High Resolution, Multi-Year Compatible Dasymetric Models of US Population

A. Dmowska[1,2], T. F. Stepinski[1], P. Netzel[1]

[1]University of Cincinnati, Cincinnati, OH, 45221, USA
Email: {dmowskaa;stepintz;netzelpl}@ucmail.uc.edu

[2]Adam Mickiewicz University, Poznan,61-680, Poland
Email: dmowska@amu.edu.pl

Abstract

We developed 30 m resolution grids of the US population in 1990, 2000, and 2010 using a multi-year compatible dasymetric model. These grids are designed to assess population change across the conterminous US at street-level spatial resolution. The model and its novel, computationally efficient implementation in R are described. The grids are available online for interactive exploration and data download using especially developed GeoWeb application SocScape (http://sil.uc.edu/webapps/socscape_usa/).

1. Introduction

The goal of this research is to provide an open and convenient access to high resolution, multi-year compatible population data. Such resource is sought after by academic, government, and industry stakeholders as it can provide access to information needed in numerous applications including social and health services, economic development, and planning. Unfortunately, the US Census data, in its original format of aggregated units, is not multi-year compatible, and, except in the densely populated urban areas, is not given at sufficiently high resolution. Boundaries of small Census aggregates, blocks, block groups, and tracts, change from one Census to another making a direct assessment of change impossible without interpolation. In the rural areas large portions of Census blocks are uninhabited with population restricted to small fragments of the blocks resulting in unrealistic estimates of population density. Numerous other shortcomings of aggregated data have also been identified (Sperling, 2012).

The solution is to calculate a grid-based model of the US population density. Surprisingly, until now no adequate grid-based model had been available. The Socioeconomic Data and Application Center (SEDAC) provides 1 km resolution (250 m for selected metropolitan areas) demographic grids, but they are constructed using an oversimplified model (areal weighting), have insufficient resolution, and are available only for 1990 and 2000. The 90 m dasymetric model (LandScan–USA) has been developed by the Oak Ridge National Laboratory (Bhaduri *et al*., 2007) but it is not available, nor is it expected to be in the public domain when and if it becomes available. Since 2014 we have developed (Dmowska and Stepinski, 2014, 2016) a series of dasymetric models covering the entire conterminous US (CONUS). The first generation of our models was the results of sharpening of SEDAC grids into smaller, 90 m cells; these grids were available only for 1990 and 2000. The second generation grid was the result of direct disaggregation of 2010 Census blocks into 30 m grid using land cover (NLCD2011) and land use (NLUD2010, Theobald 2014) as ancillary data.

Here we report on direct disaggregation of 1990, 2000, and 2010 Census blocks in a manner that makes the three resultant grids comparable to each other thus enabling change assessment. To achieve this we needed to use ancillary data that are compatible between different years. In addition, we constructed a custom dasymetric model and implemented it in R instead of GIS-specific software in order to make the entire computational pipeline efficient and completely automatic.

2. Data and methods

2.1 Population and ancillary data

The source of the population information is the 1990, 2000, and 2010 decennial Censuses data aggregated to block level. This data consist of two components: shapefiles (TIGER/Line Files), indicating blocks' geographical boundaries, and summary text files which lists population data for each block.

For ancillary data – the data used to guide disaggregation of blocks into the high resolution (hi-res) grid – we use the land cover data. This choice is dictated by the fact that land cover is the only ancillary data for which a single dataset – the National Land Cover Dataset or NLCD – covers the entire CONUS. However, the NLCD1992 (to be used for disaggregation of 1990 blocks) has a legend which is incompatible with the legends of NLCD2001 and NLCD2011 (to be used for disaggregation of 2000 and 2010 blocks, respectively). To enable a direct comparison between NLCD1992 and NLCD2001 the Retrofit Land Cover Change Product (http://www.mrlc.gov/nlcdrlc.php) was developed. In effect, this product consists of two land cover maps with number of land cover classes reduced to eight compatible categories. To transform all three NLCD maps to a common legend we reclassify the retrofit maps for 1992 and 2001 and the NLCD 2011 to just three categories: urban, vegetation, and uninhabited. These reclassified maps were used as ancillary data for dasymetric model.

2.2 Dasymetric model and its implementation in R

The spatial resolution of NLCD is 30 m and it is most convenient to disaggregate blocks to the same resolution. The overall dasymetric model follows the methodology we have developed to obtain 2010 population grid (Dmowska and Stepinski 2016) but we only use modified (see above) land cover as ancillary data. In the first step we rasterize blocks' boundaries shapefiles to 30 m grids. In the second step we determine relationship between land cover categories and population density. Representative population density for each class is established using a set of blocks (selected from the entire US) having relatively homogenous land cover (90% for urban class and 95% for vegetation class). In the third step the population in each block is redistributed to its constituent cells using block-specific weights assigned to the cells having different ancillary category. The weights are assigned based on two factors: relative density of population for each land cover category, and the area of each block occupied by each category. Once the weights are calculated the population in each cell is obtained by multiplying block's aggregated number of people by an appropriate weight. Note that cell's values are population densities in units of people/(area of the cell); they can be fractional, and occasionally quite small. Integrating cells' values over the entire block recovers the aggregated number of people.

The major challenge to calculating 30 m dasymetric model of population density for the entire CONUS is the size of input and output data – we need to disaggregate ~11 million blocks into ~8 billion cells. Traditionally, dasymetric modeling was computed in a GIS environment (ESRI ArcGIS). However, for the model of this size such approach is computationally inefficient. Instead, we implemented our calculations in R (scripts are available at http://sil.uc.edu).

R is a comprehensive computational environment that includes libraries (sp, rgrass7, raster, rgdal) to work with geospatial data as well as libraries to work with standard relational databases (DBI, RSQLite). Data is first preprocessed in GRASS GIS before R is used for actual dasymetric modeling. To manage data storage requirements and to better control the time of computation we process each county separately. We used region concept in GRASS GIS for computationally efficient division of U.S. into separate counties. Raster data for each county is read into SpatialGridDataFrame object which integrates information about its spatial content with Census data into a single relational model. The results of dasymetric modeling for each county is saved in two forms: as SpatialGridDataFrame layer (for mapping) and in a tabular form (for possible subsequent analysis). In the last step dasymetric map for each county is exported from SpatialGridDataFrame object into geotiff (library rgdal) and then maps for individual counties are joined into U.S.-wide map using GDAL tools (gdalbuildvrt, gdal_translate).

Main advantages of using R are: (1) less processing steps are required, (2) increased flexibility and automation, (3) no intermediate layers (4) easily expandable to variables other than total population.

3. Results

The results of our modeling are three population grids, for 1990, 2000, and 2010 which can be used to observe and analyze population change. The only practical mode to access maps of this size is through a GeoWeb application. SocScape (Social Landscape), available at http://sil.uc.edu/webapps/socscape_usa, provides an access to our population grids. In addition to population density grids it also includes grids of racial diversity. We used the weights from dasymetric modelling of total population to disaggregate individual racial groups and combined this information into a single product – a diversity map (Dmowska and Stepinski 2014). SocScape provides fast and intuitive exploration of population and racial diversity patterns starting from the continental scale down to the scale of an individual street. The street map is also included in SocScape, thus, by using an opacity tool, a geographical information can be overlaid on population information. With SocScape one can examine spatial dynamics of population density and racial diversity.

SocScape provides a tool for downloading data for the area of interest. Population density as seen in SocScape is classified to 11 categories but download tool access original, unclassified data. For racial diversity grids, which are classified products, the same layer is displayed and downloaded. Downloaded data are in the GeoTiff format and in the WGS84 Web Mercator (EPSG: 3857) projection. In addition to multi-year compatible data SocScape also provides alternative population and diversity maps for 2010; these maps are the result of more comprehensive dasymetric model (Dmowska and Stepinski 2016) which utilizes additional information available only for 2010. All data is also available for download on county by county and metropolitan area basis at http://sil.uc.edu/cms/index.php?id=socscape-data

Figure 1 shows an example of data downloaded from SocScape. In this example, the population grids for the Las Vegas, NV are shown for 1990, 200, and 2010 revealing spatial details of explosive population growth.

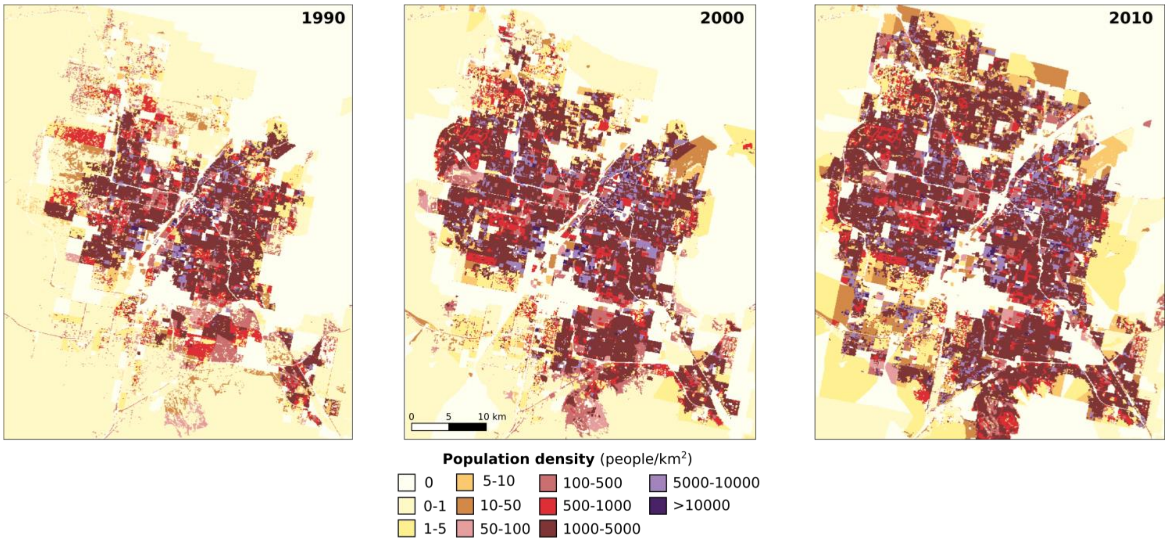

Fig. 1. Population dynamics change for Las Vegas, NV based on hi-res grids for 1990, 2000 and 2010.

4. Conclusion

Our project to provide open and convenient access to hi-res multi-year grids of US population is now completed. The grids are available for exploration and download from the GeoWeb application SocScape. Additional grids pertaining to racial diversity are also available through this application. The new grids are a significant improvement over our first generation grids. They are a valuable resource for anyone who wants to study population and racial dynamics in the US over the last two decades.

Acknowledgements

This work was supported by the University of Cincinnati Space Exploration Institute.

References

Bhaduri B, Bright E, Coleman P, Urban ML, 2007, Land-Scan USA: a high-resolution geospatial and temporal modeling approach for population distribution and dynamics. *GeoJournal,* 69(1-2): 103–117.

Dmowska A, Stepinski TF, 2014, High resolution dasymetric model of US demographics with application to spatial distribution of racial diversity. *Applied Geography*, 53: 417–426.

Dmowska A, Stepinski TF, 2016, A high resolution population grid for the conterminous United States: The 2010 edition. *Computers, Environment and Urban Systems*, in review.

Sperling J, 2012. The tyranny of census geography: Small-area data and neighborhood statistics. *Cityscape*, 14(2), 219–223.

Theobald D. M, 2014, Development and applications of a comprehensive land use classification and map for the US. *PloS One*, 9(4), e94628.