Question & Answer Session 1

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: May I know the difference of this between SEDAC?
Answer 1: SEDAC is the NASA-funded Socioeconomic Data and Applications Center, operated by CIESIN. It produces the Gridded Population of the World (GPW) data collection and partners with other organizations to produce other gridded population data (among many other data and analysis projects). POPGRID is a collaborative that CIESIN initiated, and which SEDAC supports.

Question 2: Is there a method to extract a specific area of gridded data and regrid the small area using Python?
Answer 2: It is possible, reallocation is usually accomplished by using one or more additional layers to inform the re-gridding at a more detailed level. Using arcpy or QGIS Python libraries you can extract the total population of a small area from the grids, then re-grid using a model of your choosing (dasymetric, for example). Different research studies have published their codes which can be used by the user community. We will provide more detail.

Question 3: Could you please be more specific about missing people in the census?
Answer 3: There are different levels of error in census. All population censuses have a certain level of error, meaning they do not enumerate all of the target population (e.g., everyone living in a country), not reaching 100% coverage. This can have different reasons such as homeless individuals with no permanent address, unregistered births and migration, undocumented immigrants, temporary settlements missed by census takers, or conflict/security situations that inhibit census or survey efforts.

Question 4: What is the resolution of the grids? What are the accuracy levels of the data?
Answer 4: The resolution varies by product from the most detailed, 1 arc-second (approximately 30m) to 30 arc-seconds (approximately 1km). The POPGRID website lists the spatial resolutions of the various gridded population data products. Please see some of the sources on uncertainty. Validation of population grids is very challenging

but POPGRID has made some effort on intercomparisons of population grids to better understand sensitivities if different data products are applied for the same analysis. Some examples are provided at the POPGRID website.

Question 5: How frequent is the data updated considering the fact that different countries conduct their population and housing censuses at different years?
Answer 5: This varies by product: GPW, for example, is generally updated every five years or so, incorporating all new decennial censuses that become available; other products are produced annually or biannually, such as Landscan, the World Population Estimate (WPE) and the High Resolution Settlement Layer (HRSL). The frequent updates are based on improvements in the method and on some ancillary data sources that are updated more often, such as remote sensing derived measures (e.g., settlement extent). All of the datasets adjust the data from the census years available to the target year using simple demographic interpolation or projection.

Question 6: What is the simplest method for grid reliability analysis?
Answer 6: Some of the data products publish reliability measures, although they differ. Esri's World Population Estimate has a reliability index that ranks the data quality on the pixel level. WorldPop produces a pixel-level standard deviation measure (uncertainty dataset). GPW provides a pixel-level measure of the mean administrative unit area used to grid the population data. These can be used to assess the reliability of the datasets for specific areas. More generally, the literature contains a number of papers that assess the relative accuracy of different global data sets by comparing them to an alternative population source, typically very detailed census data that are not publicly available. See for example the paper, "Estimating small-area population density in Sri Lanka using surveys and Geo-spatial data" by R. Engstrom (2020, in PLOS One). https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0237063. See also Leyk et al 2019 for a short summary (https://essd.copernicus.org/articles/11/1385/2019/)

Question 7: How are the results of the various methods of population estimation validated?
Answer 7: See question above on reliability. Validating population grids is challenging but different data providers are publishing error measures with their products. POPGRID is making efforts on data intercomparison to better understand the sensitivity if a user carries out the target analysis using different population grids (see e.g., Tuholske et al. 2019). Furthermore, the concept of fitness for use is an important

approach to assess “relative data quality” which refers to the data quality criteria in relation to the target application or purpose (see e.g., Leyk et al 2019 for a short summary).
Validation can be done when fine-resolution census or household survey data are available. Such validation analysis can be found in the dasymetric mapping literature and is usually carried out as a cross-comparison. The accuracy of population estimation methods always depends on the accuracy of the population input data (census) and the ancillary data used, as well as the model performance.

Question 8: Could you provide some examples of how popgrid data can be used for climate change adaptation and mitigation?
Answer 8: The report “Leaving no one off the map” (especially chapter 1) provides examples of the use of gridded population data https://www.unsdsn.org/leaving-no-one-off-the-map-a-guide-for-gridded-population-data-for-sustainable-development . There is more recent work done on modeling pop movement in 2030 and 2050 based on drivers, showing areas more likely to attract population. Groundswell: Preparing for internal Climate Migration https://openknowledge.worldbank.org/handle/10986/29461

Question 9: Can the population data be used through the google earth engine?
Answer 9: Some of the layers are available in Google Earth Engine (GEE): WorldPop, GPW and GHS-POP are published in GEE. HRSL is not currently available, but it should be published this year. The other datasets have licensing restrictions and as a result are not published on GEE, but you can always plug in your own datasets into GEE.

Question 10: How often is the data being referred to updated?
Answer 10: Please refer to question 5 above.

Question 11: I wonder what population information is used in these global data sets. If countries do not have updated census data, how do you estimate the population with this level of granularity? What is the interval of confidence of the estimates if you do not have block level data?
Answer 11: The quality of the population grid based estimates always depends on the input data quality. Older censuses used to create the population grid layer for that country naturally means that the grid estimates have higher error. There are research efforts on population projection and population forecasts with the goal of developing

methodological frameworks for these problems. For some datasets, the time stamp of the input data is available.

Question 12: Willing to stand corrected but presumably it's dependent on how often a sensor is run? e.g in the UK, every 10yrs? Would that be about right?
Answer 12: Basically, the quality of modeled and integrated data is as good as its inputs. The census data are the central input data source and the quality of the gridded layers is a function of these input data. Only a few countries run their census more frequently than once per decade. The further away from a census year, the more uncertainty there will be in the population estimates for small areas.

Question 13: Does GRID methodology can future estimate other variables than population counts?
Answer 13: Some data providers (e.g., SEDAC, WorldPop, and Facebook have developed estimates of additional demographic variables, e.g., age/sex breakdowns, to supplement population counts. For example, GPW includes age/sex breakdowns for the year 2010, but it should be noted that the age/sex data available from many censuses were only available for higher administrative levels and therefore had to be downscaled to the administrative level used in gridding.

Question 14: Are the pop data from Facebook available in the Mekong region yet?
Answer 14: You can see the available regions here: https://columbia.maps.arcgis.com/apps/View/index.html?appid=ce441db6aa54494cbc6c6cee11b95917

Question 15: Urban areas have multi-story buildings with varying family sizes. Moreover, commercial buildings operating round the clock with fewer employees will also consider a density. How reliable could the estimates be and how do you adjust coefficients for different types of built-up areas?
Answer 15: It is important to recognize that some population data sets focus on estimating residential population distribution (i.e., the population “at night”) and others on the population during a normal workday (i.e., the daytime population). If the data are based on detailed census inputs in a “top-down” approach, where the population is not allocated outside of the administrative unit, these areas can be assigned little to no population (it is a nighttime census population estimate). Landscan and newer products from WorldPop relax this top-down restriction and allocate people to built-up areas that are commercial or industrial. Covariate datasets (level of built density,

luminance from satellite-derived data, other measures) serve as proxies for commercial and other built-up areas that population is allocated to. There are efforts underway to include building volume derived from RADAR data in urban areas in models (HRSL, for instance, is exploring this approach) but nothing has been released to date.

Question 16: Can I use POPGRID to compare with natural disasters like floods and fires? What is the most appropriate dataset to do that?
Answer 16: Disaster assessment is one very important target application domain and you can find some studies in the literature where different population grids have been applied for such applications. It is important to understand the underlying purpose and quality of the different population grids, which may vary not only between data sets but also between regions (e.g., urban vs. rural areas; flat vs. steep regions; cloudy vs. cloud-free climates). It is also important to note that some population models may utilize ancillary data that may overlap or correlate with factors that may be relevant to disaster vulnerability, e.g., proximity to roads or slope. In addition to spatial resolution, the frequency and consistency of estimates over time, the date and completeness of the most recent census inputs, and the availability of demographic breakdowns might be factors in the selection of suitable data.

Question 17: What is the name of the same project for the buildings?
Answer 17: The POPGRID Data Collaborative includes data providers who have developed data on building footprints and potentially other building characteristics. However, the availability of these data varies.

Question 18: What was the address of the web map server in the demo (to connect through QGIS)?
Answer 18:
https://sedac.ciesin.columbia.edu/arcgis/rest/services/ciesin/popgrid_counts/MapServer

Question 19: Is it possible to obtain estimated population counts by gender?
Answer 19: Yes, WorldPop, GPW4, and the HRSL all include age and sex breakdowns.

Question 20: How can we know the teams behind POPGRID? Like what kinds of scientists contribute to developing this system?

Answer 20: You can find the list of partners and links to their websites, as well as contact information here: https://www.popgrid.org/about-us. POPGRID members come from many different disciplines including geography, demography, public health, engineering, Earth science, development, statistics, and data science.

Question 21: What is the criteria in choosing the model? Areal or dasymetric?
Answer 21: The most important factor is the intended use of the population surface. For example, if you want to examine population distribution as an independent variable, the model should not have any covariates that you are studying. Additionally, machine learning models (Random Forest, Neural Networks) are known as “black-box” models: that is, it is difficult to understand how the multi-dimensional data are combined in the model to produce a population prediction surface; this is less useful for inference related to why populations are distributed as they are, but these models can be very useful for predicting population in areas with less detailed or outdated census data. Dasymetric models are easier to understand and do not require the detailed training data that are required for machine learning and that may not be readily available (and are expensive and time consuming to collect). You do not need to train a dasymetric model. Finally, areal weighting is the simplest model to both understand and implement, and requires the least amount of input data (no covariates, just census data and sometimes Boolean surfaces such as water and settlement extent). Machine learning is less useful for inference but good for fine tuning a model. This is complex, so diving into the papers on slide 41 of the presentation would be useful.

Question 22: If comparing and sharing data are purposes of grids, we should keep the size of the cells fixed. If so, what would be the best grid size for population and dwellings applications at a global level?
Answer 22: For global data, it is common to see data at 30 arc-seconds (or roughly 1km at the equator). This is in part because this resolution makes for a global file that is easy to process quickly on almost any computer, and is a good resolution for global and large region analysis. More detailed resolutions are used for products with more detailed input data: Landsat and Sentinel data (30m and 10m respectively) are often used and the resulting product aggregated to 100 or 250m (WorldPop and GHS-POP). The HRSL, which uses 50cm - 2m input optical satellite data to detect settlement extent, is aggregated to approximately 30m. There is no standard for resolution in global products; each producer chooses the output resolution based on their view of user needs, data inputs, and model considerations. Users need to use their own judgement about whether any specific data set meets their needs and requirements on

multiple dimensions (e.g., spatial resolution, temporal frequency/currency, file size, regions of interest, population characteristics available).

Question 23: What product is best in estimating population in rural areas in developing countries, especially in areas covered by forest (where houses are often not seen in medium resolution imagery nor emit significant lights at night)?
Answer 23: No product currently does an excellent job in mapping rural population in heavily forested areas since small structures are extremely difficult to locate in optical imagery. The HRSL uses 50cm to 2m input optical data (Maxar imagery, formerly known as DigitalGlobe) and finds many more small settlements than the other gridded population surfaces that rely on lights and moderate-resolution satellite imagery. Omissions in heavily forested areas are still a problem in the HRSL data.

Question 24: I don't know anything about coding, Python, so do we need to know it thoroughly? As he used a lot of codes.
Answer 24: The population grids can be used in simple GIS based mapping efforts and GIS based overlay/analysis without any coding. The different examples you saw just demonstrate that you can use a whole range of tools including such libraries for analysis. It is possible to conduct a simple visual comparison using the POPGRID Viewer. GIS software (QGIS, ArcGIS) can be used to conduct more sophisticated analysis without any code. The data structure is pretty simple so use in QGIS is useful.

Question 25: How do you tend to solve the problem of differences in population estimates from the POP grid and the actual population in slums mentioned earlier? Slum areas underestimated
Answer 25: The use of survey data that estimates population or household size is being evaluated as a way to improve estimates in slum areas. These data are not available globally, so refined models that can predict the higher densities in slum areas are needed. Unfortunately, slum area densities are poorly known, so any updated model could still underestimate the actual population in these areas.

Question 26: What ancillary data can be taken as a proxy for sparsely dense grid and vice versa? Since, road networks, lightings, built up areas, etc. may be ambiguous to represent actual density (e.g. Slums with lesser road density will have greater pop than the organized office area with more lighting and greater road density and built-up area)!
Answer 26: It is similar to question 25 but it relates back to using survey data in order to represent these things. Research is still being conducted in this area. Relationships

between population density and potential covariates are not well understood and may vary between different regions.

Question 27: How accurate is the low elevation coastal zone data for a small island developing state like the Bahamas? Which dataset would you recommend being used?
Answer 27: SRTM is a consistent global dataset, but dated. Other elevation data are being made available from SEDAC (e.g., Altimeter Corrected Elevations, version 2, ACE2) as well as LECZ estimates (version 3 to be released soon). For Small island states it would be best to use local data, or LIDAR.

Question 28: Could you explain in more detail how the population grid accounts for areas which were not part of census data collection, but where the grid actually gives pixels with population? Will the official census data for example be also distributed over these areas, which could result in less population where the data was actually collected?
Answer 28: These are possible consequences of the population allocation process. Depending on the ancillary data used, population will be allocated to different places within the study area such that the input population is reached (pycnophylactic property). That may mean that we create errors locally which are then propagated through the target application.

Question 29: Connected to that: To what degree is it reasonable to use population grids for example in disaster management in informal settlements (most probably not accounted for in the census)?
Answer 29: This depends in part on the data product. For example, GHS-POP is based on GHSL which is built up area as ancillary variable. Thus if remote sensing based settlement layers or built up land masks would pick up the built environment, a model would be able to allocate population to any places that are classified as developed or impervious. Users may want to treat the estimates just as an indicator that population may be located there, but consider the estimated number of people to be highly uncertain. Depending on the situation, the population could consist of people who were counted in a recent nearby census or entirely undocumented individuals who were missed by a census -- or some mixture.

Question 30: What we see in all the models is that Population is not static, so how do we know per say the distribution of population between two cities? Don’t we use the gravity model out here?

Answer 30: In population allocation models that make use of dasymetric refinement, you would try to incorporate other ancillary data such as road density or topographic conditions that may help you improve estimations in such areas. The stronger the relationship between ancillary variables and population the more reliable the final estimate at the grid cell level.

Question 31: Do you think we can get gridded age information into these pop data?
Answer 31: Please see answers to questions 13 and 19 above.

Question 32: Is the input population data always Census data? What if the census data is highly flawed? E.g. in Myanmar where millions of muslim people were not included in the latest census? Do some products consider those census data quality issues and use other (what?) data inputs to improve the results?
Answer 32: The global-scale data sets are generally based on available Census data and UN estimates. In some countries, population estimates rather than census may be used. The GRID3 project is using survey data to predict population (along with other covariates) in countries without a recent census. The ancillary data are used to model spatial distribution, not to adjust total national population estimates.

Question 33: For a case study I did the validation and accuracy determination for both WorldPop and GHS grids in Greenland and a part of northern Canada.
The accuracy for both of them was about 60%. So, as they are not so accurate, how can they be used for health purposes? Because we need more accurate data. (e.g. vaccination and etc. as you told).
Answer 33: Population allocation including areal interpolation and dasymetric mapping approaches are less effective in rural areas because input census data are typically highly aggregated and ancillary data are of lower quality. As a consequence, it is not surprising that the accuracy is so low. In some cases, even poor data may be better than no data, and may be used to structure interventions to take into account higher uncertainty levels. However, a clear objective of the community is to improve accuracy across a range of conditions.

Question 34: How can you know the characteristics of each dataset to select the most suited for the application?
Answer 34: You may start with some summary characteristics as provided here: https://www.popgrid.org/data-docs-table1 And then move to each dataset documentation and metadata. The publication, Leaving No One Off the Map: A Guide for

Gridded Population Data For Sustainable Development (https://www.popgrid.org/resources), provides guidance on matching data with applications.

Question 35: In the slide no. 27, it seems World pop data is considered all the ancillary layers. Does this mean World-pop is the most realistic info to use?
Answer 35: see answer to question 36.

Question 36: In slide number 27, it seems World-Pop data has considered all the ancillary layers. Does this mean World-Pop is the most realistic info to use?
Answer 36: This means WorldPop is applying a highly intense modeling approach using multiple ancillary variables, and more complex modeling approaches (machine learning). Thus, the reliability of the resulting population estimates depends on the model performance and the relationship between population and ancillary variables. If for example, this relationship is weak in some regions, the population estimates might be less accurate. It is always a recommended approach to compare different population grids and be informed about the underlying quality of input population data, assumptions made, and model complexity. Fine resolution does not mean the estimates are more accurate but provide opportunities to represent population and its variation.

Question 37: Is this data available for the whole world?
Answer 37: All of the data resources available through the POPGRID web site are global in scale, though there may be missing countries or regions (e.g., high latitudes).

Question 38: My question is how one combines the pop gridding tools we just learned with projecting a population forward from the last census. This is important since in many humanitarian situations the last census is often more than 15 years ago.
Answer 38: Most of the data sets have utilized simple demographic projections from the last available census to harmonize estimates to the target year. Some data sets provide a consistent time series of population estimates, but others do not. Therefore, simply extrapolating population estimates for a specific area into the future may produce unrealistic results. SEDAC does distribute some future population distribution estimates developed by demographers that are based on the Shared Socioeconomic Pathways (SSPs): https://sedac.ciesin.columbia.edu/data/collection/ssp.

Question 39: It was not clear for me how these grids take into account the national census. Concretely speaking, the last census in Ukraine was made in 2001. How are these grids coping with the lack of official census at a national level for 20 years? I refer especially to the grids that provide data for 2020, for example. Thanks!
Answer 39: In the case of Ukraine, the Gridded Population of the World dataset used 2013 projections provided by Ukraine's National Statistics Office: https://sedac.ciesin.columbia.edu/downloads/docs/gpw-v4/gpw-v4-country-level-summary-rev11.xlsx. But as noted in question 38, in general, simple demographic projections are used from available census data.

Question 40: Is population grid being used to model COVID 19 affected areas?
Answer 40: No, none of the currently available population data sets have taken into account the impacts of COVID-19.

Question 41: Are the methods for creating each gridded dataset well documented? To the extent we can replicate them?
Answer 41: There are scientific publications for all population grids describing the underlying methods in detail. You can find summaries of these methods in Leyk et al. (2019) and identify key references in the bibliography (here and in this article). At the POPGRID website, you can find information on all global, continental and some national population grids, and find links to the data provider websites.

Question 42: What is the source of data for population estimates in an inter-censal period given that census data at enumerator area level is not in the public domain?
Answer 42: Population estimates for time periods between censuses are generally interpolations based on simple demographic relationships for the lowest available and consistent administrative units.

Question 43: Are there any examples of applications combining satellite data from Copernicus Sentinel and PopGrid that you can share? Many thanks.
Answer 43: POPGRID is not a specific data set; it is a collection of different data sets and a collaborative of data producers and users. The POPGRID web site provides information on different data sources used by the different data sets, including both Census and remote sensing data inputs.

Question 44: Are there plans to create the javascript population estimation tools as Esri friendly widgets that could be used in Esri web apps?
Answer 44: No, but some open data services are available such as the SEDAC Population Estimation Service: https://sedac.ciesin.columbia.edu/data/collection/gpw-v4/population-estimation-service.

Question 45: What happens or what do you use in the methodology to interpolate when instead of polygons with population data you have specific data? Or combining point population data with polygonal population data?
Answer 45: All input data used in GPW are a combination of boundary polygons and population totals.

Question 46: How are urban and rural populations defined?
Answer 46: In general, each country has their own definition of urban/rural areas and populations. The World Urbanization Prospects includes a list of definitions by country. A global approach has recently been adopted by the United Nations Statistical Commission: Degree of urbanization https://ghsl.jrc.ec.europa.eu/degurba.php
This approach represents an important contribution towards the harmonization of urban classes as they can be defined globally. However, these class definitions will differ from national census-based definitions at various degrees.

Question 47: Could this mapping infrastructure be used to find out what the current picture of cyber neutrality or use is either on a global scale or more geo-centrically refined data? E.g., what percentage of the population of London is fully digital compared to New York or Mogadishu?
Answer 47: Gridded population data have been used an input into assessments of Internet and communications accessibility. For example, there are about 30 citations of papers using SEDAC gridded population data that use the word “Internet” in their title or abstract, identified via the SEDAC Citations Database: https://sedac.ciesin.columbia.edu/citations-db.

Question 48: Is there a tutorial on how to import these POPGRID data in QGIS?
Answer 48: Once downloaded, the data can generally be added to QGIS via the add raster menu (the global data are generally available as GeoTIFF files, which QGIS can read directly). The recording of this event shows how you can add the services directly to QGIS by connecting to the MapService.

Question 49: How much variation in data accuracy should be expected within a single global dataset? I'm thinking about the US American Community Survey data vs. data from another country.
Answer 49: Examination of data quality grids provided by each data provider may be instructive as to spatial variation in accuracy within a specific data set. The SEDAC POPGRID Viewer does provide visualization of the range of estimates provided by the different data sets for specific regions of interest at different points in time.

Question 50: Hello, I would like to know which of the gridded population datasets presented so far have been adjusted with the UN World Population Prospects estimates (if any)?
Answer 50: There is a UN WPP-adjusted version of GPW4. It uses the WPP 2015 revision. WorldPop also has a UN WPP-adjusted version.

Question 51: Is there a known limitation to use the Global 1-km Downscaled Population Base Year and Projection Grids Based on the SSPs from SEDAC for change detection, and is this dataset available on the POPGRID Portal?
Answer 51: The downscaled population grids are based on relatively coarse-resolution data (½ degree), so pixel and small area change detections are not possible in a reliable manner. Larger-area changes can be identified in the data, but they are based on extrapolations, not a sophisticated predictive model. The data are available from the SEDAC web site:
https://sedac.ciesin.columbia.edu/data/set/sdp-downscaled-population-grid-b2-1990-2025

More recent work on downscaling population data may be better for change detection; e.g., see Wear et. al (2019) have completed a data set for the U.S.:
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0219242

Question 52: Field validation of gridded datasets is critical to ensure accurate decisions based on pop information. Are there apps that could easily be implemented by users to upload field data?
Answer 52: I don’t know of any apps that support field data upload for global datasets. However, in POPGRID we are discussing methods for sharing validated field data (training data of building locations, for example), so data like these could be useful! Please feel free to reach out to us if you have data to share.

Question 53: For calculating distance to nearest street, is it taking as a direct distance calculation? or calculating the route to (sub streets, alleys..) and then find the distance?
Answer 53: Distance can be calculated in different ways. Euclidean distance is often used as a simplified version. You will find alternative approaches where distance is calculated along routes/ roads or topographically adjusted.
The user should always get informed to fully understand what variables are used for population allocation and what kind of distance measures have been included. Please refer to the popgrid website to see what ancillary variables are used for creating the different population grids.
For example, WorldPop uses straight-line distance measures for roads (and other distance covariates). Landscan no longer uses distance to roads. Esri's WPE uses road intersections but does not calculate distances.

Question 54: How do you define the spatial resolution to transform polygon data into raster?
Answer 54: Such decisions are often related to existing population data (range of size of units) and ancillary data. If, for example, built up land is available at fine resolution, the analyst might feel more confident to derive fine resolution population estimates, assuming there is a strong relationship between the target variable and those ancillary variables. The final grid cell size might be a compromise between aiming for fine resolution data and robustness of estimates. For GPW, we picked 30 arc sec because it matches MODIS derived measurements.

Question 55: Would LandScan be better suited to identify informal settlements that are not captured by census data?
Answer 55: All population grids use census population data as input and thus all data products face this challenge of allocating people to places that would be missing in official census data. If such settlements would be included in larger census units the total population of the unit would be allocated to grid cells. If ancillary data used are sensitive to the built environment (e.g., built up land layer), population would be allocated to these places, but then it is possible that other (formal) settlements would be undercounted.

This is the paper mentioned in the presentation: Evaluating the Accuracy of Gridded Population Estimates in Slums: A Case Study in Nigeria and Kenya, https://www.preprints.org/manuscript/202102.0521/v1

Question 56: Are there any restrictions on exporting and publishing data brought into QGIS or ArcGIS via the ArcGIS server connection? I know LandScan, and probably others, charge for their data if you get it directly from them.
Answer 56: There are no restrictions on exporting the data at the server level. The use constraints (documented on the POPGRID website) still apply to the data. For LandScan, the data are free for research use; commercial users should contact Oak Ridge directly to access the data.

Question 57: Is it possible to get past estimates on these frequently updated grids (Facebook HSRL, for example)?
Answer 57: Yes, older versions of GPW and the HRSL are available for download, and Esri has both of its global layers online. LandScan may require a request to access older versions. GPW v3 is still online. All WorldPOP layers are online.

Question 58: Is it possible to download the information in netCDF format?
Answer 58: As far as I am aware, only GPW is available in netCDF format. If NetCDF is required, please get in touch with us at POPGRID and we can see what can be done. GeoTiff is the most common.

Question 59: It was mentioned that some pop models are better for estimating rural populations. Can you please share again which grids are better for that?
Answer 59: The HRSL uses high-resolution (50cm to 2m) optical data, which identifies more small rural settlements than the products that use moderate resolution optical data (GHS-POP, WorldPop). The rural population distribution models vary; identifying more small settlements may not produce the best population estimate, but it will be a more complete inventory of rural settlements.
The disparity of accuracy of population estimates between rural and urban settings comes mostly from input census data. The census units are commonly very large in rural areas, much smaller in urban areas. Thus inaccurate local population estimates somewhere in large areas to which few people have to be allocated to begin with, have a large impact on overall performance.

Question 60: Is it possible with your existing tool to do timeline series with one or different sources of population grids?
Answer 60: Unfortunately no, the existing services only show at most two points in time. For a timeline series, the original data would have to be downloaded for products

with multiple years and GIS (or code) used to extract the values and plot the timelines. The NASA AppEARS tool does allow this type of analysis with GPW: https://lpdaac.usgs.gov/tools/appeears/
WorldPOP has this ability on their website. (2000-2020, annually).

Question 61: The World Population Prospects (WPP) is the preferred/recommended dataset for computing and aggregating SDG indicators. How large are the discrepancies between WPP and some of these gridded datasets? Are there plans to adjust some of these datasets based on WPP estimates in support of SDG monitoring? What gridded population dataset would you recommend to use for SDG monitoring?
Answer 61:
There is a UN WPP-adjusted version of GPW4, it uses the WPP 2015 revision. WorldPop also has a UN WPP-adjusted version. The appropriateness of data set is likely to vary by SDG indicator. See the report on the POPGRID web site, Leaving No One Off the Map: A Guide for Gridded Population Data For Sustainable Development (https://www.popgrid.org/resources).

Question 62: How frequently are the population grids updated? How do I know the date of the source census data?
Answer 62: About the source census data, dates are usually available in the dataset documentation. For example, for GPW4, country level information and sources are available here: https://sedac.ciesin.columbia.edu/data/collection/gpw-v4/documentation. The POPGRID web site provides comparison tables for the major data sets: https://www.popgrid.org/data-docs-table1.

Question 63: I was interested in the API access you showed off in Python. Are there restrictions to making calls to that REST API if someone was interested in integrating that into their own application?
Answer 63:
Hundreds per minute would be too much, but lower or moderate applications is one. For high volume, please reach out. gyetman@ciesin.columbia.edu

Question 64: In the framework of a delta plan design which type of pop dataset will you advise for population projection?
Answer 64:

In theory, for a projection, you want trends from past data (over a few decades). Different models can predict, see for example, https://grid3.org/: Groundswell: Preparing for internal Climate Migration https://openknowledge.worldbank.org/handle/10986/29461. Depending on the time frame, it may also be useful to consider using downscaled population grids based on the Shared Socioeconomic Pathways (SSPs) (see question 38).

Question 65: Which data is the better to analyze trends in population (time series analysis)?
Answer 65: It depends on the time frame and spatial resolution needed. GHS-POP goes back to 1975. WorldPOP is more recent and annual. See: https://www.popgrid.org/data-docs-table1. Also note that data for different years from different versions of the same data set may not be consistent over time due to changes in methods and spatial framework.

Question 66: What kind of training data is used for weighting the model variables?
Answer 66: Training data footprint locations are evaluated and confirmed by analysts. Detailed census data is aggregated to produce lower resolution data. Using the census data as training data. Dasymetric modeling has a large amount of literature. You have to rely on areas with high resolution to compare lower to refine.

Question 67: I am currently interested in access to healthcare. In your opinion or experience is there a common gridded dataset that is currently used for assessing access to healthcare. If so, what is the specific population dataset?
Answer 67: Most common for this purpose is the WorldPOP dataset because it has a fine spatial resolution and a standard deviation surface at the pixel level. HRSL captures very small rural settlements at high resolution so if it covers your region of interest, that may be better for you. GRID3 (Geo-Referenced Infrastructure and Demographic Data for Development) is developing data for public health applications in a number of countries in Africa; see: https://grid3.org/

Question 68: Please can you show how those different data sets are integrated to improve accuracy?
Answer 68: Although the different data sets do provide a range of estimates for any particular point or area, it is important to note that the methods and input data sets used do overlap (e.g., several of them start with the same Census data inputs), so the

estimates are not statistically independent of each other. Therefore, treatment of the estimates as an “ensemble” (e.g., averaging the estimates) may be misleading.

Question 69:Please which would be the best data set(s) to study the co-existence of mixed communities (e.g., refugees and their host communities)?
Answer 69: Most of the data sets rely on Census data, which means that they may capture population distribution of host communities, but may overlook refugees and displaced populations not counted in official censuses or surveys.

Question 70: Do you offer a virtual/online training course to apply the population grid?
Answer 70: There are several recorded WorldPOP videos using their data. Recently, we did a workshop for GPW.