Using CORINE Land Cover to map population density.

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

1 Introduction

Population density data are available to the European Commission (EC) at the level of the commune (NUTS5). The size of communes is very heterogeneous across the EU. Hence this level of spatial resolution may be insufficient in many cases for planning or modelling purposes or to assess the impact of EU policies. In some countries, as in France, where most communes have a rather small area (approx. 15 km^2 in the average), the resolution may be sufficient, but it is clearly insufficient in other countries where the communes tend to be larger.

CORINE Land cover (CLC) gives useful geo-referenced information for disaggregation. This geographic database provides information that is spatially more detailed than the commune limits. A certain commune may contain for example one part of dense urban nucleus, agricultural land with some sparse population, and natural vegetation areas with very little or no population. The objective is to disaggregate population data, imputing different densities to different land cover categories. One possible approach to tackle this problem might be based on the EM algorithm (Dempster, 1977, Ambroise and Govaert, 1998), but the underlying parametric model may be debatable. Here we test a more empirical method.

2 Available Data.

The latest population data available to the Commission at commune level correspond to 1991, therefore the commune boundaries of the SABE database, version 1991 have been used (http://www.megrin.org/SABE/Sabe.html). The CLC data have been used as raster data (1 ha pixel).

For the computations hereafter, in the communes in which CORINE Land Cover is partially missing, the average population density of the commune is assigned to the areas with missing data and the general rule is appled to the rest. CLC is not available for Sweden and Finland, and the SABE commune boundaries are not available for Scotland. These areas have been excluded from the study (the column UK corresponds to England, Wales and Northern Ireland). The French département “Seine St Denis” has been also excluded because of code errors in the version of CLC used.

2.1 Communes without urban area in CORINE Land Cover.

A large number of communes appear as not having any urban area (class 1.1) according to CORINE Land Cover.

Towards agri-environmental indicators. EEA. In press.

	<1000 inhab		1000-5000 inhab		>5000 inhab		Total	
	N communes	population (*1000)	N communes	population (*1000)	N communes	population (*1000)	N communes	population (*1000)
AT	488	301	772	1353	5	40	1265	1695
DE	1860	656	185	268	0	0	2045	924
ES	1269	247	151	333	22	157	1442	737
FR	12860	2699	80	108	2	11	12942	2819
GR	2841	865	110	164	5	77	2956	1107
IE	1973	767	109	143	0	0	2082	910
IT	886	410	357	617	6	48	1249	1074
LU	12	6	1	1	0	0	13	7
NL	2	1	2	2	0	0	4	4
PT	1585	742	624	1107	10	63	2219	1912
UK	106	79	509	878	1	5	616	961
total	23882	6773	2900	4975	51	403	26833	12151

Table 1: Communes with no urban area in CORINE Land Cover

There are several possible explanations for the absence of urban area in CORINE Land Cover for a commune:

- The area of the built area is below the CORINE Land Cover threshold (25 ha). This is the explanation for the great majority of cases.
- Geometric inconsistencies between CORINE Land Cover and SABE. For example all the urban area of a commune may be in a linear development along the coast, but appears to fall outside the commune because of location inaccuracy.
- Errors in the data (CORINE Land Cover or SABE). In a few cases, crossing population data with commune borders can help focus attention on inconsistencies between land cover and population density. Some of them might turn out to be errors in data sets. Anomalies have been detected for several communes of Seine St Denis in the outskirts of Paris. Although this area has been excluded, similar errors may still remain in the data.

The 12.1 million inhabitants in this category of commune represents less than 4% of the 327 million inhabitants of the studied area, but it is a significant part of the rural population; hence this case must be considered specifically. The most populated communes without CLC urban classes correspond to areas with very scattered settlements and no major error could be detected in CLC.

3 Modified areal weighting with given coefficients.

Population data can be disaggregated with the help of CORINE Land Cover assuming that the ratio between the population density of two land cover classes is the same for any commune. This a simplified version of modified areal weighting. We can initially assume that the coefficients are known.

We call
X_m : population in commune *m*.
S_{cm} : area of land cover type *c* in commune *m*.

Towards agri-environmental indicators. EEA. In press.

Y_{cm} : density of population for land cover type c in commune m. Inside each commune Y_{cm} : assumed to be proportional to given coefficients U_c for each land cover type.

$$Y_{cm} = U_c W_m \quad (1)$$

W_m : adjustment factor to ensure that the total population in each commune matches the administrative data.

therefore $$X_m = \sum_c S_{cm} Y_{cm} \quad (2)$$

and it was assumed that $$X_m = \sum_c S_{cm} U_c W_m \quad \Rightarrow \quad W_m = \frac{X_m}{\sum_c S_{cm} U_c} \quad (3)$$

Hence the densities were computed in a first approach as $$Y_{cm} = U_c \frac{X_m}{\sum_c S_{cm} U_c} \quad (4)$$

This disaggregation has been carried out with an initial set of coefficients provided by the EEA for an aggregated CORINE Land Cover nomenclature (

grouped class	Initial coefficient U_c	CORINE Class	Label
1	32	111	Continuous urban fabric
2	25	112	Discontinuous urban fabric
3	1	121	Industrial or commercial units
4	1	122, 123, 124	Road and rail networks, ports, airports
5	1	141, 142	Green urban areas, Sport and leisure facilities
6	3	211	Non-irrigated arable land
7	3	212	Permanently irrigated land
8	1	213	Rice fields
9	5	22	Permanent crops
10	3	231	Pastures
11	5	241	Annual and permanent crops associated
12	5	242	Complex cultivation patterns
13	3	243	Agriculture, with natural vegetation
14	1	244	Agro-forestry areas
15	1	31, 324	Forest and woodland
16	1	321, 322, 323	Other natural vegetation
17	0	13, 33, 4, 5	Mine, dump and construction sites, sand, rock and burnt areas, glaciers, wetland and water

Table 2). For the other classes we assume there is no population.

grouped class	Initial coefficient U_c	CORINE Class	Label
1	32	111	Continuous urban fabric
2	25	112	Discontinuous urban fabric
3	1	121	Industrial or commercial units
4	1	122, 123, 124	Road and rail networks, ports, airports
5	1	141, 142	Green urban areas, Sport and leisure facilities
6	3	211	Non-irrigated arable land
7	3	212	Permanently irrigated land
8	1	213	Rice fields
9	5	22	Permanent crops
10	3	231	Pastures
11	5	241	Annual and permanent crops associated

Towards agri-environmental indicators. EEA. In press.

12	5	242	Complex cultivation patterns
13	3	243	Agriculture, with natural vegetation
14	1	244	Agro-forestry areas
15	1	31, 324	Forest and woodland
16	1	321, 322, 323	Other natural vegetation
17	0	13, 33, 4, 5	Mine, dump and construction sites, sand, rock and burnt areas, glaciers, wetland and water

Table 2: grouped CORINE Land Cover classes and initial coefficients

4 Search of weighting coefficients.

Assuming an approximately homogeneous behaviour of W_m , the expression

$$X_m = \sum_c S_{cm} U_c W_m \qquad (5)$$

can be interpreted as

$$X_m = \sum_c S_{cm} U_c W + \varepsilon_m \qquad (6)$$

where the residuals ε_m are small, i.e. we can write it as a simple linear regression. A simple regression gives completely unacceptable coefficients with several negative values and very high values for classes such as green urban or sport areas. This phenomenon confirms that the approach is too simplistic. In fact the weighting coefficients are not the same for all communes. Separating urban and rural areas is not sufficient to make coefficients homogeneous.

4.1 Disaggregation of regional data to assess the validity of weighting coefficients.

The best way to assess the disaggregation of the commune populations is comparing the results with data at infra-commune level, but, at the current stage, such data are generally not available to the EC. One possible way to overcome this limitation would be:

- Disaggregate regional data with CLC using a set of coefficients.
- Reaggregate the attributed population on commune basis.
- Compare with the known population per commune and compute a disagreement indicator.
- Modify the coefficients to reduce the disagreement.

X_r is the population in region *r*.
S_{cr} is the area of land cover type *c* in region *r*.
Y_{cr} is the density of population we attribute to land cover type *c* in region *r* .
W_r is an adjustment factor to ensure that the total population in each region coincides with the known total.

Thus,

$$X_r = \sum_c S_{cr} U_c W_r \quad \Rightarrow \quad \text{The densities attributed are} \quad Y_{cr} = U_c \frac{X_r}{\sum_c S_{cr} U_c} \qquad (7)$$

and the population attributed to each commune m in region r is

$$X_m^* = \sum_c S_{cm} Y_{cr} \qquad (8)$$

Towards agri-environmental indicators. EEA. In press.

Computing the ratio between the attributed population and the known population

$$\psi_m = \frac{X_m^*}{X_m} \qquad (9)$$

or an aggregated difference between attributed and real population at regional or European level

$$\delta_r = \sum_{m \in r} \left| X_m^* - X_m \right| \qquad \delta = \sum_m \left| X_m^* - X_m \right| \qquad (10)$$

It can be checked that $\delta_r \leq 2X_r$.
The maximum value of the deviation would happen when all the population is attributed to communes with real population 0.

For each region we can compute $\rho_{cr} = corr\left(\psi_m, \frac{S_{cm}}{S_m}\right)$.

If the correlation $\rho_{cr} > 0$, this would mean that a too high population has been generally attributed to communes where the CORINE Land Cover class c has a high proportion. We can try to compensate this tendency by reducing the coefficient for this region and land cover:

$$U'_{cr} = U_c \left(1 - \frac{\rho_{cr} \times \delta_r}{2 \times X_r}\right)$$

The coefficient U'_{cr} raises when the correlation is negative.

The term $\frac{\delta_r}{2 \times X_r}$ is introduced to moderate the modification of the coefficient when the attribution is close to the real population.

The coefficient adjustment can be repeated in an iterative way until the difference indicator δ becomes stable. To avoid some extreme effects on the coefficients, limits have been introduced so that the ratio between the maximum and minimum density in a commune is constrained not to exceed 10,000.

4.2 Application with 16 grouped classes.

The total deviation δ with the same coefficient to all CLC classes gives an indicator of 322×10^6 for a total population of 321×10^6 inhabitants. This corresponds to the inaccuracy of representing the average population density by NUTS2 compared with the representation by communes.

	population	deviation no weight	ratio %
AT	7796	6398	82
BE	9979	6640	67
DE	59940	49036	82
DK	5146	4574	89
ES	37182	45991	124
FR	55072	63794	116

Towards agri-environmental indicators. EEA. In press.

GR	9692	11846	122
IE	3364	3689	110
IT	56705	54779	97
LU	383	386	101
NL	14950	12349	83
PT	9371	10690	114
UK	51468	52583	102
total	321047	322755	101

Table 3: Deviation indicator with uniform disaggregation for NUTS2 regions.

With the coefficients in

grouped class	Initial coefficient U_c	CORINE Class	Label
1	32	111	Continuous urban fabric
2	25	112	Discontinuous urban fabric
3	1	121	Industrial or commercial units
4	1	122, 123, 124	Road and rail networks, ports, airports
5	1	141, 142	Green urban areas, Sport and leisure facilities
6	3	211	Non-irrigated arable land
7	3	212	Permanently irrigated land
8	1	213	Rice fields
9	5	22	Permanent crops
10	3	231	Pastures
11	5	241	Annual and permanent crops associated
12	5	242	Complex cultivation patterns
13	3	243	Agriculture, with natural vegetation
14	1	244	Agro-forestry areas
15	1	31, 324	Forest and woodland
16	1	321, 322, 323	Other natural vegetation
17	0	13, 33, 4, 5	Mine, dump and construction sites, sand, rock and burnt areas, glaciers, wetland and water

Table 2, the disagreement indicator goes down to around 241×10^6 . The application of the algorithm described above improves the deviation, so that results become stable at slightly under 137×10^6 .

Looking at the average values of the coefficients (Table 4) by region, we observe:

- The three arable land classes have similar average coefficients;
- The coefficients for permanent crops are close to the coefficients of two of the complex classes (2.4);
- Agro-forestry presents similar coefficients to those of forest and semi-natural vegetation.

Towards agri-environmental indicators. EEA. In press.

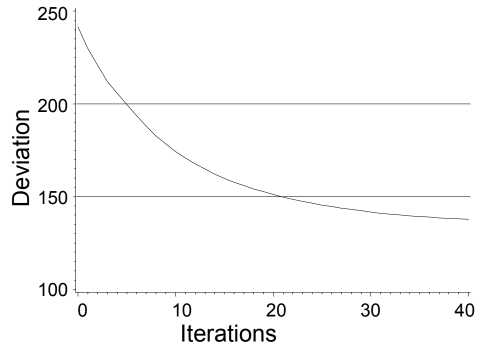

Figure 1: Total deviation δ of population attribution with 16 CORINE groups.

Land Use	
Urban dense	198.41
Urban discontinuous	176.56
Industrial and commercial	9.56
Transport	3.71
Green urban	3.42
Arable non irrigated	2.98
Arable irrigated	3.36
Rice	2.90
Permanent crops	4.95
Pastures	2.99
Arable with permanent crops	5.40
Complex agricultural	5.92
Agricultural and natural	3.24
Agroforestry	0.86
Forest	0.94
Natural vegetation	0.63

Table 4: Average coefficients after 40 iterations

4.3 Stratification and further grouping of CORINE Land Cover classes.

The ratio between the density in different land cover classes is not the same in densely populated areas and in more rural areas. Therefore communes have been stratified in each region applying a very simple criterion:

1. Dense communes: population density higher than twice the average density in its NUTS2 region;
2. Less dense: population density lower than twice the average density in its NUTS2 region, but urban area reported in CORINE Land Cover;
3. No urban: No urban area reported in CORINE Land Cover.

A cluster analysis of the 16 CLC classes based on the table of coefficients by NUTS2 after 40 iterations also gives some indication of the CORINE Land Cover classes that have a similar behaviour for different region typologies. Taking into account the results of cluster analysis

Towards agri-environmental indicators. EEA. In press.

and the meaning of classes, the classes have been regrouped into 8 (Table 5) and the iterative algorithm has been rerun. The classes "Industrial and commercial", "transport" and "green urban and sport facilities" have been aggregated with the class "urban discontinuous" with fixed weights inspired in Table 4:

Urban 2 = Urban discontinuous + 0.05 * Industrial_commercial + 0.02*transport + 0.02*green urban

Therefore 6 classes have been kept for further analysis. Applying the iteration algorithm with 3 strata and 6 CORINE Land Cover classes, a deviation indicator that approaches 90×10^6 is obtained.

grouped class	CORINE Class	Label
1	111	Continuous urban fabric
Urbdisc (2a)	112	Discontinuous urban fabric
ind (2b)	121	Industrial or commercial units
otha (2c)	122	Road and rail networks and associated land
otha (2c)	123	Port areas
otha (2c)	124	Airports
otha (2c)	141	Green urban areas
otha (2c)	142	Sport and leisure facilities
3	211	Non-irrigated arable land
3	212	Permanently irrigated land
3	213	Rice fields
4	221	Vineyards
4	222	Fruit trees and berry plantations
4	223	Olive growes
5	231	Pastures
4	241	Annual and permanent crops associated
4	242	Complex cultivation patterns
5	243	Agriculture, with natural vegetation
6	244	Agro-forestry areas
6	311	Broad-leaved forest
6	312	Coniferous forest
6	313	Mixed forest
6	321	Natural grassland
6	322	Moors and heathland
6	323	Sclerophyllous vegetation
6	324	Transitional woodland-shrub

Table 5 New grouped classes.

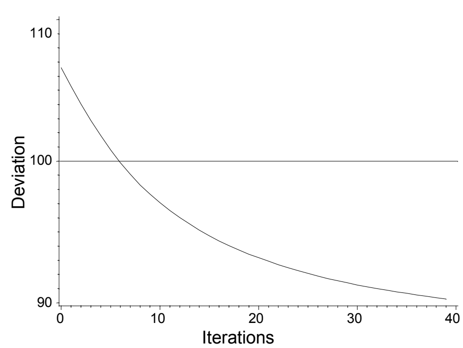

Figure 2: Total deviation δ of population attribution with 6 CORINE groups and 3 density strata per region. .

Towards agri-environmental indicators. EEA. In press.

5 Suggested disaggregation.

The algorithm gives weighting coefficients for each stratum in each NUTS2 region. For most regions the coefficients are similar, but some outliers appear. Clustering NUTS2 regions does not show a clear grouping linked with population settlement styles. Therefore a first disaggregation rule has been proposed based on a set of coefficients for each stratum that are the same for all NUTS2 regions.
The population density we attribute to land cover class c in commune m is computed as $Y_{cm} = U_c W_m$. The coefficients U_c keep their meaning if they are multiplied by any constant K and the coefficient W_m is divided by K. The values of U_c given in Table 6 correspond to a choice of K such that the median of W_m in each stratum is 1. They still cannot be interpreted as "population density for land cover class c", but as median density for each land cover class in each stratum.

Table 6: Disaggregation coefficients with 6 CLC classes and three strata of communes.

	Urban dense	Urban 2	Arable	Permanent crops and complex	Pastures	Forest & natural vegetation
Stratum 1	1445.9	619.1	10.2	15.4	5.1	3.3
2	947.4	622.4	17.4	30.9	11.3	5.2
3			32.0	69.3	22.8	8.6

Table 7: % of population in each CLC class with the suggested disaggregation.

	AT	BE	DE	DK	ES	FR	GR	IE	IT	LU	NL	PT	UK	All
Urban dense	11.8	5.7	2.5	10.1	54.1	13.2	33.2	8.4	23.8	14.8	85.3	18.3	16.1	21.9
Urban discontinuous	50.2	86.4	80.0	67.4	18.2	64.6	33.9	50.2	50.2	66.8	2.6	40.6	71.8	56.5
Indust. & commer.	0.2	0.6	0.6	0.3	0.6	0.8	0.3	0.4	0.7	0.7	0.4	0.4	0.3	0.6
Transport	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.0	0.1	0.1
Green urban	0.0	0.1	0.1	0.2	0.0	0.1	0.0	0.2	0.0	0.0	0.1	0.0	0.2	0.1
Arable non irrigated	8.6	2.4	8.1	16.5	6.2	7.1	4.4	4.1	7.9	2.5	3.1	6.1	4.8	6.7
Arable irrigated	0.0	0.0	0.0	0.0	2.2	0.0	2.7	0.0	0.5	0.0	0.0	0.1	0.0	0.4
Rice	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.2	0.0	0.0	0.3	0.0	0.1
Permanent crops	0.6	0.1	0.5	0.0	4.8	1.6	7.2	0.0	4.3	0.3	0.1	7.2	0.0	2.1
Pasture	6.6	0.7	2.2	0.2	0.3	2.5	0.0	28.9	0.3	1.4	3.7	0.0	5.1	2.4
Arable & perm crops	0.0	0.0	0.0	0.0	0.4	0.0	0.2	0.0	1.3	0.3	0.0	10.0	0.0	0.6
Complex agricultural	11.3	3.1	3.4	2.9	6.3	6.7	8.5	2.1	5.6	9.1	3.8	5.7	0.8	4.7
Agric. & natural veg.	0.5	0.4	0.4	1.4	1.3	0.8	2.3	3.0	1.6	1.4	0.3	3.7	0.1	0.9
Agroforestry	0.0	0.0	0.0	0.0	0.5	0.0	0.2	0.0	0.0	0.0	0.0	0.4	0.0	0.1
Forest	8.7	0.5	2.2	0.8	2.7	2.0	3.1	1.3	2.8	2.4	0.4	5.8	0.4	2.1
Natural vegetation	1.5	0.0	0.0	0.1	2.0	0.3	4.0	1.3	0.6	0.0	0.1	1.3	0.4	0.7

Table 7 gives the % of population that has been attributed to each CLC class in each country. The meaning of this table might need some clarification. For example "28.9" in the cell Ireland-pasture does not mean that 28.9% of the Irish population live in pasture fields, but that this amount of population has been attributed to areas coded as "pasture" by CLC. This may correspond to a large number of small urban nuclei inside the CLC "pasture" class.

Towards agri-environmental indicators. EEA. In press.

5.1 *Further coefficient tuning.*

These coefficients can be seen as a starting point for a manual tuning procedure, that allows to take into account additional knowledge on the population settlement in specific areas. An Arc-view tool is being built with the following functionalities:

- Viewing the attributed population density for a subset of land cover classes;
- Interactive selection of a subset of communes based on a set of criteria that includes a geographic window, average density or % of urban area;
- Attributing new coefficients to the selected set of communes.

6 Quality assessment of the disaggregation in Arezzo (Italy).

A first assessment has been made on the basis of the spatial behaviour of the deviation between the actual commune population and the population attributed by disaggregation of NUTS2 data. A link is observed with urban agglomerations: the population attributed to cities is generally smaller than their actual population. The opposite happens for rural areas, where the population attributed is generally larger than the actual one. This indicates that the stratification that has been made is not sufficient. The distance to major urban agglomerations seems to be another element to be considered.

A different quality assessment has been made by comparing the results of disaggregating commune data with data at sub-communal level (census sections) for a test site in Arezzo, including 27 communes subdivided into 1656 census sections (see chapter ??? in this volume for a description of the site). These data are available as a geo-referenced layer in a GIS environment with geographic limits and population. The total population of the test site is 235,630 inhabitants.

6.1 *Disagreement at pixel level and at census section level.*

For this exercise census sections and density maps have been rasterised with a resolution of 25 m. The quality of a density map ${}^{k}Y_i$ can be assessed by comparing the disaggregated density Y_{cm} (commune data + CLC) with the density by census section ${}^{1}Y_i = {}^{1}Y_s = X_s / A_s$ (assumed to be the truth), where X_s and A_s are the population and the area of the census section *s* and *i* is any pixel in section *s*. The comparison has been made in two different ways:

- per pixel: $\delta_{1k}^{pix} = \sum_i \left| {}^{1}Y_i - {}^{k}Y_i \right|$
- per census section: $\delta_{1k}^{\sec t} = \sum_s \left| X_s - {}^{k}X_s \right|$ where ${}^{k}X_s$ is the population attributed to section *s* by map ${}^{k}Y_i$

The following density maps have been compared with the "truth" ${}^{1}Y$

k=2 : Density map closest to ${}^{1}Y$ fulfilling the condition of having the same value for pixels in the same commune and the same CORINE Land Cover class with the classification into 17 classes (*k=3* for CLC in 7 classes). These maps cannot be computed using only commune data and CLC. They give an idea of the best possible result that can be achieved by maps with this condition (scale effect).

$$ {}^{2}Y_{cm} = \frac{{}^{2}X_{cm}}{S_{cm}} = \frac{\sum_j {}^{1}Y_j \, S_{cj}}{\sum_j S_{cj}} $$

k=4 disaggregation with the suggested coefficients (Table 6, Figure 3 c).

***k=5*: disaggregation with the initial coefficients (**

grouped class	Initial coefficient U_c	CORINE Class	Label
1	32	111	Continuous urban fabric
2	25	112	Discontinuous urban fabric
3	1	121	Industrial or commercial units
4	1	122, 123, 124	Road and rail networks, ports, airports
5	1	141, 142	Green urban areas, Sport and leisure facilities
6	3	211	Non-irrigated arable land
7	3	212	Permanently irrigated land
8	1	213	Rice fields
9	5	22	Permanent crops
10	3	231	Pastures
11	5	241	Annual and permanent crops associated
12	5	242	Complex cultivation patterns
13	3	243	Agriculture, with natural vegetation
14	1	244	Agro-forestry areas
15	1	31, 324	Forest and woodland
16	1	321, 322, 323	Other natural vegetation
17	0	13, 33, 4, 5	Mine, dump and construction sites, sand, rock and burnt areas, glaciers, wetland and water

Table 2);

k=6: no disaggregation, i.e. attributing uniform density in each commune (Figure 3 a);

	δ_{1k}^{sect}	δ_{1k}^{pix}
k=2	205766	235042
k=3	206737	237661
k=4	201946	240380
k=5	298676	307115
k=6	383998	425274

Table 8: Disagreement between the population density map per census section and different maps per CLC class x commune

Table 8 indicates that, **<u>for this particular site</u>**:

- There is a major disagreement between the density maps per commune and per census section.
- The disagreement is still strong with the disaggregation obtained with the coefficients of Table 8.
- Most of this disagreement is due to the different scales of CORINE Land Cover and the census sections:

Towards agri-environmental indicators. EEA. In press.

- Many census sections correspond to a small built area and are not represented in CLC
  - Different census sections in the same commune and same CLC class (specially urban) have very different population densities.
- The map we have obtained at EU-13 level (k=4) is not too far from the optimum that can be got combining commune data and CLC.

The population attributed to the section can be compared with the actual population: the ratio $\left(X_s / {}^4X_s \right)$ is represented in Figure 4. Figure 5 represents the ratio between both densities pixel per pixel (1ha). A visual inspection seems to indicate that population in urban classes is underestimated.

Systematic over/underestimation for CLC classes can be also assessed by comparing:
- the total population that has been attributed to pixels in each CLC class, and
- the population appearing in the same pixels when population density is mapped by census sections.

Towards agri-environmental indicators. EEA. In press.

Figure 3: Arezzo site: a) Population density per commune, b) CORINE Land Cover with simplified legend, c)Population density after disaggregation.

Towards agri-environmental indicators. EEA. In press.

Figure 4: Arezzo site. Ratio for each census section between actual population and attributed population.

Figure 5 density ratio per 1 ha pixel: average density per section / attributed from commune+CLC.

Towards agri-environmental indicators. EEA. In press.

The result of this comparison (see Table 9) indicates that the amount of population attributed to CLC urban classes is approximately in agreement with the data by census sections. The coefficients seem to be too high for pasture, forest and natural vegetation, and slightly underestimated for arable land, permanent crops and heterogeneous agriculture.

CORINE	Census sections	Attributed (communes +CLC)
Urban dense	9.2	8.7
Urban discontinuous	43.3	45.0
Transport infrast.	3.2	2.2
Gree urban, leisure	0.2	0.2
Arable land	10.1	7.5
Perm. crops and heterog.	24.1	22.3
Pastures	3.5	4.5
Forest and nat. veg.	6.4	9.6

Table 9: % of population appearing in each CLC class in two density maps.

7 Conclusions

CORINE Land Cover can be used to improve the mapping of population density available to the European Commission at commune level. An algorithm has been developed that combines two levels of administrative units (NUTS2 and communes in this paper) to estimate reasonable weighting coefficients for each land cover class. Results can be improved if each region is stratified grouping communes with similar characteristics. The estimation of coefficients becomes more robust if the nomenclature is simplified.

A first assessment, comparing with more detailed data for a test site in Arezzo (Italy), suggests that the estimated coefficients are approximately correct for this site, but additional checks are necessary for sites with different styles of population settlement.

REFERENCES

Ambroise C., Govaert G., 1998, Convergence of an EM-type algorithm for spatial clustering, *Pattern Recognition Letters* v 19 n 10 . p 919-927

Anderson N H, Titterington D M,1997, Some methods for investigating spatial clustering, with epidemiological applications. *Journal of the Royal Statistical Society Series A, Statistics in Society* Vol. 160; No. 1; pp. 87-105

Cerioli A, Riani M., 1999, The ordering of spatial data and the detection of multiple outliers, *Journal of Computational and Graphical Statistics*, Vol 8, No 2, pp. 239-258.

Coleman D., Dong X., Hardin J., Rocke D.M., Woodruff D.L., 1999, Some computational issues in cluster analysis with no a priori metric, *Computational Statistics and Data Analysis* Vol. 31 No 1 . pp. 1-11.

Diggle P J; Chetwynd A G, 1991 Second-order analysis of spatial clustering for inhomogeneous populations. Biometrics; Vol. 47; No. 3; pp. 1155-1163;

Everitt, B.S., 1980, Cluster Analysis, London: Heineman.

Fisher P.F., Langford M., 1995, Modelling the errors in areal interpolation between zonal systems by Monte Carlo simulation. *Environment & Planning A* Vol. 27 page 211-224.

Towards agri-environmental indicators. EEA. In press.

Flowerdew R., Green M., Kehris E., 1991, Using areal interpolation methods in GIS; *Papers in regional science*, Vol 70 No 3 pp.303-315.

Gale N., Halperin W. C., 1982, A case for better graphics the unclassed choropleth map. *The American statistician*, Vol 36, No 4, pp. 330-336.

Gebhardt F., 1999, Cluster tests for geographical areas with binary data, *Computational Statistics and Data Analysis* Vol. 31 No 1 . pp. 39-58

Hunting Technical Services, 1999, *GIS application development*. Report to Eurostat May 1999.

Le Jeannic Th., 1996, Une nouvelle approche territoriale de la ville. *Economie et Statistique* 294-295 pp. 25-45.

Macchiato-M-F; LA-Rotonda-L; Lapenna-V; Ragosta-M., Time modelling and spatial clustering of daily ambient temperature: an application in southern Italy. *EnvironMetrics*; Vol. 6; No. 1; pp. 31-53; 1995;

Tobler, W.R., 1979, Smooth pycnophylatic interpolation for geographical regions. *Journal of the American Statistical Association*, 74, pp. 519-530.

Towards agri-environmental indicators. EEA. In press.