LEAVING NO ONE OFF THE MAP

A GUIDE FOR GRIDDED POPULATION DATA FOR SUSTAINABLE DEVELOPMENT

2020

POPGRID
DATA COLLABORATIVE

TRENDS
Thematic Research Network on Data and Statistics

SUSTAINABLE DEVELOPMENT SOLUTIONS NETWORK
A GLOBAL INITIATIVE FOR THE UNITED NATIONS

LEAVING NO ONE OFF THE MAP

A GUIDE FOR GRIDDED POPULATION DATA FOR SUSTAINABLE DEVELOPMENT

A Report by the Thematic Research Network on Data and Statistics (TReNDS) of the UN Sustainable Development Solutions Network (SDSN) in Support of the POPGRID Data Collaborative

2020

DISCLAIMER

This work is a product of SDSN TReNDS in support of the POPGRID Data Collaborative. The findings, opinions, and analysis presented in this report do not necessarily reflect the views of governments, organizations, and entities described in this report. The report is available in Spanish and French.

Design By: Luciana Berger

ACKNOWLEDGMENTS

This report was written by Hayden Dahmm (SDSN TReNDS) and Maryam Rabiee (SDSN TReNDS), with input from Jessica Espey (SDSN TReNDS), Susana Adamo (Columbia University), Alex de Sherbinin (Columbia University) and Robert Chen (Columbia University). We thank Alyson Marks (SDSN TReNDS) for her assistance with copy-editing and Luciana Berger for design.

We are grateful to the many gridded population data users and experts whose inputs have made this report possible. In particular, we would like to thank Kytt MacManus (Columbia University), Greg Yetman (Colombia University), Andy Tatem (University of Southampton), Stefan Leyk (University of Colorado Boulder), Forrest Stevens (University of Louisville), Andrea Gaughan (University of Louisville), Amy Rose (Oak Ridge National Laboratory), Marie Urban (Oak Ridge National Laboratory), Jacob Mckee (Oak Ridge National Laboratory), Kelly Sims (Oak Ridge National Laboratory), Thomas Kemper (European Commission Joint Research Center), Michele Melchiorri (European Commission Joint Research Center), Sergio Freire (European Commission Joint Research Center), Charlie Frye (Esri), Ashraf Abushady (World Food Programme), Dana Thompson (University of Southampton), Lorenz Noe (Open Data Watch), and Cascade Tuholske (University of California, Santa Barbara) for sharing their knowledge and perspectives in extended interviews. We would also like to recognize the valued contributions from our colleagues at the Global Partnership for Sustainable Development Data (GPSDD), Alonso Ortiz Galán and Charu Vijayakumar, who included questions on gridded population data during their survey of government officials from Belize, Colombia, Costa Rica, Ecuador, Ghana, Nepal, Paraguay, Sierra Leone, and Tanzania. Lastly, we are grateful to the Bill & Melinda Gates Foundation who helped make this report possible with their support of the POPGRID Data Collaborative.

ACRONYMS

ADAM	Automatic Disaster Analysis and Mapping
CIAT	Centro Internacional de Agricultural Tropical
CIESIN	Center for International Earth Science Information Network
DMSP-OLS	Defense Meteorological Satellite Program Optical Linescan System
DPT	Diphtheria, Pertussis, Tetanus
ETM	Landsat Enhanced Thematic Mapper
GHSL	Global Human Settlement Layer
GIS	Geographic Information Systems
GHS-BUILT	Global Human Settlement Layer Built Up Extent
GHS-POP	Global Human Settlement Layer - Population
GPSDD	Global Partnership for Sustainable Development Data
GPW	Gridded Population of the World
GRID3	Geo-Referenced Infrastructure and Demographic Data for Development Initiative
GRUMP	Global Rural Urban Mapping Project
HRSL	High Resolution Settlement Layer
IFPRI	International Food Policy Research Institute
IHS	Integrated Household Survey
JRC	European Commission Joint Research Centre
LSMS	Living Standards Measurement Study
NSO	National Statistics Office
ORNL	Oak Ridge National Laboratory
PAGER	Prompt Assessment of Global Earthquakes for Response
PAR	Population At Risk
POPGRID	POPGRID Data Collaborative
SDGS	Sustainable Development Goals
SEDAC	NASA Socioeconomic Data and Applications Center
TReNDS	Thematic Research Network on Data and Statistics
UN	United Nations
UNPD	United Nations Population Division
UNSC	UN Statistical Commission
USGS	U.S. Geological Survey
VIIRS	Visible Infrared Imaging Radiometer Suite
WFP	World Food Program
WHO	World Health Organization
WPE	World Population Estimate

TABLE OF CONTENTS

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Each year, nearly 160 million people are impacted by natural disasters. During a natural disaster, every second counts to save lives and to ensure that critical supplies reach those in need. To respond quickly and confidently to identify impacted communities in the wake of natural disaster, emergency organizations, like the United Nations World Food Program (WFP), are using gridded population data to estimate near real-time impacts of earthquakes and tropical storms on people and infrastructure. Gridded population data have proven essential for informing emergency response efforts and minimizing suffering.

Having reliable and timely population data can make a life or death difference for individuals facing crises or living in conflict-ridden regions. These data are essential for addressing the above challenges and for critical decision-making and planning. We need to know where people are located, what conditions they are facing, what infrastructure is available, and what basic services they can access. When it comes to counting people in hard-to-reach or conflict-ridden environments, there are no silver bullets. However, gridded population data offer a promising option for delivering actionable data in difficult circumstances.

When 193 world leaders agreed upon the 17 Sustainable Development Goals (SDGs) in 2015, they promised to "leave no one behind." But without reliable and timely population data linked to location, we cannot ensure that everyone is counted and that no one will be left behind. While, most governments and policymakers depend on traditional data sources, such as household surveys and population censuses, in order to develop the necessary policies and programs to eradicate poverty and improve health, education, and other basic services, these traditional data sources present a range of geographic, temporal, and logistical challenges. For example, at the time of writing, nearly 60 countries are facing confirmed or potential delays to their census operations because of the COVID-19 pandemic (UNFPA, n.d.). Fortunately, with recent advancements in Earth observations' capabilities and statistical methods, it is now possible to obtain more frequent and more granular population estimates worldwide through the use of gridded population datasets.

Gridded (or raster) population maps represent the distribution of population in rows and columns of grid cells, typically defined by their latitude-longitude coordinates. An increasing number of data providers are combining information from censuses with satellite-derived geospatial features to redistribute populations and produce gridded population datasets. Despite this progress, there remains confusion or simply lack of awareness about gridded population data. The large number of different datasets now available can be overwhelming to users, particularly to those who lack the time and technical expertise to understand differences among the products and assess their strengths and weaknesses for potential applications.

The POPGRID Data Collaborative was established in 2018 to address many of these challenges by connecting the diverse data users, providers, and stakeholders from the public and private sectors working with georeferenced data on population, human settlements, and infrastructure.

In this report, *Leaving No One off the Map: A Guide for Gridded Population Data for Sustainable Development,* we aim to narrow this knowledge gap by helping to improve the accessibility and understanding of gridded population datasets for policymakers and other users. The report was written with two overarching questions in mind:

- How can gridded population data supplement current population data sources and support users from the sustainable development community to make timely, informed decisions?

- Which gridded population dataset is the most suitable for a user's intended use?

Drawing from an extensive literature review and interviews with key data providers and users in the POPGRID Data Collaborative, the report presents an overview, analysis, and recommendations for the use of gridded population datasets in a wide range of application areas, such as in disaster response, health interventions, and survey planning. Specifically, the report compares seven gridded population datasets from the POPGRID Data Collaborative, including an analysis of the underlying data, methods and basic assumptions, and the corresponding strengths and limitations of each dataset in simple terms. The report also presents an intercomparison assessment of the use of different datasets and their varying outputs, addresses many of the misconceptions around gridded population data, and concludes with nine guiding criteria to aid users in their selection process.

KEY MESSAGES FROM THE REPORT INCLUDE:

CENSUS DATA ARE STILL IMPORTANT; GRIDDED POPULATION DATA ARE NOT A SUBSTITUTE FOR CENSUS DATA.

The datasets featured in this report build off of national censuses as inputs to their population estimations, which can then be more regularly updated or calculated at a higher spatial frequency.

GRIDDED POPULATION DATA ARE NOT ERROR-FREE.

Although they address some of the limitations of traditional sources, the methods for producing estimates add their own sources of uncertainty. Users should be aware of and transparent about these potential uncertainties.

DATA USERS SHOULD CONSIDER A NUMBER OF FACTORS WHEN SELECTING AN APPROPRIATE DATASET FOR THEIR PARTICULAR NEEDS.

Each of the datasets are based on different underlying data and assumptions, so policymakers and researchers should examine the characteristics of available datasets carefully. These include: demographic characteristics; the spatial resolution required; the time periods of interest; data costs and rights to reuse the data; and more.

MORE VALIDATION WORK IS NEEDED TO COMPARE GRIDDED POPULATION DATA ESTIMATES AGAINST AUTHORITATIVE DATA ON POPULATION LOCATION.

There is a critical need for a more systematic analysis and objective validation of these products to further refine methods and improve their accuracy and utility; this work is underway through the POPGRID Data Collaborative.

With only ten years remaining to achieve the SDGs, we are at a crossroads. Gridded population data are already available to help fulfill these ambitious goals by improving the availability, consistency, and spatial disaggregation of SDG indicators, by helping national and international initiatives to better target their efforts to achieve the SDGs, and by identifying and locating those who might otherwise be left behind. However, these data are only as good as policymakers' understanding of their limitations, applications, and fitness for use. Furthermore, their full potential cannot be realized if they are not thoroughly validated against authoritative, real-world data. We must accelerate this important research and advancement of the application of gridded population data around the world to make sure that the SDGs are achieved and no one is left off the map.

INTRODUCTION

In 2019, the global population reached 7.7 billion (United Nations, n.d.). Although the population has nearly quintupled since the turn of the 20th Century, we unfortunately have little to no understanding of the conditions under which many people live. This is particularly acute in many of the most vulnerable and hard-to-reach communities. When the Sustainable Development Goals (SDGs) were adopted in 2015, 193 countries committed to "leave no one behind" (United Nations, 2016). Yet, without accurate geospatial data on populations, we cannot achieve these goals and ensure that everyone is counted.

Our traditional knowledge of populations is based on the census, an official accounting of all the individuals in a country. The United Nations (UN) defines it as "the total process of collecting, compiling, evaluating, analyzing and publishing or otherwise disseminating demographic, economic, and social data pertaining, at a specified time, to all persons in a country or in a well-delimited part of a country" (United Nations, n.d.). It provides invaluable information about communities, helps to direct services, can determine political representation, and much more (Dahmm, 2018). A total of 214 countries or regions conducted censuses during the 2010 census round (UNSD, n.d), yet there are still many countries with data that are decades out-of-date, especially in low-income or conflict settings. For instance, the last census in Madagascar was conducted in 2003, in 1984 in the Democratic Republic of the Congo, in 1979 in Afghanistan, in 1975 in Somalia, and in 1932 in Lebanon (Wardrop et al., 2018).[1] Additionally, even when censuses are performed regularly, there can be difficulties in achieving a fully-representative count. Census enumerators may face language barriers with indigenous communities and other minority groups, and areas impacted by disaster or war can be challenging to access (GPSDD Country Partners, 2019). In some countries, ethnic divisions may also result in systematic under-counting of minorities, and vulnerable communities are likely to be underrepresented. Also, conditions on the ground can change significantly over a decade as people are born, die, or relocate, and this can considerably impact countries' economies. In El Salvador, for example, it is estimated that the use of outdated census figures for allocating municipal funding resulted in approximately US$92 million (in real 2018 dollars) to be misdirected between 2000 and 2007 (Roseth et al., 2019). Census data are irreplaceable assets, yet additional data tools, including gridded population data, have emerged over the past two decades to help address these and other limitations.

Gridded population data products provide estimates of population figures according to uniform grid cells across the world (Leyk et al., 2019). Such data provide a better understanding of where groups of people are settled. Minimally modeled population grids redistribute census data evenly across grid cells, based on the assumption that people are homogeneously distributed. Other gridded datasets incorporate satellite imagery and a host of other data sources to refine their allocation estimates, which allows for more regular updates to the estimations.

[1] Census dates for all countries are available on United Nations Statistics Division website, https://unstats.un.org/unsd/demographic-social/census/censusdates/.

These gridded datasets offer a number of benefits:

- They provide informed and timely estimates of population distribution.
- They facilitate direct use with other sources of scientific data that are available in a gridded format.
- In areas where recent censuses have not been performed, they provide a starting point for surveys and needs assessments.

Fortunately, these innovative data products are now becoming integral to decision-making processes for a number of actors. Yet, a recent survey of National Statistical Offices (NSOs) by the Global Partnership for Sustainable Development Data (GPSDD) revealed that many officials are still largely unaware of such resources (GPSDD Country Partners, 2019). In November 2019, GPSDD surveyed country partners to gain insight on their knowledge and use of gridded population datasets (see Appendix A for survey). This research included interviews with government officials from Belize, Colombia, Costa Rica, Ecuador, Ghana, Nepal, Paraguay, Sierra Leone, and Tanzania. Country-level analysis of the survey results is included in different sections throughout the report.[2]

This report lays out what is currently known about gridded population datasets so that their potential, as well as their limitations, can be more widely recognized, and their capabilities are made more accessible to policymakers working to achieve the SDGs.

BOX 1 | CENSUS CHALLENGES AND OPPORTUNITIES

To produce a census, countries must overcome a number of challenges. These issues were explored during interviews with NSO representatives. Representatives from Colombia, Costa Rica, and Ecuador noted that the lack of financial resources limits the completeness and timeliness of census implementation. Additionally, in Colombia, accessing communities that are in remote locations or impacted by violence is difficult, and census enumerators face language barriers and resistance to participation. Meanwhile, in Belize, the NSO has struggled to employ enough enumerators, and changing population distributions have created additional logistical challenges. Nepal has also struggled with misidentifying ethnicity and caste. As a result, there can be significant accuracy issues with the census. For example, Paraguay's most recent census in 2012 may have missed a quarter of the population (GPSDD Country Partners, 2019).

Fortunately, new methods and technologies have the potential to improve the process by increasing efficiency and accuracy, and many NSOs are implementing or have plans to implement them in the future. For example, in 2018, the National Administrative Department of Statistics (referred to as DANE), integrated spatial covariates in the production of population estimates in remote and inaccessible areas. Additionally, in Tanzania, the 2022 census will use a new approach to draw maps with Geographic Information Systems (GIS), and the government intends to shift from paper to tablet-based data collection. Tablets are also being proposed for data collection in Costa Rica and Sierra Leone. Also, in Belize, the government has undertaken a building mapping exercise using satellite images to estimate changes against the previous census, and Nepal is using Google Maps to update its enumeration area map.

(GPSDD Country Partners, 2019)

[2] GPSDD colleagues captured officials' verbal responses anonymously and compiled the information by country.

Several alternative global-scale gridded population datasets have now been produced, each with their own nuances, particular advantages, and tradeoffs. The international community of population data experts has recognized a need for greater cooperation, and the POPGRID Data Collaborative (POPGRID) has brought together the leading data producers, users, and sponsors of georeferenced data on population, human settlements, and infrastructure. Together, POPGRID members aim to improve the accessibility and consistency of data, to support users by addressing pressing needs and avoiding confusion, and to encourage innovation.

In this report, we describe seven major, global gridded population datasets produced by organizations participating in POPGRID: Gridded Population of the World (GPW), Global Rural Urban Mapping Project (GRUMP), Global Human Settlement Layer (GHSL), LandScan, WorldPop, World Population Estimate (WPE), and the High Resolution Settlement Layer (HRSL). These global datasets use a top-down method to disaggregate or redistribute census-based population counts to grid cells, meaning that the products rely on census data to produce population estimates. This report focuses on datasets produced using the top-down methods outlined in Chapter 2. However, alternative methods are being developed to derive estimates that do not depend on the availability of census data, such as bottom-up or hybrid census methods which use micro-census surveys and detailed satellite imagery to derive population estimates where traditional national census data are absent (Wardrop et al., 2018; Weber et al., 2018; UNFPA, 2017).

In the following chapters, we detail the gridded population data products with global to near-global coverage and the ways they can be applied. We then discuss some of the existing work that has been done to validate these data products and explain the need for additional research. Building on all of this information, we lay out some guiding principles to use gridded population data. When used in a responsible and informed way, gridded population data can be an important asset in making sure that no one is left off the map.

BOX 2 | THE CENSUS REMAINS CRITICAL

Gridded population data are not a substitute for census data. On the contrary, all global gridded population datasets are built off of census data, and the reliability of input population data is the most important factor in the overall accuracy of the resulting estimates (Wardrop et al., 2018). Instead, gridded products are supplemental tools that help us to leverage the value of census data. Even when methods draw on satellite imagery, the census is still a necessary variable in a number of estimations, as image analysis cannot substitute for reliable on-the-ground counts. Automated imagery analysis (also known as unsupervised classification) can be used to overcome some of these challenges, however, it has several limitations. For example, dense tree coverage can obscure houses and make an area appear uninhabited, or buildings can appear residential when they are commercial or industrial. Although gridded population data can help document changes between censuses, estimate population distribution with greater resolution, and offer estimates of populations in areas where a census has not been performed, their overall utility is only strengthened with higher quality census data.

CHAPTER 1

HOW GRIDDED POPULATION DATA ARE BEING USED AROUND THE WORLD

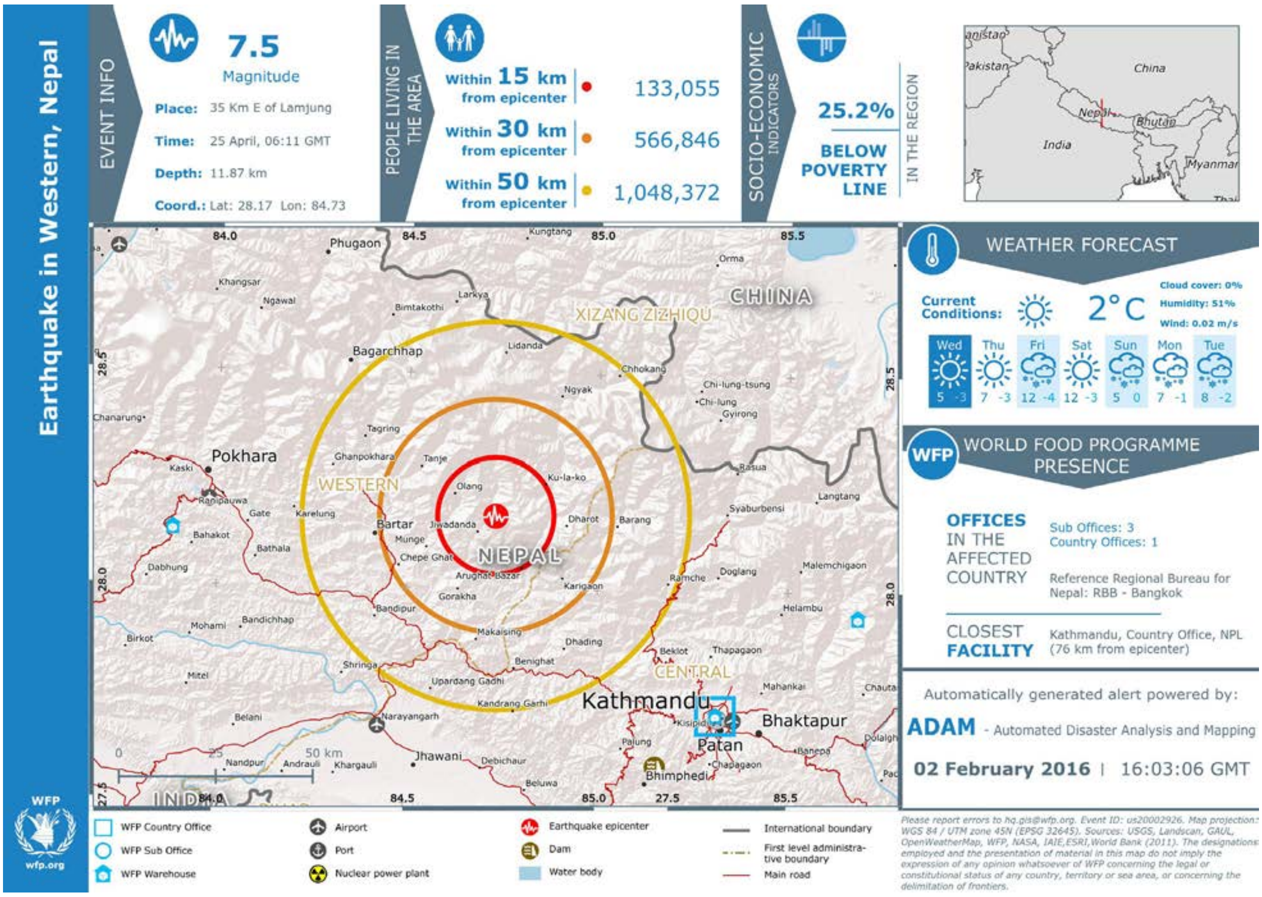
Across the international community, there is a growing movement for evidence-based decision making (Cairney, 2016). The UN has called for the production of better data and statistics for sustainable development via the data revolution to ensure that governments' decisions are evidence-based, to strengthen accountability, and to monitor progress towards the global goals (IEAG, 2014). The census is one of the most important sources of evidence we have available. Understanding where populations live and characterizing the most vulnerable communities are essential to achieving both national and global commitments. Yet, in some situations, the census can only go so far, and data users around the world are turning to gridded population data to help address urgent problems in a range of contexts; from informing responses to immediate humanitarian crises and helping to track policy commitments to predicting climate change. The diversity of applications underscores the potential of these tools to not only describe our world's population, but to improve the condition of individual people.

1.1 IMMEDIATE APPLICATIONS

Every year, natural disasters affect approximately 160 million people worldwide (WHO, n.d.) and displace some 17 million people (IDMC, 2019), while infectious diseases kill over six million (Baylor College of Medicine, n.d.). To reach those impacted, we need to quickly and reliably know where they are located. Gridded population data are being used to more rapidly mobilize and target humanitarian resources, as well as inform methods of response.

DISASTER RISK ASSESSMENT AND RESPONSE

Disaster risk assessment and response planning are among the most common applications of gridded population data (MacManus, 2019). Such data have been used to predict the potential impacts of a variety of disaster scenarios. For example, following the Fukushima Daiichi plant meltdown in Japan, researchers calculated with the Global Rural-Urban Mapping Project (GRUMP) dataset that 21 of the World's nuclear power plants had over one million people living within a 30 kilometer (km) radius (Butler, 2011), which is used to help manage the risk of similar events. Another study used the Global Human Settlement Layer (GHSL) 2015 data to estimate that over 35% of the global population is potentially exposed to earthquakes and that one billion people are potentially exposed to flooding (Ehrlich et al., 2018). When a disaster does strike, it is critical that responders reach those impacted as soon as possible, and methods have been developed to make "on the fly" estimates of the populations exposed to natural or manmade disasters (Garb et al., 2007). For example, an early study showed the potential of this approach by using demographic data to estimate mortality of the 2004 tsunami in Indonesia, demonstrating how a reasonable approximation could be derived to then inform humanitarian responses (Doocy et al., 2007).



ADAM Dashboard Example
Source: World Food Programme

The United Nations World Food Program (WFP) now actively relies on gridded population data to plan its interventions (Abushady, 2019). With only a short time after a disaster to estimate the number of resources to deploy, the WFP expedites the calculations with Automatic Disaster Analysis & Mapping (ADAM), a tool that draws on both the LandScan and WorldPop datasets to estimate near real-time impacts of earthquakes and tropical storms. The resulting maps and dashboards of the affected populations are shared with the wider humanitarian community, and the estimates directly affect WFP's decisions about initial food delivery. The population figures are quickly refined by data coming from ground operations, but the initial estimates from ADAM are still critical to preparing the appropriate level of response. Without gridded population data, the response would not be as timely or effective, leading to delays in assistance.

The United States government also relies on gridded population data in its international disaster response efforts. It has applied the data in natural disaster situations ranging from the 2010 earthquake in Haiti and the 2015 earthquake in Nepal, to the multiple hurricanes that regularly devastate the Caribbean (Rose, 2019). Additionally, the US Census Bureau developed a specialized gridded population dataset for Pakistan that considered the country's unique and varied conditions (Azar et al., 2013); this tool was then used by the US State Department's Humanitarian Unit to estimate the population affected by flooding in 2010. The U.S. Geological Survey (USGS) also uses LandScan population data in its PAGER (Prompt Assessment of Global Earthquakes for Response) system for rapid assessment of earthquake impacts.

HEALTH AND INFECTIOUS DISEASE

During the 2013-2016 West Africa Ebola outbreak, one of the greatest limitations to the health response was a lack of reliable information about the local population (Cori et al., 2017). Without recent, detailed population data, responders struggled to locate communities and accurately calculate infection rates. Gridded population data can be used to address these challenges as well as to predict infectious disease spread and inform containment strategies. For example, a study showed that the spread of Ebola could be reasonably projected using LandScan population data in combination with case counts and the distances between affected and nonaffected districts in Guinea, Sierra Leone, and Liberia (Rainisch et al., 2014). Now with the novel COVID-19 virus reaching global pandemic proportions at the time of writing, gridded population data have been important to predicting its potential spread across China and the World, and work by the WorldPop group has demonstrated the considerable impact of containment strategies to slowing the spread (Lai et al., 2020a; Lai et al., 2020b). Additionally, the producers of GPW have developed an interactive Global COVID-19 Viewer that presents statistics about virus cases in combination with population estimates by age group and sex, information that is critical to understanding the potential spread and severity of the virus (SEDAC. n.d.).

Public health studies on infectious disease are becoming increasingly reliant on gridded population data as well. In particular, they are key to calculating the 'Population At Risk' (PAR), which is a measure of the population exposed to disease risk (Linard and Tatem, 2012). For example, the World Health Organization (WHO) uses WorldPop to estimate the number of Malaria cases in areas where reliable statistics are not available, especially in Africa, where more than 90% of the cases arise (WHO, 2019). By overlaying models of disease prevalence from the Malaria Atlas Project with gridded population data, they can predict the disease burden. Building off of these estimations, the WHO has calculated that there were 228 million Malaria cases globally in 2018, which is down from 251 million cases in 2010, but unfortunately still not in keeping with international targets for Malaria prevention. Similarly, high resolution estimates of the proportion of adults with Human Immunodeficiency viruses (HIV) in Africa were generated by combining WorldPop data with surveys and information on the locations of healthcare clinics, finding that there is considerable subnational variation in the rate of change of transmission (Dwyer-Lindgren et al., 2019).

Population data are also vital to monitoring the availability of fundamental healthcare services. One study mapped diphtheria-pertussis-tetanus (DPT) vaccine coverage for children across Africa based on WorldPop and health surveys, concluding that although coverage had improved in nearly three-quarters of Africa's secondary administrative units between 2000 and 2016, only Morocco and Rwanda had met the target of reaching 80% of children (Mosser et al., 2019). Moreover, gridded population data have been also been used to support measures of child malnutrition and stunting, to develop indicators for urban health decision-making, and to estimate the spatial distribution of births and pregnancies (Local Burden of Disease Child Growth Failure Collaborators, 2020; Thomson et al., 2019; James et al., 2018).

SURVEY PLANNING

Household surveys are an important tool for planning health interventions and other time-sensitive policies. To design a locally representative sample, surveyors need to understand how the population is distributed. When census data are out of date, inaccurate, or only available at higher administration levels, planning a survey can be all the more challenging. Gridded population data can provide surveyors with updated estimates, and the grid cells in some products may be closer to the actual unit areas that are to be surveyed (Thomson, 2019). Furthermore, the uniformity of grid cells makes it much easier to perform spatial oversampling so that smaller and often underrepresented population groups can be studied in greater detail (Thomson et al., 2017). Flowminder has developed the GridSample tool to help surveyors to automatically generate sample units based on gridded population data (Thomson et al., 2017). While the use of gridded population data for survey planning may still be regarded as experimental, gridded population data have already been used to design surveys for health planning around the World, including in Rwanda and Haiti (Thomson, 2019).

1.2 MEDIUM-TERM APPLICATIONS

The world population is projected to reach 8.5 billion by 2030 (UNPD, 2019). Gridded population data can support monitoring of the economic and environmental conditions faced by this growing population and help support the achievement of the SDGs and other medium-term commitments.

STUDYING POLICY DECISIONS

An improved understanding of how people are distributed can strengthen policy planning and evaluation. At the most general level, gridded population data can help assess the outcomes of different economic strategies. Although a majority of countries have decentralized over recent decades, it is still debated whether this process has a positive economic effect. A recent economic study drew on measures of geographic fragmentation based on population data to find that increases in subnational expenditures are associated with slight increases in GDP, although this relationship is only significant in developed countries (Canavire et al., 2020). A World Bank study considered the implications of China's Belt and Road Initiative for the economies of Central Asia by creating a spatial equilibrium model that included GPW population data (Bird et al., 2019). The results suggest that the initiative's investments could benefit certain areas, where incomes are projected to rise and populations potentially double, but simultaneously disadvantage other areas. Gridded population data can also shed light on the human cost of global trade and consumption. For example, a study used LandScan data to examine the impact of fine particulate matter air pollution in Asian countries resulting from consumption in the five highest consuming countries, and concluded that it contributed to the premature death of over one million people in 2010 alone (Nansai et al., 2020).

Understanding population distribution is also critical for planning infrastructure and service delivery. First off, gridded population data can illustrate changes in the availability of basic services. One model used GPW data to document electrification rates in Africa, estimating that between 2000 and 2013, the portion of the continent's population with electricity access rose from 26.8% to 35.5% (Andrade-Pacheco et al., 2019). Likewise, a study that used WorldPop data in combination with various georeferenced surveys found that the proportion of Africans with access to improved housing more than doubled between 2000 and 2015 from 11% to 23% (Tusting et al., 2019). Beyond understanding current conditions, gridded population data can aid in the design of new infrastructure. GPWv4 data was used in a recent evaluation of solar microgrids for Malawi, where only 12% of the 18 million residents have access to the electric grid (Eales et al., 2020). By considering the size and location of communities, the study found that 37% of Malawi's population would be served most cost-effectively by solar microgrids, whereas 42% would be served more cost-effectively by extending the existing grid and the remaining 21% by solar home systems. Relatedly, an analysis of potential coal energy phaseout in China considered population and pollution patterns to recommend technically feasible strategies that would also be equitable (Cui et al., 2020). According to its producers, LandScan has been used by the private sector for designing cellular telephone networks and considering new transportation routes (Rose, 2019).

THE SUSTAINABLE DEVELOPMENT GOALS

When 193 countries committed to the SDGs, they also agreed to monitor 232 indicators relating to different aspects of sustainability. These indicators are central to the achievement of the global goals, by providing a mechanism for evaluating progress and developing and testing policies. Yet in many places, the necessary methods or data are still missing to regularly monitor the indicators. A 2018 UN survey found that in Africa and Asia, on average, data for only 20% of SDG indicators are currently available (United Nations, 2018).

Gridded population data can help provide key population inputs needed for a range of SDG indicators that describe the status or rate of change of population-related characteristics. For example, SDG indicator 1.1.1, "Proportion of population below the international poverty line, by sex, age, employment status and geographical location" (Tatem, 2019). It can also be used to help measure the proximity of populations to services and infrastructure and to identify remote or hard-to-reach communities that are sometimes left out of normal data collection methods (Tatem, 2019). Furthermore, modeled gridded population data has the potential to help monitor progress more regularly, since population counts from censuses are generally only available once per decade (MacManus, 2019).

Methods using gridded population data are already being developed for a number of SDG indicators. For example, connections to transportation networks are vital to supporting the rural poor, and the rural access index (SDG indicator 9.1.1) measures the proportion of the rural population living within walking distance of a roadway. This measure has historically been produced with household data and has only been available at the national level, which is less meaningful for development planning (Iimi et al., 2016). The World Bank has since developed a method that combines gridded population data from WorldPop with national data sources to estimate the rural access index at subnational levels (Iimi et al., 2016). Not only can the new method provide more regular measurements, but it can also offer more detailed insights. For example, the World Bank was able to describe noticeable differences in road condition and density between Africa and Asia (World Bank, 2016). As of 2018, the World Bank had calculated indices for 20 countries, and this method has been officially accepted by the UN Statistics Division (Vincent, 2018; UN IAEG-SDG, 2019).

Indicators that deal with the challenges of urban development can also benefit from gridded population data. A method was developed for measuring population-weighted particulate matter air pollution (SDG indicator 11.6.2) using GPW in combination with ground measurements and satellite remote sensing (Shaddick et al., 2017). The WHO has officially adopted the methodology, and the indicator is now available for 178 countries (UNSD, 2017a). Likewise, land use efficiency (SDG indicator 11.3.1) gives the ratio of the land consumption rate to population growth rate, and researchers from the European Commission's Joint Research Centre have submitted a method for measuring this indicator using gridded population data from GHSL (Melchiorri et al., 2019). They measured the change in land use efficiency for 10,000 urban centers between 1990 and 2015, calculating that urban population density increased globally during this period. It has also been proposed that GHSL could be used to measure the proportion of the population that has convenient access to public transport (SDG indicator 11.2.1), as well as 11 other urban indicator gaps (UN-GGIM Europe, 2019; Melchiorri et al., 2019).

While, a number of other SDG indicators rely on population figures or populated areas, many are not yet monitored at the local level (Gaughan, 2019). According to a review by TReNDS, at least 73 unique indicators require some form of population data, and 34 of these indicators still have inadequate data or undefined methods (see Table 1). Several of the most significant indicators require information about population distribution relative to areas of impact or service. This includes the proportion of people affected by disaster and their economic loss (SDG indicators 1.5.1 and 1.5.2), with access to sanitation (SDG indicator 6.2.1), and with access to electricity (SDG indicator 7.1.1). As with land use change, gridded population data may be useful in tracking many urban indicators, including the proportion of the urban population living in slums or inadequate housing (SDG indicator 11.1.1). Expanding the application of gridded population data could help foster greater accountability and awareness around the SDGs.

The United Nations Statistical Commission (UNSC) has endorsed a method to delineate urban and rural areas for international statistical comparisons based on a population grid (UNSC, 2020). The method facilitates information collection and reporting for SDG indicators requiring urban or rural disaggregation (1.1.1, 2.4.1, 3.3.1, 4.5.1, 9.1.1, 11.1.1) and helps to compare results of urban SDG indicators sensitive to how city boundaries are drawn (11.2.1, 11.3.1, 11.6.2, 11.7.1) (European Commission Joint Research Centre, 2018).

TABLE 1 SDGS THAT REQUIRE POPULATION DATA

GOALS	SDG INDICATOR	GOALS	SDG INDICATOR
1 NO POVERTY	1.1.1, 1.2.1, 1.2.2, 1.4.1, 1.4.2, 1.5.1, 1.5.3	8 DECENT WORK AND ECONOMIC GROWTH	8.1.1, 8.4.2, 8.5.2, 8.6.1, 8.7.1, 8.10.1, 8.10.2
2 ZERO HUNGER	2.1.1, 2.1.2, 2.2.1, 2.2.2	9 INDUSTRY, INNOVATION AND INFRASTRUCTURE	9.1.1, 9.2.1, 9.5.2, 9.C.1
3 GOOD HEALTH AND WELL-BEING	3.1.1, 3.2.1, 3.2.2, 3.3.1, 3.3.2, 3.3.3, 3.3.4, 3.4.1, 3.4.2, 3.5.2, 3.6.1, 3.7.1, 3.7.2, 3.8.2, 3.9.1, 3.9.2, 3.9.3, 3.A.1, 3.B.1	10 REDUCED INEQUALITIES	10.2.1, 10.3.1
4 QUALITY EDUCATION	4.2.1, 4.2.2, 4.3.1, 4.6.1	11 SUSTAINABLE CITIES AND COMMUNITIES	11.1.1, 11.2.1, 11.3.1, 11.5.1, 11.6.2, 11.7.2, 11.A.1
5 GENDER EQUALITY	5.2.1, 5.2.2, 5.3.1, 5.3.2, 5.5.2, 5.6.1, 5.B.1	12 RESPONSIBLE CONSUMPTION AND PRODUCTION	12.2.2, 12.4.2
6 CLEAN WATER AND SANITATION	6.1.1, 6.2.1	13 CLIMATE ACTION	13.1.1
7 AFFORDABLE AND CLEAN ENERGY	7.1.1, 7.1.2	16 PEACE, JUSTICE AND STRONG INSTITUTIONS	16.1.1, 16.1.3, 16.1.4, 16.2.1, 16.2.2, 16.2.3, 16.6.1, 16.6.2, 16.7.2, 16.9.1

1.3 LONG-TERM APPLICATIONS

Gridded population data are integral to assessing the long-term risks of future environmental change, including climate change-induced sea level rise, storms, and droughts. The outcomes of these assessments are critical for effective disaster preparation and resilience planning. Because the administrative boundaries that are used to report census records are geographically irregular, the raw data can be challenging to combine with other types of observations (Bai et al., 2018). Gridded population data address this issue and provide key information for researchers to understand future sustainable development risks (MacManus, 2019). Indeed, gridded products representing current population distribution have been used in multiple development scenarios to develop spatial projections of future population at multiple resolutions to better anticipate populations' vulnerability to future harm (Gao, 2017; Jones and O'Neill, 2020).

SEA LEVEL RISE

Scientific studies estimating coastal populations vulnerable to climate change and flooding often use high-resolution population datasets to capture the considerable geographic variability along the coastlines (Mondal and Tatem, 2012). Available projections suggest the impact of sea level rise on humanity could be severe. An analysis based on GRUMP suggests that up to 411 million people globally could be exposed to extreme flood risk by 2060, compared with just 189 million as of 2000 (Neumann et al., 2015). Overlaying gridded population data with measures of Gross Domestic Product (GDP) and agricultural land use has shown that developing countries face increased risk from one-meter of sea level rise (Dasgupta et al., 2011). Additionally, modeling sea level rise along with other environmental phenomena suggests that even more people may be at risk. A report from the World Bank used LandScan in combination with other tools to assess the added risk of flooding that could result from mangrove loss, finding that the culmination of sea level rise and mangrove loss could increase flood inundation across 42 developing countries by nearly a third and double the number of people facing flood risk (Blankespoor et al., 2016). More recent work, also using Landscan, found that even more people in the coastal zone may be at risk based on an improved coastal elevation layer, CoastalDEM (Kulp and Strauss, 2019), and comparative estimates have been generated in order to better capture uncertainties (CIESIN and CUNY CIDR, 2020 forthcoming). Not to mention, the economic impacts of sea level rise could be considerable. A recent model that used GRUMP data suggested that without further adaptation, a high ice-melt scenario could lead to sea level rise that could cut GDP by up to 4% (Schinko et al., 2020).

WATER AVAILABILITY

As significant as climate change might be, future water access may depend much more on population change. According to a study using gridded population data, if the worst climate projections are realized along with the greatest population growth, billions of people could experience very severe water stress by 2100. The study combined LandScan data with projected changes in precipitation to analyze the complex interactions between population and climate (Parish et al., 2012). Socioeconomic storylines were used to predict population growth for the rest of the century, and researchers then modeled per capita water availability for each of the world's 26,929 watersheds. Because water availability is a transnational issue, gridded data allowed for processes to be modeled according to natural geographies instead of artificial boundaries. While precipitation could change drastically over the 21st century, with some regions becoming much drier and others wetter, simulations with different combinations of population and climate projections suggest that per capita water availability could be more a function of population growth. Another study that also used LandScan data in a more advanced economic model calculated that potentially 8.6 billion people, nearly two-thirds of the projected population, could eventually face water stress (Kiguchi et al., 2014). Such approaches are subject to uncertainty, but studying projections of population and climate distribution together allow us to identify the most vulnerable areas in advance and plan accordingly.

1.4 EXPANDING USES & UNDERSTANDING THE OPTIONS

As demonstrated, gridded population data can be a powerful tool for understanding and improving the conditions under which people live. Not only can it be an indispensable aid when responding to disasters, it can also help us to understand the risk of disease and predict the impact of future environmental change. Moreover, it can support the monitoring of indicators that the global community has identified as necessary for achieving sustainable development. Yet gridded population data's potential is complicated by the diverse set of data products available and the nuances of how and where they might be applied. It is important that users understand the unique features of each product to ensure that they select the most appropriate option. These issues are explored in the next chapter.

CHAPTER 2

GETTING TO KNOW GRIDDED POPULATION DATASETS

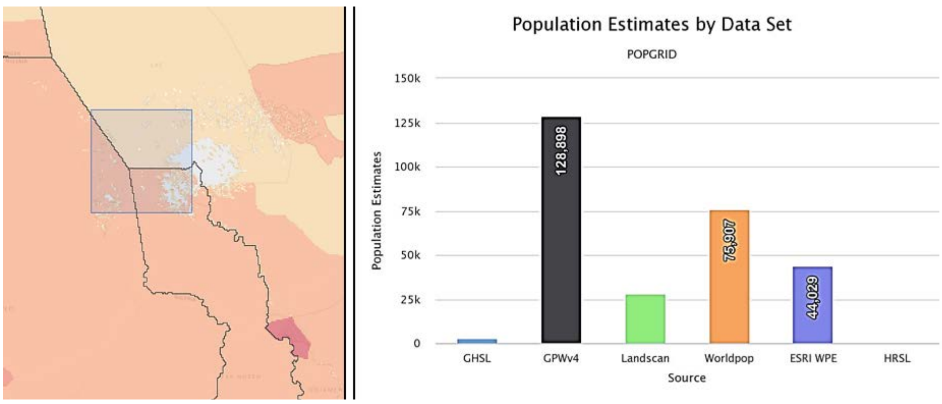
As described in Chapter 1, the WFP has employed gridded population data to mobilize its emergency responses. The ADAM platform, however, actually relies on two different gridded population datasets. After testing and comparing, WFP determined that WorldPop was best suited for most of its applications, while LandScan was preferred for island contexts. This decision underscores how – although they may all offer measures of population – these different products are fundamentally, methodologically distinct. Users should work to understand the characteristics of these products to ensure their specific data needs are being met.

Although a number of gridded population data products are now regularly used by a variety of organizations, they may each offer unique results. This is not to say that one product is systematically more appropriate than the rest. The suitability depends on the context and the type of questions that need answering. However, simply using the first available product could give misleading information (see the example highlighted below for an area bordering Chad, Nigeria, and Cameroon in Figure 1). Selection requires measured consideration, and this begins with understanding how the different data products are calculated and how they are different. In this chapter, we explain the key variations in population data models. We then go on to present an overview of seven leading products, describing the underlying data, the methods (top-down) and basic assumptions, and the corresponding strengths and limitations.

2.1 VARIATIONS IN MODEL INPUTS

The production of different gridded datasets introduced in this section generally follows similar procedures for distributing population data across space. Yet the estimations are built off of different types of data and feature different calculation methods. There are three key factors that describe how a given gridded population product is produced: 1) Input data; 2) data adjustment; and 3) modeling method.

FIGURE 1: Varying population estimates for an area on the Chad-Nigeria-Cameroon border west of Lake Chad.



Source: https://sedac.ciesin.columbia.edu/mapping/popgrid/comparison-view/

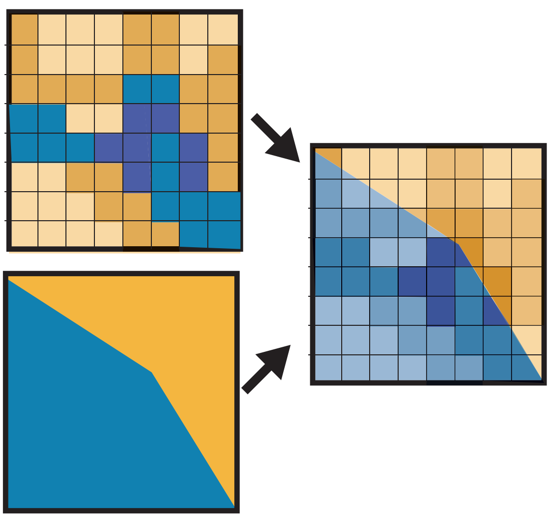
INPUT DATA

Census data, administrative boundary data, and geospatial correlates are the three major inputs used to produce gridded population datasets.

CENSUS DATA – record the official count of a particular population for an administrative and or census unit (geographic area representing a subnational entity, e.g., state, province, district, city, town, census tract), as well as the demographic information about people who live there. Censuses are meant to be carried out at least every ten years, but a number of countries struggle to do so, often due to the cost of these comprehensive surveys, the inaccessibility of certain populations, or conflict and insecurity, which leaves them with outdated or insufficient data (UNSD, 2017b). Additionally, the administrative units frequently change between censuses, including splitting and merging of units and changes in boundaries and names, as a result of urban growth, political and economic changes, and other factors. Although census data are generally collected at the household level, NSOs typically only release aggregated data for specific administrative levels—due to confidentiality and privacy norms. In some cases, they may release more detailed individual and household characteristics for anonymized samples of households (census microdata)[3].

ADMINISTRATIVE BOUNDARY DATA – consist of spatial information about the area of the Earth's surface that corresponds to each administrative unit in the census. These data are critical to spatialize the census counts and enable them and other variables to be distributed across a regular grid of geographic cells or "pixels (Figure 2)," an approximately square-shaped unit usually defined by latitude-longitude coordinates. The size of a cell can vary between tens or hundreds of square meters to tens or hundreds of square kilometers, and varies with latitude (a one-degree grid cell at the equator contains 12,364 sq km and at a 45o latitude contains 8,743 sq. km). When multiple administrative units overlap with a cell, a formula is needed to estimate a single value (e.g., total population) for that cell, such as the average of the unit values weighted by the areas of overlap. Similarly, the total population or other characteristic of an administrative unit needs to be assigned to all of the cells that overlap with it, either proportionately to the area of overlap or using some other formula. Complications may occur due to ambiguities, uncertainties, changes, and errors in boundary data, such as dynamic coastlines, disputed territories, borders defined by rivers or other dynamic features, and spatial accuracy and scale issues (census microdata).

Figure 2: Visual representation of distributing census counts across a grid of geographic cells or "pixels."



[3] A collection of census microdata is available from University of Minnesota's IPUMS International Project https://international.ipums.org/international/

GEOSPATIAL CORRELATES (ancillary inputs) – are spatial data that are known (or likely) to be correlated with population distribution, and therefore can be used to improve predictive accuracy (Lloyd et al., 2017). Rather than using simple statistical formulas that assume that populations are spread evenly within administrative units, researchers have developed a range of models that utilize these geospatial correlates to identify areas with above-or below-average population densities. The datasets featured in this section use different modeling approaches to reallocate population appropriately, while maintaining the overall population totals for known administrative units. The following variables have been used or tested in the development of gridded population data products:

- Built structures and impervious surfaces
- Known cities and settlements (e.g., populated places)
- Topographic data about elevation, slope, and coastlines
- Infrastructure data on energy, communications, health, sanitation, education, and other facilities
- Road, rail, transit, and other transportation networks
- Land cover and land use data on agriculture, industry, forests, etc.
- Night-time lights captured by satellites
- Protected areas (e.g., national parks and wildlife refuges)
- Water bodies, including rivers, lakes, seas, and oceans

DATA ADJUSTMENT

Input data for a gridded population dataset must be adjusted to match the desired time period, such as a specific day or year and the corresponding national population totals.

Since censuses are typically conducted only once per decade and for different years and dates in different countries, population estimates must be adjusted to a consistent time period to enable cross-national gridding. These adjustments are commonly based on the United Nations Population Division's (UNPD) estimates and projections, thereby ensuring at least general consistency with international standards. UNPD's estimates start with the base population for July 1, 1950, and then provide estimates of population by age and sex at five-year intervals. These estimates are based on a cohort-component method that projects rates of fertility, mortality, and migration for each group of people born in a given year (UNPD, 2017). The estimates are evaluated using a range of data sources, including censuses, post-enumeration surveys, household surveys, vital statistics and population registers (UNPD, 2017). The UNPD usually revises its World Population Prospects every two years, and the latest version is from 2019 (UNPD, 2019). Adjustments at the subnational level are generally based on the overall population change at the national level.

MODELING METHOD

The methodologies for allocating population counts across cells vary between datasets; they utilize empirical correlations between different variables and select different criteria for allocating population to cells.

AREAL WEIGHTING – is the longest standing approach to population gridding. In this method, the population of the census reporting unit is divided evenly by the number of grid cells in the unit. If cells cross boundaries between units, then those cells are allocated a smaller fraction of the population that is proportional to the land area of the cell within the unit. The sum of the population in all cells that constitute the unit will be equal to the population of the unit (Hallisey et al., 2017). Thus, the accuracy of the population count within a grid cell depends heavily on the size of the census reporting units – generally, the smaller the units, the more accurate the population counts are for each grid cell.

By contrast, most estimates provided by partners in POPGRID take a top-down approach that disaggregates population data into smaller units than those usually reported in censuses (Lloyd et al., 2019).[4] While most models follow the top-down approach, they vary in the level and method of disaggregation. Modeling methods used for the datasets in this chapter include:

DASYMETRIC MAPPING – evenly distributes population across spatial units. "No data" is allocated to cells where people are not likely to live, and the population count is allocated to cells where people are likely to live. This method can be applied in different ways:

- **Binary Dasymetric Mapping –** takes dasymetric mapping one step further, as it relies on an additional variable, an ancillary dataset, to spatially disaggregate census data from larger administrative units to smaller cells in order to develop higher resolution data products. Unlike areal weighting, this method does not assume homogeneity of population density across all cells that cover an administrative unit. This method excludes uninhabited areas when distributing population across inhabited areas; inhabited areas are determined cell by cell on a binary basis (Eicher and Brewer, 2001).

- **Multi-variable Dasymetric Mapping –** employs the same concept as binary dasymetric mapping, but relies on multiple geospatial correlates or (ancillary datasets) to produce high resolution estimations.

- **Random Forest –** the variables used in this modeling approach represent the variety of variables available to select from (e.g., the different trees available in the forest, hence the name of the method). The variety allows an algorithm to identify which one of the "random" variables is most relevant to population density. Random Forest uses variables and cross-validation techniques to assess their relative weights in any given country. Through cross-validation and training data, the algorithm "learns" which covariate or variable should have the highest weights in any given location. The covariates are, therefore, selected "randomly" for each location. The algorithm calculates population density by dividing the sum of census counts by the area of the aggregated units. Uninhabited areas (e.g., protected areas, water bodies) are excluded[5] in this method. This method offers a better correlation between population density and geographic distance-based covariates (e.g., proximity to roads) (Stevens et al., 2015).

4 It is important to note that this disaggregation depends on the size of the census units and on the cell size in the gridded product.

5 While the model excludes uninhabited areas, it does not mean there are no people living in parks and other locations that may not be commonly recognized as inhabited places.

TOC

2.2 OUTPUT CHARACTERISTIC

Gridded population products have different spatial resolutions. The spatial resolution is the size of the grid cell, though as discussed elsewhere, this spatial resolution is generally much finer than the input resolution – which is comprised of census units of varying sizes and shapes. The resolution is represented in meters or arc-seconds[6], which are common units of measurement in geographic reference systems. Thirty arc-seconds at the equator is approximately 1 km. Grids are typically produced only for the continents and islands, excluding Arctic and Antarctic regions.

2.3 REVIEW OF GLOBAL GRIDDED POPULATION DATASETS

Gridded population datasets utilize a variation of different data inputs and modeling methods to produce population estimates. For a more technical review and assessment of POPGRID datasets using a "fitness for use" lens, readers are directed to Leyk et al. (2019), "The spatial allocation of population." The following section presents a brief overview of seven datasets developed by organizations in POPGRID (Table 2). The datasets are listed by the approximate level of modeling. Note that this review only necessarily represents a snapshot in time, and readers should visit the data providers' websites to obtain documentation on the latest versions of each dataset.

6 The arc-seconds measurement assumes that the globe is divided into 360 equal parts (degrees), and each part is subdivided into 60 minutes, and subsequently each minute divided into 60 seconds. An arc-second measures the latitude or longitude traveled across the surface of the earth in one second and represents the cell size (Esri, n.d.).

GRIDDED POPULATION OF THE WORLD (GPW) V4

In 1995, the first version of GPW was developed by the geographer Waldo Tobler at the University of California-Santa Barbara, with support from the Consortium for International Earth Science Information Network, then a nongovernmental, nonprofit organization based in Michigan. Subsequent versions of GPW were developed by and disseminated through the NASA Socioeconomic Data and Applications Center (SEDAC) operated by the Center for International Earth Science Information Network (CIESIN), now a center of the Earth Institute at Columbia University based in New York. Version 4 was released in 2017, and 4.11 in 2018. With minimal input variables and modelling, GPWv4 is the input source for a number of other datasets, including Global Human Settlement Layer-Population and WorldPop.

POPULATION DATA:

2010 round of population and housing Censuses, which occurred between 2005 and 2014.

GEOSPATIAL CORRELATES:

Protected areas and water bodies.

SATELLITE IMAGERY:

None.

SOURCE OF DATA ADJUSTMENT:

Official country totals for censuses and UNPD estimates.

METHOD:

Areal Weighting.

YEARS REPRESENTED IN ESTIMATES:

2000, 2005, 2010, 2015, and 2020.

RESOLUTION:

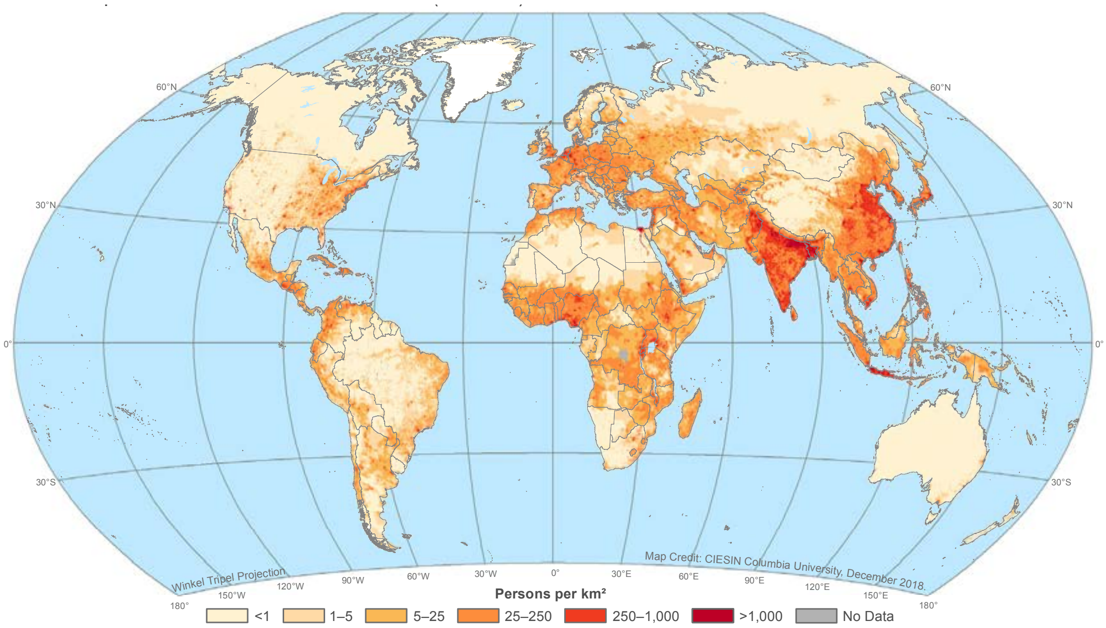
30 arc-seconds (1 km).

DISTRIBUTION POLICY:

Open access (Creative Commons Attribution 4.0 International License).

STRENGTHS AND LIMITATIONS:

GPWv4 is one of two datasets that provide estimates of basic demographic characteristics (age and sex[7]) in addition to population counts and density (the other is WorldPop). Over the years, the availability of subnational census data has significantly increased and improved the output resolution (Aubrecht et al., 2016; Doxsey-Whitfield et al., 2015). GPW is a minimally modeled dataset that remains true to the census and is, therefore, a suitable choice for large-scale analysis across regions, and for analyses that require an independent measure of population that is not affected by covariates used in most other modeling approaches. However, as a result of its minimally modeled approach, it tends to show much higher population counts in, for example, rural areas of countries that only report census data for larger geographic units (see Figure 1 above for the Nigeria-Chad-Cameroon border).



Source: CIESIN

[7] It is important to recognize that in some cases, these estimates are based on census data from larger administrative units that are downscaled to smaller units. For example, age distributions and sex ratios for a province are applied to all districts within the province. Age and sex data for small census units (e.g. enumeration areas) could be unavailable because of confidentially rules.

TOC

GLOBAL RURAL URBAN MAPPING PROJECT (GRUMP)

CIESIN, the International Food Policy Research Institute (IFPRI), the World Bank, and Centro Internacional de Agricultural Tropical (CIAT) developed the Global Rural Urban Mapping Project and released the GRUMP data collection in 2007. This dataset proportionally allocates population data to housing and settlement extents.

POPULATION DATA:
GPWv3's subnational census data for the 2000 round censuses (1995-2004).

GEOSPATIAL CORRELATES:
Nighttime lights for larger settlements, buffered points for smaller settlements, protected areas, and water bodies.

SATELLITE IMAGERY:
Defense Meteorological Satellite Program Optical Linescan System (DMSP-OLS) nighttime lights imagery for 1995.

SOURCE OF DATA ADJUSTMENT:
UNPD estimates.

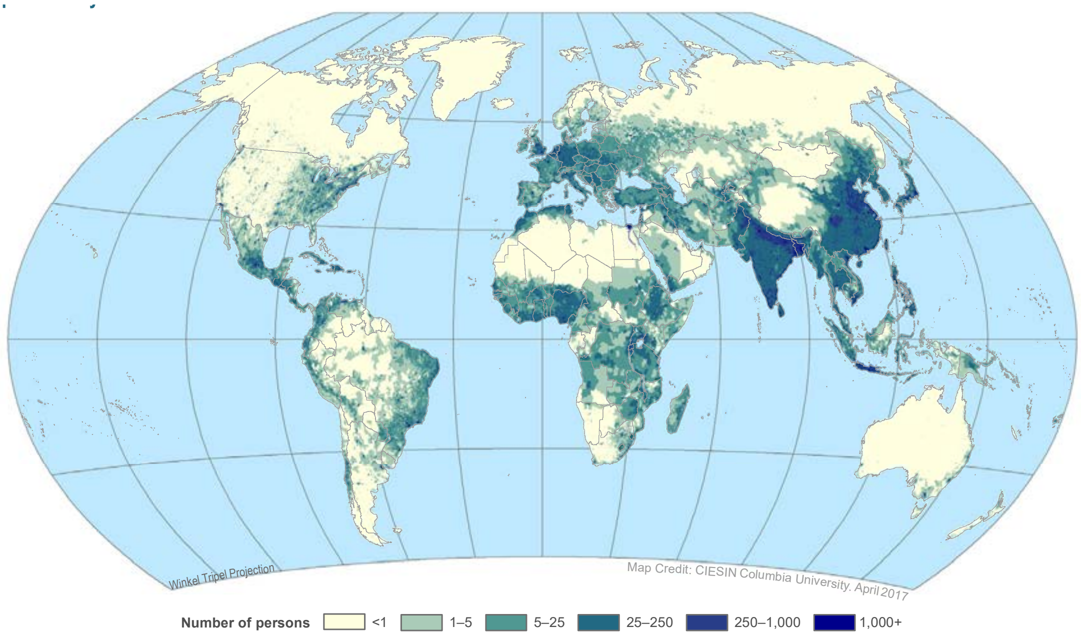
METHOD:
Binary dasymetric mapping.

YEARS REPRESENTED IN ESTIMATES:
1990, 1995, and 2000.

RESOLUTION:
30 arc-seconds (1 km).

DISTRIBUTION POLICY:
Open access (non-commercial use)

STRENGTHS AND LIMITATIONS:
GRUMP was one of the first global gridded products to transparently model urban and rural populations, and its ancillary datasets (urban extents and populated places) have been used extensively. A potential limitation is the coarse resolution and "blooming effect" of DMSP-OLS nighttime lights, which leads to overestimations of urban extents in some areas (Aubrecht et al., 2016, Gunasekera et al., 2015, Schneider et al., 2010). GRUMP has proven to be suitable for estimations of populations at risk and in low elevation coastal zones due to its finer spatial resolution and for use in many studies that required separate urban and rural estimates (Mondal and Tatem, 2012). GRUMP has since been superseded by other data products listed below.



Source: [CIESIN]

GLOBAL HUMAN SETTLEMENT LAYER - POPULATION (GHS-POP)

The European Commission Joint Research Centre (JRC) and CIESIN published the first version of the Global Human Settlement Layer - Population (GHS-POP) dataset in 2015. The latest release was in 2019. This dataset combines information from population censuses with satellite-derived built-up areas to allocate population counts from administrative units to a grid according to the presence or absence of built-up area in the grid cell. The information below is for the latest release.

POPULATION DATA:
GPWv4.10 data.

GEOSPATIAL CORRELATES:
Built-up areas mapped in the Global Human Settlement Layer Built Up Extent (GHS-BUILT).

SATELLITE IMAGERY:
Landsat imagery for 1975, 1990, 2000, and 2014.

SOURCE OF DATA ADJUSTMENT:
UNPD estimates.

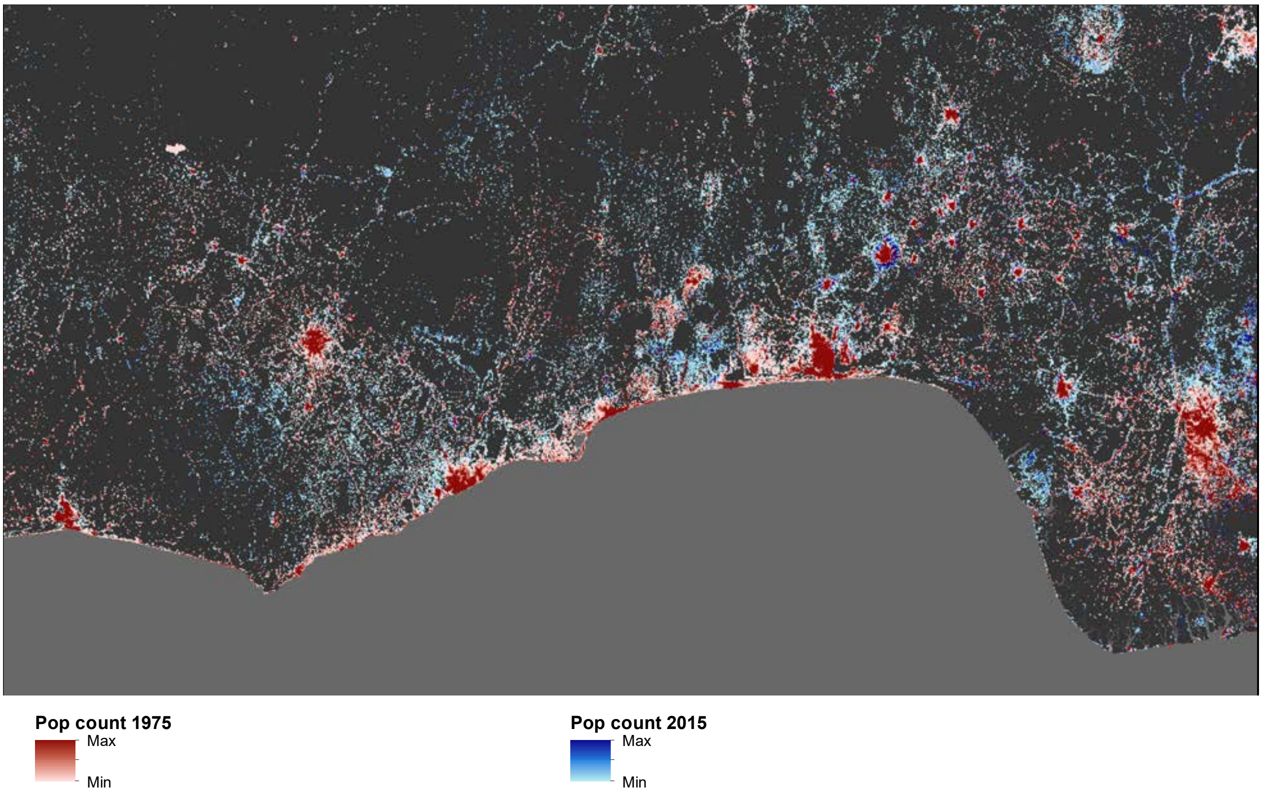
METHOD:
Binary dasymetric mapping.

YEARS REPRESENTED IN ESTIMATES:
1975, 1990, 2000, and 2015.

RESOLUTION:
250 m and 1 km in World Mollweide, 9 arc-seconds and 30 arc-seconds in World Geodetic System 1984 (WGS84).

DISTRIBUTION POLICY:
Open access (Creative Commons Attribution 4.0 International License)

STRENGTHS AND LIMITATIONS:
GHS-POP is developed on the premise that grid cells that have more built-up land will be associated with larger populations. This dataset employs a clear and reproducible method, whose primary strength is its globally-consistent time series, which makes it well-suited for analyses extending over multiple years (Freire et al., 2018). A weakness is its sole use of built-up mapping as detected from satellite imagery. Thus, the areas that are not detected as built-up only receive population counts declared in the census unit (GPWv4), even though there may be housing units that are too small to detect using moderate resolution remote sensing in some rural areas. Therefore, in some rural areas GHS-POP may show no population where in fact there are small, dispersed habitations.



Source: JRC

WORLD POPULATION ESTIMATE (WPE)

Esri's World Population Estimate (WPE) is a collection of global population datasets that were released between 2014 and 2017. WPE represents where people live (defined as the "nighttime" population) as opposed to where they are located during the day. The collection was developed for the purpose of consumer economic analysis, as where people live impacts consumer spending patterns.

POPULATION DATA:

Recent year commercial estimates based on censuses of various countries and United Nations Data.[8]

GEOSPATIAL CORRELATES:

Road intersections and populated place points from GeoNames.[9]

SATELLITE IMAGERY:

Landsat 8 imagery.

SOURCE OF DATA ADJUSTMENT:

Country-official estimates with 138 countries processed further by Michael Bauer Research GmbH.

METHOD:

Dasymetric mapping.

YEARS REPRESENTED IN ESTIMATES:

2013, 2015, and 2016.

RESOLUTION:

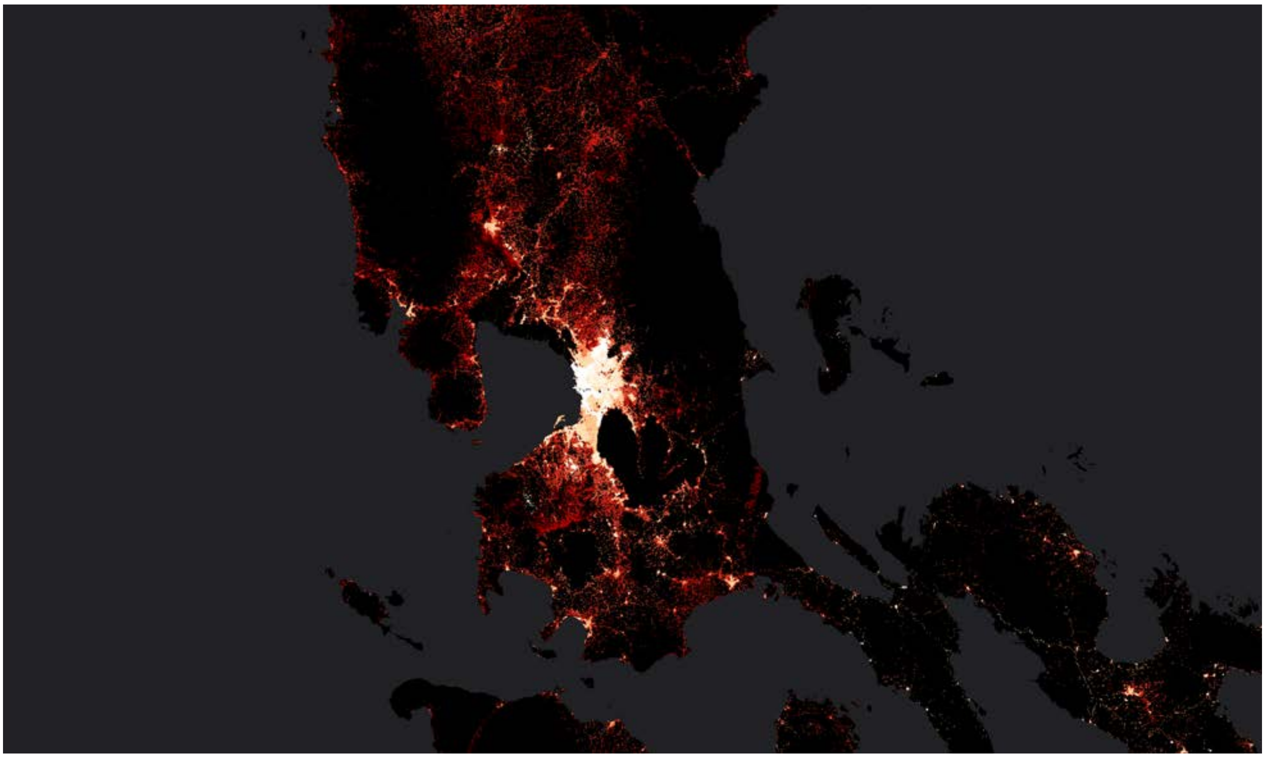
150 m in 2016 and 250 m in earlier years.

DISTRIBUTION POLICY:

Commercial (free for those with named user accounts in ArcGIS organizations).

STRENGTHS AND LIMITATIONS:

The method used for each year's estimation changes, therefore the releases cannot be compared across time. WPE is suited for large-scale analysis, economic, business, and trade analysis and estimating the distribution of party affiliations, human impact on the environment, and the location of populations affected by disasters, disease outbreak, and access to infrastructure (Frye, 2017). Detailed documentation on the methods and metadata at the country level is available for all datasets.



Source: Esri

8 The commercial estimates are provided by Michael Bauer Research GmbH, which starts with the most recent census and produces a current year estimate. WPE uses this in 138 countries and uses United Nations data for other countries.

9 Populated place points can be collected from a provider such as GeoNames, which extract 'population places' from a database of 25 million geographical names and consists of millions of unique features. Population places indicate a higher possibility of settlement within a short radius.

HIGH RESOLUTION SETTLEMENT LAYER (HRSL)

In a collaboration between Facebook, CIESIN, and the World Bank, the first version of the High Resolution Settlement Layer (HRSL) was released in 2017 and covered 30 countries. Facebook uses computer vision techniques to identify buildings from commercially available satellite images. In the next stage, CIESIN uses the building information and census data to produce population estimates. The methodology was validated in collaboration with the World Bank, using their Living Standards Measurement Study (LSMS) program, and anonymized ground truth household surveys as a basis for comparison with the gridded population dataset. An updated version was produced in 2018-2019, covering approximately 160 countries and included age and sex breakdowns.

POPULATION DATA:
GPWv4.

GEOSPATIAL CORRELATES:
Built structures.

SATELLITE IMAGERY:
DigitalGlobe (now Maxar).

SOURCE OF DATA ADJUSTMENT:
See GPWv4.

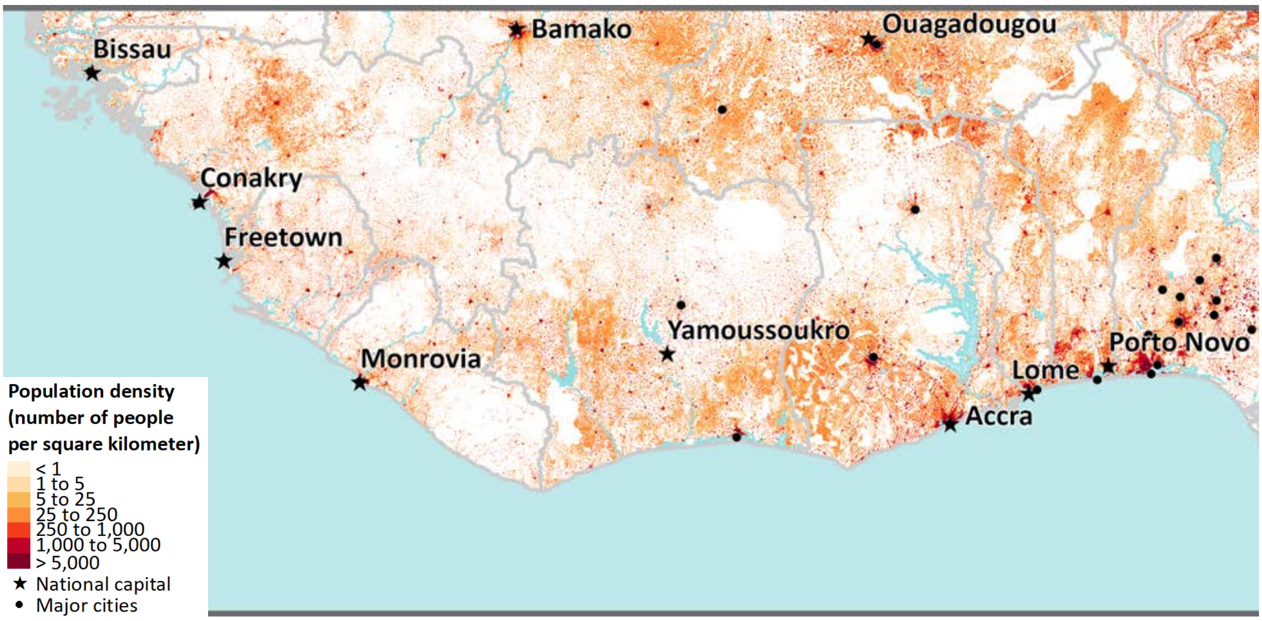
METHOD:
Binary dasymetric mapping.

YEARS REPRESENTED IN ESTIMATES:
2015.

RESOLUTION:
1 arc-second (approximately 30 m).

DISTRIBUTION POLICY:
Open access.

STRENGTHS AND LIMITATIONS:
The population grids in this dataset provide a detailed presentation of settlements in urban and rural areas and can benefit research conducted for disaster response, humanitarian planning, and other applications. It is important to note that HRSL does not differentiate between residential and commercial building features. However, in an assessment of gridded population datasets in six countries, HRSL's built area correlates proved to be the most important across all countries (Reed et al., 2018).



Source: CIESIN

LANDSCAN

Oak Ridge National Laboratory (ORNL) produced its first LandScan dataset in 1998, and has produced annual updates since then. The latest version of LandScan datasets employs a "smart interpolation" technique, in which census data and geospatial or ancillary datasets are integrated to estimate population distribution. To date, the 2018 dataset is the most recent version available.

POPULATION DATA:
Subnational census and administrative data.

GEOSPATIAL CORRELATES:
Roads, land cover, built structures, cities or urban areas, infrastructure, environmental data, protected areas, and water bodies.

SATELLITE IMAGERY:
Maxar (formerly DigitalGlobe).

SOURCE OF DATA ADJUSTMENT:
US Census Bureau International Programs estimates.

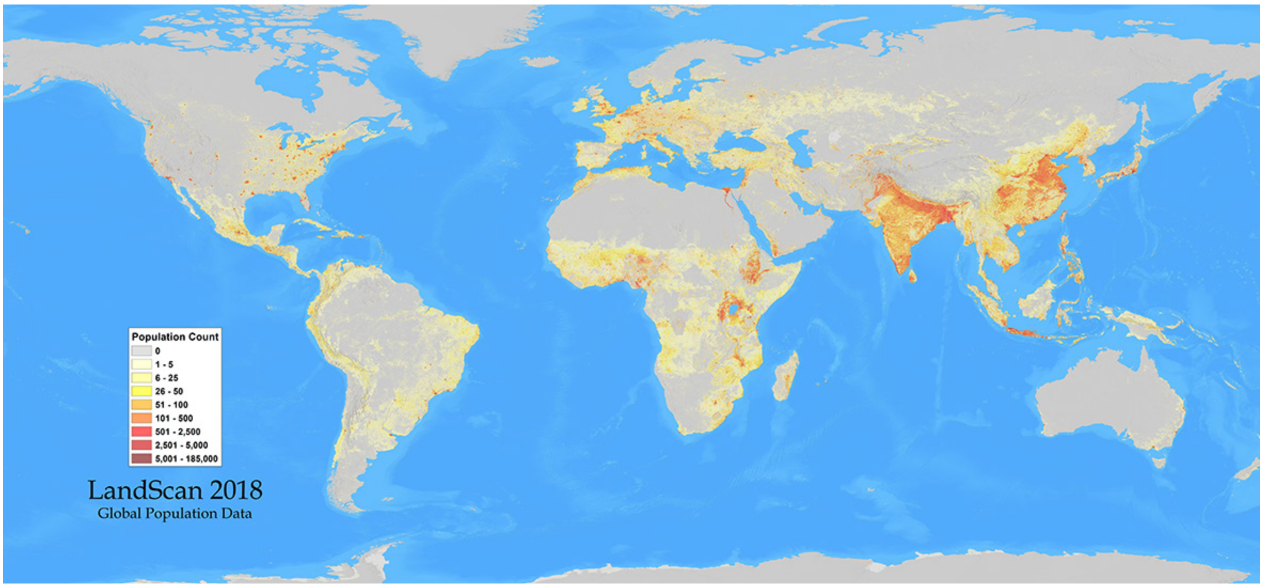
METHOD:
Multi-variable dasymetric modeling; however, the same method is not used across all countries and time.

YEARS REPRESENTED IN ESTIMATES:
Annual datasets since 1998.

RESOLUTION:
30 arc-seconds (1 km)

DISTRIBUTION POLICY:
Open access for academic research, humanitarian purposes, and for agencies or staff of the US government and mission partners. Commercial license available for other entities.

STRENGTHS AND LIMITATIONS:
One of LandScan's distinct features is that it models the "ambient population," meaning that for regions where commuting is prevalent, the population grid is somewhere between a nighttime residential population (which is what censuses measure) and a daytime population (where people work). LandScan uses complex modeling tailored to each country to weight cells for possible population presence throughout the course of a day. This can be useful for risk assessment and disaster response for hazards that may occur during the daytime. Due to these varying input data and methods, care is needed in comparing LandScan estimates with other models, especially for smaller regions or urban areas where ambient and nighttime populations may differ greatly. LandScan does include 5-year age-sex demographics at the level 1 administrative unit. LandScan 2012 also performs well in capturing urban population distribution, but is less reliable in rural areas (Gunasekera et al., 2015). This is partly due to shortcomings of satellite imagery in certain environments that make detecting settlements challenging (e.g., houses may blend into the landscape or be shrouded in trees). In addition, documentation of data and methods has been limited, and data are not fully open to all users (Gunasekera et al., 2015). However, LandScan is planning on moving to fully open access in the future. LandScan develops new models for each year's estimates that may preclude the use of the data as a consistent time series for small areas.



Source: ORNL

WORLDPOP

Initiated in 2012, WorldPop combined its AfriPop, AsiaPop, and AmeriPop population mapping projects to offer global coverage of population. WorldPop produces gridded population data using machine learning algorithms.

POPULATION DATA:
GPWv4 census data.

GEOSPATIAL CORRELATES:
Roads, land cover, built structures, cities or urban areas, nighttime lights, infrastructure, environmental, protected areas, and water bodies.

SATELLITE IMAGERY:
Landsat Enhanced Thematic Mapper (ETM), Landsat, TerraSAR-X, TanDEM-X, DMSP and Visible Infrared Imaging Radiometer Suite (VIIRS).

SOURCE OF DATA ADJUSTMENT:
Country-official estimates and UNPD estimations.

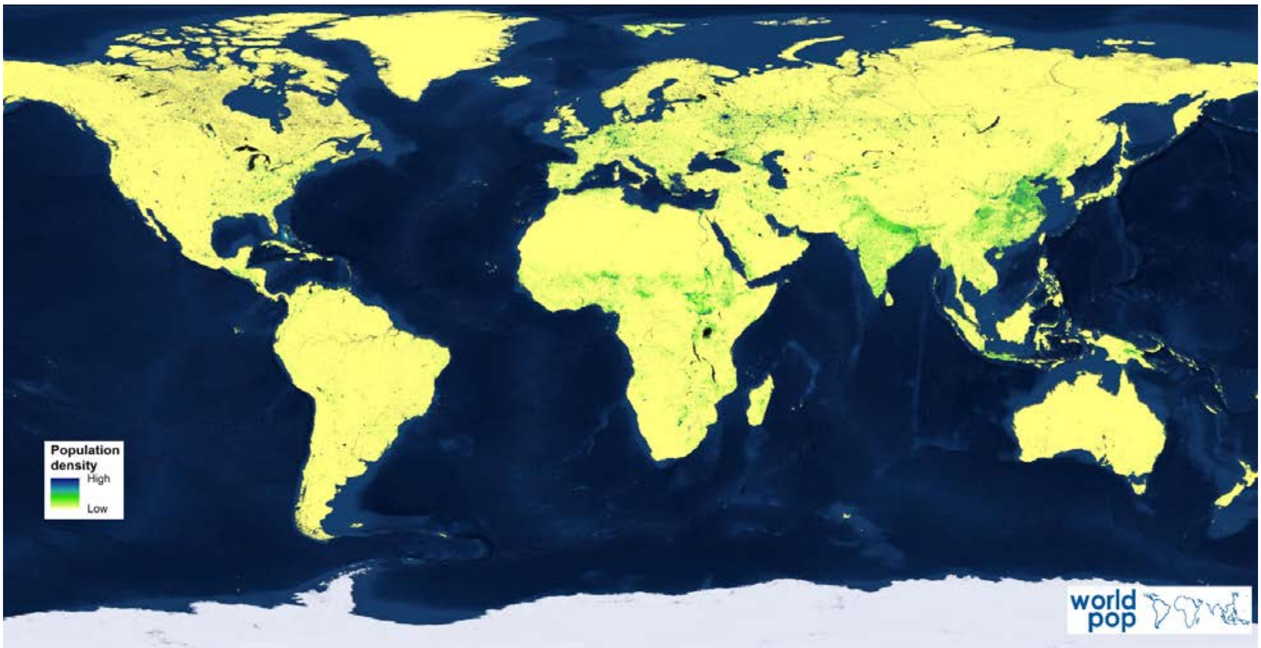
METHOD:
Random forest; however, the same method is not always used across all countries and time.

YEARS REPRESENTED IN ESTIMATES:
Every year from 2000-2020.

RESOLUTION:
3 arc-second (100 m)

DISTRIBUTION POLICY:
Open access.

STRENGTHS AND LIMITATIONS:
A strength of WorldPop's model is its ability to use machine learning to identify significant relationships from the input census data and to omit a proportion of the rural population in areas without apparent satellite-derived built-up areas; this is due to the fact that satellites will inevitably miss some populated places. WorldPop also publishes all source code and is transparent about methods, so that users can develop their own grids using different sets of assumptions. WorldPop also integrates a variety of input and ancillary data so that the model redistributes population counts using different weights across census or administrative units (Stevens et al., 2015). Similar to GPWv4, WorldPop includes demographic breakdowns according to age and sex, making it a strong contender to use for population-related SDG indicators, and also provides a 100m resolution dataset for each year from 2000 to 2020. This model does not mask non-built-up areas, so that it can produce non-zero population estimates in deserts and forests. However, because it clusters populations, the model has been criticized for excluding more uninhabited land compared to other datasets (Aubrecht et al., 2016). It is important to highlight that the global dataset described here is designed to be comparable over time.



Source: WorldPop

TABLE 2 POPGRID GRIDDED POPULATION DATASET METRICS

	GPWV4	GRUMP	GHS-POP	WPE	HRSL	LANDSCAN	WORLDPOP
POPULATION DATA	2010 Population and Housing Censuses	GPWv3's subnational census data for the 2000 round censuses	GPWv4.10 data	Censuses of various countries, commercial and UN data	GPWv4	Subnational census and administrative data	GPWv4 census data
GEOSPATIAL CORRELATES	Protected areas and water bodies	Nighttime lights for larger settlements, buffered points for smaller settlements, protected areas, and water bodies	Built-up areas mapped in the Global Human Settlement Layer Built Up Extent (GHS-BUILT)	Road intersections and populated place points from GeoNames	Built structures	Roads, land cover, built structures, cities or urban areas, infrastructure, environmental data, protected areas, and water bodies	Roads, land cover, built structures, cities or urban areas, nighttime lights, infrastructure, environmental, protected areas, and water bodies
SATELLITE IMAGERY	None	Defense Meteorological Satellite Program Optical Linescan System (DMSP-OLS) nighttime lights imagery for 1995	Landsat imagery for 1975, 1990, 2000, 2014	Landsat8	DigitalGlobe (now Maxar)	Maxar (formerly DigitalGlobe)	Landsat Enhanced Thematic Mapper (ETM), Landsat, TerraSAR-X, TanDEM-X, DMSP and VIIRS
SOURCE OF DATA ADJUSTMENT	Official country totals for censuses and UNPD estimates	UNPD estimations	UNPD estimates	Country-official estimates with 138 countries processed further by Michael Bauer Research GmbH	See GPWv4	US Census Bureau International Programs estimates	Country-official estimates and UNPD estimations
METHOD	Areal Weighting	Binary dasymetric mapping	Binary dasymetric mapping	Dasymetric mapping	Binary dasymetric mapping	Multi-variable dasymetric	Random forest
YEARS REPRESENTED	2000, 2005, 2010, 2015, and 2020	1990, 1995, and 2000	1975, 1990, 2000 and 2015	2013, 2015, and 2016	2015	Annual since 1998	2000-2020
RESOLUTION	30 arc-seconds (1 km)	30 arc-seconds (1 km)	250 m, 1 km (World Mollweide) 9 arc-seconds, 30 arc-seconds (WGS84)	150m in 2016 and 250 m in earlier years	1 arc-second (approximately 30 m)	30 arc-seconds (1 km)	3 arc-second (100 m)
DISTRIBUTION POLICY	Open access (Creative Commons Attribution 4.0 International License)	Open access (non-commercial use)	Open access (Creative Commons Attribution 4.0 International License)	Commercial (free for those with named user accounts in ArcGIS organizations)	Open access	Open access for academic research, humanitarian purposes, and for agencies or staff of the US government and mission partners. Commercial license available for other entities.	Open access

2.4 ALTERNATIVE MODELING APPROACHES

The top-down approach to modeling in the described datasets relies on the availability of census data, despite the differences in scale, quality, and temporal characteristics of the datasets. Nevertheless, some data suppliers have been exploring methods of estimation that take a bottom-up approach to compensate for lack of reliable census data (Wardrop et al., 2018; Weber et al., 2018). The bottom-up approach calculates population count and density from ancillary data and small-scale data collection on the ground (Reed et al., 2018). For instance, WorldPop employs the bottom-up approach with satellite-derived features and household surveys where census data are outdated, unreliable, or unavailable. As part of the GRID3 (Geo-Referenced Infrastructure and Demographic Data for Development) initiative, WorldPop is working to combine data from localized population counts with satellite imagery to estimate populations for areas where census data is outdated. To date, the bottom-up approach has been applied to population estimates in Nigeria, Afghanistan, Colombia, and the Democratic Republic of the Congo. LandScan HD (3 arc-second) also uses a bottom-up approach to produce population data at 3 arc-second resolution. Although the data at that resolution are not accessible to users, they are aggregated to a 30 arc-second resolution and used in LandScan Global.

BOX 3 | **USE OF NEW DATA SOURCES**

In nearly all of GPSDD's country interviews, NSO representatives indicated that they were unfamiliar with the gridded population data described in this report. The one exception was Colombia, which collaborated with the UN Population Fund and the University of Southampton to develop block-level population estimation models (using a bottom-up approach), which have the potential to reduce the uncertainty in estimations for challenging-to-reach areas.

Although gridded population data might be less widely recognized, many countries are engaging with alternative data sources. Tanzania is reportedly in conversations with the World Bank to strengthen the use of geospatial data for its National Bureau of Statistics, although this does not currently extend to population estimates. Several countries are also exploring new uses for administrative data, data that are collected as part of general record keeping but can then be repurposed to calculate statistics. For example, Paraguay currently estimates population changes based on birth, death, and migration records, and it plans to do an administrative-data-based census. Ecuador also plans to compare its 2020 census results with administrative data. Colombia, Costa Rica, and Ghana also use administrative data for estimating population changes. Additionally, Ghana's statistical service has partnered with the telecommunications company, Vodafone, to access aggregated call data records that will be used to understand changes in mobility and public service access. With novel data sources now being investigated and actively used, gridded population data can add to the growing toolbox available to decision-makers.

(GPSDD Country Partners, 2019)

2.5 FINDING THE RIGHT FIT

There are now several ways to estimate population distribution, with different data sources, underlying methods, and output resolutions. Policymakers, academics, and other stakeholders have a range of products to choose between, and comparing the modeling approaches is the first step in making this selection. Yet, understanding how the outputs of these models compare against one another and against conditions on the ground is important for determining the appropriateness of a product for a desired application. In the next chapter, we review the existing knowledge on comparing and validating these population data products. Moreover, we explain the need for additional validation work to improve confidence in different data products and their fitness for use for different applications.

CHAPTER 3

INTER-COMPARISON AND VALIDATION OF GRIDDED DATASETS

As described in Chapter 1, gridded population data can aid disaster response, household survey planning, and other key efforts. However, individuals may quickly discover inaccuracies and errors once they are actually on the ground (Thomson, 2019). Sometimes, the estimates have been substantially off the mark. Survey workers in Nepal, for example, found that WorldPop had incorrectly placed residents in a Parliament building, an airport, and even a Coca-Cola factory (Thomson, 2019). Also, in Namibia, large rock formations have been misidentified as houses, which are often overlooked in heavily vegetated parts of the world (Thomson, 2019). Experiences like these help inform future refinements, but they also remind us that model-based predictions of population are by no means perfect.

With a range of population datasets now available, it is important that we work to understand how data selection can impact research and decisions. Even more fundamentally, we need to test how accurate these data products actually are, comparing estimates against known conditions on the ground. Although uncertainty is inherent in population estimation, there is currently no accepted way to measure or communicate the levels of uncertainty associated with the data products now available. The research community has begun to tackle these questions, and their initial results offer some insights. However, there is a need for more systematic and extensive studies to develop objective measures of accuracy and uncertainty that can help users to utilize population estimates in appropriate ways. In the section below we describe how researchers are working to test, validate and improve upon their models.

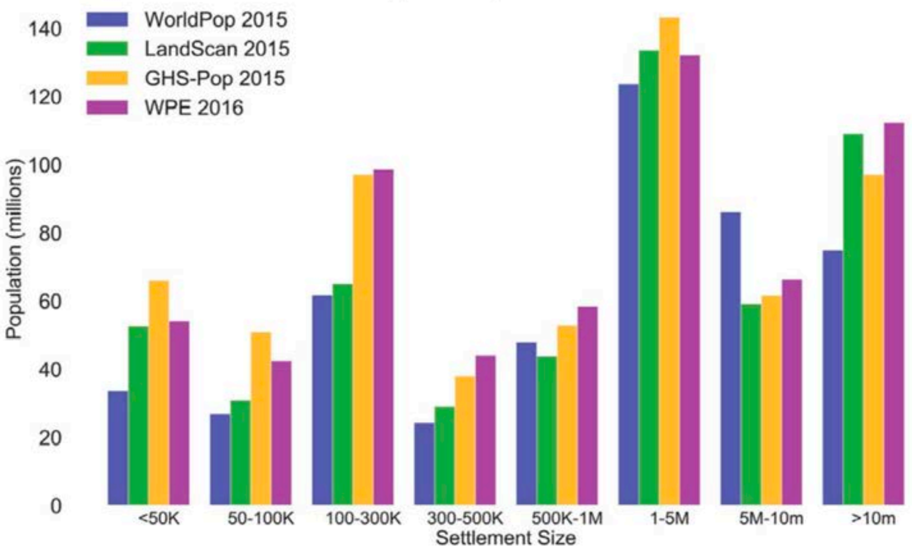
3.1 INTER-COMPARISON

Different population grids are built on alternative assumptions about how populations are distributed, so naturally the resulting estimates will be different as well. Objective comparisons of population grids can help us better understand the differences and limitations of the datasets and the nature of these differences. In the absence of authoritative validation data, it is not possible to determine which estimate is closest to the actual population, but inter-comparisons can illuminate the implications of model selection.

DESCRIBING DIFFERENCES IN ESTIMATIONS

Differences in population counts for most countries are insignificant, but they can be significant in smaller areas (Calka and Bielecka, 2019). A 2007 study compared LandScan Global 2004 and GPWv3 across multiple areas, including Washington DC, Los Angeles, Houston, London, and the country of Iran (Sabesan et al., 2007). Researchers found significant differences in downtown areas, where LandScan generally gave larger estimates, owing to its approach to measuring ambient population. Moreover, LandScan tends to have denser population clusters, whereas GPWv3 returns more spatially diffuse distributions. The authors concluded that LandScan appeared to capture greater geospatial variability, which suggests that it is more appropriate for disaster planning. In another study, researchers compared the population estimates of urban settlements in Africa across five different methods (Tuholske et al., 2019). Estimates of the total urban population for the continent range from 479.15 million for WorldPop 2015 to 608.89 million for WPE 2016 (see Figure 3). This underscores the uncertainty in identifying and measuring both urban and rural populations, specifically in developing countries.

Figure 3: Total Urban Population by Settlement Size for Africa



Source: Tuholske et al., 2019

IMPACTING OUTCOMES

The differences in population estimates can propagate through a model and affect the prediction of key decision variables. AfriPop was a predecessor to WorldPop specific to East Africa, and its developers performed a set of informal comparisons with LandScan 2008 and 2000 and GRUMP Beta (Linard et al., 2010). In particular, they tested the impact of model selection on estimations of malaria exposure in Somalia, and the contrasting models showed that GRUMP estimated a significantly higher proportion of the population would be exposed. AfriPop and LandScan gave similar spatial breakdowns, but there were large differences in absolute population figures, and GRUMP estimated a much larger number of people living in epidemic areas. Such differences can substantially impact decisions on public health measures or influence estimates of morbidity and mortality rates utilized in future planning for epidemics, among other critical decisions.

BOX 4 | ESTIMATING CLIMATE CHANGE IMPACTS

Predictions of the impact of climate change and sea level rise on vulnerable populations are also sensitive to product selection. According to Mondal and Tatem (2012), studies estimating coastal populations vulnerable to climate change and flooding often use high-resolution population datasets, but do so without considering the effects of the data selection. To test the effects, they adopted existing techniques for estimating coastal flooding and then interchanged LandScan 2008 and GRUMP 2000 v1. They found that variation in the total population at risk between GRUMP and LandScan was limited at the continental level, but there were individual countries with substantial differences. Developed countries with high-resolution input data showed little difference in estimations, while low-income countries showed larger differences. Moreover, most of the countries that were highly sensitive to model selection were small island countries, with over 25% differences in estimations. In short, the levels of uncertainty are highest for the populations that face the greatest risk from sea level rise. A more recent study examined coastal flood risk in 18 countries across Asia, Latin America, and Africa and compared the effect of selecting HRSL 2014, WorldPop (version unknown), or LandScan 2015, each of which was available at different spatial resolutions (30 m, 90 m, and 900 m respectively) (Smith et al., 2019). In all 18 countries, using the high resolution HRSL resulted in lower estimates of population exposed to a 100-year flood compared to WorldPop and LandScan, which showed reductions as high as 60%. The authors concluded that people naturally concentrate in less flood-prone areas, and such insights are necessary to plan for climate adaptation. These sensitivities can introduce considerable uncertainty to population-based research, but such uncertainty is often overlooked. Researchers should actively consider their model selections and be transparent about the associated uncertainties.

3.2 VALIDATION APPROACHES FOR POPULATION ESTIMATION

Population grids ultimately need to be validated against the population on the ground to ensure the most accurate estimates. Even if estimates are in close agreement with one another, this is not a guarantee that they are also in agreement with where people really live and work – especially in rapidly changing settings, such as peri-urban areas of developing countries. Testing this requires comparing estimates against authoritative population figures. Such a comparison, though, is challenging to undertake across many geographies. Although there have been a few validation studies at the national or local levels, validation at a global scale remains challenging in part because validation data are not currently available or accessible for many parts of the world (Calka and Bielecka, 2019; Engstrom, 2019).

COMPARING WITH CENSUS DATA

The most common validation approach in the literature is to make comparisons against available census data (Bai et al., 2018; Biljecki et al., 2016; Calka and Bielecka, 2019; Engstrom et al., 2019; Gaughan et al., 2013; Hall et. al., 2012; Hay et al., 2005; Linard et al., 2010). This approach is especially useful when census data can be made available at fine resolutions. Although population models run on aggregated census data, the outputs can be compared against more resolved data to test how accurately they distribute people. Georeferenced household census data is ideal for validation work (Hall et. al., 2012), but census agencies normally only release aggregated data (Chen, 2019). Additionally, validation is especially challenging when a model is produced at a higher resolution than the lowest administration level with available census data (Bai et al., 2018).

METHOD COMPARISON

All of the methods - whether they are lightly or more heavily modeled – will inevitably have some errors in their estimations, but we cannot characterize these errors systematically without validation. The areal weighting method used to produce GPW and GRUMP can be beneficial in that it makes minimal assumptions, but this feature can also expose it to additional estimation errors. Under the right conditions, dasymetric mapping methods can offer more realistic results (Calka and Bielecka, 2019). This dynamic is demonstrated in Poland, where researchers have tested the accuracy of the lightly modeled GRUMP data and the more intricately modeled LandScan 2012 data against a government-issued population reference grid, which was based on 2011 census data and created according to EU INSPIRE data specifications (Backer and Bloch, 2011; Calka and Bielecka, 2019; Da Costa et. al, 2017). In both cases, researchers first identified differences between the estimates and the reference data and then analyzed spatial patterns in these discrepancies. This approach revealed that GRUMP had significant disagreements with the reference, with only 1.5% of grid cells showing no difference between the two (Da Costa et al., 2017). In contrast, LandScan was in general agreement with the reference, with 10.5% of grid cells providing population estimates actually equal to the reference, and 72% of the country was covered by mostly reliable or highly reliable estimates (Calka and Bielecka, 2019). Similarly, another set of researchers combined data sources to precisely locate the one million inhabitants of Scania, Sweden, and then they up-scaled this data to a standard grid for direct comparison with GPWv3, LandScan, GRUMP, and the population density grid of EU-27+ (Hall et al., 2012). As in Poland, the LandScan estimate was on the whole closer to the reference population figures than were other methods.

Such validation work highlights patterns and issues with the estimation methods. One of the noted weaknesses of GRUMP was a tendency to overestimate the extent of urban settlements, which is a byproduct of its reliance on nighttime light data (Gaughan et al., 2013). This pattern is found in the above comparison studies. In Poland, GRUMP underestimated average population density by 50%, tending to make underestimations in city centers and overestimations in the suburbs (Da Costa et. al., 2017). LandScan has its own biases as well; for example, it overestimates Poland's city centers, while underestimating in other dense areas and the suburbs (Calka and Bielecka, 2019). It should be noted, however, that LandScan estimates the ambient (24 hour average) population, so one would expect it to give a population distribution different from the traditional census. These patterns are shifted somewhat in Scania, Sweden, where both GPWv3 and GRUMP overestimated in cities and underestimated in the suburbs, whereas LandScan's accuracy varied with geography according to factors like agricultural land use (Hall et al., 2012). These validation studies show not only how one model selection may be more appropriate for a given area or country and how the reliability can vary within the country, but also the need to improve methods and data for meaningful validation. It is also important to utilize the most recent versions of data available, given the frequent improvement in methods and data (e.g., GRUMP and GPWv3 have both been superseded by more recent versions or alternative datasets).

REGIONAL DIFFERENCES

Population models that are well adapted to highly developed countries do not necessarily translate as well to developing countries, especially where the population distribution patterns might be different or underlying data are less available. AsiaPop and AfriPop, both predecessors to WorldPop, were shown to be more accurate in the regions for which they were created. For example, a test of AsiaPop in Cambodia and Vietnam found that it provided a better overall fit and less variability in errors against reference data than did GPWv3 or GRUMPv1 (Gaughan et al., 2013). Likewise, AfriPop was in close agreement with UNDP reference data in Somalia, whereas LandScan 2008 tended to overestimate and GRUMP beta 2000 tended to underestimate in most districts, with differences especially noticeable around urban areas (Linard et al., 2010). Bai et al. (2018) tested GPWv3, GRUMPv1, WorldPop, and a China-specific population grid (CnPop) in China, calculating the errors against GIS-linked, township-level census data from the year 2000. They found WorldPop to be the most accurate out of the four. The areal weighting method, as used by GPWv3 and GRUMP, has estimation issues in the Chinese context, whereas land cover methods, like WorldPop and CnPop, appeared to be more accurate; it is worth noting, though, that GPWv3 is only available at a coarser resolution than the most current product, GPWv4.

A comparison by Engstrom et al. (2019) in Sri Lanka between several population grids and reference data from the 2011 census at the Gram Niladhari (GN) level found that LandScan was the least accurate, while Facebook's HRSL followed by GHS-POP were the most accurate (R-squares of .84 and .67, respectively). GPWv4 preformed reasonably well with an R-square of 0.4. At coarser administrative levels, agreement increased across the board. It is hypothesized that LandScan's ambient modeling approach reduced agreement with the household-based census results, though the creators also acknowledge that the underlying population data may have limitations (Bhaduri, personal communication, 2018).

THE STRENGTH OF HIGHER RESOLUTION DATA

Additional validation work demonstrates that higher-resolution input data can provide more reliable estimates. The developers of HRSL tested the accuracy of their calculations against both a human-referenced dataset and household survey data in 18 countries across Asia, Latin America, and Africa, concluding that errors seem to be minimal (Tiecke et al., 2017). Indeed, HRSL often finds buildings that other models miss, and it is particularly superior in rural areas. Engstrom et al. (2019) piloted a bottom-up method in Sri Lanka that combined household survey data with satellite data, and they tested their new method alongside existing methods against census data. They found the bottom-up estimates were more accurate than the top-down estimations from GPWv4 (2010 and 2015), GHSL 2014, LandScan 2010, and WorldPop 2010, although HRSL 2015 and WorldPop 2015 were still more accurate, as they drew on the most recent census data and incorporated higher resolution satellite imagery. Novel and geographically specific approaches can generate improved estimates as well. For example, a method piloted in China found that considering "points of interest" data improved accuracy over WorldPop (Ye et al., 2019). In a study of Pakistan, the US Bureau of the Census found that its new method was closer to census values than LandScan, GPW, or GRUMP, and it excelled at identifying large agriculture population clusters (Azar et al., 2013). Researchers should continue to explore new methods and data sources, and users should consider alternative products that may be better suited to local conditions.

ERRORS CHANGE WITH LEVELS OF GEOGRAPHY

Simply because a population grid might provide estimates at a high resolution does not mean that the estimations are more accurate at this level, as uncertainty and errors often increase as one zooms in. For instance, when validating different population grids against census data in Kenya, Hay et al. (2004) found that the models were more accurate at the provincial level, but predictive skill decreased with lower administrative levels. Yet as would be expected, they observed a sharp transition in predictive accuracy when estimating below the level of census data used to generate a given model. Indeed, the resolution of census data can often be the limiting factor in population estimation. Other work has noted that the use of ancillary data only improves accuracy if it is more detailed than the input census data (Linard et al., 2012). Yet the relationship between resolution and accuracy can be more dynamic, as shown by validation work with a newly proposed estimation method in the Netherlands (Biljecki et al., 2016). Researchers found only a 0.5 percent predictive error at the national level, yet the error increased to 9.3 percent at the largest administrative level and then dropped to 3 percent at the smallest administrative level. These results underscore the need for continuing to test population data products at different geographic resolutions. Data users might naturally want the finest resolution population data available, but they should consider their options carefully to determine if this actually addresses their needs.

3.3 ALTERNATIVE VALIDATION DATA

Outside of country-produced census data, there are other sources of data that can be used for validating population grids. HRSL was validated in part by using data from Malawi's Integrated Household Survey (IHS3) (Tiecke et al., 2017). Likewise, researchers used geospatial household survey data in Bo City, Sierra Leone to validate a new bottom-up method of population estimations (Hillson, 2014). Suggesting possibilities for population grid validation, a micro-census conducted in Northern Nigeria was also validated using reference data collected during a Polio vaccination campaign (Weber et al., 2018). Although the use of alternative data might not be able to validate an entire population grid, the technique can provide greater overall statistical confidence, and new approaches and data sources open possibilities for a wider validation effort.

3.4 THE NEED FOR ADDITIONAL VALIDATION WORK

The growing body of model comparisons and validation research has begun to document the characteristics of the different population grids and test the usefulness of their estimations. It is not yet enough, though, for us to fully understand the nuances of these products, or how their performance might vary with country and context, as well as which products are most appropriate for specific policy purposes. Past research is still too narrow and inadequate (Bai et al., 2018). There is a definite need for a more systematic comparison, one that takes a global perspective rather than considering just individual countries. However, a major limitation continues to be a lack of data for validation. Experts have proposed this could be done with micro-census data or georeferenced household survey data, similar to the above-mentioned work by Tiecke et al. (2017) and Webber et al. (2018). Such a validation approach would not require that every cell in the gridded datasets be compared, and it could draw on existing data sources (Chen, 2019). The POPGRID Data Collaborative is researching key analytic measures to be used in validating its different gridded data products to better determine the accuracy of current methods and identify how they can be improved.

Data users can still take steps to take uncertainties into account. As described, there can be significant differences between population grids, and users can work to characterize uncertainty by using multiple grids in their work, rather than simply accepting one data product as fact. Additionally, users should be mindful of the geographic characteristics where they apply a population grid, as there may be existing validation work that can help justify the selection of one product over others for that region or category of land use. Moreover, as outlined in the next chapter, there are a number of relevant questions that users should keep in mind when deciding on a model. More informed use of population grids will strengthen the quality of analysis and decision making.

CHAPTER 4

DECISION-MAKING GUIDELINES

As discussed in previous chapters, currently available gridded population datasets are each distinct, and navigating their strengths and limitations requires some background knowledge. Our interviews with data producers and users have also highlighted a gap in expectations. On the one hand, data producers were concerned with the lack of knowledge and capacity among policymakers to extract relevant information from gridded population datasets. On the other hand, data users reported that in critical times, the urgent need to extract a number for decision making takes priority over a comprehensive understanding of the datasets (Neumann et al., 2015). In some instances, logistical convenience, access issues, or political considerations determine the selection of a dataset. While we cannot fully overcome these barriers, a few key considerations will help guide the community as they manage the complexities of data selection.

A recent journal article took a "fitness for use" approach to provide general recommendations on the use of gridded population data (Leyk et al., 2019). Building on this more technical review, we outline nine key decision-making criteria to support a more accessible and informed selection process. These considerations can also help identify which datasets might not be appropriate for a given application and explain why others could provide more appropriate estimates.

1. WHERE PEOPLE ARE AT ANY GIVEN TIME MAKES A DIFFERENCE

Populations are mobile on different time scales (daily, weekly, and seasonal), and therefore many population products use different concepts for the temporal distribution of population. Most products are based on census data, which measure where people live (residential population). This is often interpreted as nighttime population, since the vast majority of people sleep at their primary residence at night and work during the day at another location. Daytime population considers where people are during the day, whereas ambient population (the temporal concept used by Landscan) is where people are likely to be throughout the course of a day (e.g., over 24 hours). The analysis of long-term population change generally is based on nighttime population, since this is what censuses measure and there are few reliable estimates of population during daytime (Leyk et al., 2019). Yet for applications, such as Emergency response, users may need information on the daytime population to better reflect where people are actually located should a disaster occur during the day (DeVille et al., 2014). GPWv4, GHS-POP, WorldPop and WPE estimate residential population, whereas LandScan estimates are an average of daytime and nighttime population count.[10] Users should first consider if their research question requires nighttime or daytime population figures and then review the corresponding options.

[10] The JRC developed the first continental population grids accounting for daily and monthly population variations by considering the presence of residents, workers per different sectors, students, and tourists, and locations of residence and activity (Schiavina et al., 2020). These grids were produced in the ENACT project for all EU-28 countries by combining official statistical data at the regional level with geospatial data from conventional and non-conventional data sources for the reference year of 2011 (most recent round of censuses).

TOC

2. BE AWARE OF THE MODELING APPROACH USED TO DISTRIBUTE POPULATION

It may be tempting to opt for the highest resolution gridded data product available, such as a product like the HRSL which is distributed at 3 arc-second (10 m) resolution. Yet each product uses different definitions of what constitute settlements, which are in turn constrained by the satellite sensors used to define them. HRSL uses a five-meter built-up area product based on a machine learning algorithm applied to Maxar imagery. In theory, the footprint of every building can be identified. GHS-POP, on the other hand, uses a 30 m product derived from the Landsat archive that defines the degree of built-up in terms of fractions. Both of these products use an allocation that strictly moves people to built-up areas, whereas WorldPop uses an algorithm that always leaves some residual population in areas not identified as built-up (on the assumption that satellite sensors don't always identify all populated areas). There are other factors that will determine fitness for use (e.g., availability of consistent time series, desire for residential vs. ambient population, etc.), so while opting for the model with the highest resolution might be tempting, a higher resolution does not guarantee greater accuracy for application. If the application is at the local level, then this suggests the use of more modeled products. Yet if the question concerns regional issues, such as climate change, then coarser resolution products might suffice (MacManus, 2019). As described in Chapter 3, predictive errors can increase as one zooms in. The resolution of input census data is a significant factor in this consideration (GPWv4 carefully documents this by country), and it is important to remember that the quality of the census can vary dramatically between countries. Finally, if population data need to be used in conjunction with other datasets, such as, elevation or slope data, then it is not appropriate to use modeled products that use these factors as an input to their models (see also #8 below).

3. DEMOGRAPHIC GROUPINGS, INCLUDING AGE AND SEX, ARE IMPORTANT

Certain demographic groups can be especially vulnerable to different socio-economic and environmental circumstances. For example, disasters often have greater socio-economic impacts on women, children, and the elderly and these groups face a higher risk of subsequent health issues, food insecurity, and poverty (McGlade et al., 2019). Population data disaggregated by age and sex can be used to assess if there are unusually high or low proportions of vulnerable groups in particular areas to help guide and improve interventions. At present, GPWv4 and WorldPop are the only two global gridded population datasets that provide gridded population estimates by sex and age. In addition, LandScan does include 5-year age-sex demographics at the level 1 administrative unit. In the case of GPWv4, only 20% of all countries have age and sex data for the country as a whole, including many small-island states. High resolution estimates by age and sex have recently been released for HRSL based on machine learning methods, but these should be considered experimental.

4. ANALYSIS ACROSS COUNTRIES IS NOT THAT SIMPLE

Population change at a global level is dependent on the number of people, annual births, and deaths. However, population change at a national level is determined by migration as well as births and deaths. To understand population dynamics across countries or regions, data users may want to use products that are adjusted based on a globally consistent set of population estimates. For example, GPWv4 has global population grids that are based on the original country census population totals and grids adjusted to the UN World Population Prospects country-by-country estimates, so that the sum of all grid cells totals to the UN estimates. It is important to note here that relying on the original census, migrant, refugee, and internally displaced populations will not always be represented. Similarly, LandScan normalizes subnational estimates to the US Census Bureau's International Programs country-level totals for a given year. For cross-country analyses, it is better to use the adjusted products.

5. IDENTIFY THE TIME PERIODS THAT ARE NEEDED

The rate of population growth has a substantial impact on food production, water management systems, land use, and climate policy. Rapid population growth, particularly in urban centers, requires informed policies and sufficient resources that will enhance the quality of life for people and protect the environment. Users generally need the most recent data that matches their application, and may want recent trends or future projections. For some gridded population datasets, estimates are provided annually, but this does not necessarily mean that these estimates can be used as a consistent time series at a fine resolution. For example, WPE and LandScan use different methods every year and are generally not appropriate for comparisons across time. Other datasets provide estimates at 5-year (or other) intervals that are specifically intended to be consistent over these periods. GPWv4, for example, uses the 2010 round of censuses as a baseline and includes a simple projection to 2020. If available, users should check the origin of census data in a country of interest to assess the quality of estimations.

6. BE MINDFUL OF THE DIFFERENCES BETWEEN URBAN AND RURAL ESTIMATES

Most of the existing global population datasets have focused on characterizing patterns and growth in urban settlements, since this is where the majority of the population lives, where census data are typically released at a finer resolution, and where infrastructure is easier to detect. For many rural areas, census administrative units are quite large in area, and rural villages and infrastructure may be harder to detect using remote sensing data due to vegetation, terrain, or building types and materials (Leyk et al., 2019). For users focused on urban population density or distribution, datasets with settlement or urban extent variables such as GHS-POP, LandScan, and WorldPop are likely to be more suitable for assessing urban population distribution. For efforts that require a more detailed understanding of rural population distribution, HRSL is an option, if it is available for the country or countries of interest because it utilizes high resolution remote sensing data that is better able to detect buildings in rural regions.

7. ENVIRONMENTAL FACTORS IMPACT POPULATION DISTRIBUTION

Users should be aware of the particular environmental setting where they are looking for a population estimate. Deserts, areas with steep slopes or heavy vegetation, and regions with high levels of cloudiness present particular challenges. This is due to limitations in the availability of usable remote sensing imagery or complications in detecting buildings in such imagery. Even the underlying census data may be less reliable in certain cases, especially for nomadic, subsistence, or otherwise hard-to-reach populations. In some cases, datasets assume that no one lives on water bodies or in protected areas, but this may lead to underestimation of populations that may be present during certain times of the year. GHS-POP does not depart from such assumptions, and in its latest release (R2019), it has critically revised areas in input census data declared as uninhabited (Freire et al., 2018).

8. CONSIDER COVARIATES

Users should consider which covariates are used in each dataset. Covariates are used to redistribute population data across administrative units more accurately. The correlation between a covariate and population density impacts population estimates. For instance, LandScan, WorldPop, and WPE use roads as a covariate in their modeling. If users are interested in understanding the relationship between roads and population density, the datasets that use roads in their population model may not be the best choice, since they are built off of a preexisting assumption about how populations cluster near roads. This is especially important to understand in the context of the SDGs, which measure access to services and facilities that have been used as a covariate to estimate the distributions of gridded population. It is important to note that the number of covariates used to produce population estimates varies between datasets and affects the level of modeling employed to produce estimates. WorldPop and LandScan use a higher number of covariates in their modeling, whereas, GHS-POP and HRSL use fewer covariates and are lightly modeled.

9. BE COGNIZANT OF THE COST AND REUSE RIGHTS

Public and nonprofit institutions do not always have the programmatic and resource commitments to financially support policymakers and other users' access to data. In some cases, users may need data for internal analysis and decision making and have the resources to purchase data that are not openly available. However, in many instances, users may want to redistribute data to partners or end users, and therefore reuse rights and costs can become an important factor. Several datasets are fully open and available, for both commercial and noncommercial use and reuse (typically labeled with the Creative Commons Attribution 4.0 International License). Others may restrict commercial use, even if a commercial entity is supporting a public good. It should be noted that the concept of "fair use" differs between countries, so users may also need to be aware of the applicable database and copyright laws in their own jurisdictions or where the data may be redistributed or used. For example, GPW4, GHS-POP, GRUMP, and WorldPop have made their data available to the public free of charge, and WPE is also available free for those with named user accounts in ArcGIS organizations. LandScan is only free and open access for humanitarian purposes, academic research, and U.S. federal government and their mission partners. It is licensed at a fee for commercial entities. However, the LandScan USA dataset is one of their resources that is free and open to the public.

The decision-making criteria presented in this section were established to assist users in selecting the most appropriate product for their intended use. However, the underlying guidelines for selection are not limited to the nine criteria mentioned above. We encourage users to also assess context-specific (socio-economic, environmental, etc.) factors pertaining to the application area when selecting a dataset. For example, in developing indicators for the SDGs, it may be appropriate for a country to try to use a single, common gridded population dataset as the basis for indicators that focus on access to services, in order to increase the consistency of indicators across goals and sectors.

BOX 5 | ADDRESSING MISCONCEPTIONS

TReNDS interviewed the POPGRID data providers presented in this report to better understand their perspective on key communication points to convey to users to foster a fit-for-purpose approach to gridded population data. The following points were raised by the providers to clarify misconceptions about population estimates and the datasets:

- **Gridded population data are a disaggregation or redistribution of census data in top-down models, not an alternative to census data.**
- **Many of the gridded population datasets use the same input census data, however, they are not homogeneous. Data providers employ different methods and assumptions which produce different outputs.**
- **Improved communication avenues between data providers and users will allow data providers to recommend appropriate datasets pertinent to the specific application area.**
- **Publishing comprehensive documentation about the dataset can foster transparency and ensure that users select the most appropriate dataset for their intended use.**
- **The POPGRID Viewer is an interactive tool that helps users visually compare different datasets and their outputs to identify and understand their differences.**

CONCLUSIONS

Confronting humanity's greatest challenges and achieving our collective goals will fundamentally depend on a detailed understanding of where people are located. Gridded population data are an important tool for improving this understanding. In the preceding chapters, we have introduced this category of data and flagged a number of aspects users should keep in mind. Already, gridded population data are powering the WFP's disaster relief efforts, helping us respond to epidemics, producing new measures of development at the World Bank, and informing critical scientific research. Based on a survey of NSOs, though, many decision-makers that stand to benefit from this resource are still unaware. We hope that our overview of the available data products and their nuances will extend the conversation and inform new applications.

Although the seven datasets we have profiled are measuring similar things, they diverge in many important ways. Users should be aware of how a dataset has been constructed, specifically with different types of input data and ancillary data, the underlying modeling method, and the way the data was adjusted to match national population figures. The features of the resulting datasets are different as well, including the spatial resolution, the available type of population disaggregation, the years represented, and the consistency of methods between versions. Differences extend beyond just general characteristics. Past work has shown that there can be significant differences in population figures between datasets. The uncertainty in population models also influences consequent predictions, such as exposure to disease or vulnerability to climate change. Data users should be deliberate about model selection and transparent about the associated uncertainties.

Initial efforts to validate gridded population data have already provided valuable insights, and it appears that some methods can indeed offer more reliable estimates. Yet users should be aware of the particular biases of available products, as estimates may perform better or worse according to geography or settlement type. Validation studies also emphasize that just because a model provides estimates at a higher resolution, does not mean it is more accurate. The POPGRID community recognizes the need for additional validation research, which would allow testing of how datasets perform in a range of contexts and ensure that tools are being used appropriately. With a new global round of censuses now underway, there are unique opportunities to access contemporary population counts and perform detailed validation studies. The multiple sources of household survey data could provide insights as well. Validation work will require strategic planning and funding, and the international community should commit to a strategy for realizing the full potential of gridded population data.

While investment in gridded population data is certainly needed, it should not be done in exchange for improving the censuses. One of the most important steps in strengthening gridded population estimates is improving the accuracy, resolution, and completeness of input data. Quality census data are essential to many of the top-down datasets we have profiled. Although the emerging bottom-up gridded population data techniques may be valuable in areas where census data are missing or out of date, they are not a replacement. Indeed, underlying census data can be the key limiting factor when forming a top-down population estimation. This issue is particularly acute in low-income settings where high-resolution data are often not available, areas of conflict,[11] or rapidly changing areas. Individual countries and the UN system as a whole must ensure that censuses and other traditional data sources remain strong, even as novel data sources continue to emerge.

Regardless of eventual investment, there are questions already available that data users should run through when selecting a population model. Do they want to consider the population during the day or night? Are they only concerned with total population figures, or are they interested in demographic breakdowns? Do they want to look at population characteristics across countries or across time? Are they interested in populations in urban or rural areas? Their answers are likely to direct them to a logical selection for their particular needs. Additional validation work will help detail other nuances to consider when choosing and evaluating datasets. For instance, users might want to query how models perform in coastal regions, in forested areas, or in other places that are challenging to characterize. The POPGRID Data Collaborative is committed to exploring these issues and advancing the discussion around gridded population data, and as our understanding evolves, relevant factors could be presented in an interactive decision tree that would support users and provide tailored guidance. With continued research and growing awareness among users, population data can help to not only document but improve the quality of lives around the World.

[11] Efforts are underway to map the "missing millions" of refugee and displaced populations using satellite remote sensing, and the International Organisation for Migration's Displacement Tracking Matrix (IOM-DTM) is making headway in geolocating populations that have been displaced (Van Den Hoek, 2018).

TOC

REFERENCES

Abushady, A. (2020, October 11). Personal interview with H. Dahmm and M. Rabiee.

Andrade-Pacheco, R., Savory, D. J., Midekisa, A., Gething, P. W., Sturrock, H. J., & Bennett, A. (2019). Household Electricity Access in Africa (2000–2013): Closing Information Gaps with Model-based Geostatistics. PloS One, 14(5). doi:10.1371/journal.pone.0214635

Aubrecht, C., Gunasekera, R., Ungar, J., & Ishizawa, O. (2016). Consistent Yet Adaptive Global Geospatial Identification of Urban–Rural Patterns: The iURBAN model. Remote Sensing of Environment, 187, 230-240. doi:10.1016/j.rse.2016.10.031

Azar, D., Engstrom, R., Graesser, J., & Comenetz, J. (2013). Generation of Fine-Scale Population Layers Using Multi-resolution Satellite Imagery and Geospatial Data. Remote Sensing of Environment, 130, 219-232. doi:10.1016/j.rse.2012.11.022

Backer, L., & Holt Bloch, V. V. (2011). GEOSTAT 1A—Representing Census Data in a European Population Grid. Final report. Retrieved from http://www.efgs.info/wp-content/uploads/geostat/1a/GEOSTAT1A-final-report.pdf

Bai, Z., Wang, J., Wang, M., Gao, M., & Sun, J. (2018). Accuracy Assessment of Multi-Source Gridded Population Distribution Datasets in China. Sustainability, 10(5), 1363. doi:10.3390/su10051363

Baylor College of Medicine. (n.d.) Introduction to Infectious Diseases. Retrieved from https://www.bcm.edu/departments/molecular-virology-and-microbiology/emerging-infections-and-biodefense/introduction-to-infectious-diseases

Bhaduri, B. (2018). Personal communication with R. Chen and A. de Sherbinin.

Biljecki, F., Ohori, K. A., Ledoux, H., Peters, R., & Stoter, J. (2016). Population Estimation Using a 3D City Model: A Multi-scale Country-wide Study in the Netherlands. PloS One, 11(6). doi:10.1371/journal.pone.0156808

Bird, J. H., Lebrand, M. S. M., & Venables, A. J. (2019). The Belt and Road Initiative: Reshaping Economic Geography in Central Asia. The World Bank. Retrieved from http://documents.worldbank.org/curated/en/836301554729486900/The-Belt-and-Road-Initiative-Reshaping-Economic-Geography-in-Central-Asia

Blankespoor, B., Dasgupta, S., & Lange, G. (2016). Mangroves as a Protection from Storm Surges in a Changing Climate. Ambio, 46(4), 478-491. doi:10.1007/s13280-016-0838-x

Butler, D. (2011). Reactors, Residents and Risk. Nature, 474, 36. doi: 10.1038/472400a

Cairney, P. (2016). The Politics of Evidence-based Policy Making. Palgrave Springer.

Calka, B., & Bielecka, E. (2019). Reliability Analysis of LandScan Gridded Population Data. The Case Study of Poland. ISPRS International Journal of Geo-Information, 8(5), 222. doi:10.3390/ijgi8050222

Canavire-Bacarreza, G., Martinez-Vazquez, J., & Yedgenov, B. (2020). Identifying and Disentangling the Impact of Fiscal Decentralization on Economic Growth. World Development, 127, 104742. doi:10.1016/j.worlddev.2019.104742

Chen, R. (2019 November 19). Personal interview with H. Dahmm and M. Rabiee.

CIESIN & CUNY CIDR. (2020 forthcoming). Low Elevation Coastal Zone (LECZ) Urban-Rural Population and Land Area Estimates, Version 3. NASA Socioeconomic Data and Applications Center (SEDAC) [Dataset].

Cori, A., Donnelly, C. A., Dorigatti, I., Ferguson, N. M., Fraser, C., Garske, T., & Van Kerkhove, M. D. (2017). Key Data for Outbreak Evaluation: Building on the Ebola Experience. Philosophical Transactions of the Royal Society B: Biological Sciences, 372 (1721), 20160371. doi:10.1098/rstb.2016.0371

Cui, R., Hultman, N., Jiang, K., McJeon, H., Yu, S., Cui, D., & Clarke, L. (2020). A High Ambition Coal Phaseout in China: Feasible Strategies Through a Comprehensive Plant-by-Plant Assessment. Center for Global Sustainability. Retrieved from https://cgs.umd.edu/sites/default/files/2020-01/AHighAmbitionCoalPhaseoutInChina_EN_Full.pdf

Da Costa, J. N., Bielecka, E., & Calka, B. (2017). Uncertainty Quantification of the Global Rural-Urban Mapping Project over Polish Census Data. Proceedings of 10th International Conference "Environmental Engineering" (Vol. 10, pp. 1-7). doi:10.3846/enviro.2017.221

Dahmm, H. (2018). Valuation of the New Zealand Census. SDSN TReNDS. https://www.sdsntrends.org/research/2018/11/16/case-study-new-zealand-census

Dasgupta, S., Laplante, B., Murray, S., & Wheeler, D. (2011). Exposure of Developing Countries to Sea-Level Rise and Storm Surges. Climatic Change, 106(4), 567-579. doi:10.1007/s10584-010-9959-6

Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F. R., Gaughan, A. E., & Tatem, A. J. (2014). Dynamic Population Mapping Using Mobile Phone Data. Proceedings of the National Academy of Sciences, 111(45), 15888-15893. doi:10.1073/pnas.1408439111

Doocy, S., Gorokhovich, Y., Burnham, G., Balk, D., & Robinson, C. (2007). Tsunami Mortality Estimates and Vulnerability Mapping in Aceh, Indonesia. American Journal of Public Health, 97(Supplement_1), S146-S151. doi:10.2105/AJPH.2006.095240

Doxsey-Whitfield, E., MacManus, K., Adamo, S. B., Pistolesi, L., Squires, J., Borkovska, O., & Baptista, S. R. (2015). Taking Advantage of the Improved Availability of Census Data: A First Look at the Gridded Population of the World, Version 4. Papers in Applied Geography, 1(3), 226-234. doi:10.1080/23754931.2015.1014272

Dwyer-Lindgren, L., Cork, M. A., Sligar, A., Steuben, K. M., Wilson, K. F., Provost, N. R., & Biehl, M. H. (2019). Mapping HIV Prevalence in Sub-Saharan Africa Between 2000 and 2017. Nature, 570(7760), 189-193. doi:10.1038/s41586-019-1200-9

Eales, A., Alsop, A., Frame, D., Strachan, S., & Galloway, S. (2020). Assessing the Market for Solar Photovoltaic (PV) Microgrids in Malawi. Hapres Journal of Sustainability Research, 2(1). doi:10.20900/jsr20200008

Ehrlich, D., Melchiorri, M., Florczyk, A. J., Pesaresi, M., Kemper, T., Corbane, C., & Siragusa, A. (2018). Remote Sensing Derived Built-up Area and Population Density to Quantify Global Exposure to Five Natural Hazards Over Time. Remote Sensing, 10(9), 1378. doi:10.3390/rs10091378

Eicher, C. L., & Brewer, C. A. (2001). Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation. Cartography and Geographic Information Science, 28(2), 125-138. doi:10.1559/152304001782173727

Engstrom, R., Newhouse, D. L., & Soundararajan, V. (2019). Estimating Small Area Population Density Using Survey Data and Satellite Imagery: An Application to Sri Lanka. The World Bank. Retrieved from http://documents.worldbank.org/curated/en/256241552483977593/Estimating-Small-Area-Population-Density-Using-Survey-Data-and-Satellite-Imagery-An-Application-to-Sri-Lanka

Esri. (n.d.) Measuring in Arc-Seconds. Retrieved from https://www.esri.com/news/arcuser/0400/wdside.html Accessed 16 April 2020

European Commission Joint Research Centre. (2018). Atlas of the Human Planet 2018. EU Science Hub. Retrieved from https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/atlas-human-planet-2018

Freire, S., Schiavina, M., Florczyk, A. J., MacManus, K., Pesaresi, M., Corbane, C., & Sliuzas, R. (2018). Enhanced Data and Methods for Improving Open and Free Global Population Grids: Putting 'Leaving No One Behind' Into Practice. International Journal of Digital Earth, 13(1), 61-77. doi:10.1080/17538947.2018.1548656

Frye, C. (2017). Esri's 2016 World Population Estimate Methodology. Esri. Retrieved from https://www.esri.com/content/dam/esrisites/en-us/media/whitepaper/Esri_2016_WPE_Methodology_WhitePaper.pdf

Garb, J. L., Cromley, R. G., & Wait, R. B. (2007). Estimating Populations At Risk for Disaster Preparedness and Response. Journal of Homeland Security and Emergency Management, 4(1). doi:10.2202/1547-7355.1280

Gao, J. (2017). Downscaling Global Spatial Population Projections From 1/8-Degree to 1-km Grid Cells. Technical Notes NCAR, National Center for Atmospheric Researcher, Boulder, CO., USA. Retrieved from https://opensky.ucar.edu/islandora/object/technotes%3A553/datastream/PDF/view

Gaughan, A. E., Stevens, F. R., Linard, C., Jia, P., & Tatem, A. J. (2013). High Resolution Population Distribution Maps for Southeast Asia in 2010 and 2015. PloS One, 8(2). doi:10.1371/journal.pone.0055882

Gaughan, A. (2019, September 24). Personal interview with H. Dahmm and M. Rabiee

GPSDD Country Partners. (2019). Personal Interviews with Alonso Ortiz Galan and Charu Vijayakumar.

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. doi:10.1016/j.earscirev.2015.08.012

Hall, O., Stroh, E., & Paya, F. (2012). From Census to Grids: Comparing Gridded Population of the World with Swedish Census Records. The Open Geography Journal, 5(1). Retrieved from https://pdfs.semanticscholar.org/aa54/496eaa46a532ba3c4353843c4a56b7712db3.pdf

Hallisey, E., Tai, E., Berens, A., Wilt, G. Peipins, L., Lewis, B., Graham, S., Flanagan, B., Lunsford, N. (2017). Transforming Geographic Scale: A Comparison of Combined Population and Areal Weighting to Other Interpolation Methods. International Journal of Health Geographics, 16:29. doi: 10.1186/s12942-0170102-z

Hay, S. I., Guerra, C. A., Tatem, A. J., Noor, A. M., & Snow, R. W. (2004). The Global Distribution and Population at Risk of Malaria: Past, Present, and Future. The Lancet Infectious Diseases, 4(6), 327-336. doi:10.1016/S1473-3099(04)01043-6

Hay, S. I., Noor, A. M., Nelson, A., & Tatem, A. J. (2005). The Accuracy of Human Population Maps for Public Health Application. Tropical Medicine and International Health, 10(10), 1073-1086. doi:10.1111/j.1365-3156.2005.01487.x

Hillson, R., Alejandre, J. D., Jacobsen, K. H., Ansumana, R., Bockarie, A. S., Bangura, U., ... & Stenger, D. A. (2014). Methods for Determining the Uncertainty of Population Estimates Derived from Satellite Imagery and Limited Survey Data: A Case Study of Bo City, Sierra Leone. PloS one, 9(11). doi:10.1371/journal.pone.0112241

IDMC. (2019). Global Report on Internal Displacement 2019. Internal Displacement Monitoring Centre. Retrieved from https://www.internal-displacement.org/global-report/grid2019/

IEAG. (2014). A World That Counts: Mobilising the Data Revolution for Sustainable Development. United Nations Secretary-General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development (IEAG). Retrieved from https://www.undatarevolution.org/wp-content/uploads/2014/12/A-World-That-Counts2.pdf

Iimi, A., Ahmed, F., Anderson, E. C., Diehl, A. S., Maiyo, L., Peralta-Quirós, T., & Rao, K. S. (2016). New Rural Access Index: Main Determinants and Correlation to Poverty. The World Bank. Retrieved from https://openknowledge.worldbank.org/handle/10986/25676

James, W. H., Tejedor-Garavito, N., Hanspal, S. E., Campbell-Sutton, A., Hornby, G. M., Pezzulo, C., &

Kerr, D. (2018). Gridded Birth and Pregnancy Datasets for Africa, Latin America and the Caribbean. Nature, 5, 180090. doi:10.1038/sdata.2018.90

Johnson, A. M. (2019. Liberia: US$18.9 Million Census Set for March 2021. FrontPageAfrica. Retrieved from https://frontpageafricaonline.com/news/liberia-us18-9-million-census-set-for-march-2021/

Jones, B., & O'Neill, B. C. (2020). Global One-Eighth Degree Population Base Year and Projection Grids Based on the Shared Socioeconomic Pathways, Revision 01. NASA Socioeconomic Data and Applications Center (SEDAC) [Dataset]. doi:10.7927/m30p-j498. Accessed 16 April 2020

Kiguchi, M., Shen, Y., Kanae, S., & Oki, T. (2014). Re-evaluation of Future Water Stress Due to Socio-Economic and Climate Factors Under a Warming Climate. Hydrological Sciences Journal, 60(1), 14-29. doi:10.1080/02626667.2014.888067

Kulp, S. A., & Strauss, B. H. (2019). New Elevation Data Triple Estimates of Global Vulnerability to Sea-Level Rise and Coastal Flooding. Nature Communications, 10(1), 1-12. doi:10.1038/s41467-019-12808-z

Lai, S., Bogoch, I. I., Ruktanonchai, N., Watts, A. G., Li, Y., Yu, J., & Li, Z. (2020a). Assessing Spread Risk of Wuhan Novel Coronavirus Within and Beyond China, January-April 2020: A Travel Network-Based Modelling Study. Health Sciences. doi: 10.1101/2020.02.04.20020479

Lai, S., Ruktanonchai, N. W., Zhou, L., Prosper, O., Luo, W., Floyd, J. R., & Tatem, A. J. (2020b). Effect of Non-Pharmaceutical Interventions for Containing the COVID-19 Outbreak: An Observational and Modelling Study. MedRxiv. doi:10.1101/2020.03.03.20029843

Leyk, S., Gaughan, A. E., Adamo, S. B., de Sherbinin, A., 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). doi:10.5194/essd-11-1385-2019

Linard, C., Alegana, V. A., Noor, A. M., Snow, R. W., & Tatem, A. J. (2010). A High Resolution Spatial Population Database of Somalia for Disease Risk Mapping. International Journal of Health Geographics, 9(1), 45. doi:10.1186/1476-072X-9-45

Linard, C., & Tatem, A. J. (2012). Large-scale Spatial Population Databases in Infectious Disease Research. International Journal of Health Geographics, 11(1), 7. doi:10.1186/1476-072X-11-7

Lloyd, C. T., Sorichetta, A., & Tatem, A. J. (2017). High Resolution Global Gridded Data for Use in Population Studies. Scientific Data, 4(1), 1-17. doi:10.1038/sdata.2017.1

Lloyd, C. T., Chamberlain, H., Kerr, D., Yetman, G., Pistolesi, L., Stevens, F. R., & Sinha, P. (2019). Global Spatio-Temporally Harmonised Datasets for Producing High-resolution Gridded Population Distribution Datasets. Big Earth Data, 3(2), 108-139. doi:10.1080/20964471.2019.1625151

Local Burden of Disease Child Growth Failure Collaborators. (2020). Mapping Child Growth Failure Across Low-and Middle-Income Countries. Nature, 577(7789), 231. doi:10.1038/s41586-019-1878-8

MacManus, K. (2019 October 9). Personal interview with H. Dahmm and M. Rabiee.

McGlade, J., Bankoff, G., Abrahams, J., Cooper-Knock, S. J., Cotecchia, F., Desanker, P., & Hirsch, F. (2019). Global Assessment Report on Disaster Risk Reduction 2019. Retrieved from https://www.undrr.org/publication/global-assessment-report-disaster-risk-reduction-2019

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 International Journal of Geo-Information, 8(2), 96. doi:10.3390/ijgi8020096

Mondal, P., & Tatem, A. J. (2012). Uncertainties in Measuring Populations Potentially Impacted by Sea Level Rise and Coastal Flooding. PloS One, 7(10). doi:10.1371/journal.pone.0048191

Mosser, J. F., Gagne-Maynard, W., Rao, P. C., Osgood-Zimmerman, A., Fullman, N., Graetz, N., & Earl, L. (2019). Mapping Diphtheria-Pertussis-Tetanus Vaccine Coverage in Africa, 2000–2016: A Spatial and Temporal Modelling Study. The Lancet, 393(10183), 1843-1855. doi:10.1016/S0140-6736(19)30226-0

Nansai, K., Tohno, S., Chatani, S., Kanemoto, K., Kurogi, M., Fujii, Y., & Lenzen, M. (2020). Affluent Countries Inflict Inequitable Mortality and Economic Loss on Asia via PM2. 5 Emissions. Environment International, 134, 105238. doi:10.1016/j.envint.2019.105238

Neumann, B., Vafeidis, A. T., Zimmermann, J., & Nicholls, R. J. (2015). Future Coastal Population Growth and Exposure to Sea-level rise and Coastal Flooding - A Global Assessment. PloS One, 10(3), e0118571. doi:10.1371/journal.pone.0118571

Parish, E. S., Kodra, E., Steinhaeuser, K., & Ganguly, A. R. (2012). Estimating Future Global Per Capita Water Availability Based on Changes in Climate and Population. Computers and Geosciences, 42, 79-86. doi:10.1016/j.cageo.2012.01.019

Rainisch, G., Shankar, M., Wellman, M., Merlin, T., & Meltzer, M. I. (2014). Regional Spread of Ebola Virus, West Africa, 2014. Emerging Infectious Diseases, 21(3), 444. doi.org/10.3201/eid2103.141845

Reed, F. J., Gaughan, A. E., Stevens, F. R., Yetman, G., Sorichetta, A., & Tatem, A. J. (2018). Gridded Population Maps Informed by Different Built Settlement Products. Data, 3(3), 33. doi:10.3390/data3030033

Rose, A. (2019 November 14). Personal Interview with H. Dahmm and M. Rabiee.

Roseth, B., Reyes, A. Yee Amezaga, K. (2019). The Value of Official Statistics: Lessons from Intergovernmental Transfers. Inter-American Development Bank (IDB). doi:10.18235/0001883

Sabesan, A., Abercrombie, K., Ganguly, A. R., Bhaduri, B., Bright, E. A., & Coleman, P. R. (2007). Metrics for the Comparative Analysis of Geospatial Datasets with Applications to High-resolution Grid-based Population Data. GeoJournal, 69(1-2), 81-91. doi:10.1007/s10708-007-9103-y

Schiavina, M., Freire, S., Rosina, K., Ziemba, L., Marin Herrera, M., Craglia, M., Lavalle, C., Kemper, T., Batista, F. (2020). ENACT-POP R2020A - ENACT 2011 Population Grid. European Commission, Joint Research Centre (JRC) [Dataset]. doi:10.2905/BE02937C-5A08-4732-A24A-03E0A48BDCDA. Accessed 15 April 2020.

Schinko, T., Drouet, L., Vrontisi, Z., Hof, A., Hinkel, J., Mochizuki, J., & Lincke, D. (2020). Economy-Wide Effects of Coastal Flooding Due to Sea level rise: A Multi-model Simultaneous Treatment of Mitigation, Adaptation, and Residual impacts. Environmental Research Communications, 2(1), 015002. doi: 10.1088/2515-7620/ab6368

Schneider, A., Friedl, M. A., & Potere, D. (2010). Mapping Global Urban Areas Using MODIS 500-m Data: New Methods and Datasets Based on 'Urban Ecoregions'. Remote Sensing of Environment, 114(8), 1733-1746. doi:10.1016/j.rse.2010.03.003

SEDAC. (n.d.). SEDAC Global COVID-19 Viewer: Population Estimates by Age Group and Sex. Socioeconomic Data and Applications Center (SEDAC). Retrieved from https://sedac.ciesin.columbia.edu/mapping/popest/covid-19/ Accessed 16 April 2020

Shaddick, G., Thomas, M. L., Green, A., Brauer, M., van Donkelaar, A., Burnett, R., ... & Gumy, S. (2017). Data Integration Model for Air Quality: A Hierarchical Approach to the Global Estimation of Exposures to Ambient Air Pollution. Journal of the Royal Statistical Society: Series C (Applied Statistics), 67(1), 231-253. doi:10.1111/rssc.12227

Smith, A., Bates, P. D., Wing, O., Sampson, C., Quinn, N., & Neal, J. (2019). New Estimates of Flood Exposure in Developing Countries Using High-Resolution Population Data. Nature Communications, 10(1), 1-7. doi:10.1038/s41467-019-09282-y

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). doi:10.1371/journal.pone.0107042

Tatem, A. (2019 September 4). Personal interview with H. Dahmm and M. Rabiee.

Thomson, D. (2019 November 12). Personal interview with H. Dahmm and M. Rabiee.

Thomson, D. R., Linard, C., Vanhuysse, S., Steele, J. E., Shimoni, M., Siri, J., ... & Georganos, S. (2019). Extending Data for Urban Health Decision-Making: A Menu of New and Potential Neighborhood-level Health Determinants Datasets in LMICs. Journal of Urban Health, 96(4), 514-536. doi:10.1007/s11524-019-00363-3

Thomson, D. R., Stevens, F. R., Ruktanonchai, N. W., Tatem, A. J., & Castro, M. C. (2017). GridSample: An R Package to Generate Household Survey Primary Sampling Units (PSUs) From Gridded Population Data. International Journal of Health Geographics, 16(1), 25. doi:10.1186/s12942-017-0098-4

Tiecke, T.G., Liu, X., Zhang, A., Gros, A., Li, N., Yetman, G., Kilic, T., Murray, S., Blankespoor, B., Prydz, E.B., & Dang, H.H. (2017). Mapping the world population one building at a time. ArXiv, abs/1712.05839. Retrieved from https://www.semanticscholar.org/paper/Mapping-the-world-population-one-building-at-a-time-Tiecke-Liu/b17d3ded924a92b31cce96d11c345a9b82367457

Tuholske, C., Caylor, K., Evans, T., & Avery, R. (2019). Variability in Urban Population Distributions across Africa. Environ. Res. Lett. 14 085009. doi:10.1088/1748-9326/ab2432

Tusting, L. S., Bisanzio, D., Alabaster, G., Cameron, E., Cibulskis, R., Davies, M., & Okumu, F. O. (2019). Mapping Changes in Housing in Sub-Saharan Africa from 2000 to 2015. Nature, 568(7752), 391-394. doi:10.1038/s41586-019-1050-5

United Nations. (2016). The Sustainable Development Goals Report 2016. United Nations. Retrieved From https://unstats.un.org/sdgs/report/2016/The%20Sustainable%20Development%20Goals%20Report%202016.pdf

United Nations. (2018). The Sustainable Development Goals Report 2018. United Nations Department of Economic and Social Affairs. Retrieved from https://www.un.org/development/desa/publications/thesustainable-development-goals-report-2018.html

United Nations. (n.d.). Population. Retrieved from https://www.un.org/en/sections/issues-depth/population/index.html Accessed 16 April 2020

UNFPA. (n.d.). Impact of COVID-19 on Census. United Nations Population Fund (UNFPA). Retrieved from https://arcgis-pilot.unfpa.org/open_data/census_cal/census_delay_list.php?pagesize=-1 Accessed 16 April 2020

UNFPA (2017). New Methodology: A Hybrid Census to Generate Spatially-Disaggregated Population Estimates. United Nations Population Fund (UNFPA). Retrieved from https://www.unfpa.org/resources/new-methodology-hybrid-census-generate-spatially-disaggregated-population-estimates

UN IAEG-SDG. (2019). Tier Classification for Global SDG Indicators 11 December 2019. United Nations Inter-agency and Expert Group on Sustainable Development Goal Indicators (UN IAEG-SDG). Retrieved from https://unstats.un.org/sdgs/files/Tier-Classification-of-SDG-Indicators-11-December-2019-web.pdf

UN-GGIM Europe. (2019). The Territorial Dimension in SDG Indicators: Geospatial Analysis and its Integration with Statistical Data. United Nations Committee of Experts on Global Geospatial Information Management (UN-GGIM). Retrieved from https://un-ggim-europe.org/wp-content/uploads/2019/05/UN_GGIM_08_05_2019-The-territorial-dimension-in-SDG-indicators-Final.pdf

UNPD. (2017). World Population Prospects The 2017 Revision: Methodology of the United Nations Population Estimates and Projections. United Nations Population Division (UNPD). Retrieved from https://population.un.org/wpp/Publications/Files/WPP2017_Methodology.pdf

UNPD. (2019). World Population Prospects Highlights. United Nations Population Division (UNPD). Retrieved from https://population.un.org/wpp/Publications/Files/WPP2019_Highlights.pdf

UNSC. (2020). A Recommendation on the Methodology to Delineate Cities Urban and Rural Areas. Items for Consideration and Decision: Demographic Statistics. United Nations Statistical Commission (UNSC). Retrieved from https://unstats.un.org/unsd/statcom/51st-session/documents/BG-Item3j-Recommendation-E.pdf

UNSD. (n.d.). Demographic and Social Statistics. United Nations Statistics Division (UNSD). Retrieved from http://mdgs.un.org/unsd/demographic/sources/census/wphc/censusclockmore.htm Accessed 16 April 2020

UNSD. (2017a). Indicator 11.6.2: Annual Mean Levels of Fine Particulate Matter (e.g. PM2.5 and PM10) in Cities (Population Weighted). United Nations Statistics Division (UNSD). Retrieved from https://unstats.un.org/sdgs/metadata/files/Metadata-11-06-02.pdf

UNSD. (2017b). United Nations Statistics Division. Principles and Recommendations for Population and Housing Censuses. United Nations Statistics Division (UNSD). Retrieved from https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Principles_and_Recommendations/Population-and-Housing-Censuses/Series_M67rev3-E.pdf

Van Den Hoek, J. (2018). Mapping the Missing Millions: A Global Analysis of Informal Settlements and the Geography of Exclusion. In AGU Fall Meeting Abstracts. Retrieved from https://ui.adsabs.harvard.edu/abs/2018AGUFMPA43E1387V/abstract

Vincent, S. (2018). Status Review of the Updated Rural Access Index (RAI). ReCAP for DFID. Retrieved from http://research4cap.org/Library/Vincent-CDS-2018-StatusReviewUpdatedRAI-FinalReport_GEN2033C-180529.pdf

Wardrop, N. A., Jochem, W. C., Bird, T. J., Chamberlain, H. R., Clarke, D., Kerr, D., & Tatem, A. J. (2018). Spatially Disaggregated Population Estimates in the Absence of National Population and Housing Census Data. Proceedings of the National Academy of Sciences, 115(14), 3529-3537. doi:10.1073/pnas.1715305115

Weber, E. M., Seaman, V. Y., Stewart, R. N., Bird, T. J., Tatem, A. J., McKee, J. J., & Reith, A. E. (2018). Census-independent Population Mapping in Northern Nigeria. Remote Sensing of Environment, 204, 786-798. doi:10.1016/j.rse.2017.09.024

WHO. (n.d.) Natural Events. World Health Organization (WHO). Retrieved from https://www.who.int/environmental_health_emergencies/natural_events/en/ Accessed 16 April 2020

WHO. (2019). World Malaria Report 2019. World Health Organization (WHO). Retrieved from https://www.who.int/publications-detail/world-malaria-report-2019

World Bank. (2016). SDG Indicator 9.1.1. Third Meeting of the IAEG-SDG, Mexico City. Retrieved from https://unstats.un.org/sdgs/files/meetings/iaeg-sdgs-meeting-03/3rd-IAEG-SDGs-presentation-WorldBank--9.1.1.pdf

Ye, T., Zhao, N., Yang, X., Ouyang, Z., Liu, X., Chen, Q., & Jia, P. (2019). Improved Population Mapping for China Using Remotely Sensed and Points-of-Interest Data Within a Random Forests model. Science of the Total Environment, 658, 936-946. doi:10.1016/j.scitotenv.2018.12.276

APPENDIX A

1. When was your country's last census conducted and how often do you conduct them?

2. What are some of the data challenges associated with the census in your country?

3. Have you considered using alternative population sources derived from EO or telecoms data that would complement traditional population sources? What sources have you looked at and where do you go to find information on them?

4. Are there instances where your government or actors in your country use gridded (spatially disaggregated (raster) population estimations derived from a variety of primary sources such as previous census data, satellite data, etc.) population data to produce statistics or inform decisions? If so, could you please provide examples?

5. What are the advantages of using gridded population data compared to admin data, sample surveys, etc.? Could you share your experience working with different data sources?

6. What gridded population sources do you use (i.e. GPW, LandScan, etc.)? Are different sources used for different applications or geographies? Are there examples you could provide?

7. Are there specific methodological, policy, or availability reasons for selecting certain gridded population sources? Were multiple data sources compared before a selection was made?

8. What are some of the gaps in the availability and accessibility of gridded population sources? What additional resources and tools do you require to fill those gaps?

POPGRID
DATA COLLABORATIVE

TRENDS
Thematic Research Network
on Data and Statistics

SUSTAINABLE DEVELOPMENT
SOLUTIONS NETWORK
A GLOBAL INITIATIVE FOR THE UNITED NATIONS