# Alternative Approaches

## Introduction

The 2030 round of censuses is likely to be one of the most challenging with which to provide timely, robust and comparable statistics. Census response rates continue to decline across many countries and regions resulting in significant undercounts. Combined with the legacy of the COVID-19 pandemic which saw many countries postpone their national census, issues with recruiting field staff, and international funding for population data collection and analysis disappearing, these factors have combined to create a complex and challenging environment within which to design, implement and disseminate a census. According to the UN Economic Commission for Africa (UNECA), undercounts and census delays due to COVID-19, conflict or financial limitations have resulted in an estimated one in three Africans not being counted during the 2020 round of censuses.

Despite this challenging environment, the availability of new data sources and technologies offer National Statistics Offices (NSOs) alternative solutions to the traditional census approach. Digital data collection for the census through tablets, smartphones and other online resources is already making census processes cheaper, more efficient and improve census quality. Recent advances in satellite imagery and AI technology have further enhanced capacities in the use of alternative methodologies to complement census data collection efforts. For example, buildings mapped from satellite imagery using AI, together with counts of people obtained at small area unit level, can help to create detailed population estimates that can either support more effective and efficient census planning and implementation, or provide alternative population data in hard to count areas where a census cannot be fully enumerated.

These statistical modelling approaches have been developed to estimate populations based on the correlation between population density and geospatial covariate data. In countries where a census is not possible, these methods can be implemented at low cost (although they require a minimum level of technical capacity) and may provide alternative ways of deriving more recent population estimates to inform census planning and implementation.

Combining geospatial and statistical data will allow NSOs to produce very-high resolution gridded population data (typically 100m or 1km grid cells) with full national coverage. A significant advantage of this approach over more traditional census dissemination is that is that it provides the flexibility to aggregate grid cells to estimate the total population within any defined boundary, but with no additional disclosure risk. There are two broad methodologies that can be applied to produce high-resolution gridded population estimates:

### Top-down Method

The [top-down method](https://wpgp.github.io/bookworm/gridded-population-estimates.html#ref-stevens2015disaggregating) disaggregates known population totals for each coarser administrative unit (e.g. states or local government areas) into finer 100 m gridded population estimates. Such reliable population totals may be obtained from national population and housing census results, thus, the top-down population estimation approach is handy where recent census totals are available. Within this method, gridded population estimates are produced using Machine Learning algorithms (e.g., random forest models) to disaggregate population totals based on relationships between population density and a set of geospatial covariates such as building density, distance to city centre, or nighttime lights intensity. The disaggregation can be applied across the entire country (unconstrained) or limited only to areas where settlements have been mapped (constrained).

The top-down population disaggregation approach allows for both [classical](https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0107042) and [Bayesian](https://www.sciencedirect.com/science/article/pii/S0143622824002212) statistical inference approaches, thereby, providing users with a means of quantifying uncertainties in parameter estimates.

### Bottom-up Method

The [bottom-up method](https://wpgp.github.io/bookworm/gridded-population-estimates.html#ref-leasure2020national)takes geolocated household survey data from a sample of locations and fits it to a statistical model to estimate population sizes for unsampled areas based on relationships with geospatial covariates. This approach applies customised statistical models to make the best use of available model input survey data, and the use of Bayesian statistical inference provides probabilistic estimates of uncertainty in the parameter estimates. This is a suitable approach requires the availability of a recent geolocated nationally representative household survey dataset or partial census data and can be used to support census preparations and implementation. It is useful tool for filling census data gaps in countries where recent census data have not been undertaken or where the existence of hard-to-reach areas meant that some places could not be visited during a recent census thereby leading to incomplete census data. Note that while this approach provides an opportunity to quantify uncertainty with ease, it also requires more detailed and reliable input population data and requires more time for model development, testing and implementation.

The methodology adopted by an NSO will be determined by the availability of the input data to the model. Where a country has complete census coverage that it wants to try and disaggregate down to a high-resolution (100m or 1km) grid, the top-down methodology will be applied. The top-down approach has been applied in Mali and Burkina Faso previously. Where a country has incomplete or no census data available, the bottom-up method will be applied. This bottom-up approach has been applied previously in multiple countries, including the Demographic Republic of the Congo, Zambia and Nigeria. It should be noted that as the different methods require different inputs, the gridded population estimates that they produce will have different characteristics.

These approaches to estimating population also align with the methods used by the UN Population Division in the production of their World Population Prospects report. These methods are therefore mature and trusted enough to provide consistent and comparable denominator data where a census cannot be fully enumerated.

### Comparison on use of alternative approaches in the non-digital versus digital census era

In the era before widespread digital technology, alternatives were constrained by the physical limitations of paper, logistics, and manual data processing. The goal was primarily to reduce the immense cost and time of a full enumeration. Digital technology, data science, and increased connectivity have unlocked powerful new alternatives. The focus has shifted from simple cost-saving to creating a dynamic, integrated data ecosystem.

## Key implementation areas in a digital census

### Census preparations

Geospatial population modelling techniques are used to support NSOs in preparation for censuses. By using recent georeferenced nationally representative household surveys and geospatial covariates, population modelling methods provide useful insights on the recent estimates of population to the NSOs thereby helping in resource allocation/management for a successful census. Specifically, it helps to estimate the required number of digital census materials (e.g., tablets and smart phones) to be deployed especially in countries that have not had census for a long time.

### Filling population data gaps in hard-to-count areas

Despite the known advantages of digital census, issues of missing data or incomplete census data could still arise, especially in contexts where individuals are unwilling to participate in the census activities or where there may be internet connectivity barriers. Such challenges are further exacerbated by conflicts and insecurity which makes it unfeasible for in-person assistance from the enumerators. In some cases, outright refusal to participate in the census processes also leads to population data gaps.

### Disaggregation of population estimates to lower administrative units

When digital census data is only available at coarser administrative units, geospatial statistical population modelling techniques, in particular top-down approach, could be used to disaggregate the population estimates to more spatially detailed small area scales. Here, Machine Learning techniques are used to provide the population density weights required to disaggregate the coarser units population data first to a 100m or 1km grid cell level before aggregating to the small area unit of interest.

## Selected Country experiences

**Burkina Faso** faced challenges during its 2019 census due to security issues in the north and east, which left some areas unenumerated. To address this, the National Institute for Statistics and Demography (INSD), in collaboration with UNFPA and WorldPop, developed predictive methods to estimate populations in inaccessible regions. These methods produced gridded population estimates for the entire country, helping to fill critical data gaps.

**Papua New Guinea** had to cancel its 2021 census due to COVID-19 disruptions. In response, the National Statistics Office partnered with WorldPop to use geospatial modelling for population estimation. This approach integrated satellite imagery, census preparation data, and malaria survey data with over 50 geospatial datasets. A two-stage modelling process was used to account for settlements hidden under forest canopies. The final model provided population estimates with 95% confidence intervals across all administrative levels. Despite technical challenges like data interoperability and boundary mismatches, the results were openly shared through global data platforms to support future census planning.

**Cameroon** has not conducted a national census since 2005 and faces significant demographic data gaps, especially in conflict-affected regions. To address this, the National Institute of Statistics collaborated with WorldPop to produce high-resolution 2022 population estimates using geospatial modelling. This method combined satellite imagery, survey data, and advanced Bayesian techniques. The resulting gridded data has been used for survey planning, vaccination campaigns, and election logistics. It also supported the creation of a digitized national sampling frame. While not a substitute for a full census, these estimates provide the only granular population data co-developed with the government and are vital for planning and service delivery.

**Zambia** postponed its 2020 census to 2022 due to the pandemic and funding issues. In the interim, the Zambia Statistics Agency (ZamStats) worked with WorldPop under the GRID3 project to develop bottom-up population estimates using household survey data and geospatial information. These estimates were used for malaria control and COVID-19 vaccination planning. Ahead of the 2022 census, ZamStats also explored household count estimates to refine enumeration areas. The 2022 census was Zambia’s first fully georeferenced census. Afterward, WorldPop helped produce top-down gridded estimates aligned with preliminary census results, offering detailed population distribution insights. These estimates will be updated with final census data in 2025, and comparisons with earlier modelled estimates are ongoing.

**Ethiopia** has not conducted an official census since 1974, resulting in major data gaps. In 2023, the Ethiopian Statistical Service (ESS) and UNFPA partnered with WorldPop to model national population estimates. Given the availability of comprehensive 2015–17 cartographic data, a top-down modelling approach was chosen. Household counts were combined with geospatial covariates like land cover, climate, infrastructure, and night-time lights. Residential characteristics were derived from global settlement data to exclude non-residential structures. The resulting national population estimates are currently undergoing validation and refinement before publication by ESS.

Link to case studies

**Burkina Faso**

Burkina Faso’s National Institute for Statistics and Demography (INSD) carried out its fifth population and housing census in late 2019, but security issues in the north and east of the country meant that some areas could not be fully enumerated.

In collaboration with UNFPA and WorldPop, the Institut National de la Statistique et de la Démographie (INSD) co-developed methods for predicting population numbers in areas where enumeration could not take place and producing gridded estimates for the full country.

**Papua New Guinea**

The National Statistics Office of Papua New Guinea (PNG) collaborated with WorldPop on geospatial modelling to estimate PNG's population after COVID-19 disrupted their 2021 census plans. The approach integrated satellite imagery to map building locations, with sample population data from census preparations and malaria bednet distribution surveys. This data was combined with over 50 different geospatial datasets including land cover types, infrastructure, and climate to identify statistical relationships between population density and environmental factors.

The modelling process employed an innovative two-stage approach to overcome the challenge of forest canopy-covered settlements that weren't visible in satellite imagery. First, they developed a model to predict settlement presence even when buildings couldn't be seen from above, then incorporated these predictions into the population estimation model. Multiple statistical models were tested to find optimal correlations between sample population data and geospatial features, with the final selection providing population estimates with 95% confidence intervals at various administrative levels (census unit, district, province, and national).

Technical challenges included data interoperability issues between government datasets and satellite imagery, boundary matching problems, and gaps in data coverage. The final model estimated PNG’s 2021 population at 11.78 million (with a 95% confidence interval of 11.64-12.03 million), broken down by sex and age classes across different administrative levels. All methods, datasets, and documentation were made openly available through WorldPop's Open Population Repository and the UN's Humanitarian Data Exchange, creating a valuable resource for PNG's future census planning.

**Cameroon**

Cameroon's last national census was conducted in 2005, and the country has yet to finalise preparations for its Fourth General Population, Housing, and Agricultural Census (GPHC). The lack of recent census data has created significant demographic gaps, particularly in the Far North, Northwest, and Southwest regions, which were excluded from the 2018 pilot census due to security concerns.

To address these data gaps, the National Institute of Statistics (INS) partnered with WorldPop to develop innovative geospatial population modelling. This approach integrates multiple data sources, including georeferenced survey data, satellite-based settlement maps, and building footprints to generate high-resolution 2022 population estimates at 100m grid cell. The method has proven resource-efficient by leveraging existing national survey data with advanced Bayesian statistical techniques to produce essential demographic estimates, particularly for crisis-affected and inaccessible areas.

The resulting gridded population data has been utilised across numerous applications, including post-stratification adjustments for household surveys, vaccination coverage assessments, and presidential election planning. Additionally, this data enabled the creation of Cameroon's first up-to-date digitised pre-census enumeration area and national sampling frame, which is now being used for major surveys including the Multiple Indicator Cluster Survey (MICS), the Cameroon Household Survey on Living Conditions (ECAM 6), and the Cameroon Demographic and Health Survey (CDHS-VI).

Though the current gridded population estimates provide valuable insights, they cannot fully replace a comprehensive national census, which remains a top priority despite technical constraints. Nevertheless, these estimates represent the only gridded population data co-developed with the Government of Cameroon and serve as the sole source of granular population figures, especially for hard-to-reach areas, supporting critical health interventions, socioeconomic planning, and census preparation efforts.

**Zambia**

Zambia had expected to conduct a national population and housing census in 2020, but due to the COVID-19 pandemic and funding constraints, the census was delayed until 2022. The previous census was conducted in 2010. WorldPop, along with other project partners, started work with the Zambia Statistics Agency (ZamStats) in 2018 during the intercensal period, as part of the [GRID3 project](https://grid3.org/about-us).

As part of the GRID3 project, WorldPop worked with ZamStats to develop [bottom-up modelled population estimates](https://grid3.org/news/zambia-partners-with-grid3-to-produce-pop-estimates). Population counts from existing household survey listings and pilot census cartography, were integrated with detailed geospatial data on buildings in a Bayesian statistical modelling framework. Modelled estimates included gridded counts of population, for grid cells of approximately 100m x 100m across the country, along with statistical estimates of uncertainty.

The completion of the bottom-up modelled population estimates in early 2020, coincided with the early stages of the COVID-19 pandemic. Following review and approval of the estimates by ZamStats, the dataset was openly published as operational population estimates, and used in a range of applications. This included [integrating the modelled population estimates into detailed microplanning maps](https://grid3.org/news/zambia-control-malaria) to support the National Malaria Elimination Centre in identifying where to deploy indoor residual spraying or insecticide-treated nets, as part of malaria control programming. Subsequently the population estimates were used [microplanning maps to support the national COVID-19 vaccination roll-out](https://grid3.org/news/maps-covid19-vaccination-zambia) by the Enhanced Programme on Immunisation and Ministry of Health.

In preparing for the 2022 census, ZamStats were also interested in exploring modelled estimates of household counts. Based on the bottom-up modelled population estimates and survey data on household size, WorldPop provided ZamStats with estimates of household counts, which were intended to support in reviewing enumeration areas (EAs). In particular, identifying where the number of households in a single EA may be too large, necessitating potential adjustment of EA boundaries.

The sixth National Population and Housing census was completed in 2022, and was Zambia's first fully georeferenced census, with the location of all residential households recorded nationally. Following the completion of the census, WorldPop worked with ZamStats to produce top-down gridded estimates from the preliminary census results. Ward-level counts of total population were spatially disaggregated using a top-down approach, with geolocated household locations used as a mask to spatially constrain population estimates to the subset of grid cells with residential households recorded. The resulting gridded population estimates are aligned with the preliminary census results on total population, but provide population estimates at 100m spatial resolution, enabling insights into the spatial distribution of population at the sub-ward level. These gridded estimates are expected to be updated based on the final census data in 2025.

Work to compare the Zambia bottom-up modelled estimates against the subsequent 2022 census results is ongoing.

**Ethiopia**

Ethiopia has not conducted as official census since 1974. They therefore have significant data gaps that make the planning and delivery of policy difficult without the detailed evidence to support it.

In 2023, the Ethiopian Statistical Service (ESS) and UNFPA approached WorldPop about support to model population at the national level based on available survey data. Both bottom-up and top-down methods were considered, but based on the availability of 2015-17 census cartography data with complete coverage for the country, a top-down methodology was selected.

The household points from the cartographic survey were combined with the administrative boundaries to produce a count of the number of households per administrative area, and these were then integrated with geospatial covariates such as distance to specific land cover classes, mean precipitation and temperature, slope and elevation, distance to coastline, schools, health facilities, market places, roads, built settlements, and night-time lights to produce the modelled estimates. Because building footprints contain different structure types (e.g. industrial areas, warehouses), information on the non-residential status of the buildings was also used with the residential characteristics of the building footprints were derived from the global human settlement layer.

Population estimates for the whole of Ethiopia were produced, and work is now taking place to validate and refine these estimates so that they can be published by ESS.

## Challenges and Lessons Learnt

* The use of modelled data does not replace the census process. Instead, modelled estimates are a supplement to a more traditional census by filling gaps in the data, and providing high-resolution estimates during the inter-censal period.
* Modelled population estimates are probabilistic. They are model-based and should therefore not be treated in the same way as true population counts. Instead, they are an opportunity to refine census planning when the most recent census is largely outdated, or fill gaps where a full enumeration cannot be delivered.
* Producing modelled population estimates relies on the availability of household surveys or microcensus data to provide the statistical input to the models. The higher the quality of the statistical input data, the higher the quality of the modelled outputs will be, and the smaller the measure of uncertainty around the estimates.
* Producing high-resolution gridded population estimates requires access to a number of geospatial covariate datasets. This is best delivered through data sharing across different parts of Government, where a national mapping agency and/or national space agency may provide access to authoritative geospatial datasets that can be used in the modelling process.
* Geospatial boundaries (administrative or census) should be maintained without topographic errors (such as slivers) to support modelling. The process will allocate points (household, buildings, address etc) to geographic areas, so any topographic errors will lead to points being allocated to either the wrong area, or no area.
* Modelling large geospatial and statistical datasets requires significant processing capacity, and consideration needs to be given to the processing environment in advance of any modelling activity. Recommendations on the hardware specifications are included in the recommendations section of this chapter.

## Recommendations

* To model large datasets, a suitable processing environment is needed. The recommended hardware specifications for modelling population estimates are at least 500GB of storage capacity, 64GB of RAM and 3.80GHz of processing speed.
* The recommended software for population modelling are R and QGIS. Alternative software such as Python or ArcGIS can be used in the process and support for users of these software packages can be offered however most of the guidance is built around R for statistical processing, and QGIS as an open source GIS solution.
* To spatially model population estimates, a strong knowledge of geospatial information systems (GIS) is recommended. Modellers should understand key GIS techniques.

## Resources

**World Population Prospects 2024 Methodology of the United Nations population estimates and projections**

<https://population.un.org/wpp/assets/Files/WPP2024_Methodology-Report_Final.pdf>

**UNFPA Technical Note on the value of modelled population estimates in the census**

<https://www.unfpa.org/sites/default/files/resource-pdf/V2_Technical-Guidance-Note_Value_of_Modeled_Pop_Estimates_in_Census.pdf>

**WorldPop Book of Methods**

<https://wpgp.github.io/bookworm/contributing-authors.html>