EBONE

European Biodiversity Observation Network:

Design of a plan for an integrated biodiversity observing system in space and time

D5.4: Report on inter-calibration for selected sites and execution of the statistical tests on added value

Ver 1.0
Document date: 2012-02-17
Document Ref.: EBONE-D5.4-1.0

Authors:
Mats Nilsson, Anna Allard, Sören Holm, and Håkan Olsson (SLU)
Hans Van Calster and Dirk Bauwens (INBO)

EC-FPV Contract Ref: ENV-CT-2008-212322

D5.4: Report on inter-calibration for selected sites and execution of the statistical tests on added value

Authors: Mats Nilsson, Anna Allard, Sören Holm, and Håkan Olsson (SLU)
Hans Van Calster and Dirk Bauwens (INBO)

Contents:

Integrating *in situ* data with Earth Observation data for estimating the area coverage of habitats: post-stratification

Precision and accuracy comparison between earth observation samples and in situ samples

Integrating *in situ* data with Earth Observation data for estimating the area coverage of habitats: Post-stratification

Mats Nilsson, Anna Allard, Sören Holm, and Håkan Olsson
Department of forest resource management
Swedish University of Agricultural Sciences (SLU)
Umeå Sweden

e-mail: mats.nilsson@slu.se

Version 1.0, Sept 30, 2011

Delivery to EBONE Task 5.3.1. Report on inter-calibration for selected sites and execution of the statistical tests on added value

Abstract

Integration of in situ data and Earth observation (EO) data for estimating the occurrence of different habitat or classes can be achieved using different approaches. In this study, the approach used is to post-stratify *in situ* data using existing land cover maps derived from satellite data. Photo-interpreted landscape elements and biotopes from the National Inventory of Landscapes in Sweden (NILS; http://nils.slu.se/) were used as *in situ* data. The mapped landscape elements and biotopes were classified into General Habitat Categories (GHCs). Five GHC classes were selected to exemplify how the precision of their area estimates was affected by using post-stratification, as compared to area estimates of the GHC classes based on the photo-interpreted data alone. The stratification was made using the Swedish version of Corine land cover 2000 (GSD Land Cover) which includes more classes and has a higher spatial resolution (1-25 ha minimum mapping unit depending on the class) than the European version of Corine land cover (CLC2000). The results show that the standard error was reduced for all tested GHC classes using post-stratification in comparison to the errors obtained without post-stratification. This shows the potential to derive improved area statistics of habitat categories by integrating *in situ* data with existing land cover maps.

Introduction

In many countries biodiversity is threatened due to, for example, present management practices in agriculture and forestry and increased impact of urban development and transport infrastructure. As a result, the extent and the quality of many habitats across Europe have decreased. This has forced national and international authorities to adapt their policies in order to improve the maintenance of regionally specific biodiversity. Thus, there is a need for up-to-date and accurate information describing the current stage and changes in biodiversity, both locally and across regions.

Ongoing inventory programs like the Countryside Survey in UK (http://www.countrysidesurvey.org.uk/), the National Inventory of Landscapes in Sweden (NILS, http://www.slu.se/nils), and the Spanish Rural Landscape Monitoring System (SISPARES, http://www.sispares.com/eng/) provides information needed to monitor biodiversity based on a sample of *in situ* observations. Reliable estimates of the occurrence of different habitats or categories can only be reported for relatively large areas due to the low sampling intensity in these inventories. To achieve reliable results for smaller areas than presently possible an increase of the field sample is needed. Adding enough *in situ* observations would, however, be very costly. Another possibility is to derive the information needed from combinations of Earth observation (EO) and *in situ* measurements. The additional cost for using EO data in relation to the cost for the additional field sample needed to obtain a certain estimation accuracy is low (Dees, 2006; McRoberts et al., 2006).
EO data from optical satellite sensors like Landsat TM and ETM+ have been used frequently to classify vegetation. The need for spatially consistent vegetation and land cover information has resulted in a number of efforts where vast areas have been mapped. In the US, the Gap Analysis Program (GAP) have produce regional and national land cover maps from Landsat data to be used in habitat analyses (e.g., Reese et al., 2002; Lowry et al., 2007). Other national mapping projects in the US have also produced national land cover datasets based on Landsat data (e.g., Vogelmann et al., 2001; Homer et al., 2007). In Canada, a land cover map of the forested areas has been produced using Landsat ETM+ data (Wulder et al., 2008). Another example is the European CORINE land cover program established in 1985 which have produced land cover maps with 25 ha minimum mapping unit for most countries (Perdigão and Annoni, 1997; Bossard et al., 2000).
The National Forest Inventory (NFI) in Finland introduced a multi-source forest inventory approach in the early 1990's in which maps with estimates of forest variables are derived for the entire country using and Landsat TM data (e.g., Tomppo, 1993; Tomppo, 2006). Maps with estimated forest variables have also been produced for countries like Sweden (Reese et al., 2003), Norway (Gjertsen, 2005), Ireland (McInerney et al., 2005), and the USA (Franco-Lopez et al., 2001; McRoberts et al., 2007).
When only statistics and not wall-to-wall maps describing the spatial pattern of different habitats or categories are needed, the combined use of EO data and data from sample-based inventories can provide accurate area estimates for various categories. Almost unbiased area estimates of habitats or classes can, for example, be obtained by combining EO data and *in situ* data using post-stratification. It has previously been shown that post-stratification or stratified estimation, where satellite images or classified satellite images are used as ancillary data, can be used to improve the accuracy of estimated forest characteristics in sample based forest inventories, for example National Forest Inventories (e.g., McRoberts et. al., 2002 and 2006; Nilsson et al., 2003 and 2009). It has also been shown that the used of CORINE land cover (CLC) data for post-stratification of *in situ* data from LUCAS improved the accuracy of area estimates for various coastal land cover classes (Galego and Bamps, 2008).

Objective

The objective of this study is to show whether or not the accuracy of area estimates of General Habitat Categories (GHCs) can be improved by post-stratifying the *in situ* data from a sample based inventory using existing land cover maps, compared to use *in situ* data alone. Photo-interpreted landscape elements and biotopes are post-classified into GHC classes defined by the BioHab project (Bunce et al., 2005) and post-stratified using a land cover map.

Materials and methods

Study area

A 7.8 million ha study area located east of the Scandinavian mountain range in the northern part of Sweden was used (Figure 1). The study area selected represents the entire stratum number 9 used in the National Inventory of Landscapes in Sweden (NILS). The terrain is undulating and the landscape is traversed by river valleys in NW-SE direction and is dominated by forests, intermixed with wetlands, lakes and streams. The dominant tree species in the area are, in order of proportional representation, Scots Pine (*Pinus sylvestris*), Norway spruce (*Picea abies*), and Birch (*Betula spp*.).

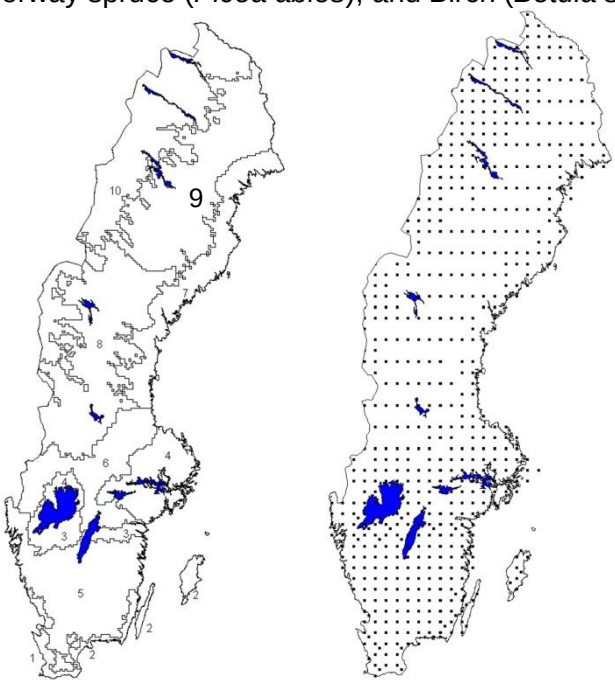

Figure 1. Location of the study area (stratum number 9) and the allocation of the 631 NILS squares.

NILS data

The National Inventory of Landscapes in Sweden (NILS; http://nils.slu.se/) program is developed to monitor and analyze conditions and trends in landscape biodiversity and land use on all terrestrial habitats across Sweden (Ståhl et al., 2010). It is based on a randomly located systematic grid of 631 permanent 5 km x 5 km squares (Figure 1) covering all terrestrial habitats in Sweden. Approximately, 120 squares are surveyed annually. The squares are inventoried with five year intervals to provide data and analyses on conditions and changes. A field inventory and interpretation of colour infrared (CIR) aerial photos are made in parallel. Both the field inventory and the photo interpretation focus on the central 1 km x 1 km square. Together the field inventory and aerial photo interpretation allows for the

analyses of various geographic scales, from large-scale landscape analyses on the spatial composition of habitats, land use and other important landscape information, to point-scale occurrence of specific species.
The *in situ* data used in this study are photo-interpreted landscape elements and biotopes from 31 of the 631 square kilometre quadrates (1x1 km) in the NILS. The photo-interpretation was carried out on 1:30 000 scale colour infrared photographs acquired from 4 600 meters flight height in the summer of 2002 and 2003. All photographs were scanned with a resolution corresponding to 0.4 meters on the ground. The reason for not using data from all NILS squares in stratum 9 was that the photo interpretation only had been completed for 31 squares.
During the photo interpretation, all landscape elements or biotopes with a size of 0.1 hectares or more are delineated. In total 87 variables are measured or interpreted for each landscape element or biotope, including information on land use, vegetation type, canopy cover, tree species, and tree height. The mapped landscape elements and biotopes were classified into BioHab General Habitat Categories (GHCs) (Bunce et al., 2005) using conversion criteria's developed within the EBONE project (Allard, 2010). The six GHC classes (groups of GHCs) presented in Table 1 were selected to exemplify how the precision of their area estimates was affected by the use of post-stratification based on an existing land cover map.

Table 1. Description of the GHCs used in the study

Class	Description	GHC numbers*
CON	Conifers, crown cover ≥ 70%	128 & 124/1282
DEC	Winter deciduous trees, crown cover ≥ 70%	1281 & 124/1281
MIX	Mixed forest. Combination: Winter deciduous trees/Conifers. Together covering ≥ 70%	12811 & 124/12811
HER/HEL	Plants that grow in waterlogged conditions	115 & 1151
HER/LHE	Leafy hemicryptophytes, broad leaved herbaceous species	116 & 1116
HER/CRY	Non-saxicolous (growing on or among rocks) bryophytes and lichens. Includes aquatic bryophytes, e.g. Spagna and Racomitrium lanuginosim	121 & 1211

* Allard (2010)

Once the mapped landscape elements had been assigned GHC classes they were converted from vector format to grid format with a 25 m x 25 m resolution, matching the resolution of the data used for stratification.

Stratification

The stratification was made using the Swedish version of CORINE land cover 2000 dataset (GSD Land Cover) which includes more classes and has a higher spatial resolution (1-25 ha minimum mapping unit depending on the class) than the European version of CORINE land cover 2000 (CLC2000). The mapping was made at a national level using classification methods ranging from visual interpretation to fully automated supervised classification (Maximum likelihood). The Landsat ETM+ data used to produce the GSD Land Cover map was provided through the European-wide Image 2000 dataset, administered by the European Environmental Agency (EEA).
In total, the GSD Land Cover Map consists of 57 classes (Engberg, 2005) of which 49 classes were apparent in the study area. These 49 classes were grouped into the 10 strata presented in Table 2. The GSD Land Cover map represents the landscape conditions in year 2000, and the grid version with a 25 m x 25 m resolution was used in the study.

Table 2. Description of strata used and the GSD Land Cover Map code defining each stratum

Stratum	Description	GSD Land cover map code*
1	Urban areas	111, 11211, 11212, 1122, 121, 122, 124
2	Sand, gravel pits, mineral extraction sites, etc	1311, 1312, 132, 1421, 331, 332, 4123
3	Arable land, pastures, etc	141,1422-1426, 211, 231, 321, 3211, 3212
4	Sea, lakes, ponds, etc	511, 5121, 5122
5	Mires	411, 4121, 4122
6	Clear-felled areas	3242
7	Young forests (tree height < 5 m)	3243
8	Deciduous forests (tree height > 5 m)	3111-3113, 322, 3241,333
9	Coniferous forests (tree height > 5 m)	31211, 312121, 312122, 3122, 3123
10	Mixed forest (tree height > 5 m)	3131-3133, 31212

* Engberg (2005)

Estimation

The area covered by each of the six selected GHC classes and corresponding standard errors were estimated both with and without the use of post-stratification. In the method used, the area covered by a specific GHC class ($\hat{Y}$) was estimated as:

$$\hat{Y} = \sum_{h=1}^{L} \hat{Y}_h = \sum_{h=1}^{L} \left(A_h \cdot \frac{\sum_{j=1}^{n_h} y_{jh}}{\sum_{j=1}^{n_h} a_{jh}} \right) = \sum_{h=1}^{L} A_h \cdot \hat{R}_h \qquad (1)$$

with $\hat{R}_h = \sum_{j=1}^{n} y_{jh} / \sum_{j=1}^{n} a_{jh}$ and where n_h is the total number of NILS squares located at least partly within stratum h; A_h is the total land area in stratum h, h=1, 2, ..., L; a_{jh} is the land area in NILS square j, stratum h; y_{ih} is the total area covered by the GHC class in NILS square j, stratum h; j is 1, 2, ..., n_h.

When estimating the variance for the area estimates of different GHC classes it was important to consider that the $\hat{R}_h$ values are correlated, which is an effect of having parts of the NILS squares that belongs to different strata. Thus, the variance has been estimated using an extended version of the standard method used for ratio estimators (e.g., Thompson, 1992) which also takes the correlation between $\hat{R}_h$ values into account (Equation 2).

$$Var(\hat{Y}) = \sum_{h=1}^{L} \sum_{k=1}^{L} A_h A_k \cdot Cov(\hat{R}_h, \hat{R}_k) \qquad (2)$$

where, with the common Taylor expansion approximation, $Cov(\hat{R}_h, \hat{R}_k)$ is estimated by:

$$Cov(\hat{R}_h, \hat{R}_k) = \frac{1}{n \cdot \bar{a}_h \cdot \bar{a}_k} \cdot \frac{\sum_{j=1}^{n} (y_{jh} - \hat{R}_h \cdot a_{jh})(y_{jk} - \hat{R}_k \cdot a_{jk})}{(n-1)} \qquad (3)$$

where the mean values $\bar{a}_h$ and $\bar{a}_k$ are calculated on all n clusters in the area (stratum 9). Equation (3) is derived in a way similar to that of the variance of a ratio estimator.

How effective the stratification was for different GHCs was evaluated using the relative efficiency (RE) which was calculated as:

$$RE = \frac{Var(\hat{Y}_{poststrat,i})}{Var(\hat{Y}_{insitu,i})} \qquad (4)$$

where $Var(\hat{Y}_{insitu,i})$ is the estimated variance for GHC class *i* based on the photo interpreted data alone, and $Var(\hat{Y}_{poststrat,i})$ is the he estimated variance for GHC class *i* obtained using post-stratification.

Results

As can be seen in Table 3, the GHC classes used were observed in most of the NILS squares. This means that the classes are quite common in the study area. It is also obvious that the forests are dominated by coniferous tree species.

Table 3. The total area of each GHC class observed on the 31 NILS squares used in the study, and the number of NILS squares in which the class occurred

	Area cover		Number of
	(ha)	(%)	NILS squares
CON	752,2	20,1	28
DEC	198,4	5,3	28
MIX	180,5	4,8	26
HER/CRY	321,5	8,6	28
HER/HEL	391,6	10,4	26
HER/LHE	379,1	10,1	28
Other	1527,8	40,7	31
Total	3751,0	100,0	31

Area estimates of the six selected GHC classes were produced, both with and without post-stratification. As can be seen in Figure 2, the area estimates are fairly similar for all GHC classes except for HER/HEL. In this case, the post-stratification yielded higher estimates than did the estimates based on photo interpreted data alone. For HER/HEL, the difference in area estimate was found to be significant between the two methods.

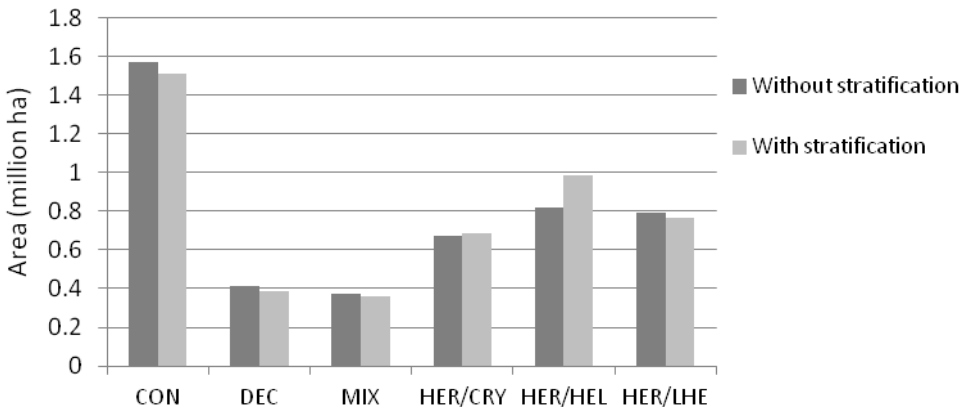

Figure 1. Area estimates of selected GHC classes, derived with and without post-stratification.

The results show that the standard error was reduced for all GHC classes using post-stratification in comparison to the errors obtained based on photo interpreted data alone (Figure 2). In terms of effectiveness, the RE of the post-stratified area estimates of the GHC classes as compared to their area estimates derived using photo interpreted data alone was found to be in the range of 1.07-2.65 [1](Table 4). The smallest improvement in terms of RE was obtained for HER/CRY and largest was obtained for DEC.

[1] An RE of 1.07 to 2.65 means that the post-stratification estimates are 7 to 265 % more efficient than the estimates derived using photo interpreted data alone (from a variance point of view).

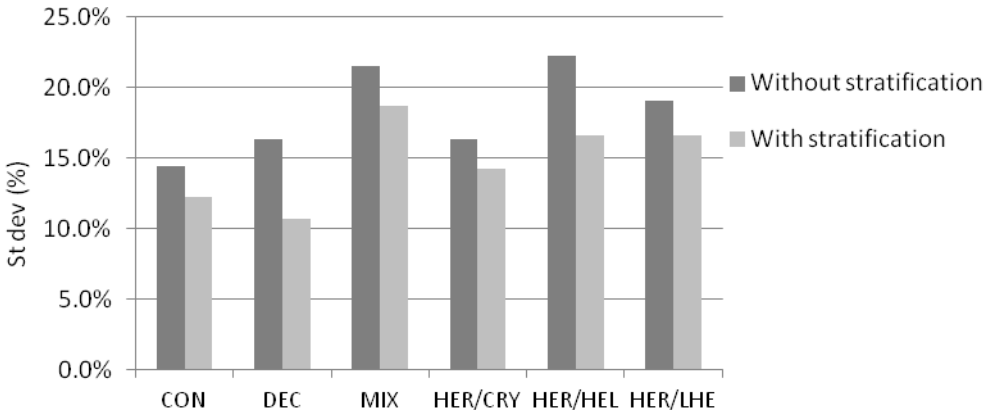

Figure 2. Standard deviation (St dev) for the estimated area of selected GHC classes, derived with and without post-stratification.

Table 4. Relative efficiency (RE) for GHC classes

GHC class	RE
CON	1.50
DEC	2.65
MIX	1.47
HER/CRY	1.07
HER/HEL	1.22
HER/LHE	1.40

Discussion

This study has shown the potential to integrate GHC classes derived from NILS and data from the GSD Land Cover map. However, the GSD Land Cover map is more detailed than CLC2000 and it was produced, to a large extent, by using automatic classification methods. The differences in resolution and the fact that GSD Land Cover was produced using automatic classification methods and CLC2000 using visual interpretation makes it difficult to draw conclusions regarding how the use of CLC2000 data for post-stratification would affect the result without further investigations. Gallego and Bamps (2008) have shown that area estimates derived directly from the CLC2000 dataset can be strongly biased, and that the estimates needs to be calibrated to give acceptable statistical results. In their study, data from CLC2000 were combined with sample data from LUCAS (land use/cover area frame survey) 2001 using post-stratification. The results show that the obtained REs were clearly above 1 for arable land classes, such as the total area of cereals which had an RE of 1.48. However, RE was found to be close to 1 for crops that cover a minor part of the total arable land. McRoberts et al (2005) have shown that classified satellite imagery, when used for stratification, can improve the precision of the estimates with little increase in cost. The precision for estimates of both forest area and volume per unit area were found to increase substantially using stratified estimation techniques. The gain in precision was higher for forest area estimates than for estimates of volume per unit area. Although the improvement in precision (RE) was higher in the study by McRoberts et al (2005), the results obtained in this study also show an increase in precision when using classified satellite images for post-stratification. Other studies have also shown that the use of classified satellite images for post-stratification of data from sample based inventories improve the precision for estimates of forest variables such as volume (e.g., McRoberts et al., 2002; Nilsson et al., 2003).

In this study, the stratification used was based on the GSD Land Cover map which has more classes and a higher spatial resolution than CLC2000. The original 49 classes were grouped into 10 strata by heuristic reasoning. It was noticed that all GHC classes were present in almost all 10 strata, although the area cover was very low in many strata. This could partially be an effect of having geometric errors, primarily in the GSD Land Cover map which was derived using Landsat ETM+ images with a positional accuracy of 25 m. The geometric errors could result in having individual photo interpreted polygons being split between

different strata. The fact that all GHC classes were apparent in most strata could also be an effect of not having grouped the GSD Land Cover map classes into strata in an optimal way. A way to improve the precision of the area estimates might be to use image classifications of the GHC classes for stratification instead of using the GSD Land Cover map. Further investigations are needed before any final conclusions can be drawn regarding how the stratification should be made to get as high precision as possible for the area estimates of the GHC classes. However, one important advantage of using products like the GSD Land Cover map or the CLC2000 map for the stratification is that they already exist. To produce new image classifications for stratification within a country or across Europe would be costly. A general conclusion is that post-stratification is an easy and straight forward method that can be used to integrate *in situ* data and EO data to derive improved area statistics for habitats or classes such as GHCs that area unbiased or at least almost unbiased. The increase in precision obtained using post-stratification also means that estimates of the area covered by different habitat classes can be presented for smaller areas than possible for estimates based on a sparse sample of *in situ* data alone without any reduction in precision. An important future research task is to test if the use of other map products than CLC2000 can improve the estimation accuracy for selected habitats. It will also be of interest to investigate how the gain in efficiency for post-stratified estimates (RE) is affected by the number of *in situ* observations used. Thus, it would be interesting for NILS to investigate how the gain in precision is affected when post-stratification is applied to both smaller and larger areas than used in this study.

Acknoweldgements

This study is carried out as a part of the EU FP7 project European Biodiversity Observation Network (EBONE, http://www.ebone.wur.nl/UK/) that addressed the development of a cost effective data collection system for biodiversity linked with extant data, both past and present, at national, regional and European levels. One of the tasks in EBONE is to provide methods that can be used to produce statistical estimates of key indicators that can be interpreted by policy makers responding to EU Directives regarding threatened ecosystems and species. We thank the NILS programme for providing the photo interpreted data used in the study.

References

Allard, A., 2010. Variables and Classifications – Conversion of the set of variables from the Swedish NILS programme into General Habitat Categotries of the European EBONE project. Swedish University of Agricultural Sciences, Dept. of Forest Resource Management, Working report 295, 43 pp.

Bossard , M. , Feranec , J. and Otahel , J., 2000. CORINE Land Cover Technical Guide — Addendum 2000, Technical Report No. 40. European Environmental Agency , Copenhagen, Denmark , 105 pp .

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., Ková, 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.

Dees, M., 2006, Kombination von Fernerkundung und Stichprobeninventur bei betrieblichen und nationalen Waldinventuren. Freiburg: Fakultät für Forst und Umweltwissenschaften der Albert-Ludwigs-Universität Freiburg. Schriftenreihe Freiburger Forstliche Forschung, Bd. 37. 194 p.

Engberg, A., 2005. Produktspecifikation av Svenska CORINE Marktäckedata (in Swedish). Lantmäteriet, Gävle, Document No SCMD-0001. Available online at: http://www.lantmateriet.se/upload/filer/kartor/kartor_och_geografisk_info/GSD-Produktbeskrivningar/SCMDspec.pdf

Franco-Lopez, H., Ek, A.R., and Bauer, M.E., 2001, Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. *Remote Sensing of Environment* 77: 251-274.

Galego, J and Bamps, C., 2008. Using CORINE land cover and the point survey LUCAS for area estimation. *International Journal of Applied Earth Observation and Geoinformation* 10: 467–475.
Gjertsen, A., 2005, Accuracy of forest mapping based on Landsat TM data and a kNN-based method. *Remote Sensing of Environment* 110: 420-430.
Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., Herold, N., McKerrow, A., VanDriel, J.N. & Wickham, J., 2007. Completion of the 2001 National Land Cover Database for the conterminous United States. *Photogrammetric Engineering and Remote Sensing* 73: 337-341.
Lowry, J., Ramsey, R.D., Thomas, K., Schrupp, D., Sajwaj, T., Kirby, J., Waller, E., Schrader, S.,Falzarano, S., Langs, L., Manis, G., Wallace, C., Schulz, K., Comer, P., Pohs, K., Rieth, W.,Velasquez, C., Wolk, B., Kepner, W., Boykin, K., O'Brien, L., Bradford, D., Thompson, B. & Prior-Magee, J., 2007. Mapping moderate-scale land-cover over very large geographic areas within a collaborative framework: A case study of the Southwest Regional Gap Analysis Project SWReGAP). *Remote Sensing of Environment* 108: 59-73.
McInerney, D., Pekkarinen, A., and Haakana, M., 2005, Combining Landsat ETM+ with field data for Ireland's National Forest Inventory – a pilot study for Co. Clare. *In* H. Olsson (Ed.): Proceedings of ForestSat 2005. Borås, Sweden, 31 May – 3 June 2005. pp. 12-16.
McRoberts, R.E., Wendt, D.G., Nelson, M.D. & Hansen, M.H., 2002. Using a land cover classification based on satellite imagery to improve the precision of forest inventory area estimates. *Remote Sensing of Environment* 81: 36-44.
McRoberts, R.E., Holden, G.R., Nelson, M.D., Liknes, G.C., and Gormanson, D.D, 2006, Using satellite imagery as ancillary data for increasing the precision of estimates for the Forest Inventory and Analysis program of the USDA Forest service. *Canadian Journal of Forest Research*, 36, pp. 2968-2980.
McRoberts, R.E., Tomppo, E.O., Finley, A.O., and Heikkinen, J., 2007, Estimating areal means and variances of forest attributes using the k-Nearest Neighbor technique and satellite imagery. Remote Sensing of Environment 111:466-480.
Nilsson M., Folving, S., Kennedy, P., Puumalainen, J., Cincinati, G., Corona, P., Marchetti, M., Olsson, H., Ricotta, C., Ringvall, A., Ståhl, G., and Tompoo, E., 2003. Combining remote sensing and field data for deriving unbiased estimates of forest parameters over large regions. *In*: Corona, P., Köhl, M., and Marchetti, M. (eds.) Advances in Forest Inventory and for Sustainable Forest Management and Biodiversity Monitoring. Kluwer Academic Publishers, Forestry Sciences pp. 19-32
Nilsson, M., Holm, S., Wallerman, J., Reese, H. and Olsson, H., 2009. Estimating annual cuttings using multi-temporal satellite data and field data from the Swedish NFI. *International Journal of Remote Sensing*, 30:5109-5116.
Perdigão , V. and Annoni , A., 1997 Technical and Methodological Guide for Updating CORINE Land Cover Database . European Communities , Luxembourg, Germany , 140 pp .
Reese, H.M., Lillesand, T.M., Nagel, D.E., Stewart, J.S., Goldmann, R.A., Simmons, T.E., Chipman, J.W. & Tessar, P.A., 2002. Statewide land cover derived from multiseasonal Landsat TM data - A retrospective of the WISCLAND project. *Remote Sensing of Environment* 82: 224-237.
Reese, H., Nilsson, M., Pahlén, Hagner, O., Joyce, S., Tingelöf, U., Egberth, M., and Olsson, H., 2003, Countrywide estimates of forest variables using satellite data and field data from the National Forest Inventory. *Ambio* 32, pp. 542-548.
Ståhl, G., Allard, A., Esseen, P.-A., Glimskär, A., Ringvall, A., Svensson, J., Sundquist, S., Christensen, P., Gallegos Torell, Å., Högström, M., Lagerqvist, K., Marklund, L., Nilsson, B., & Inghe, O., 2010. National Inventory of Landscapes in Sweden (NILS) - scope, design, and experiences from estabilshing a multiscale biodiversity monitoring system. Environmental Monitoring and Assessment [Published online 17 March 2010].
Thompson, S.K., 1992, Sampling, New York: John Wiley & Sons, pp. 59–70.
Tomppo, E. 1993. Multi-Source National Forest Inventory of Finland. *In*: Proceedings of Ilvessalo Symposium on National Forest Inventories, August 17-21, Finland. pp. 52-59.
Tomppo, E., 2006, The Finnish multi-source national forest inventory – small area estimation and map production. Forest inventory – methodology and applications. Dordrecht, The Netherlands. Springer. pp. 195-224.
Vogelmann, J.E., Howard, S.M., Yang, L.M., Larson, C.R., Wylie, B.K. & Van Driel, N., 2001. Completion of the 1990s National Land Cover Data set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. *Photogrammetric Engineering and Remote Sensing* 67: 650-662.

Wulder, M.A., White, J.C., Cranny, M., Hall, R.J., Luther, J.E., Beaudoin, A., Goodenough, D.G. & Dechka, J.A. (2008). Monitoring Canada's forests. Part 1: Completion of the EOSD land cover project. *Canadian Journal of Remote Sensing* 34: 549-562.

Contribution to WP5: Precision and accuracy comparison between earth observation samples and in situ samples

Hans Van Calster and Dirk Bauwens
INBO

Precision and power calculations for the EBONE sampling design

EBONE monitoring objectives

In order to do power calculations or to calculate the precision of parameters we need to be explicit about the most important aspect to monitor (the monitoring objective). For the EBONE project, a major goal is to be able to make stock and change estimates for the general habitat categories, using the BIOHAB field protocols (in situ samples).
In addition, we want to explore how earth observation samples compare to in situ samples regarding precision, bias and power for stock estimates (see 0).
There are two spatial scales at which estimates for stock and change are particularly relevant. First, at the European scale. Second, at the scale of biogeographical zones within Europe. Due to the procedure the sample was drawn, the sample size is allocated proportionately to the area of biogeographical zones.

EBONE sampling design

The proposed EBONE sampling scheme is inspired by the UK Countryside Survey. The sampling design is discussed in detail in Brus et al. (2011). In short, the sample is a random, but spatially balanced sample. Each sample location is a 1 km x 1 km square. Field sampling is done according to BIOHAB field protocol, hence, the result is a 1 km x 1 km map of the spatial distribution of General Habitat Categories within each square. Samples are permanent and will be revisited once in every monitoring cycle. For the EBONE sampling design, it was decided that a sample size of 10000 1 km x 1 km squares was the maximum sample size achievable. It represents approximately 0.25 % of European surface area belonging to the sample frame (for the exact definition of the sampling frame: see Brus et al. 2011). This population fraction is roughly comparable to that of the UK Countryside Survey.

More specifically, the EBONE sampling design has the following properties:

- **Sampling unit**: 1 km x 1 km square
- **Sampling methodology**: mapping of General Habitat Categories within 1 km x 1 km squares
- **Sample size**: n = 10000
- **Area of the sample frame**: 4027782 km²
- **Sample selection**: spatially balanced random sample (i.e. assuring that samples are distributed evenly across the sample frame but in a random fashion, not in systematic fashion). The sample is first stratified by Environmental strata (EnS; about 80 strata) and sample size within each EnS (n_{EnS}) proportional to the area of the EnS. Spatial balance is ensured by subdividing each EnS in so-called geostrata by means of a clustering algorithm that results in geostrata that have approximately the same area and have an edge-to-area ratio as high as possible. Within each geostratum five 1 km x 1 km squares are selected at random (see space-time design) (the number of geostrata is $n_{EnS}/5$. EnS can be aggregated to Environmental zones (EnZ). The number of samples within an EnZ (n_{EnZ}) is proportional to the area of the EnZ (see Table 1). EnZs are an important spatial scale for reporting on stock and change estimates. This is because EnZs largely correspond with the biogeographical zones defined in the Habitat Directive (92/43/EEC) (Jongman et al. 2006).
- **Space-time design**: The space-time design is serially alternating with a cycle of five years and yearly sampling. Samples are permanent and will be revisited once in each cycle. Because each year one out of five random samples within each geostratum is chosen

randomly without replacement, each year-panel (the samples collected in a given year) has the desirable property of being a spatially balanced sample.

Environmental Zone	Sample Size (n_EnZ)	Area (km²)
Alpine North (ALN)	580	231351
Boreal (BOR)	1075	435626
Nemoral (NEM)	400	161850
Atlantic North (ATN)	395	159188
Alpine South (ALS)	610	245390
Continental (CON)	2125	854814
Atlantic Central (ATC)	1045	420500
Pannonian (PAN)	730	294792
Lusitanian (LUS)	495	198660
Mediterranean Mountains (MDM)	910	244159
Mediterranean North (MDN)	1045	423632
Mediterranean South (MDS)	895	357820
First quartile	516	206833
Median	813	270091
Mean	859	335649
Third quartile	1045	422849
Standard deviation	471	191676
Interquartile range	529	216016

Table 1 Sample size and area per environmental zone, along with some summary statistics.

Since the sampling scheme compares well with the UK Countryside Survey, it is worthwhile to look at the data they have collected in the past to learn about the empirical estimates for total area of broad habitat types and their associated precision. This is what we will cover in the next section (section 0). For two reasons this is an important first step before performing calculations of the actual precision and power for the EBONE sampling design . First, it will allow us to get an insight in plausible ranges of values (area within a km-square, proportion of samples where the habitat is present, …). Second, it will provide a benchmark against which our procedures to calculate precision and power can be checked.

Empirical precision of stock and change estimates based on UK Countryside Survey

We used data from Howard et al. (2003). Howard et al. (2003) give estimates for stock and change based on data gathered in the 1990 and 1998 UK countryside survey. We focus first on the stock estimates for 1998. These estimates are based on a stratified random sample (strata are ITE land use classes) from 576 1 km x 1 km squares. This is about 0.26% of the total land surface. The total surface area of the sample frame (UK) equals 224280 km².
We first look at the actual field data and precision achieved on area estimates (stock estimates). Next, we simulate the precision for the 1998 CS and check if our results match the empirical estimates.

The empirical precision of stock estimates for the UK Countryside Survey is depicted in Figure 1. The upper panel depicts the relative precision, the lower panel the absolute precision.

Figure 1 suggests that the relative margin of error is smaller for broad habitat types that have a larger total surface area and/or have a more even distribution (hence, a higher percentage of samples in which the habitat was found). Indeed, it is to be expected that a less common habitat type (say one that covers less than 10% of the UK) with a uniform spatial distribution can be estimated more precisely than one with a clustered spatial distribution (hence, absent from most samples).

According to these estimates, the broad habitat types cover between 0.2% (Montane habitats) and 24% (Improved grasslands) of the UK land surface,. The percentage of 1 km x 1 km squares where the habitat was present ranges from 3% (Montane habitats) to 79% (Broadleaved, mixed and yew woodland).

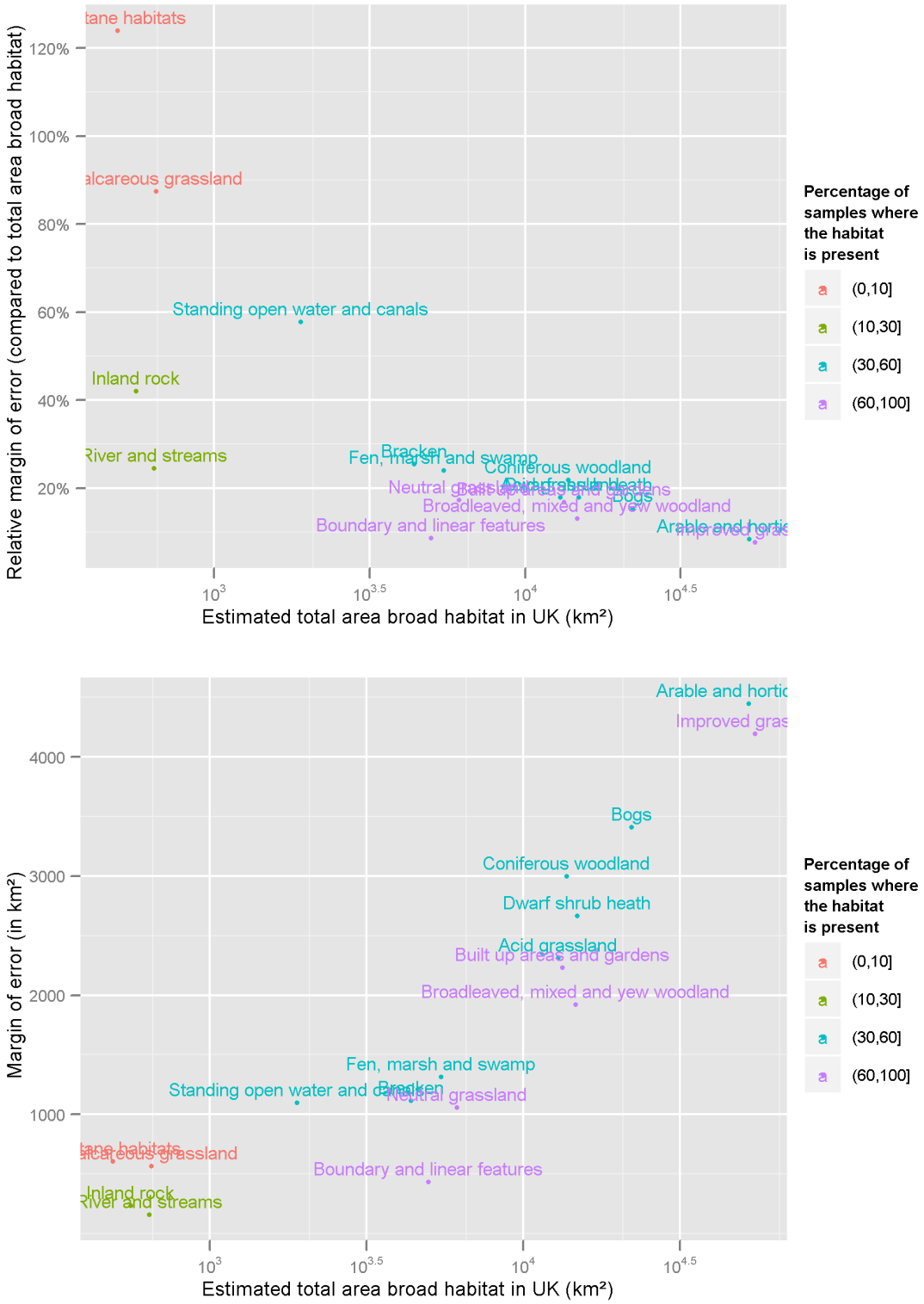

Figure 1 Empirical precision as calculated from data obtained in the 1998 Countryside Survey for the broad habitat types. Each point represents one broad habitat type. Precision is expressed as the half width of the 95% confidence interval (= margin of error). Upper: relative values for precision; Lower: absolute values for precision.

Given these observations based on the Countryside Survey sampling scheme, we developed a procedure to realistically simulate data to estimate the stock precision. Comparing the simulated data with the empirical estimates will tell us if the simulations are indeed realistic.

For the simulation, we ignore the stratification and calculate the standard error as if the data come from a simple random sample. The variance estimator for a simple random sample is less efficient than the variance estimator for stratified random sample (e.g. Cochran 1977). Thus, the simulated precision estimates are an overestimate. The simulation procedure was as follows:

- Determine plausible range of values for total area of a broad habitat type. The largest broad habitat type in the UK was improved grassland, which covered about 25% of the UK. Montane habitats were least extensive with a cover of only 0.2% of the UK land area.
- Determine plausible range of values for the percentage of samples where a habitat is present. Broadleaved, mixed and yew woodland was present in 79% of the samples. Only 3% of samples had Montane habitat.
- Determine plausible combinations of total habitat area (or proportion) and proportion of samples with that habitat. The following rule was used, which ensured that the average area of a habitat within a 1 km x 1 km square cannot be greater than 0.5 km² (or 50% of the square):
    - the proportion of habitat area / the proportion of samples where that habitat is present ≤ 0.5
    - for the UK, the habitat type with the largest average area in samples where it was present, was equal to 0.41 km² for arable and horticultural land
- For every one of these plausible combinations, we proceeded as follows:
    - Select *n* (= 576) random values from a binomial distribution with the number of Bernoulli trials (flip of a coin; 0 or 1) equal to 1 and the probability of success (= 1) equal to the proportion of samples where the habitat type is present
    - If the habitat is present (= 1), replace “1” with the proportion of the 1 km x 1 km square covered by that habitat. To do this, we took one random sample from a beta distribution (a value between 0 and 1) where the mean was set to the proportion of habitat area in the UK divided by the proportion of samples where the habitat occurred (= the average amount of the habitat type in the 1 km x 1 km squares where that habitat occurs). The beta distribution is flexible in that it allows different variances given a mean (hence different shapes of the probability density distribution). For these simulations, we set the parameter *a* to 1 (see Bolker 2008 pp 133-135; see Figure 2 for the effect of tuning the variance given a mean: parameter *a* = 1 and mean = 0.5 corresponds with a uniform distribution).
    - The resulting 576 response values (*z*) represent fictive field data for one habitat type. Based on these values, the Horvitz-Thomson estimator for the total of the response values for the case of a simple random sample (i.e. each response value is weighted (*wgt*) by the total land area divided by the number of samples) can be calculated. The standard error (*SE*) for this total is an estimate for the precision. We multiplied the standard error with 1.96 () to obtain the margin or error (*ME*) (half width of 95% confidence interval).
        - $SE = \sqrt{\mathrm{var}(wgt \times z) \times n}$
        - $ME = SE \times 1.96$
    - The above steps were repeated 20 times

Box I

The beta distribution is a continuous analogue for the binomial distribution. It is especially useful when proportions need to be drawn at random from a continuous distribution, because it is bounded between 0 and 1. The beta distribution depends on two shape parameters *a* and *b*, that determine the mean, variance and shape of the distribution. Parameter *a* – 1 represents the number of successes and *b* – 1 the number of failures. This interpretation in terms of successes and failures only holds for *a* and *b* values larger than 1 (Bolker 2008). Figure 2 gives the effect of parameter *a* on the variance of the beta distribution for a given mean. The mean value should be understood as the average amount of a habitat type in the 1 km x 1 km squares where that habitat occurs. Based on data in Howard et al. (2003) this average is unlikely to exceed 0.5 km² (the largest average was for arable and horticular land and equalled 0.4 km²).

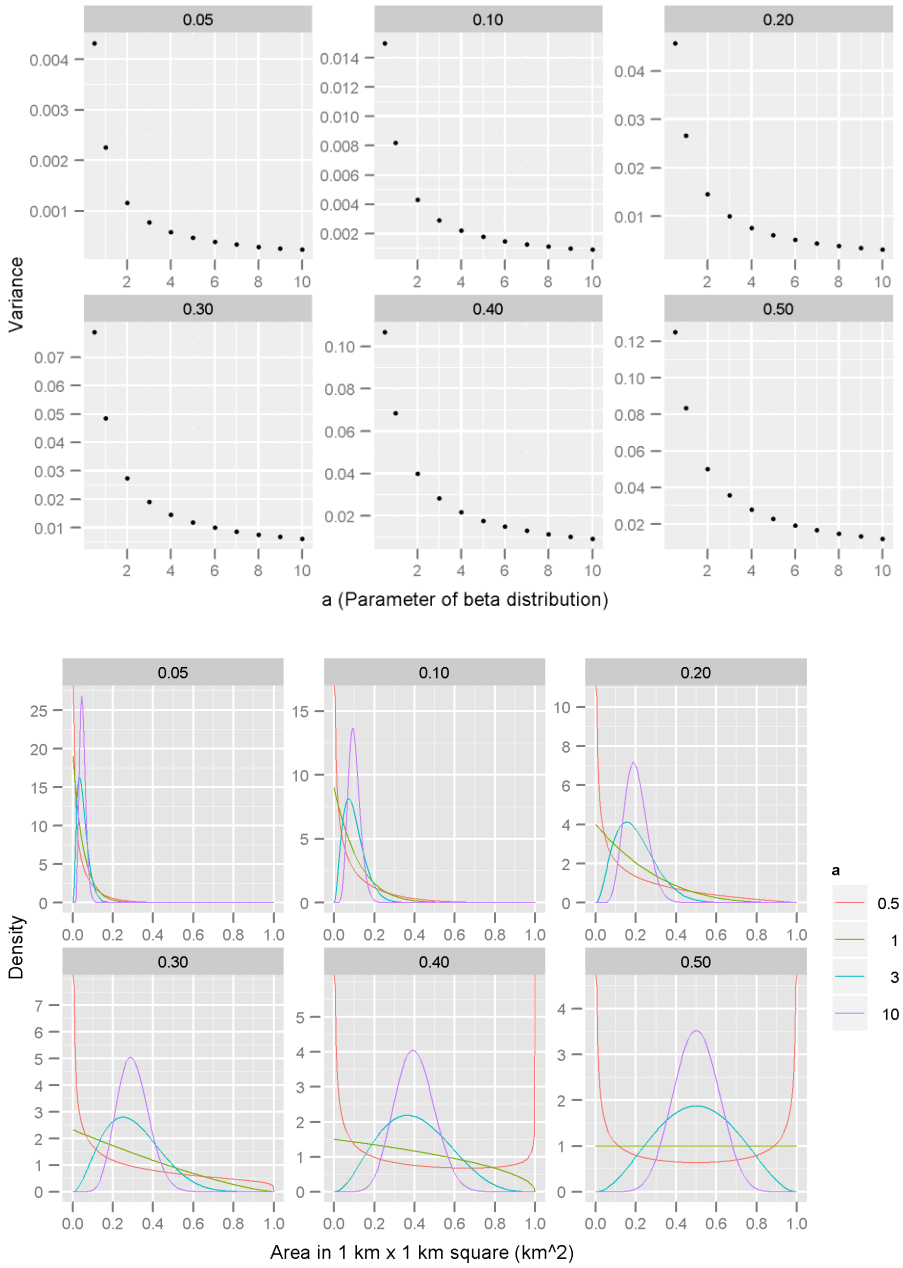

Figure 2 Influence of the shape parameter *a* on the variance of the beta distribution for a given mean (ranging from 0.05 to 0.50). Top: variance as a function of the mean and parameter *a*; these points appear on a straight line when the axes are scaled on a logarithmic base. Bottom: corresponding density distribution plots for *a* = 0.5, 1, 3 or 10.

The results of the simulation are given in Figure 3. Comparing Figure 1 and Figure 3, we learn that the precision estimates based on the simulation procedure largely overlap with the empirical values. It is not a perfect match, and we could have fine-tuned the simulations more by adjusting the parameter of the beta distribution that determines the amount of variability between 1 km x 1 km sample squares where the habitat is present (see Box I Figure 2). However, as this would only complicate things, since the variability will be habitat specific, we decided not to do this. However, we should be aware that the results could give a more optimistic picture than the reality will turn out to be. Only European wide data collected in an implementation or pilot phase of the monitoring schedule could provide certainty about the robustness of the results.

In the next section (0), we will use this simulation procedure to assess statistical performance of the EBONE sampling design and cross-check the results with the design-based estimators for precision given in Brus et al. (2011).

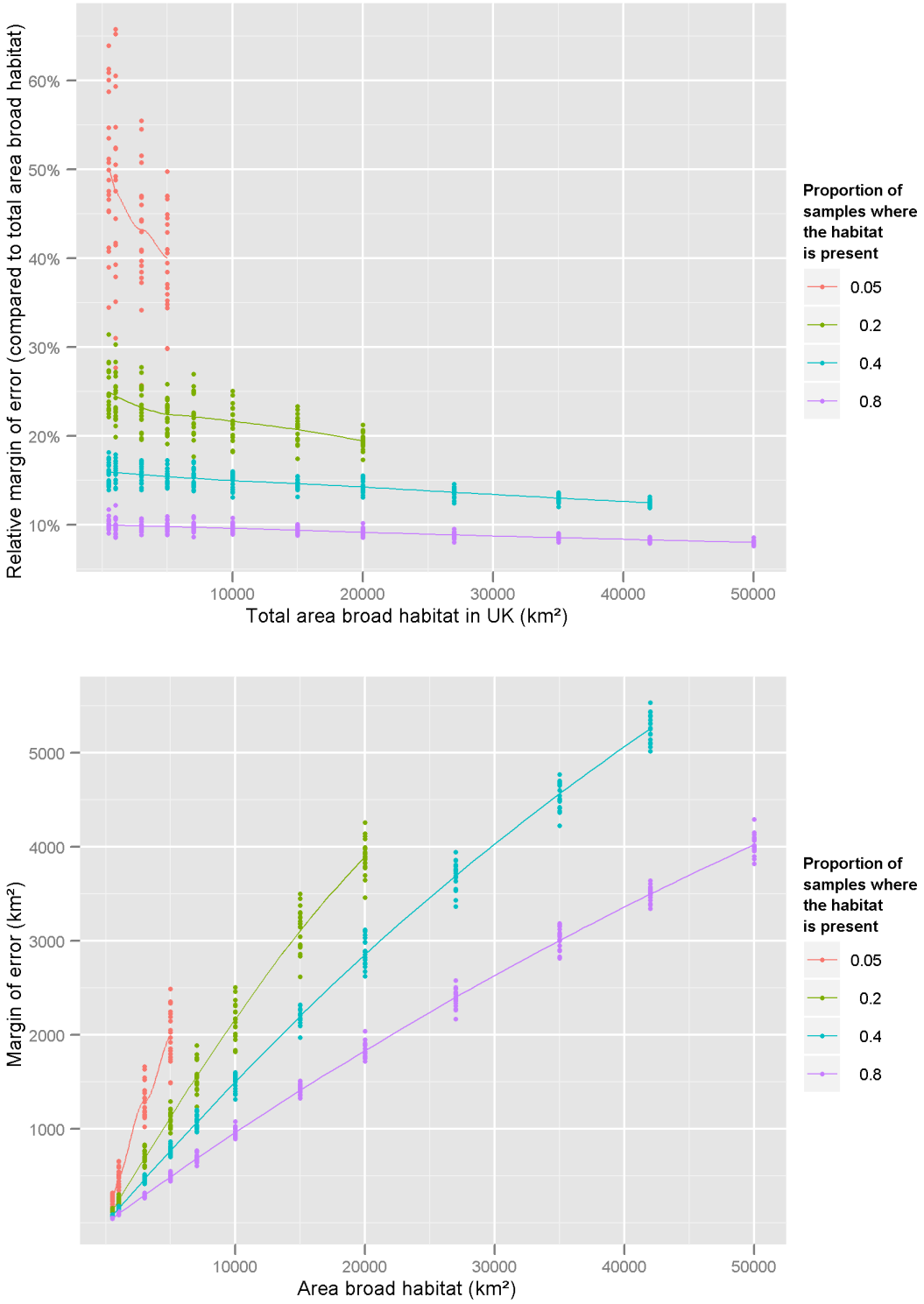

Figure 3 Simulation results showing the precision with which the total area of a broad habitat category can be detected. Twenty simulations for each combination of total habitat area and proportion of samples where the habitat is present. Each point represents one simulation and is based on 576 random samples from a distribution that reflects both variation associated with habitat presence or absence and variation in the amount of habitat area between samples where habitat is present. Top panel: Margin of error relative to the total area of broad habitat; Bottom panel: The same data but raw values in km².

Precision of stock and change for the EBONE sampling design

Precision of stock estimates

We based our simulations on sample sizes of 625, 2500 and 10000 samples. Based on the data from the UK CS, we also let the proportion of samples where habitat was present vary between 0.05 and 0.8. We did not explore different scenarios for the effect of differences in variability (given a mean area) in area between samples where habitat was present (i.e. the a parameter for the beta distribution was fixed at 1).
We consulted Evans (2006) and Bloch-Petersen et al (2006) to have a rough indication of total area estimates in GHC classes or super-GHC classes. Bloch-Petersen et al (2006) give data for an individual 1 km x 1 km square (located in the Taagerup area in Denmark), the GHC percentage of total area ranges from 0.1% to 72.3% (crops) and for super-GHC classes (aggregation of GHC classes in broader habitat categories) from 0.4% to 72.7%. Evans (2006) give total area estimates for the group and sub-group levels of Annex I habitat types in the Natura 2000 network. At the sub-group level (for instance code 91 for forests of temperate Europe), these estimates range from 500 km² to 37 000 km² (resp. 0.06% to 2% of Natura2000 network, and 0.01% to 0.36% of Europe). However, the area outside the Natura2000 network is unknown. As these publications could not give us a clear understanding of ranges of total area of broad habitat in Europe, we returned to the estimates of total area in the UK Countryside Survey and took as lower limit 1000 km². This is only 0.025% of the European area frame. However, we were also interested in precision estimates relevant at the scale of a biogeographical zone. Therefore, this low limit is relevant, but we have to remember that biogeographical zones have a smaller sample size (the average biogeographical zone will have roughly 860 samples, see Table 1). As upper limit, we simply take the maximum relative total area in the UK (24%), which, for the European case, represents about 1 000 000 km² (≈ 25% cover of a habitat type in Europe).

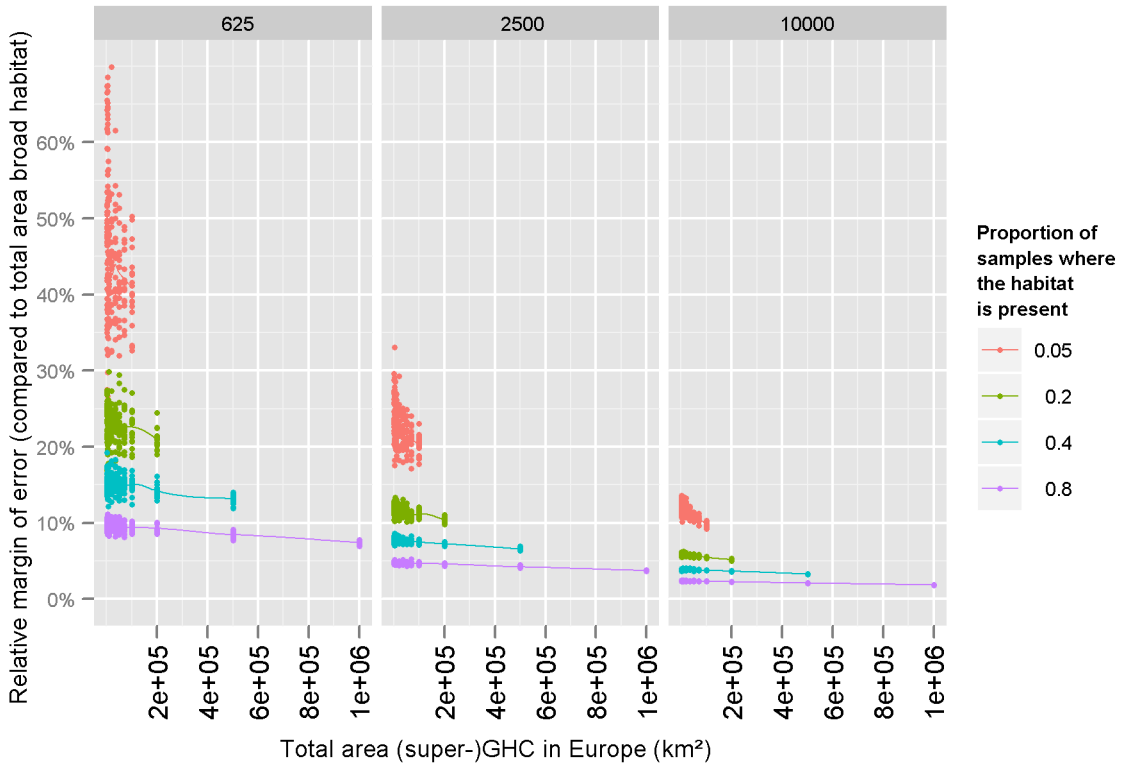

Figure 4 Simulation results showing how precise the total area of a GHC category can be detected. The Y-axis gives the relative margin of error compared to the total area of broad habitat. The X-axis gives the total area of broad habitat. Twenty simulations for each combination of total habitat area and proportion of samples with habitat. Each point represents

one simulation and is based on 625 (left), 2500 (middle) or 10000 (right) random samples. Colours indicate differences in proportion of samples where the habitat is present.

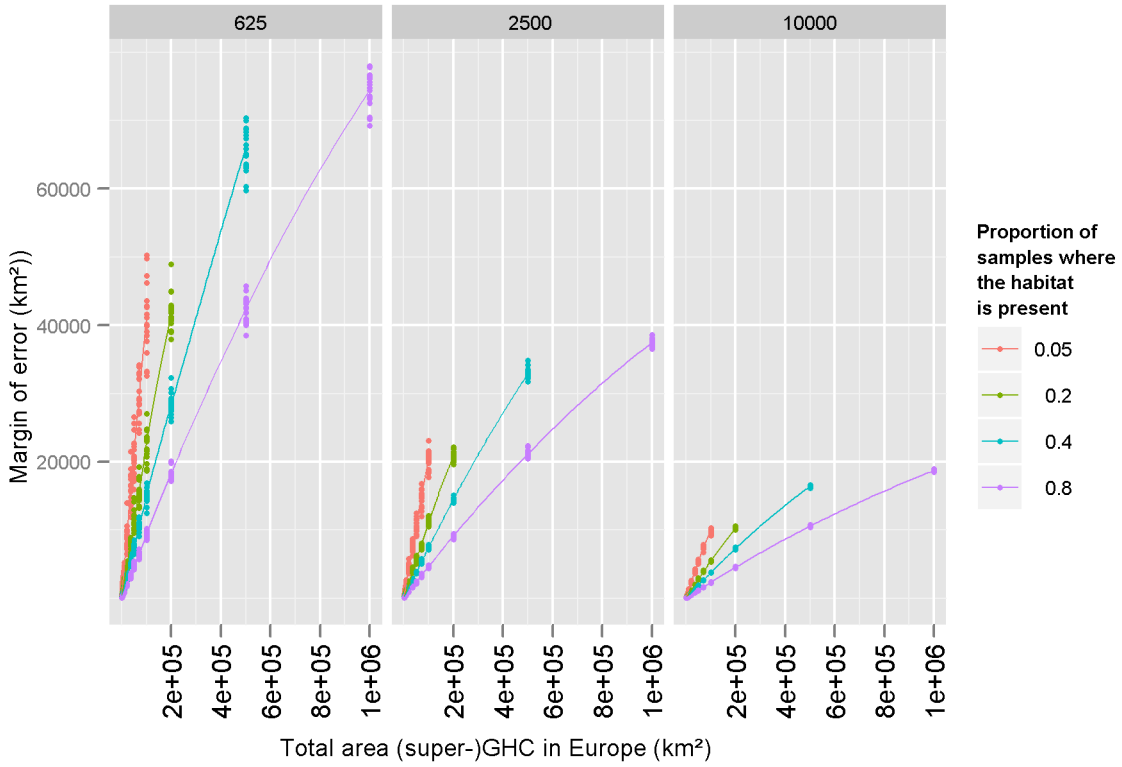

Figure 5 The same data as in Figure 4, but axes are expressed as absolute values in km².

Why would we go through all the trouble of simulating the data, when we can simply use formula to analytically calculate the precision? Brus et al. (2011) present formulas to do exactly this. The formula to calculate the margin of error only needs the following data:

- The significance level (α set to 0.05)
- The sample size (n = 625, 2500, 10000)
- The population size, which is the total number of 1 km x 1 km squares that can cover the sample frame (N = 4027782)
- The estimated spatial variance $\check{S}^2(y)$ (= population variance)

The formula to estimate the margin of error (ME) for a stock estimate – if we approximate the EBONE sampling design as a simple random sample – than equals (Brus et al. 2011):

$$ME = \frac{N\sqrt{\check{S}^2(y)}u_{1-\alpha/2}}{\sqrt{n}}$$

where $u_{1-\alpha/2}$ is the (1 - α/2) quantile of the standard normal distribution.

There is, however, one problem: we need to estimate the population or spatial variance. If we want to base this spatial variance on the zero-altered beta distribution, which we used to simulate data from, we faced the problem that the analytical derivation for the variance of this distribution is unknown to us. There is however an alternative, but computationally less optimal, way to derive a good estimate of the spatial variance. Since the sample variance gradually converges to the spatial variance when n increases to N, we can simply draw a sample of sufficiently large size and use this sample variance as an estimate of the spatial variance. In practice, we needed a sample size as large as N/10 in order to have a stable spatial variance estimate. This meant that, in this case, more simulations had to be done to apply the formula (N/10 ≈ 400 000 > 20 simulations based on a sample size of 10 000). The results match perfectly with our previously obtained results (compare the smoothed lines in Figure 3, Figure 4 and Figure 5 with the lines in Figure 6 and Figure 7), so that we can have

confidence in our simulation procedure. These estimators are also design unbiased (Brus et al. 2011).

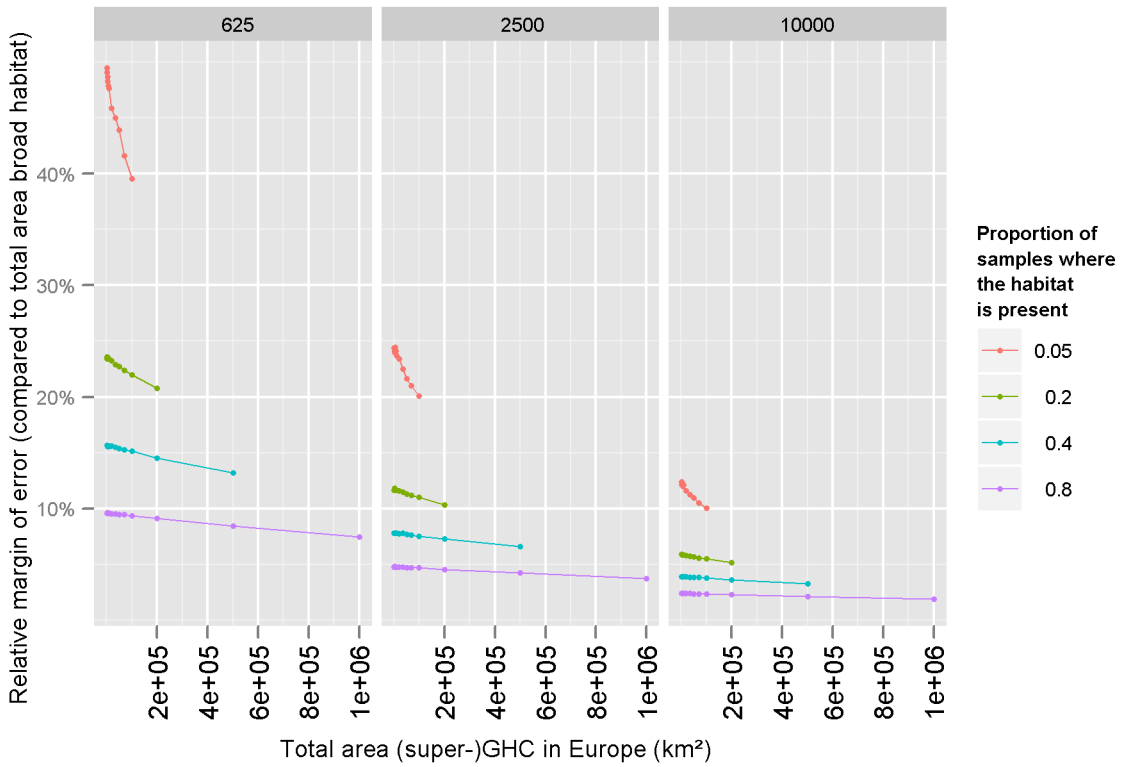

Figure 6 Formula-based results for the relative margin of error given the (expected) total area and the proportion of samples where the habitat is present. The population (or spatial) variance is from the corresponding zero-altered beta distribution.

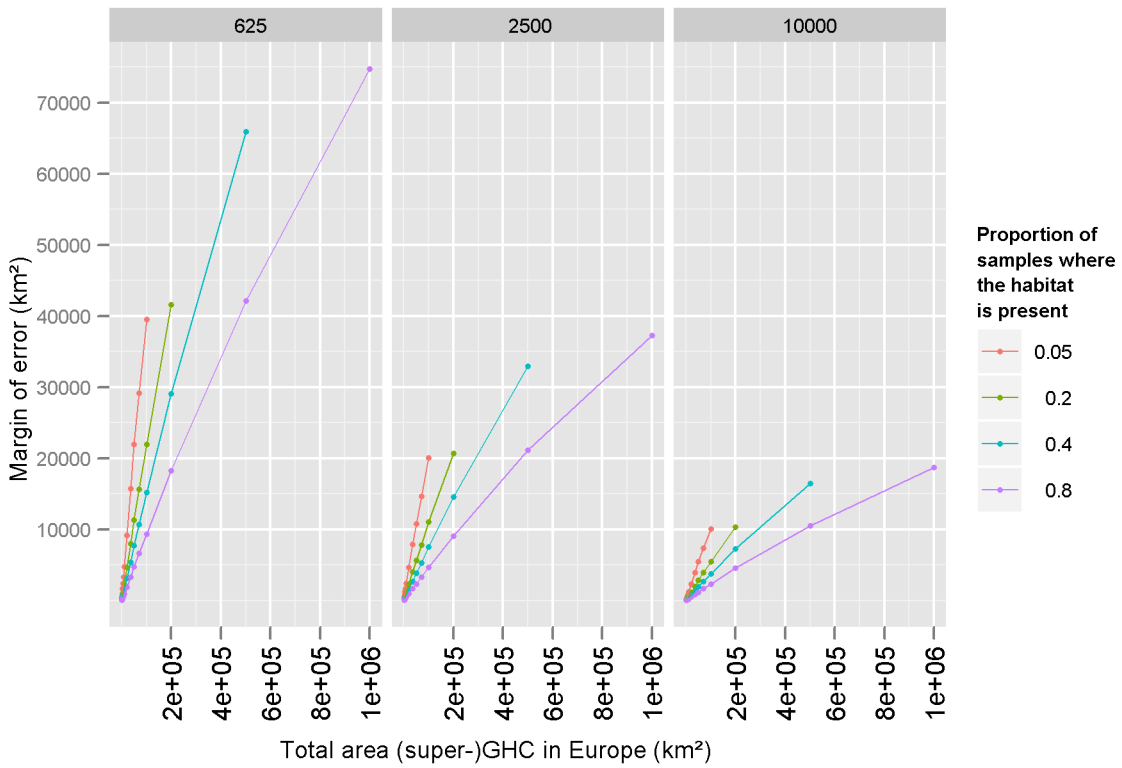

Figure 7 The same data as in Figure 6, but axes are expressed as absolute values in km².

Precision and accuracy of stock estimates based on earth observation samples

We can now ask if we could increase the quality of the estimates of habitat extent by sampling additional 1 km² samples by earth observation (EO). But, how do we incorporate the reduced accuracy of EO samples compared to in-situ samples?

In order to have an indication of the accuracy of EO samples compared to in-situ field measurements (GHC methodology), we can rely on accuracy matrices. An accuracy matrix is an N x N matrix of "observed" and "classified" cells corresponding to N land cover classes (derived from GHC classification). The matrix depicts the land cover classification category derived from earth observation versus the field-observed land cover type. The diagonal cells indicate correct observations, i.e. classified correctly according to the field observations. Any observation off the diagonal indicates a misclassification. Especially, it would be very helpful to have realistic values for both producer and user accuracy for several of the broad habitat categories. Producer accuracy tells us, from the point of view of someone who produces a map based on earth observation data, how well the producer classified the earth observation data (i.e. how many times does the producer mistakenly classifies a pixel to habitat *A*; false negative = omission errors). User accuracy takes the point of view of users of the map and tells them how well the EO-derived map performs in the field (i.e. how many times does the map tell us we are standing in habitat *A*, while in reality we are not; false positives = commission errors).

We used data from Morton et al. (2007) to see the range of values that can be expected for user and producer accuracy. The authors discuss the first UK land cover map (LCM2007) with land parcels (the spatial framework) derived from national cartography by a generalisation (simplification) process. LCM2007 is the first land cover map to provide continuous vector coverage of UK Broad Habitats derived from satellite data. The data from Morton et al. (2007) allow the comparison between the field data gathered during the 2007 UK Countryside Survey (591 1 km x 1 km squares in Great-Britain) and aerial photos with high spatial resolution (25 m). Note that the UK CS squares were not used as ground reference points/data to produce the LCM2007 map. The comparison was made *ex post* as a sort of quality check to gain confidence in the LCM2007 map. A separate set of ground reference points, which is not of interest for our purpose here, was used to calibrate and validate the LCM2007 classification.
Table 2 shows the accuracy results when the EO map (LCM2007) is compared with the CS 1km survey squares (data from Morton et al. 2007).

	Countryside survey in 2007											
LCM2007	Broadleaved woodland	Coniferous woodland	Arable and horticulture	Improved grassland	Semi-natural grassland	Mountain, heath, bog	Saltwater	Freshwater	Coastal	Built up areas and gardens	Sum of columns (ha)	User accuracy
Broadleaved woodland	1212	205	42	131	245	24	2	4	2	92	1959	0.62
Coniferous woodland	134	2503	9	23	84	85	1	3	1	26	2869	0.87
Arable and horticulture	103	41	8643	1944	676	31	3	3	3	424	11871	0.73
Improved grassland	194	38	654	7533	2769	50	4	3	3	371	11619	0.65
Semi-natural grassland	135	80	170	1066	2660	1377	5	10	22	153	5678	0.47
Mountain, heath, bog	84	155	13	43	1068	3479	5	9	5	49	4910	0.71
Saltwater	0	0	0	0	0	0	172	0	2	0	174	0.99
Freshwater	1	2	2	1	3	2	54	346	5	3	419	0.83
Coastal	3	0	42	20	63	3	335	0	87	8	561	0.16
Built up areas and gardens	27	1	48	24	36	1	0	2	2	1109	1250	0.89
Sum of columns (ha)	1893	3025	9623	10785	7604	5052	581	380	132	2235	41310	
Producer accuracy	0.64	0.83	0.90	0.70	0.35	0.69	0.30	0.91	0.66	0.50		0.67

Table 2 Accuracy table for the 2007 countryside survey compared to the UK land cover map. Data from Morton et al. (2007).

Figure 8 retakes some of the data in Table 2. It shows the range of observed data for user and producer accuracy. Most broad habitat types do not deviate far from the line of no difference. Coastal and Saltwater deviate much from the 1:1 line but in opposite directions, which probably just indicates that they represent two sides of the same coin: the omission errors made in one class are offset by commission errors in the other class. Furthermore, we observe that (when Coastal and Saltwater are excluded) the user accuracy range is from 0.5 to 0.9, while producer accuracy ranges from 0.35 to 0.9.

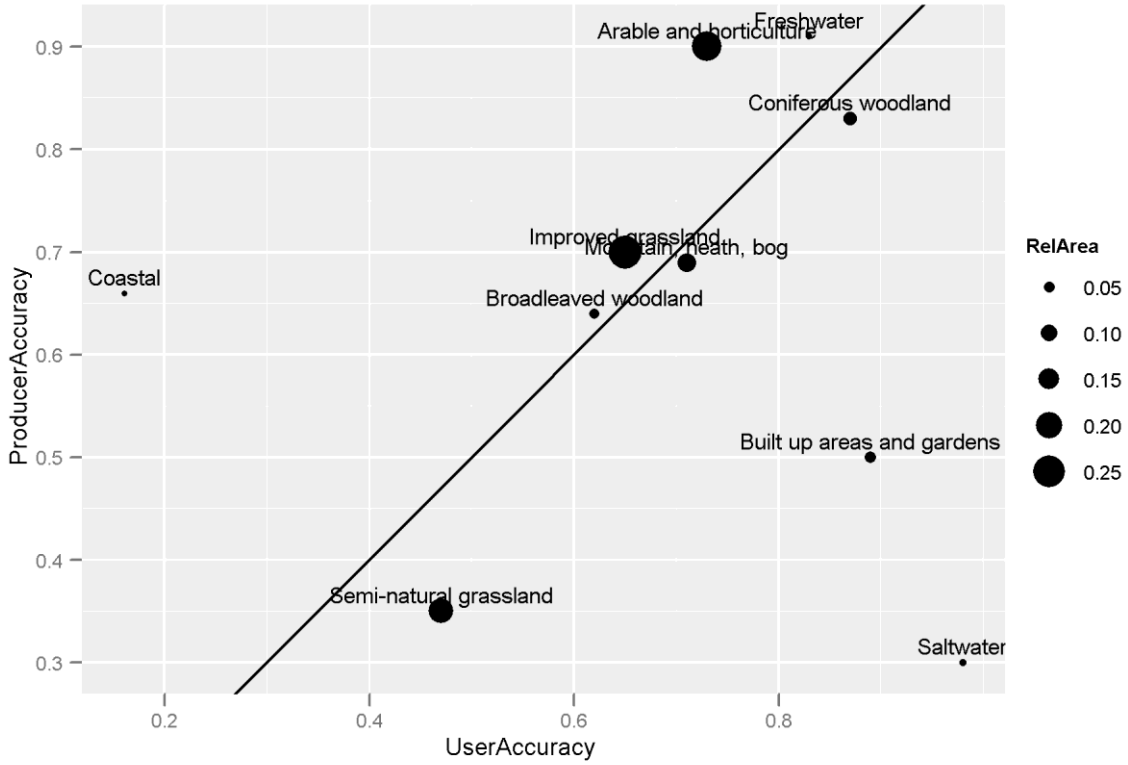

Figure 8 Scatterplot of producer versus user accuracy. The size of each circle (broad habitat type) is proportional to the estimated total area for that habitat type.

We can single out each broad habitat type in Table 2 and construct a new table that looks like the one presented in Table 3.

	In situ samples (GHC methodology)			
Earth observation	Habitat A	Not habitat A	Sum	User accuracy
Habitat A	TP	FP	TP + FP	TP / (TP + FP)
Not habitat A	FN	TN	FN + TN	
Sum	TP + FN	FP + TN		
Producer accuracy	TP / (TP + FN)			

Table 3 A simplified accuracy matrix that compares a focal broad habitat type against all non-focal habitat combined. TP = true positives, FN = false negatives, FP = false positives and TN = true negatives.

From the perspective of the earth observation map producer, we have:

$$Producer\ accuracy = P(EO = A|IS = A) = Sensitivity = True\ positive\ fraction$$

Thus, the producer accuracy equals the conditional probability of observing the focal habitat with earth observation given the in situ observed focal habitat. When the latter is observed without error, the producer accuracy is also referred to as the sensitivity or the true positive fraction (Quataert 2011).

From the perspective of the users of this map, we have:

$$User\ accuracy = P(IS = A|EO = A) = Positive\ predictive\ value$$

Thus, the user accuracy equals the conditional probability of observing the focal habitat in the field given the distribution of the focal habitat on the earth observation map. This fraction is also called the positive predictive value (Quataert 2011).

We will now use both fractions to explore the precision and bias of EO samples. But before we do this, we make the following assumptions:

- When a habitat is present in a km-square sample, both earth observation and in situ sampling will detect it
- We assume that the in situ samples, represent ground truth. They represent the true state without measurement error.

We can use Bayes rule to relate the producer accuracy to the user accuracy:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$P(EO = A|IS = A) = \frac{P(IS = A|EO = A)P(EO = A)}{P(IS = A)}$$

If we rearrange this, we can estimate the unconditional probability of observing habitat *A* with earth observation as a function of both user and producer accuracy and the ground truth probability for habitat *A*:

$$P(EO = A) = \frac{Producer\ accuracy}{User\ accuracy} P(IS = A)$$

Equation 1

The following conclusions can be drawn from Equation 1 regarding potential bias of the EO-samples:

1. When PA = UA, the earth observation estimates will be unbiased
2. When PA ≠ UA, the earth observation estimates will be biased
    a. PA > UA, results in a positive bias (overestimation)
    b. PA < UA, results in a negative bias (underestimation)
3. When PA and UA can be estimated, we have a means to correct for bias. The bias-corrected estimates can be calculated as follows:
$$P(EO = A)\frac{User\ accuracy}{Producer\ accuracy}$$
4. Given the previous conclusion, we need to make sure that PA and UA estimates can be obtained from the combined IS and EO sample design
5. Even when PA and UA are much smaller than 1, unbiased estimates can be obtained (for the average or total area in the sample); however, the produced EO maps will have very large spatial errors regarding the spatial location of the habitat and will therefore be useless.

Equation 1 can also be used in simulation of the effect that earth observation has on precision and accuracy (bias). We took the same simulation procedure as before, but simply replaced the simulated values (habitat area in a 1 km x 1 km square) with the value that would result if the sample were observed with earth observation (Equation 1). We used values for user and producer accuracy from 0.6 to 0.8 in steps of 0.1. To be able to compare with the result from in situ samples we also included user and producer accuracy equal to 1 (i.e. the situation where EO and in situ produce exactly the same result: 100% accuracy).

The conclusions regarding bias – discussed above – are depicted in Figure 9.

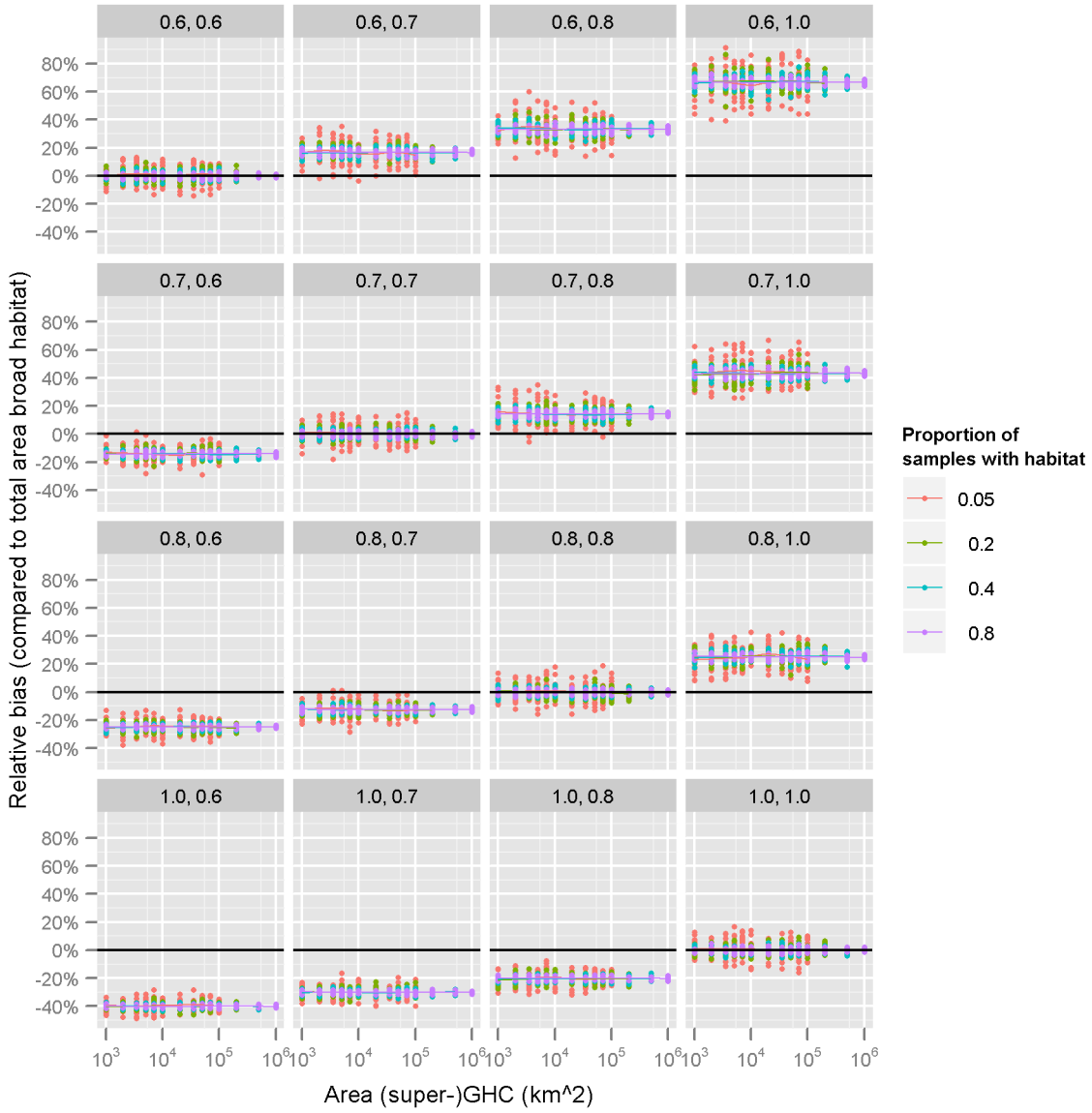

Figure 9 Relative bias as a function of user and producer accuracy. The heading above each panel gives user accuracy and producer accuracy respectively. The lower right panel corresponds with the situation where in situ samples and earth observation samples give identical results (i.e. 100% accuracy; for comparison purposes only). The sample size is equal to 10000.

Inspection of Figure 10 allows us to draw the following conclusions regarding precision (which is inversely related to the relative margin of error; the margin of error equals the half-width of a 95% confidence interval):

1. When PA = UA, the precision is the same as that obtained from an IS sample (sampling errors only; no measurement errors assumed)
2. When PA ≠ UA, the precision is different from that obtained from an EO sample:
    a. PA > UA: a less precise estimate will be obtained
    b. PA < UA: a more precise estimate will be obtained

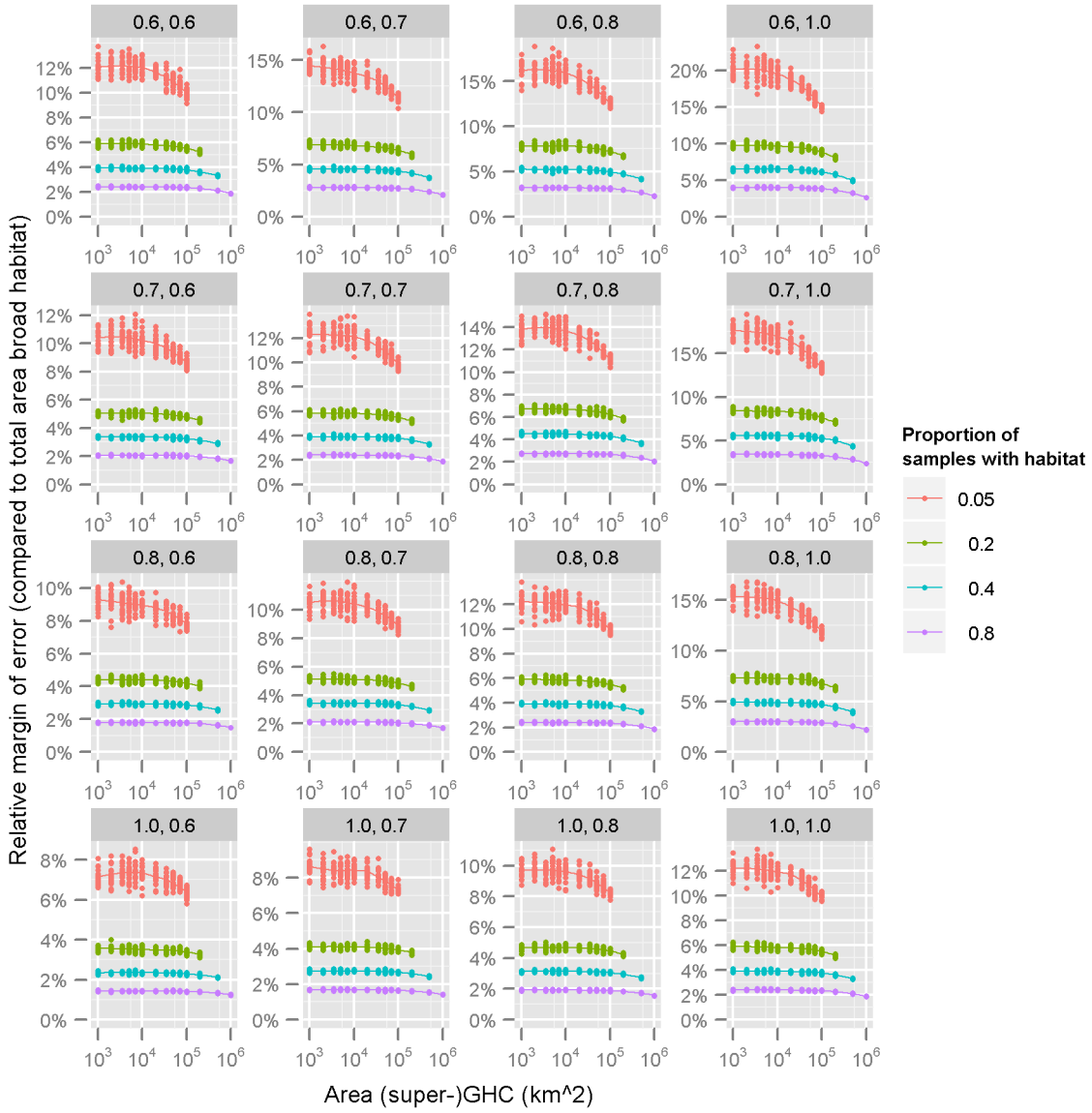

Figure 10 Relative margin of error as a function of user and producer accuracy. The heading above each panel gives user accuracy and producer accuracy respectively. The lower right panel corresponds with the situation where in situ samples and earth observation samples give identical results (i.e. 100% accuracy; for comparison purposes only). Sample size used was 10000.

It is clear from the above discussion that the effect that earth observation samples has on precision or bias will depend crucially on differences in user and producer accuracy. The most important take home message is probably that we can correct for possible systematic bias if and only if the earth observation sample and the in situ sample partly overlap. This overlap, however, should be sufficiently large to ensure that user and producer accuracy themselves can be estimated precisely and without bias. In this respect, it is also crucial that the overlapping part of both samples is a spatially balanced, random sample to avoid bias.

Cost-effectiveness of the sampling design

Starting from the sampling design scenario's, we can explore the implications for a cost-effective monitoring design. The problem we want to solve is how to achieve a good balance between output quality of the design and available monetary budget or alternatively, the constraint could be formulated in terms of time. The effectiveness can often be related to statistical concepts, such as the margin of error or the sampling variance. Which measure for effectiveness will be most useful will depend on the question at hand. For estimation of a mean or a total, higher effectiveness is related to a narrower confidence interval, as we have shown. For trend detection, the effectiveness will depend on the power to detect a trend, and thus this will depend on the magnitude of the trend that needs to be detected.
For a given sample size, we can thus assess effectiveness. The pilot data gathered during the EBONE project also allow us to get insights into time requirements for field work. Confronting the time requirements with the effectiveness yields a first rough approximation of cost/time-effectiveness. In order to compare with earth observation, we would also need estimates of the cost associated with earth observation. Especially, the relative difference in cost between in situ sampling and earth observation will be important. Data to make this comparison were however lacking.

Conclusions

We see this exercise as a first approximation to performance that can be expected from the EBONE sampling design and earth observation samples. As a general rule, precision and power calculations are not sacrilegious. The details of the monitoring design are just as important, as well as the institutional and other factors that will be crucial for implementing the design. We therefore also recommend that, an implementation phase will be foreseen, and that data gathered during that phase will be analysed again in the light of statistical performance, field procedures, data storage, cost-effectiveness etc.

References

[12, 16, 17], [3], =[10, 11]=[10, 11], [7], [14], [8], [9], [5], [15], [4], [1], [6], [13], [2]

[1] BLOCH-PETERSEN, M., BRANDT, J., AND OLSEN, M. Integration of european habitat monitoring based on plant life form composition as an indicator of environmental change and change in biodiversity. *Geografisk Tidsskrift - Danish Journal of Geography 106*, 2 (2006), 61–74.

[2] BOLKER, B. M. *Ecological Models and Data in R*. Princeton University Press, Princeton, NJ, 2008. ISBN: 0691125228.

[3] ELZINGA, C. L., SALZER, D. W., WILLOUGHBY, J., AND GIBBS, J. *Monitoring Plant and Animal Populations*. Blackwell Science, 2001.

[4] EVANS, D. The habitats of the European Union Habitats Directive. *Biology & Environment: Proceedings of the Royal Irish Academy 106*, 3 (2006), 167–173.

[5] GARDNER, T. A., BARLOW, J., ARAUJO, I. S., Ã?VILA PIRES, T. C., BONALDO, A. B., COSTA, J. E., ESPOSITO, M. C., FERREIRA, L. V., HAWES, J., HERNANDEZ, M. I. M., HOOGMOED, M. S., LEITE, R. N., LO-MAN-HUNG, N. F., MALCOLM, J. R., MARTINS, M. B., MESTRE, L. A. M., MIRANDA-SANTOS, R., OVERAL, W. L., PARRY, L., PETERS, S. L., RIBEIRO-JUNIOR, M. A., DA SILVA, M. N. F., DA SILVA MOTTA, C., AND PERES, C. A. The cost-effectiveness of biodiversity surveys in tropical forests. *Ecology Letters 11*, 2 (2008), 139–150.

[6] HOWARD, D. C., WATKINS, J. W., CLARKE, R. T., BARNETT, C. L., AND STARK, G. J. Estimating the extent and change in broad habitats in great britain. *Journal Of Environmental Management 67*, 3 (Mar. 2003), 219–227.

[7] JONGMAN, R., BUNCE, R., METZGER, M., MÃ¼CHER, C., HOWARD, D., AND MATEUS, V. Objectives and applications of a statistical environmental stratification of europe. *Landscape Ecology 21* (2006), 409–419.

[8] JOSEPH, L. N., MALONEY, R. F., AND POSSINGHAM, H. P. Optimal allocation of resources among threatened species: a project prioritization protocol. *Conservation Biology 23*, 2 (2009), 328–338.

[9] LAYCOCK, H., MORAN, D., SMART, J., RAFFAELLI, D., AND WHITE, P. Evaluating the cost-effectiveness of conservation: The UK Biodiversity Action Plan. *Biological Conservation* (2009).

[10] LINDENMAYER, D., AND LIKENS, G. Adaptive monitoring: a new paradigma for long-term research and monitoring. *Trends In Ecology & Evolution 24* (2009), 482–486.

[11] LINDENMAYER, D. B., AND LIKENS, G. E. The science and application of ecological monitoring. *Biological Conservation 143* (2010), 1317–1328.

[12] ONKELINX, T., VERSCHELDE, P., WOUTERS, J., BAUWENS, D., AND QUATAERT, P. Ontwerp en evaluatie van meetnetten voor het milieu- en natuurbeleid. steekproefgrootteberekeningen en analyse van de kosteneffectiviteit. Tech. rep., Vlaamse overheid, Departement Leefmilieu, Natuur en Energie, Instituut voor Natuur- en Bosonderzoek, Brussel, 2008.

[13] QUATAERT, P. *Methodological and statistical aspects of indices of biotic integrity to assess the ecological condition of surface waters*, vol. INBO.T.2011.1. Instituut voor Natuur en Bosonderzoek (INBO), Brussel, 2011.

[14] SCHILPEROORT, T., AND GROOT, S. Design and optimization of water quality monitoring networks. Tech. rep., Waterloopkundig Laboratorium Delft Hydraulics Laboratory, 1983.

[15] TULLOCH, A., POSSINGHAM, H. P., AND WILSON, K. Wise selection of an indicator for monitoring the success of management actions. *Biological Conservation 144*, 1 (2011), 141 – 154.

[16] WOUTERS, J., BAUWENS, D., ONKELINX, T., AND QUATAERT, P. Ontwerp en evaluatie van beleidsgerichte meetnetten voor het milieu- en natuurbeleid. Leidraad voor de meetnetontwerper. Tech. Rep. INBO.M.2008.8, Instituut voor Natuur- en Bosonderzoek, Brussel, 2008.

[17] WOUTERS, J., QUATAERT, P., ONKELINX, T., AND BAUWENS, D. Ontwerp en handleiding voor de tweede regionale bosinventarisatie van het Vlaamse Gewest. Tech. rep., Instituut voor Natuur- en Bosonderzoek, Brussel : Belgium, 2008.