Validation of Copernicus Land Monitoring Services and Area Estimation.

Javier Gallego[1], Christophe Sannier[2], Alexandre Pennec[2], Hans Dufourmont [3]
[1]Joint Research Centre of the European Commission.
[2]SIRS: Systèmes d'Information à Référence Spatiale.
[3]European Environment Agency.
DOI: 10.1481/icasVII.2016.g42b

ABSTRACT

Small photographs.

Keywords: Land cover area estimation, High resolution satellite images, Area frame sampling.

1. Introduction

The Land Monitoring Service of the EU (European Union) Copernicus programme (land.copernicus.eu) includes High Resolution Layers (HRL) that provide information on specific land cover characteristics produced from 20 m resolution satellite imagery. The main 5 themes covered are: imperviousness, (sealed soil), tree cover density and forest type, permanent grasslands, wetlands and water bodies. Pixels of 20 by 20 m are aggregated into 100 by 100 m grid cells for final products. The imperviousness layer was the first one to be produced in 2006-2008. New imperviousness layers have been produced for 2009 and 2012. They cover the 33 Member States of the European Environment Agency (EEA) and 6 associated West-Balkan countries representing a total of 6 million km². Some countries are missing at the time of drafting this paper (Spain, Greece, Cyprus and the French overseas regions are missing). The imperviousness HRL captures the spatial distribution of artificially sealed areas, with a degree 1-100% is produced using an automatic algorithm based on a calibrated normalised difference vegetation index (NDVI). The methodology was described by Gangkofner et al. (2010) for the 2009 update and by Lefebvre et al (2013) for the 2012 update. Similar methods have been also applied in the USA for the development of the National Land Cover database (Xian et al. 2011). A density threshold of 30% was used to derive a 0-1 mask. In Europe, it is currently estimated that artificial areas represent less than 5% of the total EEA39 (Büttner et al. 2012) and impervious surfaces only represent a subset of this area. Because of this relatively low share, commission and omission errors may be high even when the overall accuracy of the layer is apparently good.

2. Sampling Scheme.

Similar products are provided by the National Land Cover Database of the Conterminous United States (Wickham et al. 2013) and a recent study from Hansen et al. (2014). Wickham et al. (2013) applied a stratified random sampling approach using the land cover classes as strata. For the European Imperviousness layer, a similar approach was applied but based on a stratified systematic sampling approach using the LUCAS sampling frame (Gallego and Delincé 2010) targeting also imperviousness changes for which relatively low accuracies are reported, in part because the area increase occupies a very small proportion of the total land area. Using the LUCAS based approach improves traceability and sample sharing for assessing several products. Estimation domains are countries or groups of countries with an area greater than 90,000km². In each domain 6 strata were determined as follows:

- Commission 2006-2009-2012: Imperviousness Degree 30-100% in 2006-2009-2012. This category is divided into two strata defined by intersection with the CORINE Land Cover (CLC) artificial and non-artificial classes.
- Omission High Probability 2006-2009-2012: Imperviousness Degree 0-29% & CLC impervious classes 2006-2009-2012
- Omission Low Probability 2006-2009-2012: Rest of the area 2006-2009-2012
- Commission Change 2006-2009: all changes
- Commission Change 2009-2012 : all changes

A sample of 20,164 PSUs (primary sampling units) was selected. PSU were squares of 100 m × 100 m. It was decided to select a minimum of 50 PSU per stratum in each of the 23 zones. In each PSU a grid of 5 x 5 Secondary Sample units (SSUs) with a 20 m step is selected and photo-interpreted on orthophotos (figure 1). If a point falls on the boundary of an impervious element, a shifting rule is applied so that roughly half of the points in this situation are classified as impervious.

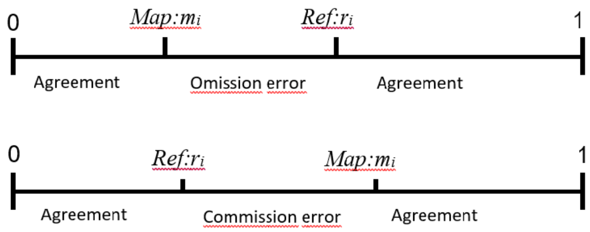
Figure 1: *Example of PSU with a grid of points.*

3. Thematic accuracy.

The usual validation scheme for land cover maps or satellite image classification is based on confusion matrices that assume sharp categories (Congalton and Green, 1999): each point is allocated to one single category as well as each validation unit. When both validation data and product under validation are continuous, a confusion matrix approach can be used by applying a threshold to produce a mask. For the Copernicus imperviousness layer, a threshold of 30% had been foreseen. However better alternatives can be found in the modelling literature to deal with quantitative products (Wilmott, 1981, Legates and Mc Cabe, 1999, Duveiller et al., 2016). Agreement indicators for quantitative parameters have been also widely used in the remote sensing literature (Ji and Gallo, 2006, Silván Cárdenas and Wang, 2008, Meroni et al., 2013).

Among the indicators that have been proposed, we have chosen to quantify the disagreement at the PSU level as the difference between the map value m_i and the reference r_i. If the map value is larger than the reference value for a PSU, it will contribute to the commission error, but it will contribute to the omission error if the oposite happens (Figure 2).

Figure 1: *Commission and omission errors with continuous data between 0 and 1.*



The commission φ and omission ψ errors would be computed as:

$$\varphi = \frac{\sum_i w_i \, pos(m_i - r_i)}{\sum_i w_i \, m_i} \qquad \psi = \frac{\sum_i w_i \, pos(r_i - m_i)}{\sum_i w_i \, r_i} \qquad (1)$$

and the overall accuracy

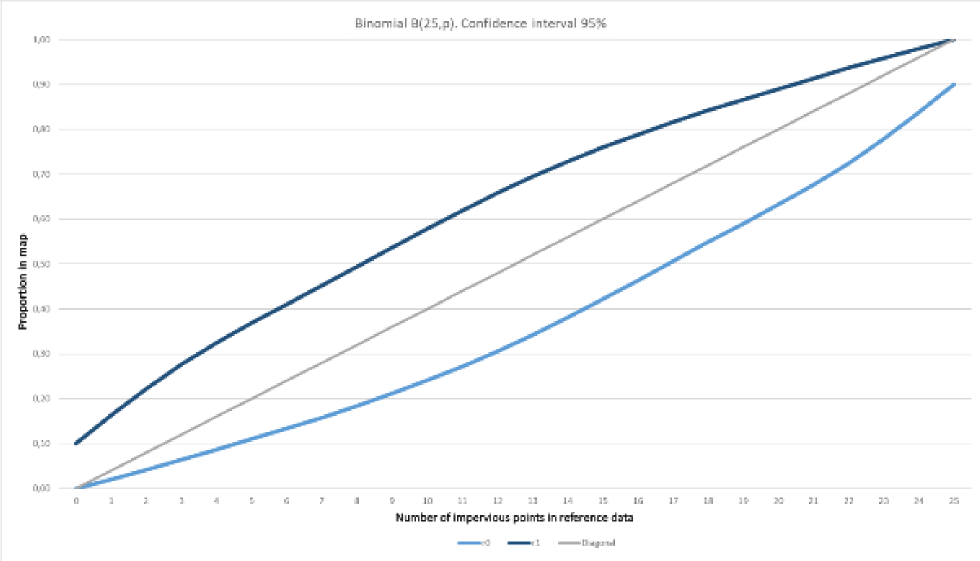
$$\theta = \frac{\sum_i w_i \left(min(r_i, m_i) + min(1 - r_i, 1 - m_i) \right)}{\sum_i w_i} \qquad (2)$$

where w_i is the extrapolation weight (inverse of the sampling probability) and $pos(x)$ is the positive part. With these criteria we have a commission error of 21.9% and an omission error of 40.2%, even if the overall accuracy is 98.4%. Part of the disagreement between reference comes from the incomplete information on the cell for our reference data: if we assume that the map and the reference data are in perfect agreement, there would be still a difference between m_i and r_i because the spatial support is different: m_i refers to the whole PSU while r_i refers to a sample of points inside. Our r_i has a probabilistic distribution that we can approximate by a binomial B(25, p_i) with a somewhat lower variance due to the systematic sampling. This has an impact on the expected commission and omission errors. One possible way to take this into account is delineating a confidence band under the null hypothesis $m_i = p_i$ (figure 3). For any pair (r_i, m_i) inside this band we are not reasonably sure that $m_i \neq p_i$. A possible way to deal with this source of uncertainty is quantifying the commission and omission errors:

$$\varphi = \frac{\sum_i w_i \, pos(m_i - r_{0i})}{\sum_i w_i \, m_i} \qquad \psi = \frac{\sum_i w_i \, pos(r_{1i} - m_i)}{\sum_i w_i \, r_i} \qquad (3)$$

where r_{0i} and r_{1i} are the closest points in the confidence band. With these criteria the commission error is around 10% and the omission error around 20%.

Figure 3: *95% confidence band (r_i,m_i) under the hypothesis $m_i = p_i$.*



4. Area estimation.

A naïf approach for area estimation using classified satellite images is pixel counting or equivalent for classifications that focus on the proportion of a parameter, as it happens in our case:

$$\tilde{A} = \sum_i m_i \qquad (4)$$

The result for this estimator for the zones covered at the moment (excluding Spain, Greece, Cyprus and French overseas territories) is 2.06%, approximately 107,000 km^2. It is well known that this area estimation has a bias that can be very large (Bauer et al, 1978, Houston and Hall, 1984, Czaplewski, 1992) and roughly corresponds to the difference between the commission and the omission errors (Carfagna and Gallego, 2005). A simple reasonable idea may be correcting the bias with the difference between the commission and omission errors computed on a confusion matrix. Unfortunately it is still frequent that remote sensing practitioners compute confusion matrices on the basis of a purposive set of units instead of a proper probability sample. In the scientific literature purposive samples have nearly disappeared thanks to the effort of a number of authors (Congalton and Green, 1999, Stehman, 2009, Olofsson et al., 2014), but they are still frequent in reports of projects carried out for various institutions and remain unpublished. We can simulate a purposive sample using our sample by omitting the weights in equations (1). We would get table 1 that suggests us reducing the estimate by approximately the difference between 26.6% (apparent commission error) and 20.8% (apparent omission error) to obtain an estimate just above 100,000 km^2. If we apply the correct weights we get the confusion matrix in table 2, in which we find that the omission error (40.7%) is much higher than the commission error (21.9%). The figures of the confusion matrix can be interpreted as area in km^2 and we can obtain an area estimation by adding something that is close to the difference (40.7-21.9)%. A better estimate is obtained by simply adding the difference between the off-diagonal terms: 107,000+57,000-23,200 ~ 140,000 km^2, i.e the 2.74% of the territory.

Table 1: *Unweighted confusion matrix.*

		Map			
		Impervious	Other	Total	Omission error
Reference	Impervious	4280.4	1124.1	5404.5	**20.8%**
	Other	1553.3	13041.2	14594.5	
	Total	5833.7	14165.3	19999	
	Commission error	**26.6%**			

Table 2: *Weighted confusion matrix.*

		Map			
		Impervious	Other	Total	Omission error
Reference	Impervious	82859	56981	139840	40.7%
	Other	23232	4961592	4984824	
	Total	106091	5018573	5124664	
	Commission error	21.9%			

The most reasonable principle is combining the two sources of information we have: more accurate data on a sample and less accurate but nearly-exhaustive information from the image classification (i.e. the imperviousness layer we are validating). We have given above a coarse view on starting with remote sensing based estimates using reference data on a sample. The opposite approach is more frequent: a sample-based estimate modified using the co-variate provided by a classified image. At the first step of this approach we compute an estimate of 139.800 km^2 with a coefficient of variation of 3%. The estimate is very close to the estimate obtained with a simple correction on commission-omission errors computed from the weighted confusion matrix. The observation of the variances by stratum, not reported here, suggests that major improvements can be obtained by applying a Neyman allocation in the future.

A better estimator is generally obtained combining the more accurate data on a sample (reference data) with less accurate exhaustive data obtained from satellite images. If both data are quantitative, the standard technique to do so is the regression estimator (Cochran, 1977). For the impervious areas in 2012 in the zones considered for the Copernicus validation, table 3 reports the naïf estimator obtained directly from the image-derived values, the stratified sampling estimator from the reference data and the stratified regression estimator. Just by chance the overall regression estimator for the study area nearly coincides with the extrapolation from the reference data (139,800 km^2) with a coefficient of variation of 2.77%. This means that the relative efficiency is around 1.17, a rather modest value.

Table 2: *Different estimates of impervious areas in 2012 (in km^2).*

	Naïf from image	From reference data	Regression estimator	CV regression (%)
Turkey	6342	11032	10914	23
France	15566	21111	21827	5
Sweden	2156	5223	5196	15
Germany	18346	16999	17269	4
Finland	2010	2472	2448	15
Norway	895	1395	1457	11
Poland	7831	8354	8239	9

Italy	11449	19894	19776	7
UK-Ireland	9611	16212	15641	7
Romania	3915	4044	3864	14
Bulgaria	2154	1680	1692	6
Iceland	142	81	82	3
Hungary	2987	2830	2708	10
Portugal	2719	2778	2798	7
AT-CH-LI	2877	4397	5629	7
BENELUX- DK	7237	9535	8650	7
AB-MN-MK-SB-KO	2194	1833	1784	8
SI-HR-BH	2116	2488	2284	13
CZ-SK	3989	5410	5350	9
Baltic Republics	2426	2065	2225	15

5. Conclusions.

The direct area estimation by pixel counting in a classified image, or equivalent approaches, is known to have a bias because commission and omission errors. We have illustrated with the example of the estimation of impervious areas in Europe that a simple correction of bias using a confusion matrix gives acceptable results if the confusion matrix has been properly weighted with the inverse of sampling probabilities. In exchange if these weights are ignored, the correction can be completely wrong and even "correct" the estimates in the wrong direction. In our example the bias of the naïf (direct) estimation of impervious area from classified satellite images is above 20%, even if the overall accuracy of the classification is above 98%. This observation has an implication on the use of remote sensing for area estimation that is not new, but is worth reminding:

- the risk of bias in direct area estimation from classified images is particularly strong if the targeted classes occupy a small proportion of the geographic area.
- Bias correction with a sample of more accurate and approximately unbiased data requires applying the correct weights from the sampling plan.

REFERENCES.

Bauer, M.E., Hixson M.M., Davis B.J., Etheridge J.B. (1978), Area estimation of crops by digital analysis of Landsat data. *Photogrammetric Engineering and Remote Sensing*, 44: 1033-1043.

Büttner G (2012) *Guidelines for verification and enhancement of high resolution layers produced under GMES initial operations (GIO) Land monitoring 2011 – 2013*, EEA, 47p.

Büttner G., Kosztra B., Maucha G. and Pataki R. (2012) *Implementation and achievements of CLC2006*, ETC-LUSI, EEA, 65p.

Carfagna E., Gallego F.J., 2005, The use of remote sensing in agricultural statistics, *International Statistical Review* 73, 3, 389-404

Cochran W., 1977, *Sampling Techniques*. New York: John Wiley and Sons

Czaplewski R.L. (1992), Misclassification bias in areal estimates. *Photogrammetric Engineering and Remote Sensing*. 58 (2): 189-192.

Duveiller, G., Fasbender, D., & Meroni, M. (2016). Revisiting the concept of a symmetric index of agreement for continuous datasets. *Scientific Reports, 6* doi:10.1038/srep19401

Gallego, F.J., Delincé J. (2010) The European Land Use and Cover Area-frame statistical Survey (LUCAS), in *Agricultural Survey Methods*, ed: R. Benedetti, M. Bee, G. Espa, F. Piersimoni, Ch. 10. pp. -168, John Wiley & sons.

Gallego, J., & Delincé, J. (2010). The European Land Use and Cover Area-Frame Statistical Survey. *Agricultural survey methods*, 149-168.

Gangkofner, U., Weichselbaum, J., Kuntz, S., Brodsky, L., Larsson, K., & De Pasquale, V. (2010). Update of the European High-resolution Layer of Built-up Areas and Soil Sealing 2006 with Image2009 Data. 30th *EARSeL Symposium 2010: Remote Sensing for Science, Education and Culture*

Hansen, M. C., Egorov, A., Potapov, P. V., Stehman, S. V., Tyukavina, A., Turubanova, S. A., ... & Kommareddy, A. (2014). Monitoring conterminous United States (CONUS) land cover change with web-enabled landsat data (WELD). *Remote sensing of Environment*, 140, 466-484.

Houston A.G., Hall F.G. (1984), Use of satellite data in agricultural surveys. *Communications in Statistics Theory and Methods*, 23, pp. 2857-2880.

Ji, L., & Gallo, K. (2006). An agreement coefficient for image comparison. *Photogrammetric Engineering and Remote Sensing, 72*(7), 823-833. Retrieved from www.scopus.com

Lefebvre, A.; Beaugendre, N.; Pennec, A.; Sannier, C.; Corpetti, (2013) T. Using data fusion to update built-up areas of the 2012 European High-Resolution Layer Imperviousness. In *Proceedings of the 33rd EARSeL Symposium Conference*, Matera, Italy, 3–6 June 2013; pp. 321–328.

Legates, D. R., & McCabe Jr., G. J. (1999). Evaluating the use of 'goodness-of-fit' measures in hydrologic and hydroclimatic model validation. *Water Resources Research, 35*(1), 233-241. doi:10.1029/1998WR900018

Meroni, M., Atzberger, C., Vancutsem, C., Gobron, N., Baret, F., Lacaze, R., . . . Leo, O. (2013). Evaluation of agreement between space remote sensing SPOT-VEGETATION fAPAR time series. *IEEE Transactions on Geoscience and Remote Sensing, 51*(4), 1951-1962. doi:10.1109/TGRS.2012.2212447

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

Silván-Cárdenas J.L., Wang L., 2008, Sub-pixel confusion-uncertainty matrix for assessing soft classifications, Remote Sensing of Environment, 112 1081-1095

Stehman, S. V. (2009). Sampling designs for accuracy assessment of land cover. *International Journal of Remote Sensing, 30*(20), 5243-5272.

Wickham, J. D., Stehman, S. V., Gass, L., Dewitz, J., Fry, J. A., & Wade, T. G. (2013). Accuracy assessment of NLCD 2006 land cover and impervious surface. *Remote Sensing of Environment, 130*, 294-304.

Willmott, C. J. (1981). On the validation of models. *Physical Geography, 2*(2), 184-194. doi:10.1080/02723646.1981.10642213

Xian, G., Homer, C., Dewitz, J., Fry, J., Hossain, N., & Wickham, J. (2011). Change of impervious surface area between 2001 and 2006 in the conterminous United States. *Photogrammetric Engineering and Remote Sensing, 77*(8), 758-762.