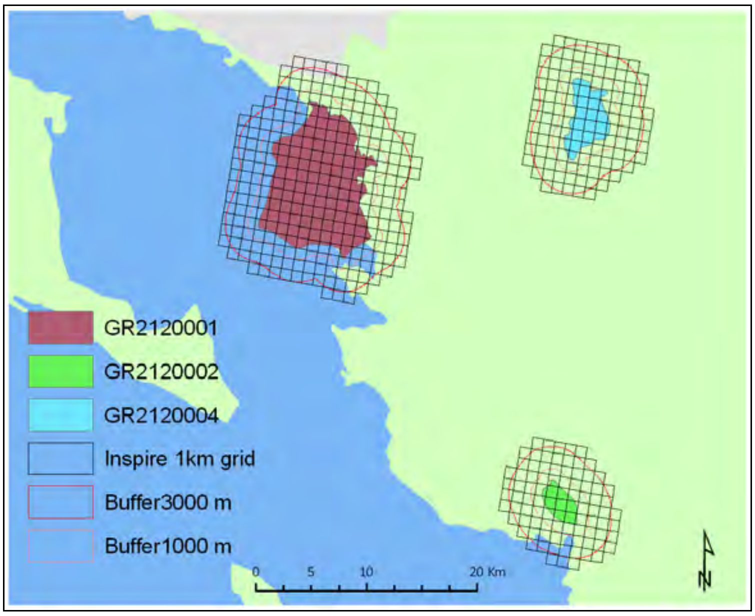

Figure 4.8.1 - Stratified random sampling within a cell of 1km standard regular grid (INSPIRE).

4.9 Stratified sampling for GHC and biodiversity in Indian sites

The two Indian sites are located in the Western Ghats biodiversity hotspot region of India (Myers, 2000). Currently there is no existing national sampling protocol and designated grid for the whole country as is available in the European Sites. Thus an appropriate sampling methodology will be developed for BIO_SOS project and followed in both sites.

Indian Site 1 (IN1): The Biligiri Rangaswamy Temple (henceforth, BRT) Wildlife Sanctuary (henceforth, WLS) area lies in between the Eastern and Western Ghats region of India, and is highly biodiverse (1400 species of higher plants, Ramesh, 1989; Kammathy *et al*., 1967) (Figures 4.9.1 and 4.9.2). The area surrounding the park is densely populated with agriculture as the main occupation of the residents. There is high dependence on natural vegetation in this surrounding area for fuelwood and fodder, the scrub forests having long been converted to barren areas (Barve et al, 2005). A land use/cover map for the region is available from previous mapping exercises (Ramesh and Menon, 1997), which has been used to overlay a grid for sampling (Figure 4.9.1). As the surrounding area has very high human impacts, remnant vegetation here is extremely sparse and only seen in places that are not already covered by agriculture. Thus, only the protected area has been considered for detailed sampling.

ATREE has been involved in long term vegetation monitoring of the forests in BRT WLS (Murali and Setty, 1998). Systematic sampling has been carried out here as follows: The BRT WLS was overlaid by a uniform 2 x 2 km grid, and a 5 x 80 m plot was laid at the center of each grid cell. In previous vegetation sampling exercises, within each 5 x 80 m plot, the diameter at breast height (DBH) of all living woody vegetation ≥1 cm DBH was recorded. In case of multi-stemmed plants, all stems ≥1 cm DBH were counted, and their DBH was measured. Individuals of all woody species were identified in the field whenever possible; samples of unidentified species were brought back to the field station and identified using floras and herbarium records. Vegetation in these 122 plots has last been sampled between August 2007 and January 2008 (Sundaram, 2011).

This dataset will provide an invaluable base for study, as it allows us to investigate the biodiversity in IN1 for well over a decade. For the purposes of GHC identification and sampling, we will select a subset of cells in the 2x2 km grid. The selection will be based on stratified random sampling, weighted on the basis of disturbance and pressure observed from previous field studies. Grid cells with greater pressure and disturbance, requiring higher amount of monitoring, will be surveyed for vegetation and landuse/cover type to comply with the GHC categories. Thus, more cells will be located in the areas that are more accessible (such as those near major roads, at lower elevations, near encroachments and plantations, and near the periphery of the park), where most anthropogenic disturbance is anticipated. In particular, since invasive plant species represent a prominent type of disturbance in this region that is of special interest to researchers, local community forest users and forest managers, we will also focus on identifying grid cells in areas expected to have variations in the presence and density of invasive plants.

In each of these identified cells, we will select a 100 x 100 m plot at the centre of the grid cell for GHC mapping. The choice of the 100 x 100 m plot size is suggested keeping in mind that GHCs have not been as yet verified for India, and it is anticipated that this protocol will be more time consuming and complex in the sub-tropical Indian sites where biodiversity is substantially higher, and landscape heterogeneity and forest structure also considerably more complex than in the Mediterranean and north European sites. Thus, generating GHCs for ¼ by ¼ km plots

will be an extremely challenging task, while a 100 x 100 m plot seems more feasible. However, this is a tentative plan and will need to be flexible so it can be modified in the field if necessary, depending on our field experiences.
For the purposes of data recording for biodiversity, in order to relate our plots to the previously available time series data on biodiversity in the park since 1998, we will lay rectangular 80x5 m plots at the centre of the GHC plots, within which we will record data on plants using the protocols for previous datasets as outlined above.

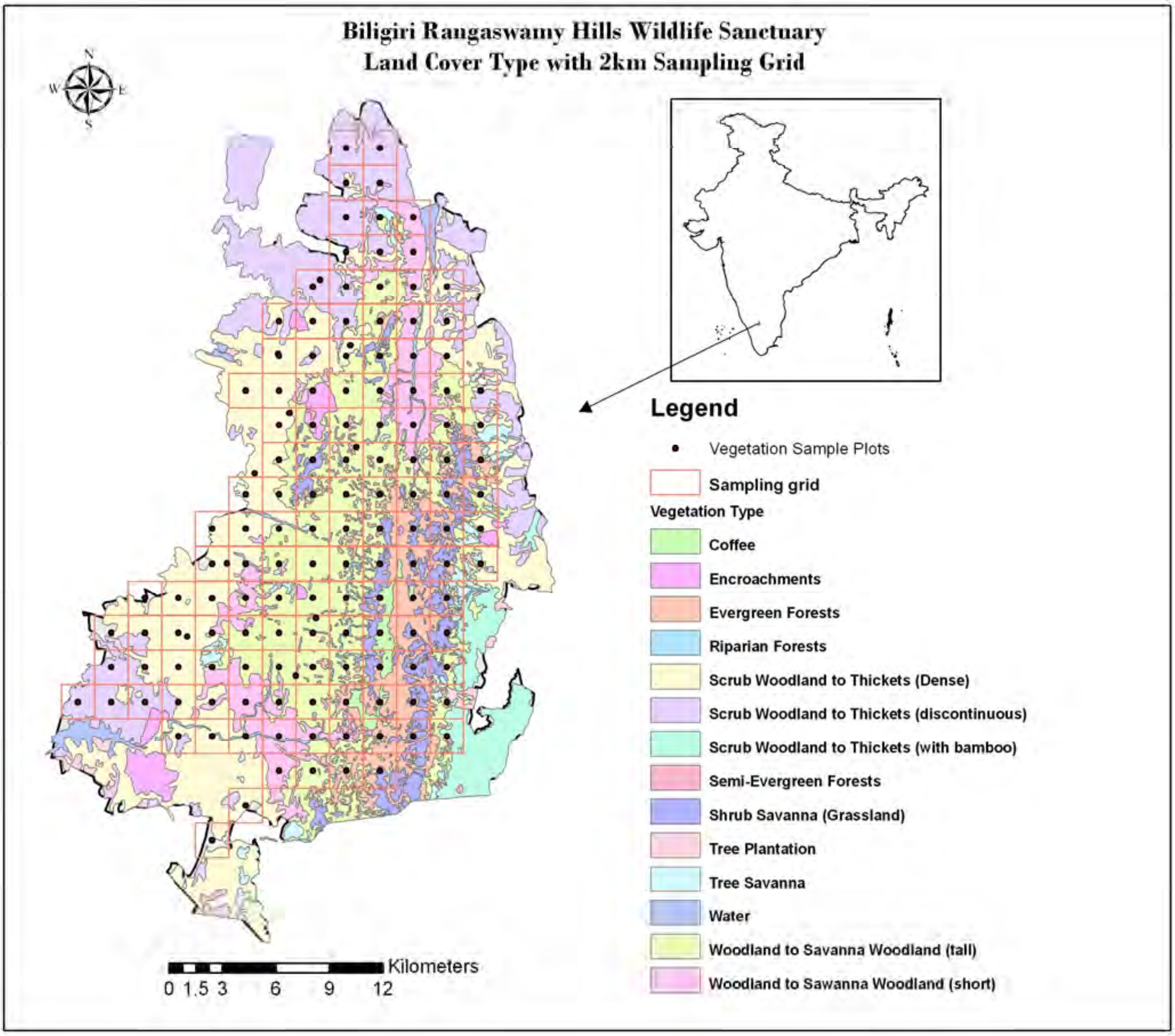

Figure 4.9.1 - A vegetation type map of the study site, with grid overlay for sampling. The center points of each grid cellindicate the location of the 80 x 5 m plots.

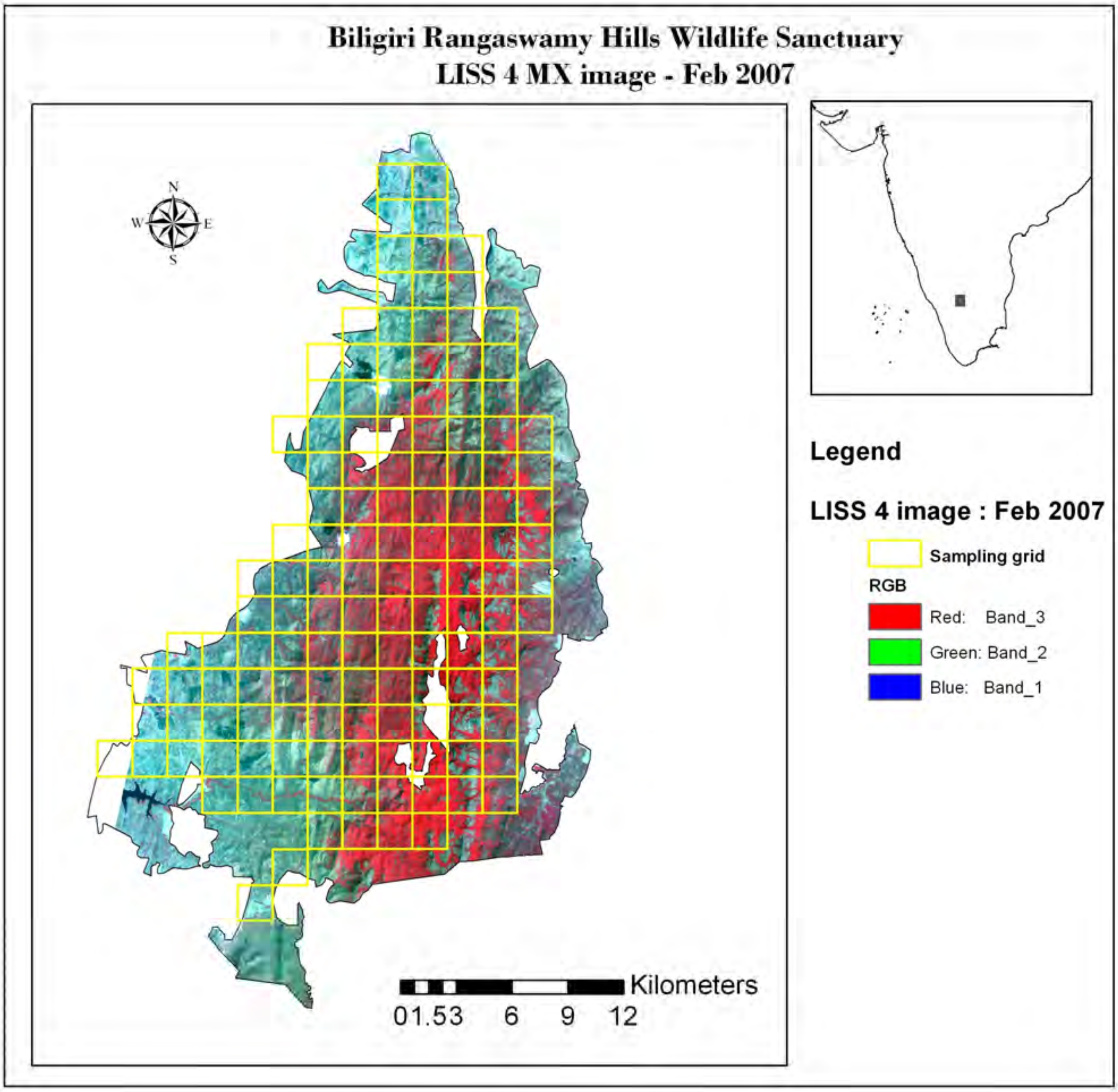

Figure 4.9.2 - A LISS 4 MX image (spatial resolution 4 m) for the study area in February 2007.

Indian Site 1 (IN2)**:** The Netravali Wildlife Sanctuary (henceforth, WLS) is situated in the coastal state of Goa, in the Central Western Ghats region (Figures 4.9.3 and 4.9.4). This site also has very high biodiversity and endemicity of plants as well as fauna. However, inspite of notifying it as a WLS in 1999, there is very little documentation and almost no detail studies conducted in this region. There is no previous detailed landuse/cover map easily available for the WLS on which the stratified sampling protocol can be based (FSI, 2009). Ancillary data such as roads, location of settlements (both at 1:50,000), and soil maps (1:500,000) are available for the area. Digital Elevation Model from ASTER is available at 30 m resolution. In addition to the WLS, buffer areas will be also be monitored in locations where natural vegetation exists, as several kinds of anthropogenic pressures including agriculture and mining prevail in the buffer area surrounding the park.
Following the same methodology as described for IN1, a sampling grid overlay of 2 km x 2 km will be overlaid on the site which includes the WLS, and the areas of the

buffer where native forest is present. For GHC mapping, stratified random sampling (ranked set sampling) of grid cells will be undertaken within this grid based on a disturbance index developed from field knowledge, such that disturbed areas are sampled to a greater extent than relatively undisturbed, inaccessible areas that are of less interest from the point of view of mapping anthropogenic pressures and responses. Within each cell, a 100 x 100 m plot will be located for GHC sampling. Again, as with IN1, the size of this plot is only tentative and may need modification based on field experience, as GHC mapping has not been conducted in the more diverse sub-tropical forests of India before, and we do not know what field challenges we can expect while conducting such sampling, or the time and effort that this may take. Vegetation of the area will be studied and GHC categorization will be developed on these studies. Within these 100 x 100 m GHS plots, central 80 x 5 m plots will be used to record information about plants following the same protocol as outlined for IN1.

An intensive study will be undertaken for the pressure of invasive species in both the study sites. The incidence of invasive shrub species *Lantana camara* and *Chromolaena odorata* has increased tremendously in the overall region (Sundaram, 2011, Murali and Setty, 2001), and reports from the local management personnel (Personal Communication) have indicated some establishment even in the northern study site (IN2). Focused field sampling for this biotic pressure phenomenon will be undertaken in particularly vulnerable locations within the sites.

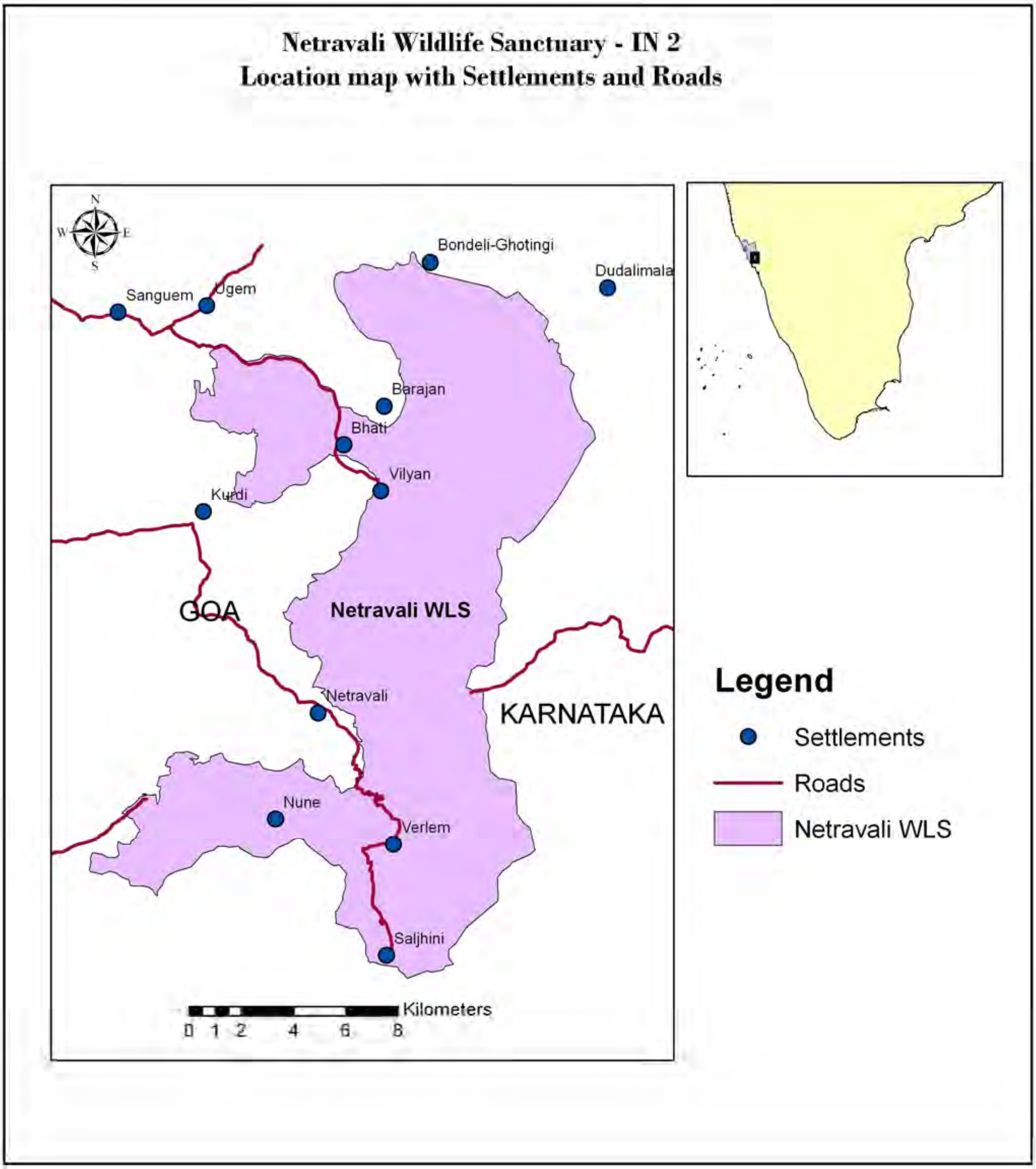

Figure 4.9.3 - Netravali WLS location map showing locations of settlements and roads.

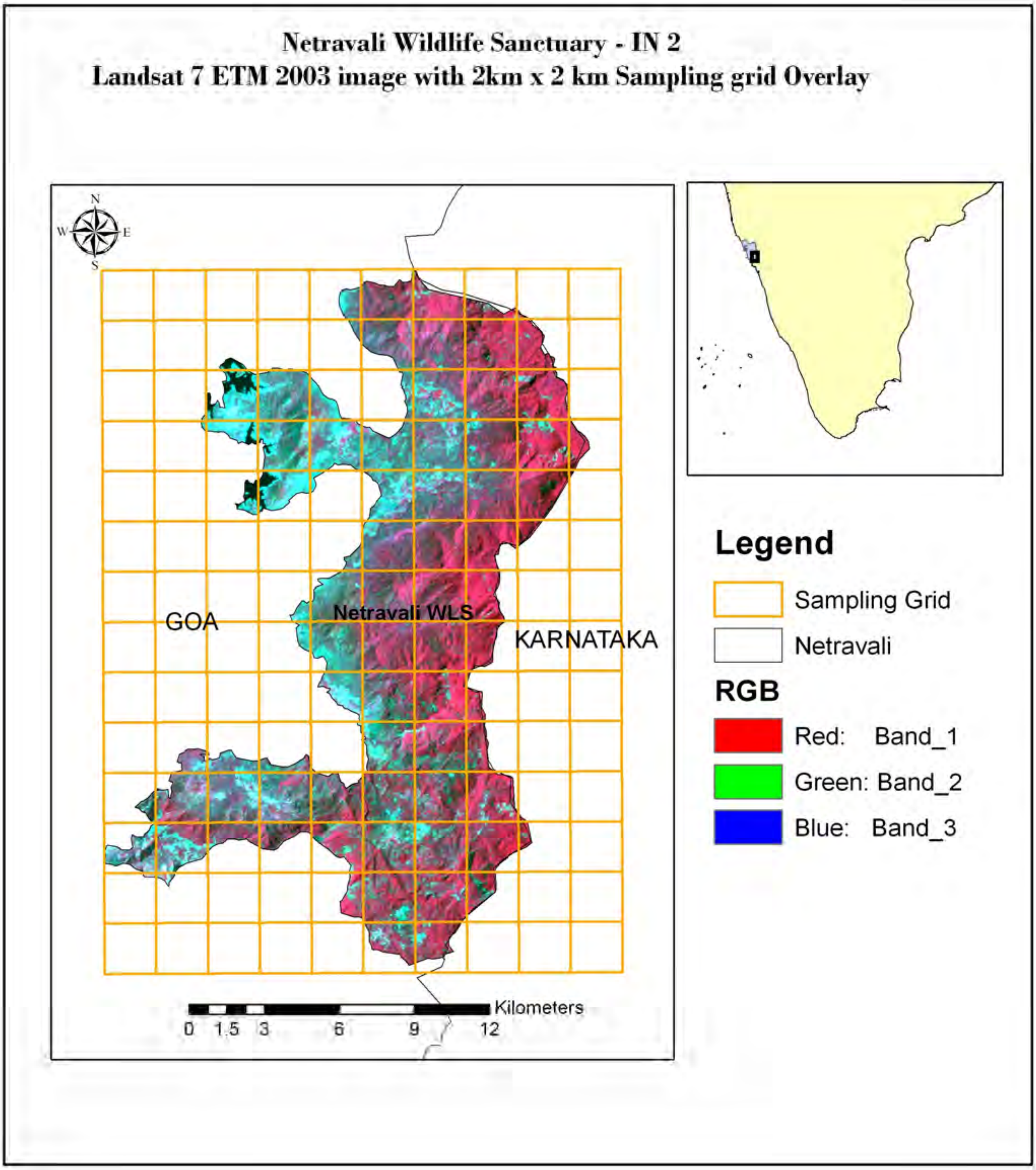

Figure 4.9.4 - Landsat 7 ETM image for Netravali WLS for February 2003.

5. Measurements protocols

5.1 General Habitat Category mapping and recording

The instructions of habitat mapping and recording in this deliverable are based on the Manual for Habitat and Vegetation Surveillance and Monitoring: Temperate, Mediterranean and Desert Biomes by Bunce *et al.* 2011. For a full understanding of the method we recommend to read this carefully.

A preliminary general training session on the methodology was organised in Italy (April 18-20 2011) by P1 and P8 and led by P4 (Bob Bunce and Marion Bogers) in two BIO_SOS Italian sites (IT3 and IT4). Another demonstration, also led by Bob Bunce and Marion Bogers followed during the Wales meeting and more formal and site specific training sessions will be organised for both Italian and Greek sites in the forthcoming months.

5.1.1 General instructions

The survey area: The basic survey area is1 km^2, ¼ km^2 or 1/16 km^2 within which areal and optional linear and point elements are recorded.

Time window for survey: For monitoring, the recording of the GHCs should be made in a time window as close as possible to the height of the growing season. This window is likely to be before maximum biomass in the Mediterranean. Local flexibility may be required for variations in the weather.

The field team: A field team should consist of at least two people for safety and for consultation. Mixed teams, preferably with a botanist and an experienced mapper or GIS expert, are needed to ensure that the team is balanced. Adequate field training is required for all surveyors.

Quality control is essential and involves regular liaison with staff in the field, and direct supervision and consultation. The Manual must be referred to continually in order to optimise field performance. Adequate on-site training is essential for the quality of the output.

Database checks is a first stage to carry out automated checking. It is also essential to carry out manual checks with an expert observer to ensure that the data are as consistent as possible.

5.1.2 Preparations

Output from Remote Sensing exercise: Preparatory work on delineation of the major elements within the survey area from the aerial photograph, map or satellite images will be provided by the Remote Sensing teams.

Aerial photographic prints at a scale of 1:5,000. Aerial photographs should preferably be ortho-photos or else geometrical properties need to be assessed.

With **permanent markers**, on sheets that are copied from the most recent 1:10,000 scale base map including topographic and/or cadastral information, enlarged to 1:5,000 scale or on transparent overlay sheets placed on aerial photographs.

Separate **sheets or overlays** are to be used for the mapping of areal and of linear elements.

Aerial photographs with digital outlines or equivalent maps of elements can be held on the field computer.

Application of **field computers:** Major advances are taken place in the application of field computers for the recording of habitat data. Various options are now available. It is important to note that all systems involve previous interpretation of different types of aerial photographs to produce parcel outlines which are then validated in the field.

Photo camera: It is strongly recommended that a photograph is taken including a GPS position to illustrate the local conditions at the time of recording and as input for later quality assessment.

It is strongly recommended to prepare **a list of dominant species** of the area to be surveyed including the appropriate Life Form.

5.1.3 Mapping of individual elements

The size of individual elements: An areal element is 400m^2 with minimum dimensions of 5 x 80m or 20 x 20m. If the element is smaller than 5m. it is recorded as a linear element with a minimum length of 30m. Elements that do not pass the criteria for either areal or linear elements can be mapped and recorded as point elements or as proportions of a larger element. If an area is less than 400m^2 in the survey square but belongs to a larger element outside, it should be mapped as an area. Areal elements are drawn on a separate map from the linear elements.

Assigning Alpha codes: Elements are assigned alpha codes as identification codes that are the same on the map and on the corresponding recording sheet. Capital letters of the Latin alphabet are used for the alpha code. "I", "O" and "X" and should not be used. Once all the letters of the alphabet have been used then double codes are used: e.g., AA, AB, AC etc. The alpha code for an areal element should be placed as closely as possible to the centre of the element, as shown in the worked examples.

Separate mapping elements that have identical data coding have the same alpha code, but they should be recorded as A1, A2, etc on the map.

Rules for separating map elements (i.e. new Alpha codes): A new areal or linear element will be mapped and separated from adjacent or surrounding elements if any one of the following nine rules is true:

- A change in GHC.
- A change of more than 30% of a cover of a GHC.
- A change in environmental qualifier.
- A change in site or global qualifier.
- A change in the occurrence of point elements.
- A change in management qualifier e.g., a fence line or age of forest trees.
- A change of at least 30% in the cover of an individual species over the whole element
- A change of at least 30% in any of the TRS layers, if they are being recorded under forest canopies.
- A change in any other specified European habitat, especially the habitats of Annex I of the Habitats Directive.
- A change in the proportion in the Annex I habitats.

In agricultural land separate fields should be mapped individual, even though the boundaries may or may not be delimited by fence lines or grass strips. These data are required for subsequent spatial analyses.

5.1.4 Recording of areal elements

In all cases the field surveyor should make many decisions as possible in the field. All fields of the recording form must have an entry in order to ensure that subsequent database management can identify that an entry has not been omitted in error. There is a separate recording format to be used for areal and linear and point elements. The recording form has alpha codes as identification codes that are the same on the map.

The recording form:

1. The first field is for entry of the General Habitat Category (Annex1 in Bunce et al 2011)
2. The second field is for entry of the global qualifiers and the environmental qualifier (Annex2 in Bunce et al 2011);
3. The third field is for entry of the site qualifiers (Annex2 in Bunce et al 2011);
4. The fourth field is for entry of the management qualifiers (Annex2 in Bunce et al 2011);
5. The fifth field is for entry of the full list of habitats within the GHCs together with the major species and percentages ;
6. The sixth field is for entry of Annex I habitats;

Using the field computer: this Access based software (figure 5.1.1) is freely available and can be downloaded from the BIO_SOS ftp-site.

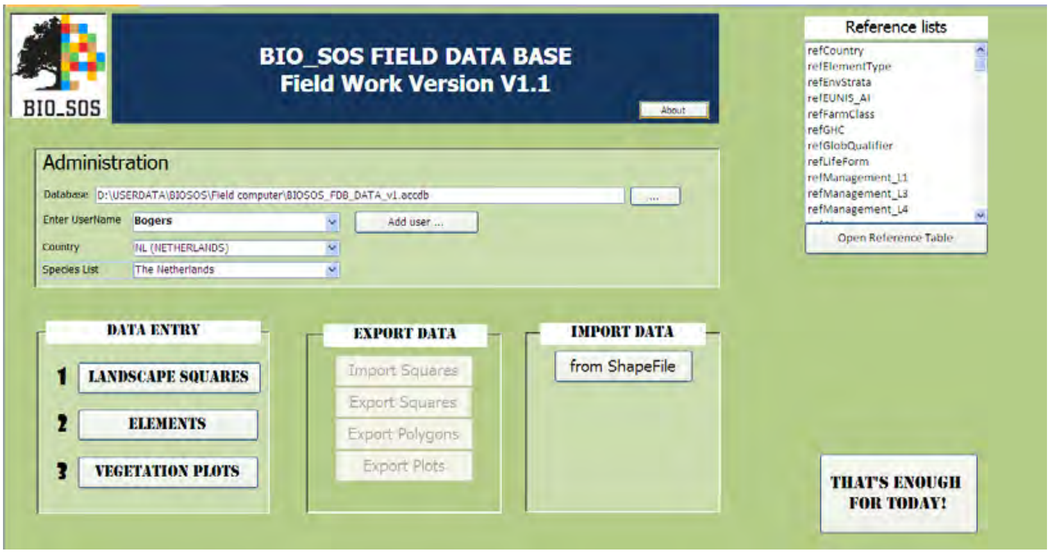

Figure 5.1.1 - Access based software for field recording.

5.1.5 Recording of linear elements (optional)

For the recording of linear elements a predefined list is used. The appropriate GHC is added as well. For BIO_SOS it is recommended to record only the categories in bold. Other categories can be added if needed, depending on local conditions and requirements.

1. Lines of trees	**LTR**
2. Hedges	**HED**
Species Rich Hedge	SRH

Lines of scrub	LSC
3. Water edges	**WAT**
Herbaceous strips	HST
Grass strips	GST
Annual strip	ANN
Walls	WAL
Banks	BAN
4. Tracks	**TRA**
5. Roads	**ROA**
Lines of SPV	LSV

5.1.6 Vegetation plot protocol (optional)

Vegetation plots are needed to The A main vegetation or X plot is 100 m2 in the centre of the GHC and is set up using survey poles with the strings forming the diagonals of the square as shown in figure 5.1.2, and in figure 4.3.10. If the plot is in a patch that is less than 10 m wide then the shape of the plot must be altered to fit within the patch.

All species must be recorded from the inner nested plot first. When the inner plot has been completed the second nested plot will be examined and any additional species will should be recorded. Cover estimates (in steps of 5%) are only made for the whole plot when all sizes have been completed. Species with less than 5% cover are given a nominal cover of 1%. Total cover maybe over a 100% if several layers are present. All vascular plants, and optionally bryophytes or lichens, are recorded. Only plants rooted in the plot are recorded, including trees.

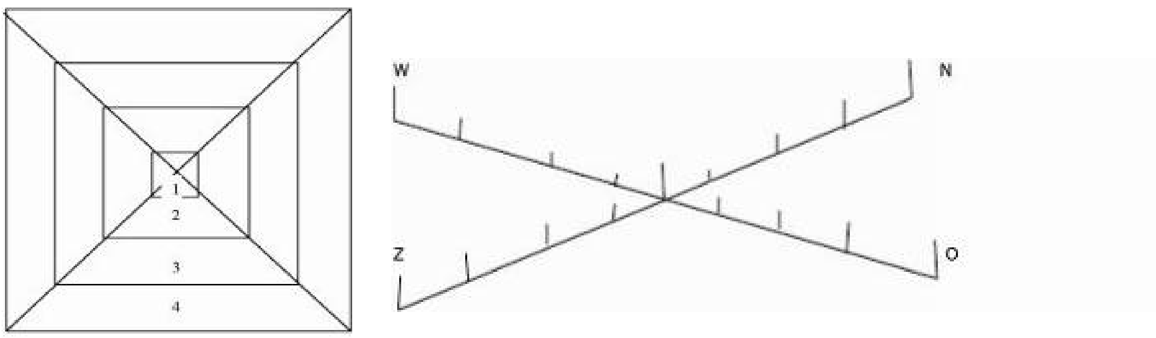

Figure 5.1.2 - Design of the X plot (After GB Countryside Survey Manual 2007). The half diagonals are 1.42 m, 3.54 m, 5.00 m and 7.07 m .This produces nested plots of respectively 4 m2, 25 m2, 50 m2 and 100 m^2.

If the plot falls in a field with a growing crop or hayfield, then the plot should be moved to the edge of the field starting 3m into the plot to avoid any edge effect. Access should be made using drill lines where possible and causing minimum disturbance to the crop or hayfield.

5.1.7 Sampling of the surrounding area

The surrounding area (1-3 km wide) will also be surveyed in the field, but can be done with less samples. No squares will be placed in Urban.

5.1.8 Digitising protocol and data management

Scale: ground resolution between 0.2 to 1m, Ortho-rectified or geo-referenced.

Age of the aerial photograph: less than 5 years old

ArcGIS 9/10 BingMaps Aerial or Google Earth, which is sometimes the best available. Set the snapping environment in GIS.

Format: ArcInfo Shapefile format. Areal, line and point in separate shapefiles. Auxiliary files (attribute table and projection). All send to the data manager together (Files: .shp, .shx, .dbf, .prj)!

Geo-referencing: Mandatory that all data are geo-referenced and all projection and datum information stored as metadata following the "Geodata format policy" definened in BIOSOS_D.8.5 Project Management and Quality Assessment Plan (section 7.1)
Data management: ArcInfo Shapefile format. Areal, line and point habitat in separate shapefiles. Auxiliary files (attribute table and projection). All send to the data manager together (Files: .shp, .shx, .dbf, .prj)

Naming conventions: Layers: 2 letters for country + Square number+ Habitat type, e.g., FR_0011_A.shp; FR_0011_L.shp. Elements: use alpha coding to name elements; preferably using single part element (i.e. each record have its own code). Use the same coding for recording into the field database.

5.2 Biodiversity

In field data collection for biodiversity at both species and community levels (flora, fauna, vegetation and animal communities) will be carried out making references to well known methodologies.

5.2.1 Floristic surveys

Depending on specific objectives or pressures to be assessed by each partner, floristic surveys will be performed in BIO_SOS test sites in order to: (i) record full species pools (for vascular plants and selected animal groups) at the landscape level; (ii) collect estimates of species richness (or diversity, if abundances are recorded) and of its patterns and components across habitat types and landscape mosaics; and (iii) record locations of rare species and collect estimates of their local abundances and patterns of distribution across habitat types and landscape mosaics.

Concerning sampling strategies for floristic surveys, three general approaches will be followed, depending on specific objectives: (i) performing a thorough wall-to-wall floristic survey of each landscape mosaic, according to specific protocols and considering standard sampling efforts; (ii) conducting a probabilistic survey (stratified according to LC/GHC classes) targeted at individual species or diversity indicators; or (iii) collecting additional vegetation plots (see section 5.1.6) in un-surveyed (or under-surveyed) GHC classes, which will complement the ones already collected for a more complete survey of floristic diversity of landscape mosaics.

From floristic surveys, three main results may be derived and used as response input variables for modelling tasks in WP6: (i) estimates for total (i.e. gamma) species richness as well as of its additive alpha and beta components (Crist et al. 2003); (ii) patterns of distribution of taxonomic and functional diversity across habitat types; and (iii) patterns of occurrence and local abundance of rare species within and across landscape mosaics.

5.2.2 Vegetation surveys

5.2.2.1 General overview of motivations and approaches

Motivations for the on-site collection of vegetation data in BIO_SOS range from the classification of vegetation units for support to 92/43/EEC Directive Annex I habitat mapping, to the biological, ecological and environmental characterization of vegetation units to support image classification and habitat recording. Therefore, procedures for describing vegetation will be planned following both methodological and spatial scale categorization.
Different data categories will be recorded for different sites and within different sample unit’s categories: 1, 1/4 or 1/16 km^2 (X plots, section 5.1.6) and those placed within 100 m radius circular units (line transects).

5.2.2.2 Vegetation data collection

Vegetation data (species composition and relative cover) in X plots will be recorded according to the GHC standard procedure (section 5.1.6) in 1, 1/4 or 1/16 km^2 sample units This will be common for all sites.

Vegetation surveys in the 100 m radius sample units, which might be carried out, will be more intensive than X plots, also aiming at the collection of vegetation structure data (e.g., number and type of layers and relative height/cover) besides those relevant to species composition and relative cover. Such data will be collected in randomly laid line transects (50-100 m line-intercept). In this kind of sample units also abiotic variables (e.g., soil pH, aspect, slope angle, rock abundance) might be recorded in order to collect data on independent (with respect to vegetation data) environmental heterogeneity variables at the local scale. This kind of data will also be used for improving the 3-D LC/LU Class Description/explanation/definition in terms of (2-D) appearance properties in the 2-D RS image domain. This protocol might be optionally adopted in the case of X plots falling within a GHC corresponding to a stratification class especially in the sites when no 100 m sample units are to be carried out.

Depending on specific objectives to be assessed by each partner, measures of environmental heterogeneity, which as independent variables are appropriate for broader scales (e.g., elevation, productivity or proxies, LC/CLU), will be derived either from ancillary data (e.g., DEM), or from LCCS base maps, or from EO images (e.g., NDVI), as well as from the landscape pattern analysis (section 5.3.4). Those independent variables will be used to test environmental variability at multiple scales as predictors of local scale habitat, vegetation and animal communities' conditions (Gould and Walker 1997; Costanza *et al.* 2011).

5.2.2.3 Phytosociological surveys

Phytosociological surveys of natural and semi-natural vegetation will also be conducted in some test sites to support the identification of 92/43/EEC Directive Annex I habitats and the evaluation of their conservation status/ecological integrity using floristic and phytocoenotic indicators *(*European Commission 2006; Honrado et al. 2007; Neto et al. 2007; Panitsa et al. 2011*)*, as already described in 4.2.1.

Habitat and compositional qualifiers

Habitat mapping is based on field surveys and measurements, such as plant ecological surveys and vegetation mapping. Permanent plots or transects are necessary to detect fine-scale changes in habitats (Bakker et al. 1996) e.g., in species composition or relative species abundances.

Taking into consideration that the majority of the 92/43/EEC Directive Annex I habitat types are defined by vegetation syntaxa, the collection of phytosociological data corresponding to vegetation syntaxa, habitat types, GHCs and LCCS units, is recommended in modern habitat monitoring methodological frameworks (Bunce et al. 2008). This information will be recorded from a number of plots representing various habitat types in the study sites. Additionally, selected environmental (e.g., soil data) and ecological (e.g., landscape context) variables will also be recorded during the field survey, according to the standardized protocols described below.

With the overall objective of habitat monitoring, herein we plan how to describe and understand the state and changes in habitat-relevant aspects of biodiversity in the training and test sites of the BIO_SOS project, focusing on the collection of data and parameters for biodiversity (flora and vegetation). Sampling procedures in current projects of vegetation survey move towards standardization, as well as towards a chance to standardize plot sizes besides the sampling design itself. Nevertheless, it is difficult to propose standard plot sizes for the whole of Europe due to the high diversity of vegetation types across the continent and variable field methods traditionally used in different countries. As concerns the sampling procedure, it is recommended to implement the template for common data standards for vegetation surveys, as suggested by Mucina et al. (2000), which includes area and form of the plot, distinction of vegetation layers and use of any scale for species quantities which can be converted to the ordinal transform scale.

A representative sampling area, for each vegetation type, will be selected according to the following criteria: (1) plot size large enough to capture the local species pool adequately, (2) relatively constant environmental and ecological conditions, and (3) homogeneous vegetation. Plot sites in the field will be positioned in vegetation stands that are relatively homogeneous in terms of structure, species composition, and environment, so that variation is minimized within and maximized between plots (van der Maarel 2005, Dengler et al. 2008, Alexandridis et al. 2009); all the sample plots will be selected within all the GHCs and the LCCS unit inside and outside the Natura 2000 sites, following the sampling strategy described in section 3.8 of the current deliverable. If possible, this sampling should also try to identify the locations and thus to replicate a similar sampling scheme performed in previous decades, so as to provide some insight on the biodiversity changes occurring in the meantime.

Regarding plot sizes, and following Chytrý and Otýpková (2003), we recommend to adopt four (4) plot sizes as standards for all BIO_SOS vegetation samples (phytosociological relevés), which seem to fit closest to the established tradition:

4 m^2 – All types of aquatic vegetation and low terrestrial vegetation
16 m^2– Most types of herbaceous vegetation
50 m^2– Shrub vegetation
200 m^2 – Boreal, temperate and Mediterranean woodlands.

Plot size for tropical natural and semi-natural vegetation is not being specified as yet and will be determined based on field trials in the tropical sites.

Ideally, these plots should always have a square shape, but in instances of linear habitat types (e.g., riparian forests) an elongated, rectangular shape with similar area will be adopted.

At the end, all vegetation patches in selected landscape mosaics will have been classified into one of the habitat types according to the habitat legend adopted in the project (for inter-calibration and mapping purpose) and will have been qualified according to the ecological and environmental data provided by all plant species occurring within it (for condition and trend evaluation). The total vegetation dataset will be stored in a database managed by the TURBOVEG software package and the vegetation samples (phytosociological relevés) will be classified into discrete community types, i.e. plant communities corresponding to the habitat types occurring in each of the study areas, as well as to the habitat legend (habitat types, GHCs, LCCS) adopted in the project for inter-calibration with remote sensing data.

Floristic composition of sample plots will be used as a source of indicators of environmental condition and habitat quality and hence can provide early signals of changing environmental conditions (Mueller-Dombois & Ellenberg 1974; ; Pignatti et al. 1996). In order to do so, detailed knowledge about both ecological characteristics of the species and their distribution and abundance in specific landscapes is necessary and will accompany each species. Plant species diversity will be studied in order to detect qualitative and quantitative changes in the vegetation. Some parameters will be attributed to each plant species that describe their ecological state in relation to environmental factors, their abundance in the landscape, as well as their establishment and adaptation to different climatic conditions. For each plant species, a number of indices will be evaluated, such as: ecological indicator values (Ellenberg, 1974); life forms, which are also good categories as habitat qualifiers (Raunkiaer, 1934; Bunce et al. 2008); life strategies (e.g., Grime, 1974, 2001); chorological type (i.e. geographical distribution); abundance in the landscape (Hoffman, 1998); time of establishment; and hemerobic level.

For each habitat type, a specific set of plant species, which are highly representative of the specific habitat type (e.g., characteristic species; Ellenberg, 1988, typical species; European Commission, 2006), are good indicators for favourable habitat quality, e.g., by indicating presence of a wider group of species with specific habitat requirements and/or are sensitive to environmental changes, will be selected and evaluated/monitored with regard to their spatial distribution within plant communities. As species distribution, abundance and function within a habitat vary geographically, the “typical species” of a specific habitat type are often not constant throughout the natural range of that habitat type in the EU, or even in any one country. The “typical species” are therefore better defined at regional or national level for the purpose of assessing conservation status of a habitat type.

Phenological qualifiers

Phenology is the study of periodic biological events as influenced by the environment. When developing a sampling method, it is necessary to consider the seasonal variations occurring in each vegetation-habitat type in a given geographic / environmental context. Vegetation phenology follows the seasonal variation of a large number of environmental factors, such as temperature, radiation, precipitation, etc. As a general rule, in mid and high latitudes, with vegetation-rest in winter and an active growing period in summer, air temperature has the greatest influence on phenology (Schwartz, 2003). This is especially true for spring phenological phases. In the Mediterranean, summer drought also plays an important role in the phenological cycle of many vegetation types. Moreover, for many types of aquatic or flooded vegetation, seasonal fluctuations in water levels play a crucial role in their phenology.

Many plant communities have distinct seasonal peaks of growth and flowering activity and different components of the vegetation often grow at different times of year. An extreme example of this phenomenon refers to annual grassland where live vegetation is present for only part of the year. The seasonal variation can markedly affect spectral reflectance, so that the satellite images are used to complement traditional methods for phenological studies on a large or global scale (Reed et al, 1994; Studer et al, 2007). Seasonal changes can be used to differentiate between herbaceous vegetation and woody vegetation or among different woody (or herbaceous) vegetation types with different phenological patterns. Therefore, the acquisition of satellite images in BIO_SOS will have to take account of seasonal variations of each vegetation-habitat type (see deliverable D4.4). The seasonal course of reflectance for the various types of vegetation can be useful to achieve a better discrimination.

The phenological spectrum of the species occurring in the sample plots of the selected communities-habitat types could be calculated by applying syn-pheno-ecological methods (Dierschke 1994). Phenological phases of a set of “typical” species, selected for each habitat type, will be recorded according to a BBCH (Biologische Bundesanstalt, Bundessortenamt, Chemische Industrie) code (Meier

1997). The pheno-ecological stages constitute the basis for the seasonal mapping of the sites. Pheno-ecological maps will be produced and compared for the same habitat type in the different test sites along the different partners of the project.

Environmental and management qualifiers

Habitat monitoring always involves collecting additional information on internal properties of habitat patches such as habitat quality (e.g., naturalness, degradation, pollution etc.), environmental parameters (soil type, climate) and potential drivers and pressures (land use, human impact). For the habitat patches, located within the GHC large samples, at least two additional types of habitat qualifiers will be collected: environmental and management qualifiers (Bunce et al. 2008). The collection of environmental and management qualifiers will take place when taking phytosociological relevés using sample plots inside the large samples implemented for GHC mapping, but also outside them.

Environmental qualifiers will include variables such as land use types, geological structure, soil formation and type, slope, aspect and altitude of the plots, to be recorded during the field survey (Berberoglu et al. 2004). The assessment of impact and land use is an important feature for vegetation and conservation status of habitat types, and the recording of management qualifiers will therefore be performed in all the representative vegetation types (plant communities) within all the GHCs and the LCCS units. Therefore, both land use (livestock grazing, which livestock, mowing, woodcutting, etc.) and intensity will be assessed. Land use intensity may be subdivided into the categories ‘low’, ‘moderate’ or ‘high’. A shift of habitats towards higher human pressure conditions can result in considerable loss of indigenous species, so the collection of data on habitat qualifiers related to land management is crucial for adequate (and predictive) habitat monitoring.

The recording of conservation management qualifiers will include data on the type of management and evidence of whether management is active, recent or distant, indicative of abandonment (Bunce et al. 2008). The Conservation Status Assessment of the habitat types serves as a suitable basis for the recognition of negative trends at current management and environmental conditions. The general scheme for the conservation status of habitat types is composed of three parameters: Species Inventory (habitat-specific species), Habitats and Structures (habitat structures typical of the habitat type), Impacts and three value categories or status degrees (A, B and C), which are based on the EU Directive 97/266/EG. Value categories are estimated for each of the three parameters and then combined to a total value. From the calculations made on the basis of an algorithm, the conservation status of the habitat type, as expressed in the respective area unit, is at one of the following levels: A: excellent conservation status, B: good conservation status, C: conservation status restricted or average (Dimopoulos et al. 2005, 2006).

5.2.3 Animal communities surveys

The procedure for animal communities surveys focuses on four taxonomic categories, birds (Aves), butterflies (Lepidoptera), grasshoppers (Orthoptera) and soil beetles (Coleoptera), which are known to be excellent environmental and diversity indicators in managed grasslands (Baldi *et al.* 1997, Lövei et al. 1996, McGeoch 1998, New 1997).

Animal surveys will be carried out within each 100 m radius sample unit either by means of the line intercept method, e.g., for Insecta, or the point count method, e.g., for Aves, (Bibby et al. 1992).

Species richness and abundance of selected taxa will be recorded, then it would be possible to estimate information about community structure indices (e.g., diversity), species habitat selection and distribution.
Breeding birds will be counted in the central point of the GHC sample unit, recording all birds heard or seen in a 15 min period (Bibby et al. 1992). Each count will be conducted in early morning and repeated during the breeding season.
All insects will be collected along 50-100 m long linear transects.
Adults of Lepidoptera and Orthoptera will be collected walking along transects and using a butterfly-net (Pollard, 1979) during flying season and with good weather conditions.
Soil Coleoptera will be collected using a pitfall-trap in each habitat patch.
Each transect will be stratified on the basis of vegetation structure, so that local animal data can be referred to local habitat conditions.

5.3 Pressures

5.3.1 Soil degradation

Soil degradation in the area is due to human activities including most cropland and pasture, soil reclamation to agriculture using rock fragment crushing and/or removal and land levelling, rill and gully erosion, surface mass movements. In these cases the new soil chemical and physical characteristics will be determined and differences with respect to preserved soils defined in order to understand whether habitat elements will be positively/negatively affected. Soil sealing due urban expansion will also be examined in this case as loss of soil. Based on the amount of analysis needed and data type necessary for a decent characterization of the differences it is foreseeable the use of pedotransfer functions after a calibration using ad hoc measurements and existing soil data. Potential degradation shall be considered but it needs to be evaluated in the context of flow (water and sediment) connectivity at landscape scale, within hydrological basins. Among the main effects of a changed connectivity is the change of local catchment which can modify water distribution at detailed scale. The connectivity part can be developed and expanded in cooperation with 4.3.2 and 4.3.3. it will be based on the methodology developed by Borselli et al. (2008), adapted to the local situation and updated on the basis of more recent research results.

5.3.2 Local pattern of land use and land abandonment

The general protocol for land use changes involves the use of spatial pattern analysis tools to assess changes from pre-existing LC time series as well as from novel data to be collected during the project (e.g., GHC maps from ¼ km^2 samples). Two specific types of analyses will be conducted, one for land abandonment processes, and a second one for scenario-based projections of management options.

For land abandonment, time series of land cover and of indicators of ecosystem function (e.g., those related to phenology) will be used to identify those areas where abandonment is more likely to be occurring. These will be submitted to detailed GHC mapping and also to socio-ecological surveys aimed at evaluating local patterns of abandonment in space and time. Previous models of vegetation dynamics (Honrado 2003) will be used to "convert space into time", linking vegetation types to dynamic processes and to "time since changes in disturbance regimes. The several stages of succession will then be surveyed for floristic and structural data (see section 5.2).

Ecological outcomes of scenario-based LC projections related to alternative management options can be assessed under a generally similar protocol, including (i) the implementation of scenarios in the form of alternative future LC/GHC maps, (ii) the calibration (and validation) of image classifications targeted at all relevant LC/GHC classes, (iii) the field survey of ecological features (see section 5.2) of all relevant LC/GHC classes occurring in test sites or neighbouring areas, and (iv) the simulation of ecological outcomes from distinct future management options.

A specific protocol for wildfires and related LC changes might be developed and implemented, including: (i) the analysis of wildfire time series and the identification of "hot spots"; (ii) wildfire risk modelling and forecast; (iii) field mapping of small wildfires, not captured by national databases; (iv) the establishment of successional models; (v) the survey of floristic and structural data along fire regime gradients; and (vi) the analysis of fine ecological impacts of wildfires in the landscape context.

5.3.3 Invasive species

For **alien invasive species**, the procedure includes the use of models for individual species (Vicente et al. 2011) and for landscape invasibility by multiple species (Vicente et al. 2010). Pre-existing and new distribution data (from EO images and from field surveys) will be used to calibrate models and project the potential distribution of the most threatening invaders (e.g., *Acacia dealbata*). A ranked set sampling design based on extent of invasion can then be used to select a subset of ¼ km^2 GHC samples where floristic and structural data are to be collected in order to evaluate the impacts of invasion on habitats and on native biodiversity.

5.3.4 Landscape pattern analysis

In order to accomplish WP6 task 6.2 activity 1, an analysis of landscape pattern (LP), i.e. the spatial arrangement of land cover types or the landscape configuration (structure and composition) (D6.2), will be carried out (e.g., for the 1 km^2 sample set identified in BIO_SOS sites). The grain at which the analysis will be performed will be defined based on the nominal scale of the LCCS maps provided by WP5.
With regard to the procedures for landscape pattern analysis (LPA), a set of indices will be identified and computed, which are referred to spatial attributes at patch, class landscape levels. Such indices consist in landscape pattern metrics developed for categorical maps and aimed at the characterization of the geometric and spatial (toplogical) properties of categorical map patterns represented at a single scale (grain and extent).
These indices are usually employed in Landscape ecological studies (Forman 1995; Franklin e Forman, 1987; Haines–Young e Chopping 1996; McGarigal *et al.* 2002) and reputed as able to capturing the modes ant the trends of spatially explicit changes in a given area (refer to D6.2 for more detail). A set of indices is required as it is well known that neither an individual index, nor a single gradient (e.g., derived from a Principal Component Analysis –PCA– performed on a set of indices) can adequately describe landscape configuration (McGarigal and McComb 1995). In addition to such discrete approaches also the application of fragmentation measures based on continous data (e.g., NDVI or similar) will be tested in order to be able to capturing within patch variability in environmental heterogeneity. Among the possible approaches (D6.2) reference will be made to those relying on indices of local spatial autocorrelation (e.g., Seixas 2000, Pearson 2002, Read and Lam 2002, Southworth et al. 2004).
Attention has to be paid to the assessment of habitat relative amount, fragmentation *per se* (Fahrig 2003, i.e. independently than habitat loss), habitat vs non-habitat contrast, and landscape heterogeneity in order to explore the existence of 1) an habitat fragmentation gradient between protected, partially protected and non protected areas; 2) the relations between habitat fragmentation and landscape relative heterogeneity; 3) test the relations between community attributes (dependent variables) at the local scales to independent variables at the class/landscape levels (section 5.2.2).
The set of indices suggested is the same as those in McGarigal and McComb (1995). However the inclusion of those indices related to core area has to be considered carefully, due to the difficulty of objectively identifying edge width for each class. To this set the "effective mesh size" (Jaeger, 2000) index should be added. This index is proven to a) monotonically decrease with increasing fragmentation and be consistent throughout the phases of the fragmentation process as defined by Jaeger (2000) based on Forman (1995); b) mathematically "intensive", meaning that it can be interpreted as quantifying an intrinsic landscape feature (Jaeger, 2000); c) mathematically "area proportionately additive", meaning that it is suitable for

comparing fragmentation of regions of different extent and for assessing the influences of a part of a region to the fragmentation of the whole region (Jaeger, 2000). The indices will be statistically treated as in McGarigal and McComb (1995) in order to rank samples according to a fragmentation gradient.

For the purposes of WP6 task 6.2 activity 2, the described LPA protocol will be iterated for randomly selected (e.g., ranked set sampling) habitat patches falling within 1 km^2 samples . Here the scale will be defined based on both the extent of the habitat patch and the grain allowed by VHR sensors, and reference will be made to the most detailed level of the LCCS taxonomy available.

In the forthcoming months the opportunity of comparing (for the 1km^2 ranked set of samples) the LPA protocol to the one adopted in EBONE (EBONE D5.3, Estreguil and Caudullo, 2010) will be evaluated. Such procedure is based on the combination of maps derived from the implementation of a mathemathical morphology based method providing a standard and unambiguous pixel level spatial pattern classification for a focal class (Soille and Vogt, 2009) with maps of the landscape mosaic index (Wickham and Norton 1994, Riitters et al., 2000 and 2009), providing the (pixel level) landscape context (e.g., natural, agricultural, urban) of a focal habitat class. The resulting similarity index (EBONE ID5.3, Estreguil and Caudullo, 2010) is expected to quantify proportion of edges in an anthropogenic (agriculture, urban) or natural-context provide a proxy for and therefore landscape permeability quality.
This possibility is presently beyond the scope of the BIO_SOS project, but it might provide the opportunity to test at more detailed scales procedures envisaged for the continental scale, thus representing a challenge to BIO_SOS.

6. Appendix 1 - List of information primitives for a semantic network representation of the world model

A semantic network representation of the prior geospatial knowledge about the 2-D thematic class models consists of the following information primitives.

a. A discrete and finite set of target 2-D geospatial thematic classes as network nodes provided with attributes. According to Congalton and Green, the discrete and finite set of target 2-D thematic classes must be (Congalton et al. 1999):
    - Mutually exclusive (i.e., crisp) thematic maps are generated. In other words, each mapped area falls into one and only one category.
    - Totally exhaustive (i.e., each mapped area is assigned with a semantic label. In practice, this condition implies that class 'outliers' must be dealt with explicitly by the 2-D thematic class taxonomy. It is noteworthy that the definition of a rejection rate is a well-known objective of any RS image classification system, e.g., refer to (Swain et al. 1978; p. 185). Nonetheless, in RS common practice image classifiers are often applied without any outlier detection strategy.

Mutually uncorrelated attributes of the discrete and finite set of target 2-D thematic classes are partly inherited and in part transformed from the attributes of the sensor-independent 3-D thematic classes (refer to this text above). They are listed below (Baraldi et al. 2010).

i. A hierarchical class index (numerical identifier).

ii. A class name and acronym.

iii. An (unequivocal) description/explanation/definition in terms of (2-D) appearance properties in the 2-D RS image domain. This description can be accomplished by a combination of surface type attributes (e.g., tree percent cover > 60% and tree height > 2 m and mixture of forest types none of which exceeds 60% of landscape).

iv. *Pictorial* (appearance, visual) attributes belonging to the following taxonomy.

   I. Locational (spatial) properties.
   - Non-oriented Minimum Enclosing Rectangle (MER).
   - Oriented MER (as a function of the convex hull).

   II. *Photometric* properties (chromatic and achromatic).
   - Min, Max, Mean, Standard deviation.
   - Average contrast of a region along its boundary.

   III. *Geometric and shape properties*.
   - Area.
   - Perimeter length.
   - Angle of the oriented MER.
   - Compactness (proportional to a scale-invariant area over perimeter ratio) (Nagao et al. 1980).
   - Rectangularity (Nagao et al. 1980).
   - Elongatedness (Nagao et al. 1980).
   - Straightness of boundaries (Nagao et al. 1980).

- List *of skeleton endpoints with attributes: position, angle, inter-endpoint distance (Shackelford et al. 2003).*

IV. *Morphological properties*.
- Top-hat of opening (bright object over dark background) (Baraldi et al. 2010).
- Top-hat of closing (dark object over bright background) (Baraldi et al. 2010).

V. Textural properties.
- Period.
- Orientation.
- Multi-scale Gabor wavelet-based 3° order statistics: (i) contrast, (ii) entropy / energy $\in$ [0, 1] (Baraldi et al. 1996).
- Autocorrelation (estimated from a Differential Morphological Profile (Pesaresi et al. 2001).

b. Spatial relations, either topological or non-topological, and non-spatial relations between classes of 2-D objects as edges (links, arcs) in the network provided with attributes.

7. Appendix 2 - RS image constraints in thematic mapping

When a thematic map is generated from RS imagery, the selection of RS images requires the transformation of the thematic map project requirements into RS image constraints defined as follows.

1. Geographic area(s) of interest (in terms of geographic lat-long coordinates) within the project target surface area.
2. Spectral resolution.
3. Spatial resolution.
4. Radiometric resolution (typically, 1 byte, i.e. 256 gray levels).
5. Temporal resolution (observation, i.e. revisit, frequency, e.g., daily, weekly, number of times per year).
6. Instantaneous coverage: swath width and acquisition length.
7. Observation timing (e.g., seasonal considerations).
8. Duration of each observation period (in days, weeks, months or years).
9. Required spatial quality = Co-registration/orthorectification. Multi-temporal RS image co-registration is mandatory in all classification and change detection techniques. A quantitative measure of co-registration quality is the root mean square (RMS) of the Euclidean distance between the location of ground control points (GCPs) in image pairs. For example, due to the pixel-by-pixel nature of the change detection analysis, it is recommended that the RMS error between any two date images should not exceed 0.5 pixels. In addition, a registration accuracy < 1/5 of a pixel is required to achieve a change detection error < 10% (Baraldi et al. 2010).
10. Geo-coding (geographic projection), e.g., UTM.
11. Required radiometric quality = Radiometric calibration = Transformation of digital numbers into a community-agreed radiometric unit of measure (e.g., TOARD, TOARF, SURF, See Section 2). This is a necessary, although not sufficient condition, for RS data to be processed by an automatic system, see Section 2 (Baraldi et al. 2006; Baraldi, 2009).
12. Required image enhancement in terms of topographic correction, bidirectional reflectance function (BRDF) effect removal, etc.

8. References

8.1 References to PART 1

Baraldi A., Puzzolo V., Blonda P., Bruzzone L., Tarantino C., Sept. 2006. Automatic spectral rule-based preliminary mapping of calibrated Landsat TM and ETM+ images, *IEEE Trans. Geosci. Remote Sensing*, vol. 44, no. 9, pp. 2563-2586

Baraldi A., Durieux L., Simonetti D., Conchedda G., Holecz F., Blonda P., March 2010. Automatic spectral rule-based preliminary classification of radiometrically calibrated SPOT-4/-5/IRS, AVHRR/MSG, AATSR, IKONOS/Quickbird/OrbView/GeoEye and DMC/SPOT-1/-2 imagery – Part I: System design and implementation, *IEEE Trans. Geosci. Remote Sensing*, vol. 48, no. 3, pp. 1299 – 1325

Baraldi A., Durieux L., Simonetti D., Conchedda G., Holecz F., Blonda P., March 2010. Automatic spectral rule-based preliminary classification of radiometrically calibrated SPOT-4/-5/IRS, AVHRR/MSG, AATSR, IKONOS/Quickbird/OrbView/GeoEye and DMC/SPOT-1/-2 imagery – Part II: Classification accuracy assessment, *IEEE Trans. Geosci. Remote Sensing*, vol. 48, no. 3, pp. 1326 – 1354

Baraldi A., Jan 2011. Fuzzification of a crisp near real-time operational automatic spectral rule-based decision-tree preliminary classifier of multi-source multi-spectral remotely-sensed images," *IEEE Trans. Geosci. Remote Sensing,* accepted for publication

Baraldi A., Wassenaar T., Kay S., Sept. 2010. Operational performance of an automatic preliminary spectral rule-based decision-tree classifier of spaceborne very high resolution optical images*, IEEE Trans. Geosci. Remote Sensing,* vol. 48, no. 9, pp. pp. 3482 - 3502

Baraldi A., June 2009. Impact of radiometric calibration and specifications of spaceborne optical imaging sensors on the development of operational automatic remote sensing image understanding systems, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 2, no. 2, pp. 104-134

Baraldi A., Bruzzone L., Blonda P., Apr. 2005. Quality assessment of classification and cluster maps without ground truth knowledge, *IEEE Trans. Geosci. Remote Sensing*, vol. 43, no. 4, pp. 857-873

Baraldi A., 2011. Vision goes symbolic without loss of information within the preattentive vision phase: The need to shift the learning paradigm from Machine-Learning (from examples) to Machine-Teaching (by rules) at the first stage of a two-stage hybrid remote sensing image understanding system, Part I: Introduction (to be continued), in Earth Observation, InTech Open Access Publisher, ISBN: 978-953-307-655-3.

Baraldi A., 2011. Vision goes symbolic without loss of information within the preattentive vision phase: The need to shift the learning paradigm from Machine-Learning (from examples) to Machine-Teaching (by rules) at the first stage of a two-stage hybrid remote sensing image understanding system, Part II: Introduction (continued), novel developments and conclusions, in Earth Observation, InTech Open Access Publisher, ISBN: 978-953-307-655-3.

Baraldi A., Parmiggiani F., May 1996. Combined detection of intensity and chromatic contours in color images, *Optical Engineering*, vol. 35, no. 5, pp. 1413-1439,

Beauchemin M., Thomson K., 1997. The evaluation of segmentation results and the overlapping area matrix, *Int. J. Remote Sensing,* vol. 18, no. 18, pp. 3895-3899

Bruzzone L., Persello C., July 2009. A novel context-sensitive semisupervised SVM classifier robust to mislabeled training samples, *IEEE Trans. Geosci. Remote Sensing*, vol. 47, no. 7, pp. 2142-2154

Bruzzone L., Persello C., 2010. A novel protocol for accuracy assessment in classification of very high resolution imagery, *IEEE Trans. Geosci. Remote Sensing,* Vol. 48, No. 3, pp. 1232-1244

Bunce R.G.H., Bogers M.M.B., Roche P., Walczak M., Geijzendorffer I.R., Jongman R.H.G., Manual for Habitat Surveillance and Monitoring and Vegetation in Temperate, Mediterranean and desert Biomes, European Biodiversity Observation Network (EBONE), Alterra-EBONE_Handbook_v20110131.doc.

Congalton R. G., Green K., 1999. *Assessing the Accuracy of Remotely Sensed Data*, Lewis Publishers: Boca Raton

Di Gregorio A., Jansen L., Land Cover Classification System (LCCS): Classification concepts and user manual, FAO Corporate Document Repository, [Online]. Available: http://www.fao.org/DOCREP/003/X0596E/X0596e00.htm

Foody G. M., 2002. Status of land cover classification accuracy assessment, *Remote Sensing of Environment*, vol. 80, pp. 185-201

GEO/CEOSS, Sept. 2008. A Quality Assurance Framework for Earth Observation, version 2.0, [Online]. Available: http://calvalportal.ceos.org/CalValPortal/showQA4EO.do?section=qa4eoIntro

Growe S., June 1999. Knowledge based interpretation of multisensor and multitemporal remote sensing images, *International Archives of Photogrammetry and Remote Sensing*, Vol. 32, Part 7–4–3 W6, Valladolid, Spain, 3–4. [Online].
Available:
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.18.3343&rep=rep1&type=pdf

Gutman G. et al., 2004. *Land Change Science*, Eds., Dordrecht, The Netherlands: Kluwer Academic Publishers

Liang S., 2004. *Quantitative Remote Sensing of Land Surfaces*. Hoboken, NJ: J. Wiley and Sons

Biging G.S., Colby D.R., Congalton G., 1999. Sampling Systems for Change detection Accuracy Assessment. In *Remote Sensing Change Detection: Environmental Monitoring Methods and Applications*. Lunetta R.S., Elvidge C.D. Eds., London, UK: Taylor and Francis.

Maniates D., Mollicone D., 2010. Options for sampling and stratification for national forest inventories to implement REDD+ under the UNFCCC, *Carbon Balance and Management*, vol. 5, no. 9, pp. 1-14

Marr D., 1982. Vision. New York: Freeman and C.

Matsuyama T., Shang-Shouq Hwang V., 1990. *SIGMA – A Knowledge-based Aerial Image Understanding System*. New York: Plenum Press

Nagao M., Matsuyama T., 1980. *A Structural Analysis of Complex Aerial Photographs*. New York: Plenum Press

Page-Jones M., 1988. *The Practical Guide to Structured Systems Design*. Englewood Cliffs, NJ: Prentice-Hall

Pakzad K., Bückner J., Growe S., 1999. Knowledge Based Moorland Interpretation using a Hybrid System for Image Analysis, Proc. International Society for Photogrammetry and Remote Sensing (ISPRS) Conf., Munich, Germany, Sept. 8-10, [Online]. Available: http://www.tnt.uni-hannover.de/papers/view.php?ind=1999&ord=Authors&mod=ASC

Pesaresi M., Benediktsson J. A., Feb. 2001. A new approach for the morphological segmentation of high-resolution satellite imagery, *IEEE Trans. Geosci. Remote Sens*., vol. 39, no. 2, pp. 309–320

SEVENTH FRAMEWORK PROGRAMME, THEME [SPA.2010.1.1-04]: Stimulating the development of GMES services in specific areas], "BIOdiversity Multi-

Source Monitoring System: from Space TO Species", Grant agreement no: 263435
Shackelford K., Davis C. H., Oct. 2003. A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas, *IEEE Trans. Geosci. Remote Sensing*, vol. 41, no. 10, pp. 2354 – 2363
Stehman S. V., Czaplewski R. L., 1998. Design and analysis for thematic map accuracy assessment: Fundamental principles, *Remote Sens. Environ.*, vol. 64, pp. 331-344
Stehman S. V., 1997. Selecting and interpreting measures of thematic classification accuracy, *Remote Sens. Environ.*, vol. 62, pp. 77-89
Swain P. H., Davis S. M., 1978. *Remote Sensing: the Quantitative Approach*. New York: McGraw Hill

8.2 References to PART 2

Amoruso N., Tarantino C., Baraldi A., Blonda P. 2009 Two-stage stratified classification of Quickbird images: automatic olive groves recognition and characterization. *International Conference on Space Technology (ICST*), August 24-26, Thessaloniki , Greece,.
Baraldi A. Bruzzone L., Blonda P. 2005 Quality assessment of classification and cluster maps without ground truth knowledge *IEEE Trans. Geosc. Remote Sensing*, vol. 43, no. 4, pp. 857-873, Apr. 2005.
Sanz R.T. 2006 Texture orientation and period estimator for discriminating between forests, orchards, vineyards, and tilled field. *IEEE Trans. on Geosc. and Remote Sensing*, vol.44, no.10, pp. 2755-2760, Oct. 2006.
Benediktsson J.A., Pesaresi M., Arnason K., 2003 Classification and feature extraction for Remote Sensing Images from urban areas based on morphological transformation. *IEEE Trans. Geosc. Remote Sensing*, vol. 41, no. 9, pp. 1940-1949, Sept 2003.
Soille P,. Pesareis M., 2002 Advanced in mathematical morphologies applied to Geoscience and Remote Sensin. *IEEE Trans. Geosc. Remote Sensing*, vol. 40, no. 9, pp. 2042-2055, Sept . 2002.

8.3 References to PART 3

Alexandridis TK, Lazaridou E, Tsirika A, Zalidis GC, (2009). Using Earth Observation to update a Natura 2000 habitat map for a wetland in Greece. *Journal of Environmental Management* 90 (7): 2243–2251.
Bakker JP, Olff H, Willems JH, Zobel M (1996). Why do we need permanent plots in the study of long-term vegetation dynamics? *Journal of Vegetation Science* 7: 147-156.
Baldi A., Kisbenedek T., (1997). Orthopteran assemblages as indicators of grassland naturalness in Hungary. Agriculture, Ecosystems and Environment 66: 121-129
Barve, N., Kiran, M.C., Vanaraj, G., Aravind, N.A., Rao, D., Shaanker, R.U., Ganeshaiah, K.N., Poulsen, J.G. (2005) Measuring and mapping threats to a wildlife sanctuary in southern India. Conservation Biology, 19, 122-130.
Bibby C.J., Burgess N.D., Hill D.A. 1992 *Bird census techniques*. British Trust for Ornithology - The Royal Society for the Protection of Birds. London
Borselli L., Cassi P., Torri D., 2008. Prolegomena to sediment and flow connectivity in the landscape: A GIS and field numerical assessment. Catena 75, 268–277 DOI-10.1016 J.Catena.2008.07.007.

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., Kovárˇ 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 DOI 10.1007/s10980-007-9173-8

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

Chytrý M, Otýpková Z (2003). Plot sizes used for phytosociological sampling of European vegetation. *Journal of Vegetation Science* 14: 563-570.

Cochran, W.G., and G.M. Cox. 1957. *Experimental Designs*. John Wiley & Sons, New York.

Costanza J, Moody A, Peet RK (2011) Multi-scale environmental heterogeneity as a predictor of plant species diversity. Landscape Ecology 26:851-864

Crist TO, Veech JA, Gering JC, Summerville KS (2003). Partitioning Species Diversity across Landscapes and Regions: A Hierarchical Analysis of Alpha, Beta, and Gamma Diversity. The American Naturalist 162(6): 734-743.

Dengler J, Chytry M, Ewald J (2008) *Phytosociology*. Encyclopedia of Ecology (eds S. E. Jørgensen & B. D. Fath), pp. 2767-2779. Elsevier, Oxford.

DeVries, P. G. 1979. Line intersect sampling statistical theory, applications, and suggestions for extended use in ecological inventory. In: *Sampling from biological populations* (R. M. Cormack, G. P. Patil & D. S. Robsen, eds.) p. 1-70.

Dierschke H (1994). *Pflanzensoziologie.* Verlag Eugen Ulmer, Stuttgart.

Dimopoulos P, Bergmeier E, Fischer P (2006). Monitoring and Conservation Status Assessment of Habitat Types in Greece: fundamentals and exemplary cases. *Annali di Botanica* V: 7-20.

Dimopoulos P, Bergmeier E, Fischer P (2006). Natura 2000 Habitat Types of Greece evaluated in the light of distribution, threat and responsibility. *Biology and Environment* 106B (3): 175-187.

Eberhardt, L. L. 1978. Transect Methods for Population Studies. *Journal of Wildlife Management.* 42(1) 1978.

Ellenberg H (1988). *Vegetation Ecology of Central Europe.* Cambridge University Press, Cambridge – translation of 'Vegetation Mitteleuropas mit den Alpen', the 5th ed. of which appeared in 1996 at Eugen Ulmer Verlag, Stuttgart.

EPA QA/G-5S 2002 Guidance on Choosing aSampling Design for EnvironmentalData Collection for Use in Developing a QualityAssurance Project Plan United States Office of Environmental EPA/240/R-02/005 Environmental Protection Information December 2002 Agency Washington, DC 20460

Estreguil C and Caudullo G (2010) ID 5.3 (Intermediate Deliverable, July 2010): Pattern Related Measures and Indicators for Selected Sites at Varying Spatial Scales in Selected Environmental Strata EBONE http://www.ebone.wur.nl/NR/rdonlyres/298E8A25-FEB4-4E0C-B7E3-1C725D3F7006/143155/EBONED53_EstreguilandCaudullo2010_final.pdf

European Commission (2006). *Assessment, monitoring and reporting under Article 17 of the Habitats Directive: Explanatory Notes & Guidelines.* Technical Document, Final Draft – October 2006.

Fahrig L. 2003 Effects of Habitat Fragmentation on Biodiversity Ann. Rev . Ecol. Syst. 34:487-515

Forest Survey of India, (2009). India - State of the Forest Report, Goa State. 78-81

Forman R.T.T. 1995 *Land Mosaics. The ecology of Landscapes and Regions*. Cambridge University Press. Cambridge.

Franklin J.F., Forman, R.T.T. 1987 Creating landscape pattern by forest cutting: ecological consequences and principles. Landscape Ecology 1:5-18

Gilbert, R.O. 1987. *Statistical Methods for Environmental Pollution Monitoring*. Van Nostrand Reinhold, New York.

Gilbert, R.O. 1995. Ranked Set Sampling. *DQO Statistics Bulletin, Statistical Methods for the Data Quality Objectives Process*, Volume 1, Number 1. PNL-SA-26377. Pacific Northwest National Laboratory, Richland, WA.

Gould WA, Walker MD (1997) Landscape-scale patterns in plant species richness along an arctic river. Can J Bot 75:1748–1765

Grime JP (1974). Vegetation classification by reference to strategies. *Nature* 250: 26-31.

Grime JP (2001). *Plant Strategies, Vegetation Processes and Ecosystem Properties.* John Wiley & Sons Ltd, Chichester, pp. 417.

Guisan A, Broennimann O, Engler R, Vust M, Yoccoz NG, Lehmann A, Zimmermann NE (2006). Using Niche-Based Models to Improve the Sampling of Rare Species. Conservation Biology 20(2): 501-511.

Haines-Young R., Chopping M. 1996 Quantifying landscape structure: a review of landscape indices and their applications to forested landscapes. Progress in Physical Geography 20:418-445

Haines-Young RH, Barr CJ, Black HIJ, Briggs DJ, Bunce RGH, Clarke RT, Cooper A, Dawson FH, Firbank LG, Fuller RM, Furse MT, Gillespie MK, Hill R, Hornung M, Howard DC, McCann T, Morecroft MD, Petit S, Sier ARJ, Smart SM, Smith GM, Stott AP, Stuart RC, Watkins JW (2000) Accounting for nature: assessing habitats in the UK countryside. DETR, London

Honrado J (2003). *Flora e Vegetação do Parque Nacional da Peneda-Gerês.* Unpublished PhD thesis, University of Porto (in Portuguese).

Honrado J, Alves P, Lomba A, Torres J, Caldas FB (2007). Ecology, Diversity and Conservation of Relict Laurel-Leaved Mesophytic Scrublands in Mainland Portugal. *Acta Botanica Gallica* 154(1): 63-77.

Jaeger J.A.G.(2000) Landscape division, splitting index, and effective mesh size: new measures of landscape fragmentation. Landscape Ecology 15:115-130

Jongman RHG, Bunce RGH, Metzger MJ, Mücher CA, Howard DC & Mateus VL (2006) Objectives and applications of a statistical Environmental stratification of Europe. Landscape Ecology 21: 409-419.

Kammathy, R.V., Rao, A.S., Rao, R.S. (1967) A contribution towards a flora of Biligirirangan Hills. Mysore State Bulletin: Botanical Survey of India, 9

Lomba A, Pellissier L, Randin C, Vicente J, Moreira F, Honrado J, Guisan A (2010). Overcoming the rare species modelling paradox: A novel hierarchical framework applied to an Iberian endemic plant. *Biological Conservation* 143(11): 2647-2657.

Lövei G.L., Sunderland K.D. 1996. Ecology and behavior of ground beetles. Annual Review of Entomology 41: 231–256.

McGarigal and McComb 1995 Relationships between landscape structure and breeding birds in the Oregon coast range Ecological Monographs 65:235-260.

McGarigal K., Cushman S.A., Neel M.C., 2002 FRAGSTATS: Spatial pattern analysis program for categorical maps. Computer software produced by the authors at the University of Massachusetts, Amherst. Available at the following web site: www.umass.edu/landeco/fragstats.html.

McGeoch M.A., 1998. The selection, testing and application of terrestrial insects as bioindicators. Biol. Rev. 73

Metzger MJ, Bunce RGH, Jongman RHG, Mucher CA, & Watkins JW (2005) A climatic stratification of the environment of Europe. Global Ecology and Biogeography 14: 549-563.

Mode NA, Conquest LC, Marker DA (1999). Ranked set sampling for ecological research: accounting for the total costs of sampling. Environmetrics 10:179-194.

Mucina L, Schaminée JHJ, Rodwell JS (2000). Common data standards for recording relevés in field survey for vegetation classification. *Journal of Vegetation Science* 11 (5): 769-772.

Mueller-Dombois D, Ellenberg H (1974) *Aims and Methods of Vegetation Ecology*. John Wiley and Sons, New York.

Murali, K.S., Setty, R.S. (2001) Effect of weeds Lantana camara and Chromelina odorata growth on the species diversity, regeneration and stem density of tree and shrub layer in BRT sanctuary. Current Science, 80, 675-678

Murali, K.S., Setty, R.S., Ganeshaiah, K.N., R, U.S. (1998) Does forest type classification reflect spatial dynamics of vegetation? An analysis using GIS techniques. Current Science, 75, 220-227.

Neto C, Costa JC, Honrado J, Capelo J (2007) - Phytosociological associations and Natura 2000 habitats in Portuguese coastal sand dunes. *Fitosociologia* 44(2), suppl. 1: 29-35.

New T.R., 1997. Are Lepidoptera an effective 'umbrella group' for biodiversity conservation?. Journal of Insect Conservation 1:5-12

Panitsa M, Koutsias N, Tsiripidis I, Zotosa A, Dimopoulos P (2011). Species-based versus habitat-based evaluation for conservation status assessment of habitat types in the East Aegean islands (Greece). *Journal for Nature Conservation*, doi:10.1016/j.jnc.2011.04.001

Pearson, D.M. (2002) The application of local measures of spatial autocorrelation for describing pattern in north Australian landscapes. Journal of Environmental Management 64: 85–95.

Pignatti S, Ellenberg H, Pietrosanti S, (1996). Ecograms for phytosociological tables based on Ellenberg's Zeigerwerte. *Annali di Botanica (Roma)* 54: 5-14.

Pollard E. (1979). A national scheme for monitoring the abundance of butterflies: the first three years. British entomological and Natural History society. Proceedings and Transactions 12: 77-90.

Ramesh, B. R., and Menon, S. (1997). Map of Biligiri Rangaswamy Temple Wildlife Sanctuary, vegetation types and land use (Pondicherry: French Institute and Bangalore: ATREE).India Biodiversity Portal (http://indiabiodiversity.org/map?layername=lyr_160_brt_vegetation&) accessed on 18th July 2011.

Ramesh, B.R. (1989) Flora of Biligirirangan Hills. Madras University, Chennai.

Raunkiaer C, (1934). *The Life Forms of Plants and Statistical Plant Geography.* Clarendon Press, Oxford.

Read, J.M., Lam, N.S.-N. (2002). Spatial methods for characterising land cover and detecting land-cover changes for the tropics. International Journal of Remote Sensing 23: 2457-2474.

Reed BC, Brown JF, VanderZee D, Loveland TR, Merchant JW, Ohlen DO, 1994. Measuring phenological variability from satellite imagery. *Journal of Vegetation Science* 5:703–714.

Riitters K.H., Wickham J.D., O'Neill R.V., Jones K.B., Smith E.R., 2000. Global-scale patterns of forest fragmentation. Ecol Soc (formerly Cons Ecol) 4(2):3.

Riitters, K.H., J.D. Wickham, and T.G. Wade., 2009. An indicator of forest dynamics using a shifting landscape mosaic. Ecological Indicators 9:107-117

Seixas, J. (2000). Assessing heterogeneity from remote sensing images: the case of desertification in southern Portugal. International Journal of Remote Sensing 21: 2645–2663.

Soille P. and Vogt P., 2009. Morphological segmentation of binary patterns. Patterns Recognition Letters. doi:10.1016/j.patrec.2008.10.015

Southworth, J., Munroe, D., Nagendra, H. (2004). Land cover change and landscape fragmentation: comparing the utility of continuous and discrete analyses for a study area in Western Honduras. *Agriculture, Ecosystems and Environment* 101: 185-205.

Sundaram, B. (2011). Patterns and processes of Lantana camara persistence in South Indian tropical dry forests. PhD Thesis, Manipal University.

Vicente J, Alves P, Randin C, Guisan A, Honrado J (2010). What drives invasibility? A multi-model inference test and spatial modelling of alien plant species richness patterns in Northern Portugal. *Ecography* 33: 1081-1092.

Vicente J, Randin CF, Gonçalves J, Metzger MJ, Lomba A, Honrado J, Guisan A (2011). Where will conflicts between alien and rare species occur after climate and land-use change? A test with a novel combined modelling approach. *Biological Invasions* (DOI: 10.1007/s10530-011-9952-7).

Wickham, J.D. and Norton D.J., 1994. Mapping and analyizing landscape patterns. Landscape Ecology 9:7-23.

Wolfe DA (2010). Ranked set sampling. *Wiley Interdisciplinary Reviews: Computational Statistics* 2: 460-466.

Abbreviations and Acronims

ASTER = Advanced Spaceborne Thermal Emission and Reflection Radiometer
BIO_SOS = Biodiversity Multisource Monitoring System: from Space TO Species
BIOHAB = A Concerted Action of the Fifth Framework – A framework for the coordination of Biodiversity and Habitats
BRDF = Bidirectional Reflectance Function
CCW = Countryside Council for Wales
CEOS = Committee on Earth Observation Satellites
CHM = Canopy Height Model
CORINE = Coordinate Information on the Environment
DBH = Diameter at Breast Height
DoW= Description of Work
DSM = Digital Surface Model
DTM = Digital Terrain Model
EAGLE =Name of the improved version of the AISA (Airborne Imaging Spectrometer for Applications)
EBONE = European Biodiversity Observation Network
EnS = Environmental strata of Europe
ENVI = ENvironment for Visualizing Images (software)
EnZ = Environmental zones of Europe
EO = Earth Observation
EODHaM = Earth Observation Data for Habitat Monitoring
EPA= Environmental Protection Agency
ETM = Enhanced Thematic Mapper
EUNIS = European Nature Information System
GEO = Global Earth Observation
GEOSS = Global Earth Observation System of Systems
GHC = General Habitat Category
INSPIRE = Infrastructure for Spatial Information in Europe
LANDSAT = Land Satellite
LC/LU = Land Cover / Land Use
LCCS = Land Cover Classification System
LCM = Land Cover Map
LGN-6 = The Netherlands National land cover database
LiDAR = Laser Imaging Detection and Ranging
LPA = Landscape pattern analysis
MER = Minimum Enclosing Rectangle
MS = Multi Spectral
N2K = Natura 2000
NDVI = Normalized Difference Vegetation Index
NP = National Park
OAMTRX = Overlapping area matrix
PAN = Panchromatic
PANSH = Pansharpened
PCA = Principal Component Analysis
pOA = Overall Accuracy Probability
QA4EO = Quality Assurance for Earth Observation
QB = QUICKBIRD
REDD = Reducing Emissions from Deforestation and Degradation of Forests
RGB = Red Green Blue
RS = Remote Sensing
SE = Structuring Element
SIAM™ = Satellite Image Automatic Mapper™

SPOT = Système Probatoire d'Observation de la Terre
SRSRG = Stratified Random Sampling within a Regular Grid
SSS = reference Sample Set Size
TM = Thematic Mapper
TOARF = Top-of-Atmosphere Reflectance
VHR = Very High Resolution
WLS = Wildlife Sanctuary
WP = Work Package
WT = Workplan Table
WV-2 = WORLDWIEW-2