Summary of product development and application of gridded population and settlement data

Sergio Freire
European Commission's Joint Research Centre (JRC)

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

Open and free data sources for open and free information

Data on global population distribution are a strategic resource currently in high demand in an age of new Development Agendas that call for universal inclusiveness of people. These development agreements, their targets and monitoring needs (i.e. indicators) present an opportunity not only for producing more data, but also for improving existing datasets. Traditionally, population distribution grids rely on two major types of input datasets in their construction: (i) georeferenced population statistics and (ii) geospatial proxies associated to the presence of population at the selected spatio-temporal scale (covariates).

Remote sensing technology has demonstrated potential for improvements in both domains mentioned above (Freire et al. 2018), and imagery and methods for information extraction have been evolving towards constituting a more detailed, objective and independent data source on human presence on the Earth surface. The combination of new cost-effective, automated and fully replicable data classification methods (e.g. machine learning) with the synoptic capacities of satellite Earth Observation imagery, made accessible in a public, full open-and-free frame, can contribute to fill information gaps and supplement existing statistics by mitigating some major shortcomings in population data (e.g. currency of data, extreme omission or commission errors). This is especially true in poor, remote, unsafe, disputed, very large, and/or highly dynamic areas of the globe where conventional data gathering and updating is challenging.

At global level, these developments have enabled global mapping of built-up areas and settlements with unprecedented spatio-temporal detail, consistency, and temporal coverage – in essence capturing the local scale with global coverage. These new remotely-sensed global datasets include the Global Human Settlement Layer (GHSL) (Pesaresi and Ehrlich, 2009, Corbane et al. 2017), mining both optical and radar imagery, and the Global Urban Footprint (GUF), generated from radar satellite constellation of TerraSAR-X and TanDEM-X (Esch et al., 2017).

Making these datasets available open and free helps to increase access, promotes transparency, and ensures accountability of the information produced. GHSL's enhanced spatio-temporal mapping of buildings and density of built-up areas offers a suitable proxy for the location of people and, as such, is being used for enhancing multitemporal global population grids, such as GHS-POP (covering the period 1975-2015) and WorldPop (covering years 2000-2020), and to create time series of population estimates that can be used to fill in data gaps between national census surveys that are commonly taken at decadal intervals. In the case of WorldPop, built-up areas are being used in combination with other spatial covariates (Stevens et al. 2015). The current provision of population grids mapping additional demographic variables such as sex and age group seems of high value and potential for the PERN community. Also, the WorldPop project has mapped internal migration for all low and middle income countries.

At continental level, new big earth data analytics are also enabling sustained mapping of built-up areas at high spatial resolutions, such as the European Settlement Map (ESM). Detailed mapping of built-up areas (at 10 m) from ESM was combined with enhanced LULC map to disaggregate detailed census 2011 data for production of the most detailed pan-European grid of resident population (100 m) spanning 43 countries. ESM is now in its third version, with Symbolic Machine Learning (SML) supervised classifier, the PANTEX and the morphological image features being employed to map buildings for 39 countries from multi-sensor very high resolution optical imagery (multispectral and panchromatic) for reference year 2015, totaling ~127TB of data. An added-value to the previous versions is the improved automatic detection of buildings, automatic extraction of water, extraction of building typology (residential vs. non-residential) and an information layer allowing to derive city indicators (Sabo et al. 2019). These are expected to benefit modeling of population distribution.

Other developments in population and settlement mapping are progressing on a country basis, such as the High Resolution Settlement Layer (HRSL) (Tieke et al. 2019). CIESIN is producing and providing population distribution grids at a resolution of 1 arc-second (~30 m) for 2015, based on settlement extents extracted from very high-resolution (0.5 m) satellite imagery by the Connectivity Lab at Facebook, using convolutional neural networks (CNN).

Increasing temporal resolution and integrating unconventional data sources

Census figures register where people reside and usually sleep, although their spatial distribution varies widely over daily, seasonal and long term time scales. Due to the diverse locations of human activities and the displacements they induce, the spatial distribution of population is strongly time-dependent, particularly in metropolitan areas. Especially for disaster risk management in urban areas, it is the population variation during the daily cycle that is particularly relevant.

Twenty years ago, the LandScan project (Dobson et al. 2000) has successfully incorporated the temporal dimension in global population distribution by mapping “ambient population”, a temporally averaged measure of population density that accounts for human activities (such as sleep, work, study, transportation, etc.). Meeting the need for more temporally-explicit representations, subsequent population distribution databases having higher temporal (day- vs nighttime) and spatial detail were developed for the territory of the USA (McPherson and Brown 2004; Bhaduri et al. 2002). Outside of the US, such improvements were mostly limited to a few ad-hoc efforts at local level and even fewer at national scale.

In recent years, a number of emerging data sources and technologies were explored for direct mapping of population or as alternative proxies for its disaggregation, but mainly as proof-of-concept. These include mobile phones, crowdsourcing/Volunteered Geographic Information (VGI), and location-based social media (LBSM). LBSM appeared especially valuable for dynamic population distributing modeling by providing information both on time usage and on target zone characterization (Aubrecht et al. 2017). However promising, a number of issues involve these data and technologies, e.g. sustainability of approaches, data access and ownership, privacy and anonymity, and representation bias. The main question is how to scale up localized case studies to wide geographical areas (continents, world) and provide datasets open and free?

Modeling the spatio-temporal distribution of population is typically subject to a number of trade-offs between geographical coverage, spatial resolution, temporal resolution, and uncertainty. Mapping the dynamics of population is especially challenging for large, multi-national areas and typically involves intensive work of identifying, mining, combining and harmonizing heterogeneous data from disparate sources. This requires significant investment in data integration and harmonization, to overcome limitations of remote sensing and conventional maps.

The combination of conventional (remote sensing, LULC maps) with unconventional geospatial data sources (big data, VGI) is allowing overcoming some of the previous challenges and gaps, such as in the project ENACT (ENhancing ACTivity and population mapping). This project succeeded in producing the first EU-wide (28 countries), consistent, seamless, high-resolution and validated population density grids that take into account both daily and seasonal variations: day- and nighttime grids per month of the year were produced at 1 km resolution, for 2011 (Batista e Silva et al. 2018). This effort was aided by technical advances and increased availability and access to various big data sources that enabled the detailed mapping of land use and activities (Rosina et al. 2018).

Sample applications:

In the context of the **2030 Agenda for Sustainable Development/Sustainable Development Goals** (SDGs), the new generation of geospatial population and settlement data are expanding the range of applications across the domains of research, decision and policy-making. The consistent production of detailed and compatible time series of built-up and population grids has enabled developing global modeling and classification of human settlements in space and time (i.e. the GHS-SMOD), deriving an harmonized delineation of cities, and building the Urban Centre Database - UCDB (Florczyk et al. 2019). In turn, these datasets are meeting requests for supporting SDG monitoring at a range of scales, as in analysis of Land Use Efficiency (SDG goal 11.3.1) (Fig. 1).

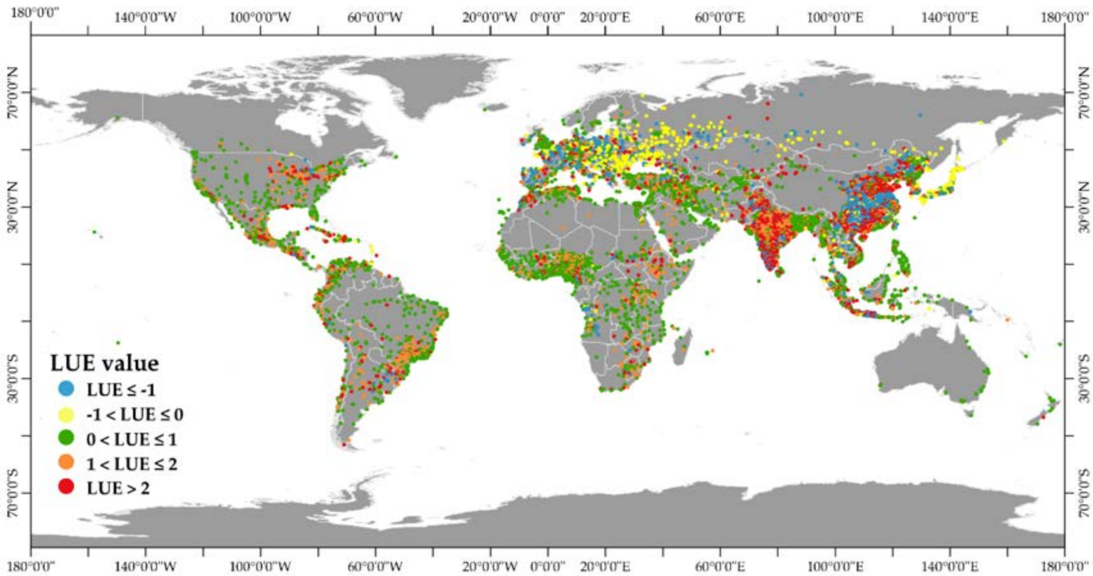

Fig. 1. Land Use Efficiency (LUE) value for circa 10,000 urban centers computed in the period 1990–2015. Source: Melchiorri et al. 2019.

In the context of **risk analysis and DRM**, population dynamics can greatly modify the patterns and assessment of population exposure, particularly for rapid onset hazard events such as earthquakes, tsunamis, and floods. By combining the ENACT population grids with the most recent pan-European hazard data on flood hazard, we are able to map and quantify variations of population exposure, to study their spatio-temporal patterns, and to eventually identify potential daily and seasonal exposure hot spots (Fig. 2).

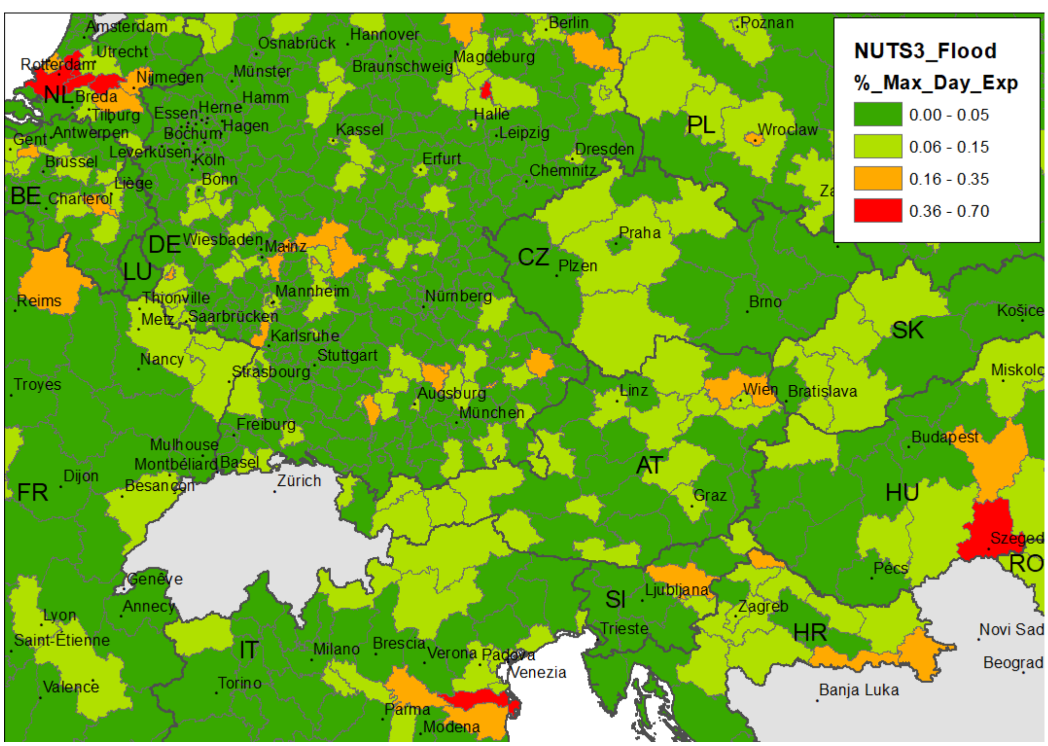

Fig. 2. Population potentially exposed to floods in daytime period as percentage of population in NUTS3, in Central Europe in 2011. Source: Freire et al. 2018b

Despite all the recent advancements, some challenges remain – these include:

- At global scale, major improvements are expected from classifying building functions and estimating their height (in urban areas), and from better detection and mapping of dispersed settlement patterns (in rural areas).
- There is further need for more realistic representations and increased temporal resolution - how to overcome classical trade-offs in modeling population distributions at high spatio-temporal resolutions?
- Assessing, representing (mapping), and communicating uncertainty remain important challenges in population distribution modeling that require more attention.
- Related to previous point, accuracy assessment of population grids suffers from lack of standardized approach or metrics, and scarcity of independent reference data for external validation and indication of accuracy for final products.

References

Aubrecht C., D. O. Aubrecht, J. Ungar, S. Freire, K. Steinnocher, 2017. VGDI - Advancing the concept: Volunteered Geo-Dynamic Information and its benefits for population dynamics modeling. Transactions in GIS, SI: 'Role of Volunteered Geographic Information in Advancing Science', 21(2):253–276.

Bhaduri B, Bright E, Coleman P, Dobson J (2002) LandScan: Locating People is What Matters. Geoinformatics, Vol. 5, No. 2, pp. 34-37.

Batista e Silva, F., K. Rosina, M. Schiavina, M. Marin, S. Freire, L. Ziemba, M. Craglia, & C. Lavalle, 2018. From place of residence to place of activity: towards spatiotemporal mapping of population density in Europe. Proceedings of the 21th AGILE Conference on Geographic Information Science. Lund, Sweden, June 12-15, 2018.

Corbane Christina, Martino Pesaresi, Panagiotis Politis, Vasileios Syrris, Aneta J. Florczyk, Pierre Soille, Luca Maffenini, Armin Burger, Veselin Vasilev, Dario Rodriguez, Filip Sabo, Lewis Dijkstra & Thomas Kemper (2017) Big earth data analytics on Sentinel-1 and Landsat imagery in support to global human settlements mapping, Big Earth Data, 1:1-2, 118-144.

Dobson JE, Bright EA, Coleman PR, Durfee RC, Worley BA (2000) A Global Population Database for Estimating Population at Risk. Photogrammetric Engineering & Remote Sensing 66(7):849-857.

Esch, T., Heldens, W., Hirner, A., et al. (2017). Breaking new ground in mapping human settlements from space – the global urban footprint. ISPRS Journal of Photogrammetry and Remote Sensing, 134(December), 30–42.

Florczyk, Aneta; Corbane, Christina; Schiavina, Marcello; Pesaresi, Martino; Maffenini, Luca; Melchiorri, Michele; Politis, Panagiotis; Sabo, Filip; Freire, Sergio; Ehrlich, Daniele; Kemper, Thomas; Tommasi, Pierpaolo; Airaghi, Donato; Zanchetta, Luigi (2019): GHS Urban Centre Database 2015, multitemporal and multidimensional attributes, R2019A. European Commission, Joint Research Centre (JRC) [Dataset] PID: https://data.jrc.ec.europa.eu/dataset/53473144-b88c-44bc-b4a3-4583ed1f547e

Freire S., M. Schiavina, A. J. Florczyk, K. MacManus, M. Pesaresi, C. Corbane, O. Borkovska, J. Mills, L. Pistolesi, J. Squires & R. Sliuzas, 2018a. Enhanced data and methods for improving open and free global population grids: putting ‘leaving no one behind’ into practice. International Journal of Digital Earth, DOI: 10.1080/17538947.2018.1548656

Freire S., M. Schiavina, Batista e Silva, F., K. Rosina, 2018b. Advancing exposure and risk assessment in the EU by modeling population distribution in daily and seasonal cycles. EGU (European Geosciences Union) General Assembly 2018. Geophysical Research Abstracts, Vol. 20, EGU2018-19691-1. Vienna, Austria.

McPherson TN, Brown MJ (2004) Estimating daytime and nighttime population distributions in U.S. cities for emergency response activities. Preprints: 84th AMS Annual Meeting, AMS, Seattle, WA, 10 pp.

Melchiorri, M.; Pesaresi, M.; Florczyk, A.J.; Corbane, C.; Kemper, T., 2019. Principles and Applications of the Global Human Settlement Layer as Baseline for the Land Use Efficiency Indicator—SDG 11.3.1. ISPRS Int. J. Geo Inf. 2019, 8, 96.

Pesaresi, M., and D. Ehrlich. 2009. “A Methodology to Quantify Built-up Structures from Optical VHR Imagery.” In Global Mapping of Human Settlement Experiences, Datasets, and Prospects, ch. 3, edited by P. Gamba and M. Herold, 27–58. Boca Raton, FL: CRC Press.

Rosina, K., F. Batista e Silva, P. Vizcaino, M. Herrera, S. Freire & M. Schiavina, 2018. Increasing the detail of European land use/cover data by combining heterogeneous data sets. International Journal of Digital Earth, DOI: 10.1080/17538947.2018.1550119

Sabo, F., Corbane, C., Politis, P., Pesaresi, M., & Kemper, T., 2019. Update and improvement of the European Settlement map. 2019 Joint Urban Remote Sensing Event (JURSE), 1–4. https://doi.org/10.1109/JURSE.2019.8808933

Stevens, F. R., Gaughan, A. E., Linard, C., Tatem, A. J., 2015. Disaggregating census data for population mapping using Random forests with remotely-sensed and ancillary data. PloS One 10(2).

Tieke, T., X. Liu, A. Zhang, A. Gros, et al. 2017. Mapping the world population one building at a time. Available at https://arxiv.org/abs/1712.05839