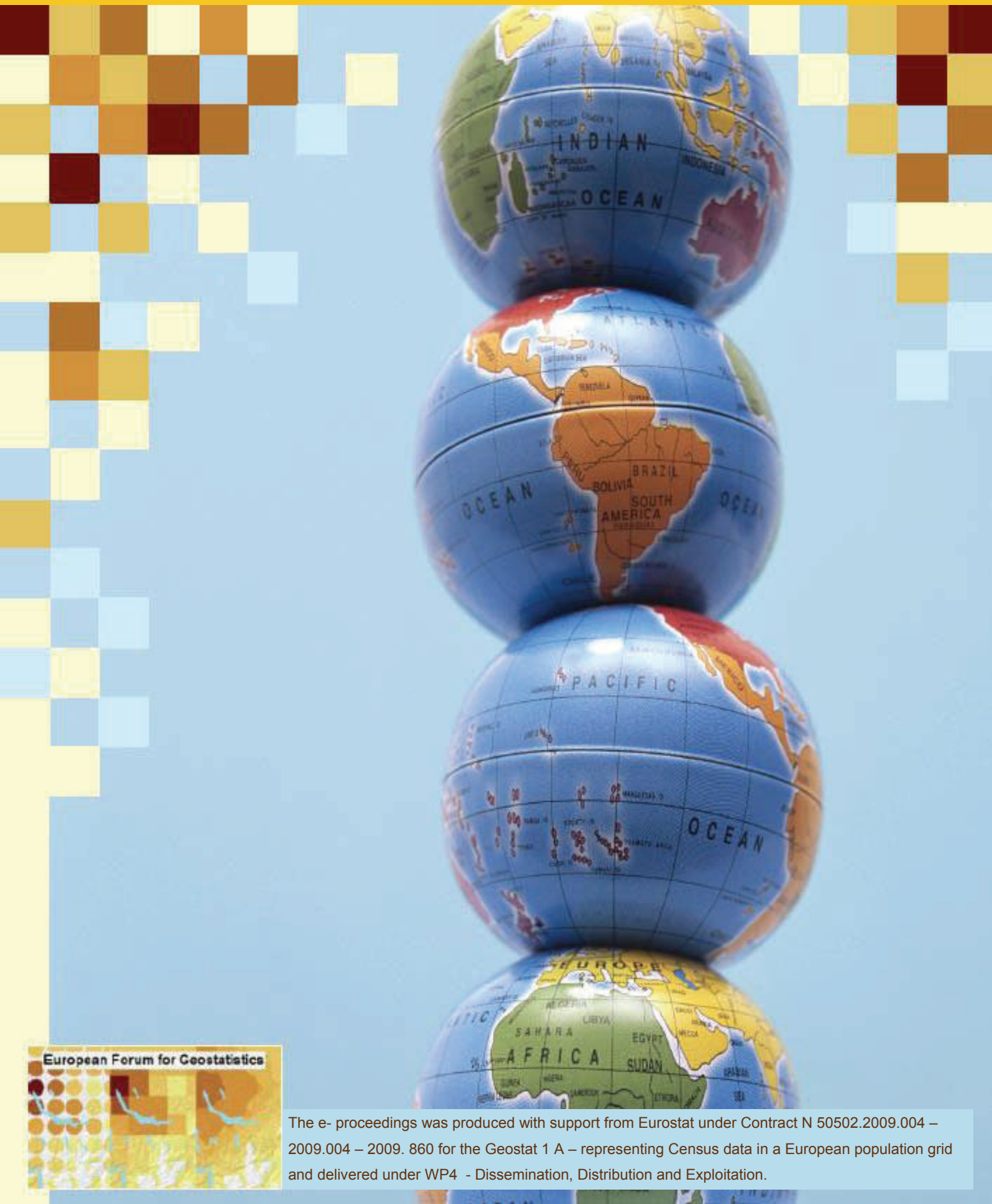


# e – Proceedings of European Forum for Geostatistics Conference

5 - 7 October, 2010 Tallinn, Estonia

[www.efgs.info](http://www.efgs.info)



The e- proceedings was produced with support from Eurostat under Contract N 50502.2009.004 – 2009.004 – 2009. 860 for the Geostat 1 A – representing Census data in a European population grid and delivered under WP4 - Dissemination, Distribution and Exploitation.

## Construction of Gridded Population and Poverty Data Sets from Different Data Sources

Deborah Balk\*, Gregory Yetman\*\* and Alex de Sherbinin\*\*

\*Baruch College, City University of New York,\*\*CIESIN, The Earth Institute at Columbia University

The Center for International Earth Science Information Network (CIESIN) was, together with partners at the National Center for Geographic Information Analysis (NCGIA), the first group to produce and distribute gridded census products. CIESIN and NCGIA released Gridded Population of the World (GPW) v.1 in 1995. Since that date, CIESIN has produced a large number of gridded socioeconomic data products, including two updates to GPW (versions 2 and 3), a 1km resolution urban-rural grid (GRUMP), US census grids, global infant mortality and malnutrition grids, and gross domestic product grids. The baseline gridding is done using a proportional allocation algorithm on the highest spatial resolution census or survey data available, and for GRUMP, population are reallocated based on urban-rural boundaries as defined by night time lights and estimates of settlement population size. This paper describes the products and the methods used to produce them.

### Introduction

Global or broad-scale inquiry on the relationship between population and the environment is intrinsically spatial, however, much of the analysis occurs in a spatial vacuum. While notable exceptions exist, especially at the local scale two key barriers have contributed to the lack of spatially-oriented analysis: (1) the methods of analysis require some knowledge of geographic data and tools for analysis; and (2) population data, at a global scale, tend to be recorded in national units rather than those that would permit cross-national, sub-national analysis. These barriers have been slowly eroding. On the demand side, demographers are becoming more familiar with geographic constructs, data and technology (and the technologies are becoming more relevant - e.g., in terms of spatial analysis - to demographers). On the supply side, data and tools are becoming increasing available. This paper describes recent developments in rendering global population and poverty data at the scale and extent require to facility broad-scale human-environment inquiry. It first describes the Gridded Population of the World (GPW), then turns to the Global Rural - Urban Mapping Project (GRUMP), US Census Grids, and two poverty mapping products: infant mortality rates and child malnutrition.

Nearly twenty years have passed since the first efforts to render population data, primarily from censuses, on a latitude-longitude grid on a global scale (Tobler et al., 1997; Clark and Rind, 1992). In those years, several key advances have been made: The spatial resolution of administrative boundary data is improving; national statistical offices and spatial data providers and related institutions are becoming more open with their data; population and spatial data providers are increasingly

aware of (or collaborate with) one another; and lastly, computing capacity to manage, manipulate, and process increasingly large data sets is continually expanding. This bodes well for the future of gridding.

Gridded data are useful for a range of application areas related to human-environment interactions, natural hazards, and environmental health (see for example Balk et al. 2006 and 2005, de Sherbinin 2009, Dilley et al. 2005), and CIESIN has collected more than 400 citations to journal articles citing GPW alone. But the focus of this paper is on construction of these data sets.

### Gridded Population of the World (GPW)

The basic methods, developed for GPW v1 (Tobler et al., 1997) and modified slightly for GPW v2 (Deichmann et al., 2001), remained more or less the same in the development of GPW v3. Population data are transformed from their native spatial units, which are usually administrative and of varying resolutions, to a global grid of quadrilateral latitude-longitude cells at a resolution of 2.5 arc minutes, which equates to approximately 4km on a side at the equator (Figure 1). Slight modifications have been made to the processing, and the increases in input resolution have meant that GPW v3 relied more heavily of interpolations of population data that rely on spatial hybrids (e.g., growth rates between states in 1990 and 2000 are applied to the spatial distribution of population in municipalities in the year 2000).

The steps used to develop GPW include the following:

1. Find tabular population counts
2. Match these to geographic boundaries (census or administrative units)
3. Estimate the population for target years (e.g. 1990, 1995 and 2000)
4. Transform to grids

Depending on the country, steps 1 and 2 often requires substantial effort to reconcile tabular counts with boundaries that may have been produced by a third party. The matching can be especially difficult where new census or administrative units are created from subdivisions of units from earlier time periods. Any assumptions that went into the assignment of population counts into subdivided units are specified on the individual country pages of the GPW web site (go to "Data and Maps" and click on the country name to access these pages).

Figure 1: Transforming census units to a 2.5min grid



As for Step 3, estimating the population to target years, the method is relatively straightforward.

Annual rates of change are calculated as follows:

$$r = \frac{\log e \left( \frac{P_2}{P_1} \right)}{t}$$

Population estimates are adjusted to target years:

$$P_x = P_2 e^{rt}$$

Adjustment factors for matching national estimates to UN estimates are calculated as follows:

$$a = (P_{un} - P_x) / P_{un}$$

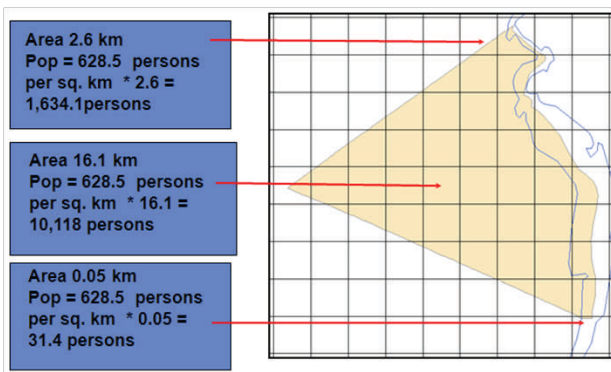
Adjustment factors are applied at the national level :

$$P_{adj} = P_x * a$$

Where  $r$  is the annual rate of growth,  $P_{1,2}$  is the census estimate,  $t$  is the number of years between census enumerations,  $P_x$  is the year of estimate,  $P_{un}$  is the UN Estimate,  $P_{adj}$  is the adjusted estimate, and  $a$  is the adjustment factor.

The last step (Step 4) entails a proportional allocation algorithm, as illustrated in Figure 2. Assuming the administrative unit has an average population density of 628.5 persons per sq. km, each grid square is allocated a population count from that unit proportional to the area of the unit located in that grid cell. If the grid cell covers another unit, then the population in proportion to the area covered in that unit is added to the grid cell.

Figure 2: The proportional allocation algorithm



GPW is an effort to amass information on the distribution of human population without modelling. However, there are many good reasons for modelling. For example, census data typically represent a decennial, residential picture of population distribution. It does not indicate daytime or seasonal distribution, non-residential patterns such as transportation zones, or built-up industrial and commercial areas. Another reason for modelling is that GPW's accuracy is closely related to that of the accuracy of census data. If these data are old (i.e., no new census in many years), coarse (national or coarse-level only), or believed to otherwise be of poor quality, additional information may be very useful in estimating the distribution of human population. Thus, over the past decade, many efforts have focused on efforts to model population distribution. These have ranged from lightly modelled approaches, with urban areas (addressed below; CIESIN et al., 2004) or roads (UNEP et al., 2001) or heavily modelled with these and other inputs to reallocation population (e.g., LandScan, see Dobson et al., 2000). We argue that these modelled datasets are complementary to GPW's heuristic method.

Over time, the greatest investment has been made in increasing the number of input units. Table 1 describes the number of input units by continent for GPW v3. In 1994, the first GPW database was developed using about 19,000 units, and rendered at an output resolution of 5 minutes; whereas the second version had nearly 120,000 input units, about half of which were due to the inclusion of tract-level data for the United States. The third version has over 375,000 inputs units, with no improvement to the resolution of the inputs for the United States (although higher resolution data are available), but substantial improvements for other countries including both geographically large and small entities: South Africa (80,000), Indonesia (60,000), France (36,000), Malawi (9,000) and Brazil (5,500). These along with the U.S., account for 70% of the units in the database, 17% of the global land area and roughly 13% of the population.

**Table 1:** Summary information on input units for GPW v3, by continent

Continent	Modal Level*	Total Number of Units	Average Resolution	Average Persons per Unit
Africa	2	109,138	73	166
Asia	2	88,782	53	276
Europe	2	91,086	25	112
North America	2	74,421	29	83
Oceania	1	2,153	25	27
South America	2	10,919	68	49
Global	2	376,499	46	144

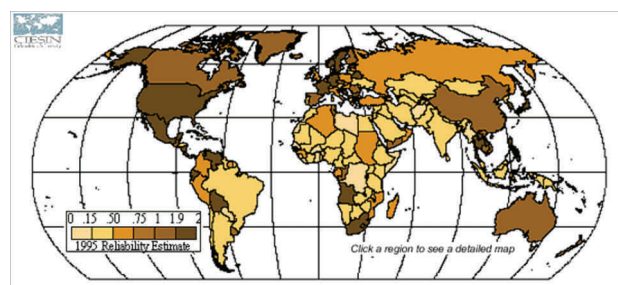
Earlier versions of GPW had less motivation to gather higher resolution inputs because the output resolution of 2.5 minutes rendered finer input resolution redundant. GPW v3, however, was also used as an input to the Global Rural Urban Mapping Project (GRUMP) population grid (see below) that includes reallocations towards urban area and whose output resolution is 30 arc seconds; at this resolution, the effort to find higher resolution spatial inputs was justified. Often, these new inputs had to be heads-up digitized, since digital versions of these data were not available. For countries that are comprised of island chains, the improvements consisted of collecting island-level population data, and then assigning population to existing spatial inputs. GPW v2 had 41 level-0 countries, 31 of which were islands, which had an average resolution of 46. In version 3, fewer than half of these countries remain (with a slightly smaller share of them being islands) with an average resolution of 22.

The ideal resolution for GPW administrative units is somewhere close to the size of a few grid cells (i.e., for a 2.5 arc-minute cell at the equator, this would be an administrative unit area of 85 square km). For GRUMP, which has a resolution of 30 arc-seconds, the ideal administrative unit would have an area of only 4 square km (CIESIN et. al., 2004). Where high-level boundary data (level 4 or greater) are available, the area of administrative units in densely populated areas exceeds the GPW ideal resolution and, in some areas, even that of GRUMP. In low-density areas, even where the highest-level boundary data are available, the administrative units are much larger than these ideal sizes. However, administrative units this detailed over sparsely inhabited regions would be inefficient to process (they would comprise over 2 million units for GPW), they would add little or no additional information to the distribution of population, and they would be infeasible to maintain.

In terms of temporal resolution, GPW v3 provides estimates for 1990, 1995, 2000, 2005, and 2015. Most countries of the world have now experienced two census in their recent history and with the exception of Africa and some parts of the middle East, West Asia and East Europe, most countries have had a census taken recently, since or in the year 2000. Figure 3 describes the relative reliability for 1995, with countries in darker col-

ours having more recent censuses and those in lighter colours having older data only.

**Figure 3:** Recency of data for GPW v3



When higher resolution data become available, often the associated population are only available for a single (recent) time period, although in some exceptional cases population (e.g., France) estimates are given for a range of dates. It is not uncommon for the relevant statistical offices to not know how the current thematic population map matches to one from a prior time period. Thus, much of the work of preparing this database is to reconcile such differences in geographies resulting from temporal change. Aside from war torn countries, which often to lack current data altogether, countries undergoing periodic and medium to large-scale political or administrative reorganization pose the greatest challenge. This is a more general issue, however, because it is a normal part of geographic and administrative change, and it tends to occur most commonly at a fine-scale (i.e., state boundaries change much less frequently than higher-resolution boundaries like *municipios* or counties). To the extent future efforts to amass data at the current scale are undertaken, it will persist.

In terms of methodological advances, all information is couched on correspondence between geographic units, which means if there were large changes in spatial units (e.g., Namibia or the former Soviet Republics) that some of the spatial specificity of population change over time may be lost. For example, boundaries in 2001 that differ from most of those for in 1991 require construction of artificial regions to generate growth rates to interpolate and extrapolate to the target years. Transformations of this nature are clearly documented on the GPW web site on a country-by-country basis. Although we create a correspondence between the two geographies (where



available) for interpolating population values to target years, we only use one year of boundary data for creating the population grids. In this manner, the best spatial resolution can be retained while incorporating sub-national population change information via the correspondence. In cases where the two geographies are at the same level (e.g., Canada and the United States), only the most recent geography is used for gridding. This reduces the labour in preparing the data and the amount of processing time required for gridding.

Because countries vary between each other and internally on the size of the administrative areas, analysis of the data may benefit from more information about the administrative area underlying each unit in the output grid. Thus, for GPW version 3 we constructed a population-weighted administrative unit area layer. This layer allows the determination, on a pixel-by-pixel basis, of the mean administrative unit area that was used as an input for the population count and density grids. For grid cells (pixels) that are wholly comprised of one input unit, the output value is the total area of the input unit. Where grid cells are comprised of multiple input units, the output value is the population-weighted mean of all of the inputs.

There have also been improvements in production methods. Quality in production has become more standardized, thus allowing for the identification of anomalies and errors introduced in processing.

There are several barriers that limited improvements for GPW v3. Most of the former Soviet republics underwent redistricting in the 1990s, but few of them make their spatial data available, either freely or for a fee. Recently war-torn countries take a while to implement new censuses, although they may be the places most susceptible to population movements. In some instances, official population data are available while official boundary information are not. In such instances, if unofficial boundary information is available (e.g., Bosnia Herzegovina) is incorporated, if at all possible.

Several countries were just too expensive to purchase census or spatial data. Many of the former British colonies sell licenses to use their fine-resolution census data rather than release it freely. This meant that it would have cost thousands of dollars to update Australia and New Zealand at the level that we had undertaken for GPW v2. Because the last reference year for population data for version 2 were in 1996 at high resolution for these countries, they were updated at a coarser resolution—using the hybrid method described above—for which the data were publicly available.

### Global Rural Urban Mapping Project (GRUMP)

There are actually three products included in GRUMP – an urban settlements points data set of more than 70,000 unites with populations >1,000 persons, an urban extents “mask” of more than 27,500 urban areas

with populations >5,000 persons, and a population grid with urban reallocation at 30 arc-second resolution. It is the latter that is the focus of this paper. The GRUMP population grid is a 30-arc second population distribution raster dataset that was developed by combining population data from the census administrative units and from the urban extent mask. To create the population surface, we developed a mass - conserving algorithm called GRUMPe (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: (1) Administrative polygons, containing the total population for each administrative unit; (2) Urban areas, containing the urban population for each area.

These two data sets are combined in such a way that an intermediate (polygon) data set representing the urban and rural areas, but which does not assign populations into those areas, is produced. This intermediate dataset is then passed to GRUMPe, a stand-alone model written in C, that assigns population to each new polygon and labels it as rural or urban. Typically, the algorithm works on a country-by-country basis and uses the following pieces of information: The size and population of each urban area, denoted by a unique urban area identifier, the size and population of each administrative area denoted by a unique administrative identifier, the size of the intersect areas where the urban and administrative areas overlap, and the UN national estimates for the percentage of the population in urban and rural areas (UN, 2002).

The goal of the algorithm is to reallocate the total population in each administrative unit into rural and urban areas while reflecting the UN national rural-urban percentage estimates closely as possible. The algorithm was designed to have few constraints and to make the constraints simple and reflect common sense. There are 6 constraints in total: (1) The total population (urban + rural) within any given administrative units remains constant; (2) The urban population density in any given administrative unit must be greater than the rural population density in that administrative unit; (3) The rural population density in any given administrative unit cannot be lower than a national minimum rural population density threshold; (4) The rural population density in any given administrative unit cannot be greater than a national maximum rural population density threshold; (5) The urban population density in any given administrative unit cannot not be greater than a national maximum urban population density threshold; (6) The urban population density in any given administrative unit cannot not be lower than a national minimum urban population density threshold.

The algorithm works on each administrative unit in turn, and checks the urban and rural populations within that administrative unit against constraints 2 to 5. If any of the constraints are not met, then the rural and/or urban populations are adjusted iteratively to meet them while ensuring that constraint 1 is met. These constraints

and the national population density thresholds are controlled by parameters that are passed to the algorithm. If no parameters are specified then the algorithm will assign fixed values that have been empirically determined to be good first estimates.

The adjustment in population is trivial when there are no or one urban area per administrative unit, and where the urban area lies wholly within the administrative unit. It becomes increasingly complex however when there are more than one urban area, and urban areas overlap more than one administrative area (e.g., Cali, Colombia), and large urban areas contain more than one administrative area (e.g., Quito, Ecuador). All of these are common situations, and may require successive iterations to meet all the constraints. The algorithm can also be run on a region-by-region basis (such as states or other first-level administrative units), such that the national constraints (3 to 6) now become regional constraints and will better reflect the state-level variation in rural/urban population percentages in large countries like the USA. This approach was employed for most of the largest countries or countries with very large numbers of administrative units (e.g., South Africa).

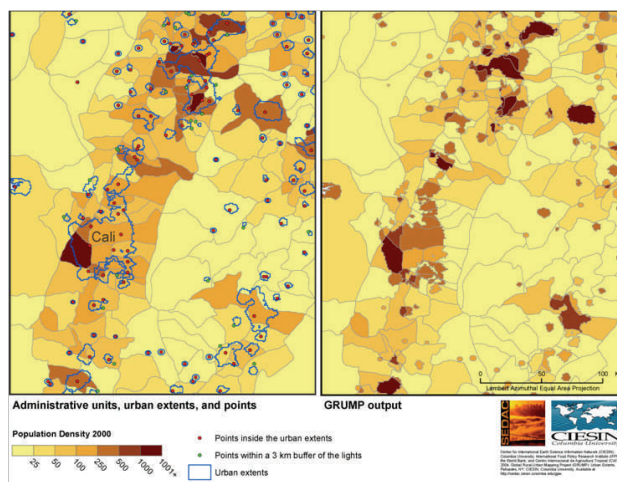
The resulting map, Figure 4 - a close-up of Cali, Colombia - shows the data before and after running GRUMPe. Note how, where urban areas are present in a given administrative unit, the density of the GPW administrative units decreases after GRUMPe because people are reallocated into their respective urban areas. The final results from each country are then compared to the UN urban population estimates. Although the UN totals are useful as a benchmark, they are only that. Not only have recent studies shown the uncertainty associated with UN urban estimates (NRC, 2003), there are many reasons why our estimates may differ considerably from that of the UN's. For example, our data stream may have included many more small settlements, including those below the urban threshold either given by the country, or implied by the region, in which case we would expect the comparison between percentages of the population living in urban areas to be quite different between the two. We estimate that in X% of the countries, we had a priori reasons to expect much different outcomes from the UN estimates (mostly but not always for the better), and in another Y% for them to match rather closely because our data streams matched closely those which they also report. In the remainder of the countries, we had no information either way to predict the closely to those estimates.

The final stage is to convert the output coverage from GRUMPe into a grid, at 30 arc-seconds resolution.

### US Census Grids

The U.S. Census Grids are created by taking population and housing counts at the block level and proportionally allocating the counts in the census blocks to a latitude-longitude quadrilateral grid (Seirup et al. 2006).

**Figure 4:** Demonstration of the urban reallocation for an area near Cali, Colombia



Grid cell size for the country is 30 arc-seconds (approximately 1km), and for metropolitan areas it is 7.5 arc-seconds (250 meters). It uses the same proportional allocation algorithm as GPW v.3. If a grid cell contains 40% of the area of one census block and 30% of the area of a second census block, the population count for that grid cell will be 40% of the population of the first census block and 30% of the population of the second census block.

U.S. Census Grids, 2000 uses the 2000 TIGER/Line files for the census block boundaries and 2000 SF1 and SF3 tables for the demographic and socioeconomic characteristics of each census block. SF1 data are based on the census short form and therefore include counts for the total population. The SF3 data are based on the census long form, which is sent to approximately one out of every six households.

For the 30 arc-second grids using SF1 data, the relevant fields are extracted from the SF1 tables. Some of the grids contain data from a single field, such as the number of non-Hispanic whites, which is taken from SF1 table P8, field P008003. Other grids use data from several fields. This is true of all the age grids, which are derived from SF1 tables P12 and P14. These tables report the number of people in each age category by gender. The grid for the population under age one uses SF1 table P14, field P14003 (the number of males under age one) plus SF1 table P14, field P14024 (the number of females under age one). The Variable Catalog contains a list of the fields used for each census grid.

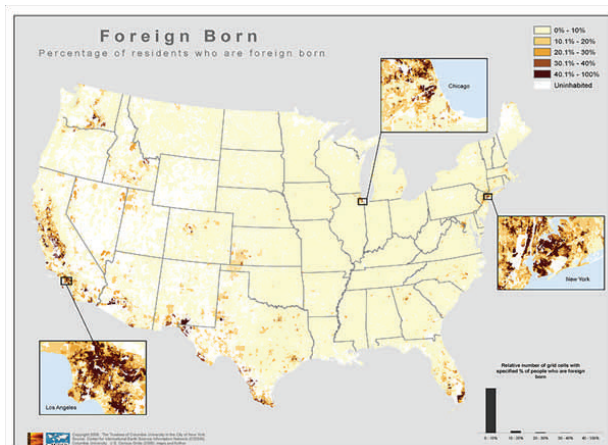
The TIGER/Line files are converted to ArcInfo coverages and joined to the SF1 field variables. The density of the variable being gridded is calculated for each census block, for example the number of foreign born residents per square kilometer. A 30 arc-second quadrilateral grid is intersected with the census block coverage. This divides each census block into pieces that fit into the grid cells. The total count for the grid cell is calculated by taking the area of each census block piece within the grid cell, multiplying it by the density of the variable

being gridded, and summing these values for all census block pieces in the grid cell.

The lowest level of geography for which SF3 data are released is the census block group. For the SF3 grids, these data are proportionately allocated to census blocks using the distribution of the underlying SF1 population. For instance, if 35% of the block group's population aged 25 and older lives in a given census block, as reported in the SF1 tables, 35% of the block group's population aged 25 and older with a high school diploma, as reported in the SF3 tables, is assigned to that census block. Once the SF3 data have been allocated to the census block level, the gridding process is the same as described above.

The metropolitan statistical area (MSA) grids are created by selecting all census blocks within the MSA and gridding those blocks using a 7.5 arc-second quadrilateral grid. An example of a US Grid map is found in Figure 5.

**Figure 5:** Foreign born population, 2000



## Poverty Mapping

CIESIN's poverty mapping work was developed in connection with the Millennium Development Project (MDP), a research project led by Jeffrey Sachs at the Earth Institute at Columbia University that was designed to help inform governments on how to best meet the Millennium Development Goals (MDGs). To assist the MDP poverty and hunger task forces, CIESIN developed two datasets: a global map of infant mortality rates (Figure 6) (Storeygard et al. 2008), and a global map of child malnutrition (Figure 7).

Infant mortality rates (IMRs) were chosen because they serve as a useful proxy for overall poverty levels and are highly correlated with metrics such as income, education levels, and health status of the population (Balk et al., 2006). This metric is particularly good for distinguishing poverty levels at the lower end of the income ladder.

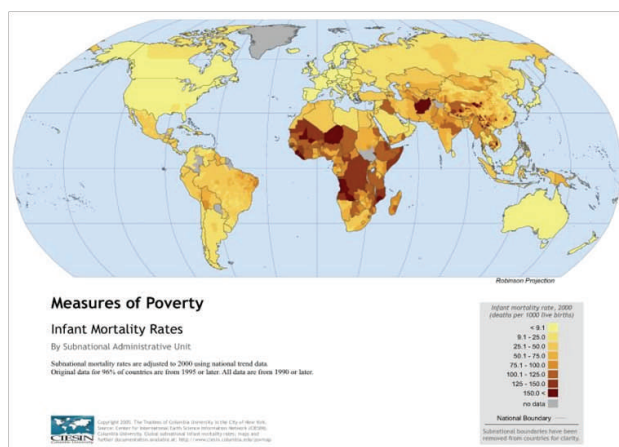
Unlike the Census gridding, the sources of data for infant mortality rates were largely survey data. Sample sizes dictated that the sub-national units are generally larger in size, since results can only be reported at the geographic scale at which they are still robust. The sources for the IMR map include Demographic and Health Surveys (DHS) (39 countries), Multiple Indicator Cluster Surveys (MICS) (5 countries), National Human Development Reports (14 countries), and National Statistical Offices (18 countries). There are only 6,494 spatial units in the global data base, 82 percent of which are in Brazil and Mexico (5,372 units). There are 74 other countries with subnational data, with an average of 22 subnational units per country. Finally, for 115 countries we only had national level data from UNICEF, and 36 countries had no data.

For each country the subnational IMR values were adjusted to be consistent with national UNICEF 2000 IMR values. The data were gridded using the same proportional allocation algorithm as the population grids. Note that the grids are actually at a higher spatial resolution than the shape files because for some countries subnational administrative boundaries could not be distributed. We also converted rates to counts of infant deaths. To do this, for each subnational unit we estimated live births and infant deaths. These were calculated based on *gridded* population, *national* fertility data, and *subnational* IMR data.

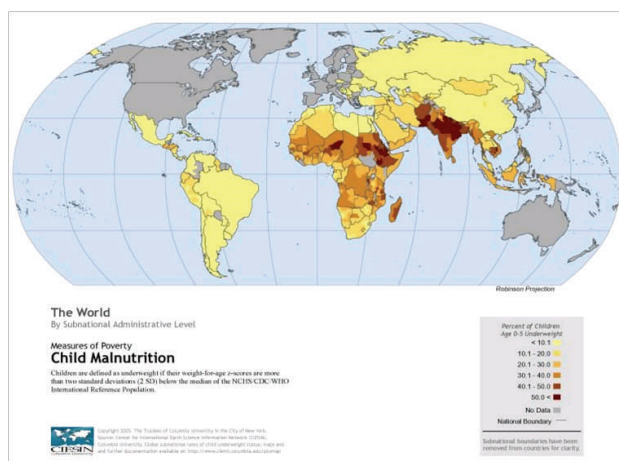
For child malnutrition, we used anthropometric data found in household surveys. The metric chosen is percent of children underweight, with underweight defined as being two standard deviations or more below the mean weight for a given age when compared to an international reference population. Although there are alternative measures of malnutrition, such as stunting (low height for age) and wasting (low weight for height), the percentage of children underweight was chosen because the MDG Target is to "halve the prevalence of underweight children by 2015."

DHS and MICS data were aggregated to the spatial units at which the surveys report, based on raw data where it was available, and published reports by UNICEF otherwise. There were a total of only 369 spatial units, but it should be mentioned that developed countries were omitted because of very low levels of child malnutrition. These spatial units are typically equivalent to first level administrative regions or aggregations thereof. Geospatial boundary files that match those spatial units were located or created in order to match the reporting regions of the surveys as closely as possible. In many cases, the survey reports contained maps detailing the survey regions. Elsewhere, matches were purely name-based. Note that the dates of the surveys varied and no effort was made to standardize to a specific year (see Table 2).

**Figure 6: Infant mortality rates of the World, 2000**



**Figure 7: Child malnutrition, circa 2000**



**Table 2: Source data for child malnutrition map**

Country	Year	Hunger Source Abbreviation	Units with Hunger Data
Algeria	2000	MICS	4
Angola	2001	MICS	6
Benin	2001	DHS	6
Botswana	2000	MICS	10
Burkina Faso	1998-99	DHS	4
Burundi	2000	MICS	5
Cameroon	1998	DHS	4
Cape Verde	1994	UNICEF	1
Central African Rep.	2000	MICS	17
Chad	1996-97	DHS	1
Comoros	2000	MICS	3
Congo	1998-99	No data	1
Congo, Dem. Rep.	2001	MICS	11
Côte d'Ivoire	1994	DHS	10
Djibouti	1996	No data	1

**Table 2 continues:**

Egypt	2000	DHS	26
Equatorial Guinea	2000	MICS	2
Eritrea	2002	DHS	6
Ethiopia	2000	DHS	11
Gabon	2000	DHS	4
Gambia	2000	MICS	8
Ghana	1998	DHS	10
Guinea	1999	DHS	5
Guinea-Bissau	2000	MICS	9
Kenya	2000	MICS	7
Lesotho	2000	MICS	10
Libya	1995	ANDI	7
Madagascar	1997	DHS	6
Malawi	2000	DHS	3
Mali	2001	DHS	6
Mauritania	2000-01	DHS	5
Mauritius	1995	UNICEF	1
Morocco	1992	DHS	7
Mozambique	1997	DHS	11
Namibia	2000	DHS	13
Niger	2000	MICS	6
Nigeria	1999	DHS	5
Rwanda	2000	MICS	12
Sao Tome and Principe	1996	UNICEF	1
Senegal	2000	MICS	10
Seychelles	No data	UNICEF	1
Sierra Leone	2000	MICS	4
Somalia	2000	MICS	3
South Africa	1995	ANDI	9
Sudan	2000	MICS	16
Swaziland	2000	MICS	4
Togo	1998	DHS	5
Tunisia	2000	MICS	7
Uganda	2000-01	DHS	4
Tanzania	1996	DHS	22
Zambia	2000-01	DHS	9
Zimbabwe	1999	DHS	10



## Conclusion

Efforts to grid socioeconomic data have progressed substantially since the first efforts at gridding census data by CIESIN and the National Center for Geospatial Information and Analysis (NCGIA) in the early 1990s. CIESIN has been at the forefront of these efforts, but it is heartening to see many other groups taking up the cause. The number of gridded census products has increased markedly in the past decade. New thrusts include adding the temporal dimension – both in terms of diurnal and seasonal population movements, and longitudinal data of night – time population (as measured by standard censuses). Additional census and survey variables are also being gridded.

Looking to the near future, for GPW v4 we expect to make the following improvements. There will be a continued emphasis on higher resolution inputs. We will collect and grid more census variables, including age and sex distribution and urban/rural distribution. The proposed output resolution will be 30 arc-second grids. CIESIN may also create a time series back to 1980.

Many barriers to data collection and processing have been overcome since the early versions of GPW to enhance our understanding of population distribution. The role of international technical assistance for population census taking and georeferencing enumerator area maps, has no doubt played an important part in improving spatial accuracy in geolocating populations. Along with these improvements come the possibility of new data streams and integrations, such as using satellite information to detect urban areas along with population information from censuses on human settlements. Such new efforts (see Balk et al., 2004) build strongly on GPW's efforts. Undoubtedly, there will continue to be the need for information at different scales, extents, and resolutions, and that which is simple and that which is modelled. GPW – and its underlying data infrastructure – are critical foundations for future efforts.

CIESIN is always looking for collaborators and data sharing partners to lighten the work load, so feel free to contact the authors if you feel you have something to offer.

## Acknowledgements

This paper draws heavily from earlier papers, including Deichmann, Balk, and Yetman (2001) and Balk et al. (2004), as well as documentation available on the US Census Grids Web site (Seirup et al. 2006). The paper was produced with support from National Aeronautics and Space Administration under Contract NNG08HZ11C for the Continued Operation of the Socioeconomic Data and Applications Center (SEDAC) at CIESIN at Columbia University. For more information visit <http://sedac.ciesin.columbia.edu>

## References

- Balk, Deborah. 2009. "More than a name: Why is Global Urban Population Mapping a GRUMPy proposition?" In P. Gamba and M. Herold, (eds.) *Global Mapping of Human Settlement: Experiences, Data Sets, and Prospects*, (Taylor and Francis): 145-161. <http://www.crcnetbase.com/doi/abs/10.1201/9781420083408-c7>
- Balk, Deborah, Glenn Deane, Marc Levy, Adam Storeygard, and Sonya Ahamed. 2006. *The Biophysical Determinants of Global Poverty: Insights from an Analysis of Spatially Explicit Data*. Paper presented at the 2006 Annual Meeting of the Population Association of America, Los Angeles, USA.
- Balk, Deborah, Adam Storeygard, Marc Levy, et al., 2005. "Child hunger in the developing world: An analysis of environmental and social correlates," *Food Policy*, 30: 5-6 (2005) 584–611. Available at: <http://www.sciencedirect.com/science/article/B6VCB-4HHWWG9-2/2/2f25e9cce26e94fa5b9aff7f6b95db62>
- Balk, Deborah, Francesca Pozzi, Gregory Yetman, Uwe Deichmann, and Andy Nelson. 2004. "Distribution of People and the Dimension of Place: Methodologies to Improve the Global Estimation of Urban Extents," Available at [http://sedac.ciesin.columbia.edu/gpw/docs/UR\\_paper\\_webdraft1.pdf](http://sedac.ciesin.columbia.edu/gpw/docs/UR_paper_webdraft1.pdf)
- Balk, Deborah and Gregory Yetman. 2004. *Transforming Population Data for Interdisciplinary Usages: From census to grid*. Available at: [http://sedac.ciesin.columbia.edu/gpw/docs/gpw3\\_documentation\\_final.pdf](http://sedac.ciesin.columbia.edu/gpw/docs/gpw3_documentation_final.pdf)
- Center for International Earth Science Information Network (CIESIN), Columbia University; and Centro Internacional de Agricultura Tropical (CIAT), 2004. *Gridded Population of the World (GPW), Version 3*. Palisades, NY: Columbia University. Available at <http://beta.sedac.ciesin.columbia.edu/gpw>.
- Center for International Earth Science Information Network (CIESIN), Columbia University; International Food Policy Research Institute (IFPRI), the World Bank; and Centro Internacional de Agricultura Tropical (CIAT), 2004c. *Global Rural-Urban Mapping Project (GRUMP): Gridded Population of the World, version 3, with Urban Reallocation (GPW-UR)*. Palisades, NY: CIESIN, Columbia University. Available at: <http://beta.sedac.ciesin.columbia.edu/gpw>.
- Clark, John and David Rind, 1992. *Population Data and Global Environmental Change*. The International Social Science Council with the assistance of UNESCO, ISSC/UNESCO Series 5.

de Sherbinin, Alex. 2009. "The Biophysical and Geographical Correlates of Child Malnutrition in Africa" Population, Space and Place Vol.15, available at <http://dx.doi.org/10.1002/psp.599>.

Deichmann, Uwe, Deborah Balk and Gregory Yetman, Oct. 2001. "Transforming Population Data for Interdisciplinary Usages: From Census to Grid," available at <http://sedac.ciesin.columbia.edu/plue/gpw/GPWdocumentation.pdf>.

Dilley, Max, Robert Chen, Uwe Deichmann, Arthur L. Lerner-Lam and Margaret Arnold, with Jonathan Agwe, Piet Buys, Oddvar Kjekstad, Bradfield Lyon and Gregory Yetman. 2005. "Natural Disaster Hotspots: A Global Risk Analysis." Available at: <http://sedac.ciesin.columbia.edu/hazards/hotspots/synthesisreport.pdf>

Seirup, Lynn, Greg Yetman, and CIESIN at Columbia University. 2006. U.S. Census Grids, 2000. Palisades, NY: Socioeconomic Data and Applications Center (SEDAC), Columbia University.

Storeygard, Adam, Deborah Balk, Marc A. Levy and Glenn Deane. (2008). "The global distribution of infant mortality: A subnational spatial view," Population, Space and Place, 14 (3): 209-229. <http://onlinelibrary.wiley.com/doi/10.1002/psp.484/pdf>

Tobler, Waldo, Uwe Deichmann, Jon Gottsegen and Kelly Maloy. 1997. "World Population in a Grid of Spherical Quadrilaterals," International Journal of Population Geography, 3:203-225.

## Geospatial data in EEA assessments – status and requirements

Andrus Weiner  
European Environment Agency

*The analysis and views presented in this paper should be taken as the personal perspective of the author and cannot be regarded as the official position of the European Environment*

### Introduction

The European Environment Agency (EEA) is an agency of the European Union. It has a task to provide sound, independent information on the environment and provide a major information source for those involved in developing, adopting, implementing and evaluating environmental policy, and also the general public. Currently, the EEA has 32 member countries. EEA has supported the EU Sixth Environment Action Programme (6th EAP) across its four priority areas for action: climate change; biodiversity; environment and health; and sustainable management of resources and wastes (EC, 2001), including development and analysis of geospatial information related to priority actions of the EAP.

EEA strategy for 2009-2013 targets improving of our knowledge base. In particular, the Information and communication technology strategy towards 2013 is relevant to overall development of spatial data management. The main types of actions are:

- Enhancing the EEA's capabilities around spatial data, assuring INSPIRE implementation and development of Shared Environmental Information System;
- Increasing EEA capacity to handle new types of data, such as near-real time data, satellite data, citizen observations (through mobile devices), models;
- Strengthen role of EEA as European Environmental Data Centre and contribute to the European Spatial Data Infrastructure.

Following these principles the EEA is managing its spatial data by applying a wider concept of Spatial Data Infrastructure distinguishing between four focus areas:

- Institutional framework and organization
- Technical standards and specifications
- Geospatial data sets and metadata
- Spatial information services

Besides producing some geospatial data EEA is a user of data produced by other organisations. To get maximum benefit from integration and assimilation of these data sources, the EEA has continuously worked on its user requirements.