Gridded Population density in the EU from Commune Data and Land Cover Information

F.J. Gallego
Agrifish Unit, IPSC, JRC, 21020 Ispra (Varese), Italy
e-mail: javier.gallego@jrc.it

Extended Summary

1 Introduction

Harmonised population density data for the European Union (EU) are available at the level of the commune. Some countries have more detailed geo-referenced data, but for EU-wide studies the communal level is the most detailed available. This level of spatial resolution may be insufficient in many cases for planning or modelling purposes. There is a need to downscale population density, i.e. to represent it in smaller geographical units. There is a range of possible approaches for downscaling (Flowerdew et al, 1991, Blook et al, 1996, Xie, 1995, Wu and Murray, 2005). We have chosen a empirical approach based on land cover information as proxy variable. The result of our work is a GIS layer in raster format with 100 m resolution. We attribute to each 1 ha pixel an estimated population density.

We first present an iterative method applied by Gallego and Peedell (2001) and then a modification to exploit additional data from a land use survey on a point sample.

2 Data

Several layers of information are combined for this exercise: Commune data (population and geographic boundaries), a land cover map, and a fine scale point survey: The study covers an area of 4.3 Million km2 with more than 480 Million inhabitants. Population data come from the 2001 census at commune level (more than 114.000 communes)

CORINE Land Cover 2000 (CLC) is a land cover map with a nomenclature of 44 classes. The minimum mapping unit of CLC is 25 ha; smaller units are included in the dominant land cover type around or grouped in an area coded as heterogeneous. We have used a raster version of CLC with a pixel size of 1 ha. A raster version with cells of 1 ha has been used. CLC is complemented by LUCAS-2001 (Land Use/Cover Area-frame Survey) with on-the-ground observations for nearly 100,000 points grouped in clusters of 10. 2245 LUCAS points were residential (2.4% of the total sample). Therefore LUCAS estimates the area with residential use in EU15 to be around 75,000 km2.

3 Modified areal weighting with given coefficients.

We suppose that the population density Y_{cm} for land cover type c in commune m, can be written:

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

Where the coefficient U_c depends on the land cover class of CLC and W_m is a factor that ensures that the total population in each commune matches the known commune population.

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

Where X_m is the population in commune m and S_{cm} is the area of land cover type c in commune m. For a given set of coefficients U_c we easily get W_m: and Y_{cm}:

$$X_m = \sum_c S_{cm} U_c W_m \quad \Rightarrow \quad W_m = \frac{X_m}{\sum_c S_{cm} U_c} \quad \Rightarrow \quad Y_{cm} = U_c \frac{X_m}{\sum_c S_{cm} U_c} \qquad (3)$$

This disaggregation was been carried out with an initial set of coefficients provided by the European Environment Agency (EEA) for an aggregated nomenclature of CORINE Land Cover (Table 2).

To assess the disaggregation of the commune populations, we would need to compare the results with data at infra-commune level, but such data are not available at EU level. One possible way to overcome this limitation is:

- Pretend that we only know the data at some regional level (larger than the communes).
- Disaggregate regional data with CLC using a given set of coefficients U_c .
- Estimate the commune population by aggregating the downscaled density.
- Compare with the known population per commune and compute a disagreement indicator.
- Modify the coefficients to reduce the disagreement and disaggregate again.

For a region *r*, $X_r = \sum_c S_{cr} U_c W_r$ and the densities attributed are $Y_{cr} = U_c \frac{X_r}{\sum_c S_{cr} U_c}$

The population attributed to each commune m in region r is $X_m^* = \sum_c S_{cm} Y_{cr}$

We can compute the ratio between the attributed and the known population $\psi_m = \frac{X_m^*}{X_m}$

or an aggregated difference for the region or at EU level: $\delta_r = \sum_{m \in r} \left| X_m^* - X_m \right|$ $\delta = \sum_m \left| X_m^* - X_m \right|$

For each region, a correlation $\rho_{cr} = corr\left(\psi_m, \frac{S_{cm}}{S_m} \right) > 0$. would mean that a too high population has been attributed to communes class c has a relatively high proportion. We can try to compensate this tendency by reducing the coefficient for this region and land cover. We have empirically chosen the next formula to reduce the disagreement:

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

Where k is a tuning coefficient:

This procedure was run separately on strata, defined inside each region:

1. Dense communes: population density higher than twice the average density in its region;
2. Less dense: population density lower than twice the average density in its region. Still some urban area is reported in CLC;
3. Sparse population: No urban area reported in CLC .

The disagreement generally became stable after 40 iterations. For most regions the coefficients are similar, but some outliers appear. To make the results more stable, the median values have been applied to all regions (table 1)

Table 1: Disaggregation coefficients with 6 CLC classes and three strata of communes.

	Urban dense	Urban discontinuous	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

A quality assessment was made by comparing the results of disaggregating commune data with data at sub-communal level (census sections) for a test site in the province of Arezzo, Italy. The comparison suggested that the coefficients are reasonable but the information provided by CLC is another limitation of the disaggregation accuracy, for example no information is available on height of buildings, possible dominance of non-residential buildings (offices), etc.

4 Reviewing coefficients with LUCAS data

Overlaying the approx 96,000 point of the LUCAS sample on the CLC map, we get a contingency table crossing CLC classes with fine scale land cover types (Gallego, 2003). In particular we can estimate the proportion of each CLC class that has residential use. CLC classes were clustered on the basis of the proportion of residential use to get a better simplified nomenclature in 9 classes (table 2). The coefficients in the right column where approximately derived from the % of residential land for the non-urban classes. For the urban classes, the % of residential area is not such a good proxy and we simply modified the coefficients derived in the previous version on the basis of subjective perception of the results in a number of known areas.

Table 2 : Proportion of residential area in CLC classes using LUCAS

		LUCAS points		suggested
CORINE Land Cover class	residential	total	% resid.	coefficients
Urban dense	60	132	45.4	2000
Urban discontinuous	1085	2609	41.6	500
Other urban	72	748	9.6	150
Artificial non residential	2	241	0.8	0
Agricultural	576	31956	1.8	30
Heterogeneous	272	9492	2.9	50
Forest and agroforestry	142	30826	0.5	8
Natural vegetation	19	13339	0.14	2
Open spaces and water	15	7087	0.21	0
Total	2243	96430	2.3	

We can expect that the same non-urban CLC class (e.g. "agricultural") has more dense population in areas with higher average density, i.e. the commune coefficients W_m are higher for communes with higher average density D_m but W_m does not grow linearly with the average population density. A simple regression between logarithms gives $W_m \cong c\ D_m^{0.53}$

An application of logit regression suggests a lower exponent, of the order of 0.4. This may mean that density in non urban areas of very populated communes is overestimated. This might lead to a revision of the coefficients in table 2.

5 Geographic patterns of scattered population

In many European areas we can see numerous scattered houses outside the urban nuclei. In other areas the population is concentrated in large or small nuclei with very few houses in between. The density of LUCAS residential points in non-urban CLC classes is being used to map these different behaviours. We use an indicator $\frac{r_i^*}{a_i^*}$ where r_i^* is the number of residential points in non-urban CLC classes for a geographical unit i and $a_i^* = \sum_c a_{ci} \frac{n_{c+}}{r_{c+}}$ is a weighted average of the areas of CLC classes, that indicate the expected number of residential points if the EU territory was homogeneous.

6 Some examples of applications

The range of applications of a population grid layer is very wide. Two examples are presented:

The population that lives within 10 km from the coast. The figure obtained is about 19% of the total population in EU25. In some countries the coastal population density is around 5 times the average of the country (Spain, Portugal, Finland, Sweden, Lithuania), while in other countries it is approximately the same (UK, Belgium, Netherlands, Germany).

The definition of rural and urban areas. Many organisations use the so-called OEDC definition, based on a threshold of 150 inhab/km^2 at the level of the commune. The heterogeneous sizes of communes leads to some abnormal classifications, in particular for very large communes. We study an alternative based on GIS operations on the gridded population density.

Acknowledgements: César de Diego and Edwin Schaaf (Eurostat), Roger Milego, Oscar Gómez, Stefan Kleeshule (European Topic Centre for Terrestrial Environment), Jean Dusart, Bettina Baruth, Giorgio Libertà (JRC), Tomas Soukup (GISAT, Prague).

REFERENCES

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

Gallego J., Peedell S., 2001, Using CORINE Land Cover to map population density. *Towards Agri-environmental indicators*, Topic report 6/2001 European Environment Agency, Copenhagen, pp. 92-103. http://reports.eea.eu.int/topic_report_2001_06/en

L. M. Bloom, P. J. Pedler and G. E. Wragg, 1996, Implementation of enhanced areal interpolation using MapInfo *Computers & Geosciences, Volume 22, Issue 5, Pages 459-466*

Wu Ch., Murray A. T., 2005, A cokriging method for estimating population density in urban areas, *Computers, Environment and Urban Systems*, Volume 29, Issue 5, September 2005, Pages 558-579.

Xie, Y. 1995, The overlaid network algorithms for areal interpolation problem. *Computers, Environment and Urban Systems, Volume 19, Issue 4, July-August 1995, Pages 287-306*