

EUROPEAN FORUM FOR GEOSTATISTICS SOFIA CONFERENCE 2013

23 - 25 October, Sofia, Bulgaria "The proof of the apple is in the eating!"



Using the European Grid "ETRS89/LAEA_PT_1K"

as the foundation

for the new Portuguese Sampling Infrastructure

Ana Santos ana.msantos@ine.pt Bart-Jan Schoenmakers bart.schoen@ine.pt





INSTITUTO NACIONAL DE ESTATÍSTICA Statistics Portugal





The presentation

- Background
 - Master Sample 2001
- National Dwellings Register (FNA)
 - Methodology
- Geographical Component
- Using the European GRID
- Tools
- Final Remarks







- Since 1979 household Surveys are based on a large sample of housing units, commonly referred to as "Master Sample" (MS)
- The Master Sample is selected after the completion of each Census and is maintained over a decade updated to a greater or lesser degree

Situation until 2013 - The 2001 Master Sample (MS2001)

- A stratified, one-stage cluster sample selected with probability proportional to size
- Designed taking into account the provisional results of the 2001 Population and Housing Census
- Initially planned for a 5 year period
- Between 2006 and 2010 Statistics Portugal conducted a fieldwork to guarantee that the MS2001 could be used until 2013, the year of the transition to the new sampling frame







The distribution of the 2001 MASTER SAMPLE

by NUTS II (1408 areas)

EUROPEAN FORUM

FOR GEOSTATISTICS





- In 2011, Statistics Portugal obtained a national databasis comprising all the georeferenced buildings from the 2011 Population and Housing Census
- This geographical basis has been used to reference census data at point level and to support the creation of a **National Dwellings Register (FNA)**
- **FNA** can be updated by taking advantage of data available in different sources, like statistical surveys conducted by Statistics Portugal and also the following administrative sources:

Source	Entity	Register	Frequency
ADENE	Agency for Energy	Efficiency and Indoor Air Certification	Monthly
IMI	Tax Authority	Property Tax	Annual / Monthly
SS	Ministry of Labor and	Social Security File	Annual
	Social Security		
IRS	Tax Authority	Income Tax Singular	Quarterly
BDIC/CCIC	Institute of Registries	Base Civil Identification	Annual
	and Notaries, IP		







Conceptual Model







Architecture of the proposed model for Household Surveys









Spatial Data for the FNA









New Geographical Referential

GeoStat 's 1Km2 Grid and georeferenced **Census buildings (points)** are used for the creation of the sampling frame and the extraction of optimized samples

Official GRID (Grid_ETRS89_LAEA_1K) - 94 265 cells of 1 Km² for Portugal
3 547 318 Census buildings



GRID and Census Buildings at Faial Island, Azores





Using the rectangular cells of the European GRID allows:

- A uniform representation of the buildings regardless of the administrative division
- The whole system will be independent from any future changes in the boundaries of the administrative units (LAU1 and LAU2) and statistical units (sections and subsections)







The system is composed of several geographical datasets and the 2011 census population data, all organized within the Datawarehouse of Statistics Portugal and an ArcSDE geographical database

Geographical features

- Centroids of the georeferenced buildings
- The European "ETRS89_LAEA_1K" GRID
- The administrative division at parish level (NUTS5)
- Geographical reference data like road information, digital imaginary and the statistical division





Methodological Aspects:

- i. Relate the Census Buildings to the European GRID
 - A spatial overlay to obtain the corresponding GRID Cell of each building
 - The European GRID is in the LAE-ETRS projection and has been projected for the 4 different projections used for the Portuguese territory.
 - Only 43% of the cells have occupied homes
- ii. Order of cells in each NUTS III area

EUROPEAN FORUM

FOR GEOSTATISTICS

iii.Definition of Primary Sampling Units (PSU) consisting of a subset of the GRID used for different surveys





N° of occupied households by grid cell

EUROPEAN FORUM

FOR GEOSTATISTICS



STATISTICS PORTUGAL



II. Spatial Order of cells in each NUTS III area

- NUTS III areas are used as strata within the sampling process
- Within each stratum, cells are selected in an order which should assure their contiguity/proximity
- Each cell within the NUTS III area is assigned a sequential number (other than the ID)
- The methodology used for this spatial sort is the "PEANO" method (using a space filling curve algorithm, also known as the **Peano curve**), a functionality implemented within the ArcGIS software

Example Ordering Method "PEANO" Methodology







Results

Darker Cells have a higher Rank order

EUROPEAN FORUM

FOR GEOSTATISTICS

 One cell can be represented more than once, when it is located at a NUTS III boundary. The same cell has a different sequential numbers in each distinct NUTS III area









an

03

INSTITUTO NACIONAL DE ESTATÍSTICA Statistics Portugal



GRID Cell with is located in different NUTS III areas and parishes



TATÍSTICA



III. Selection of GRID cells which constitutes the PSU





INSTITUTO NACIONAL DE ESTATÍSTICA Statistics Portugal





Application to visualize and manage samples

- WEB Application developed in AGS 10.0 – Flash
- Visualization:
 - Roads and places names
 - Buildings of the households of the extracted Sample (Possibility to upload from an EXCEL file)

– Functionalities:

- Identification of Coordinates
- Search for Addresses
-









Final Remarks

- This work is an important development for the integration of geographical information to support the sampling infrastructure used for the household surveys conducted by Statistics Portugal
- European GRID is an important component of the National Dwellings Register
- Shows the potential for the use of the European GRID as one of the principal components within the sampling process (assuring independence)
- We recognize that it is important to develop other specific geographic analysis: management of the fieldwork for the different surveys, proximity analysis, definition of specific geographical groups and optimization of samples





EUROPEAN FORUM FOR GEOSTATISTICS SOFIA CONFERENCE 2013 23 - 25 October, Sofia, Bulgaria

"The proof of the apple is in the eating!"

EUROPEAN FORUM FOR GEOSTATISTICS

Thank you

Ana Santos ana.msantos@ine.pt Bart-Jan Schoenmakers bart.schoen@ine.pt





| INSTITUTO NACIONAL DE ESTATÍSTICA | STATISTICS PORTUGAL