



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

PORTUGUESE POPULATION GRIDS FOR THE YEARS 2001, 2006 AND 2011.

Thursday, November 24, 2011

FRAMEWORK

The “Geographic Information Referencing Base” (BGRI) is the geographic reference base used for the 2001 and 2011 Census operation. The BGRI divides each basic administrative unit, “Parish” (LAU2), into smaller statistical areas – statistical sections and subsections:

- Statistical Section - Territorial unit representing the continuous area of a single Parish with approximately 300 dwellings ;
- Statistical Subsection - Territorial unit, whether built-up or not, within the statistical section. It represents a block in urban areas, a locality or part of a locality in rural areas or residual areas which may or may not have dwellings.



In 2011 Statistics Portugal created a complete coverage with the location of all the residential buildings of the Portuguese territory.

A Geographical Information Referencing Base was also used for the years 1981 and 1991 Census, however only the 1991 base covered the whole of the National territory. The following table shows the characteristics of the different years:

Census	Sections (enumeration areas)	Subsections (building blocks)	100% territorial coverage	Digital format	Buildings geocoding
1981	Major cities	Major cities	No	No	No
1991	13 705	106 626	Yes	Yes	No
2001	16 094	177 893	Yes	Yes	Partial – 400 000
2011	18 074	265 955	Yes	Yes	3 549 435

Statistics Portugal doesn't have experience working with grids. However, due to the spatial data infrastructure and the characteristics of census 2001 and 2011 data, it was possible to produce a population grid data for both years using the BGRI 2001 and 2011. The 2006 population Grid was built using Municipality (LAU1) administrative units areas and population estimation techniques.

INPUT DATA

The data exists for the following 4 geographical distinct areas, each area has its own geographical reference system.

Geographical region	Datum
Portugal mainland	ETRS89
Madeira Islands	WGS 1984
Eastern Azores Island groups	ITRF 1993
Western Azores Island groups	ITRF 1993

The following datasets for the different geographical areas have been used:

- 2001 Subsections (BGRI2011)
- 2011 Subsections (BGRI2001)
- 2011 Residential Buildings
- 2006 Municipal Borders

The population data

Year	
2001	Residential Population at subsection level
2006	Annual estimates of resident population at municipality level (LAU1)
2011	Residential Population at subsection level

Reference GRID: INSPIRE 1km2 reference GRID in LAEA projection with ETRS89 datum.

METHODOLOGY

GENERAL REMARKS ABOUT THE METHODOLOGY

For the years 2001 and 2011 the data at subsection level was used to determine the population values for the INSPIRE population GRID. Information on the location of the 2011 buildings was used to calculate the population distribution within the grid.

The 2001 and 2011 population distribution by municipality has been used to estimate the population values for the INSPIRE population GRID for the year 2006.

At the moment the data for the year 2011 is only available for the major variables at the subsection level. Only with the publishing of the definitive 2011 Census results, by the end of 2012, it will be possible to create a population GRID for the year 2011 using a 100% bottom-up methodology. It will also be possible to distinguish the buildings that already existed in 2001 and these can be used to get a better distribution for the 2001 population values within the Grid. When publishing the 2011 definitive results the number of dwellings and population within a building will also be known.

CREATING 2011 GRID

The 2011 population GRID was created using the following methodology.

1. Projection of the 2011 subsections (BGRI2011) to the European LAEA projection;
2. Union between BGRI2011 and the INSPIRE population GRID, creating polygons for the combination of features within both layers;
3. Spatial join with the geographical location of 2011 residential buildings to determine the number of buildings within each polygon;
4. Calculate population values for the INSPIRE grid cells:
 - a. If the subsections are not split across grid cells then the population grid cell is the sum of the population within each subsection;
 - b. If the subsections are split across grid cells then the population grid cell depends on the percentage breakdown of residential buildings within each subsection.

This methodology is based on the following assumption:

- Each residential building has the same number of individuals.

The population for each grid cell (G) is given by:

$$G_j = \sum_{i=1}^{N_j} S_{ij}$$

Where:

S_{ij} = Population of subsection i \times $\frac{\text{Number of 2011 residential buildings within 2011 subsection } i \text{ in grid } j}{\text{Number of 2011 residential buildings in subsection } i}$

N_j = Number of subsections in grid j

No error-measures have been calculated for the grid cell population estimates.

Improvements after definitive Census results:

- The exact population for each building will be known and therefore it will be possible to determine the exact population within each grid-cell.

CREATING THE 2001 GRID

The values for the 2001 population GRID will be created using the subsequent methodology.

1. Projection of the 2001 subsections (BGRI2001) to the European LAEA projection;
2. Union between BGRI2001 and the INSPIRE population GRID, creating polygons for the combination of features within both layers;
3. Spatial join with the geographical location of 2011 residential buildings to determine the number of buildings within each polygon;
4. Calculate population values for the INSPIRE grid cells, using:
 - a. If the subsections are not split across grid cells then the population grid cell is the sum of the population within each subsection;
 - b. If the subsections are split across grid cells then the population grid cell depends on the percentage breakdown of residential buildings within each subsection;

This methodology is based on the following assumption:

- Each residential building has the same number of individuals.

The population for each grid cell (G) is given by:

$$G_j = \sum_{i=1}^{N_j} S_{ij}$$

Where:

$$S_{ij} = \text{Population of subsection } i \times \frac{\text{Number of 2011 residential buildings within 2001 subsection } i \text{ in grid } j}{\text{Number of 2001 residential buildings in subsection } i}$$

$$N_j = \text{Number of subsections in grid } j$$

No error-measures have been calculated for the grid cell population estimates.

Improvements after definitive Census results:

- The 2001 buildings can be distinguished;
- The number of dwellings within each building will be known (it will be assumed that the population distribution will be the same at the dwelling level).

CREATING THE 2006 GRID

The population data for non Census years is only available on municipality level (LAU1). These data and the 2001 and 2011 population distribution will be used to estimate the 2006 population for each grid cell. The population for each grid cell for 2006 is obtained as follows:

1. Projection of the municipality (LAU1) borders of 2006 to the European LAEA projection. In 2001 and 2011 the same municipalities existed;
2. Union between municipalities polygons and the INSPIRE population GRID, creating polygons for the combination of features within both layers;
3. Determination of the percentage breakdown of population in 2001 and 2011 within each municipality that reside in a particular grid;
4. Estimate the 2006 grid cell population as a linear combined estimates of the previous percentages with fixed weights multiplied by the 2006 population in the municipality;
For example, assume that in 2001 and 2011 the percentage population of a given municipality residing in a particular grid cell is 10% and 15%, respectively. If the total population in 2006 in that municipality is 1000, then the population grid cell is (assuming a weight=0,5): $[(0,5*0,1)+(0,5*0,15)]*1000 = 125$.

This methodology has the following assumptions:

- The changes of the population within each municipality is linear over time;
- The same assumptions are valid as well for the for the 2001 and 2011 population estimates for each grid cell.

The population for each grid cell is give by:

$$G_j = [\alpha P_{ij}^{2001} + (1 - \alpha) P_{ij}^{2011}] \times M_i^{2006}, \quad 0 < \alpha < 1$$

Where:

$$P_{ij} = \frac{M_{ij}}{M_i}$$

M_{ij} = Population of municipality i in grid j

M_i = Population of municipality i

$\alpha = 0.5$

No error-measures have been calculated for the grid cell population estimates.

Improvements after definitive Census results:

- The real population value of each cell in 2011 will be known;
- The estimates of the 2001 population for each grid cell will be improved.