

Methodologies to Improve Global Population Estimates in Urban and Rural Areas

Deborah Balk, Francesca Pozzi, Gregory Yetman, Andy Nelson, and Uwe Deichmann

ABSTRACT

This paper discusses some of the technical issues involved in the development of a georeferenced population database, including urban extents, currently being developed by CIESIN and partner organizations, using ArcGIS. The database will be made available as three separate data products: i) human settlements database (points), ii) urban extents (polygons) derived from satellite imagery and additional geographic data sources (such as DCW), and iii) urban-rural surface (grid). To produce the urban-rural surfaces, we developed a mass-conserving algorithm that reallocates people into urban areas, within each administrative unit, providing a more accurate representation of the distribution of human population.

1. INTRODUCTION

The growth of cities is an intrinsically spatial issue. By 2007, it is estimated that over half of the world's population will reside in urban areas (UN, 2002). Historically, urban residence has been associated with increased mortality (Wood, 2003), although recent urbanization is associated with greater well-being. Despite increasing knowledge about the characteristics of urbanization, little is known about its spatial dimensions. For example, what share of the world's population live in places of 20,000 persons or more? Or, what share of the world's inhabited land area is urbanized? The growth of cities is an intrinsically spatial issue: (1) how does the area of a city change as city population size changes?

Although there is ample research on urban growth as separate geographic and demographic phenomenon, there is little research or dataset in which these parameters are integrated (even though they are implicitly integrated: population growth is a problem in part because it forces geographic expansion or geographic growth occurs because population changes). A better quantification of these processes, and empirically linking them, would go a long way to more thorough analysis of socioeconomic, demographic, and environmental change in urban areas. This study proposes a new methodology to foster this integration.

The first attempt to generate a consistent global georeferenced population dataset was the Gridded Population of the World (GPW, CIESIN et al, 2003), produced at the National Center for Geographic Information Analysis (NCGIA) in 1995 (Tobler et al, 1995), and updated by CIESIN in 2000. The study presented in this paper forwards a new methodology to extend those efforts. The study has been undertaken by Columbia University's Earth Institute's Center for International Earth Science Information Network (CIESIN), in collaboration with the International Food and Policy Research Institute (IFPRI), the World Bank, and the International Center for Tropical Agriculture (CIAT). It is the first systematic effort to produce a global population dataset that also incorporates urban extents. This paper describes the data sources and the methodology followed to produce those urban extents and the urban-rural population surface. It will also present estimates from the newly created database, compare such estimates to UN ones, and make estimates of a spatial nature far beyond those which could be calculated from UN estimates alone.

2. DATA SOURCES

To create the human settlement dataset we used information on population size and spatial extent of urban areas, as described below.

a Population

Population data were gathered primarily from official statistical offices (census data) and secondarily from other web sources, such as Gazetteer (www.gazetteer.de) and CityPop (www.citypop.de), or from specific individual databases when official statistical databases were not available. Based on the data available and applying UN growth rates, we estimated population in 1990, 1995, and 2000. In some cases, the records for cities and town included latitude and longitude coordinates. For those where coordinates were not available, we matched the settlement name and administrative units with the National Imagery and Mapping Agency (NIMA) database of populated places (gnswww.nima.mil/geonames/GNS/index.jsp).

The resulting database constitutes what we will call “points”.

b Settlements extent

The physical extent of settlements has been derived both from raster and vector datasets, in particular:

- **Night-time lights**, produced using time series data from the Defense Meteorological Satellite Program (DMSP) Operational Linescan System (OLS) for the period 1 October 1994 to 30 April 1995, where the pixel values are measurements of the frequency with which lights were observed normalized by the total number of cloud-free observations. To delineate the physical extent of human settlements we used the World Stable Lights dataset (“cities” component).
- **Digital Chart of the World (DCW)’s Populated Places**: an ESRI product originally developed for the US Defense Mapping Agency (DMA) using DMA data and currently available at 1:1,000,000 scale (1993 version). The “populated places” coverage is available for most countries and contains depictions of the urbanized areas (built-up areas) of the world that are represented as polygons at 1:1,000,000 scale.
- **Tactical Pilotage Charts (TPC)**: standard charts produced by the Australian Defense Imagery and Geospatial Organization, at a scale of 1:500,000, originally designed to provide an intermediate scale translation of cultural and terrain features for pilots/navigators flying at very low altitudes. Each chart contains information on cultural, drainage/hydrography, relief, distinctive vegetation, roads, sand ridges, power lines, and topographical features. Settlements are reported both as polygons and points. Polygons and points were digitized for a number of countries, especially where night-lights and DCW data did not adequately delimit urban areas.

All the sources of urban extent (night-lights, Landsat-derived polygons, DCW and TPCs) were combined in order to obtain the maximum possible coverage for each country. The methodology followed to combine the different sources was to use the night-time lights as baseline (due to its global coverage), and then add polygons from other sources that did not intersect any existing light. Therefore the total number of urban polygons in each country is the number of lighted areas plus all the other polygons identifying settlements that were not already identified by the lights. In this case the resulting database will be called “polygons”. These polygons do not have attribute data, such as population, associated with them, but are characterized only by the basic geographic attributes, such as area and perimeter.

3. METHODOLOGY

To create the human settlement database from all the different sources we developed a hierarchical process, as follows:

Assign population from the points to the settlement extents, based on a 3 km buffer distance. If multiple points were present, as in the case of an urban agglomeration, the sum of the population of all points was assigned to the polygon. The name of the most populous city within the buffer was also assigned to the polygons. Polygons with no points in the 3 km buffer were not assigned a population value. These polygons were removed at a later stage.

Estimate areal extents for points without polygons, based on a relationship between population size and areal extents for the points with known parameters. Using a logarithmic regression, we predicted the expected geographic size of a place, given its population size. Based on these area values, we created circles, centered on the points. Intersecting circles were merged, and circles intersecting existing polygons, or having more than half its area outside the country boundaries were slightly moved to a more appropriate location. We then added these newly created polygons to the existing ones to create a complete urban extent coverage for each country. This coverage is defined as Human Settlements Mask.

Create the urban-rural surface, by developing a mass-conserving algorithm, GRUMP (Global Rural Urban Mapping Programme) that reallocates people into urban areas, within each administrative unit. In particular we used data inputs from two vector sources:

- A map of administrative polygons, containing the total population for each admin unit.
- A map of urban areas, containing the urban population for each area.

Figure 1 shows the data required for GRUMP in the case of Ghana.

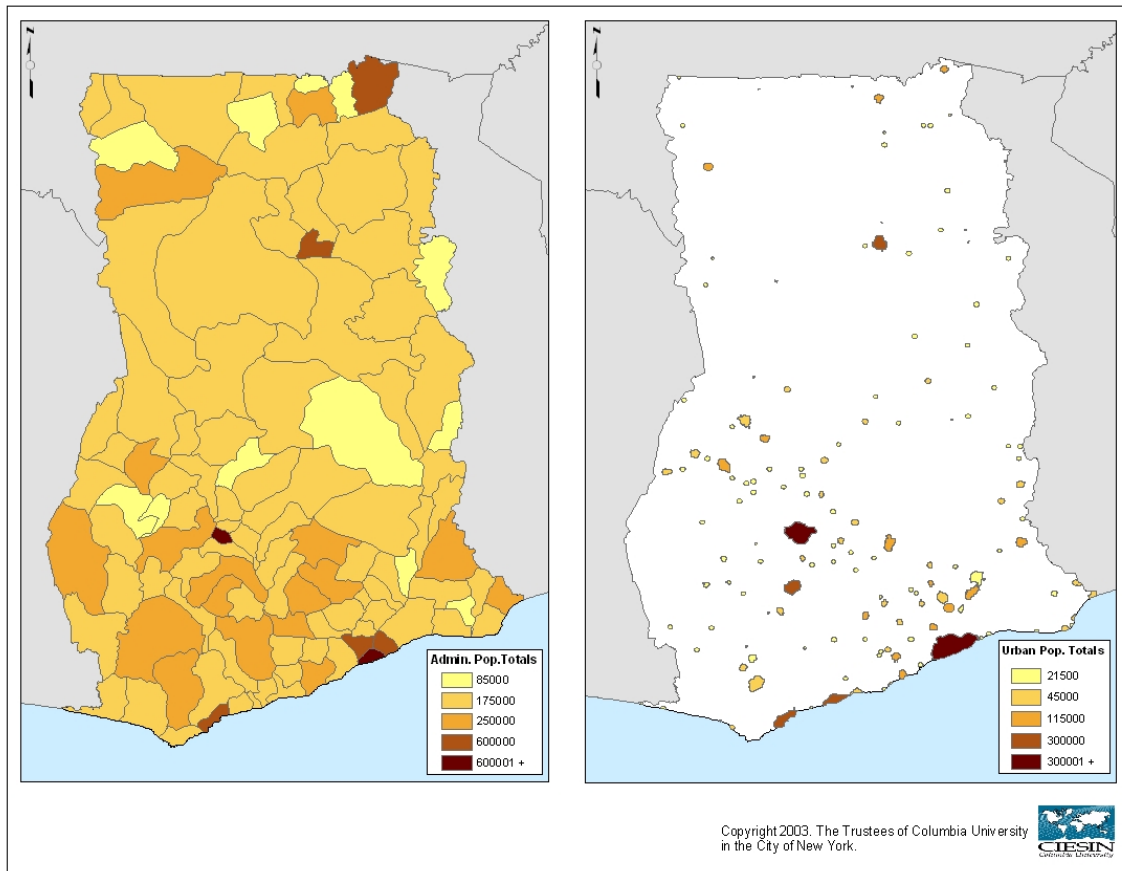


Figure 1 Population data required for GRUMP: admin totals (l) and urban areas (r).

These two data sets are combined using the ArcInfo command IDENTIFY. The new dataset is then passed to GRUMP (a stand alone model written in C) that assigns population to each new polygon and labels it as rural or urban.

The calculation of the reallocation of people within urban areas is based on the assumption that the total population remains constant within a given administrative unit but the rural and urban distributions change while these conditions are met:

- No polygon has zero population unless the input population was also zero
- Urban density is less than an upper bound derived from the country average urban density
- Rural density is less than an upper bound derived from the country average rural density, and less than the current urban area density (if applicable)
- Rural density is greater than a lower bound derived from the country average rural density

These four conditions are controlled by parameters that are passed to the AML on the command line. If no parameters are specified then the AML will assign fixed values that have been empirically determined to be good first estimates.

The resulting map is shown below (Figure 2), with a close up of Accra showing the data before and after running GRUMP.

The final results from each country are recorded in an excel spreadsheet to compare the output rural and urban population totals to the UN totals.

The output coverage from GRUMP is then converted to a grid, at 30 arc-seconds resolution.

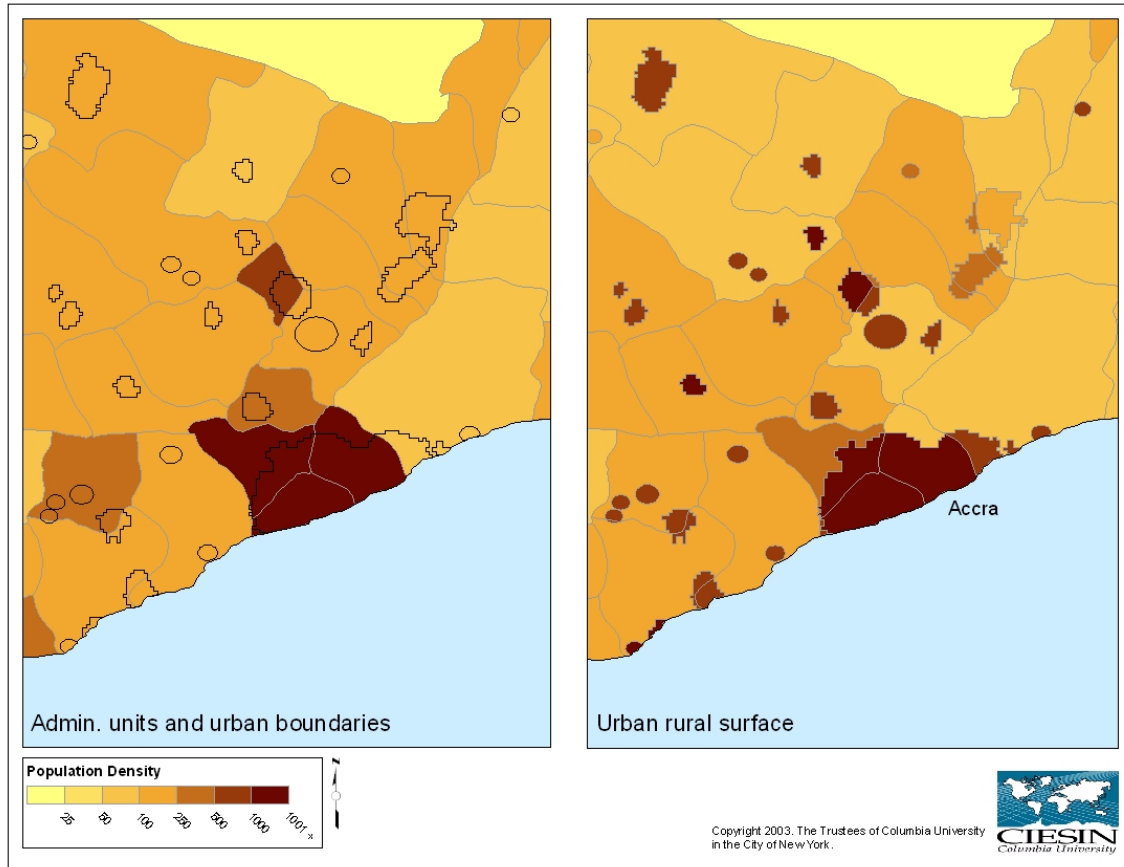


Figure 2. Close-up view of Accra and the southern portion of Ghana, showing population density of the original administrative units along with the human settlement mask (l) and the population density resulting from the reallocation process (r).

4. RESULTS

As shown in the Ghana maps above, the population distribution of Ghana is more concentrated than estimated by the GPW approach where population values are averaged over the administrative areas. For large cities, such as Accra, the administrative units tend to be smaller such that identification of the more populated localities was possible even prior to this new methodology. Nonetheless, even at the edges of large cities, where human settlements tend to have peri-urban or peri-rural characteristics, this methodology allows for a significant delineation of the urban extent and thus a reallocation of population within the administrative area. For the vast majority of human settlements, however, the administrative area estimates alone produce estimates of population density that may lead to mis-estimation of the land area occupied by densely populated settlements. In Ghana, for example, GPW would estimate that only 1 percent of the land area is inhabited at a density of 500 persons/km² or more, whereas the urban-rural surface suggests that approximately twice that amount of land area is inhabited at this density.

These differences may be important to our understanding of urban processes including land use and ecosystem change.

5. REFERENCES

Center for International Earth Science Information Network (CIESIN), Columbia University. 2003. *Gridded Population of the World, version 3 (beta)*. (Data available at <http://beta.sedac.ciesin.columbia.edu/gpw/>)

Tobler, W, Deichmann, U., Gottsegen, J. and Maloy, K. (1995). The Global Demography Project, Technical Report TR-95-6, National Center for Geographic Information and Analysis, Santa Barbara.

Woods, Robert. 2003. Urban-Rural Mortality Differentials: An Unresolved Debate. *Population and Development Review*, March 2003, 29(1), pp. 29-46.