Population density grid of EU-27+, version 5

Summary of the downscaling method

For further details, send a mail to Javier.gallego@jrc.ec.europa.eu

Grid main characteristics

Geographic coverage: EU27 + Croatia. Some islands and overseas territories missing.
Resolution: 100m (1 ha pixels)
Values correspond to density in inhabitants/km2. to obtain the estimated population in a polygon, divide the sum of pixel values by 100.
Projection: Lambert-Azimuthal equal area (INSPIRE-recommended)
Downloadable from http://dataservice.eea.europa.eu/dataservice/ or by request from *Javier.gallego@jrc.ec.europa.eu*

Data used:

Population by commune. Census 2001. Provided by Eurostat.
Commune boundaries. geographic boundaries of the communes of the SABE GIS layer (http://www.megrin.org/SABE/Sabe.html , © Eurogeographics)
CORINE Land Cover (CLC-2000), version 9, reclassified into 9 classes (table 1).
LUCAS-2001 point data provided by Eurostat. (used to tune the coefficients of CLC classes).
Population density grids 2006 (1 km resolution) for Austria, Denmark, Finland, Netherlands, Northern Ireland (2001) and Sweden. Provided by National Statistical Institutes members of the European Forum for Geostatistics ((http://www.efgs.ssb.no/)

grouped class	CORINE Class	Label
1	111	Continuous urban fabric
2	112	Discontinuous urban fabric
3	121, 133, 14	Other urban
4	122-124, 131-132	Low population artificial
5	21, 22, 23	Agriculture
6	241-243	Heterogeneous
7	244, 31	Forest
8	32	Natural vegetation
9	33, 4, 5	Bare land, wetland and water

Table 1: CLC Nomenclature aggregation

Method.

The methods applied for the previous versions of the population disaggregation (or downscaling) are summarised in (Gallego, 2009)

We call:
X_m population in commune m .
S_{cm} area of land cover type c in commune m .
Y_{cm} density of population we attribute to land cover type c in commune m .

The procedure starts with the attribution of a uniform density to the whole territory of the commune: $Y_{cm}^{0} = D_m = X_m / S_m$.

- Land cover classes are ranked and the subindex c is renumbered from highest to lowest presumed population.
- Density thresholds θ_{cm} are progressively applied starting with the land cover class c=9 with the lowest threshold in commune m.
- If the density attributed to class c is above the threshold: $Y_{cm}^{9-c} < \theta_c$, the population in excess $S_{cm} \times \left(Y_{cm}^{9-c} - \theta_c \right)$ is redistributed among the more dense classes: $Y_{cm}^{9-c+1} = \theta_c$ and $Y_{c'm}^{9-c+1} = Y_{c'm}^{9-c} + \frac{S_{cm} \times \left(Y_{cm}^{9-c} - \theta_c \right)}{\sum_{c'<c} S_{c'm}}$ for c'<c.
- If there is still some excess population at the end of the process, It is redistributed to all classes proportionally to the thresholds. This happens when $X_m > \sum_c \theta_{cm} S_{cm}$. In this case the final attributed density is proportional to the thresholds: $Y_{cm} = \frac{\theta_{cm} X_m}{\sum_{c'} \theta_{c'm} S_{c'm}}$

A key issue in the application of this method is the computation of the thresholds θ_{cm} . In the version of the method suggested by Eicher and Brewer (2001), the thresholds are computed as the 70 percentile of the population density of administrative units with a “pure” land cover type. The application of this rule was initially problematic in our case because of an insufficient number of “pure” communes for several CLC land cover grouped classes.

The availability of reference data for 1 km^2 grids for Austria, Denmark, Finland, Netherlands, Northern Ireland and Sweden allowed applying the limiting variable method in two versions:

- limiting variable method “simple”: The threshold θ_{cm} is constant for all communes $\left(\theta_{cm} = \theta_c \right)$. The threshold for each land cover class is computed as the mean of the of the population density $Y_{cj,ref}^{ref}$ in the reference grids for pure 1 km cells (only land cover class c is reported in the cell). The 70th percentile suggested by Eicher and Brewer (2001) was also tested, but gave slightly worse results. The map obtained with this method is not distributed because it is slightly worse than the version 5.

- limiting variable method “tuned”: This is the method used for version 5.
  The threshold θ_{cm} changes with the average density of the commune: $\theta_c = B_c \; D_m^{A_c}$. To avoid anomalies in high density communes, an absolute threshold is applied if the estimated threshold is higher.
  Thresholds are estimated by a simple log-log regression $\log\left(Y_{c,ref}^{*} \right) = \log\left(B_c \right) + A_c \; \log\left(D_{cm,ref}^{*} \right)$

  $Y_{c,ref}^{*}$ and $D_{cm,ref}^{*}$ are computed averaging reference density and commune density in subsets of cells defined as 10-percentile intervals in the set of pure cells of class c: from the minimum to percentile 10, etc. to get 10 pairs of data for the regression.
  If the regression slope is not significantly different from 0, the threshold θ_c computed for the “limiting variable simple” method

  The potential efficiency of stratification of communes was explored by an analysis of variance. Only for the CLC class “agriculture” there was a significantly different link between $Y_{c,ref}^{*}$ and $D_{cm,ref}^{*}$ for strata defined as sets of communes for which CLC reports or does not report urban areas (strata 0 and 1). Therefore the thresholds for the class agriculture were estimated separately for the two strata.

CLC	$\log(B_c)$	A_c	Absolute threshold
Urban discontinuous	5.95	0.51	n.a.
Urban discontinuous	5.73	0.37	9000
Other artificial	n.s.		70
Scarcely populated artificial	n.s		1
Agricultural (stratum 0)	1.87	0.53	70
Agricultural (stratum 1)	1.34	0.44	30
heterogeneous	2.55	0.31	80
forest	-1.21	0.67	10
Natural vegetation, bare land, water			0

Table 2: parameters to compute the thresholds in the limiting variable method applied for version 5.

Accuracy Assessment

The disagreement indicator was computed as:

$$\Delta_m = \sum_j \left| Y_{j,m} - Y_{j,ref} \right| \qquad (2)$$

Where j refers to the pure cells of the reference grids.

Notice that the maximum theoretical value is twice the population. The values obtained for the disagreement are reported in table 3. This table indicates that disaggregation of commune-level population density significantly reduces the disagreement with reference data, but is far from eliminating it. The tuned LimVar (Limiting variable) method used for version 5 behaves better than previous versions.

Dasymetric map	***Austria***	***Denmark***	***Finland***	***Sweden***	***Netherlands***	***Northern Ireland***
Country population	*8.03*	*5.35*	*5.18*	*8.88*	*15.99*	*1.69*
Communes (choropleth)	8.96	6.08	6.79	12.48	18.72	1.12
CLC-iterative	4.55	4.07	5.44	8.05	7.13	0.72
CLC-LUCAS simple	4.39	3.97	5.06	8.09	9.03	0.71
CLC-LUCAS logit	4.35	3.95	5.03	8.07	7.08	0.64
CLC EM	4.50	3.98	5.12	8.08	9.29	0.68
CLC LimVar simple	4.36	2.36	3.38	6.71	6.14	0.65
CLC LimVar tuned	4.37	2.19	3.20	6.34	6.18	0.60

Table 3: Disagreement of different dasymetric maps with reference data in 5 countries

REFERENCES

Eicher C., and Brewer, C. (2001) Dasymetric mapping and areal interpolation: implementation and evaluation, Cartography and Geographic Information Science 28 125–138

Gallego F.J., 2009, A population density grid of the European Union, Submitted to Population and Environment.