Population Association of America - 2011 Annual Meeting Program

Washington, DC – March 31 – April 2 – Marriot Wardman Park

Session 803: Measurement and Methods in Population and Environment Research

Using regular grids for spatial distribution of census data for population and environment studies in Brazil

Álvaro de Oliveira D'Antona[1]

Maria do Carmo Dias Bueno[2]

Ricardo de Sampaio Dagnino[3]

Extended abstract

The spatial distribution of demographic variables is essential in Population and Environment Studies, especially in those regarding issues of Land Use and Land Cover Change (LULCC), in favoring the investigation of the reciprocal effects between the population characteristics and biophysical or structural characteristics where such populations live.

In studies conducted in small areas, supported by specific surveys and samples that respect the population and landscape attributes, it is relatively simple to establish the relationship between a set of variables and the data can be aggregated in territorial units that make sense to the analysis – such as protected areas, colonization projects, and watersheds, among others.

However, in regional analysis it is noticed that secondary data, like the ones obtained in demographic census, do not meet the needs of users who would like to have disaggregated data in the smallest spatial unit possible. Usually, these data are spatially distributed in territorial units defined by operational criteria (census tracts) or by political-administrative boundaries (municipalities) that are presented as if they have

---

[1] Professor at the School of Applied Sciences and the Post-Graduate Program in Demography, researcher at the Population Studies Center, all at the University of Campinas (Unicamp).

[2] GIS Analyst of the Brazilian Institute of Geography and Statistics – IBGE. Master Degree in Geomatics at the Rio de Janeiro University and Doctoral student of the Post-Graduate Program in Demography at the University of Campinas (Unicamp).

[3] Bachelor and Master degrees in Geography, Doctoral student of the Post-Graduate Program in Demography at the University of Campinas (Unicamp). Scholarship from the National Council of Scientific and Technological Development (CNPq - Brazil).

internal homogeneity. This assumption is not always true: there is no guarantee that the demographic data distribution is homogeneous within these units and it is hard to believe that the units make sense from the point of view of the landscape's variables.

Another difficulty in using these data is the temporal comparability, since the boundaries of the territorial units change over time, so that a census tract or a political-administrative region are not necessarily the same as in previous periods. Although there are several methods to transfer data from one area to another, these techniques are limited by sampling strategies and the data confidentiality (Bueno, 2010).

In Brazil, census tracts constitute themselves as the smallest territorial units used for data dissemination of the basic questionnaire of census surveys. For dealing with boundaries delimited because of operational purposes, their use brings the previously mentioned problems. The census tracts are designed based on the capacity of the enumerator to collect the data in that particular territorial unit in the appointed deadline, having an average of 300 households in urban areas and 200 in rural areas. In sparsely populated rural areas, like in the Amazon Region, a census tract can have an excessively large area, passing through distinct types of land use and administrative units. Besides, when constrains are considered to ensure data confidentiality, many census tracts cannot make the data publicly available. In this way, it becomes difficult to use such spatial units in Population and Environment studies, due to the unavailability of the data as well as the incompatibility of the limit of the territorial units used to disseminate demographic data and those relevant to an environmental point of view. For example, in 2000, the largest census tract in the municipality of Altamira occupied a portion of the Baú Indigenous Land and the Menkragnoti Indigenous Land, which had 27 inhabitants distributed over 18,353.14 km², an area three times larger than the Federal District, which had over two million inhabitants in 2000 (Dagnino et al, 2010). Another problem detected in 2000 Census data is the existence of tracts with no information available, since IBGE restricts the access to data aggregated by census tracts, as a way to avoid disclosure risk always when there are less than five households in a unit (IBGE, 2003:20).

In Brazil, in 2007, the data collection of Population Count and Agricultural Census were made using handhelds equipped with Global Position System (GPS), enabling the capture of coordinates of the visited units in the rural areas. This technological upgrade increases the data analysis capacity, since it enables several types of spatial analysis. However, there are still restrictions regarding the legal requirements for preserving the identities of informants. Nowadays there is an intense discussion at IBGE on the ways of making the data publicly available, in an attempt to find a balance between the growing demand of the users for geo-referenced and disaggregated data and the restrictions of a legal and technical nature.

Some possible alternatives to be implemented by the statistical offices that allow the use of more detailed data are the provision of an internal and/or external service where users can work with the micro database without directly accessing it, or the implementation of a confidentiality engine to prevent the identification of the informant (United Nations, 2001) or, the dissemination of data based on regular grids (Tammilehto-Luode, 2003) or the use of output areas drawn artificially based on similarity criteria (Vickers and Rees, 2007).

In the present work, a regular grid was used for the aggregation of some variables of the micro data from the Population Count 2007 for the rural areas, as the quantity of residents by gender and age. Several sizes of cells were tested for the grid, in an attempt to achieve the more detailed dimension that would not compromise confidentiality. From this grid a continuous surface was generated for each variable, and that will enable the estimation of data for any territorial unit.

The data were processed and analyzed using data tabulation and geoprocessing software. As a comparative test, in a GIS environment, the grids with census data were overlapped on vectors that represent territorial units relevant for population and environment studies: protected areas, highway buffer, hydrographic buffer and rural plots. The results were compared with the ones obtained by the overlapping of demographic data per census tracts to those relevant territorial units.

The results show that the technique used leads to results that are closer to reality and more accurate, regarding the distribution of data in territorial units more adequate to population and environment studies. Compared to other methods that enable the spatial distribution of demographic data in grids, the one presented here is based on the real distribution of the households and not on spatial statistics techniques. The results are promising as for the possibility of comparison with new surveys conducted by IBGE, like the Demographic Census 2010, which will be made available in the following year, and with previous surveys (for example, Demographic Census 2000 and 1991), through the use of the same grids generated for 2007, through direct comparison, as in the case of 2010 data or using mathematical models to spatial distribution in the case of previous data.

References

Bueno, M.C.D.. 2010. A proposal for the dissemination of Brazilian 2010 Population Census. Conference of European Forum for Geostatistics, 05–07 October, Tallinn, Estonia. (Accepted to presentation).

Dagnino, Ricardo; D'Antona, Álvaro; Carmo, Roberto. 2010. Espacialização de dados demográficos de setores censitários do Censo 2000: um exercício com os dados de Altamira e São Félix do Xingu, Estado do Pará. Anais do XVII Encontro Nacional de Estudos Populacionais, Caxambu, MG, setembro de 2010. Avaliable on http://www.abep.nepo.unicamp.br/encontro2010/docs_pdf/tema_1/abep2010_2450.pdf

IBGE - Instituto Brasileiro de Geografia e Estatística. 2000. Censo Demográfico 2000 - Agregado por Setores Censitários dos Resultados do Universo (2ª edição), Documentação do Arquivo. Rio de Janeiro, 2003. Avaliable on ftp://ftp.ibge.gov.br/Censos/Censo_Demografico_2000/Dados_do_Universo/Agregado_por_Setores_Censitarios/. Acesso em 19 nov. 2009.

IBGE - Instituto Brasileiro de Geografia e Estatística. 2008. Censos 2007 – Inovações e impactos nos sistemas de informações estatísticas e geográficas do Brasil. Rio de Janeiro, 2008. http://censos2007.ibge.gov.br/Censos2007_Inovacoes_web.pdf

Tammilehto-Luode, M. 2003. GIS for Dissemination of Census Data. *EuroGrid Workshop*. Italy, 27-29 October.

United Nations. 2001. Guidelines on the Application of new information technology to population data dissemination. Economic and Social Commission for Asia and the Pacific.

Vickers, D., Rees, P. 2007. Creating the UK National Statistics 2001 output area classification. Journal of the Royal Statistical Society, A (Statistics in Society). Vol.170, Number 2, pp. 379-403.