Question & Answer Session Part 2

Please type your questions in the Question Box. We will try our best to get to all your questions. If we don't, feel free to email Brock Blevins (brock.blevins@nasa.gov).

Question 1: Do we have older age (>65y) population of the world (especially India) as a population grid data?
Answer 1: GPWv4 has age/sex grids by 5 year age groups up to >85 (where available) for the year 2010.
https://sedac.ciesin.columbia.edu/data/set/gpw-v4-basic-demographic-characteristics-rev11
WorldPop has global age and sex data as well. They break down pop into 5-year age classes (they also have a <1 age class).
https://www.worldpop.org/project/categories?id=8
HRSL has age and sex data for broad categories: under 5, working age, women of childbearing age, elderly.
https://dataforgood.fb.com/docs/high-resolution-population-density-maps-demographic-estimates-documentation/

Question 2: How did you validate those Demographic data within GPW?
Answer 2: From the data set homepage: "To estimate the male and female populations by age in 2010, the proportions of males and females in each 5-year age group from ages 0-4 to ages 85+ for the given census year were calculated. These proportions were then applied to the 2010 estimates of the total population to obtain 2010 estimates of male and female populations by age." The age estimates are consistent with the total population which is adjusted to the United Nations World Population Prospects. Note that GPW uses an intentionally simple disaggregation method, as a first approximation of population distribution based only on census data and geography. Validation of estimates from GPW and other sources is difficult since "actual" data on population location and demographic characteristics are difficult to obtain and intercompare.

Question 3: Are there methods available in R for spatial disaggregation of counts?
Answer 3: Yes, there are methods in R for this type of work.
WorldPop has example code for their models on GitHub

https://github.com/wpgp

Question 4: What is the best population dataset out there suitable for flood mapping in urban coastal cities like Accra, Ghana vs flood mapping in island countries like Mauritius which consist of both urban and rural areas. What are some key limitations to consider when using the public dataset for both cases?

Answer 4: There's not really a clear answer of what is "best"--different data sets vary due to different inputs and methods, which can vary by country or region. The scale of analysis is also important, compared with the resolution of the underlying census data and covariates used in the models. We are beginning to find that all of the gridded population data sets may underestimate populations actually present in urban slums. Validation data for slums can be lacking. The POPGRID Viewer allows you to see how the global data sets differ (variation) for specific areas of interest. For a more detailed discussion, see the publication Leaving No One Off the Map mentioned in the first session and available from the POPGRID.org site.

Question 5: What data should I use in a project that involves tea production management?

Answer 5: This depends on the question(s) you are addressing, e.g., whether you are interested in the population living in or near tea production areas, or looking at potential markets accessible to tea producers, or some other aspect of tea production.

Question 6: I am currently using Worldpop raster data to assess the population impacted by gully erosion in urban areas of DR Congo. I have noticed that in large cities in DR Congo the population pixel values are highly differentiated (which is good) while for small affected urban areas it is the opposite and the population is grossly underestimated. Is there a way to overcome this problem?

Answer 6: Maybe experiment with using unconstrained versus constrained Worldpop data. There are different advantages to each method. For example, the WP website states that top-down constrained datasets should be used "where accurate identification of rural populations and uninhabited areas is a priority". Just an idea! (https://www.worldpop.org/methods/top_down_constrained_vs_unconstrained)

GHS POP also allocates population only to satellite derived built up areas which may result in more accurate counts in urban areas.

Question 7: What is the Geopackage format and what are its merits and demerits?

Answer 7: A geopackage is a format for transferring geospatial information. An SQL container. In my limited experience, they can be difficult to work with but can contain TONS of data. The OSM roads data used in the RAI presentation were downloaded as geopackages. You can work with geopackages in Python (i.e. geopandas package) and arcpro.

Question 8: How do you delineate urban and rural areas? Is there a global acceptable definition for urban and rural areas?
Answer 8: Every country has its own definition for urban and rural areas. This can affect cross-country comparisons for indicators that are based on these delineations. This is why using a global settlement model like GHS S-MOD is useful. It is based on population counts and densities and satellite-derived built up areas.

Recently the UN accepted the Degree of Urbanization (DEGURBA) as the standard for delineating cities/urban areas. This methodology "classifies local units as 1) "cities" or "densely populated areas", 2) "towns & semi-dense areas" or "intermediate density areas" and 3) "rural areas" or "thinly populated areas" based on population density, population size and contiguity using 1 km2 grid cells."
(https://unstats.un.org/unsd/statcom/51st-session/documents/BG-Item3j-Recommendation-E.pdf).
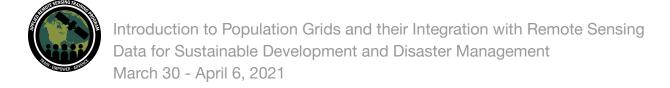
The UN will often refer to metropolitan areas, urban agglomerations, or city proper.

SMOD is a refinement of the Eurostat Degree of Urbanization Method. It has more classes at L2 which can be aggregated to L1 (as outlined in the data set documentation) which match the DEGURBA units.

Note that the DEGURBA approach was developed by the EC Joint Research Centre under the auspices of the GEO Human Planet Initiative.

Question 9: What does the UN-adjusted World Pop dataset mean? Are the other population datasets also UN-adjusted?
Answer 9: UN-adjusted means that the population totals in the grids (as summarized by a zonal statistics operation) are checked against official UN national totals reported in the World Population Prospects (every two years) and an adjustment factor is calculated and applied to the population estimates in every grid cell so that the national totals in the final grid match the UN totals. Not all grids are adjusted to the UN. Both

GPW and WorldPop provide population grids that are either adjusted or not adjusted. This is usually explicit in the data product name.

Question 10: Hello, do you have examples of using gridded population data to assess disaster damage/risk and its impact by gender?
Answer 10: On the topic of inferring potential or actual building damage, the use of gridded population for inference of number of buildings will ultimately yield damage in terms of number of buildings by damage state. There are examples of overlaying that information onto demographic data to explore the impact by gender after a disaster, for example following the 2015 earthquake in Nepal, but it is not typically addressed intrinsically in this type of analysis.

That said, if you go to the SEDAC Citations Database, https://sedac.ciesin.columbia.edu/citations-db, you can search on publications that cite "gpw-v4-basi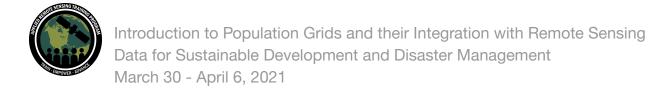c-demographic-characteristics-rev10" and "gpw-v4-basic-demographic-characteristics-rev11", which contain the gender and age structure data. This currently produces 15 citations, with several that appear to be hazard and health applications, e.g., on heatwaves, tornadoes, floods, air quality, and COVID-19. One would have to look at the cited papers to find out if the authors used the data to address gender issues.

Question 11: Is there foundation of buildings considered to come up with building exposure information?
Answer 11: Foundation can be considered statistically at an aggregate level if needed. For example, the first floor elevation above the foundation is often used for estimating damage in flood modeling, which implicitly includes the height of the foundation. Certain types of foundations are also more vulnerable to earthquake damage.

Question 12: Is there the real earth surface (slope, land cover etc.) considered to buffer the distance of 2 km. of all season road?
Answer 12: Not currently, but that is definitely something to look into further. Currently it is just a 2 km geodesic buffer around the subset of all-season roads. In the World Bank methodology, road condition data is used to get at the "real surface" aspect of this indicator. Unfortunately there is no global road condition data set and this information is lacking in the OSM data. We tried to use the fclass attribute as a proxy for our global analysis.

Question 13: Can a type of dwelling unit be used as a proxy for poverty levels in developing nations?
Answer 13: The type of dwelling as inferred by the development pattern is used as an indicator of the estimated replacement cost of structures. This is largely based on the negative correlation between the construction quality of housing and poverty that we see universally. I think that there would have to be adjustment on a country by country basis, but that a case could likely be made for developing such a proxy, but it is not used currently as an indicator.

Ideamaps Network: https://ideamapsnetwork.org/ is doing some interesting work in this area.

Question 14: Which resource do you recommend for the imperviousness density, and how to obtain it from any US environmental gov. agency? Thanks!
Answer 14: SEDAC provides access to a data set on impervious surfaces (area) developed by NASA scientists:
https://sedac.ciesin.columbia.edu/data/set/ulandsat-gmis-v1.

Question 15: Are these algorithms open code? If yes, where can we access them?
Answer 15: The algorithms discussed by ImageCat in the session are not open, but many of the datasets are. See: https://meteor-project.org/data/
Typically, each project requires a fresh methodological approach, and it is more an art that an algorithm. ImageCat aims to be transparent with methods and has developed new ways of communicating methods through metadata.

Question 16: How can we identify people in households through EOs data?
Answer 16: This needs to be determined through census. Building footprint data can effectively override estimates of people per households regionally, or EO can be used to adjust, but typically estimates of persons per households are used directly.

Question 17: I have a question, as buildings are very heterogeneous in nature and their unit replacement cost varies a lot depending on spatial location, what do you think should be the factor (or say line) delineating the areas that can be considered to have similar unit rates for replacement cost?
Answer 17: The goal in loss estimation is to try to characterize typical building stock. In my opinion, any building types that represent less than 5% of buildings are not worth characterizing, in most circumstances.

Introduction to Population Grids and their Integration with Remote Sensing
Data for Sustainable Development and Disaster Management
March 30 - April 6, 2021

Question 18: My question is how do we best deal with different resolutions of ancillary variables vis-à-vis population gridded data?
Answer 18: Data fusion is the key to the art. Very generally, integrate statistically by development pattern. Higher resolution ancillary data are essentially used to help model/infer potential population variations at finer spatial scales.

Question 19: Is data available from Hispaniola?
Answer 19: Data for Haiti is available from https://meteor-project.org/data/.

Question 20: I just want to make sure I correctly understood one can "simply" add its own particular data to POPGRID and perform analysis. Also, one could exchange data? Let's say replace WPP data for own personal population estimates?
Answer 20: To clarify, the POPGRID viewer only allows users to identify a specific area of interest (polygon, circle, rectangle) and obtain estimates of population in that area for the 6 major global-scale georeferenced data sets (at a couple of different points in time). We do not currently offer any service to perform regridding using user-defined parameters or census data inputs, or to, say, analyze a user's dataset in conjunction with population data (e.g., to create population-weighted averages).
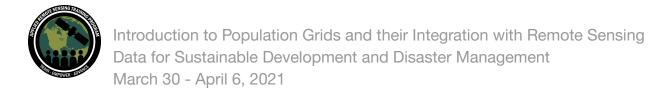
Question 21: My first question regards the possibility of obtaining nocturnal Landsat datasets. Could you suggest any website in particular? And could the thermal constants of OLI imagery return useful for any application? Are the radiometric corrections also required as pre-processing stage?
Answer 21: I'm not an expert on Landsat, but this website seems to describe how to search for nighttime Landsat imagery:
https://www.usgs.gov/faqs/how-do-i-search-and-download-ascending-nighttime-landsat-scenes?qt-news_science_products=0#qt-news_science_products
If you are interested in night-time lights imagery, you may want to check out this NASA site that provides access to data from the Visible Infrared Imaging Radiometer Suite (VIIRS) instrument on the Suomi National Polar-orbiting Partnership (NPP) satellite:
https://blackmarble.gsfc.nasa.gov/.

Question 22: In regards the application of PM10 Landsat measurements. Could you provide any useful algorithm/resource, whose exposure does not work on large scale rather for small urban cases?

Answer 22: This might be of interest:
https://appliedsciences.nasa.gov/sites/default/files/D2P3_AODPMEx.pdf

Question 23: Why does the metric for RAI need to be in relation to "all-season roads" if not all locations experience all seasons?
Answer 23: An all-season road just means that the road is "motorable all-year round by the prevailing means of rural transport". They are also often referred to as "all-weather roads". The term itself is a bit misleading, but it really is just referring to durable, reliable roads. (https://unstats.un.org/sdgs/metadata/?Text=&Goal=9&Target=9.1)

Question 24: Who created/maintains the GADM data?
Answer 24: The data set originated through a collaboration between Robert Hijmans at UC-Davis, International Rice Research Institute, and UC-Berkeley. The website doesn't currently indicate the data creator or who is currently maintaining or updating it, but does indicate: "Data hosting provided by the Center for Spatial Sciences at the University of California, Davis." Robert Hijmans research website at UC Davis is not responding.

Some info from the Esri hub entry for version 2 of the data set:
Source: Global Administrative Areas, www.gadm.org, version 2.0, January 2012 by Robert Hijmans (UC-Davis), International Rice Research Institute, and UC-Berkeley.

GADM was developed to support various activities, including georeferencing of textual locality descriptions (the BioGeomancer project) and for mapping census type data. The database is currently developed by Robert Hijmans. Major contributions have been made by Nell Garcia, Arnel Rala, and Aileen Maunahan at the International Rice Research Institute and by John Wieczorek and Julian Kapoor at the University of California, Berkeley, Museum of Vertebrate Zoology.

https://hub.arcgis.com/datasets/bff789530b964a76938de17824165ceb_2?geometry=149.063%2C-89.998%2C-149.063%2C-79.510

The data can be downloaded here: https://gadm.org/data.html

Question25: When using GADM, are topographical maps included?
Answer 25: GADM is administrative data only. No additional topographic data is included. The dataset is topologically integrated (no gaps/overlaps).

Question 26: Question to Bob Chen: Are the GPW data for different years comparable in terms of pop growth or were there methodological changes?

Answer 26: Within GPW version 4, the gridded data sets for 2000, 2005, 2010, 2015, and 2020 (a projection) are treated as consistently as possible, so that overall population changes correspond to known population trends reflected in the available censuses. This does not eliminate the possibility of inconsistent time series at high resolution resulting from changes in administrative boundaries and census reporting changes. Note that other data providers deal with time series differently, since, for example, model covariates are not always consistently available over time.

Question 27: When computing RAI what are the processes in the methodology that take the most time to compute?
Answer 27: Buffering the roads and dissolving the buffers takes the most time. Some countries can take multiple days to finish this step.

Question 28: Are any subsurface structures included in DRR processes?
Answer 28: I'm not 100% clear what is meant by subsurface structures, but typically the objective is to capture the general building stock, rather than exotic or unique structure. Sports stadiums and convention centers, for example, may be excluded. In some cases, as with flood, the number of floors under ground level might be important when analyzing individual structures or characterizing the building stock as a whole. For example, percentage of residential buildings with basements is nearly zero in California, and high in the midwest of the United States and other areas with cold climates.

Question 29: How could one use the data collated by POPGRID?
Answer 29: POPGRID tries to provide easier access to the different data sets produced by different groups. Each group has developed their data to meet their own community's needs, and users from those communities may wish to consult the source's site for guidance on the use of their specific data sets. POPGRID aims to help users who aren't sure which data set is appropriate for their needs, or who wish to use multiple data sets to address an issue, by providing more consistent metadata and documentation and tools for intercomparison across data resources, and by encouraging validation and intercomparison efforts.

Question30: How do you see this research feeding into policies and evidence based decisions by relevant stakeholders? Is there a dissemination strategy in place for it?
Answer 30: Loss estimation research is mature in the sense that exposure products are developed into disaster loss estimation systems that inform mitigation/adaptation

strategies and real time response, as well as insurance premiums. Mitigation strategies, for example. The data is disseminated through a variety of web portals, including the UKSA METEOR project website, but are more frequently built into applications by stakeholders themselves.

I can add that one of objectives of the POPGRID Data Collaborative is to promote awareness and uptake by stakeholders in the sustainable development community. SDSN TReNDS is helping the Collaborative members to better match the supply of data with the demand and need for these types of data.

Question 31: How do you integrate AI protocols using sensor systems to validate population data estimated from satellite imagery (e.g., roads, buildings)?
Answer 31: AI is used in the segmentation process to classify the development patterns in our research. It is also being used increasingly to extract building-level data in lieu of detailed site surveys (e.g., masonry buildings, soft-stories, number of stories). I am not familiar with any research into using AI for validation.

Question 32: Are the datasets available for any of this research?
Answer 32: The UKSA project METEOR has data for download covering 47 developing countries. https://meteor-project.org/. Access to the range of georeferenced population data discussed previously is available through https://popgrid.org.

Question 33: Would the zoning data of the cities be useful to estimate the development patterns or at least one could limit heights of buildings, or maybe as an input for classification algorithms to estimate buildings?
Answer 33: Zoning data, when available, can be used either to replace development patterns or support/train them. When codes are followed, they can inform vulnerability. EO data is most valuable when a risk study is regional.

Question 34: I am looking for advice on the Modifiable Areal Unit Problem (MAUP). What are some of the ways to address the MAUP in your research?
Answer 34: We follow a stratified sampling approach whereby we either disaggregate population to a gridded level using a dataset like HRSL or use a product like WorldPop. We then sample building types by our custom building development patterns and extrapolate statistical properties at the gridded level. The data is not aggregated back up at that point. In effect, the whole process is an attempt to intelligently disaggregate using EO data in a way that undoes the bias introduced by large and uneven administrative units. That said, this problem is not addressed explicitly.

Question 35: For nations/regions where in general population based information on various domains remain largely absent, what has been the general approach to fill the data gap (for example, in assessing the risks and vulnerability of population from climate extremes)?

Answer 35: This work seeks to provide a fundamental layer on population distribution that can fill in gaps due to limited censuses and other sources. By creating these global scale georeferenced population layers, we are trying to identify where people and settlements are, using EO data and other sources. These pop layers can be the first step in efforts to develop additional socioeconomic layers important for assessing climate risks and vulnerability, e.g., poverty, proximity to coastlines, access to resources, or quality of infrastructure.

Question 36: I think the satellite-based approaches are very realistic and practical to estimate the risk of some disaster... Have you ever applied those to predict the level of pollution worldwide? For instance, have you ever estimated the urban infra maintenance priority by using the pops >> building index, the RDI and the subsurface infrastructure status, etc.?

Answer 36: We (ImageCat) have not because of where we sit in the commercial sector (pollution). But we have looked at hazardous materials, sources, etc. We do a lot of work looking at lifelines (electrical power, pipelines)

Question 37: Is the earliest year of the population grid? 2015?

Answer 37: There is a lot of variation. See: https://www.popgrid.org/data-docs-table1.