Time Series Population Data Harmonization in Deforestation Research

María José Andrade-Núñez
Faculty of Natural Science, University of Puerto Rico
e-mail: mj_andradenunez@yahoo.com

PERN Cyberseminar on Application of Gridded Population and Settlement Products in Geospatial Population-Environment Research
14 – 18 October 2019

Forest cover and human population maps

The increase in human population is recognized as one the drivers of global change, therefore determining human population changes in urban and rural areas is crucial to understand the underlying causes of land-use changes and specifically forest cover changes over time. Forest change can be related to human population change; however, its relationship is not always direct. For instance, in many areas of Latin America, reforestation has been associated with rural-urban migration (i.e. decrease in population) (Aide & Grau, 2004); however, reforestation also can occur in highly populated areas (Baptista, 2008), and areas with an increasing human population (Redo, Aide, & Clark, 2013). Similarly, deforestation trends have been identified in areas with low human population and large-scale agriculture (Aide et al., 2013). Therefore, spatial human population datasets coupled with other socio-economic, climatic and biophysical information are important sources of information to understand global and regional patterns of land changes (Aide & Grau, 2008).

Deforestation and reforestation studies have been conducted at different spatial and temporal scales (from local to global assessment, and for yearly to multi-year and decadal time series). In addition, different classification methods, and different satellite images with differences in temporal and spatial resolution (e.g., Landsat, MODIS, Aster) have been used to map forest changes. Despite the variety of methods, images, and temporal and spatial resolutions, the information produced (i.e., land use map) with this type of data can be easily assessed for accuracy.

Nevertheless, the development of spatial population datasets and specifically gridded-population datasets produce more uncertainties and errors than forest cover maps. These uncertainties and errors are related to the intrinsic characteristics of the methods used to acquire population data. Population data mostly rely on national census counts, which vary greatly among countries in the temporal and spatial scales and in the methods used to count people. Additionally, the process of population data harmonization and the methods to create a population grid from irregular vectors (i.e., administrative boundaries) involves a series of assumptions that diminish the spatial accuracy of the data. Despite the uncertainty associated with the development of large-scale gridded population data products, they provide valuable spatial information that has been used in a wide array of scientific research topics.

There are many large-scale gridded population data sets, but depending on the purpose of a research project the information and methods of one may be the most appropriate. Based on my experience I think that the selection of the spatial population dataset (gridded or not) must be based on: the research objective, and the spatial scale of the analysis (e.g. municipality level, 1 km grid, 250 m grid) and time series. In addition, it is important to understand how the spatial population dataset is going to be used, analyzed, and integrated into other spatial variables in the research. These three points are important for defining the type of spatial population dataset to be used. Furthermore, it is important to understand the accuracy and precision of the spatial population dataset to be used and indicate when necessary its limitations. In this regard, the discussion paper (Leyk et al., 2019) summarizes the important aspects of the available large-scale gridded population datasets such as data properties, the methods used to create them, and quality aspects. I consider that the discussion paper is an important contribution to researches/users to help in the process of selecting the most suitable data set for his/her research.

Dealing with spatial population dataset: from national census data to gridded population products

In my opinion, and aligned with the discussion paper (Leyk et al., 2019) several aspects need to be considered in the process of population data harmonization, aspects that potential users of the gridded population products should have in mind:

1) Different national definitions of what is urban and rural.
2) Years of each national census.
3) Number and size of national administrative units.
4) Accuracy of country field method used to count people.
5) The methods used to deal with changes in administrative unit boundaries between censuses. This is a critical step in the process of harmonizing population data.
6) Use of population extrapolation estimates to provide time-series data. This step is specifically relevant when census years are not the same among countries.
7) The time (year/s) of gridded data.
8) The spatial resolution of gridded data.
9) The spatial allocation of population counts to grid cells: Simple (e.g., uniform allocation of the population in grids) vs. Complex (e.g., aggregate allocation based on dasymetric mapping using ancillary data) methods.
10) Aggregation issues related to the process of producing gridded data from administrative unit's census data.
11) The spatial and temporal resolution of the ancillary data used in dasymetric and hybrid methods.

Examples of using spatial population datasets in scientific research related to forest change and infrastructure expansion

The first example is a study, in which I was a co-author, that explored the patterns of deforestation and reforestation at the municipality level in Latin American and the Caribbean countries between 2001-2010 (Aide et al., 2013). In this study the scale of the analysis was the municipality (or second level administrative unit); therefore, land-use and population data were obtained at this spatial scale. We did not use a gridded-population data set, but we performed part of the process of data harmonization needed to create a gridded-population data set. Specifically, we performed the following steps: 1) gathered total population data from national census data at the municipality scale (n=16,050), 2) when a municipality boundary change between census we used the boundary of the first census and combined the population data from the two new municipalities, 3) extrapolate official population data using the same method as the one applied in the Gridded Population of the World, v. 3 (CIESIN [Center for International Earth Science Information Network] & CIAT [Centro Internacional de Agricultura Tropical] 2005).

In the second study (Andrade-Núñez & Aide, 2018) the objective was to assess the patterns of built-up expansion across the South American continent between 2001 and 2011. We classified municipalities based on built-up and population net changes between 2001 and 2011, and then assigned each municipality to one of four built-up/population categories: (1) shrink-age, (2) densification, (3) rural expansion, and (4) urban expansion. We used spatial population data at municipality scale because of the lack of gridded population data for the interested time period (2001-2011). The study period was defined based on nighttime light (NTL) data availability which was used to characterize and map built-up expansion. We basically performed the same procedures explained in the first example; however, in this study we were able to use urban and rural population data.

Future improvements of gridded-population data sets

In the future it would be important to produce finer scale gridded population data sets to match global forest cover maps produced at high resolution (e.g. Landsat image 30 m, Hansen et al., 2013). It would also be useful to have yearly gridded population data sets (or at least increase the frequency of censuses) so users can easily match population with other data products (e.g. economic, NLT, forest cover). In addition, new techniques and technologies are needed to improve gridded population data sets that can spatially identify with accuracy population in urban and rural areas. This kind of dataset would provide valuable information to improve the understanding of the drivers of deforestation and land use change.

References

Aide, M., & Grau, H. R. (2008). Globalization and land use transitions in Latin America. *Ecology and Society*, *13*(2), 60p. https://doi.org/10.1057/9780230603554

Aide, T.M., & Grau, R. H. (2004). Globalization, Migration, and Latin American Ecosystems. *Science*, *305*, 1915–1916. https://doi.org/10.1126/science.1103179

Aide, T.M., Clark, M. L., Grau, H. R., López-Carr, D., Levy, M. A., Redo, D., … Muñiz, M. (2013). Deforestation and Reforestation of Latin America and the Caribbean (2001-2010). *Biotropica*, *45*(2). https://doi.org/10.1111/j.1744-7429.2012.00908.x

Andrade-Núñez, M. J., & Aide, T. M. (2018). Built-up expansion between 2001 and 2011 in South America continues well beyond the cities. *Environmental Research Letters*, *13*, 084006. https://doi.org/10.1088/1748-9326/aad2e3

Baptista, S. R. (2008). Metropolization and forest recovery in Southern Brazil: A multiscale analysis of the Florianópolis City-Region, Santa Catarina State, 1970 to 2005. *Ecology and Society*, *13*(2). https://doi.org/5

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., … Townshend, J. R. G. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, *342*(6160), 850–853. https://doi.org/10.1126/science.1244693

Leyk, S., Gaughan, A. E., Adamo, S. B., de Sherbinin, A., Balk, D., Freire, S., … Tatem, A. J. (2019). The spatial allocation of population: a review of large-scale gridded population data products and their fitness for use. *Earth System Science Data Discussions*, *11*, 1385–1409. https://doi.org/10.5194/essd-2019-82

Redo, D., Aide, T. M., & Clark, M. L. (2013). Vegetation change in Brazil's dryland ecoregions and the relationship to crop production and environmental factors: Cerrado, Caatinga, and Mato Grosso, 2001-2009. *Journal of Land Use Science*, *8*(2), 123–153. https://doi.org/10.1080/1747423X.2012.667448