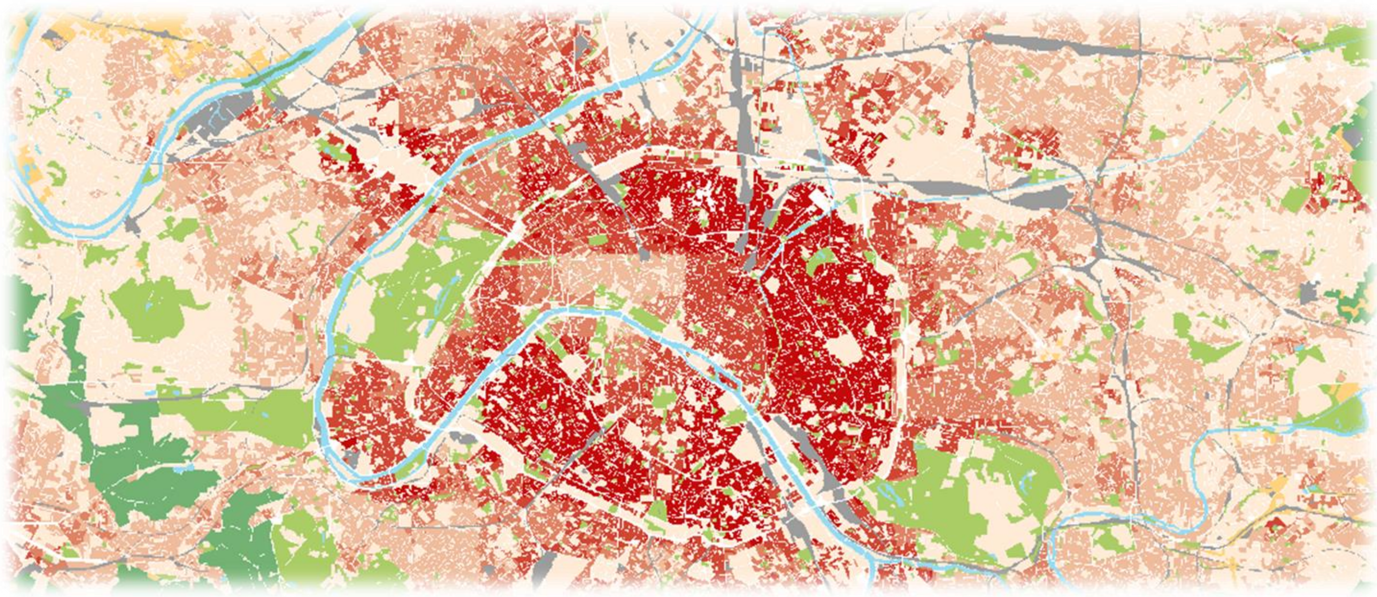
European Commission

JRC TECHNICAL REPORTS

Population Estimation for the Urban Atlas Polygons



Filipe Batista e Silva (DG Joint Research Centre)

Hugo Poelman (DG Regional Policy)

Veerle Martens (DG Regional Policy)

Carlo Lavalle (DG Joint Research Centre)

2013

Joint Research Centre

Report EUR 26437 EN

European Commission
Joint Research Centre
Institute for Environment and Sustainability

Contact information
Carlo Lavalle
Address: Joint Research Centre, Via Enrico Fermi 2749, TP 272, 21027 Ispra (VA), Italy
E-mail: carlo.lavalle@jrc.ec.europa.eu
Tel.: +39 0332 78 5231

http://www.jrc.ec.europa.eu/
http://ies.jrc.ec.europa.eu/our-activities/scientific-achievements/Land-Use-Modelling-Platform.html

This publication is a Reference Report by the Joint Research Centre of the European Commission.

JRC 87300

EUR 26437 EN

ISBN 978-92-79-35089-4 (pdf)

ISSN 1831-9424 (online)

doi:10.2788/54791

Luxembourg: Publications Office of the European Union, 2013

Cover: Population Density in Paris, 2006.

TABLE OF CONTENTS

1. INTRODUCTION

1.1. Scope and objectives

The aim of this technical report is to describe the methodology and source data used to estimate the residential population in each built-up polygon of the Urban Atlas land use/cover dataset[1].

Such estimation is done by an areal interpolation procedure whereby population counts at a given 'source' geometry (regular grids, census tracts or commune boundaries) are transferred to the 'target' geometry, i.e. the Urban Atlas polygons. The transfer of the population counts from the source to the target geometry is done by means of GIS operations, which we describe in detail in this report.

As final outcome of the procedure, a new attribute to the Urban Atlas polygons is added. This attribute – residential, or night time population – will broaden the range of uses of the Urban Atlas dataset, contributing to new analyses and assessments in different thematic fields, e.g. urban quality of life (accessibility to recreational areas; exposure to sources of noise); urban morphology (population density gradients).

1.2. Data used

The data used in this project can be categorized in three main typologies: 'source', 'target' and 'ancillary'. Source data refers to the original population values. These are usually reported at a given geographical zoning system (e.g. regular grids, census tracts or commune boundaries). For each country of the EU-27, the finest source of population was used, subject to availability. As preferred option, high resolution bottom-up grids (< 1 km^2 cell size) were chosen as source data. Bottom-up grids refer to population values assigned to fine grid systems of regular squared cells, usually based on geo-referenced register data or obtained by field surveys/census produced by national statistical offices. Because such grids were not available for the majority of the countries, other sources of population data were sought, namely census tracts and commune boundaries. The spatial resolution/detail of the source data greatly influences the accuracy of the final disaggregation.

[1] http://www.eea.europa.eu/data-and-maps/data/urban-atlas

Target data is herein defined as the set of geographical entities (polygons) for which population needs to be estimated. In this case, the target data are the polygons of the Urban Atlas dataset. The Urban Atlas is a set of land use/cover maps[2] covering over 300 medium to large size European cities and their agglomerations[3].

Lastly, ancillary data is any kind of spatially explicit data that informs the population disaggregation/downscaling process. For the work presented here, a corrected version of the Soil Sealing Layer was used. The soil sealing was assumed as a proxy for building density and, consequently, a proxy for population density. Conceptually, the information on the imperviousness of the soil captures well the horizontal built-up density, but it misses the vertical built-up density, i.e. does not inform on the average heights of buildings. So far, however, the soil sealing layer is still the finest available proxy for population density at EU-level, with a high spatial resolution of 20 meters. EU-wide information on building heights is not expected to be available any soon. Table 1 summarizes the data used in the entire downscaling process.

Table 1. Data used

Data category	**Description**		**Reference year**	**Coverage**
Source	Residential population	Type 1: High resolution bottom-up grids (<1km)	2006 +/- 1	Denmark; Finland; Sweden; Slovenia.
		Type 2: Census tracts		Belgium; England and Wales; Netherlands, Spain*.
		Type 3: Medium resolution bottom-up or hybrid grids (1km)		Austria; France; Portugal.
		Type 4: Commune boundaries		Remaining EU-27 countries.
Target	Urban Atlas polygons (only the polygons presumed to be populated)		2006 +/- 1	All EU-27
Ancillary	Soil Sealing Layer (adjusted version used for the production of the Urban Atlas)		2006	All EU-27

Notes:

* For the Larger Urban Zone of Madrid, bottom-up population data (residential registry points) were aggregated to the Urban Atlas polygons. For the Larger Urban Zone of Seville, a hybrid 1 km grid was used as source data for the disaggregation.

[2] The Urban Atlas mapping guide and class description can be found here: http://www.eea.europa.eu/data-and-maps/data/urban-atlas/mapping-guide

[3] Larger Urban Zones (LUZ) with more than 100.000 inhabitants as defined by the Urban Audit.

2. METHOD

2.1. Main assumptions

2.1.1. Populated and non-populated classes

When transferring source population to the target geometry, the Urban Atlas classes were classified into three major categories relevant to population distribution. This pre-classification is the first main assumption.

- Category 1: Classes assumed to contain most of the resident population:
  - 'Urban fabric classes' (1.1.X.X), ranging from high to very low density urban fabric, and including the class 'isolated structures' (1.1.3.0);
- Category 2: Classes assumed to contain only residual amounts of resident population:
  - 1.2.1.0 (Industrial, commercial, public, military and private units);
  - 1.2.3.0 (Port areas);
  - 1.4.2.0 (Sports and leisure facilities);
  - 2.0.0.0 (Agricultural areas, semi-natural areas and wetlands).
- Category 3: Classes assumed to have no resident population:
  - All remaining classes.

2.1.2. Weighting scheme

The different land use/cover classes are assumed to have different population densities. To perform the disaggregation, each populated polygon needs to be assigned a population 'weight' that accounts for the differences in population density between different land use/cover classes.

The soil sealing degree refers to the soil imperviousness and is expressed as a percentage. It was assumed that the sealing degree is a proxy for building density and, consequently, for population density. Therefore, the population weights were attributed to each polygon of each class according to the observed average soil sealing degree. With few exceptions[4], the weights attributed to each polygon were within the soil sealing thresholds as defined in the Urban Atlas mapping guide for each class. This reasoning applies only to the land use classes of category 1. In what regards the classes of category 2, the population weights were attributed arbitrarily in order to force the allocation of residual amounts of population in such polygons. The overall weighting scheme is summarized in table 2.

Table 2. Weighting scheme used to distribute population according to the soil sealing levels.

Urban Atlas class	**Description (from the Urban Atlas Mapping Guide)**	**Weight**	
		Method	**Value***
11100	Continuous urban fabric (S.L. > 80%)	Directly derived from the average soil sealing degree of polygons	80-100
11210	Discontinuous dense urban fabric (S.L. 50% - 80%)		50-80
11220	Discontinuous medium density urban fabric (S.L. 30 - 50%)		30-50
11230	Discontinuous low density urban fabric (S.L. 10% - 30%)		10-30
11240	Discontinuous very low density urban fabric (S.L. < 10%)		4-9**
11300	Isolated structures		4-9***
12100	Industrial, commercial, public military and private units	Attributed arbitrarily	1
12210	Fast transit roads and associated land		0
12220	Other roads and associated land		0
12230	Railways and associated land		0
12300	Port areas		0.1
12400	Airports		0
13100	Mineral extraction and dump sites		0
13300	Construction sites		0
13400	Land without current use		0
14100	Green urban areas		0
14200	Sports and leisure facilities		1
20000	Agricultural areas, semi-natural areas and wetlands		0.1

[4] Exceptions to this general rule occurred due to some discrepancies between the observed average soil sealing at the polygon level and the respective class definition. Such discrepancies are usually artifacts originated by the size and shape of the polygons when calculating the average sealing.

30000	Forests		0
50000	Water		0

Notes:

* Indicative thresholds. Miner differences in the actual weights attributed may vary in individual polygons. See footnote 4.

** The lower threshold for the class 11240 was obtained through empirical analysis of observed sealing values.

*** For the class 11300 it was assumed a population weight equal to the one used for the class 11240. The only difference is that the polygons of this class are not contiguous to other urban fabric polygons.

2.2. Disaggregation

To redistribute population from its source geometry to the target spatial units, source and target geometries are firstly intersected geometrically through a GIS operation, resulting in a third geometry which we will refer to as 'transitional' geometry.

The next step is to estimate the population for each polygon of the 'transitional' geometry. The following formulation was used:

$$P_i' = P_s . \left(\frac{A_i . W_i}{\sum_i^n A_i . W_i} \right)$$

where:

P_i' corresponds to estimated population of a given polygon i of the transitional geometry;

P_s is the known population in the source zone s;

A_i is the area of polygon i;

W_i is the weight assigned to polygon i, corresponding to the average soil sealing value;

n corresponds to the number of transitional polygons within each source polygon.

Finally, the estimated population for each Urban Atlas polygon t is simply:

$$P_t' = \sum_i^j P_i'$$

where j corresponds to the number of transitional polygons within each target polygon of the Urban Atlas dataset.

2.3. Software used

The procedure was implemented using ArcGIS geoprocessing tools. A script written in Python programming language, and accessible as a tool within the ArcGIS environment was created to facilitate the processing. This script deals with all kinds of source data as input for the disaggregation, and allows batch processing. Figure 1 shows the interface of the script. This approach allows an easy re-run of specific cities whenever finer source data are available.

The processing time for each task (disaggregation of population for each Large Urban Zone) varied from 2 to 40 minutes, depending on the size of the Larger Urban Zone as well as on the number of polygons of source and target geometries.

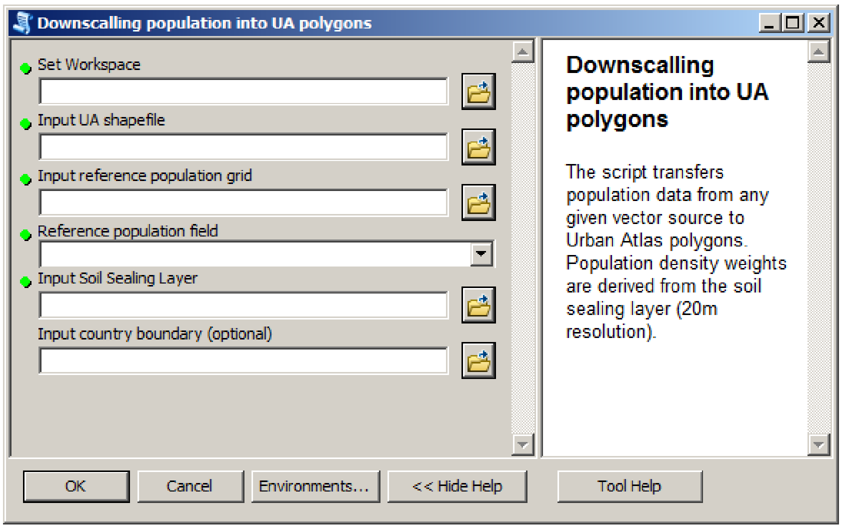


Figure 1. Interface of the script within ArcGIS environment.

3. RESULTS

The final output is a new attribute field ('POP06') of the original Urban Atlas polygons. As a result of the approach, fractions of inhabitants (decimal values) were attributed to the polygons. In order to remove this effect, values were rounded for all spatial units. This means that polygons which were attributed less than 0.5 persons were modified to 0 persons. As a result of this procedure, slight underestimations of total population usually occurred at the level of the Larger Urban Zones (fewer allocated persons compared to the initial value given by the source data)[5].

Figure 2 depicts an example of the disaggregation operated for Rennes, France, where the gain in spatial detail is self-evident.

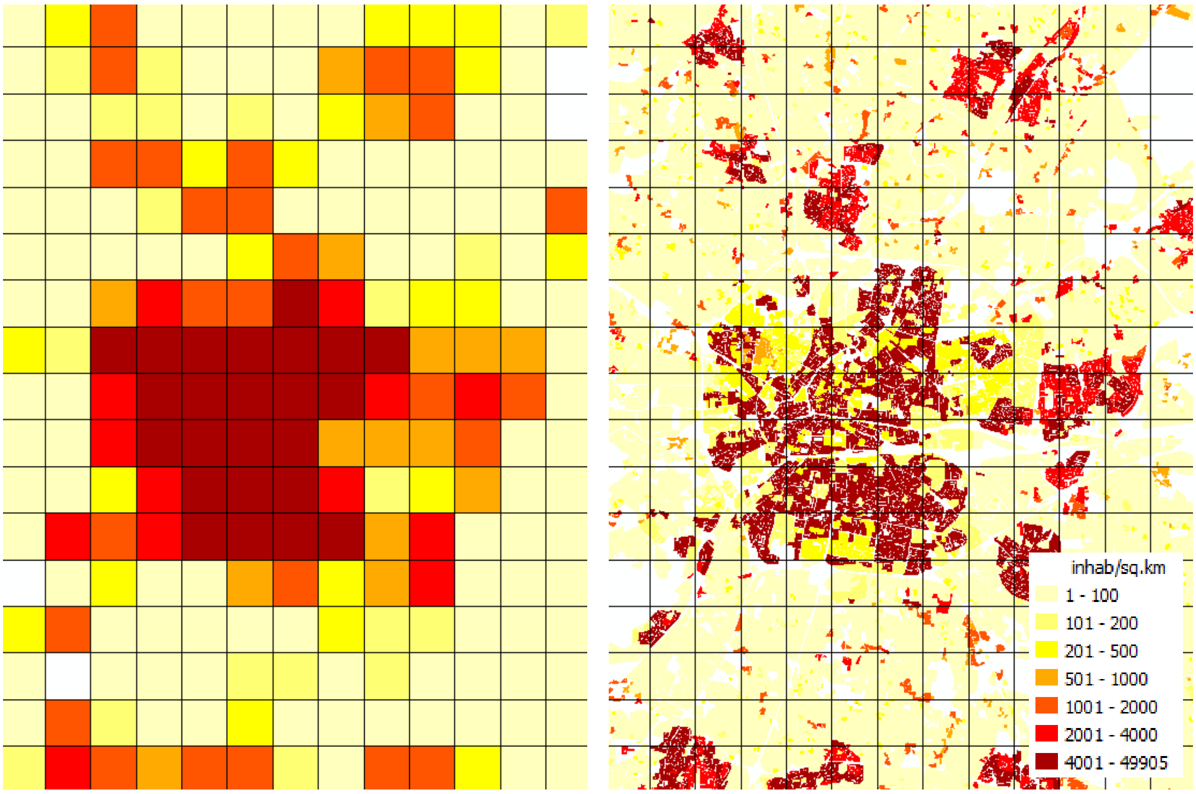


Figure 2. Disaggregation operated for Rennes, France. On the left, source population reported at the level of 1 km grid cells. On the right, the final result of the disaggregation. For visual comparison reasons, population is expressed as nr. of inhabitants per square Kilometre in both left and right.

[5] Average loss estimated on 0.033%. The observed total losses per LUZ were never greater than 0.5%.

4. VALIDATION

In a disaggregation procedure such as the one herein described, the outcome of the process is never less accurate than the original source data. By disaggregating numerical data from one coarse geometry to a finer geometry, we always gain detail and approximate ground truth without the risk of deteriorating the source information. The degree to which the disaggregation approximates reality, however, varies greatly, and it depends chiefly on: 1) the quality of the ancillary data and 2) the appropriateness of the disaggregation algorithm and its parameters.

To assess the quality of the final product we carried a validation exercise for a sample of countries/regions for which very high spatial resolution ground truth data were available (point/building data): Austria, Finland, Portugal, and Madrid. The validation is a relatively straightforward process by which the actual known residents at point/building level are aggregated at the level of the Urban Atlas polygons, and then compared to the disaggregation estimate. The Total Absolute Error (TAE) was chosen to measure the overall disagreement observed per country. In the formula below, P' and P are the estimated and known population, respectively, for each Urban Atlas polygon *t*.

$$TAE = \sum_{t} |P_t' - P_t|$$

By definition, the TAE varies within the range of $[0 , (2 * \sum_t P_t)]$, i.e. from zero to twice the total population of the study area. For easier interpretation, the TAE can be made relative to the total population of the study area:

$$RTAE = \frac{\frac{TAE}{\sum_t P_t}}{2}, \in [0 , 1]$$

If RTAE = 0, then the disaggregation can be considered perfect, i.e. as accurate as the ground truth. If RTAE = 1, then the disaggregation is completely wrong. The latter would happen if, for the surface within each source zone, people were allocated where it does not exist while residential areas were left uninhabited by the disaggregation produced. These extremes value are, of course, rare to obtain through disaggregation exercises. An

RTAE of 0.5 obtained for a given study area would mean that, on average, half of the population is misplaced within each source zone.

Table 3 shows the results of the validation for the LUZs within Finland, Austria, and Portugal. For Finland, the source data consisted of bottom-up grids of 250 meter resolution, while for Austria and Portugal 1 km bottom-up grids were used. For the Madrid study area, three different source zone systems were tested.

For Finland the RTAE varies between 0.12 and 0.15 among 5 validated Large Urban Zones. For Austria and Portugal the RTAE varies consistently between 0.22 and 0.29. The gap between the Finish and the Austrian and Portuguese RTAE values clearly demonstrates the influence of the source zone resolution in the success of the disaggregation. The case of Madrid is eloquent, as it shows, for the same study area, that increased resolution of the source data leads to improved accuracy of the disaggregation process.

Table 3. Validation results.

UATL City			Source data		RTAE [0-1]
Code	**Country**	**Name**	**Type**	**Median unit size (sq. Km)**	
FI001	Finland	Helsinki	Bottom-up	0.25	0.12
FI004	Finland	Oulu	Bottom-up	0.25	0.13
FI002	Finland	Tampere	Bottom-up	0.25	0.15
FI003	Finland	Turku	Bottom-up	0.25	0.14
AT001	Austria	Wien	Bottom-up	1.00	0.22
AT002	Austria	Graz	Bottom-up	1.00	0.27
AT003	Austria	Linz	Bottom-up	1.00	0.28
AT004	Austria	Salzburg	Bottom-up	1.00	0.26
AT005	Austria	Innsbruck	Bottom-up	1.00	0.22
PT001	Portugal	Lisboa	Bottom-up	1.00	0.25
PT002	Portugal	Porto	Bottom-up	1.00	0.24
PT003	Portugal	Braga	Bottom-up	1.00	0.25
PT004	Portugal	Funchal	Bottom-up	1.00	0.26
PT005	Portugal	Coimbra	Bottom-up	1.00	0.26
PT006	Portugal	Setubal	Bottom-up	1.00	0.23
PT007	Portugal	Ponta Delgada	Bottom-up	1.00	0.24
PT008	Portugal	Aveiro	Bottom-up	1.00	0.23
PT009	Portugal	Faro	Bottom-up	1.00	0.29
ES001	Spain	Madrid	Communes	33.49	0.35
ES001	Spain	Madrid	Bottom-up	1.00	0.26
ES001	Spain	Madrid	Census tracts	0.05	0.19

Table 4 shows results for Austria, Finland and Portugal per land use class. The 'urban fabric' classes are affected, in general, by overestimation (chiefly class 11220), while classes such as 11240, 11300, 12100, 14200 and 13300 are affected by significative absolute and relative underestimation. These misplacements of population between land use classes are likely driven by different factors such as temporal mismatches between the population and the land use data; poor calibration of the relationship between soil sealing levels and population density; missing information regarding building height or building volume.

Table 4. Absolute and relative error per country and land use categories.

Land Use Class		Austria		Finland		Portugal		
Code	**Label**	Absolute Error	% Error	Absolute Error	% Error	Absolute Error	% Error	
11100	Continuous Urban Fabric	47,484	4.7%	7,597	2.9%	-111,635	-4.2%	
11210	Discontinuous Dense Urban Fabric	5,177	0.4%	18,043	4.6%	237,074	18.4%	
11220	Discontinuous Medium Density Urban Fabric	116,785	14.4%	21,012	4.4%	88,644	26.1%	Weights derived from SSL
11230	Discontinuous Low Density Urban Fabric	19,354	7.5%	8,369	1.7%	9,376	7.6%	
11240	Discontinuous Very Low Density Urban Fabric	-6,023	-39.6%	-27,126	-11.0%	-6,188	-39.1%	
11300	Isolated Structures	-27,676	-48.3%	4,303	7.6%	-19,398	-53.8%	
12100	Industrial, commercial, public, military and private units	-141,065	-83.4%	-35,612	-49.8%	-121,755	-81.6%	
12300	Port areas	-135	-61.0%	-104	-51.8%	-53	-32.9%	Ad-hoc weights
14200	Sports and leisure facilities	-8,766	-53.7%	2,887	229.3%	-1,815	-29.5%	
20000	Agricultural + Semi-natural areas + Wetlands	154	0.6%	3,020	33.6%	-42,627	-67.3%	
12400	Airports	-124	-100.0%	-37	-100.0%	-46	-100.0%	
13300	Construction sites	-4,215	-100.0%	-1,741	-100.0%	-26,343	-100.0%	No pop. assigned
13400	Land without current use	-951	-100.0%	-257	-100.0%	-5,234	-100.0%	

5. CONCLUSIONS AND WAY FORWARD

The work herein presented resulted in a valuable dataset that is currently being used internally in the European Commission services. It displays population with a very high spatial detail, enabling diverse analyses at fine scale for a set of +300 European Large Urban Zones. By merging land use and population data, additional spatial detail is gained without deteriorating the original population data sources. Countries for which the dataset is the most reliable are all those for which the downscaling was performed from bottom-up grids or census tracts, as reported in table 1, and others like Ireland, Bulgaria and some regions in Germany and Czech Republic where the communes have small sizes on average.

The disaggregation process was fairly straightforward, and was informed by the land use and soil sealing. The validation, while showing satisfactory results for the sampled countries, also revealed misplacements of population between land use classes. This could have been caused chiefly by a simple parameterization of the relationship between the soil sealing information and the population density, but also by missing information on building heights/volumes. Future updates of this dataset will expand upon the current work and take on board new data and methodological improvements.

ACKNOWLEDGEMENTS

For providing detailed bottom-up population data (grid cells < 1 km²):

- Francisco Goerlich (University of Valencia)
- Madrid Regional Statistical Institute
- Statistics Denmark
- Statistics Slovenia
- Statistics Finland
- Statistics Sweden

For collaborating in the validation of the results by comparing our estimates with population point-data

- Ana M. Santos (Statistics Portugal)
- Rina Tammisto and Markku Koivula(Statistics Finland)
- Ingrid Kaminger (Statistics Austriaa)

APPENDIX 1: BORDER ADJUSTMENTS

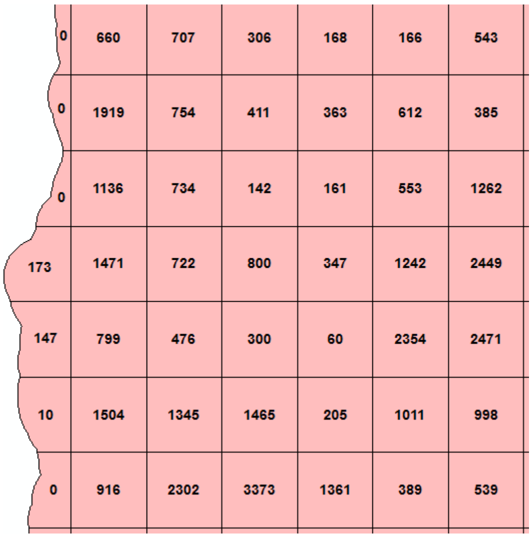
The source population data comes in various forms/geometries: bottom-up grids, census tracts and commune boundaries. The geometry of the source data may not always coincide with the target geometry, particularly when the source data comes as a regular grid. The mismatch is troublesome along the sea line and country borders.

It is common to find populated cells which have a portion of area on top of the sea, but whose reported population refers only to the actual land surface of the cell. Similarly, in a grid of cells reporting population for country A, some of those cells will eventually include a portion of area of a neighbouring country B. In such situations, the source data has to be clipped by the boundary of the country, thus removing the unpopulated surface from the cell.

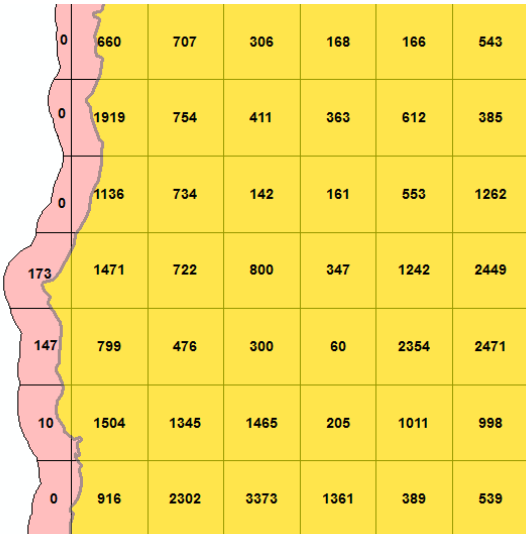
In addition, the spatial extent of the Larger Urban Zones is smaller than the extent of the source data. Therefore, remaining source cells ought to be clipped by the border of the Larger Urban Zone and its population adjusted. The adjustment is done through a simple areal weighting rule.

These preparatory steps of the source data are part of the script. The following sequence of images and respective labels illustrate the adjustments mentioned above.

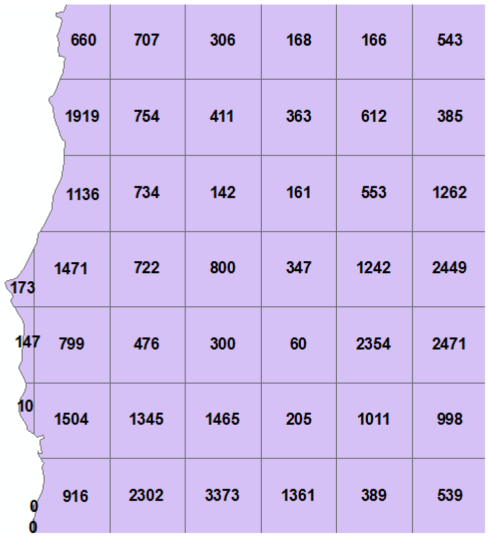
Original source data (grid), with respective population per cell.



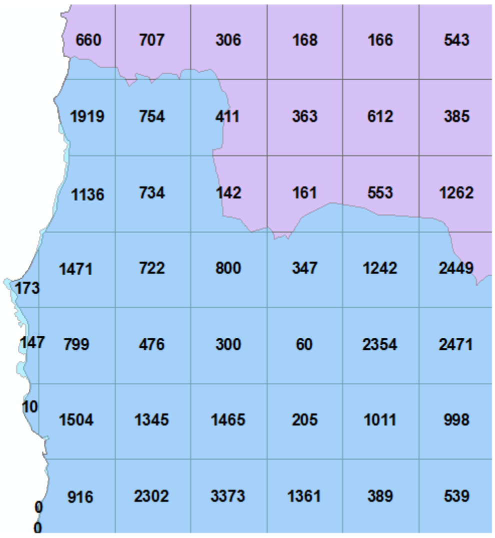
The actual land area, in yellow.



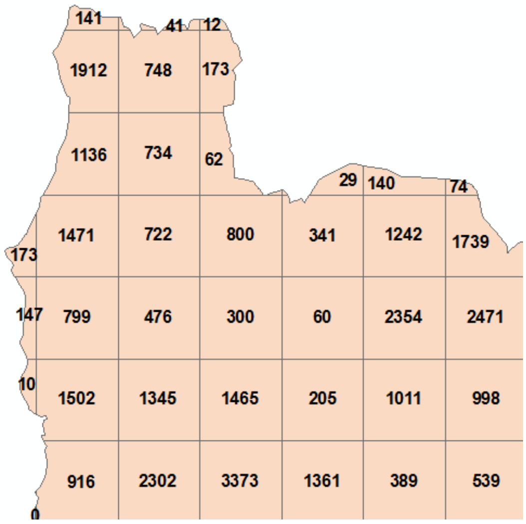
Step 1: The grid cells are clipped by the country border. Population values are kept the same.



Border of the Larger Urban Zone, in blue.



Step 2: The grid cells are clipped by the border of the Larger Urban Zone. Population in clipped cells is adjusted through simple areal weighting. The resulting grid is used in the subsequent disaggregation steps.



APPENDIX 2: SOURCE DATA

Belgium

Total population, 31/12/2006, of statistical sectors (v. 01/10/2001).

Coordinate system: Lambert belge 1972

Federal Government Office of Economy

Denmark

Total population, 01/01/2006, on grid cells of 100 * 100 m (Danish national grid)

Coordinate system: UTM zone 32, EUREF89

Statistics Denmark

Germany

Total population, 2006, of LAU2 units

Data collected by Eurostat

Exception:

Berlin (only city of Berlin):

Population, 31/12/2006, of LOR Planungsraüme (447 sub-LAU2 units)

Government of Berlin

Spain

Total population, 2006, of census tracts

Coordinate system: ETRS89, Lambert Azimuthal Equal Area

National Statistical Institute (INE)

Exceptions:

Madrid:

Registered population, 2006, aggregated to Urban Atlas polygons

Coordinate system: ETRS89, Lambert Azimuthal Equal Area (EPSG 3035)

Population data source: Madrid Regional Statistical Institute

Aggregation: Departamento de Análisis Económico, University of Valencia

Sevilla:

Estimate of total population, 2006, on grid cells of 1 * 1 km

Coordinate system: ETRS89, Lambert Azimuthal Equal Area (EPSG 3035)

Departamento de Análisis Económico, University of Valencia

France

Estimate of total population, 2006, on grid cells of 1 * 1 km

Coordinate system: ETRS89, Lambert Azimuthal Equal Area (EPSG 3035)

INSEE

Netherlands

Total population, 01/01/2006, of neighbourhoods (wijken en buurten)

Coordinate system: Rijksdriehoekstelsel_new, GCS_Amersfoort

Statistics Netherlands

Cyprus

Estimate of total population, 2006, of LAU2 units

Coordinate system: ETRS89, Lambert Azimuthal Equal Area (EPSG 3035)

National Statistical Institute, PRIO (Peace Research Institute Oslo) and DG REGIO estimates.

Austria

Total population, 2006 on grid cells of 1 * 1 km

Coordinate system: ETRS89, Lambert Azimuthal Equal Area (EPSG 3035)

Statistics Austria

Portugal

Estimate of total population, 2006, on grid cells of 1 * 1 km

Coordinate system: ETRS89, Lambert Azimuthal Equal Area (EPSG 3035)

Statistics Portugal

Slovenia

Total population, 2006, on grid cells of 100 * 100 m

Coordinate system: Transverse Mercator, GCS_MGI_1901 (EPSG 3912)

Statistical Office of the Republic of Slovenia

Finland

Total population, 31/12/2006, on grid cells of 250 * 250 m

Coordinate system: Finnish coordinate system KKJ3 (EPSG 2393)

Statistics Finland

Sweden

Total population, 31/12/2006, from Statistics Sweden's Total Population Register (TPR), on grid cells of 250*250 m in urban areas and 1*1 km in rural areas

Coordinate system: national Swedish projection SWEREFF99 (EPSG 3006)

Statistics Sweden

United Kingdom *(England and Wales only)*

Estimate of total population, mid-year 2006, of LSOA (Larger Super Output Areas)

Coordinate system: British national grid (EPSG 27700)

Office for National Statistics / Ordnance Survey

Other countries/areas

Population data (2006) at LAU2 level

National Statistical Institutes, data collected by Eurostat

European Commission
EUR 26437 – Joint Research Centre – Institute for Environment and Sustainability

Title: Population Estimation for the Urban Atlas Polygons

Authors: Filipe Batista e Silva, Hugo Poelman, Veerle Martens, Carlo Lavalle

Luxembourg: Publications Office of the European Union

2013 – 20 pp. – 21.0 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN 1831-9424 (online)

ISBN 978-92-79-35089-4 (pdf)

doi:10.2788/54791

Abstract

The aim of this technical note is to describe the methodology and source data used to estimate the residential population in each built-up polygon of the Urban Atlas land use/cover dataset, and to document a validation exercise of the dataset.
The final outcome of the procedure is a new attribute in the Urban Atlas polygons with the potential to broaden the range of uses of the Urban Atlas dataset, contributing to new analyses and assessments in different thematic fields, such as urban quality of life (accessibility to recreational areas; exposure to sources of noise); urban morphology (population density gradients), among others.

LB-NA-26437-EN-N

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new standards, methods and tools, and sharing and transferring its know-how to the Member States and international community.

Key policy areas include: environment and climate change; energy and transport; agriculture and food security; health and consumer protection; information society and digital agenda; safety and security including nuclear; all supported through a cross-cutting and multi-disciplinary approach.

Publications Office