Comparing CORINE Land Cover with a more detailed database in Arezzo (Italy).

Javier Gallego JRC , I-21020 Ispra (Varese) ITALY e-mail: javier.gallego@jrc.it

Keywords: land cover, accuracy assessment, area estimation

Abstract

The Italian Statistical Institute (ISTAT) promoted a pilot project for a land use / land cover database at 1:25.000 scale over a test area in the Province of Arezzo, 2000 km^2 wide. The nomenclature was derived from Corine Land Cover with two additional levels.

In this paper we analyse the impact of the difference in scale of maps and minimum size of mapping units between CORINE Land Cover and the ISTAT land cover map.

A straightforward computation of confusion matrices gives a pessimistic view of the agreement between both land cover maps, but most of the apparent disagreement is due to different scales. An assessment with a simplified nomenclature shows that there is a very good agreement for main classes with two exceptions: pasture and non-forest natural vegetation, that represent only a limited area (less that 1% for pasture). The disagreement for these two categories is partly due do the difficulties in interpreting the nomenclature.

1 Introduction

When CORINE Land Cover (CLC) is compared with land cover information at a different scale, premature conclusions could be drawn if a straightforward comparison is seen as a quality assessment. If the comparison between two geospatial land cover data layers takes into account their different scales, the conclusions change. We give a procedure to perform such a comparison trying to remove the effects of different scales and collocation inaccuracy.

We also illustrate the fact that CLC, should not be directly used for area estimation by direct polygon area measurement. However CLC, used as covariable, is extremely useful for land cover area estimation (Gallego et al, 1999)

1.1 The Arezzo ILC database and CORINE Land Cover

ISTAT, the National Statistical Institute of Italy, has developed a land use-land cover database on an area of 200,000 hectares in the Arezzo province, in central Italy, with the geometric accuracy of 1:25,000 scale. Below we call this land cover map ILC. The main purpose of this database was testing the possible extension at country level (ISTAT, 1998). The pilot project was partially funded by the European Union (EU).

The territory is divided into polygons classified into land use - land cover classes. The nomenclature has five levels; the three first levels coincide with the CORINE land-cover nomenclature (EC, 1993); a fourth level is added for urban areas and a forth and a fifth level are added for forests and semi-natural areas.

The project used digital black and white ortho-photos with a resolution of one meter and three Landsat-TM quarter-scenes, taken on the following dates: 10/05/97, 26/10/97, 06/05/98. The photo-interpretation was carried out drawing polygons directly on screen with data at the level of census sections as ancillary information. The minimum mapping unit was 1,56 hectares (1 hectare for urbanised areas).

The polygon layer was integrated with a vector layer composed of linear and point features, referring to railways, highways, major roads and rivers and topographic names. Linear and point features were digitised on screen taking into account the digital ortho-images geometry, after identification on 1:250.000 scale maps. The database has undergone several quality checks, including independent photo-interpretation of samples of arcs, polygons and points (Napolitano et al, 2000).

Figure 1 gives a visual overview of both land cover maps (ILC and CLC) and Table 1 shows their main differences. In the pilot area there is a dominant presence of forest with significant agricultural and urban areas in the valleys.

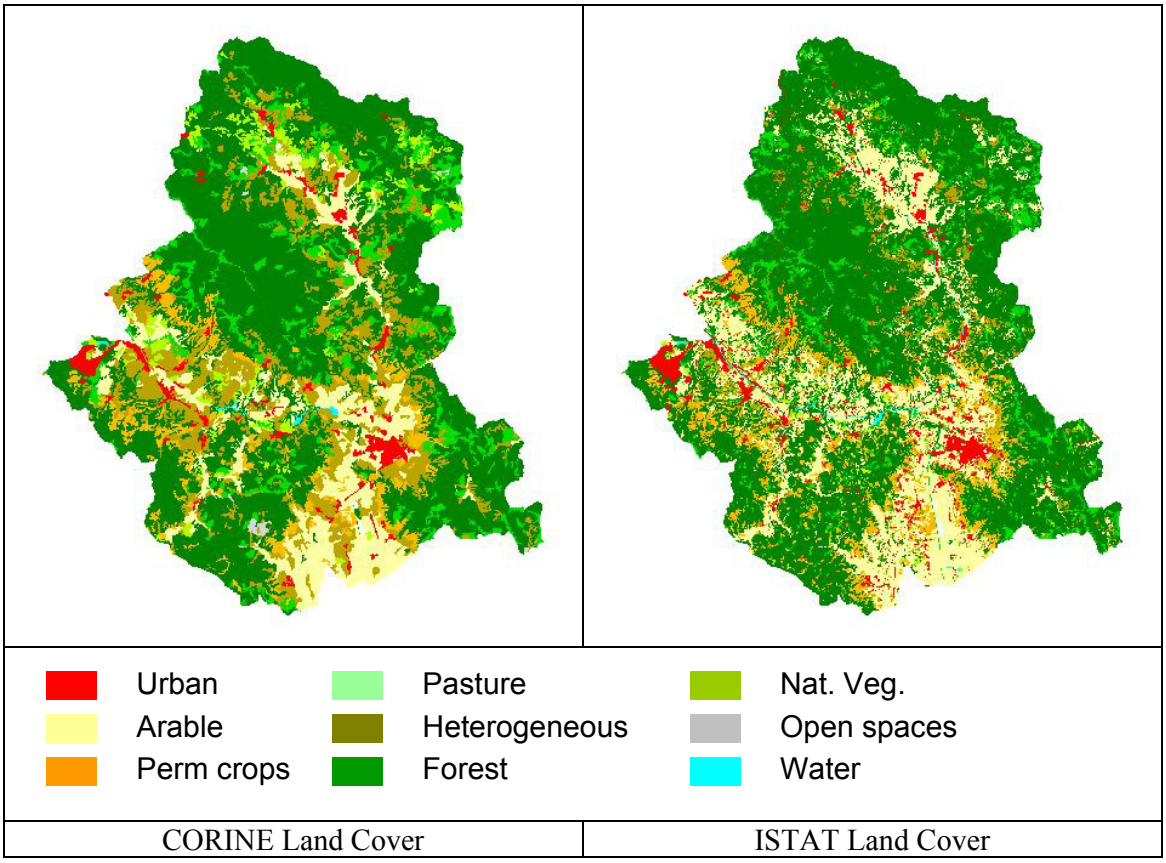

Figure 1: CORINE and ILC land cover maps recoded into 9 classes

	CORINE Land Cover	Arezzo ILC
minimum mapping unit	25 ha	1.56 ha (1 ha for urban areas)
location accuracy	100 m	25 m
Area covered	>3 Million km^2	2000 km^2 . Extension to 300,000 km^2 proposed.
Nomenclature	3 levels, 44 classes	same with 2 additional levels.

Table 1: Main differences between CLC and ILC

2 Comparing total area per class

Land cover maps are sometimes used for area estimation of a land cover class by simply adding the area of the polygons labelled as belonging to that class. This approach is rather naïve and can lead to a serious bias if the mapping scale is not detailed enough or the thematic accuracy is not very high (Gallego et al, 1999). We can compare the total area obtained by this method from CORINE Land Cover and the ILC map for some groups of land cover classes (Table 2).

On the other hand **interpreting the strong differences that appear for some classes as a sign of poor accuracy of CLC would be also naïf**. These differences appear mainly because CLC respects its scale specifications. For example urban areas under 25 ha do not appear in CLC but may appear in ILC and certainly should be counted for statistics. Interpreting the total area of ILC as statistics is also inadequate. Isolated buildings under 1 ha should not appear in ILC and should be counted as artificial for land cover statistics.

	ILC	CORINE
Artificial	84.9	57.1
Arable land	375.9	234.8
Permanent crops	141.0	99.7
Pastures	21.8	66.9
Heterogeneous agriculture	110.8	328.2
Forest	1117.6	1029.5
Natural vegetation	132.4	167.2
Open spaces and wetland	1.7	5.5
Water	5.3	2.3

Table 2: Total area in km^2 per land cover class from CORINE and ILC land cover maps

Part of the disagreement can be attributed to the fact that the images were taken on different dates, part can be due to the scale effect and part to different thematic accuracy levels. Compared with CORINE Land Cover, the area labelled as “heterogeneous agriculture” is significantly reduced and attributed to “pure” classes. This would explain, at least partly, the increase of arable land, permanent crops and forest. However, if we remove the polygons coded in CORINE Land Cover as “heterogeneous”, we still have 26.000 ha classified as arable land by ILC, i.e. 11% more than CORINE Land Cover. We shall see below that the thematic disagreement for arable land is very low; therefore a significant part of the difference in polygon area comes from the scale effect.

Another part of the disagreement may be due to a different interpretation of the nomenclature or to photo-interpretation errors, but it is difficult to know the impact of each source of disagreement without a suitable ground survey. We try below to resolve the impact of each source of disagreement.

3 Statistical analysis of the differences between the two maps

Both land cover maps have been rasterised with a resolution of 50 m for the present analysis. These 50m × 50m cells are called "pixels". A visual inspection of the overlay of both land cover maps in UTM 32 indicates that the co-location accuracy is generally within 50 m (Figure 2), and nearly always within 100 m, i.e. within CORINE Land Cover specifications.

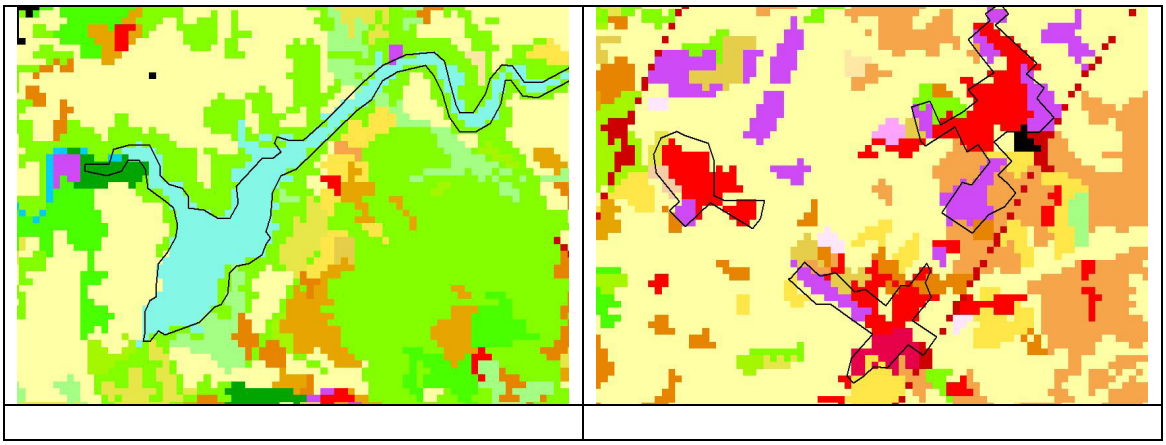

Figure 2: CORINE Land Cover poygons with ILC map rasterised at 50 m.

If we disregard co-location inaccuracy, we can describe the disagreement through a matrix (often named confusion matrix) with elements $A_{cc'}$ = area of the pixels with code *c* in CLC and *c'* in ILC. This confusion matrix would be a naïf description of the disagreement. To eliminate the part of the disagreement due to co-location tolerance (100 m), we eliminate a buffer 2 pixels wide on each side of the CORINE Land Cover polygon borders. This corresponds to eliminating a 200m wide corridor around the polygon limits (Figure 3).

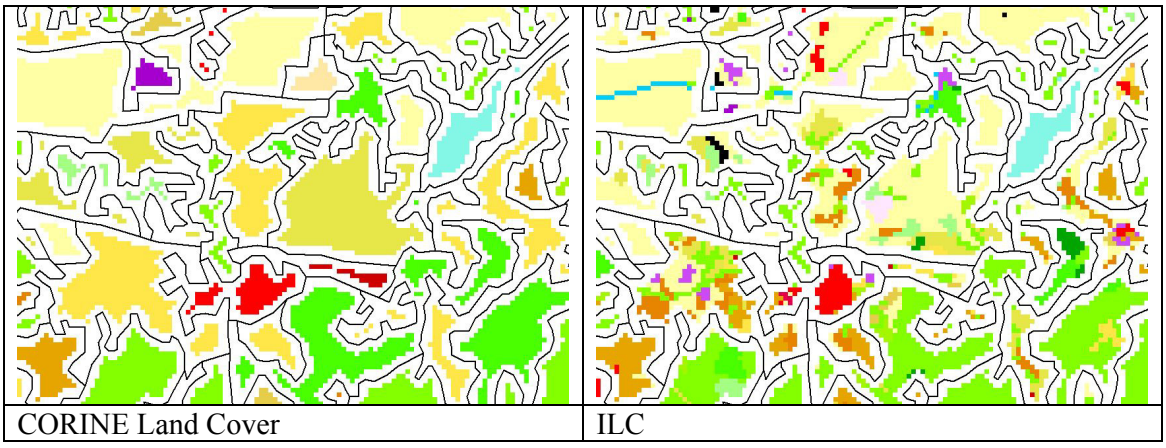

Figure 3: Buffer of 2 pixels eliminated around CORINE Land Cover limits with rasterised CORINE Land Cover

3.1 Pixelwise (naïf) commission and omission disagreement.

The accuracy of a rasterised land cover map can be assessed through the confusion matrix, that gives the number of pixels $A_{cc'}$ for which the true land cover is *c* and the land cover given by the map is *c'*. The omission error for class *c* refers to the proportion of pixels for which the truth is *c* and the map gives a different class. The commission error is the proportion of pixels represented as *c* by the map, but having a different true land cover.

Here the usual terms "confusion matrix" and "commission and omission errors" are substituted by "disagreement matrix" and "commission and omission disagreement" to stress the fact that two different land cover maps are compared: none of the maps can be considered as the truth and part of the disagreement can be explained by the different scales; in this case it should not be considered as an error.

The agreement between two land cover maps can be computed ast he overall % of pixels with coinciding codes, i.e. in the main diagonal of the disagreement matrix. The kappa statistic (Bishop et al, 1975) is a better measure of agreement: its value is close to 0 when the coincidences are random and 1 when both maps coincide perfectly. Table 3 shows both agreement measures before and after removing buffers. The comparison gives an idea of the part of the disagreement that can be attributed to co-location inaccuracy.

	Total area *1000 ha	% coincident	kappa
no buffer	199	56	0.42
buffer 200 m	112	67	0.52

Table 3: Agreement from a confusion matrix between CORINE and ILC.

	Commission		Omission	
	No buffer	buffer 200m	No buffer	buffer 200m
Urban dense	89	87	46	37
Urban discont.	59	62	60	57
Indus-comm	64	62	36	21
Road-rail	96	99	90	88
Mines	40	32	17	3
Arable non irrigated	51	39	24	19
Vineyards	73	76	54	43
Fruits	46	47	30	22
Olive	49	41	35	24
Pastures	76	78	92	92
Arable+perm	80	81	94	94
Complex agr	59	55	87	87
Agric + nat.veg	78	75	92	92
Forest broad leaved	27	16	14	10
Coniferous	60	58	33	18
Forest mixed	63	64	87	85
Nat. Grass	86	88	60	67
Moors-heath	69	70	63	55
Wood-shrub	87	85	96	95
Water bodies	45	63	30	0

Table 4: Commission and omission disagreement by pixel (in %).

Table 4 reports commission and omission disagreements with and without buffer for categories that have at least 100 ha for both maps. We should stress again that **pixelwise disagreement can be very misleading:** both maps can be perfectly consistent with each other, and have a high % of disagreement by pixel because they represent the same reality at different scale. We shall see below that both maps have a very good agreement when the scale effect is removed.

3.2 Recoding into 9 classes.

To simplify the analysis of different types of disagreement, we have recoded both maps into 9 major classes. **Figure 1** shows that both land cover maps have a similar pattern, with the obvious smoothing effect for CORINE Land Cover. Tables 5, 6 and 7 below report main agreement parameters when a pixelwise overlay is performed.

	Total area *1000 ha	% coincident	kappa
no buffer	199	68	0.51
buffer 200 m	125	77	0.61

Table 5: Agreement between CORINE and ILC with 9 classes.

	Commission No buffer	Commission buffer 200m	Omission No buffer	Omission buffer 200m
Urban	47.4	44.4	22.3	8.2
Arable	50.4	40.2	20.1	14.7
Permanent crops	51.6	49.9	33.5	23.6
Pastures	76.4	79.5	92.3	92.3
Heterogeneous	43.5	35.8	80.8	80.4
Forest	17.3	9.2	10.0	6.0
Other nat. veg.	61.5	63.6	69.5	66.3
Open spaces	80.2	93.8	42.2	18.4
Marsh and water	63.9	75.4	35.0	32.0

Table 6: Commission and omission disagreement with 9 classes (%)

		ISTAT Land Cover									
	ha	Urban	Arable	Perm. crops	Pastures	Heterogeneous	Forest	Other nat.veg.	Open spaces	Water	Total
CORINE	Urban	**2601**	105	10	0	46	34	29	0	10	2834
	Arable	532	**13355**	292	89	513	545	200	5	120	15650
	Perm. crops	126	258	**3160**	18	219	299	53	0	4	4136
	Pastures	66	1597	25	**201**	277	179	273	5	2	2623
	Heterogeneous	1043	6442	2223	406	**3162**	2188	660	7	9	16141
	Forest	180	393	527	163	435	**72071**	2916	16	8	76710
	Other nat.veg.	129	178	68	104	262	3960	**2405**	25	3	7133
	Open spaces	0	0	0	0	0	1	0	**4**	0	5
	Water	0	1	0	0	0	0	17	6	**51**	74
	Total	4677	22329	6305	981	4914	79277	6553	68	207	125306

Table 7: Pixelwise disagreeement matrix with 9 classes and 200 m buffer.

Off-diagonal figures correspond at first sight to disagreements, but they do not take into account the scale effect. For example the 2 ha in the case "*pastures×water*" correspond to a small pond in a pasture area, and CORINE Land Cover respects its specifications not reporting it; therefore it cannot be considered as a real disagreement. There are hundreds of other false disagreements in the test site that distort Table 7.

3.3 Scale-corrected thematic disagreement by CORINE Land Cover polygon.

If a CLC polygon has 80% of arable land, it is correctly classified according to the specifications of this land cover map. Similar criteria apply for the main CLC categories, excepting the ones that correspond to heterogeneous landscapes.

We have followed the procedure described below to remove the part of the disagreement between both land cover maps due to the different scales:

- The assessment is made on the nomenclature grouped into 9 classes. The class "heterogeneous agriculture" does not follow the rules below. We eliminate as well the class "burnt areas";
- A buffer of 200 m is removed around CORINE Land Cover polygon borders (100 at each side) to avoid counting location inaccuracy as disagreement;
- If more than 70% of the pixels have the same code for the ILC map, the whole polygon is considered in agreement;
- If the % of different code is between 30 and 70, the polygon is considered partially in disagreement;
- If less than 30% of the pixels have a different code for the ILC map, the whole polygon is considered in disagreement;

For example if a polygon of 100 ha (after removing the buffer) has been coded "arable" in CORINE Land Cover and ILC reports:

- 85 ha arable and 15 permanent crops, it will no contribute to the disagreement.
- 60 ha arable and 40 permanent crops, it will contribute as 40 ha disagreement
- 20 ha arable and 80 permanent crops, it will contribute as 100 ha disagreement

CORINE Land Cover	area excluding buffer (Kha)	% disagreement
Urban	2.8	2.2
Arable	15.2	1.4
Permanent crops	3.9	6.4
Pastures	2.3	93.6
Heterogeneous	13.0	.
Forest	76.2	0.3
Other nat. veg.	6.8	68.6
Open spaces	0.2	0
Marsh and water	0.1	32

Table 8: Thematic disagreement with a grouped nomenclature.

Table 8 reports the disagreement rates. This table confirms several expected facts:

- The agreement is very good for urban, arable, forest, and open spaces.

- There are major discrepancies on pasture and natural vegetation, that represent 7.5% of the area (excluding buffers). This divergence probably comes from the difficulty to interpret the nomenclature.

Other figures are more surprising but may have some explanation:

- The disagreement for "marsh and water" comes from a single marsh polygon.
- The low disagreement of permanent crops increases our confidence in CLC for this land cover class, especially difficult to photo-interpret.

References

Bishop Y., Fienberg S., Holland P., 1975, *Discrete Multivariate Analysis*, M.I.T. press, Cambridge, Ma.

EC, 1993, *CORINE Land Cover; guide technique*, Report EUR 12585EN. Office for Publications of the European Communities. Luxembourg,. 144 pp

Gallego, F.J. Carfagna, E., Peedell S., 1999, The use of CORINE Land Cover to improve area frame survey estimates. *Research in Official Statistics*, Vol 2, no 2, pp. 99-122

ISTAT, 1998, - *Capitolato tecnico per il progetto pilota del database sull'uso e la copertura del suolo in scala 1:25.000 su un'area test.* Rome

Napolitano P., Carbonetti G., Gallego J., 2000, Accuracy Assessment and Validation of a Land Use-Land Cover Database in Arezzo (Italy). Proceedings of *Accuracy 2000*, Amsterdam, July 12-14, 2000, pp. 233-236.