Preprints.org

Review

Not peer-reviewed version

Advances in Small Area Population Estimation in the Absence of National Census Data

Attila N. Lazar *, Gianluca Boo, Heather R. Chamberlain, Chibuzor Christopher Nnanatu, Edith Darin, Douglas R. Leasure, Ortis Yankey, Assane Gadiaga, Sabrina Juran, Luis de la Rua, Jessica Espey, Andrew J. Tatem

Posted Date: 17 December 2025

doi: 10.20944/preprints202506.1625.v2

Keywords: population modelling; census; geostatistical models; surveys; remote sensing

Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 December 2025 doi:10.20944/preprints202506.1625.v2

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

Advances in Small Area Population Estimation in the Absence of National Census Data

Lázár AN [1,*], Boo, G [1], Chamberlain HR [1], Nnanatu CC [1], Darin E [2], Leasure, DR [2], Yankey O [1], Gadiaga A [1], Juran S [3], de la Rua L [3], Espey, J. [1] and Tatem AJ [1]

[1] WorldPop, Department of Geography and Environment, University of Southampton, UK
[2] Demographic Science Unit, Nuffield Department of Population Health, University of Oxford, UK
[3] Latin America and the Caribbean Regional Office, UNFPA, Panama City, Panama
* Correspondence: a.lazar@soton.ac.uk

Abstract

Population data at small area scales are essential for effective decision-making, influencing public health, disaster response, and resource allocation, amongst others. While national censuses remain the cornerstone of population data, they are often constrained by high costs, infrequent collection cycles, and coverage gaps, which can hinder timely data availability. To address these challenges, geospatial statistical approaches using limited microcensus surveys have been demonstrated as a reliable source, but the field has advanced substantially in recent years, with significant developments in both data sources and modelling methodologies. New approaches now leverage routine health intervention campaign data, satellite-derived settlement maps, and bespoke modelling approaches to produce reliable small area population estimates where enumeration is difficult or outdated. Various countries are applying these techniques to support census operations, health program planning, and humanitarian response. This manuscript reviews recent advances in 'bottom-up' population mapping approaches, highlighting innovations in input data, modelling methods, and validation techniques. We examine ongoing challenges, including partial observation of buildings under forest canopy, population displacement, and institutional uptake. Finally, we discuss emerging opportunities to enhance these approaches through better integration with traditional data ecosystems, capacity strengthening, and co-production with national institutions.

Keywords: population modelling; census; geostatistical models; surveys; remote sensing

1. Introduction

In a world where data increasingly shape policies, planning, and resource allocation, significant gaps persist in our knowledge of basic human demographics. Central to these gaps is the lack of small area data on population counts, arguably the cornerstone of planning and decision-making for governmental, non-governmental, and private organisations [1–4]. Vital registries (records of births, marriages and deaths) and the population and housing census, hereafter referred to as 'census', are traditionally the most extensive demographic data sources. However, both face practical challenges that can limit their capacity to provide actionable data.

Censuses are typically conducted on a decadal basis, with results often released at a coarse spatial resolution, due to challenges with accurate geographic boundaries and to protect the privacy of individuals. Moreover, censuses require substantial human and financial resources, and ensuring full coverage can be particularly complex in remote or marginalized areas [5,6]. Data-rich countries may use population registers that are updated with administrative data collected at service points, but these require robust administrative systems which are limited or lacking in most low and lower-middle income countries [7]. Urbanization, conflicts, migration and climate change are leading to rapid changes in population distributions and demographics, with an increasing need for regularly updated demographic data. The United Nations' 2030 Agenda for Sustainable Development [8]

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 December 2025 doi:10.20944/preprints202506.1625.v2

underscores the need for high-quality, timely, and small area population data. Half of all Sustainable Development Goal (SDG) indicators, such as those on poverty, health, education, and gender equality, require small area population data as a denominator, disaggregated by sex, age. However, millions of people remain invisible due to outdated censuses, undercounts, or inability to reach hard to count populations. Developing complementary population estimation methods becomes both a technical necessity and a human rights imperative to achieving the SDG principle of leaving no one behind.

Small area population estimation methods have increasingly been recognized as critical tools not only for addressing data gaps during intercensal periods, but also for strengthening census operations themselves [9]. These methods can help address coverage issues, cartographic updates, and support demographic reconciliation processes post-enumeration. Importantly, such methods should not be seen as replacements for a national census, but rather as a complementary source that can enhance accuracy, equity, and responsiveness of statistical systems. The integration of these approaches into national data systems requires technical capacity, institutional coordination, and political will.

In 2018, Wardrop, et al. [10] envisaged a census-independent 'bottom-up' estimation approach to produce timely, subnational demographic estimates in countries where demographic data sources from censuses or registries were outdated, incomplete, or unavailable. The 'bottom-up' approach leverages datasets from small area enumerations and high-resolution geospatial datasets derived from satellite imagery and other sources. The envisaged modelling methods aimed to estimate population counts and demographic characteristics at high spatial resolution across unsurveyed areas, together with measures of statistical uncertainty. Due to the variability in available demographic data, with differences in geolocation accuracy, sample sizes and spatial coverage, bottom-up models were anticipated to require bespoke design and development.

Since the publication of Wardrop et al., conflicts, political changes, natural disasters, and the COVID-19 pandemic have resulted in census delays and cancellations in many countries, as well as increasing rates of undercounting in many censuses conducted in the 2020 round [11–13]. Fuelled by these issues and the substantial increase in geospatial and non-census data availability, the bottom-up approach has seen important advances in the past five years. Here we review the current state of these methods, with a focus on resource-poor settings. We present an overview of novel input demographic and geospatial datasets and modelling frameworks and the validation of outputs. Finally, we discuss current challenges and future opportunities facing the field.

2. Approaches to Census-Independent Bottom-Up Population Estimation

This section reviews data sources and bottom-up modelling approaches for small area population estimation with examples. Figure 1 provides an overview illustration of the main components and options for population modelling.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 December 2025 doi:10.20944/preprints202506.1625.v2

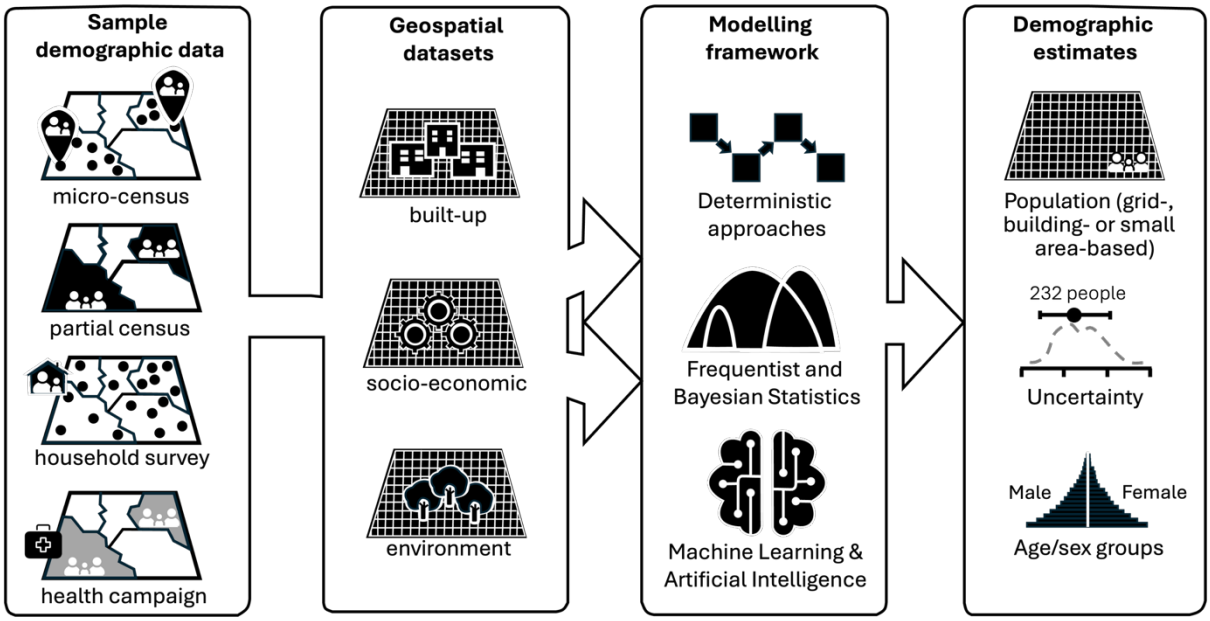

Figure 1. Components of the bottom-up approach to small area population estimation. Different sample demographic datasets are combined with ancillary geospatial datasets covering the entire study area in bespoke modelling frameworks to estimate population distributions and demographics.

2.1. Demographic Datasets

Bottom-up demographic estimation methods leverage observed population data to predict demographic attributes, such as population counts and age and sex structures, across unsampled locations. In recent years, digital collection and geolocation have become standard survey tools for censuses, household surveys, and health campaigns [11]. This has resulted in significant improvements in attribute accuracy, enabling easier integration of these datasets into modelling processes. These demographic datasets can have different levels of spatial completeness (Figure 1]. Wardrop et al. introduced the concept of 'microcensus' surveys for bottom-up population modelling, stressing the advantage of rapid data collection [10,14–17]. These surveys enumerate fully a relatively small, representative sample of well-defined small areas, typically of approximately three hectares, collecting geolocated total household size and age/sex distributions. They can accommodate more complex sampling designs, such as stratified and probabilistic samples [18], but do require considerable human and financial resources — albeit a fraction of a full census — and face similar challenges to a census in terms of inaccessibility and insecurity.

While some bottom-up model applications have relied on microcensus survey data, a wider set of models have since been developed to use recent, partial census enumeration to fill geographic gaps in census data collection [19,20] or to support census planning processes [21]. These data consist of demographic attributes associated with household coordinates or enumeration areas, have very large sample sizes, but do not provide a random sample of the population. Gaps in enumeration typically occur due to accessibility issues [6,22], undercounts in certain types of communities, such as high-rise buildings and gated communities [23], and among certain demographic groups [5,24]. Population modelling approaches can help to identify and address such gaps.

Another demographic data source is household listings collected in preparation for national household surveys. These surveys are designed to be nationally representative by fully enumerating a stratified random sample at selected locations. Repurposing the household listing offers the advantage of frequent and low-cost estimates updates [25–27]. Such datasets typically include geographic data on the cluster locations and, in some cases, the coordinates of de-identified individual households and the sampling weights [28].

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 December 2025 doi:10.20944/preprints202506.1625.v2

A different type of population data can be obtained from service planning and delivery activities targeting every household in a given area. For example, insecticide treated nets (ITNs) distribution campaigns often use a house-to-house distribution model, with relevant demographic attributes collected at the household level. Population counts collected as part of ITN distribution campaigns have been used in several bottom-up applications [29,30], and have the advantage that semi-regularly collected campaign data can enhance sample coverage and frequency, offering a practical complement to censuses or microcensus surveys. However, the household-level population data collected during distribution campaigns are not designed to be representative samples and details of the data collection protocol are often not documented. Thus, these datasets are more prone to specific biases, such as omitted or duplicated of households, or inflated household sizes, requiring careful quality assessment and adjustment.

2.2. Geospatial Datasets

Geospatial datasets typically representing phenomena related to population distribution are an essential component of bottom-up population modelling. Geospatial datasets have advanced substantially since Wardrop et al. in quality, spatial detail and update frequency, particularly those derived from satellite imagery. These developments are associated with increasing computing power and advances in machine learning algorithms, enabling classification and feature extraction at scale. To be able to predict into areas missing demographic data, geospatial datasets must have complete coverage of the area of interest. Such ancillary datasets are summarised within small areas, administrative or census units, or in a regular gridded format, with spatial resolution matching the target population estimates — e.g. approximately 10m [31] or 100m [32].

The most important geospatial datasets for population modelling describe built-up areas [33] that have seen major advances in recent years: now mapping individual buildings (e.g. footprint outline and estimated height) or settled areas with near-global coverage from multiple sources [31,34–37]. Building footprints can be used to derive, for example, building count, area, perimeter, volume, and other morphological characteristics that can be valuable predictors (e.g. settlement types, neighbourhood types, residential/non-residential classes) [31,38–42]. Building footprints are now becoming available with annual timestamps, such as the Google 2.5D Open Buildings dataset, enabling temporal alignment with demographic datasets [43].

Other geospatial datasets leverage the growing availability of volunteered geographic information (VGI). OpenStreetMap for example [44], provides detailed geographic locations for different point of interests (e.g., health and education facilities) associated with population distribution [45]. Regional and global datasets mapping health and education facility [46], violent events [47], and building damage [48,49] also offer potential value as predictors of population at small area scales. These point location datasets can be summarised using focal windows to map push-pull effects of different phenomena on population distributions. Additionally, digital trace data from mobile devices have been used as indicators of population presence, including call detail records [50] and smartphone app data [51].

Population distributions are also associated with environmental factors, such as topography, climate, water bodies, and land cover [33] — available globally at high spatial resolution and with frequent updates [52]. Similarly, remote sensing images of nighttime light brightness can capture features relating to human presence, such as areas with electrification and fires, though need careful processing to remove features like gas flares and volcanic eruptions [53].

2.3. Modelling Frameworks

Wardrop et al. suggested simple regression and geostatistical approaches as potential bottom-up modelling frameworks. Other approaches to bottom-up population estimation have been developed to accommodate available input data, including deterministic approaches, frequentist and Bayesian statistical models, machine learning, and artificial intelligence (AI). These approaches all

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 December 2025 doi:10.20944/preprints202506.1625.v2

leverage relationships between the demographic and geospatial datasets to estimate demographic attributes at both sampled and unsampled locations.

The simplest bottom-up model combines building or settlement information with average values of population densities (e.g. people per km^2 or people per building) applied uniformly across the study area. Such models can be improved when the calculation differentiates residential/non-residential buildings, households per building, or the average household size [54]. While accounting for local variations might be limited in such approaches, the availability of high resolution observed building density, building height and building type datasets can dramatically improve the precision of outputs compared to solely using the mapped presence of built settlement [42].

Leasure et al. [16] used Bayesian hierarchical models to estimate population counts as a Poisson process using the product of settled area and estimated population densities. Population densities are modelled as a lognormal process to better capture variability in the input data. The model is designed to draw information from different scales, such as coarse administrative units (e.g., regions or districts) or functional areas (e.g., settlement classes), to define the model parameters, including intercepts, covariate (i.e. geospatial data) effects, and variance terms [16,26]. Hierarchical intercepts account for stratified sampling designs as well as potential patterns in population density. Moreover, by using settlement type or administrative unit fixed effects, key information is shared among various spatial scales, thereby, leveraging observations from neighbouring clusters to predict population densities across clusters with no observations. Finally, the settlement-class-specific variance term captures the residual variation. More recent hierarchical model designs now use building footprints to define these parameters enabling more accurate population distribution predictions [17,19,25], with further developments allowing joint estimation of building count and population distributions [18,55].

To enable the use of household surveys for population modelling [25] required a crucial study by Leasure et al. [28] that demonstrated that a population-weighted sample (i.e. national household surveys with a probability proportional to size design) results in biased estimates of population densities. They explored various Bayesian approaches and concluded that weighted-likelihood models can recover unbiased estimates of population densities and population totals from a non-random, population-weighted sample.

Spatial variation in bottom-up population models is needed, because populations are not randomly or uniformly distributed in space, and thus ignoring spatial autocorrelation can lead to biased estimates. Nnanatu et al. [27] used Bayesian geostatistical models to explicitly account for spatial autocorrelation not captured by covariates through leveraging distance-dependent covariance matrices. This included a two-stage Poisson-Gamma geostatistical regression. Similarly to the hierarchical models, this design also assumes that population density is related to a set of geospatial covariates through a linear regression model of covariate effects, but estimates population counts as the product of the predicted population density and the corresponding building counts. However, the Poisson-Gamma approach utilises a Gamma probability density instead of the lognormal of the hierarchical models above to model the population density. Geostatistical models specify spatial autocorrelation using a triangulation (or mesh) of the entire continuous spatial domain. These models have been also applied where, due to permanent canopy covers, building footprints were only partially available [29].

Machine learning (ML) and AI can be used to estimate intercensal population counts at high spatial resolution directly from satellite imagery. Landscan HD (56] uses a data fusion approach to merge relevant land use and building layers in a deep learning framework to generate gridded population estimates. Convolutional neural networks with visual geometry group architecture and other ML models (e.g. gradient boosting) are also used for deep learning directly from a concatenation of low-resolution Landsat-7/8, Sentinel-1, and satellite-derived night-time light intensity [57–60]. These new methods leverage frequent and freely accessible satellite imagery updates to produce frequently updated population estimates. The use of very high-resolution satellite imagery, such as the Maxar VIVID 2.0, in conjunction with a small sample of ground truth

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 December 2025 doi:10.20944/preprints202506.1625.v2

dataset (i.e. microcensus), appears promising to help with the identification of inhabited areas and non-residential buildings, and thus estimating population distributions more accurately [61].

Most of the bottom-up modelling frameworks outlined above estimate the total population distributions, but service planning, resource allocation and exposure assessments typically require data on specific sex and age subgroups. The most widespread method for age- and sex-disaggregation uses observed or projected subnational population pyramids to deterministically disaggregate total population count estimates [62]. However, when the input demographic datasets contain age and sex information, bottom-up models can also approximate sex and age subgroups. There are examples where the proportion of the population under the age of five years is estimated at high spatial resolution using a Bayesian spatio-temporal model[63], or where population pyramids are estimated as Dirichlet-multinomial processes [17], or by using a series of Bayesian binomial processes [64]. Field campaigns and development planning often require household level information that can also be modelled at small area scales using Bayesian geostatistical and hierarchical models [27,65] or household characteristics modelled through machine learning methods [66].

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 December 2025 doi:10.20944/preprints202506.1625.v2

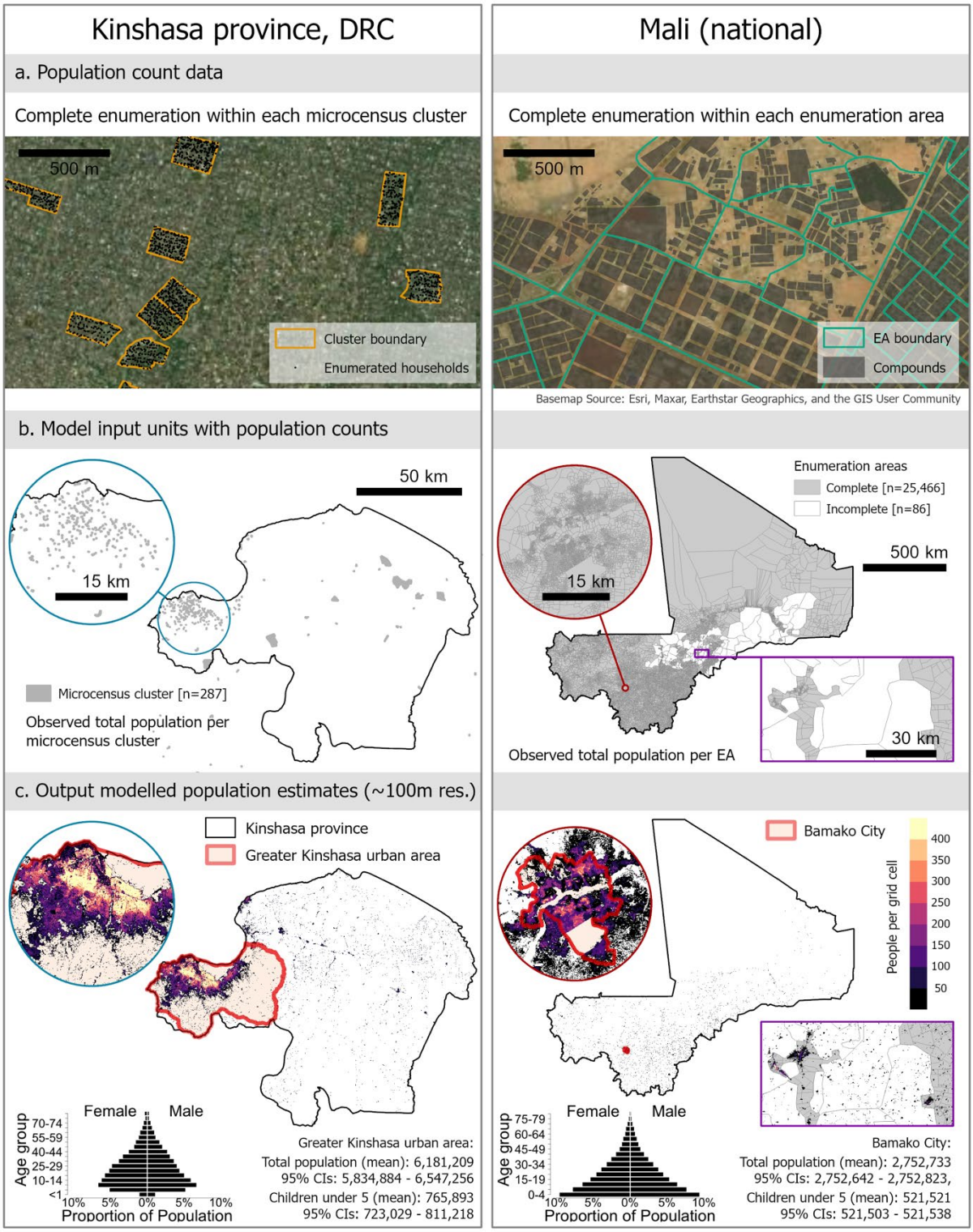

Figure 2. Example bottom-up model applications: field observations (panel a), input data (panel b), gridded population estimates with age/sex and uncertainty information (panel c). Data sources: Democratic Republic of the Congo (DRC) microcensus [67], DRC population estimates [17], Mali census cartography [68], Mali population estimates [21].

3. Outputs and Validation

Bottom-up population estimates have been typically produced for regular grid cells, with a spatial resolution of approximately 100m (Figure 2]. This format provides a continuous surface enabling the flexible aggregation to different geographic decision-making units, such as administrative districts, health zones, or summarized by features of interest, such as urban extents,

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 December 2025 doi:10.20944/preprints202506.1625.v2

flooded areas, or hurricane tracks. This format also facilitates integration with other geospatial datasets, for instance, health facilities, conflicts, schools, and polling booths, for geographic accessibility analyses. Furthermore, bottom-up population estimates have been used to calculate social-distancing indicators [69], map out-of-school children [70,71], support disaster response [72,73], create survey sampling frames [74,75], and plan vaccination campaigns [76].

An understanding of the confidence in modelled population estimates is important for decision makers and planners. ML models generally capture model uncertainty by using boot-strap resampling of the training data, that is to repeatedly draw random samples from a single original training dataset to create many new, simulated datasets (59, 66]. This tests that the model can generalise well to unseen areas. Other uncertainty estimation uses grid search, a hyperparameter optimization technique that exhaustively tries every possible combination of hyperparameter values from a user defined grid [61]. Validation of ML models include leave-one-out cross-validation, quantifying how close the observed and predicted probabilities or counts are to each other.

Bayesian methods quantify statistical uncertainty. Model assumptions, guided by expert knowledge and data availability can be tested in simulation studies [28,29], assessing input data and model structure. Model uncertainty is quantified by assessing the spread of the parameters' posterior distribution. These posterior distributions are used to calculate metrics such as the mean value of the predicted population counts, the standard deviation, and 95% credible intervals (CIs). Model uncertainties are propagated through the Bayesian CIs. For example, Boo et al [18] propagated uncertainties of observed buildings through the modelled population density and the gridded population estimates to the various levels of spatial aggregations of the outputs. The results highlighted that the uncertainty in the modelled estimates varies spatially and increases with granularity. Knowledge of spatial variation of uncertainties can enable more informed decisions. For example, where there is a specific target for vaccination coverage of children under 5 years e.g. 90%, the upper CI can provide an estimate of the maximum number of children expected to be found, guiding vaccine amounts required to achieve target coverage. Spatial maps of uncertainty can also be used to prioritise where additional demographic data collection can maximise improvement of modelled estimates.

Model validation of population models is generally done thorough model diagnostics, goodness-of-fit assessment, and external comparisons against an independent set of observations. Model diagnostics include amongst others: checking assumption of linear regression, checking parameters convergence, posterior predictions, grid-cell predictions, undertaking cross-validation to check the robustness of the design, as well as uncertainty visualisation [16]. Goodness-of-fit assessments evaluate differences between the population estimates and the input data through different metrics, such as root mean square error (RMSE), mean average error (MAE), correlation coefficients (Person's correlation or R^2], and spatial autocorrelation (global and local Moran's I) [27].

Plausibility checks against population projections or other external sources of population data are also widely used to examine estimates of population models. Chamberlain et al. [77], for example, compared the 2022 national Zambian census counts with bottom-up modelled estimates representing the year 2019. Their analysis showed a strong correlation between the two datasets with some degree of regional positive bias, however the temporal difference between the modelled estimates and the census also complicated the comparison. To counter this external data triangulation issue, Darin et al. (55] sampled the 2018 Colombian census data to create an artificial data set and thus making the sample and the validation data directly comparable. They tested two Bayesian and one ML model for accuracy and found that the ML model was the best suited for unbiased, complete observations, whereas the Bayesian methods were the best for coarse or biased observations and for sparsely populated regions.

4. Challenges and Future Directions

Recent advances in data sources and methods have addressed many challenges outlined by Wardrop et al. [2018], but some remain, and new ones have emerged as applications involve

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 December 2025 doi:10.20944/preprints202506.1625.v2

increasingly diverse data sources. These challenges can be summarised into i) data input challenges, ii) methodological considerations, and iii) implementation barriers (Figure 3].

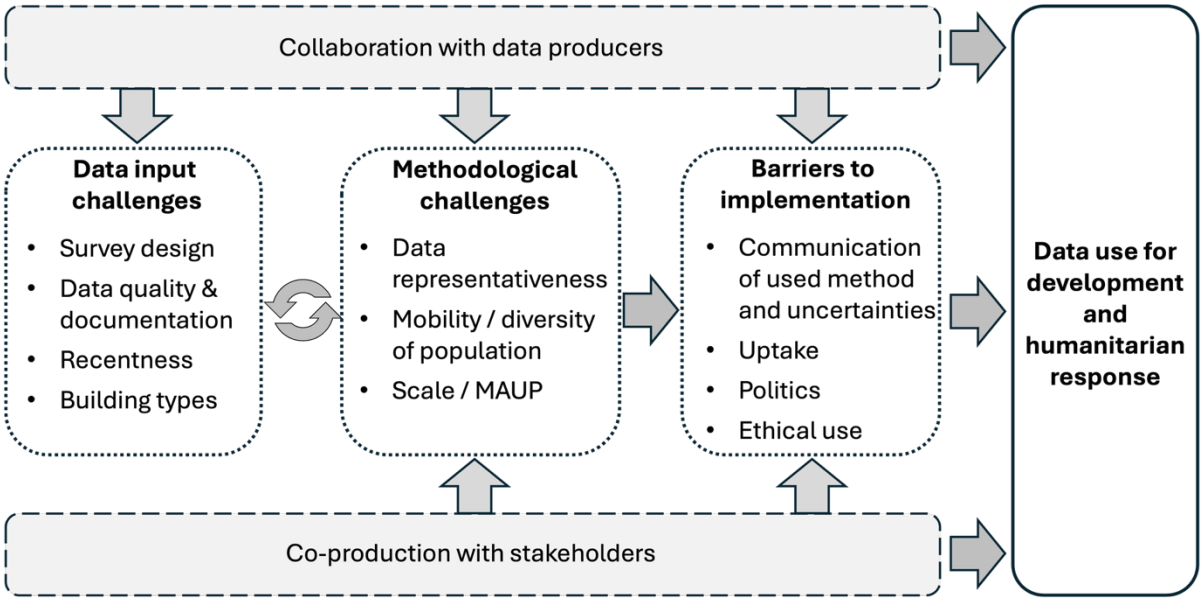

Figure 3. Challenges (white boxes) and opportunities (grey boxes) towards sustainable data production and use. MAUP = Modifiable Areal Unit Problem.

5. Data Input Challenges

Challenges persist regarding demographic data quality and suitability. Field data collection has benefited from GPS-enabled devices that facilitate navigation and location recording, but practical limitations remain, especially around accessibility and safety concerns for enumerators [78,79]. Local community scepticism or distrust around data collection efforts can result in non-response or incomplete coverage, particularly among hard to count populations [80]. Enumeration data collected through health distribution campaigns can result in biases in numbers upwards or downwards, depending on incentives in receiving services or resources [81]. Though field data collection software enables near real-time quality checks, data quality ultimately depends on survey design, field protocols, data management and team training [79]. This underscores the importance of thorough documentation of survey methodologies, particularly as financial resources for field data collection become increasingly constrained [82].

For increasing the temporal relevancy of population estimates, digital trace data from mobile phone call detail records [83], smartphone location history [84], mobile applications [85], and social media platforms [86,87] have demonstrated potential for measuring population distributions and movements in near real-time. These data sources are valuable for their volume and frequency of updates, though spatial resolution varies along urban-rural gradients and representativeness across population subgroups remains uncertain. Methods to correct for these biases using survey data are emerging [88], but access challenges due to private ownership persist.

Despite major developments in building footprint and settlement datasets, in terms of global coverage and availability, challenges remain around temporal and spatial coverage due to cloud cover, tree canopy obstruction, and certain areas not being systematically mapped in open datasets due to conflicts or other sensitivities. Documentation of these datasets can be limited, making it difficult to ascertain the currency of imagery used in their construction, classification criteria, and methodological assumptions. Substantial variations exist between datasets [89–91], likely translating into significant differences in bottom-up population estimates.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 December 2025 doi:10.20944/preprints202506.1625.v2

6. Methodological Considerations

Estimating remote rural population numbers at small area scales remains challenging [92]. Moreover, at the other end of the population density spectrum, the accurate estimation of informal settlement population numbers in urban areas presents particular difficulties [93]. Combining satellite imagery data, statistical modelling, and community engagement is however offering a promising solution [94]. Mobile populations also present a significant challenge. According to UNHCR, 117.3 million people worldwide were forcibly displaced at the end of 2023, excluding those displaced by environmental disasters and climate stressors [95]. These rapid-onset events driving displacements and other localized movements, including nomadic populations, are generally not captured in the types of static population models primarily discussed in this paper. Recent model applications have attempted to address this by creating displacement-adjusted layers that redistribute populations based on known displacement patterns [96]. Feature extraction algorithms have been developed to map structures in refugee camps [97], and humanitarian administrative data has been integrated with satellite-derived building footprints to map registered refugee populations [98]. However, many types of mobile populations remain inadequately represented in both current bottom-up modelling approaches and traditional demographic datasets in general.

Spatial autocorrelation in input data can require attention as it can alter observation variance and increase sensitivity to certain covariates [99,100]. Furthermore, spatial autocorrelation is particularly important when there are spatial dependences in the data that cannot be accounted for by the model covariates, i.e. when spatial autocorrelation remains in the residuals of a fully specified model. The use of geostatistical models that handle spatial autocorrelation can address this.

The modifiable areal unit problem (MAUP) is the statistical bias arising from the choice of spatial units used to summarise point-based observations [101]. MAUP represents a crucial modelling consideration, particularly when the spatial scales of observations differ from modelling units (e.g. grid cells). This mismatch, often called ecological fallacy, can reduce explained variance, increase residual spatial autocorrelation, and potentially lead to erroneous conclusions [102,103]. Simulation studies can help determine appropriate scale choices [104], while Bayesian models that incorporate the sampling frame components can better account for such population distribution uncertainties [102].

7. Implementation Barriers and Opportunities

While a growing number of applications show that bottom-up population estimates provide valuable support for planning and decision-making during intercensal periods, they often face barriers to acceptance by decision makers, who may be reluctant to use modelled data and to collaborate with external partners to produce population estimates that may conflict with official projections. Population statistics have significant implications for resource allocation [105], political representation [106], and national economic performance and SDG indicators [107], making their official endorsement politically sensitive but key to ensure data ownership [108]. To facilitate greater governmental uptake, several approaches have shown promise. There is a need to have a supportive data policy and regulatory environment, which is often a stumbling block for national statistical offices and ministries to collaborate with external third parties. Furthermore, enhanced communication of model methods, results and associated uncertainties [109] and crucially, co-development with decision-makers and local ownership of the modelling process [110]. Local ownership ideally means that when external partners are brought in to support the design and use of these new methods, the approach is co-developed with end users, such as national statistical offices, and is accompanied by appropriate capacity strengthening for sustainable data generation and use [111]. Medium- to long-term investment is needed to develop and retain institutional skills and strengthen national university programs to cultivate the next generation of geospatial statisticians. Moreover, building links with the rapidly growing ML and AI expertise in start-ups and NGOs can drive local capacity strengthening and sustainable adoption.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 December 2025 doi:10.20944/preprints202506.1625.v2

A growing concern is that high-resolution population estimates might expose vulnerable groups to unwanted surveillance or security risks and might impact equitable resource allocations [112,113]. Ensuring responsible use of digital technologies, as well as the publication and use of population data are crucial. Good practice includes seeking approval from independent ethics boards of the data provider, government and/or universities, de-identification of observations, and publishing of only aggregated statistical estimates. At the UN level, the publication of guidance for countries utilising these modelling approaches and population estimates on effective practice and good governance should be considered. It would empower countries to embrace alternative methodologies to fill crucial population gaps, whilst showing how such modelling approaches can be appropriately and confidently applied.

8. Conclusions

Reliable, timely, and detailed population estimates are increasingly required for effective decision-making, development planning, SDG reporting, operational campaign planning, and humanitarian response. Since Wardrop et al. [10] presented their perspectives on census-independent population estimates, significant advances have been made in both data sources and methodological approaches. The emergence of building footprint datasets, complex data integration with Bayesian and ML modelling frameworks, and better quantification and communication of statistical uncertainty have enabled improved bottom-up population estimates and uptake.

In Burkina Faso [114], Mali [115], Papua New Guinea [116], and South Sudan [117], bottom-up population estimates have been adopted by national statistical offices for national planning. In Colombia, the national statistical office implemented these methods to address census enumeration gaps [20]. UN agencies including UNFPA and the United Nations Children's Fund (UNICEF) are actively promoting and applying these approaches [9,118], as well as practical implementation tutorials [119,120] to support field operations and data sustainability by building in-country capacity to use and update the models.

Population models successfully capture population distributions when using good quality, representative observations. Challenges, however, remain in capturing urban diversity, informal settlements, mobile populations (e.g. homeless, nomadic), and areas with persistent heavy vegetation. Looking forward, the field is poised to benefit from increasing availability of large-scale non-traditional demographic (e.g. digital trace) data, advances in statistical methods, including ML and AI to leverage such big data sources rapidly. Caution is always needed however to address biases in survey data, digital traces and other inputs to models.

As demographic data gaps increase [121], the complementary role of bottom-up population estimation approaches will likely expand. Their ability to address some of the technical challenges likely to be faced in the 2030 census round [11], and their integration with national demographic data systems offers a path toward more responsive and cost-efficient statistics, particularly in contexts of census disruption, under-coverage or demographic reconciliation for planning and monitoring.

Continuing methodological advances and stronger partnerships and co-production between data producers, modelling teams, and stakeholders are needed to further enhance the utility and acceptance of these approaches. For example, in Latin America and the Caribbean, a collaborative partnership model involving UNFPA, Economic Commission for Latin America and the Caribbean (ECLAC), regional statistical entities such CARICOM, and national statistical offices has proven effective [122]. These partnerships leverage contextual knowledge and statistical expertise to co-develop robust estimates. Such rigorous partnerships enable reliable applications, proper institutionalization of the methods, and the achievement of data-driven, inclusive development.

References

1. M. Borowitz, J. Zhou, K. Azelton, I.-Y. Nassar, Examining the value of satellite data in halting transmission of polio in Nigeria: A socioeconomic analysis. *Data & Policy* **5**, e16 [2023].

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 December 2025 doi:10.20944/preprints202506.1625.v2

2. S. P. Cumbane, G. Gidófalvi, Spatial Distribution of Displaced Population Estimated Using Mobile Phone Data to Support Disaster Response Activities. *ISPRS International Journal of Geo-Information* **10**, 421 [2021].
3. P. G. Greenough, E. L. Nelson, Beyond mapping: a case for geospatial analytics in humanitarian health. *Conflict and Health* **13**, 50 [2019].
4. T. A. Robin et al., Using spatial analysis and GIS to improve planning and resource allocation in a rural district of Bangladesh. *BMJ Global Health* **4**, e000832 [2019].
5. USCB [2022] Post-Enumeration Survey and Demographic Analysis Help Evaluate 2020 Census Results. in *Census Bureau Releases Estimates of Undercount and Overcount in the 2020 Census* (United States Census Bureau).
6. S. Randall, Where have all the nomads gone? Fifty years of statistical and demographic invisibilities of African mobile pastoralists. *Pastoralism* **5**, 22 [2015].
7. UNECE, *Guidelines on the Use of Registers and Administrative Data for Population and Housing Censuses* (Geneva, 2018].
8. UN, *The Sustainable Development Goals Report* (United Nations, ISBN: 978-92-1-003135-6, New York, USA, 2024].
9. UNFPA [2020] The Value of Modeled Population Estimates for Census Planning and Preparation. Technical Guidance note. (United Nations Population Fund (UNFPA), https://www.unfpa.org/sites/default/files/resource-pdf/V2_Technical-Guidance-Note_Value_of_Modeled_Pop_Estimates_in_Census.pdf).
10. N. A. Wardrop et al., Spatially disaggregated population estimates in the absence of national population and housing census data. *Proceedings of the National Academy of Sciences* **115**, 3529-3537 [2018].
11. S. Tadesse [2025] The Evolving Census Landscape: Lessons from the 2020 round and anticipated trends for the 2030 round. in *United Nations Statistical Commission, 56th Session, Side Event: Advancing Population and Housing Censuses in the 2030 Round; New York, USA; 5 March 2025* (United Nations Statistics Division (UNSD); https://unstats.un.org/UNSDWebsite/statcom/session_56/side-events/HPC-side-event-5Mar2025/UNSD-PHCs-2020-lessons-and-2030-trends.pdf).
12. E. Jensen, T. Kennel [2022] Detailed Coverage Estimates for the 2020 Census Released Today. in *America Counts: Stories* (United States Census Bureau, March 10, 2022, https://www.census.gov/library/stories/2022/03/who-was-undercounted-overcounted-in-2020-census.html).
13. United Nations Economic and Social Council [2024] Future of population and housing censuses and lessons learned from past and current experiences (E/ESCWA/C.1/2024/4]. (UN, New York).
14. E. M. Weber et al., Census-independent population mapping in northern Nigeria. *Remote Sensing of Environment* **204**, 786-798 [2018].
15. R. Hillson et al., 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**, e112241 [2014].
16. D. R. Leasure, W. C. Jochem, E. M. Weber, V. Seaman, A. J. Tatem, National population mapping from sparse survey data: A hierarchical Bayesian modeling framework to account for uncertainty. *Proceedings of the National Academy of Sciences* 10.1073/pnas.1913050117, 201913050 [2020].
17. G. Boo et al., High-resolution population estimation using household survey data and building footprints. *Nature Communications* **13**, 1330 [2022].
18. G. Boo et al., Tackling public health data gaps through Bayesian high-resolution population estimation. *PLOS Global Public Health* **5[9]: e0005072** [2025].

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 December 2025 doi:10.20944/preprints202506.1625.v2

19. E. Darin, M. Kuépié, H. Bassinga, G. Boo, A. J. Tatem, La population vue du ciel : quand l'imagerie satellite vient au secours du recensement. *Population (french edition)* **77**, 467-494 [2022].
20. L. M. Sanchez-Cespedes et al., Social cartography and satellite-derived building coverage for post-census population estimates in difficult-to-access regions of Colombia. *Population Studies* **78**, 3-20 [2024].
21. WorldPop, Institut National de la Statistique du Mali [2022] Census-cartography-based gridded population estimates for Mali [2020], version 1.0. WorldPop, University of Southampton. doi: 10.5258/SOTON/WP00745; https://wopr.worldpop.org/?MLI/Population/v1.0
22. IPAC [2019] NUMBERS MATTER: THE 2020 CENSUS AND CONFLICT IN PAPUA. (Institute for Policy Analysis of Conflict Jakarta).
23. T. A. Sullivan, "Who, What, When, and Where of the Census" in Census 2020: Understanding the Issues, T. A. Sullivan, Ed. (Springer International Publishing, Cham, 2020], 10.1007/978-3-030-40578-6_2, pp. 17-31.
24. A. M. Wazir, A. Goujon [2019] Assessing the 2017 Census of Pakistan Using Demographic Analysis: A Sub-National Perspective. in *Vienna Institute of Demography Working Papers No. 06/2019* (Vienna Institute of Demography, Vienna).
25. C. A. Dooley et al. [2021] Description of methods for the Zambia modelled population estimates from multiple routinely collected and geolocated survey data, version 1.0. WorldPop, University of Southampton. doi: 10.5258/SOTON/WP00707.
26. R. Engstrom, D. Newhouse, V. Soundararajan, Estimating small-area population density in Sri Lanka using surveys and Geo-spatial data. *PLOS ONE* **15**, e0237063 [2020].
27. C. C. Nnanatu et al. [2025] Efficient Bayesian Hierarchical Small Area Population Estimation Using INLA-SPDE: Integrating Multiple Data Sources and Spatial-Autocorrelation. in *Preprints; 10.20944/preprints202501.0588.v1*.
28. D. R. Leasure, C. A. Dooley, A. Tatem [2021] A simulation study exploring weighted likelihood models to recover unbiased population estimates from weighted survey data. University of Southampton; doi:10.5258/SOTON/WP00706.
29. C. C. Nnanatu et al., Estimating small area population from health intervention campaign surveys and partially observed settlement data. *Nature Communications* **16**, 4951 [2025].
30. C. C. Nnanatu et al., Modelled gridded population estimates for Nigeria 2025 version 3.0. WorldPop,
31. University of Southampton; https://dx.doi.org/10.5258/SOTON/WP00782.
32. M. Pesaresi, P. Politis, GHS-BUILT-S R2023A - GHS built-up surface grid, derived from Sentinel2 composite and Landsat, multitemporal [1975-2030].
33. D. Woods et al., WorldPop high resolution, harmonised annual global geospatial covariates. Version 1.0. University of Southampton: Southampton, UK; DOI:10.5258/SOTON/WP00772.
34. J. J. Nieves et al., Examining the correlates and drivers of human population distributions across low- and middle-income countries. *Journal of The Royal Society Interface* **14**, 20170401 [2017].
35. W. Sirko et al. [2021] Continental-scale building detection from high resolution satellite imagery. arXiv:2107.12283.
36. Microsoft, Worldwide building footprints derived from satellite imagery (GitHub Repository); https://github.com/microsoft/GlobalMLBuildingFootprints. https://github.com/microsoft/GlobalMLBuildingFootprints.
37. Ecopia, Global Feature Extraction: Building footprints; https://www.ecopiatech.com/products/global-feature-extraction.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 December 2025 doi:10.20944/preprints202506.1625.v2

38. X. X. Zhu, S. Chen, F. Zhang, Y. Shi, Y. Wang, GlobalBuildingAtlas: An Open Global and Complete Dataset of Building Polygons, Heights and LoD1 3D Models. *arXiv:2506.04106* [2025].
39. R. Hillson et al., Stratified Sampling of Neighborhood Sections for Population Estimation: A Case Study of Bo City, Sierra Leone. *PLoS One* **10**, e0132850 [2015].
40. W. C. Jochem et al., Classifying settlement types from multi-scale spatial patterns of building footprints. *Environment and Planning B: Urban Analytics and City Science* 10.1177/2399808320921208, 2399808320921208 [2020].
41. C. T. Lloyd et al., Using GIS and Machine Learning to Classify Residential Status of Urban Buildings in Low and Middle Income Settings. *Remote Sensing* **12**, 3847 [2020].
42. L. Tomás, L. Fonseca, C. Almeida, F. Leonardi, M. Pereira, Urban population estimation based on residential buildings volume using IKONOS-2 images and lidar data. *International Journal of Remote Sensing* **37**, 1-28 [2016].
43. F. Schug, D. Frantz, S. van der Linden, P. Hostert, Gridded population mapping for Germany based on building density, height and type from Earth Observation data using census disaggregation and bottom-up estimates. *PLOS ONE* **16**, e0249044 [2021].
44. W. Sirko et al., High-Resolution Building and Road Detection from Sentinel-2. arXiv:2310.11622, https://sites.research.google/gr/open-buildings/temporal/.
45. OSM, OpenStreetMap. https://www.openstreetmap.org/#map=5/54.91/-3.43.
46. B. Herfort, S. Lautenbach, J. Porto de Albuquerque, J. Anderson, A. Zipf, The evolution of humanitarian mapping within the OpenStreetMap community. *Scientific Reports* **11**, 3037 [2021].
47. GRID3, GRID3 Data Hub. https://data.grid3.org/.
48. ACLED, ACLED Data. https://acleddata.com/.
49. C. Scher, J. Van Den Hoek, Nationwide conflict damage mapping with interferometric synthetic aperture radar: A study of the 2022 Russia-Ukraine conflict. *Science of Remote Sensing* https://doi.org/10.1016/j.srs.2025.100217, 100217 [2025].
50. S. Al Shafian, D. Hu, Integrating Machine Learning and Remote Sensing in Disaster Management: A Decadal Review of Post-Disaster Building Damage Assessment. *Buildings* **14**, 2344 [2024].
51. L. Dong et al., Defining a city — delineating urban areas using cell-phone data. *Nature Cities* **1**, 117-125 [2024].
52. N. N. Patel et al., Improving Large Area Population Mapping Using Geotweet Densities. *Transactions in GIS* **21**, 317-331 [2017].
53. C. F. Brown et al., Dynamic World, Near real-time global 10 m land use land cover mapping. *Scientific Data* **9**, 251 [2022].
54. D. Stathakis, P. Baltas, Seasonal population estimates based on night-time lights. *Computers, Environment and Urban Systems* **68**, 133-141 [2018].
55. D. R. Leasure, C. A. Dooley, M. Bondarenko, A. J. Tatem [2021] peanutButter: An R package to produce rapid-response gridded population estimates from building footprints, version 1.0.0. . (WorldPop, University of Southampton, 10.5258/SOTON/WP00717, https://github.com/wpgp/peanutButter).
56. E. Darin, D. R. Leasure, R. Kashyap, How accurate are high resolution settlement maps at predicting population counts in data scarce settings. https://doi.org/10.31219/osf.io/pz38h_v1. [2025].
57. J. V. Tuccillo et al., LandScan HD: a high-resolution gridded ambient population methodology for the world. *Population and Environment* **47**, 42 [2025].
58. P. Doupe, E. Bruzelius, J. Faghmous, S. G. Ruchman [2016] Equitable development through deep learning: The case of sub-national population density estimation. in *Proceedings of the 7th Annual*

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 December 2025 doi:10.20944/preprints202506.1625.v2

Symposium on Computing for Development (Association for Computing Machinery, Nairobi, Kenya), p Article 6.

59. W. Hu et al. [2019] Mapping Missing Population in Rural India: A Deep Learning Approach with Satellite Imagery. in *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (Association for Computing Machinery, Honolulu, HI, USA), pp 353–359.
60. C. Robinson, F. Hohman, B. Dilkina [2017] A Deep Learning Approach for Population Estimation from Satellite Imagery. in *Proceedings of the 1st ACM SIGSPATIAL Workshop on Geospatial Humanities* (Association for Computing Machinery, Redondo Beach, CA, USA), pp 47–54.
61. N. Ahmed et al., Artificial Neural Network and Machine Learning Based Methods for Population Estimation of Rohingya Refugees: Comparing Data-Driven and Satellite Image-Driven Approaches. *Vietnam Journal of Computer Science* **06**, 439-455 [2019].
62. I. Neal, S. Seth, G. Watmough, M. S. Diallo, Census-independent population estimation using representation learning. *Scientific Reports* **12**, 5185 [2022].
63. C. Pezzulo et al., Sub-national mapping of population pyramids and dependency ratios in Africa and Asia. *Scientific Data* **4**, 170089 [2017].
64. V. A. Alegana et al., Fine resolution mapping of population age-structures for health and development applications. *Journal of The Royal Society Interface* **12**, 20150073 [2015].
65. C. C. Nnanatu, S. Chaudhuri, A. N. Lazar, A. J. Tatem [2025] jollofR: A Bayesian statistical model-based approach for disaggregating small area population estimates by demographic characteristics. R package version 0.3.0, https://github.com/wpgp/jollofR/. .
66. C. Nnanatu et al. [2024] Modelled gridded population estimates for Cameroon 2022. Version 1.0, University of Southampton, 17 Jun 2024, DOI: 10.5258/SOTON/WP00784, https://data.worldpop.org/repo/wopr/CMR/population/v1.0/.
67. M. Viljanen, L. Meijerink, L. Zwakhals, J. van de Kassteele, A machine learning approach to small area estimation: predicting the health, housing and well-being of the population of Netherlands. *International Journal of Health Geographics* **21**, 4 [2022].
68. UCLA-DRC Health Research and Training Program, Kinshasa School of Public Health, Kinshasa, Kongo Central and former Bandundu microcensus survey data [2017-18].
69. Institut de la Statistique du Mali, Cartographie du RGPH5.
70. H. R. Chamberlain, A. N. Lazar, A. J. Tatem, High-resolution estimates of social distancing feasibility, mapped for urban areas in sub-Saharan Africa. *Scientific Data* **9**, 711 [2022].
71. V. A. Alegana, C. Pezzulo, A. J. Tatem, B. Omar, A. Christensen, Mapping out-of-school adolescents and youths in low- and middle-income countries. *Humanities and Social Sciences Communications* **8**, 213 [2021].
72. P. M. Macharia et al., Modelling geographic access and school catchment areas across public primary schools to support subnational planning in Kenya. *Children's Geographies* **21**, 832-848 [2023].
73. A. Smith et al., New estimates of flood exposure in developing countries using high-resolution population data. *Nature Communications* **10**, 1814 [2019].
74. F. Hierink, N. Rodrigues, M. Muñiz, R. Panciera, N. Ray, Modelling geographical accessibility to support disaster response and rehabilitation of a healthcare system: an impact analysis of Cyclones Idai and Kenneth in Mozambique. *BMJ Open* **10**, e039138 [2020].
75. S. H. Qader et al., Using gridded population and quadtree sampling units to support survey sample design in low-income settings. *International Journal of Health Geographics* **19**, 10 [2020].
76. J. Cajka, A. Safaa, R. Jamie, J. and Allpress, Geo-sampling in developing nations. *International Journal of Social Research Methodology* **21**, 729-746 [2018].

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 December 2025 doi:10.20944/preprints202506.1625.v2

77. O. Borkovska et al., Developing High-Resolution Population and Settlement Data for Impactful Malaria Interventions in Zambia. *Journal of Environmental and Public Health* **2022**, 2941013 [2022].
78. H. R. Chamberlain, C. A. Dooley, A. J. Tatem, Assessing the accuracy of census-independent small area modelled population datasets
79. . (in preparation).
80. J. M. Davis, M. Wilfahrt, Enumerator Experiences in Violent Research Environments. *Comparative Political Studies* **57**, 675-709 [2024].
81. DESA [2019] Guidelines on the use of electronic data collection technologies in population and housing censuses. (United Nations, New York, January 2019, Department of Economic and Social Affairs Statistics Division).
82. H. Hogan, Distrust in the Governments Brings Risk to the Census. *Harvard Data Science Review* **2**, https://doi.org/10.1162/99608f99692.99611f99603e99977 [2020].
83. H. B. Aguma et al., Mass distribution campaign of long-lasting insecticidal nets (LLINs) during the COVID-19 pandemic in Uganda: lessons learned. *Malar J* **22**, 310 [2023].
84. A. J. Tatem, J. Espey, Global population data is in crisis – here's why that matters. http://dx.doi.org/https://theconversation.com/global-population-data-is-in-crisis-heres-why-that-matters-251751 (March 26, 2025].
85. P. Deville et al., Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences* **111**, 15888-15893 [2014].
86. N. W. Ruktanonchai, C. W. Ruktanonchai, J. R. Floyd, A. J. Tatem, Using Google Location History data to quantify fine-scale human mobility. *International Journal of Health Geographics* **17**, 28 [2018].
87. M. Sinclair et al., Assessing the socio-demographic representativeness of mobile phone application data. *Applied Geography* **158**, 102997 [2023].
88. D. R. Leasure et al., Nowcasting Daily Population Displacement in Ukraine through Social Media Advertising Data. *Population and Development Review* **49**, 231-254 [2023].
89. G. Chi, G. J. Abel, D. Johnston, E. Giraudy, M. Bailey, Measuring global migration flows using online data. *Proceedings of the National Academy of Sciences* **122**, e2409418122 [2025].
90. Flowminder Foundation, R. Hosner, Z. Strain-Fajth, V. Lefebvre [2023] Using survey data to correct for representation biases in mobility indicators derived from mobile operator data to produce high-frequency estimates of population and internal migration. in *Netmob 2023, https://doi.org/10.5281/zenodo.8414414*.
91. H. R. Chamberlain et al., Building footprint data for countries in Africa: To what extent are existing data products comparable? *Computers, Environment and Urban Systems* **110**, 102104 [2024].
92. M. Minghini, S. Thabit Gonzalez, L. Gabrielli, Pan-European open building footprints: analysis and comparison in selected countries. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **XLVIII-4/W12-2024**, 97-103 [2024].
93. H. R. Chamberlain et al., Assessing the impact of building footprint dataset choice for health programme planning: a case study of indoor residual spraying (IRS) in Zambia. *International Journal of Health Geographics* **24**, 13 [2025].
94. C. Visée et al., Addressing bias in national population density models: Focusing on rural Senegal. *PLOS ONE* **19**, e0310809 [2024].
95. D. R. Thomson et al., Evaluating the Accuracy of Gridded Population Estimates in Slums: A Case Study in Nigeria and Kenya. *Urban Science* **5**, 48 [2021].
96. M. Owusu et al., Towards user-driven earth observation-based slum mapping. *Computers, Environment and Urban Systems* **89**, 101681 [2021].

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 December 2025 doi:10.20944/preprints202506.1625.v2

97. UNHCR [2024] Global Trends: Forced displacement in 2023. ed U. N. H. C. f. Refugees (United Nations High Commissioner for Refugees, Copenhagen, Denmark).
98. C. A. Dooley, W. C. Jochem, A. Sorichetta, A. N. Lazar, A. Tatem, Description of methods for South Sudan 2020 gridded population estimates from census projections adjusted for displacement, version 2.0. WorldPop, University of Southampton. doi: 10.5258/SOTON/WP00710. [2021].
99. J. A. Quinn et al., Humanitarian applications of machine learning with remote-sensing data: review and case study in refugee settlement mapping. *Philos Trans A Math Phys Eng Sci* **376** [2018].
100. E. Darin et al., Mapping refugee populations at high resolution by unlocking humanitarian administrative data. *Journal of International Humanitarian Action* **9**, 14 [2024].
101. C. Donegan, Y. Chun, A. E. Hughes, Bayesian estimation of spatial filters with Moran's eigenvectors and hierarchical shrinkage priors. *Spatial Statistics* **38**, 100450 [2020].
102. K. D. Mets, D. Armenteras, L. M. Dávalos, Spatial autocorrelation reduces model precision and predictive power in deforestation analyses. *Ecosphere* **8**, e01824 [2017].
103. S. Openshaw, The modifiable areal unit problem. *Concepts and techniques in modern geography* [1984].
104. J. Paige, G.-A. Fuglstad, A. Riebler, J. Wakefield, Spatial aggregation with respect to a population distribution: Impact on inference. *Spatial Statistics* **52**, 100714 [2022].
105. D. Manley, "Scale, Aggregation, and the Modifiable Areal Unit Problem" in Handbook of Regional Science, M. M. Fischer, P. Nijkamp, Eds. (Springer Berlin Heidelberg, Berlin, Heidelberg, 2014], 10.1007/978-3-642-23430-9_69, pp. 1157-1171.
106. M. O. Kimberley, M. S. Watt, D. Harrison, Characterising prediction error as a function of scale in spatial surfaces of tree productivity. *New Zealand Journal of Forestry Science* **47**, 19 [2017].
107. Project on Government Oversight [2023] Dollars and Demographics: How Census Data Shapes Federal Funding Distribution. (POGO: Washington DC).
108. T. Krieger, D. Meierrieks, Population size and the size of government. *European Journal of Political Economy* **61**, 101837 [2020].
109. C. Tuholske et al., Implications for Tracking SDG Indicator Metrics with Gridded Population Data. *Sustainability* **13**, 7329 [2021].
110. E. Darin, Demographic figures at risk in the digital era: Resisting commodification, reclaiming the common good
. *Big data & society* **Doi: 10.1177/20539517251403948** [2025].
111. A. M. van der Bles et al., Communicating uncertainty about facts, numbers and science. *Royal Society Open Science* **6**, 181870 [2019].
112. V. Kudakwashe Paul, H. Kate, G. Matthew, P. Helen, K. Lynda, Local ownership of health policy and systems research in low-income and middle-income countries: a missing element in the uptake debate. *BMJ Global Health* **4**, e001523 [2019].
113. B. Knittel et al., Critical barriers to sustainable capacity strengthening in global health: a systems perspective on development assistance. *Gates Open Res* **6**, 116 [2022].
114. B. Harrell-Bond, E. Voutira, M. Leopold, Counting the Refugees: Gifts, Givers, Patrons and Clients *. *Journal of Refugee Studies* **5**, 205-225 [1992].
115. C. N. Mayemba et al., A Short Survey of Human Mobility Prediction in Epidemic Modeling from Transformers to LLMs (ver 1, 25 April 2024]. *arXiv:2404.16921* [2024].
116. Institut National de la Statistique et de la Démographie du Burkina Faso, VOLUME I : EVALUATION DE LA QUALITE DES DONNEES, ETAT, STRUCTURE ET DYNAMIQUE DE LA POPULATION;

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 December 2025 doi:10.20944/preprints202506.1625.v2

https://www.insd.bf/fr/resultats
.

117. Institut de la Statistique du Mali [2023] Resultats Globaux du RGPH5, https://www.instat-mali.org/laravel-filemanager/files/shares/rgph/rapport-resultats-globaux-rgph5_rgph.pdf.
118. PNG National Statistical Office (Population Estimates 2021. https://www.nso.gov.pg/statistics/population/.
119. C. Ninrew [2023] S. Sudan population is 12.4 million - govt estimates. https://www.eyeradio.org/s-sudan-population-is-12-4-million-govt-estimates/.
120. WHO, UNICEF [2023] Geo-Enabled Microplanning Handbook: A product of the WHO-UNICEF COVAX GIS Working Group.
121. UNICEF [2025] Reach the Unreached - Geospatial modelling mapping methods. (https://github.com/unicef-drp/reach-the-unreached?tab=readme-ov-file).
122. E. Darin, D. R. Leasure, A. J. Tatem [2023] Statistical population modelling for census support. (United Nations Population Fund (UNFPA), Leverhulme Centre for Demographic Science, University of Oxford, and WorldPop, University of Southampton. https://wpgp.github.io/bottom-up-tutorial/, doi:10.5281/zenodo.7945266].
123. J. M. Espey, A. J. Tatem, D. R. Thomson, Disappearing people: A global demographic data crisis threatens public policy. *Science* **388**, 1277-1280 [2025].
124. A. Gutierrez [2024] ECLAC and UNFPA approach to model populations in Latin America and the Caribbean, IAOS-ISI, Mexico City, https://www.inegi.org.mx/eventos/2024/iaos-isi/doc/34.pdf.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.