

## TREE COVER AREA ESTIMATION IN EUROPE BASED ON THE COMBINATION OF IN SITU REFERENCE DATA AND THE COPERNICUS HIGH RESOLUTION LAYER ON TREE COVER DENSITY

Christophe Sannier<sup>1</sup>, Javier Gallego<sup>2</sup>, Tobias Langanke<sup>3</sup>, Usue Donezar<sup>3</sup>, & Alexandre Pennec<sup>4</sup>

<sup>1</sup>IGN FI, 7 rue Biscornet, 75012 Paris, France

<sup>2</sup>Independent Consultant, formerly JRC Ispra, Italy

<sup>3</sup>EEA, Kongens Nytorv 6, 1050 Copenhagen, Denmark

<sup>4</sup>CLS Lille, 61 rue de la Cimaise, 59650 Villeneuve d'Ascq, France

**KEY WORDS** : activity data, change detection, EEA, statistical inference, forest cover

### ABSTRACT:

There is a natural tendency from the remote sensing community to extract area statistics (i.e. "Pixel counting") from EO based geospatial products to produce statistical indicators for various purposes. However, geospatial map products suffer from misclassification errors and "pixel counting" can only be justified when the accuracy of such map products reaches a level when these misclassification errors can be considered negligible, but this is possible only in very specific circumstances. Nevertheless, there has been some effort in the Remote Sensing community to assess the accuracy of map products against some form of reference data to ensure that the maps could reach a sufficient level of accuracy. However, there is generally a lack of standards and guidelines and how to perform rigorous map accuracy assessment and rigorous methods for assessing map accuracies and extracting statistics are still lacking as highlighted by McRoberts (2011). Despite substantial advances in this topic in the scientific literature in recent years notably with the paper from Olofsson et al (2014), this has yet to be fully implemented in operational projects. In addition, even if map accuracy assessment is performed correctly, high accuracy does not necessarily mean that area statistics can be directly extracted from a map.

This study is focused on developing the rigorous and appropriate use (i) of geospatial map products from satellite imagery and (ii) statistically sound methods for reporting area estimates and their associated uncertainty.

from Olofsson et al (2014), this has yet to be fully implemented in operational projects. In addition, even if map accuracy assessment is performed correctly, high accuracy does not necessarily mean that area statistics can be directly extracted from a map.

## 1 INTRODUCTION

There is a natural tendency from the remote sensing community to extract area statistics (i.e. "Pixel counting") from EO based geospatial products to produce statistical indicators for various purposes. However, geospatial map products suffer from misclassification errors and "pixel counting" can only be justified when the accuracy of such map products reaches a level when these misclassification errors can be considered negligible, but this is possible only in very specific circumstances. Nevertheless, there has been some effort in the Remote Sensing community to assess the accuracy of map products against some form of reference data to ensure that the maps could reach a sufficient level of accuracy. However, there is generally a lack of standards and guidelines and how to perform rigorous map accuracy assessment and rigorous methods for assessing map accuracies and extracting statistics are still lacking as highlighted by McRoberts (2011). Despite substantial advances in this topic in the scientific literature in recent years notably with the paper

To illustrate this, the Table below shows a confusion or error matrix that was produced from the comparison of a set of reference data collected from the visual interpretation of sample units and a corresponding map product of Forest (F), Non-Forest (NF) and Change (C). The error matrix below is based on a probability sample (i.e. the probability of inclusion of each sample unit is known and unequal sampling intensity from different strata was corrected to ensure that each sample unit carries an equal weight in the error matrix) and different accuracy metrics can be extracted from the matrix such as overall accuracy for the entire matrix / map and user and producer accuracies for individual thematic classes which are indicative respectively of commission (i.e. over-classification) and omission (i.e. under-classification) errors. Confidence intervals can also be added to assess whether the differences between the different metrics are statistically significant at a certain confidence level.

**Table 1. Error Matrix of a Forest (F), Non-Forest (NF) and Change map**

		Reference			Total	User's accuracy
		Stable NF	Stable F	Change		
Map	Stable NF	276	6		283	0,977
	Stable F	20	691	1	712	0,970
	Change	1	2	2	5	0,427
Total		297	699	3	1000	
Producer's accuracy		0,930	0,988	0,651	Overall Accuracy	0,969

In the example shown, the overall accuracy is relatively high with a value close to 97%. However, the overall accuracy can be misleading in that it can hide some imbalance for individual map classes. This is particularly true for small classes such as the change class for which the accuracy is often lower when in fact changes are the most important indicators to be generated from a policy perspective. In this particular case changes are overestimated by a factor of 2/3 as compared to the reference data considering that the proportion of each classes in the matrix is properly respected. In other words, if the total area considered is 100,000km<sup>2</sup>, the area of change from the reference data would represent 300km<sup>2</sup> whereas it would represent 500km<sup>2</sup> from the map.

The issue of the impact of map inaccuracies or rather imbalance between omission and commission errors has been known for some time when satellite-based products were used in combination with reference data for crop area statistics first in the USA during the 1980's (Allen 1990) and then in Europe as part of the Monitoring of Agriculture with Remote Sensing programme during the late 1980's and 1990's (Taylor et al 1997). More recently, similar approaches were developed and adapted for tropical forest monitoring to support reporting of activity data for the Forest sector as part of the United Nations Framework Convention on Climate Change (UNFCCC). Under UNFCCC, there are strict requirements that estimates should neither over or underestimate reality and that their uncertainty should be known. In this context, "pixel counting" violate this statement in that there is no way of knowing (i) if the estimate is realistic (i.e. neither over or under estimation) and (ii) the associated uncertainty. The approaches from Olofsson et al (2014), but also Sannier et al (2014) were identified as suitable methods to estimate activity data and its associated uncertainty (GFOI, 2016). Methods are therefore well known and documented, but there is still a lack of widespread implementation among the remote sensing community. On the other hand, remotely sensed based products are still not very widely used as part of traditional statistical reporting systems whereas there could potentially be strong synergies between traditional purely sample based methods and the use of remotely sensed based products. Notably map products produced from satellite imagery can be:

- A cost-efficient basis for stratification thus contributing to reducing the uncertainty
- Combined with reference data through a model assisted approach to further reduce the uncertainty of the estimate

Therefore, this study is focused on developing the rigorous and appropriate use (i) of geospatial map products from satellite imagery and (ii) statistically sound methods for reporting area estimates and their associated uncertainty.

## 2 TECHNICAL APPROACH AND METHODOLOGY

The sections below provide a detailed description of the approach that was applied based on the extensive experience of the project team in the validation of Copernicus HRL products.

### 2.1 Area estimation and map validation

Both problems may seem different to each other at first sight, but they are deeply linked to each other. Most map validation indicators are computed from the confusion matrix (Table 1). Each term  $A_{gc}$  of a confusion matrix should be extrapolated to the area that belongs to the ground class  $g$  and has been classified as  $c$ . However, for the area as well as map accuracy estimates calculated from a confusion matrix, each sample unit included in the matrix must be based on a probability sample for which the sampling intensity will need to be accounted for in case of unequal sampling intensities, which may occur if a stratification has been applied. Therefore, estimating a confusion matrix is a problem of area estimation as outlined by Olofsson et al. (2014). More details are provided in the following sections.

### 2.2 Sample design

#### 2.2.1 Choosing sampling units: clusters or unclustered points?

Samples for assessing the accuracy of map products and area estimation are usually drawn from area frames as opposed to list frames to provide a better representation of the population (Gallego, 2004a). In an area frame, sample units can be points, lines (often referred to as transects) or areas (often referred to as segments, described by Gallego, 1995). The first step is to define the geographical area for which the accuracy assessment is to be reported and the type of sample units. Point samples are very often used, but clusters or segments have advantages in many cases: they reduce the distortion generated by co-registration and may improve the outcome when not only thematic accuracy need to be reported, but also the geometry of mapped objects such as for CLC.

A two-stage sampling approach is implemented by further selecting Primary sampling units (PSUs) and secondary sampling units (SSU) within PSUs in the

second stage. Two-stage sampling is considered suitable for accuracy assessment of land cover maps or area estimation (Stehman, 2009) and can be adopted in certain cases because it represents a good compromise between the ease of data collection and a good geographic distribution.

In the case of the validation of HRL density product for EEA a 2-stage sampling approach was applied as shown below in which 1ha square PSUs were selected and a grid of 5x5 or 10x10 SSUs was applied to facilitate the data collection as shown in Figure 1. The same approach is suggested to be adopted here although the implementation or not of the second stage is discussed in the response design section.



**Figure 1: Example of SSUs organised in a 5x5 20m grid**

### 2.2.2 Stratification

A probability sampling design is essential for map validation and area estimation (Stehman and Foody, 2008, Olofsson et al., 2014), but this is less clear for collecting training data. In exchange, a purposive selection of locations for training data can be an acceptable choice, for example collecting data along roads (Gallego et al, 2014).

Simple random, stratified random, clustered random and systematic designs are all examples of probability sampling designs. In simple random designs, classes covering a small portion of the population may not be adequately sampled. Cluster sampling is often used to reduce the costs of the collection of reference data, but does not resolve geographic distribution problems. Stratified approaches overcome this drawback; therefore, the stratified random sampling of points is one of the most common approaches to assess map accuracy. Stratified systematic with random origin also overcome this drawback and has the advantage of enhancing traceability. The main limitation of systematic sampling is that there are no unbiased estimators of the variance, although the estimators of the target variance is unbiased (Bellhouse, 1988). In practice, the random sampling estimator of variance is used even though it may slightly overestimate the true variance. In addition, there is also a strong advantage of a systematic approach in that there is better traceability as compared with a random approach: a sample drawn systematically cannot be changed since its location is known a priori. For

example, LUCAS is based on a stratified systematic sample.

In many cases, the land cover map to be assessed or validated is directly used as stratification (Lowell and Jatón, 1999). This is a good approach to estimate commission errors in a binary classification, but may be weak for other cases. If we have some information that quantifies the likely proportion of errors, for example a measure of landscape complexity, it can be a more efficient stratification for all types of errors: we shall choose a higher sampling rate in more difficult areas, where both the errors and their variances are higher.

Experience has shown that a too complex stratification will not bring major improvement and that there needs to be a clear case for stratifying the Area of Interest (AoI). A clear case is when there are marked geographical differences. Experience in Europe from the Regional Inventory programme of the MARS project in the 1990s from (Taylor, 1997) showed that the more complex the analysis on which the stratification is based, the less efficient it tends to be. In addition, there should not be too many strata because even if the stratification will make it possible to reduce the number of sample unit needed for a single stratum, the benefits will be lost if too many strata are present. (Cochran, 1977, p. 134) recommends no more than 6-8 strata in total.

If we are looking at commission errors in a binary map (impervious-non impervious, or forest-non forest) and we have decided to use both classes as strata and unclustered points as sampling units, it is possible to estimate a suitable sample size for each stratum based on the expected acceptable error rate. The standard error of the error rate can be calculated as follows:  $\sigma_h = \sqrt{\frac{p_h(1-p_h)}{n_h}}$  where  $n_h$  is the sample size for stratum  $h$  and  $p_h$  is the expected error rate. This can be reworked to express the sample size  $n_h$  as a function of  $p_h$  and desired standard error  $\sigma_h$ :  $n_h = \frac{p_h(1-p_h)}{\sigma_h^2}$ .

### 2.2.3 Sample allocation

When using stratified sampling, the main issue to maximise the efficiency of the stratification is to optimize the sample allocation per strata. A simple way is the use of equal allocation, such as is used e.g. for the verification of the 2012 HRLs, but this is usually not very efficient. Proportional allocation is an option, but it will give disappointing results for important classes covering a small proportion (e.g. impervious land). If we have a high priority for a class, the Neyman allocation rule is a better alternative (Cochran, 1977):

$n_h = n * (N_h * \sigma_h) / [\sum (N_i * \sigma_i)]$  where  $n_h$  is the sample size for stratum  $h$ ,  $n$  is the total sample size,  $N_h$  is the population size for stratum  $h$ , and  $\sigma_h$  is the standard deviation of stratum  $h$ . According to Stehman (2009), Neyman optimal allocation should be preferred for estimating area of change as well as overall accuracy, whereas equal allocation is effective for estimating user's accuracy.

In practice (Särndal et al., 1992, p. 267 and 407) recommend a minimum within-stratum sample sizes of 10-20; (Cochran, 1977, p. 134) recommends minimum within-stratum sample sizes of 20; and for temperate forest inventories (Westfall et al., 2011) recommend within-stratum sample sizes of at least 20. Therefore, for this work, if the use of a stratification approach is suitable, the maximum number of strata should not exceed 6-8 strata and the smallest number of sample units should be not be less than 20 for a single stratum.

### 2.2.4 Selected approach

The approach for drawing sample units as part of the validation undertaken at pan-European level by EEA was based on the LUCAS grid which makes comparison with LUCAS results easier. However, considering the different characteristics and class definition of the HRLs the thematic information of LUCAS points was not directly used even though the LUCAS photos proved particularly useful for identifying land cover / use classes. A similar approach is suggested to be adopted here using the LUCAS grid as the sampling frame corresponding to approximately 1,100,000 points throughout the European Union where land cover or land use type is observed. Using LUCAS points ensures traceability and coherence between the different layers. In the case of the HRL, the LUCAS point is then used to define the origin of the 1ha plot to be selected as described above.

LUCAS points are located every 2 km on a regular grid, as illustrated below. A set of 81 points located on an 18x18 km square constitutes a group in which every point is associated with a number comprised between 1 and 81 (the numbers do not follow each other spatially). The same pattern with the same numbers allocation is repeated all over the grid. A replicate refers to the points with the same number selected on the whole LUCAS grid.

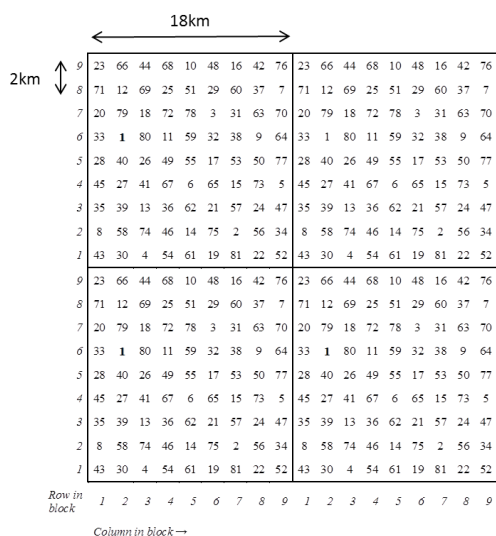


Figure 2. LUCAS points located on a regular grid

At first, the number of samples to allocate to each stratum (or thematic class) is calculated as a function of

their area. In this manner the sampling design is not only systematic but also stratified. The number of sample units per stratum is to be defined to ensure sufficient level of precision at reporting level: The number of replicates to be selected for a stratum depends on its area and the number of LUCAS points intersecting the stratum.

For thematic classes covering a large proportion of the study area, 1 replicate may already exceed the defined number of samples for this class. To solve this problem, replicates are split into four sub-replicates, as illustrated by the blue numbers in the Figure below.

The opposite problem is encountered for land cover classes covering a small proportion of the study area: even by selecting 81 replicates (the maximum number), the intersecting area between the stratum and LUCAS points may be too small to reach the required number of samples. Therefore LUCAS grid can be densified by creating one point every 200 m.

Based on the lessons learnt, from previous validation exercises, the stratification procedure is simplified to only include a commission (tree cover mask 1-100) and an omission stratum (rest of the area) in order to ensure the efficiency of the approach. Stratification is based on the strata defined as follows:

- Commission: tree cover density 1-100% (minimum of 75 PSUs per country / group of countries)
- Omission: tree cover density 0% (minimum of 75 sample units per country / group of countries)
- Commission Change: all changes from 2012-2015 & 2015 -2018

If the minimum of 75 PSUs per country / group of countries is not reached at the first level of stratification, a second level is applied, per country to ensure the minimum number of PSUs.

## 2.3 Response design

### 2.3.1 Overview

Response design is the methodology used to obtain the reference data from the sample units (Stehman & Czaplewski, 1998). Stehman (2009) asserted that accuracy assessment is often made relative to some "higher quality determination of land cover". Czaplewski (2003) indicated that visual interpretation is acceptable if the spatial resolution of EO data is sufficiently better compared to the thematic classification system. For this exercise, visual interpretation of selected sample units will be used as a basis of the response design.

In a quality assessment of an EO product, the response design is a critical element: it includes all aspects of the quantification of the agreement to determine whether the map and reference data on the selected sample. The

response design should closely follow the definition of the product to be assessed both in terms of geometric and thematic characteristics.

When we discuss the use of EO for statistics, the response design is still relevant especially in this case for which the use of the map products can be combined with the sampled observation in a model assisted regression estimation. Therefore, it will be important to ensure that the following characteristics are followed and closely monitored as part of the visual interpretation of selected sample units and that discrepancies between the map and sampled reference data are due to actual thematic errors:

- Minimum Mapping Unit (MMU)
- Minimum Mapping Width (MMW)
- Class definition: what is an impervious surface and what is forest?
- Ensure that the image data used is as closed as possible temporally to that used for the map production

A 2-stage sampling approach was applied:

- 1 ha square PSUs were selected based on LUCAS points as describe in section 2.2.4
- SSUs were selected in a systematic grid of 5x5 points

### 2.3.2 Bias Estimation and mitigation

The bias of pixel counting is the difference of the extrapolated commission and omission errors. With the notation introduced in section 2.3:

$$B = \varphi - \psi$$

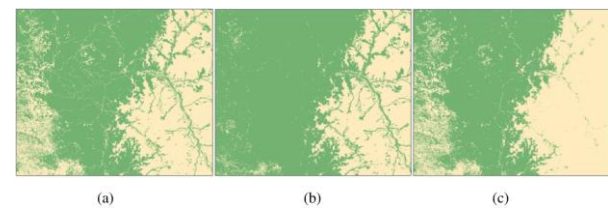
The reason why classification algorithms introduce a bias is that there is no theorem than ensures that the mathematical expectation of commission and omission errors compensate each other. There is nothing similar to the *Law of Large Numbers* in sampling theory. Example of bias are provided in **Erreur ! Source du renvoi introuvable.**

**Table 2: Estimates Forest cover/change proportion forest from segments, figures in bold indicates estimates not significantly different from 0 (Source: Sannier et al., 2014).**

	Forest Cover			Forest Cover Change		
	1990	2000	2010	1990-2000	2000-10	1990-2010
Direct mean	0.8568	0.8534	0.8532	-0.0034	<b>-0.0002</b>	-0.0036
Direct SE	±0.0185	±0.0186	±0.0186	±0.0009	±0.0008	±0.0012
Map estimate (pixel counts)	0.8633	0.8616	0.8613	-0.0017	-0.0003	-0.0020
Bias estimate	0.0027	0.0042	0.0049	0.0015	0.0007	0.0023
MAR Adjusted estimate	0.8606	0.8574	0.8564	-0.0032	<b>-0.0010</b>	-0.0043
MAR Adjusted SE	±0.0028	±0.0026	±0.0025	±0.0009	±0.0008	±0.0012

In some algorithms the result of pixel counting can be adjusted by tuning explicit parameters. In particular the classical maximum likelihood classifier uses the *a priori probability* of each class to push the total number of pixels of that class to higher or lower numbers. It is empirically well known that a uniform *a priori probability* usually introduces a negative bias for large classes and positive for small classes, while a proportional *a priori probability* makes the opposite.

More complex algorithms, such as random forests, neural networks and support vector machines are usually more accurate, although less robust. In these complex algorithms the parameters are hidden and cannot be tuned by the user, but the total number of pixels can be adjusted by cleaning training data or adding/removing the amount of training data in each class. Moreover, most current classification algorithms usually provide an additional probability layer which can be used to adjust the area classified by a given thematic class. These layers are not part of the products delivered for the forest layer. However, the binary layers used for the area estimation are based on density layers, the tree cover density for forest. Different threshold can be applied locally to the density values to adjust the area covered by the layers and reduce the amount of bias. Such a procedure was applied for forest cover in Gabon by Sannier al (2016b) using the global forest change dataset from Hansen et al (2013) concluding that at national level, a 70% threshold applied to the dataset was a closer depiction of the actual 30% threshold from the national forest definition, but for some areas, a more localised threshold adaptation was needed as illustrated in Figure 3 below.



**Figure 3: Comparison of (a) the national forest cover map, (b) the Global Forest Cover based forest cover map with a 30% tree cover threshold and (c) the Global Forest Cover based forest cover map with a 70% tree cover threshold over a 100 x 80 km extract in Eastern Gabon (Source Sannier et al, 2016)**

### 2.3.3 Unbiased area estimates

The most usual bias correction methods are regression and calibration estimators. It should be mentioned that the “calibration estimators” used in remote sensing is not exactly the same as the “calibration estimators” as defined by Deville and Särndal (1992), more familiar to official statisticians.

Small area estimators are a larger family of more complex methods (Rao, 2003). For this study we do not intend to describe in depth small area estimators, although a brief description is foreseen for the training course.

One of the main purposes of this exercise is to produce reliable indicators of area and area change for a range of land cover/use types at selected NUTS levels based on Copernicus and map products. In addition, the efficiency of the stratification and the contribution of the map products in terms of improvements of the precision of the area and area change estimates can be evaluated by

calculating the **stratification efficiency** (Gallego, 1995) and the **relative efficiency** of the maps combined with reference data (Taylor et al. 1997). Finally, this approach can also be used to identify the level of thematic accuracy (there is always a strong relation between thematic accuracy and precision levels) that would be required to get reliable and precise estimates of area and area change, which for the latter is always the most difficult to obtain. This level could then be used as a target for the improvement of Copernicus Land products in terms of the following:

- Minimum level of thematic accuracy to reliably detect changes for a range of Copernicus Land products and typical themes
- Minimum duration between observation periods to reliably detect change, given a specified thematic accuracy target

Area estimates can be derived directly from the field data alone using the so-called direct expansion method as long as the data has been collected based on a probabilistic sample. The estimate of proportion ( $y$ ) of class ( $c$ ) and its variance are given by:

$$\bar{y}_c = \frac{1}{n} \sum_{i=1}^n y_i \text{ and } var(\bar{y}_c) = \left(1 - \frac{n}{N}\right) \frac{1}{n(n-1)} \sum_{i=1}^n (y_i - \bar{y}_c)^2$$

where:  $y_i$  is the proportion of segment  $i$  covered by class  $c$ ,  $N$  is total number of segments in the region,  $n$  is number of segments in the sample. The proportion of the study region sampled ( $n/N$ ) is the sample fraction. The variance calculation above assumes a single stage sampling, if a 2 stage sampling is applied, this needs to be accounted for as described by Sannier et. (2014). The estimate of class area ( $Z$ ) and variance in study area ( $D$ ) are as follows:

$$\hat{Z}_c = D * \bar{y}_c \text{ and } var(\hat{Z}_c) = D^2 * var(\bar{y}_c)$$

where  $D$  is the area of the region. It is better to compute the estimates first as proportions rather than as absolute areas because this automatically takes account of errors resulting from small localised variations in the scale of segment maps and drawing or digitising errors. The Direct expansion estimators are calculated for each stratum present in the AOI and the total estimate just correspond to the weighted average of the proportions according to the area covered by each stratum. The standard error for the whole area is then the square root of the sum of the variance times the square of the area for each stratum:

$$\sigma_{Total} = \sqrt{\sum D_h^2 \cdot Var_h}$$

where  $D_h$  is the stratum area. The 95% confidence interval is +/- 1.96.  $\sigma_{Total}$ .

However, the confidence interval of the estimate derived from direct expansion is likely to be relatively broad. To improve the precision of the estimates the sampling fraction needs to be increased or field segment data can be combined with classified satellite imagery.

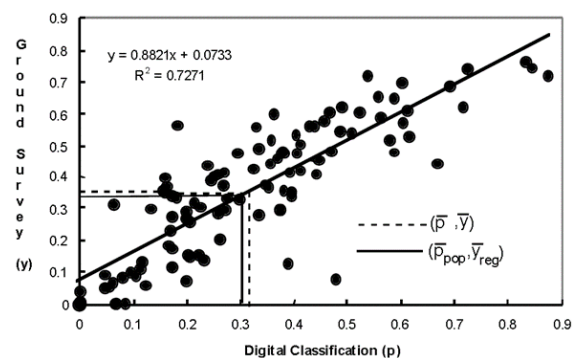
A so-called Model Assisted Regression (MAR) estimator can be applied, this is more reliable than any other area estimation methodology as it provides both an area estimation per cover type together with an indication of its uncertainty.

As a summary, it relies on the combination of area estimates made at the segment level for both reference data and classified satellite imagery. The observation is paired, and a regression analysis is performed as illustrated in Figure 4.

The regression estimator methodology is fully described by Taylor et al. (1997).

The estimation of land cover type areas can be very variable from pixel counts because image classification is affected by misclassification errors affecting classes differently causing omission and commission errors. Area estimates derived from the regression estimator method are corrected from misclassification errors whilst exhibiting a more precise precision estimate than that of the direct expansion estimate thanks to the complete coverage provided by the image classification.

In summary, direct expansion estimates are unbiased, but suffer from high sampling error, pixel counts from classified satellite imagery are biased but have no sampling errors and the combination of ground data and classified imagery are unbiased and exhibit a reduced sampling error.



**Figure 4: Relationship between digital classification and ground survey for wheat in the UK (after (Taylor, 1997)).**

The so-called regression estimator  $y_{reg}$  is calculated based on the regression formulae:

$$y_{reg} = \bar{y} + b * (\bar{p}_{pop} - \bar{p})$$

where  $\bar{y}$  is the mean field data sample value,  $b$  is the slope of the regression line,  $\bar{p}_{pop}$  is the proportion of pixels classified as the land cover in the whole of the region of interest and  $\bar{p}$  is the classified image mean sample value. The variance of the estimate is calculated as:

$$var(\bar{y}_{reg}) = \frac{1}{n} var(y)(1 - r_{py}^2)$$

where  $r_{py}^2$  is the regression coefficient. The variance calculation above assumes a single stage sampling, if a

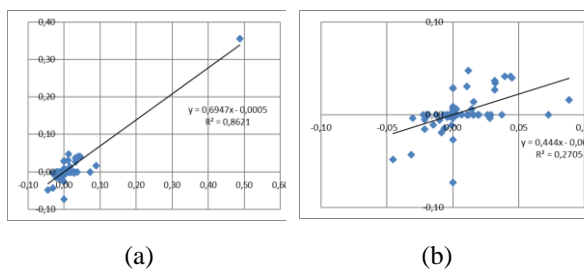
2-stage sampling is applied, this needs to be accounted for as described by Sannier et. (2014).

Therefore, the higher the regression coefficient the smaller the variance and as a result the precision of the estimate. Experience for the MARS programme (Taylor, 1997) showed that high classification accuracy was correlated with high regression coefficient.

The efficiency of the regression estimator is estimated by the relative efficiency, which is the ratio of the radiance from the regression estimator method and the direct expansion estimate:

$$n_{reg} = \frac{1}{1-r_{py}^2}$$

During the MARS project, it was shown that relative efficiency above 2 could easily be obtained, but with multi-temporal image coverage, higher classification accuracy could be achieved as part of this project resulting in reduction of variance by a greater factor. For land cover types that are very distinct such as tropical rainforest, very high relative efficiency i.e. reduction of the uncertainty can be obtained with the Regression Estimator as shown by (Sannier et al., 2016) for forest cover in Gabon for which relative efficiency close to 60 were obtained. This means that the variance of the regression is reduced by a factor of 60 as compared to the direct expansion estimate. In other words, to obtain the same level of uncertainty with the direct expansion i.e. without using the EO-based forest cover map, the sample size would need to be increased by a factor of 60. The 95% confidence interval of the direct expansion estimate is around 2% of the total area of Gabon when it is reduced to 0.25% for the regression estimate. It is also worth noting that this is a rare case for which the pixel count from the map is relatively close to unbiased estimates considering that it is invariably contained within the bounds of the, most precise, regression estimates.



**Figure 5: Illustration of the importance of the quality of the linear regression with (a) all observations and (b) one observation removed.**

One of the issues with the regression estimator as described above is that it is potentially very sensitive to the quality of the linear regression which is sometimes poorly reflected by the  $r^2$  as shown in the example below (see Figure 5).

A slightly adapted estimator is also available in the form of the GREG estimator described by (Särndal et al., 1992, sec. 6.5) which focuses on the average of the

differences from each observation to the mean of observations:

$$\hat{\mu}_{greg} = \frac{1}{N} \sum_{i=1}^n \hat{y}_i - \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)$$

with variance:

$$v\widehat{ar}(\hat{\mu}_{greg}) = \frac{1}{n(n-1)} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2$$

where N is the number of map units, n is the reference set sample size,  $y_i$  is the observation for the  $i^{th}$  reference set sample unit,  $\hat{y}_i$  is the map class,  $\varepsilon = \hat{y}_i - y_i$ , and  $\bar{\varepsilon} = \frac{1}{n} \sum_{i=1}^n \varepsilon_i$ . The variance calculation above assumes a single stage sampling, if a 2-stage sampling is applied, this needs to be accounted for as described by Sannier et. (2014).

It has the advantage of not being sensitive to the quality of the linear regression but could potentially provide an estimate with a larger confidence interval compared with the direct method when the EO-based map is of poor quality for the land cover class considered. Such an approach was applied by Sannier et al. (2014) in Gabon and was applied here over Europe for Tree Cover (no forest definition was applied).

### 3 RESULTS AND CONCLUSIONS

The Stratified Random Sampling or Direct Estimates as well as the GREG or Model Assisted Regression estimates were produced for the entire area of the EEA 38+UK countries covered by the Copernicus High Resolution layer on Tree Cover Density, the estimates are shown in the table below for 2012, 2015 and 2018.

**Table 3. Tree cover area proportion estimates at EEA38+UK level for 2018, 2015 and 2012**

	2018		2015		2012	
	Estimate	CI95%	Estimate	CI95%	Estimate	CI95%
Direct mean	0,305	0,004	0,290	0,004	0,273	0,005
Map (pixel counts)	0,250	n/a	0,269	n/a	0,250	n/a
MAR	0,287	0,002	0,294	0,002	0,256	0,004
Relative Efficiency	1,42		3,32		5,34	

It appears that there has been a significant increase in Tree cover between 2012 and 2015. Which seems to have stabilised in 2018. It should be noted that the combination of sample and map observations provides estimates with reduced uncertainty and that the quality of the HRL TCD appears to have improved substantially between 2012 with a relative efficiency of just 1.42 in 2012 and over 5 in 2018.

#### 4 REFERENCES

- Allen, J.D., 1990. A Look at the Remote Sensing Applications Program of the National Agricultural. *J. Off. Stat.* 6, 393–409.
- Bellhouse D.R., 1988, Systematic sampling, *Handbook of Statistics*, vol. 6, ed. P.R. Krishnaiah, C.R. Rao, pp. 125-146, North-Holland, Amsterdam
- Cochran W., 1977, *Sampling Techniques*. New York: John Wiley and Sons
- Czaplewski, R. L. (2003). Chapter 5: accuracy assessment of maps of forest condition: statistical design and methodological considerations, pp. 115–140. In Michael A. Wulder, & Steven E. Franklin (Eds.), *Remote sensing of forest environments: concepts and case studies*. Boston: Kluwer Academic Publishers (515 pp.).
- Deville J.C., Särndal C.E., 1992, Calibration estimators in survey sampling, *Journal of the American Statistical Association*, Vol 87, pp. 376-382
- Gallego, J. (1995). *Sampling Frames of Square Segments*. Official Publication of the European Communities, EUR 16317, 72 pp.
- Gallego, J. (2004a). Area Frames for Land Cover Estimation: Improving the European LUCAS Survey. *Third International Conference on Agricultural Statistics*. November 2-4, 2004, Mexico.
- GFOI (2016), *Integration of remote-sensing and ground-based observations for estimation of emissions and removals of greenhouse gases in forests: Methods and Guidance from the Global Forest Observations Initiative*, Edition 2.0, Food and Agriculture Organization, Rome.
- Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend. (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* 342 (15 November): 850–53. Data available on-line from: <http://earthenginepartners.appspot.com/science-2013-global-forest>.
- Lowell K., Jaton A. (ed), 1999, Spatial Accuracy assessment: Land information uncertainty in natural resources, Ann Arbor Press, Chelsea, Michigan, USA. 443 pp.
- McRoberts, R. E. (2011). Satellite image-based maps: Scientific inference or pretty pictures?. *Remote Sensing of Environment*, 115(2), 715-724.
- Olofsson, P., Foody, G. M., Stehman, S. V., & Woodcock, C. E. (2013). Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sensing of Environment*, 129, 122–131. doi:10.1016/j.rse.2012.10.031
- Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., Wulder, M.A., 2014, Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment* 148, 42 - 57
- Rao, J.N.K. (2003). *Small Area Estimation*. Wiley, New York.
- Sannier, C., McRoberts, R. E., Fichet, L. -, & Makaga, E. M. K. (2014). Using the regression estimator with landsat data to estimate proportion forest cover and net proportion deforestation in gabon. *Remote Sensing of Environment*, 151, 138-148.
- Sannier, C.; Gallego, J.; Dahmer, J.; Smith, G.; Dufourmont, H.; Pennec, (2016a) A. Validation of Copernicus high resolution layer on imperviousness degree for 2006, 2009 and 2012. Spatial Accuracy Assessment, Montpellier, 5–8 July 2016. <http://www.spatial-accuracy.org/system/files/Accuracy%202016%20Proceedings.pdf>
- Sannier, C., McRoberts, R. E., & Fichet, L. V. (2016b). Suitability of Global Forest Change data to report forest cover estimates at national level in Gabon. *Remote Sensing of Environment*, 173, 326-338.
- Särndal, C.-E., Swensson, B., Wretman, J.H., 1992. *Model assisted survey sampling*. Springer, New York.
- Selkowitz, D. J., & Stehman, S. V. (2011). Thematic accuracy of the National Land Cover Database (NLCD) 2001 land cover for Alaska. *Remote Sensing of Environment*, 115(6), 1401–1407. doi:10.1016/j.rse.2011.01.020.
- Stehman, S. V. (2009). Sampling designs for accuracy assessment of land cover. *International Journal of Remote Sensing*, 30(20), 5243-5272.
- Stehman, S. V., & Czaplewski, R. L. (1998). Design and analysis for thematic map accuracy assessment: fundamental principles. *Remote Sensing of Environment*, 64, 331–334.
- Stehman, S. V., & Foody, G. M. (2008). Accuracy assessment. *The SAGE handbook of remote sensing* (pp. 297-314) doi:10.4135/9780857021052.n21
- Taylor J., Sannier C., Delincé J, Gallego F.J. (1997), *Regional Crop Inventories in Europe Assisted by Remote Sensing: 1988-1993. Synthesis Report* . EUR 17319 EN, Office for Publications of the EC. Luxembourg. 71pp
- Westfall, J.A., Patterson, P.L., Coulston, J.W., 2011. Post-stratified estimation: within-strata and total sample size recommendations. *Can. J. For. Res.* 41, 1130–1139. <https://doi.org/10.1139/x11-031>