

A population density grid of the European Union

Francisco Javier Gallego

Published online: 1 May 2010
© Springer Science+Business Media, LLC 2010

Abstract This paper describes four methods used to produce dasymetric population density grids combining population data per commune with CORINE Land Cover, a map available for all countries of the European Union. An accuracy assessment has been carried out for five countries for which a very reliable 1-km population density grid exists; the improvement, compared with the choropleth map per commune, ranges between 20% for the weakest result in Finland and 62% for the best result in the Netherlands. The best results are obtained with a method using logit regression to integrate information from the point survey LUCAS (Land Use/Cover Area frame Survey); however, performance differences between methods are moderate. The dasymetric grid is distributed free of charge by the European Environment Agency, for non-commercial use.

Keywords Population density · Dasymetric mapping · Downscaling · CORINE Land Cover (CLC) · Land Use/Cover Area frame Survey (LUCAS)

Introduction

Many European countries produce detailed geo-referenced population data, but few of them make such data available to the research community. Member states of the European Union (EU) provide Eurostat with population data per commune, which is too coarse for many studies linking population with the environment; the area of communes is very heterogeneous, but is mostly between 10 and 100 km². This paper describes four areal interpolation methods applied to produce a dasymetric (downscaled) population density layer with 1-ha resolution. The techniques we use can be called “dasymetric mapping” or “downscaling” a term used more often in

F. J. Gallego (✉)
IPSC, JRC, 21020 Ispra, Varese, Italy
e-mail: javier.gallego@jrc.ec.europa.eu

environmental studies (Bierkens et al. 2000). The European Environment Agency (EEA) distributes the dasymetric population grid, free of charge for non-commercial purpose, through its data warehouse (<http://dataservice.eea.europa.eu/>). In 2008 it was downloaded 1636 times; this indicates the interest that it has raised. As Langford (2007) points out, complex methods to produce dasymetric population maps are a major obstacle for many users. The target of this paper is to help overcoming such inconvenience by documenting the method applied to obtain the grid distributed by EEA.

Examples of the application of dasymetric population maps include the damage assessment of natural disasters, such as floods (Tralli et al. 2005). Fine-scale population data are not essential to assess damage caused by natural disasters with a geographically smooth distribution, but are important for more localized events. Chen et al. (2004) conclude that downscaled population data make little difference to assess the damage caused by an earthquake but considerably improve the assessment of damage from hailstorms.

The layer presented in this paper has some conceptual differences with the worldwide products LandScanTM population density grid (Dobson et al. 2000; Bhaduri et al. 2002) and the Gridded Population of the World (GPW) of the Center for International Earth Science Information Network (CIESIN 2005): in our case the area covered is smaller but the spatial resolution is finer. LandScan refers to the “ambient population”, a time-weighted average of the number of people in a given area, while our grid locates each person in his/her dwelling at the time of a population census or register.

The best way to produce a gridded map of population density is collecting individual data with coordinates of the dwellings and counting the number of people in each cell of the grid (bottom-up approach). A few European countries, grouped in the European Forum for Geostatistics (EFGS: <http://www.efgs.ssb.no/>), have produced bottom-up grids, in most cases with a resolution of 1 km; however, the distribution of the grids is limited by confidentiality rules (http://www.statistik.at/web_en/statistics/regional/regional_breakdown/statistical_grids/).

An alternative solution when the bottom-up approach cannot be used or the data are not accessible is downscaling available data. There is a range of possible approaches for downscaling; Eicher and Brewer report three alternative solutions all based on dasymetric mapping principles:

- The binary method (Langford and Unwin 1994) assigns the whole population to one land cover class (usually urban or artificial land cover).
- The three-class method attributes some density to agricultural and forest classes.
- The limiting variable method attributes first the same density to all land cover classes in each administrative unit; densities are then modified by applying thresholds to each land cover class and redistributing the excess to other classes.

In the case study conducted by Eicher and Brewer, the limiting variable method gave the best results.

Mennis (2003) applied a method to downscale population data from census tracts to a 100-m grid to five counties in south-east Pennsylvania. In this example, three land cover classes are considered: high-urban, low-urban and non-urban. The

method attributes to each land cover class a population that depends on the area of the land cover class and on a “population density fraction”, a factor that is estimated from a sample of census tracts that only contain one of the three land cover classes. Other methods (Langford et al. 1991; Yuan et al. 1997; Briggs et al. 2007) use a regression model to obtain the population density of each class; coefficients are applied later to adjust the total population assigned to each administrative unit and make it coincide with the known population (pyncophylatic constraint). Wu and Murray (2005) use a cokriging method in a small test area in Ohio. This method gives a variance of the estimated density in each location, but presents computational problems for large data sets.

Some authors have produced more precise downscaled population density layers using street and road networks in a small area, such as a county (Xie 1995). A similar approach is adopted by Reibel and Bufalino (2005) and by Mrozinski and Cromley (1999), who assume that the population is concentrated in a buffer around a road network. Chen et al. (2004) also use street buffers from StreetWorksTM to downscale the population in the area of Sydney.

Another approach to the downscaling problem is based on the EM algorithm (Expectation–Maximization, Dempster et al. 1977). We present below the results of the EM algorithm applied to the EU data. An alternative method, that we call here CLC-iterative, is applied by Gallego and Peedell (2001). The resulting dasymmetric map has been assessed by Thielen et al. (2006), who conclude that the map gave realistic population figures for the areas flooded in Germany in the flood events of 1999 and 2002; however, they find that the method overestimates the population in non-urban land cover classes. The International Committee on Aviation Environmental Protection has used this population grid to assess the impact of noise around airports (Vinkx and Visée 2008). The European Commission (EC) has also used it to compute indicators of accessibility to services in rural areas (Dijkstra and Poelman 2008).

Bracken and Martin (1989, 1995) produced population density grids for the UK, based on the censuses of 1981 and 1991. The problem they tackled was different: their starting point was a detailed database of population for Enumeration Districts (ED). A population-weighted centroid of each ED was known, but no land cover data were available. The approach consisted of distributing the ED population from the centroid to neighbouring pixels with an adaptive kernel function. Martin (1996) highlights the importance of the pyncophylatic constraint for these surface models. Martin (1998) also analysed the design of EDs optimizing its compatibility with different applications, to minimize the need of downscaling.

Table 1 gives the main characteristics of some existing population grid maps. In general maps are more accurate if the units of the initial census data are smaller.

Data

The area covered by the study is EU27 (the 27 Member States of the EU): approximately 4.3 million km² with 480 million inhabitants. Some overseas territories are excluded. Population data from the 2000/2001 round of censuses were

Table 1 Comparison of some population density grids

Grid	Landscan	EFGS national grids	UK (Bracken and Martin)	EEA-JRC*
Coverage	World	1.4 mill km ² , 62 mill inhab.	245,000 km ² , 60 mill inhab.	4.3 mill km ² 480 mill inhab.
Location concept	Ambient population	Census address	Census address	Census address
Latest Ref date	2007	2006	1991	2000/2001
Cell size	30'' (~800 m)	1 km ²	200 × 200 m	100 × 100 m
Approx average population per unit in the raw data	>100,000 inhab.**	45 inhab.	300–400 inhab.	~4000 inhab.

* The grid presented in this paper

** Landscan metadata are not very explicit on this point

provided by Eurostat for nearly 115,000 communes. “Commune” refers to the EU Local Administrative Units, level 2 (LAU2). Country equivalences can be found at http://ec.europa.eu/eurostat/ramon/nuts/introannex_regions_en.html.

The average commune area per country ranges from less than 15 km² in Slovakia, Czech Republic and France to more than 1500 km² in Sweden. Table 2 gives a coarse description of the size of communes in the EU. For example, 6.2% of the communes have an area of more than 100 km², accounting for 32.6% of the population and 49.4% of the total area.

The land cover map used is the 1-ha resolution raster version of CORINE Land Cover 2000 (CLC). CLC has been produced, with common rules in all EU countries (EEA-ETC/TE, 2002), by photo-interpretation of Landsat ETM+ (Enhanced Thematic Mapper) satellite images. The nomenclature of CLC has 44 classes, which were simplified for this work, and the minimum mapping unit is 25 ha. Smaller patches are included in polygons labelled with the dominant land cover type. If there is no clearly dominant land cover type in a polygon, it is coded as “heterogeneous”; this occurs in 11% of the EU area. CLC can be also downloaded from the EEA warehouse (<http://dataservice.eea.europa.eu/>), where researchers can find a large number of data sources for Europe. For a high number of communes (29%), CLC does not report any urban area because they do not contain any urban patch larger than 25 ha. These communes require a separate treatment for population mapping.

Table 2 Heterogeneity of commune sizes in the EU

Commune area (km ²)	% Number communes	% Area	% Population
0–10	36.6	5.4	14.6
10–100	57.2	45.3	52.8
100–500	5.7	28.3	27.3
>500	0.5	21.1	5.3

LUCAS (Land Use/Cover Area frame Survey) covered in 2001 the 15 countries that were Member States at that time (EU15). It was based on a sample of points with a two-stage systematic design (Delincé 2001; Bettio et al. 2002). Primary Sampling units (PSU) were selected with a systematic grid of 18 km without stratification. Each PSU was a cluster of 10 points following a 5×2 rectangular pattern with a 300 m step. LUCAS has a double nomenclature: land cover and land use. For this work, we have only considered the land use class “residential”. The residential points in LUCAS were 2.4% of the sample; thus, the area with residential use in EU15 was estimated at approximately 75,000 km².

Downscaling methods compared

An iterative method to estimate land cover coefficients (CLC-iterative)

This method was used for the first version of the population density grid. We give brief indications on the approach; more details can be found in (Gallego and Peedell 2001). Communes have been stratified in each of the 272 NUTS2 regions in EU27. “NUTS” stands for “Nomenclature des Unités Territoriales Statistiques”. Each NUTS2 region usually contains 100–1,000 communes, with an area between 2,000 and 50,000 km² and a population between 500,000 and 5,000,000. The strata were defined as follows:

1. Dense communes: population density higher than twice the average density in the region;
2. Less dense: population density lower than twice the average density in the region, but urban area is large enough to be reported in CLC;
3. Sparse population: No urban area in the commune is reported by CLC.

The model assumes that the population density in each grid cell can be expressed as

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

Where Y_{cm} is the density of population for land cover type c in commune m that belongs to stratum h . The coefficient U_{ch} depends on the land cover class. W_m is a factor that ensures that the total population attributed to pixels in each commune matches the known commune population (Pycnophylactic constraint, Tobler 1979).

$$W_m = \frac{X_m}{\sum_c S_{cm}U_{ch}} \quad (2)$$

Where X_m is the population in commune m and S_{cm} is the area of land cover c in commune m . The same pycnophylactic constraint formula is applied to the other methods mentioned below.

This model implies two simplifying assumptions:

- (a) the population density is supposed to be the same for all pixels in the same commune and same CLC class.

- (b) the ratio between the population density of two land cover classes is constant for all communes in the same stratum.

Coefficients U_{ch} were estimated with an iterative algorithm with the following scheme:

1. Pretend first that we only know the population data at regional level (NUTS2).
2. Disaggregate regional data with CLC using a given set of coefficients U_{ch} . In the first iteration the coefficients are subjectively chosen.

Consider now the communes and compute the population X'_m attributed to the territory of commune m in step 2.

3. Compare X'_m with the known population per commune X_m and compute the disagreement.
4. Modify the coefficients U_{ch} to reduce the cumulated disagreement and restart step 2 until the results become stable.

Straight parameter estimation from LUCAS (CLC-LUCAS simple)

A CLC class, for example “arable land”, is generally not pure, mainly because of scale limitations: an arable land polygon may contain patches of non-arable land smaller than 25 ha. If we zoom into a fine scale, a small percentage of the class “arable land” is residential. Overlaying the LUCAS sample (approximately 96,000 points) on the CLC map, we can estimate the proportion of each CLC class with residential use. This proportion was used as a proxy of the population density (except for the urban-dense class), assuming that the residential area per person is approximately homogeneous in each commune. A set of coefficients was derived for a version of the downscaled density map, called here “CLC-LUCAS simple”. For the “urban dense” class the coefficients were based on those of the CLC-iterative method.

Logit regression on LUCAS data (CLC-LUCAS logit)

An analysis of the commune coefficients W_m computed with the previous methods shows that they slowly grow with the average density of the commune D_m , i.e. population density in a given CLC class tends to be slightly higher in communes with higher average density. We have used a logit regression to model the proportion of residential land p_{cm} (that we assume proportional to the population density) as a function of D_m . The logit function $\text{logit}(p) = \log(p/(1-p))$ is frequently used to model probabilities or proportions avoiding predicted values outside the interval $[0,1]$.

$$\text{logit}(p_{cm}) = \log(p_{cm}/(1-p_{cm})) = \alpha + \sum_c \beta_c J_c + \gamma \log(D_m) + \varepsilon_{cm} \quad (3)$$

Where J_c is an indicator function of CLC class c . (1 if the point is in class c and 0 otherwise).

The stratification used for the CLC-iterative method was modified later to obtain a better fit.

- Stratum a: communes with some “urban dense” CLC class.
- Stratum b: communes with some CLC artificial area, but no “urban dense”.
- Stratum c: communes without any CLC artificial area.

Some CLC classes were excluded from the logit regression because they had too few residential points; the urban-dense class was excluded because the residential area was not a good proxy of the population density. Thus, for the “urban dense” class we needed a complementary rule. We first attributed a population density for the other land cover classes and the remaining population of the commune was attributed to the class “urban dense”. To avoid abnormally high-density values, thresholds were imposed to each CLC class. If there were a sufficient number of CLC-pure communes (only one class in the commune), the threshold was the 90th percentile; otherwise a subjective choice was made.

The grid produced with this method is the version 4 distributed by the EEA. Figure 1 illustrates the difference between the choropleth map (homogeneous density in each commune), the downscaled grid and the reference data in the area of Innsbruck (Austria). Readers can notice that the grid cell size is different in the dasymetric map (1 ha) and the reference data (1 km²). Finer resolution reference data are often considered confidential by National Statistical Offices, but its publication becomes acceptable for less accurate model-derived dasymetric maps.

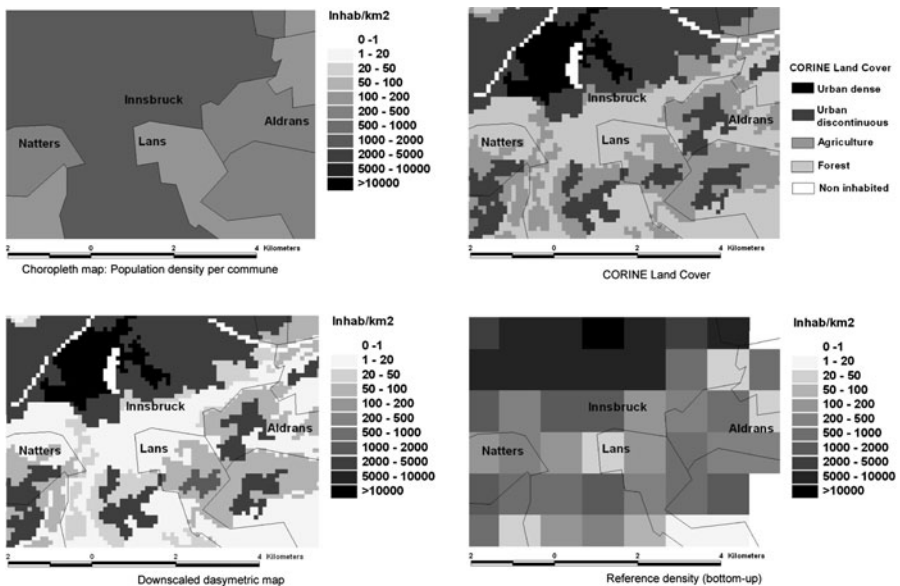


Fig. 1 Illustration of choropleth, dasymetric and reference density maps in the area of Innsbruck (Austria)

EM algorithm

Flowerdew et al. (1991) have proposed to apply the EM algorithm to estimate disaggregation coefficients. The EM method (Dempster et al. 1977) assumes an underlying probabilistic model. We have followed the suggestion of Flowerdew et al. assuming that the population X_{mc} in land cover class c for the commune m follows a Poisson distribution. The algorithmic details are out of the scope of this paper.

Results and validation

We omit in this paper the values of the coefficients obtained with each algorithm and give instead the average population density attributed to each aggregated CLC class (Table 3), which are easier to interpret.

The performance of the disaggregation procedures has been compared with the help of reference data provided by five National Statistical Offices. These national reference data were obtained by aggregation of individual dwellings on the basis of the 2005 population census (bottom-up) and presented as 1-km resolution grids in national cartographic coordinates. Unfortunately, these five countries do not adequately represent the variety of landscapes and settlement structures in the EU and were only chosen due to data availability at the time of writing this paper. Northern Ireland, possibly the whole UK, and Slovenia might be added in the near future, but doubts on the representativeness will remain. It is unlikely that Mediterranean countries or countries in the former Eastern Block (except Estonia) will provide reference data in the near future.

The five gridded maps (choropleth per commune and four downscaled maps) were first produced as raster grids in the EU standard Lambert-Azimuthal Equal Area projection (Annoni et al. 2001) with 1-ha resolution, then re-projected into the national coordinates and generalized to 1 km² with the same geometry of the reference grid. The disagreement indicator was computed as:

Table 3 Average population density attributed to CLC classes with different methods

CLC class	Area (km ²)	Choropleth (non disaggregated)	CLC-Iterative	CLC-LUCAS simple	CLC-LUCAS logit	CLC-EM
Urban dense	5840	2504	8241	10739	15388	8067
Urban discontinuous	130571	738	2550	1973	2225	2272
Infrastructure.	31694	976	314	780	352	2250
Agriculture	1571318	106	36	51	34	28
Heterogeneous	449523	118	60	91	66	48
Forest	1381860	58	8	11	5.9	1.7
Natural vegetation	517328	40	7	2.3	1.1	0.4
Open spaces and water	261933	45	6	0	0	0.4

Table 4 Disagreement of different dasymetric maps with reference data in 5 countries (in millions)

Dasymetric map	Austria	Denmark	Finland	Sweden	Netherlands
Country population	8.03	5.35	5.18	8.88	15.99
Communes (choropleth)	8.96	6.08	6.79	12.48	18.72
CLC-iterative	4.55	4.07	5.44	8.05	7.13
CLC-LUCAS simple	4.39	3.97	5.06	8.09	9.03
CLC-LUCAS logit	4.35	3.95	5.03	8.07	7.08
CLC-EM	4.50	3.98	5.12	8.08	9.29

$$\Delta_m = \sum_j |Y_{j,m} - Y_{j,ref}| \quad (4)$$

Notice that the maximum theoretical value is twice the population. The values obtained for the disagreement are reported in Table 4. This table indicates that downscaling commune-level population density significantly reduces the disagreement with reference data, but is far from eliminating it because the heterogeneity of population density for areas in the same commune and same land cover type in CLC is not captured by any of the methods. The improvement is similar for all disaggregation procedures, except in the Netherlands. The best results are obtained with the introduction of LUCAS data with a logit model to tune the behaviour of CLC classes. This method, however, presumes that suitable point survey data are available. If such LUCAS-like data are not available, our empirical results suggest that the CLC-iterative method is better, in particular for densely populated countries.

A brief analysis of the disagreement

The reason why the Netherlands have a different behaviour in Table 4 can be understood, to some extent, by comparing the densities estimated for different CLC classes. For this purpose we have looked into the 1 km² cells of the reference data that are “quasi-pure”, in the sense that a CLC class covers more than 80% of the area.

Table 5 reports some average densities for the Netherlands (densely populated) and Finland (scarcely populated). All methods underestimate the population attributed to “urban discontinuous” but the underestimation is smaller for the CLC-iterative and the CLC-LUCAS-logit methods. Better results obtained from these two methods in urban classes are decisive in the Netherlands, because the urban class has a strong weight. The opposite happens for agricultural and forest classes, all methods tend to overestimate the density in these classes. In many cases the methods that behave better for urban classes, behave worse for agriculture and forest. The consequence is that total disagreement per country with different methods is similar in thinly populated countries.

All methods overestimate the population in agricultural, heterogeneous and forest areas. The overestimation is often smaller for the CLC-EM method, but this approach heavily overestimates the population in the class “infrastructure”. The

Table 5 Average population density in different types of 1 km² cells of the Netherlands and Finland according to the dominant land cover type

CLC type	Reference data	Choropleth (non-disaggregated)	CLC-iterative	CLC-LUCAS simple	CLC-LUCAS logit	CLC-EM
Netherlands						
Urban	5331	1640	4614	4125	4725	3775
Infrastructure	340	1194	268	1081	600	2495
Agricultural	75	282	90	139	101	87
Heterogeneous	104	332	154	218	166	129
Forest and natural vegetation	27	365	30	55	38	28
Finland						
Urban	3358	1252	2970	2708	3108	2484
Infrastructure	135	980	107	486	184	1823
Agricultural	9	17	15	13	14	10
Heterogeneous	18	11	13	10	14	13
Forest and natural vegetation	0.2	3	0.7	0.4	0.4	0.1

reason being that the EM algorithm attributes a high density to this class because it appears mainly in very populated communes.

Discussion and further research

Population data of the 2000/2001 censuses have been merged with CLC to produce four versions of a fine-scale dasymetric map. The point survey LUCAS has been used to tune the downscaling coefficients. The validation by comparison with a reference population density grid, available for five countries, has shown that the inaccuracy of the homogeneous density representation per commune is substantially reduced thanks to the introduction of CLC. The validation results suggest, however, that:

- The downscaling is far from perfect.
- The density attributed to non-urban classes is generally overestimated.
- The inaccuracy of the different methods tested changes only moderately.

This confirms the observation made by Martin et al. (2000): the quality of the land cover map, and more generally, of the proxy variables available, is more important than the choice of the downscaling algorithm.

The results reported in this paper relate to CLC. For example, when we talk about population in agricultural land, this refers to the area reported as agriculture by CLC. Part of the inaccuracy that cannot be removed with the reported methods is due to limitations of CLC, in particular to the relatively coarse scale (minimum mapping unit of 25 ha). The dasymetric map is formally homogeneous (1-ha

resolution grid), but its quality is not: it is poorer in the areas where communes are larger, and the size of the communes is extremely heterogeneous. This limitation is inherent to the initial data and may be overcome if the efforts of the EFGS lead to more homogeneous quality in the data provided by National Statistical Offices.

The communes are polygons, and CLC is also originally defined as a set of polygons. In a small area it would have been more logical to build a vector layer of polygons defined by intersection of communes with CLC. In a large area the computational advantages of a raster layer have determined the choice of a grid output. However, the impact of the conversion to raster might be important for small communes and needs to be investigated; in particular for approximately 1300 communes smaller than 1 km² (100 grid cells). Their total area is tiny, but they account for more than 1% of the total population.

Ongoing and future research can be mentioned in two different directions: improving the dasymetric map and developing additional applications.

Improvements might be achieved with an adaptation of the limiting variable method (Eicher and Brewer 2001), in particular to limit the density overestimation in non-urban classes. The geographical analysis of the residuals of the logit model can give additional hints to locally modify the parameters to compute the dasymetric maps. Positive residuals suggest that the proportion of land with residential use for a given CLC class is higher than in other areas with similar average population density. This may be due to a higher amount of scattered houses in the landscape or to different criteria applied by the photo-interpreters to delineate CLC polygons.

The comparison with bottom-up reference data indicates that the population density can be very different for different grid cells in the same commune and the same CLC class. Such heterogeneity is difficult to catch with the land cover data we have used. The way to improve the results should be introducing new information layers: Satellite images of night-time light emissions (Briggs et al. 2007) might be part of the answer, although the coarse resolution of the available night-time light products is probably a serious limitation. Another limitation is the unclear link between night-time light and night-time population.

The EEA has recently launched a study to produce a EU-wide soil sealing information map from satellite images. This layer should have a better spatial resolution than CLC (pixels of 30 m or smaller) and should become, when available, a useful input for population mapping. The EU urban atlas (http://www.gmes-gseland.info/com/promo/200709_EuropUrbAtlas.pdf) is producing a detailed land cover map for 185 urban areas. This should also improve the population density grid.

The grid would be more useful for social applications if demographic information (age, sex) could be added, but these data are not available to Eurostat at commune level. The modification of the method to produce an “ambient” or a “day time” population grid may be another major evolution.

The number of downloads of the data set from the EEA data service suggests that this data layer is being widely used, although the projects for which the layer has been downloaded are not recorded. We mention a few examples:

Assessing the human pressure to the environment in the Natura-2000 network (Weber and Christophersen 2002), that provides a GIS-based description of protected areas in the EU (<http://ec.europa.eu/environment/nature/natura2000/>), although its completeness has been contested (Rosati et al. 2008) and updates should follow. Most Natura-2000 sites are smaller than a commune; therefore, population data per commune are not sufficiently detailed. If a study focuses on one specific site, it may be easy to produce an accurate population data layer for the neighbouring area, but for studies involving the whole Natura-2000 network in the EU or a large number of sites, the grid presented in this paper is being used.

The recently launched Geoland2 project (<http://www.gmes-geoland.info/>) is also using the population density grid to stratify a sampling area frame for land cover change estimation. The principle is that a high population density is likely to generate a change from natural or agricultural to artificial land, while a too low population density may be linked with agricultural abandonment.

The European Forest Fire Information System (Barbosa et al. 2008) is using the population layer to quantify the impact of population density on forest fire risk, even if it is likely to be less important as a risk factor in Europe than social and climatic issues.

The studies of the International Committee on Aviation Environmental Protection mentioned at the beginning of this paper (Vinkx and Visée 2008) have lead to the project of an EU noise map that has produced at the moment a prototype viewer for large agglomerations (<http://noise.eionet.europa.eu/>). The downscaled population density will be used to quantify the impact of noise on the population.

As final conclusion we can say that the population grid presented in this paper has significant limitations, but will hopefully provide a useful tool for researchers that study the relationship of European population with the environment.

Acknowledgments Eurostat provided most of the data for the study. We are particularly grateful to César de Diego, Daniele Rizzi, Albrecht Wirthmann, Daniel Rase, Torbjorn Carlquist and Edwin Schaaf. The European forum for Geostatistics provided reference data for validation thanks to Ingrid Kaminger, Rina Tammisto, Niek van Leeuwen, Erik Sommer and Lars Backer. Roger Milego, Oscar Gómez, Stefan Kleeshulte and Tomas Soukup, from the European Topic Centre of Terrestrial Environment (ETC/TE) made useful suggestion. Mette Lund managed the distribution of the results through the EEA dataservice. Brooke Tapsall and four anonymous reviewers made useful suggestions to improve the paper.

References

- Annoni, A., Luzet, C., Gubler, E., & Ihde, J. (2001). *Map projections for Europe*. Report EUR 20120 EN, JRC-Ispra (Italy), 131 pp.
- Barbosa, P., Camia, A., Kucera, J., Liberta, G., Palumbo, I., & San-Miguel-Ayanz, J. (2008). Assessment of forest fire impacts and emissions in the European Union based on the European Forest Fire Information System. *Dev Environ Sci*, 8, 197–208.
- Bettio, M., Delincé, J., Bruyas, P., Croi, W., & Eiden, G. (2002). Area frame surveys: Aim, principals and operational surveys. Building Agri-environmental indicators, focussing on the European Area frame Survey LUCAS. EC report EUR 20521, pp. 12–27. <http://agrienv.jrc.it/publications/ECpubs/agri-ind/>.
- Bhaduri, B., Bright, E., Coleman, P. H., & Dobson, J. (2002). LandScan: Locating people is what matters. *Geoinformatics*, 5(2), 34–37. <http://www.ornl.gov/sci/landscan/>.

- Bierkens, M. F. P., Finke, P. A., & de Willigen, P. (2000). *Upscaling and downscaling methods for environmental research* (p. 190). Dordrecht: Kluwer.
- Bracken, I., & Martin, D. (1989). The generation of spatial population distributions from census centroid data. *Environment & Planning A*, 21(4), 537–543.
- Bracken, I., & Martin, D. (1995). Linkage of the 1981 and 1991 UK censuses using surface modelling concepts. *Environment & Planning A*, 27(3), 379–390.
- Briggs, D. J., Gulliver, J., Fecht, D., & Vienneau, D. M. (2007). Dasymetric modelling of small-area population distribution using land cover and light emissions data. *Remote sensing of Environment*, 108, 451–466.
- Center for International Earth Science Information Network (CIESIN). (2005). Gridded Population of the World (GPW), Version 3. CIESIN, Columbia University, Palisades, NY. <http://sedac.ciesin.org/plue/gpw>.
- Chen, K., McAneney, J., Blong, R., Leigh, R., Hunter, L., & Magill, C. (2004). Defining area at risk and its effect in catastrophe loss estimation: A dasymetric mapping approach. *Applied Geography*, 24, 97–117.
- Delincé, J. (2001). A European approach to area frame survey. *Proceedings of the Conference on Agricultural and Environmental Statistical Applications in Rome (CAESAR)*, June 5–7, vol. 2 pp. 463–472. <http://www.ec-gis.org/>.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B*, 39, 1–38.
- Dijkstra, L., & Poelman, H. (2008). Remote rural regions: How proximity to a city influences the performance of rural regions, *Regional Focus* n.1 (2008) EC-DG REGIO.
- Dobson, J. E., Bright, E. A., Coleman, P. R., Durfee, R. C., & Worley, B. A. (2000). LandScan: A global population database for estimating populations at risk. *Photogrammetric Engineering and Remote Sensing*, 66(7), 849–857.
- EEA-ETC/TE. (2002). Corine land cover update, Technical guidelines, http://www.eea.europa.eu/publications/technical_report_2002_89.
- Eicher, C., & Brewer, C. (2001). Dasymetric mapping and areal interpolation: Implementation and evaluation. *Cartography and Geographic Information Science*, 28, 125–138.
- Flowerdew, R., Green, M., & Kehris, E. (1991). Using areal interpolation methods in GIS; *Papers in regional science*, 70(3), 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.
- Langford, M. (2007). Rapid facilitation of dasymetric-based population interpolation by means of raster pixel maps. *Computers, Environment and Urban Systems*, 31(1), 19–32.
- Langford, M., Maguire, D. J., & Unwin, D. J. (1991). The area transform problem: Estimating population using remote sensing in a GIS framework. In I. Masser & M. Blakemore (Eds.), *Handling geographical information: Methodology and potential applications* (pp. 55–77). London, UK: Longman.
- Langford, M., & Unwin, D. J. (1994). Generating and mapping population density surfaces within a geographical information system. *Cartographic Journal*, 31(1), 21–26.
- Martin, D. (1996). An assessment of surface and zonal models of population. *International Journal of Geographical Information Systems*, 10(8), 973–989.
- Martin, D. (1998). Optimizing census geography: The separation of collection and output geographies. *International Journal of Geographical Information Science*, 12(7), 673–685.
- Martin, D., Tate, N. J., & Langford, M. (2000). Refining population surface models: Experiments with Northern Ireland Census data. *Transactions in GIS*, 4(4), 343–360.
- Mennis, J. (2003). Generating surface models of population using dasymetric mapping. *Professional Geographer*, 55(1), 31–42.
- Mrozinski, R., & Cromley, R. (1999). Singly—and doubly—constrained methods of areal interpolation for vector-based GIS. *Transactions in GIS*, 3, 285–301.
- Reibel, M., & Bufalino, M. (2005). Street-weighted interpolation techniques for demographic count estimation in incompatible zone systems, *Environment and Planning A*, 37(1), 127–139.
- Rosati, L., Marignani, M., & Blasi, C. (2008). A gap analysis comparing Natura 2000 vs National Protected Area network with potential natural vegetation. *Community Ecology*, 9(2), 147–154.
- Thielen, A., Müller, M., Kleist, L., Seifert, I., Borst, D., & Werner, U. (2006). Regionalisation of asset values for risk analyses. *Natural Hazards and Earth System Sciences*, 6, 167–178.

- Tobler, W. R. (1979). Smooth pycnophylatic interpolation for geographical regions. *Journal of the American Statistical Association*, 74, 519–530.
- Tralli, D. M., Blom, R. G., Zlotnicki, V., Donnellan, A., & Evans, D. L. (2005). Satellite remote sensing of earthquake, volcano, flood, landslide and coastal inundation hazards. *ISPRS Journal of Photogrammetry and Remote Sensing*, 59(4), 185–198.
- Vinkx, K., & Visée, T. (2008). Usefulness of population files for estimation of noise hindrance effects. ICAO Committee on Aviation Environmental Protection. CAEP/8 Modelling and Database Task Force (MODTF). 4th Meeting. Sunnyvale, USA, pp 20–22, February 2008.
- Weber, N., & Christophersen, T. (2002). The influence of non-governmental organisations on the creation of Natura 2000 during the European Policy process. *Forest Policy and Economics*, 4(1), 1–12.
- Wu, CH., & Murray, A. T. (2005). A cokriging method for estimating population density in urban areas. *Computers, Environment and Urban Systems*, 29(5), 558–579.
- Xie, Y. (1995). The overlaid network algorithms for areal interpolation problem. *Computers, Environment and Urban Systems*, 19(4), 287–306.
- Yuan, Y., Smith, R. M., & Limp, W. F. (1997). Remodeling census population with spatial information from Landsat TM imagery. *Computers, Environment and Urban Systems*, 21(3–4), 245–258.