Local Data for Local Decisions[1]

Brian Blankespoor
Analytics and Tools Unit
Data Department, The World Bank
Washington DC USA
bblankespoor@worldbank.org

Panel Contribution to the PERN Cyberseminar on the Application of Gridded Population and Settlement Products in Geospatial Population-Environment Research[2]
14 to 18 October 2019

Population data represent an **essential** component in studies focusing on economic development, disaster risk assessment, transportation and environment. Collection of population data typically occurs during the census and population-based national household surveys. However, the frequency of census is at least every decade and household surveys offer representativeness at a limited number of geographic areas. This highlights an important trade-off between accuracy and spatial resolution that must be resolved before these data can be applied at the local level for stakeholders.

As the **demand for local level statistics** increases, high resolution spatial data and statistical models enable the mapping of characteristics like population down to the small area level. Modern statistical approaches, such as dasymetric modeling, interpolation, cross-entropy, machine-learning and small area estimation, can help tackle the problem of providing feasible estimates for smaller geographical areas, while explicitly acknowledging uncertainty associated with the estimates.

Several recent efforts deploying advanced methods have produced global- and continental-extent **gridded population data**, which are becoming increasingly popular among various communities in development (Leyk et al. 2019). However, these data products, which are of very

[1] The author would like to thank Alex De Sherbinin (Columbia University), Rashmin Gunasekera (World Bank), Siobhan Murray (World Bank), members of the World Bank's Geospatial Community of Practice and POPGRID Data Collaborative participants for previous discussions. Also, the author thanks the PERN organizing committee including Andrea Gaughan (University of Louisville) for the opportunity to participate on the panel. The findings, interpretations, and conclusions expressed in this panel cyberseminar note are entirely those of the author. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.
[2] https://www.populationenvironmentresearch.org/cyberseminars/10680

different characteristics and based on different modeling assumptions, have not previously been systematically reviewed and compared, which may impede wider adoption.

To that aim, the POPGRID Data Collaborative provides a viewer and comparison table of the multiple global gridded population datasets. This information provides potential data users with the knowledge base needed to make informed decisions about the appropriateness of the gridded population data products available in relation to the target application and for critical analysis.

Population Data Applications at the World Bank

The World Bank has been using gridded population datasets in its work for more than a decade. Recognizing that growth in cities, increasing mobility of people and ever more specialized products are integral to economic development, the seminal World Development Report 2009: Reshaping Economic Geography presents three dimensions of economic development: density, distance and division.[3] Other applications include: disaster risk management, measurement of indicators and economic analysis.[4] A few highlights described and referenced below include: population movement from climate change, the spatial definition of urban, the use of household survey to validate population models, and quantifying the number of people exposed to a hazard in disaster risk management.

Measurement matters

The measurement of population informs spatial patterns of development. Roberts et al. (2017) reexamine whether Latin America and Caribbean countries differ from the rest of the world in the basic stylized facts of urbanization, development, and structural transformation. The analysis makes use of two alternative methodologies for the consistent definition of urban areas across countries that rely on globally gridded population data sets as input: the Agglomeration Index methodology (Uchida and Nelson 2008) and a methodology based on the identification of dense spatially contiguous clusters of population (Dijkstra et al. 2018). We use GHS-POP and Landscan in order to understand the results of the analysis from different population datasets and urban definitions. They find that when levels of urbanization are instead measured using a consistent set of criteria across countries derived from gridded population, this apparent "over-urbanization" of Latin American countries relative to their levels of development and the importance of the agricultural sector to their economies is revealed to be an "urban myth." Another study using population census data, Ferré et al. (2012) examine the inverse relationship between poverty and city size in eight countries. As a robustness check to the heterogenous definitions of urban area across countries, the authors also employ non-administrative definition of urban areas from Global Rural-Urban Mapping Project (GRUMP). Population data not only inform analyses with urban areas but are also used in many cross-country analyses as the

[3] Reports that focus on World Bank regions include: Poor places, thriving people (Middle East and North Africa) and Reshaping Economic Geography in Latin American Caribbean.
[4] For example, population data was used in the Natural Disaster Hotspots analysis by Dilley et al. 2005

denominator. The Data Department at the World Bank hosts one of the primary collections of development data on Development Data Hub as the World Development Indicators (WDI), compiled from officially-recognized international sources. The WDI data are used frequently in cross-country analyses and increasingly are incorporating spatial data summarized at the country level such as the indicator of population living in areas where elevation is below 5 meters (% of total population) from GRUMP (CIESIN 2013). These examples underscore the importance of measurement when making research conclusions and policy.

Integrate population data for new insights

New insights are possible by integrating population data with other data both in response to major current events and in preparation of the future. For example, the Global RApid post-disaster Damage Estimation (GRADE) approach developed at the World Bank aims to address specific damage information needs in the first few weeks after a major disaster and how it complements the more comprehensive post-disaster needs assessment (PDNA) process. GRADE was used to assess direct damages to property; direct damage estimations by economic sector; potential impacts on gross domestic product (GDP) and the economy; and, in the case of earthquakes, estimations of human casualties. The assessment uses gridded population data (e.g., LandScan, Global Human Settlement Layer, WorldPop) along with many other data sources[5] to identify the **distribution of damage** and to **quantify sectoral damages** and **human casualties** (particularly in the case of earthquakes in densely inhabited areas). The GRADE approach has been successfully used after more than four disasters, including Madagascar (after Cyclone Enawo in March 2017), Haiti (after Hurricane Matthew in October 2016), Ecuador (after the earthquake on April 16, 2016), and Nepal (after the earthquake on April 25, 2015) (Gunasekera et al. 2019).

The World Bank not only needs to respond rapidly to natural disasters, but countries need sustainable economic growth and good development outcomes; climate change puts both at risk. The recent World Bank report on **climate induced internal migration**, which is called Groundswell : Preparing for Internal Climate Migration, uses a spatially explicit model to estimate **population distribution changes** applying demographic (e.g. gridded population data), socioeconomic and climate impact data. Some key messages are: the scale of internal climate migration will increase in the absence of mitigation, adaptation and development interventions, and countries can expect "hotspots" with out-migration from climate sensitive livelihood systems and in-migration to cities and more suitable (or less impacted) agricultural areas. Another study examined coastal population, where adaptation to climate change includes addressing sea-level rise (SLR) and increased storm surges in many coastal areas. Mangroves can substantially reduce vulnerability of the adjacent coastal land from inundation but SLR poses a threat to the future of mangroves. This paper quantifies **coastal protection services** of mangroves for population and

[5] Other data include: remotely sensed data for damage and consequences (e.g., UNOSAT, EU-Copernicus), social media updates, local situation reports, and other relief-related information flows, as well as pre-existing scientific, engineering, and socioeconomic datasets and loss damage statistics

economic activity in 42 developing countries in the current climate, and a future climate change scenario with a 1-m SLR and 10 % intensification of storms (Blankespoor et al. 2017).

Validation

Given the importance of measurement and data integration on subsequent analysis, an understanding of data and model quality are imperative to sound policy recommendations. The World Bank produces many household surveys, which are an important source of socio-economic information in the developing world, that can both be used as part of **validation** of model performance and as an **input** to models. Working with Facebook Connectivity Labs and Columbia University, Tiecke et al. (2017) evaluated the model performance of a recent very high resolution population dataset (High Resolution Settlement Layer) with Living Standards Measurement Study survey location data. Another study combines survey data and satellite data to estimate population density and compare their model results with global gridded population datasets such as GPW, Landscan, GHSL, GHS-POP and WorldPop (Engstrom et al. 2019). The opportunities to validate population models with survey data will likely increase as the geo-enabled Computer-Assisted Personal Interview technology, such as the World Bank's Survey Solutions, becomes even more widely used (already in 162 countries) to assist governments, statistical offices and non-governmental organisations in conducting surveys using tablet devices.

Complementarities and final thoughts

The 2015 Policy Research Report on A Measured Approach to Ending Poverty and Boosting Shared Prosperity states that "Data and measurement are vital to achieving the World Bank Group's twin goals of ending poverty by 2030 and promoting shared prosperity, but investments in data quality and national statistical systems are needed to better inform national policy and to help international partners identify gaps and prioritize actions." The report goes on to highlight the complementarity of population data with other data.

I sit in the Data Department at the World Bank where we put data to work for development. We deliver for World Bank Group clients and our staff by improving the quality, accessibility and use of development data through innovation, solid technical expertise and effective partnerships. As we continue to innovate to produce new data products and tools in response to the increasing demand for local level socio-economic statistics, we rely on population data as a key input. For example, we developed a local level high resolution **global dataset of total economic activity and agricultural economic activity** that combined the best available economic statistics with high resolution spatial data with regards to crops, livestock, forestry, hunting and fishing. The gridded data structure provides a method for data integration given the spatial mismatch between natural disaster events and administrative boundaries (UNEP PREVIEW 2010, Gunasekera et al. 2015, Thomas et al. 2019).

Much progress has been made in the development of gridded population data, ease of access, documentation of methods and its applications. As we move forward as a population data

community, key challenges continue such as the need to invest in statistical systems to produce high quality data (including population), applications that are fit for purpose, methods to assess the quality and statistical capacity.

References:

Blankespoor, B., Dasgupta, S., & Lange, G. M. (2017). Mangroves as a protection from storm surges in a changing climate. *Ambio*, *46*(4), 478.

Center for International Earth Science Information Network - CIESIN - Columbia University. (2013). Low Elevation Coastal Zone (LECZ) Urban-Rural Population and Land Area Estimates, Version 2. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4MW2F2J.

Coulibaly, S., Deichmann, U., Freire, M.E., Gill, I., Goh, C., Kopp, A., Lall, S.V., Montenegro, C. E., Packard, T.G., Uchida, H., & Ross-Larson, Bruce [editor] (2008). World development report 2009 reshaping economic geography (English). World development report. Washington, DC : World Bank Group.

Dijkstra, L., Florczyk, A. J., Freire, S., Kemper, T., Pesaresi, M., & Schiavina, M. (2018). Applying the degree of urbanization to the globe: a new harmonised definition reveals a different picture of global urbanisation. *Journal of Urban Economics. In review*.

Dilley, M., Chen, R.S., Deichmann, U., Lerner-Lam, A. L., Arnold, M., Agwe, J., Buys, P., Kjevstad, O., Lyon, B., Yetman, G. (2005). *Natural disaster hotspots: A global risk analysis (English)*. Washington, DC: World Bank.

Engstrom, R., Newhouse, D. L., & Soundararajan, V. (2019). *Estimating Small Area Population Density Using Survey Data and Satellite Imagery: An Application to Sri Lanka*. World Bank Policy Research Working Paper 8776. World Bank.

Ferré, C., Ferreira, F. H., & Lanjouw, P. (2012). Is There a Metropolitan Bias? The relationship between poverty and city size in a selection of developing countries. *The World Bank Economic Review*, *26*(3), 351-382.

Gunasekera, R., Ishizawa, O., Aubrecht, C., Blankespoor, B., Murray, S., Pomonis, A., & Daniell, J. (2015). Developing an adaptive global exposure model to support the generation of country disaster risk profiles. *Earth-Science Reviews*, *150*, 594-608.

Gunusekera, R., Daniell, J., Pomonis, A., Donoso Arias, R. A., Ishizawa, O., & Stone, H. (2018). Methodology Note on the Global rApid post-disaster Damage Estimation (GRADE) approach. *Washington, DC: Global Facility for Disaster Reduction and Recovery*.

Leyk, S., Gaughan, A. E., Adamo, S. B., Sherbinin, A. D., Balk, D., Freire, S., ... & Comenetz, 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*, *11*(3), 1385-1409.

Rigaud, K. K., Jones, B., Bergmann, J., Clement, V., Ober, K., Schewe, J., ... & Midgley, A. (2018). Groundswell: Preparing for Internal Climate Migration. Washington, DC: World Bank.

Roberts, M., Blankespoor, B., Deuskar, C., & Stewart, B. (2017). *Urbanization and development: is Latin America and the Caribbean different from the rest of the world?*. World Bank Policy Research Working Paper. World Bank.

Thomas, T. S., You, L., Wood-Sichra, U., Ru, Y., Blankespoor, B., & Kalvelagen, E. M. F. (2019). Generating Gridded Agricultural Gross Domestic Product for Brazil: A Comparison of Methodologies. *World Bank Policy Research Working Paper*, (8985). Washington, DC: World Bank.

Tiecke, T. G., Liu, X., Zhang, A., Gros, A., Li, N., Yetman, G., ... & Dang, H. A. H. (2017). Mapping the world population one building at a time. *arXiv preprint arXiv:1712.05839*.

Uchida, H., & Nelson, A. (2009). Agglomeration index: towards a new measure of urban concentration. World Bank.

World Bank. (2009). *Reshaping Economic Geography in Latin America and the Caribbean A Companion Volume to the 2009 World Development Report (English)*. Washington DC: World Bank.

World Bank. (2011). *Poor places, thriving people : how the Middle East and North Africa can rise above spatial disparities (English)*. MENA development report. Washington DC: World Bank.

World Bank. (2015). *A Measured Approach to Ending Poverty and Boosting Shared Prosperity: Concepts, Data, and the Twin Goals*. *Policy Research Report.* Washington, DC: World Bank.

BRIAN BLANKESPOOR is the GAIA Team Leader in the Analytics and Tools Unit in the Development Data Group at the World Bank. The Geospatial Analytics, Integration and Applications (GAIA) team works on the production of geospatial development data, tools and methodological innovation for survey and economic analysis. Over the past 12 years at the World Bank, his skill in geospatial, computational methods and economics, coupled with cross-cutting sector experience, has enabled him to contribute to many projects. He is an author of articles in peer reviewed journals on issues of transportation, urban development, disaster risk management, transboundary water and the environment. He is a World Bank focal point to the United Nations Geographic Information Working Group, co-lead of the World Bank's Geospatial Community of Practice and an active member of the POPGRID Data Collaborative. Prior to the World Bank, he held analytical positions in the private sector and World Wildlife Fund. He holds a M.Sc. in Applied Economics from The Johns Hopkins University, a M.Sc. in Geography from Oregon State University and a B.Sc. and B.A. in Computer Science and IDIS Geography and Chemistry, respectively, from Calvin College. He is a certified GIS Professional (GISP) since 2011.