# Remote Sensing and Gridded Population Data: Considerations for the Population-Environment Research Community

Panel contribution to the Population-Environment Research Network Cyberseminar,
"People and Pixels Revisited"
(20-27 February 2018)
https://populationenvironmentresearch.org/cyberseminars/10516

By Andrea Gaughan*, Catherine Linard**, Forrest Stevens*, and Andy Tatem***
* University of Louisville, Kentucky, USA
** Université Libre de Bruxelles, Belgium
*** Southampton University, UK

## Introduction

Moore's law, an observational reflection, noted in a 1965 publication by Intel co-founder Gordon Moore, predicted an exponential growth in technological development that came to shape the silicon chip industry for decades to come. That rate of development, arguably, has slowed in recent years. However, the sheer magnitude of technological advancement, including remote sensing sensors and geospatial platforms and techniques, has shaped the direction of how people and pixels are linked for understanding human-environment interactions. Specifically, since the seminal work, *People and Pixels: Linking Remote Sensing and Social Science* (NRC, 1998), one of the most progressive research agendas in "socializing the pixel" has been to literally take population data and grid the counts into a spatially-explicit human denominator.

Along with increasingly fine-grained and accessible remote sensing data and techniques for analyzing the environmental dimension, there has been a research community pushing the bounds for how to best quantify and grid human populations and their demographic information from local to global scales. Gridded population information provides a consistent, comparable areal unit to represent the human denominator that has appropriate spatial representation relative to the other sets of information in analysis. Gridded population data provides a base spatial denominator to identify specific populations at risk, quantify burdens, and inform our understanding of human-environment systems, both from a theoretical and applied perspective.

Gridded population data led to key advancements, especially in public health. While 20 years ago, remote sensing was mostly used to map health risks, and more particularly the spatial distribution of vector-borne diseases through the mapping of vector habitats, it is now commonly used to also integrate the human factor. The spatial distribution of populations, their demographic and socio-economic characteristics, and their connectivity all have a considerable impact on disease dynamics. Gridded population data are also particularly useful for measuring progress towards international health and development goals, to plan vaccination needs and estimate infectious disease burden.

## Advances in modeling human population and the role of remote sensing data

Traditional approaches to grid population rely on national censuses which can provide a comprehensive and relatively unbiased source of information at a single time point, and when linked with accurate boundary data, provide a spatially detailed evidence base on population. The techniques and underlying methods that produce gridded population modeling evolved from early efforts of simply areal weighting census counts tied to GIS-defined administrative units. The *areal weighting* approach equally distributes the total count tied to a unit across all grid cells within the boundary of that administrative area (Doxsey-Whitfield et al., 2015). Slightly more evolved is a *pycnophylatic* approach which weights the redistribution of census count to smooth out edges between census units (Tobler, 1979). With increasing availability of spatially explicit data, including the advancements in remote sensing and its techniques, the *dasymetric* mapping approach, which relies on ancillary data to disaggregate census counts at a coarser resolution to a finer scale (Eicher and Brewer, 2001, Balk et al., 2006), became very common. Starting with data like satellite-derived, urban/rural redistribution of populations (Balk et al., 2006), most modern techniques use a variety of data, mainly derived via remote sensing (Stevens et al 2015).

Thus, disaggregating census data through integration with higher spatial resolution ‘covariate’ datasets in modelling frameworks can then disaggregate these boundary-linked counts to consistent gridded representations (Fig.1, Stevens et al 2015, Balk et al 2006, Azar et al 2010, Bhaduri et al 2007, Sorichetta et al 2015). Different gridded products currently exist to the end-user, all based on different underlying techniques (Stevens et al., 2015, Balk et al., 2006, Dobson, 2000). But, with increasing reliance on these products as the human denominator from which a multitude of other research and policy initiatives rely, making not only the accuracy of the population data important but also a need to understand how that population data was created and appropriate applications for use.

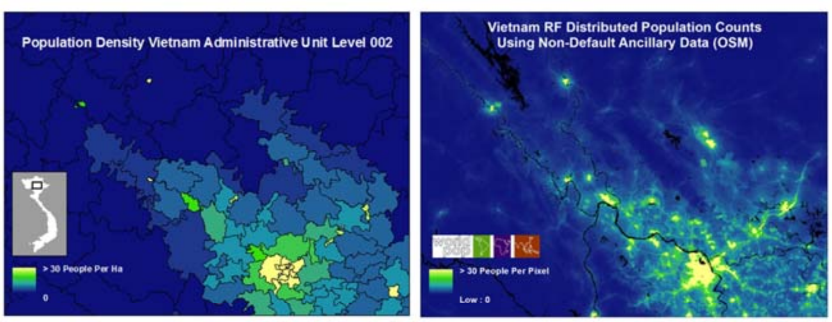


*Fig.1 WorldPop population mapping example: (left) Population density from census data for each administrative level 2 unit in an area of northern Vietnam, (right) WorldPop population modelling methods take the census data as input, then use machine learning methods to exploit the relationship between population density and high resolution landscape features, such as those from land cover and satellite data, to predict population densities for each 100x100m grid cell on the landscape.*

Methods to grid population data continue to evolve in parallel with remote sensing developments. While early population estimation work relied on hand-drawn land cover classifications and were guided by “controlled guesswork” (Wright, 1936), most modern techniques use very spatially detailed settlement layers now available at the global level (Esch et al 2013, Pesaresi et al 2013). The increasingly finer grain of built and settlement areas specifically makes a binary dasymetric approach appealing and competitive in terms of assessment compared to more complex statistical approaches.

What’s interesting about a more straightforward binary dasymetric approach (versus a statistical model) is the variety of ways one might apply the dasymetric constraint to redistributing counts.

Binary dasymetric techniques rely on one ancillary data set to inform which grid cells receive source-to-target values (e.g. administrative unit to pixel) (Mennis and Hultgren, 2006). Naturally, the quality of the population map will be tied to the quality of the underlying ancillary dataset. It was generally acknowledged that there are limitations for using imagery to estimate population in sparsely settled areas, especially in regions where buildings are made of the same materials as surrounding landscapes and therefore difficult to detect on satellite images. The continued improvements in new sensors and new techniques that leverage the value of optical and radar-based to improve on both error rates of omission and commission will certainly push for continued efforts in the comparison of population modelling approaches.

Arguably, the best approach could be a hybrid of the two, creating a weighting layer based on statistical model but masking out any area other than built areas. The best approach could also vary within and between countries. For instance, a different approach could be used in rural vs. urban areas in data-scarce countries, as urban land uses (e.g. planned/unplanned residential, industrial, commercial, etc.) are important ancillary information for improved mapping of population density variations within urban built pixels.

**Alternative data and approaches**

The top-down modeling approach described above uses remote sensing data for spatially disaggregating national census data. This necessitates a solid and regularly updated understanding of not only how many people live in a country, but where the people are, and who they are. Such requirements and the deficiencies of national census data mean that other data sources are increasingly being explored in efforts to produce estimates at different geographical scales and time periods.

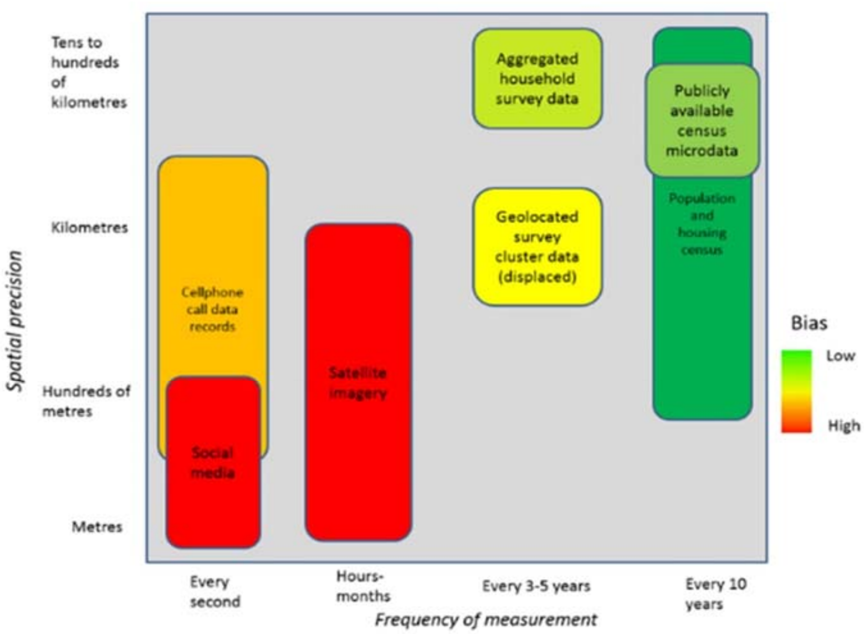


*Fig.2 Examples of datasources used in the construction of high resolution population maps in low/middle-income countries, and their features*

Figure 2 highlights some of those data currently being utilised in the population research community to complement census data in the detailed mapping of populations and their characteristics across timescales. Though increasingly prone to bias through measurement of smaller sample sizes (e.g. geolocated household survey clusters), specific demographic groups (e.g. social media) or simply factors related to population densities (e.g. satellite imagery), each source has advantages over census data in terms of the frequency of measurement and spatial precision (Fig.2). Moreover, their utilization represents a gradual shift from 'top-down' approaches where census data counts are maintained and disaggregated to small areas, to more 'bottom-up' approaches, where estimates are made independent of census data.

## Uncertainties and Considerations

There are an increasing variety of different gridded population products available to the end-user. Informed decisions should include knowledge about the underlying characteristics of the data used to produce the gridded map. What is the quality (e.g. how recent, granularity) of the census data? What is the accuracy of any ancillary data sources used in producing the population map and is it contemporaneous with the scale of interest? This necessitates extensive source and metadata information.[1] One of the greatest challenges we still face is appropriate means of validating final model outputs. The nature of a source (census unit) to target (pixel) top-down modeling approach, makes it difficult to have reliable, accurate validation data from which to assess model fit. Even more challenging with a model developed at regional or global scales.

The incredible evolution of remote sensing techniques and population modelling methods over the last decades could lead to a data revolution in data-scarce countries. For example, in a country where recent census data are outdated, unavailable, or difficult to measure and collect, a combination of remote sensing data with micro-census-derived population counts allows putting people on the map using more sophisticated modeling approaches. Applications of such approaches might be finding neglected populations who have been excluded from aid distribution, vaccination programs, voting, etc. over the past decade or more.

While challenges remain, spatial reasoning and the importance of space (and time) for linking environmental changes to the distribution, movement and concentration of human population is implicit for tying the "why" to the "where," oftentimes requiring a multi-scalar perspective. Gridded population provides an important input source of information for spatially-explicit analyses detailing some type of pattern about a phenomenon as related to human population. Continued efforts to provide well-validated and well-documented spatial demographic datasets will be required to obtain the full benefits from these powerful methodologies.

## References

Azar D, Graesser J, Engstrom R, Comenetz J, Leddy RM, Schechtman NG, et al. 2010. Spatial refinement of census population distribution using remotely sensed estimates of impervious surfaces in Haiti. Int J Remote Sens.

Bhaduri B, Bright E, Coleman PR, Urban ML. 2007. LandScan USA: a high-resolution geospatial and temporal modeling approach for population distribution and dynamics. GeoJournal. 69 (1/2):103–17.

Balk DL, Deichmann U, Yetman G, Pozzi F, Hay SI, Nelson A. 2006. Determining global population distribution: methods, applications and data. Adv Parasitol. 62:119–56.

[1] The POPGRID initiative unites the major gridded population and settlement data providers in providing consistent documentation and tools for identifying data appropriate to different use cases. See www.popgrid.org.

Dobson JE, Brlght EA, Coleman PR, Worley BA. 2000. LandScan: A Global Population Database for Estimating Populations at Risk. Photogramm Eng Remote Sens. 66(7):849–57.

J. Mennis, T. Hultgren. 2006/ Intelligent dasymetric mapping and its application to areal interpolation Cartogr. Geogr. Inf. Sci., 33, pp. 179-194, 10.1559/152304006779077309.

Doxsey-Whitfield, E. *et al.* 2015. Taking advantage of the improved availability of census data: A first look at the Gridded Population of the World, Version 4 (GPWv4). *Papers in Applied Geog* 1, 226–234.

Eicher, Cory L., and Cynthia A. Brewer. 2001. Dasymetric mapping and areal interpolation: Implementation and evaluation. Cartography and Geographic Information Science 28 (2):125–38.

Liverman D, Moran EF, Rindfuss RR, Stern PC, eds. 1998. People and Pixels: Linking Remote Sensing and Social Science. Washington, DC: Natl. Acad.

Esch, T., Marconcini, M., Felbier, A., Roth, A., Heldens, W., Huber, M., Schwinger, M., Taubenböck, H., Müller, A., and S. Dech. 2013. Urban footprint processor—fully automated processing chain generating settlement masks from global data of the TanDEM-X mission
IEEE Geosci. Remote Sens. Lett., 10, pp. 1617-1621, 10.1109/LGRS.2013.2272953.

Pesaresi, M., Huadong, G., Blaes, X., Ehrlich, D., Ferri, S., Gueguen, L., Halkia, M., Kauffmann, M., Kemper, T., Lu, L., Marin-Herrera, M.A., Ouzounis, G.K., Scavazzon, M., Soille, P., Syrris, V., and L. Zanchetta 2013. A global human settlement layer from optical HR/VHR RS data: concept and first results IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 6, pp. 2102-2131, 10.1109/JSTARS.2013.2271445.

Stevens, F.R., Gaughan, A.E., Linard, C., and A.J. Tatem. 2015.Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data
PLoS One, 10, Article e0107042, 10.1371/journal.pone.0107042.

Sorichetta A, Hornby GM, Stevens FR, Gaughan AE, Linard C, Tatem AJ. 2015. High-resolution gridded population distribution datasets of Latin America in 2010, 2015, and 2020. Sci. Data 2, 150045.

Tobler, W. 1979. Smooth pycnophylactic interpolation for geographic regions. Journal of the American Statistical Association 74:519–30.

Wright, J.K. 1936. A method of mapping densities of population: with cape cod as an example
Geogr. Rev., 26, pp. 103-110.