Development of new open and free multi-temporal global population grids at 250 m resolution

Sergio Freire
European Commission – Joint Research Centre (JRC)
Via E. Fermi, 2749
21027 Ispra (VA), ITALY
sergio.freire@jrc.ec.europa.eu

Kytt MacManus,
Columbia University CIESIN
Geoscience, PO Box 1000
61 US 9w
Palisades, NY 10964 USA
kmacmanu@ciesin.columbia.edu

Martino Pesaresi
European Commission – Joint Research Centre (JRC)
Via E. Fermi, 2749
21027 Ispra (VA), ITALY
martino.pesaresi@jrc.ec.europa.eu

Erin Doxsey-Whitfield
Columbia University CIESIN
Geoscience, PO Box 1000
61 US 9w
Palisades, NY 10964 USA
edwhitfi@ciesin.columbia.edu

Jane Mills
Columbia University CIESIN
Geoscience, PO Box 1000
61 US 9w
Palisades, NY 10964 USA
jmills@ciesin.columbia.edu

Abstract

Global population grids are increasingly required and used for countless applications in analysis, modeling, and policy-making. However, better and comparable global information requires improved geospatial data on population distribution and densities, in particular concerning temporal and spatial resolution and capacity for change assessment.

This paper presents the development of improved global multi-temporal population grids, by combining best-available population estimates for 1975, 1990, 2000 and 2014, with best-available assessment of the spatial extents of human settlements as inferred from Landsat satellite data for same periods. Using a dasymetric mapping approach, population is disaggregated from finest census or administrative units to built-up areas. These 250-m grids represent population changes in time, having also higher spatial resolution than those previously available.

The novel population grids constitute currently the *de facto* state-of-the-art in terms of open global geospatial population data, potentially enabling advances in many of the fields where this information is relevant.

Keywords: Population estimates, population grids, dasymetric mapping, global analysis, GPW, GHSL.

1 Introduction

Global population grids are essential to support analyses and inform policy-making in a wide range of fields, from environmental assessment to disaster risk analysis and reduction. Although such grids have been available since the 1990s, their usability is still limited by some constraints [1, 2]. Among these constraints are their still coarse spatial resolution, insufficient discrimination of urban areas (especially smaller ones), and the opacity of methods and parameters used in their construction. Also, for global change detection and monitoring, a clear, coherent and robust spatio-temporal production model is necessary to ensure users' confidence and allow objective comparisons in space and time.

Amongst the best known global population grids, LandScan [3] outputs surfaces of 'ambient population' at 0.5 arc-minutes which are produced using a complex (and proprietary) multi-variable dasymetric method. Although LandScan is reproduced annually and the methods constantly revised, the annual improvements made to model and underlying spatial variables advise against comparison of versions.

The History Database of the Global Environment - HYDE 3.1 project [4] departs from LandScan 2006 to estimate global population densities for the period 10,000 BC to 2005 AD (from 1700 AD by decade). Despite being a remarkable achievement, its coarse spatial resolution (5 arc-min.), the fact that it departs from a highly modelled 'ambient population' surface and the nature of the model, result in significant uncertainties that make it inappropriate for many applications.

Recently, the production of a 250-m global grid was announced [5], but the census data used in input range from 1990 to 2013 and access to the layer and use of data is strictly commercial.

The Gridded Population of the World (GPW) project [6, 7, 8] has employed a simple, transparent approach to create open and free residence-based population grids for different years, using areal weighting. Although globally, the nominal cell size (currently 0.5' in GPWv4) is not representative of the effective average spatial resolution (ASR) of input census units (ASR = 33 km) [9], there is significant variation locally and a large number of countries with very high resolution input geometries. Even though there are variations in the resolution of input units, the collection and integration of census data and geographies in GPWv4 results in the most detailed and coherent census-based geodatabase available globally.

In most population gridding approaches, significant effort and cost is expended combining and modeling geographic variables in order to obtain an appropriate proxy layer to support population disaggregation. An ideal covariate would be the location and characteristics of buildings, where most human activities take place.

The Global Human Settlement Layer (GHSL) project [10] maps the distribution and density of the world's built-up (BU) areas with unprecedented spatio-temporal coverage and detail. Such geoinformation constitute a good covariate for population disaggregation [11], especially in combination with detailed, matching census data.

This paper presents the development of new multi-temporal global population grids (GHS-POP), based on best-available global built-up and census geospatial data and following a simple and clear modeling approach.

2 Data and methods

2.1 Data sets

Input data for this work has included vector-based population estimates and continuous grids of built-up areas (Table 1).

Table 1: Main input datasets used in modeling population distribution.

Data set	Source	Dates	Data type
GHS-BUILT	EC-JRC, IPSC	1975 1990 2000 2014	Raster (250m)
Population estimates	CIESIN-GPWv4	1975 1990 2000 2015	Vector polygon

The Global Human Settlement Layer built-up grids (GHS-BUILT) were recently produced from Landsat imagery collections (epochs: 1975, 1990, 2000 and 2013-2014) [12]. GHSL technology relies on automatic analysis of satellite imagery to produce unprecedented fine scale maps quantifying built-up structures in terms of their location and density. The image processing technology exploits structure (texture, morphology, and pattern) as key information, outputting a texture-derived "built-up presence index". The distribution of built-up areas is expressed as their proportion (ratio) of occupied footprint in each cell. GHS-BUILT was aggregated from its native output resolution (38 m) to that selected for this population model (250 m), in World Mollweide projection (EPSG 54009), an equal-area representation of land masses.

Population estimates were produced and made available for processing by the Center for International Earth Science Information Network (CIESIN). These consist of country-based layers (one for each of 241 countries) of census and administrative polygons containing estimated residential population for target years 1975-1990-2000-2015, as produced for GPWv4 [13]. The production procedure is explained in more detail below.

The two input variables involved (BU and population estimates) display unusually high temporal coherence, especially when global analyses are concerned.

2.2 Production of population estimates for 1975, 1990, 2000, and 2015

While in version 4 GPW has evolved in terms of increases in spatial resolution (now 30 arc-second or ~1km), and use of contemporary population estimates from the 2010 round of population censuses, the essential data inputs to GPW have remained population census tables and cartography, as these provide the most complete and exclusive geographic coverage of a country's population at a given point in time.

2.2.1 Population data sources

The two basic data inputs to GPW are tabular population estimates corresponding to enumeration or administrative areas, and spatially-explicit geographic boundary data. Population input data were collected at the highest possible spatial resolution from the results of the 2010 round of censuses (Figure 1). For a full description of sources please refer to Doxsey-Whitfield et al [13].

Figure 1: GPWv4 input administrative level by country.

Source: CIESIN-GPWv4.

2.2.2 Geometry matching

For GPW, the integration of census data and geography is accomplished in a process that compares attribute data on hierarchically nested area names from input population tables to those present in the input boundary source. Unmatched data is divided into two categories based on their source. Unmatched rows from the input population estimates are reconciled though primary research of alternative boundary sources which may include paper maps or legal descriptions. Unmatched units in the boundary data are commonly reconciled by historical research on territorial or name changes. In some cases research is not able to reveal the reason for a unit mismatch. In those situations the units are aggregated to a coarser spatial resolution.

Data integration along international boundaries presents another set of issues. Contested territories, or differences in surveying techniques may often lead to gaps or overlaps along shared international borders. In order to reconcile these issues, a global framework of international boundaries was used. The Global Administrative Areas version 2.0 (GADMv2) [14] data set was selected as the international framework because it is frequently used in the research community, and includes many countries at a high spatial resolution. The international boundaries of census geography data sets were adjusted to the GADMv2 framework when required.

2.2.3 Estimating population for target years

The GPWv4 data set includes population estimates for years 2000, 2005, 2010, 2015, and 2020. To support integration of GPWv4 with GHSL, population estimates for a series of additional target years conforming to the GHSL temporal coverage (i.e. 1975, 1990) were derived. In all cases exponential growth rates were calculated for each administrative unit by matching the total population from the input data to those from a previous census enumeration. In cases where matching high resolution geographic units in the two points in time were impossible, census data were matched and growth rates were calculated at a coarser resolution. Annualized rates of change were then calculated using standard formulas and procedures.

The extrapolation of exponential annualized growth rates is problematic when the composition of population in a given country experiences sudden or dramatic changes, or when the rate is applied over an extensive period. Additionally, in some cases, official statistics are hampered by errors in reporting, lack of timeliness, or incompleteness in coverage. In order to account for these inherent limitations and to improve the certainty of population estimates, adjustments are needed. The UN Population Division invests substantial effort in adjusting national census data based on these, and other issues. The UN World Population Prospects (WPP): The 2015 Revision is the source of the most current official national-level UN population estimates [15]. In order to produce a set of adjusted estimates, national level factors were calculated by comparing the GPWv4 source data to the WPP estimates. These factors are then applied uniformly to geographies within each country so that the final estimated country population matches the UN's WPP estimates.

2.3 Population disaggregation model

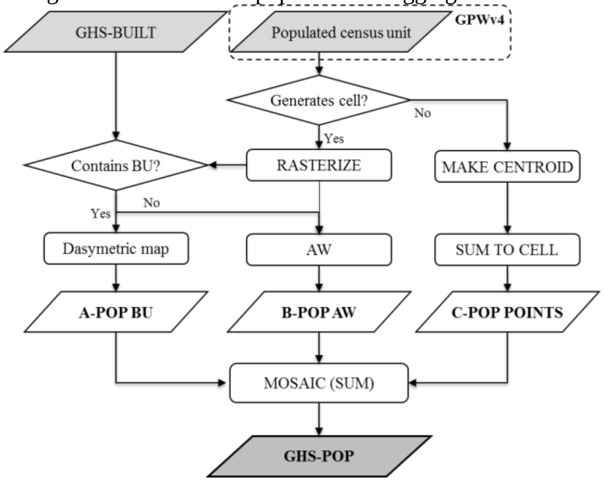
To produce the new population grids, a simple model was developed and implemented. The approach is based on raster-based dasymetric mapping [16, 17] relying on GHS-BUILT as proxy to restrict and refine the distribution of people and inform on the respective density. For each period, population grids were produced following a clear volume-preserving dasymetric mapping approach (Figure 2): given a census layer and a GHSL BU raster layer, for a populated polygon (source zone),

A) if the polygon generates 250 m cells and contains BU, disaggregate the population in proportion to density of BU (POP BU);

B) if the polygon generates 250 m cells and does not contain BU, disaggregate population using areal weighting (POP AW);

C) if the polygon does not generate its own 250 m cell, convert polygon to point (centroid within), sum all points within a cell, and sum to mosaic of a) and b) (POP POINTS).

Figure 2: Flow chart of population disaggregation model.



Option C) was required because some populated source zones (polygons) were smaller than a 250 x 250 m cell. The rasterization method used was the cell center (polygon which overlaps the center of the cell yields the attribute to assign to the cell).

The developed model allows preserving the full resolution of the GPW input population units and conserves the total input population.

3 Results and discussion

3.1 Production of population estimates for 1975, 1990, 2000, and 2015

Table 2 shows overall results of population estimation stage that was used as input to population disaggregation.

Table 2: Total population and number of populated units in 1975-2015 (millions).

Year	1975	1990	2000	2015
Population	4,062	5,310	6,127	7,351
No. of pop. units	7.968	7.991	7.996	7.996

3.2 Population disaggregation model

Using the approach presented above, four global 250-m population distribution grids were produced for the years 1975, 1990, 2000, and 2015. Such raster datasets depict the distribution of resident-based population in each period, expressed as the estimated number of people per cell. These are residential-based population estimates in built-up – and not 'residential population' or 'population in their place of residence', for which consideration of land use would be required. Since these grids are produced and made available in Mollweide projection, values represent both population counts as well as densities. The four GHS-POP 250-m grids are illustrated in Fig. 3 for Bangkok, Thailand.

Figure 3: GHS-POP 250-m grids for Bangkok, Thailand (1975: red, 1990: green, 2000: violet, 2015: blue)

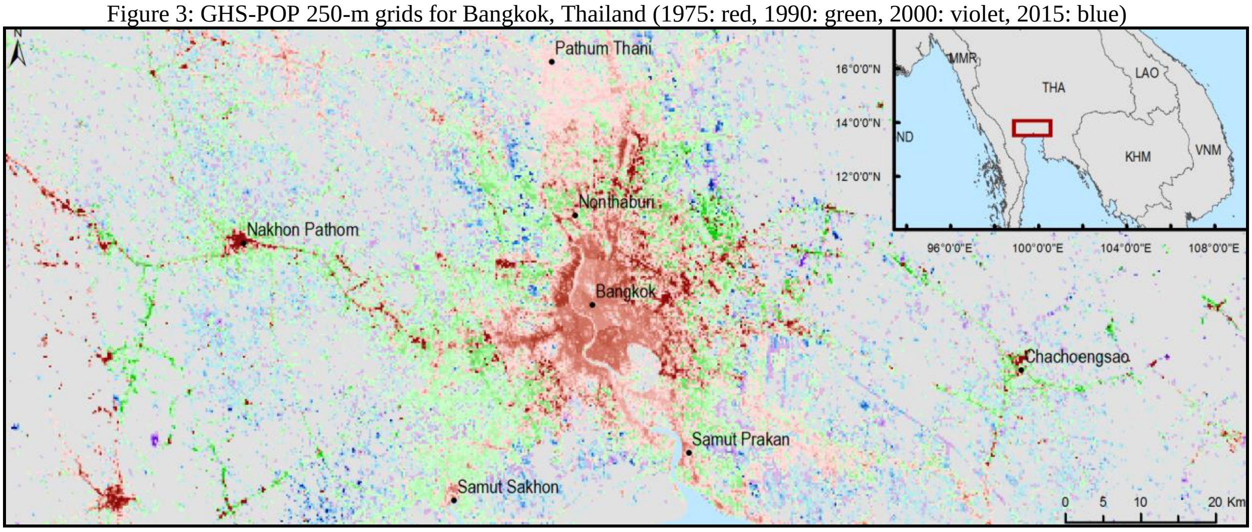


Fig. 4 illustrates GHS-POP 2015 in comparison to LandScan 2014 for Bangkok, Thailand.

For 1975, greater uncertainties in estimating small-scale populations farther from census year combine with larger difficulties in mapping built-up (from lower resolution Landsat MSS sensor) to make this surface less reliable. This could potentially be mitigated by merging fine census units to a lower level prior to population disaggregation.

A consequence of method C-Points is the allocation of population from these (typically small) polygons to only one cell, and occasionally a slight displacement of population to neighboring 250 m cells, but such effect can be considered negligible.

Concerning the disaggregation method effectively applied in each grid, Table 3 shows that in three periods analysed almost all the global population (around 95%) was disaggregated into and in proportion to mapped BU, and only small shares were rasterized via points (which may be in BU cells) or areal weighted (AW). This shows that GHS-BUILT allowed for a very significant and effective increase in the spatial resolution of input population layers.

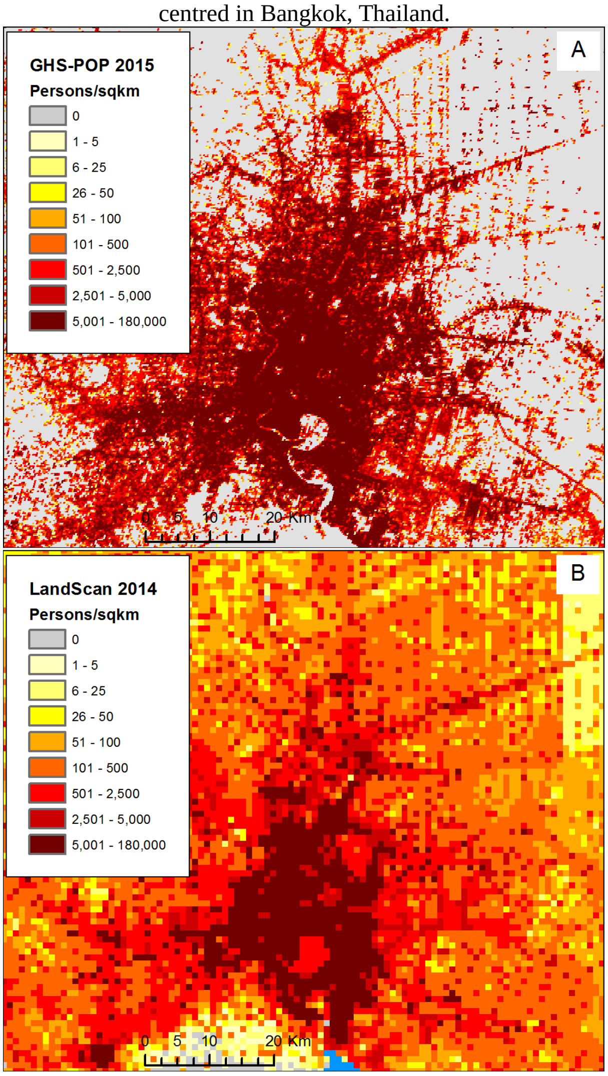
Table 3: Results of population disaggregation according to method applied in model (in percent of total population).

Disaggregation method	1990	2000	2015
A-POP BU	94.2	95.5	95.5
B-POP AW	4	2.8	2.2
C-POP Points	1.8	1.7	2.3
Total	100	100	100

Verification and validation

Quality control was conducted to ensure that all input population was disaggregated and totals were preserved. A strict quality assessment procedure was limited by lack of independent but compatible (i.e. concerning date, concept) reference data. Nevertheless, benchmarking of results was made using official GEOSTAT 2011 resident population [18], by selecting from that database 18 European countries without disclosure control (total pop. = 249,200,695; ASR = 3.1 km). After aggregating the 2015 GHS-POP to the same vector 1-km populated cells (n = 1,111,646), correlation analysis was performed, yielding r = 0.83.

Figure 4: GHS-POP 2015 (A) and LandScan 2014 (B) centred in Bangkok, Thailand.



4 Conclusions

Better and finer global analyses of population-environment interactions require enhanced geoinformation on population distribution and densities, in particular concerning temporal and spatial resolution and capacity for change assessment.

Novel and improved global multi-temporal population grids were produced at 250 m, by combining best-available population estimates for 1975, 1990, 2000 and 2015, with best-available assessment of the spatial extents of human settlements as inferred from Landsat satellite data for same periods. The use of a robust, consistent global approach for population disaggregation into built-up areas mapped from satellite imagery enables analyses of changes in time. These grids have also higher spatial resolution than those previously available, enabling advances in many of the fields where this information is relevant.

Simplicity and clarity of procedures facilitates future improvements, as well as favors inter-operability with other geoinformation, and therefore widens potential range of applications.

Future developments will include increasing the spatial resolution of output grids, as well as consideration of land use/functional characterization of buildings and their height in the process of population disaggregation.

References

[1] C. Linard and A. Tatem. Large-scale spatial population databases in infectious disease research. *International Journal of Health Geographics*, 11:7, 2012.

[2] P. Mondal and A.J. Tatem. Uncertainties in Measuring Populations Potentially Impacted by Sea Level Rise and Coastal Flooding, PLoS ONE, 7(10), 2012.

[3] J. Dobson, E.Bright, P.Coleman, R.Durfee, and B.Worley. LandScan: A global population database for estimating populations at risk. Photogramm. Eng. Remote Sens., 66(7):849–857, 2000.

[4] K.K Goldewijk, A. Beusen, M. de Vos and G. van Drecht. The HYDE 3.1 spatially explicit database of human induced land use change over the past 12,000 years. *Global Ecology and Biogeography*, 20(1):73-86, 2011.

[5] Esri, World Population Estimate, 2014.

[6] W. Tobler, U. Deichmann, J. Gottsegen, and K. Maloy. World population in a grid of spherical quadrilaterals. *International Journal of Population Geography* 3:203-225, 1997.

[7] U. Deichmann, D. Balk, and G. Yetman. *Transforming Population Data for Interdisciplinary Usages: From Census to Grid*. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC), CIESIN, Columbia University, 2001.

[8] D.L. Balk, U. Deichmann, G. Yetman, F. Pozzi, S.I. Hay, and A. Nelson. Determining global population distribution: methods, applications and data. *Advances in Parasitology* 62:119-156, 2006.

[9] E. Doxsey-Whitfield, K. MacManus, S.B. Adamo, L. Pistolesi, J. Squires, O. Borkovska and S.R. Baptista. Taking Advantage of the Improved Availability of Census Data: A First Look at the Gridded Population of the World, Version 4. *Papers in Applied Geography*, 1(3):226-234, 2015.

[10] M. Pesaresi and D. Ehrlich (2009) A methodology to quantify built-up structures from optical VHR imagery. In P. Gamba and M. Herold, editors, *Global Mapping of Human Settlement Experiences, Datasets, and Prospects*, CRC Press, Boca Raton, FL, ch. 3, 27–58, 2009. [11] S. Freire, T. Kemper, M. Pesaresi, A.J. Florczyk, and V. Syrris. Combining GHSL and GPW to Improve Global Population Mapping. *2015 IEEE International Geoscience & Remote Sensing Symposium Proceedings*, 2541-2543, 2015.

[12] M. Pesaresi, D. Ehrlich, S. Ferri, A. Florczyk, S. Freire, M. Halkia, A. Julea, T. Kemper, P. Soille, and V. Syrris. Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014. *JRC Technical Report*, EUR 27741 EN, 2015.

[13] Center for International Earth Science Information Network - CIESIN - Columbia University. Gridded Population of the World, Version 4 (GPWv4). Palisades, NY, 2014. http://www.ciesin.columbia.edu/data/gpw-v4.

[14] GADM, v.2.0 (Global Administrative Boundaries Database). Available at: http://www.gadm.org/.

[15] United Nations, Department of Economic and Social Affairs, Population Division. World Population Prospects: The 2015 Revision, DVD Edition, 2015. Available at: http://esa.un.org/unpd/wpp/DVD/

[16] J. K. Wright. A method of mapping densities of population, *The Geographical Review*, 26(1):103–110, 1936.

[17] J. Mennis. Generating surface models of population using dasymetric mapping. *Prof. Geogr.*, 55:31–42, 2003.

[18] Eurostat, GEOSTAT 2011 V2, 2015. Available at: http://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/population-distribution-demograph.