Review

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: 19 June 2025

doi: 10.20944/preprints202506.1625.v1

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

Review

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

Attila N. Lazar [1,*], Gianluca Boo [1], Heather R. Chamberlain [1], Chibuzor Christopher Nnanatu [1], Edith Darin [2], Douglas R. Leasure [2], Ortis Yankey [1], Assane Gadiaga [1], Sabrina Juran [3], Luis de la Rua [3], Jessica Espey [1] and Andrew J. Tatem [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: author: a.lazar@soton.ac.uk

Abstract

Population data at small area scales are essential for effective decision-making, affecting public health, disaster response, and resource allocation, amongst other issues. While national censuses remain the foundational source of population data, they may face challenges such as high cost, infrequency, and coverage gaps, which can hinder timely data availability. Geospatial statistical approaches to address these challenges using limited microcensus surveys have been demonstrated, and the field has advanced substantially in recent years, with significant developments in both data sources and modelling methodologies. New approaches now enable the use of routine health intervention campaign data, improved satellite-derived settlement mapping, and novel Bayesian frameworks to produce high-quality small area population estimates where traditional enumeration is difficult or outdated. Various countries are increasingly applying these techniques to support census operations, health program planning, and humanitarian response. This Perspective reviews recent advances in 'bottom-up' population mapping approaches, highlighting innovations in input data, statistical methods, and validation techniques. We examine ongoing challenges, including partial observation of buildings under forest canopies, population displacement detection, and institutional adoption. Finally, we discuss emerging opportunities to enhance these approaches through better integration with traditional data systems, capacity strengthening and co-production with national institutions, often facilitated by UN organisations, such as United Nations Population Fund (UNFPA), and the use of novel data streams to develop more timely, accurate, and useful small area population estimates for planning and humanitarian purposes.

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

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 [1–5]. Vital registries (records of births, marriages and deaths) and the population and housing census — hereafter referred to as 'census' — are traditionally the most detailed demographic data sources. However, both face practical challenges — such as infrequent collection and resource demands — that can limit their capacity to provide timely, granular data. Censuses are typically conducted only every 10 years or more, with results often made publicly available only at coarse spatial resolution to protect the privacy of individuals. Moreover, conducting a census requires substantial human and financial resources, and ensuring full coverage — particularly in remote or marginalized areas — can be complex and resource-intensive [6,7]. Data-rich countries may turn to population registers that are updated with

 doi:10.20944/preprints202506.1625.v1

administrative data collected at service points, but these require robust administrative systems which are lacking in most low and lower-middle income countries [8]. Urbanization, conflicts, migration and climate change are leading to rapid changes in population distributions and demographics, and thus the need for regularly updated demographic data is growing, prompting the exploration and development of alternative approaches to estimate and map populations at small area scales.

The 2030 Agenda for Sustainable Development [9] underscores the urgent need for high-quality, timely, and disaggregated population data to monitor progress toward the Sustainable Development Goals (SDGs). Many SDG indicators—such as those on poverty, health, education, and gender equality—require population data as a denominator, disaggregated by sex, age, geographic location, and other dimensions of inequality. However, millions of people remain statistically invisible due to outdated censuses, undercounts, or a lack of data systems capable of reaching marginalized populations. In this context, developing complementary population estimation methods becomes not only a technical necessity, but also a human rights imperative. Enhancing statistical visibility at small area scales is fundamental 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. These models can serve to evaluate and adjust for coverage issues, inform 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 of demographic intelligence that can enhance the accuracy, equity, and responsiveness of statistical systems — particularly in settings affected by conflict, environmental hazards, or rapid population shifts. 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 recent, small area demographic estimates in countries where demographic data sources from censuses or registries were outdated, incomplete or unavailable. The 'bottom-up' approach aimed at generating high-resolution population estimates using datasets from small area population surveys — referred to as 'microcensus' data — and high-resolution geospatial covariates derived from satellite imagery and other sources. The envisaged statistical modelling methods aimed to estimate population counts and demographic characteristics at high spatial resolution across unsurveyed areas, together with measures of 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 Wardrop *et al.* was published, conflicts, resource limitations, 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]. Partly as a result, there has been increasing demand for alternative approaches for the provision of small area population estimates. Fuelled by this need, together with increased data availability and recent statistical modelling developments, the bottom-up approach has seen substantial advances in the past five years. These complementary estimation methods are also increasingly being explored in combination with traditional censuses to enhance coverage, support pre-enumeration planning, and improve responsiveness to dynamic demographic conditions.

Here we review the current state of census-independent bottom-up models for the production of small area population estimates. We explore their application to challenges concerning the estimation of population distributions and demographics, particularly in resource poor settings. We present an overview of the modelling frameworks that have been developed to date, the range of input demographic and geospatial datasets they utilise, and approaches to the validation of outputs. Finally, we discuss current challenges and future opportunities facing the field.

 doi:10.20944/preprints202506.1625.v1

Approaches to Census-Independent Small Area Population Estimation

New data types and sources, as well as demographic challenges, which have required novel approaches to spatial coverage, data quality and human mobility issues, have accelerated development of small area population methods in recent years. Figure 1 illustrates these different approaches and data used for different scenarios – filling gaps in an existing census, providing new estimates in the absence of a recent census, addressing undercounts, and supporting census preparations, among other applications. In the following sections we review each component of Figure 1, providing examples throughout.

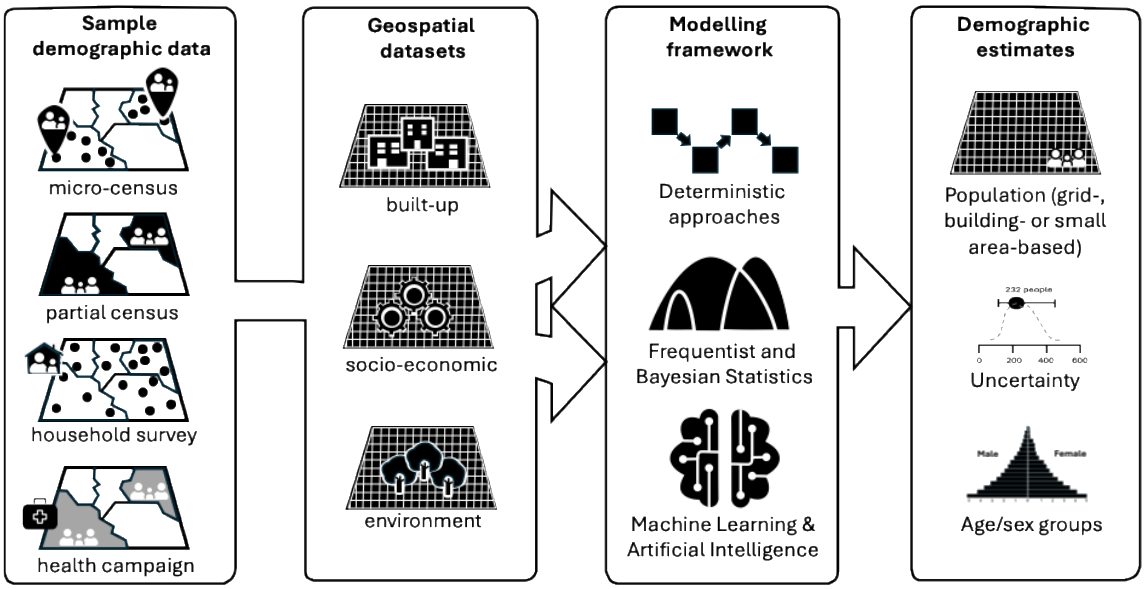

Figure 1. Components of the bottom-up approach to population estimation. Different sample demographic datasets are combined with ancillary geospatial datasets covering the entire study area in bespoke modelling frameworks to estimate population distributions and demographics.

Demographic Datasets

Bottom-up demographic estimation methods rely on a training set of population data to predict demographic attributes, such as population counts and age and sex structures, across unsampled locations. In recent years, digitally collected data and GPS devices have become standard for survey teams, both for the census and household surveys and for those delivering health campaigns [11]. This has resulted in significant improvements in attribute accuracy and geolocation, which makes for more straightforward integration of these datasets into modelling processes. These demographic datasets can have different levels of spatial completeness (Figure 1), from purposefully designed microcensus surveys to partial censuses, household surveys and specific sectoral interventions, like health campaigns.

Wardrop *et al.* (ibid) presented the use of microcensus surveys for population modelling in the absence of census observations, offering the advantage of rapid data collection (10, 14-17). Microcensus surveys typically enumerate a random, representative sample of locations, fully covering a small area; around three hectares each of a single settlement type. The sample locations must be geographically defined, with clear spatial locations and extents. These bespoke surveys can also accommodate more complex sampling designs, such as stratified and probabilistic samples, and usually involve limited sample sizes [18]. Like traditional censuses, they still require considerable human and financial resources — albeit a fraction of a full census — and face traditional challenges related to inaccessibility and insecurity.

Since these early applications, a wider set of models have been developed to deal with situations where recent enumeration data already exists. Among those discussed here, the most complete form

 doi:10.20944/preprints202506.1625.v1

of demographic sample can be derived from partial census enumerations which provide observations across entire regions, enabling the filling of geographic gaps in census data collection [19,20]. There are also examples of using census cartography data for population modelling to support census planning processes [21]. These data consist of demographic attributes associated with household coordinates or enumeration areas, but do not provide a random sample of the population due to accessibility limitations. Nevertheless, they generally have very large sample sizes. In addition to gaps in enumeration due to conflict or remoteness [7,22], there exist growing numbers of examples of undercounts in both rural and urban settings, including associated with high rise buildings, gated communities [23] and among certain demographic groups [6,24,25]. Population modelling approaches can help complement census data by identifying and addressing such gaps.

Another source of georeferenced enumeration data can be found through the process of undertaking national household surveys to monitor demographic, socio-economic and health characteristics (e.g. the Demographic and Health Surveys, Household Income and Expenditure Surveys, Multiple Indicator Cluster Surveys). These surveys are designed to be nationally representative by fully enumerating a stratified random sample of the population at selected locations. Repurposing the household listing observations for population modelling is a novel way of conducting population estimation, potentially enabling frequent and low-cost model updates [26–28]. Such datasets typically include geographic data on the cluster locations and, in some cases, even the coordinates of individual households and the sampling weights [29].

A different type of sample population enumeration data can be obtained from activities that aim to deliver a commodity or service to all households, in a similar way to the census, aiming to reach every household in a given area. For example, campaigns to distribute insecticide treated nets (ITNs) as part of malaria control measures, often adopt a house-to-house distribution model, with demographic attributes relevant to the campaign recorded at the household level. Previous applications have used health campaign data (Figure 1), but the potential is not limited to data collected for the delivery of health interventions. Such semi-regularly collected campaign data can enhance sample coverage and frequency, offering a practical complement to censuses or microcensus surveys [30,31]. While not originally intended for robust statistical analysis, these data may contain collection biases that require careful quality assessment and adjustment. For example, random (i.e. missed and ghost households) and systematic (i.e. inflated household sizes) biases likely to be more prevalent than in well-planned surveys. In addition, the lack of clarity in the designs of these operations could also introduce estimation biases.

Geospatial Datasets

The development of bottom-up population modelling approaches has been closely linked with the availability and growth of geospatial datasets derived from satellite imagery [14,32] and other geospatial data sources. All modelling approaches discussed here leverage ancillary geospatial datasets that describe the human landscape (Figure 1) and thus correlate with the geographic distribution of demographic attributes across an area of interest. To be able to predict into gap areas more accurately, geospatial datasets must have complete coverage of the area of interest. Such ancillary datasets can be prepared as small area summaries at administrative or census unit level, but they are often utilised in a regular gridded format, with the spatial resolution matching the desired estimated population output – e.g. approximately 10m [33] or 100m [34].

Wardrop *et al.* highlighted advances in geospatial data, citing satellite-derived building area, land use, counts of dwelling units, spectral radiance and other socio-economic and physical characteristics. Today, geospatial datasets representing factors relating to the human landscape have advanced substantially due to increasing computing power, a wider range of data sources, and advances in AI algorithms that can extract relevant features. These have greatly improved the quality, spatial detail and regularity of relevant available data. As a result, the availability of near-global open geospatial datasets on factors relating to population distributions have become commonplace, driving the production of more detailed, more accurate and more recent small area estimates.

 doi:10.20944/preprints202506.1625.v1

The most important geospatial datasets for population modelling describe built-up areas [35]. Today, near-global building footprint datasets exist going beyond simply defining the extents of major settlements, and instead mapping individual buildings (e.g. outline, height, roof type) with unprecedented accuracy [36–39]. Building footprints in population models are typically aggregated to provide summaries of the built environment at the scale of the model output, for instance 100x100m grid cells. These aggregations include, for example, building footprint counts and area, morphology (e.g., building footprint perimeter and volume), and other locational characteristics (e.g. building orientation, neighbourhood types) [40–44]. Some datasets also exist that stratify building footprints into residential/non-residential classes [33]. Road network data can also be extracted from satellite imagery [45] and converted into population model inputs using similar metrics described above. Finally, satellite-derived temporally explicit building datasets are now becoming available with annual timesteps, such as the Google 2.5D building dataset, enabling spatio-temporal assessment and better temporal alignment of covariates with ground enumeration samples [46].

Other geospatial datasets can capture additional valuable covariates. Leveraging the growing availability of volunteered geographic information (VGI), OpenStreetMap data for example [47], provides detailed geographic locations for different point of interests (e.g., health and education facilities) associated with population distributions [48]. Regional and global datasets mapping health and education facility locations [49], armed conflicts [50], and building damage [51–53] can also provide useful ancillary covariate data. These point location datasets can be converted into gridded or other formats using focal windows of different sizes to count them and thus mapping potential push-pull effects of different features on population distributions. Additionally, digital trace data from mobile devices have been used as indicators of the presence and densities of humans, including call detail records [54,55] and smartphone app data [56].

The presence and distribution of humans and their settlements have always been driven by or associated with environmental factors, such as topography, climate, water bodies and land cover [35], and an increasing number of high spatial resolution datasets with global coverage and frequent updates are now available. Similarly, lights visible from nighttime satellite imagery are often used as a potential indicator of human presence, indicating areas with electrification and fires [57], while open near real-time global land cover mapping datasets are now also routinely produced [58].

Modelling Frameworks

Wardrop *et al.* suggested simple regression and geostatistical approaches as potential bottom-up model frameworks. Since then, many different frameworks for bottom-up population estimation have been developed and implemented. These frameworks are specifically designed to accommodate the available input data and span from deterministic approaches, through frequentist and Bayesian statistical modelling frameworks, to machine learning and artificial intelligence (AI) techniques (Figure 1). In essence, these approaches assess relationships between the demographic and geospatial covariate datasets outlined above that are observed at sampled locations (e.g., administrative units, census enumeration areas, microcensus clusters, household locations) to estimate demographic attributes at both sampled and unsampled locations using the geospatial covariates as predictors (Figure 2). The simplest form of small area population involves mapping constant values of population densities across the study area. Through estimates of the number of residential buildings, households per building, and the average household size, population numbers can be estimated deterministically at small area scales. While such approaches produce highly uncertain outputs that do not account for local variations, recent advances in the availability of datasets like building footprint maps, building heights and settlement types are improving the precision of outputs [44,59].

Leasure *et al.* [16] used Bayesian hierarchical models to estimate population counts as a Poisson process with exact counts derived from the product of settlement datasets and estimated population densities. In a hierarchical modelling framework, population densities are generally modelled as a lognormal process to capture variabilities in the expected population densities. Hierarchical models draw information from all scales, such as coarse administrative units (e.g., regions or districts) or

 doi:10.20944/preprints202506.1625.v1

functional areas (e.g., settlement classes), to define different model parameters, including random intercepts, random effects, and variance components [16,28]. The hierarchical random intercept is a key component of model structure that helps to account for the complex sampling designs used in the data collection as well as the geographic patterns in population density. Moreover, by using random effects, key information is shared among various spatial scales, thereby, leveraging observations from neighbouring clusters to predict population densities across clusters with no observations. More recent hierarchical model applications utilise satellite image-derived building footprints enabling more accurate population distribution prediction [17,19,26] and some even extended to estimating building count distributions [18,60]. These approaches train the model parameters using the observed population data and then predict into other unsampled areas based on the trained parameters in line with the model assumptions.

Bayesian geostatistical models have also been developed to explicitly account for spatial autocorrelation not captured by covariates. Geostatistical models employ the use of distance-dependent covariance matrices to describe the spatial autocorrelation between observations. This includes a two-stage Poisson-Gamma geostatistical regression [61]. Similar to those described for hierarchical approaches, these models also assume that the population density is related to a set of geospatial covariates through a linear predictor, and then produce estimates of population as a function of the product of the predicted population density and the corresponding building counts. However, the Poisson-Gamma approach utilises Gamma probability density instead the lognormal to model the population density in a more flexible manner whilst utilising the integrated nested Laplace approximations (INLA) [62] techniques in conjunction with stochastic partial differential equation (SPDE) [63] approaches. The use of the INLA approach enables Bayesian statistical inference based on the posterior marginal distribution, thereby ensuring both accuracy and computational speed. An extra layer of computation speed is achieved via the use of the SPDE approach which specifies spatial autocorrelation using a triangulation (or mesh) of the entire continuous spatial domain so that the usual computationally expensive dense matrix of the Gaussian Process is approximated via a sparse matrix, represented through the Gaussian Markov Random Field. These geostatistical models have been extended to estimate population counts where no building footprints were available and the indicators of human settlements were only partially observed [30].

An important development is the use of machine learning and AI, including deep learning, to estimate intercensal population counts at high spatial resolution directly from satellite imagery. Landscan HD, for example, uses a data fusion approach to merge relevant land use and building layers in a deep learning framework to generate gridded population estimates [64]. Another approach is to use a convolutional neural network with visual geometry group architecture for deep learning directly from a concatenation of low-resolution Landsat-7/8, Sentinel-1 and night-time light satellite images [65,66]. Key advantages of these methods are the frequent and freely accessible satellite image updates, enabling rapid population map updates and scalability in data scarce regions. The use of very high-resolution satellite images such as the Maxar VIVID 2.0 in conjunction with a small sample of ground truth dataset, appears promising to help with the identification of inhabited areas and non-residential buildings [67].

Most of the types of population models outlined above estimate the total population, but development and humanitarian interventions often require specific age and sex information (Figure 2). The most widespread method for age and sex disaggregation is using observed or projected subnational population pyramids to deterministically disaggregate the total population estimates [68]. There are however, hierarchical model applications where the proportion of under-five population is estimated at 1km resolution using a Bayesian spatio-temporal model [69], or the entire age and sex structure of each simulation unit is statistically estimated using a Dirichlet-multinomial process [17]. A recent methodology developed uses a flexible multi-stage Bayesian statistical modelling approach based on ‘sequential’ Binomial probability mass functions to produce disaggregated structured estimates of population counts and population proportions at both administrative units and high-resolution grid cell levels [70]. The methodology can be used to

 doi:10.20944/preprints202506.1625.v1

estimate age/sex classes, and other socio-economic groups such as ethnicity, occupation, education. The key input datasets are the demographic structures (e.g. age, sex, ethnicity, etc) which can come from various sources such as census/microcensus, household surveys (e.g. DHS), and administrative records. The methodology is implemented via INLA-SPDE approach, thus capable of rapidly producing structured estimates of population and population proportions at small area scales including at locations without observations, along with the corresponding estimates of uncertainties. In addition to age/sex information, field campaigns often require household level information that can also be modelled statistically [27].

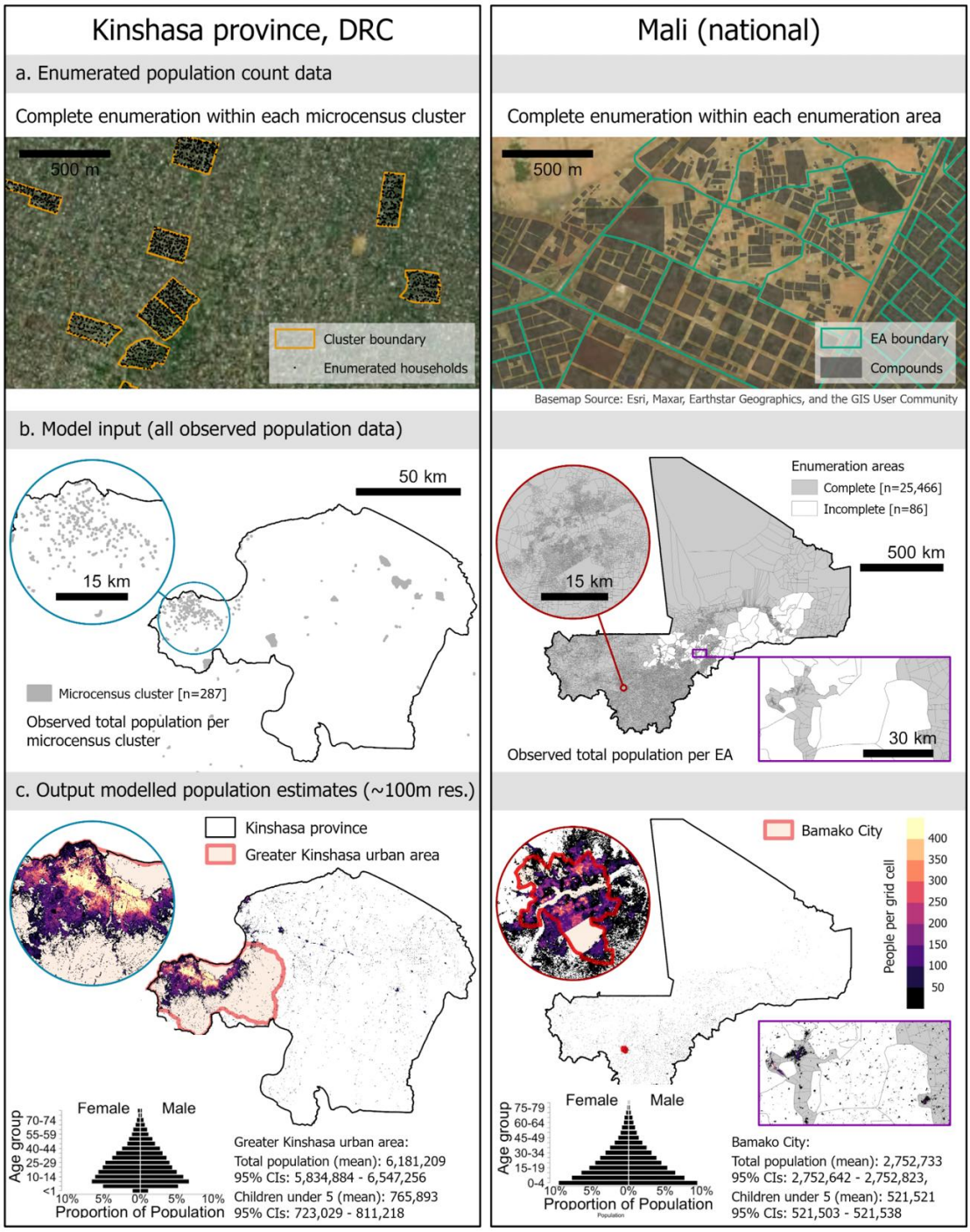

Figure 2. Example population model applications: field observations (panel a), model input (panel b), settlement constrained model outputs with age/sex and uncertainty information (panel c). Data sources: DRC microcensus [71], DRC population estimates [17], Mali census cartography [72], Mali population estimates [21].

 doi:10.20944/preprints202506.1625.v1

Outputs and Validation

Small area estimates produced using the modelling approaches outlined above are typically output at the grid square scale (Figure 2). This format is advantageous because it provides a continuous surface to support visualisation and understanding of small-scale landscape variation. Moreover, the grid format enables flexible aggregation and summation of population estimates to different decision-making units, such as districts, wards or health zones, or by features of interest, such as urban extents, flooded areas, or hurricane tracks. Moreover, the format enables ease of integration with other geolocated datasets, such as the locations of health facilities, conflicts, schools or polling booths. Gridded population estimates have been used to calculate social-distancing indicators [73], map out-of-school children [74,75], support disaster response [76,77], create survey sampling frames [78,79], and prepare for vaccination campaigns [80]. Accuracy is however challenging to assess at the grid level unless geolocated household data - with GPS accuracy better than the grid resolution - is available. Gridded estimates require technical expertise to aggregate to more useful spatial units for field operations planning such as settlement or admin boundaries.

Population estimates are often reported as a single 'best' estimate of residential population for a specific timepoint, however, they cannot capture accurately the full complexity of our world. Uncertainty in population estimates varies place to place, originating from both the observations used in their construction and from the model design and assumptions. Observation uncertainties include non-representativeness (e.g. too small sample size), sampling bias that under-represents some populations (e.g. slum and informal settlement dwellers), and measurement error (e.g. missed populations, incorrectly recorded data). Partial, unobserved or outdated observations of the settled areas have a direct effect on the accuracy of the estimates and thus on their usability. However, it is difficult to translate the estimated uncertainty across spatial scales.

Using a Bayesian method enables the consideration of these errors during population estimation and provides the most likely estimates with quantified uncertainties. Model designs, guided by the local context and the accessible observations can be tested in a simulation study [29,30], assessing the importance of various inputs and the sensitivity of the model structure. Uncertainties are then quantified for the final model design through the posterior distribution of model parameter values. These posterior distributions are used to calculate metrics such as the mean value of the predicted population in an area, the standard deviation, and 95% credible intervals. In addition to communicating varying levels of uncertainty in modelled estimates, the use of these measures has the potential to also help make more informed decisions. For example, (i) using the upper confidence interval for health intervention campaign planning to ensure confidence that enough resources are acquired for each delivery area (ii) to identify where to conduct additional data collection to improve population estimates, or (iii) mapping model uncertainty to identify important predictors that may be missing from the modelling process that could aid better prediction in areas with high uncertainty.

Model validation is vital to ensure that model outputs are accurate and can be trusted by data users and decision makers. This is generally done though model diagnostics, goodness-of-fit assessment and external validation against an independent set of observations. Model diagnostics include checking convergence issues, posterior predictions, grid cell predictions and undertaking cross-validation to check the robustness of the model design, as well as uncertainty visualisation (e.g. 16). Goodness-of-fit assessments compare the outputs against the training data used by calculating metrics such as root mean square error and correlation coefficients (e.g. 61).

Out-of-sample cross validation is considered the gold standard way of validating population models [81,82], though data triangulation with other sources is often important for model validation. For example, aggregating the outputs to various administrative units enables cross-checking the results against population predictions and other external data sources. Modelled estimates are however often constructed precisely because of the lack of reliable ground enumeration data, making such comparisons rare and challenging to undertake. Even if an independent, reliable data source exists, there are also always differences that make comparisons difficult and potentially uninformative, such as the time difference between the reference year of the modelled and the

 doi:10.20944/preprints202506.1625.v1

independent datasets, the survey design and/or methodology, the population groups being counted and time of year of enumeration.

There have been very few bottom-up model validations using extensive small area enumeration data published so far. Darin *et al.* [60] used the 2018 Colombian census to test the accuracy of two Bayesian model designs against a machine learning methodology at enumeration area and municipality scales, while also varying the use of six different input settlement maps. They sampled from the census data to create an artificial data set, so the sample and the validation data were directly comparable and thus avoiding the issues described above relating to external data triangulation. They highlighted the importance of building footprints as the best ancillary information source for settlements. For unbiased, complete observations, the machine learning approach was the best performing method, whereas coarse or biased observations for sparsely populated regions were best explained by Bayesian methods because of their ability to correct biases. They emphasised that uncertainties varied vastly across different landscapes and at different spatial scales. For example, the median inaccuracy was 32%, but at finer spatial scales it rose to 148%. They highlighted that there is a limit to the information that can be derived from building footprint data, with no apparent increase in accuracy following an increase in sample size.

Chamberlain *et al.* [83] compared the 2019 census-independent modelled population estimates for Zambia [26] with the population counts from the 2022 national census at province, district and ward-level. Their analysis showed a strong correlation between the modelled population estimates and the census-enumerated counts (r=0.98 at district-level and r=0.95 at ward-level). In province-level comparisons, census counts were within the 95% credible interval (CI) of the modelled estimates for five out of ten provinces, with census counts less than the lower 95% CI for the other five provinces, indicating some degree of positive bias in the modelled estimates. At district- and ward-level, census counts were within the 95% CIs for 64% of districts, and 52% of wards. The difference between the modelled estimates and enumerated census counts is likely in part due to some degree of over-estimation associated with non-residential buildings, however the temporal difference between the modelled estimates and the census also complicated the comparison.

Breuer *et al.* [84] compared gridded population estimates with population counts for a sample of slum areas in eight major cities of the Global South and found that slum populations were significantly underestimated with large spatial variation: only estimating 48 percent on average (range of 8-147 percent) of the true population. Thomson *et al.* [85,86] similarly found that population models only estimated a fraction (11-39 percent) of slum residents in Nigeria, Kenya and Namibia, potentially omitting 0.75 to 1.5 people for every estimated person. They concluded that this underestimation is due to (i) the 100m resolution being too coarse to capture the vastly different demographic and building patterns in slum areas, and (ii) insufficient observations in slum areas for model training.

Challenges and Future Directions

The significant developments in data sources and methods described above have enabled the generation of geospatial modelled population estimates in a range of countries and contexts. These advances have addressed many challenges outlined by Wardrop *et al.* [2018], but some challenges remain, and new ones have emerged as applications involve increasingly diverse sources of data. These challenges can be categorized into data input challenges, methodological considerations, and implementation barriers (Figure 3).

 doi:10.20944/preprints202506.1625.v1

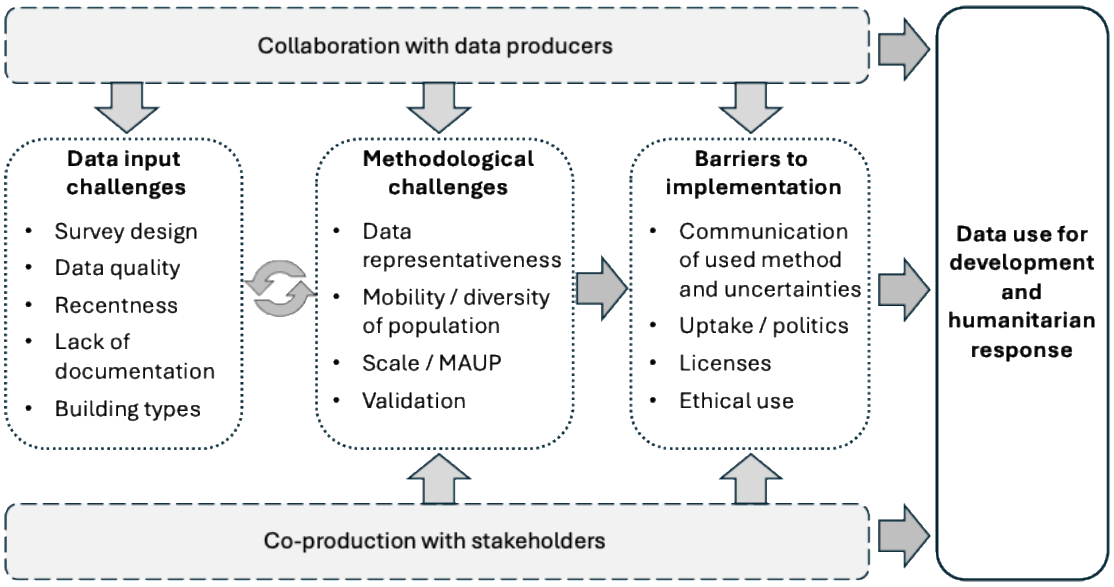

Figure 3. Challenges (white boxes) and opportunities (grey boxes) towards sustainable data production and use. MAUP = Modifiable Areal Unit Problem.

Data Input Challenges

Challenges persist regarding input enumeration data quality and suitability for training and validating models. Field data collection has benefited from GPS-enabled devices that facilitate navigation and location recording, but practical limitations remain, including battery life, internet access in remote areas, and safety concerns for enumerators in insecure regions [87,88]. Community distrust in data collection efforts can result in non-response or incomplete coverage, particularly among marginalized populations [89]. Enumeration data collected through questionnaires linked to the distribution of health interventions, such as bednets or vaccines, can result in biases in numbers upwards or downwards, depending on incentives in receiving these interventions [90]. Though field data collection software enables near real-time quality checks, data quality ultimately depends on survey design, field protocols, and team training [88]. This underscores the importance of thorough documentation of survey methodologies, particularly as financial resources for field data collection become increasingly constrained [91].

Building footprint and settlement datasets have seen substantial improvements in global coverage and public availability, but challenges remain around temporal and spatial coverage due to cloud cover, tree canopy obstruction, and certain areas not being 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 limitations. Moreover, substantial variations exist between datasets [92], likely translating into significant differences in population predictions in many areas. Capturing rural populations needs further attention [93] and mapping urban slums also presents particular difficulties [85]. Thomson *et al.* demonstrated that slum populations are significantly underestimated by all types of gridded population datasets, primarily due to a lack of recent and reliable population enumeration data from such settings, as well as challenges in mapping the structures and extents of such areas consistently. Multiple efforts are ongoing to improve the mapping of structures and populations in these areas by combining Earth observation data, statistical modelling, and community engagement [84,86,94].

Mobile populations present another 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 static

 doi:10.20944/preprints202506.1625.v1

population models. 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 current approaches.

Methodological Considerations

Novel data sources offer promising avenues for increasing the temporal-resolution of population estimates and thus increased accuracy in context of mass movement. Digital trace data from mobile phone call detail records [54], smartphone location history data [99], mobile applications [100], and social media platforms [101,102] have demonstrated potential for measuring near real-time population distributions and movements. 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 [103], but access challenges due to private ownership persist.

Population model design also raises other important methodological considerations. Spatial autocorrelation of inputs requires particular attention for accurate uncertainty estimation, as it can alter observation variance and increase sensitivity to certain covariates [104,105]. This is particularly important when there are spatial dependences in the data that cannot be accounted for by readily available covariates, i.e. when spatial autocorrelation remains in the residuals of a fully specified model.

The modifiable areal unit problem (MAUP) - error arising from the choice of spatial units used to summarise point-based observations - represents another crucial consideration, particularly when the spatial scales of observations differ from pixel-level estimation units [106]. This mismatch can reduce explained variance, increase residual spatial autocorrelation, and potentially lead to erroneous conclusions [107,108]. MAUP requires careful harmonization of all scales—observations, human populations, and statistical model predictions [109]. For population models that rely heavily on satellite-derived information, even the gridding process itself may impose artificial discretization on the underlying continuous landscape [110]. Simulation studies can help determine appropriate scale choices [111], while spatial aggregation models that incorporate both data response and sampling frame components can better account for population distribution uncertainties [108].

Implementation Barriers and Opportunities

While a growing number of applications show how small area modelled population estimates provide valuable support for planning and decision-making during intercensal periods, they often face barriers to acceptance by governments, who may be reluctant to use modelled data, geospatial covariates, or to collaborate with external partners to produce official population estimates. Population statistics have significant implications for resource allocation [112–114], political representation [115], and economic and SDG indicators [116], making their official endorsement politically sensitive. To facilitate greater governmental uptake, several approaches show promise: improved model validation, enhanced communication of model methods, results and associated uncertainties [117] and crucially, local ownership of the modelling process [118]. 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 national statistical offices, and is accompanied by appropriate capacity strengthening for sustainable data generation and use [119,120]. Although geographic information systems are now commonly used to analyse and display population data, spatial statistical expertise is often limited in government institutions across low-income settings. Medium- to long-term investment is needed to identify in-country champions, develop institutional

 doi:10.20944/preprints202506.1625.v1

skills, and strengthen national university programs to cultivate the next generation of geospatial experts.

Beyond the uptake of model results by various stakeholders, there are other important barriers for population modelling, such as data licenses. Various licence types are available for public data products (e.g. Creative Commons Attribution, Open Data Commons Open Database License). Some of them are fully open, some of them are restricted one way or another. Some geospatial data products have strict restrictions on combination with similar data products and have even restrictions on data dissemination of the derived outputs. Others, such as the OSM license, only require the final product to be made publicly available. The licence of the available data must therefore be considered during model design and publishing.

Another concern is that small area estimates, especially high-resolution population estimates, might expose vulnerable groups to unwanted surveillance or security risks and might impact equitable resource allocations [121,122]. Ensuring responsible use of modern technologies and publication and use of population data are crucial. Good practice includes seeking approval from (and if necessary, establishing) independent ethics boards of the data provider, government and/or universities, de-identification of observations, and publishing only aggregated statistical estimates and not the original observations. At the international level, the UN Statistical Commission should consider developing guidance for countries utilising these approaches on effective practice and good governance. In doing so it would empower countries to innovate and try alternative methodologies to fill crucial gaps, whilst showing how this can be done safely and sustainably.

Conclusions

Reliable, recent, and detailed small area population data are increasingly required for effective decision-making, development planning, Sustainable Development Goal reporting, operational campaign planning, and humanitarian response. Since Wardrop *et al.* [10] presented their perspectives on spatially disaggregated population estimates, significant advances have been made in both data sources and methodological approaches. The emergence of building footprint datasets with near-global coverage, applications of Bayesian hierarchical modelling frameworks to handle complex data integration, and development of methods to quantify and communicate uncertainty have substantially improved the quality and reliability of small area population estimates in the absence of census data.

These advances have enabled practical applications across diverse contexts. In countries including Burkina Faso [123], Mali [124], Papua New Guinea [125], and South Sudan [126], modelled population estimates have been adopted by national statistical offices for country planning activities. In Colombia, the national statistical office has developed capacity to implement these methods themselves, using them to address census enumeration gaps [20]. UN agencies including UNFPA and UNICEF (United Nations Children's Fund) are actively promoting and applying these approaches [127–129] and computer algorithms [130,131] to support field operations and data sustainability.

Despite these successes, challenges remain in capturing urban diversity, informal settlements, mobile populations, and populations in areas with persistent cloud cover or heavy vegetation. Continuing methodological advances and stronger partnerships between data producers, data scientists, and stakeholders are needed to further enhance the utility and acceptance of these approaches. Increasing transparency in data products, promoting collaboration and co-production, and investing in capacity strengthening would significantly enhance uptake and sustainability.

As demographic data gaps increase [91], the complementary role of small area population estimation approaches will likely expand. Their integration with traditional demographic data systems offers a path toward more responsive and cost-efficient statistics, particularly in contexts of census disruption or under-coverage. These methods can play a crucial role in demographic reconciliation, helping to detect and adjust for omissions or overcounts and providing robust, disaggregated data for planning and monitoring. Looking forward, the field is poised to benefit from

 doi:10.20944/preprints202506.1625.v1

increasing availability of digital trace data, advances in AI for building characterization, and improvements in statistical methods for data integration and uncertainty quantification.

The elaboration of such models, however, requires more than technical expertise. It depends on the sustained collaboration between national statistical offices, academic institutions, and international partners. In Latin America and the Caribbean, a collaborative partnership model involving UNFPA, ECLAC (Economic Commission for Latin America and the Caribbean), and national statistical offices has proven effective [132]. These partnerships leverage local contextual knowledge, statistical leadership, and external modelling capacity to co-develop robust estimates. In the Caribbean, coordination with regional statistical entities such as CARICOM (Caribbean Community) has also emerged as a key enabler for institutionalizing these approaches.

To ensure the long-term utility and trust in these estimates, methods and results must be transparently documented, presented with clear uncertainty and goodness of fit metrics, and integrated into national processes such as census reconciliation and population projections. With proper institutionalization and regional coordination, small area estimation methods can become a strategic asset for achieving data-driven, inclusive development.

Acknowledgments: The authors would like to acknowledge that the work summarised in this paper is the result of many years of research and collaboration, with a wide variety of partners including UNFPA, UNICEF, The Gates Foundation, GAVI, GRID3, the UK Foreign, Commonwealth and Development Office. Particular thanks to the many national governments, and their statistical and data offices, who have worked with us to develop and improve upon many of the methodologies here described.

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).
2. J. Bryant (2021) Digital mapping and inclusion in humanitarian response. in *HPG working paper. London: ODI (https://odi.org/en/publications/digital-mapping-and-inclusion-in-humanitarian-response).*
3. 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).
4. P. G. Greenough, E. L. Nelson, Beyond mapping: a case for geospatial analytics in humanitarian health. *Conflict and Health* **13**, 50 (2019).
5. 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).
6. 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).
7. S. Randall, Where have all the nomads gone? Fifty years of statistical and demographic invisibilities of African mobile pastoralists. *Pastoralism* **5**, 22 (2015).
8. UNECE, *Guidelines on the Use of Registers and Administrative Data for Population and Housing Censuses* (Geneva, 2018).
9. UN, *The Sustainable Development Goals Report* (United Nations, ISBN: 978-92-1-003135-6, New York, USA, 2024).
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/events-details/UNSC56-population-housing-census-5Mar2025/).
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* **https://verixiv.org/articles/2-8/v1** (under review).
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. Statistics South Africa (2022) Post Enumeration Survey - Statistical Release P0301.5. (Pretoriam, Stats SA).
25. 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 (VID), Vienna).
26. 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.
27. 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/.
28. 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).
29. 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.
30. C. C. Nnanatu *et al.*, Estimating small area population from health intervention campaign surveys and partially observed settlement data. *Nature Communications* **16**, 4951 (2025).
31. C. Nnanatu *et al.*, Bottom-up gridded population estimates for Maniema province in the Democratic Republic of Congo (2022), version 4.1. https://dx.doi.org/10.5258/SOTON/WP00773
32. S. Leyk *et al.*, The spatial allocation of population: a review of large-scale gridded population data products and their fitness for use. *Earth Syst. Sci. Data* **11**, 1385-1409 (2019).
33. P. M., P. P., GHS-BUILT-S R2023A - GHS built-up surface grid, derived from Sentinel2 composite and Landsat, multitemporal (1975-2030).
34. 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.

35. 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).
36. W. Sirko *et al.* (2021) Continental-scale building detection from high resolution satellite imagery. arXiv:2107.12283.
37. Microsoft, Worldwide building footprints derived from satellite imagery (GitHub Repository); https://github.com/microsoft/GlobalMLBuildingFootprints. https://github.com/microsoft/GlobalMLBuildingFootprints.
38. Ecopia, Global Feature Extraction: Building footprints; https://www.ecopiatech.com/products/global-feature-extraction.
39. 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).
40. 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).
41. 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).
42. 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).
43. 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).
44. 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).
45. Microsoft, Road detections from Microsoft Maps aerial imagery. https://github.com/microsoft/RoadDetections?tab=readme-ov-file.
46. W. Sirko *et al.*, High-Resolution Building and Road Detection from Sentinel-2. arXiv:2310.11622, https://sites.research.google/gr/open-buildings/temporal/.
47. OSM, OpenStreetMap. https://www.openstreetmap.org/#map=5/54.91/-3.43.
48. 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).
49. GRID3, GRID3 Data Hub. https://data.grid3.org/.
50. ACLED, ACLED Data. https://acleddata.com/.
51. 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).
52. S. Wiguna, B. Adriano, E. Mas, S. Koshimura, Evaluation of Deep Learning Models for Building Damage Mapping in Emergency Response Settings. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **17**, 5651-5667 (2024).
53. 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).
54. P. Deville *et al.*, Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences* **111**, 15888-15893 (2014).
55. L. Dong *et al.*, Defining a city — delineating urban areas using cell-phone data. *Nature Cities* **1**, 117-125 (2024).
56. N. N. Patel *et al.*, Improving Large Area Population Mapping Using Geotweet Densities. *Transactions in GIS* **21**, 317-331 (2017).
57. D. Stathakis, P. Baltas, Seasonal population estimates based on night-time lights. *Computers, Environment and Urban Systems* **68**, 133-141 (2018).
58. C. F. Brown *et al.*, Dynamic World, Near real-time global 10 m land use land cover mapping. *Scientific Data* **9**, 251 (2022).

59. 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).
60. 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).
61. 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*.
62. H. Rue, S. Martino, N. Chopin, Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **71**, 319-392 (2009).
63. F. Lindgren, H. Rue, J. Lindström, An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **73**, 423-498 (2011).
64. J. V. Tuccillo *et al.*, LandScan HD: A High-Resolution Gridded Ambient Population Methodology for the World. 09 April 2025, PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-6396722/v1]. (2025).
65. 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 Symposium on Computing for Development* (Association for Computing Machinery, Nairobi, Kenya), p Article 6.
66. 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.
67. I. Neal, S. Seth, G. Watmough, M. S. Diallo, Census-independent population estimation using representation learning. *Scientific Reports* **12**, 5185 (2022).
68. C. Pezzulo *et al.*, Sub-national mapping of population pyramids and dependency ratios in Africa and Asia. *Scientific Data* **4**, 170089 (2017).
69. 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).
70. 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/. .
71. UCLA-DRC Health Research and Training Program, Kinshasa School of Public Health, Kinshasa, Kongo Central and former Bandundu microcensus survey data (2017-18).
72. Institut de la Statistique du Mali, Cartographie du RGPH5.
73. 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).
74. 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).
75. 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).
76. A. Smith *et al.*, New estimates of flood exposure in developing countries using high-resolution population data. *Nature Communications* **10**, 1814 (2019).
77. 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).
78. 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).
79. J. Cajka, A. Safaa, R. Jamie, J. and Allpress, Geo-sampling in developing nations. *International Journal of Social Research Methodology* **21**, 729-746 (2018).

80. 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).
81. A. Gelman *et al.*, Bayesian workflow. *arXiv preprint arXiv:2011.01808* (2020).
82. P. B. Conn, D. S. Johnson, P. J. Williams, S. R. Melin, M. B. Hooten, A guide to Bayesian model checking for ecologists. *Ecological Monographs* **88**, 526-542 (2018).
83. H. R. Chamberlain, C. A. Dooley, A. J. Tatem, Assessing the accuracy of census-independent small area modelled population datasets. (in preparation).
84. J. H. P. Breuer, J. Friesen, H. Taubenböck, M. Wurm, P. F. Pelz, The unseen population: Do we underestimate slum dwellers in cities of the Global South? *Habitat International* **148**, 103056 (2024).
85. 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).
86. D. R. Thomson *et al.*, Improving the accuracy of gridded population estimates in cities and slums to monitor SDG 11: Evidence from a simulation study in Namibia. *Land Use Policy* **123**, 106392 (2022).
87. J. M. Davis, M. Wilfahrt, Enumerator Experiences in Violent Research Environments. *Comparative Political Studies* **57**, 675-709 (2024).
88. 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).
89. H. Hogan, Distrust in the Governments Brings Risk to the Census. *Harvard Data Science Review* **2**, https://doi.org/10.1162/99608f99692.99611f99603e99977 (2020).
90. 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).
91. 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).
92. 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).
93. C. Visée *et al.*, Addressing bias in national population density models: Focusing on rural Senegal. *PLOS ONE* **19**, e0310809 (2024).
94. M. Owusu *et al.*, Towards user-driven earth observation-based slum mapping. *Computers, Environment and Urban Systems* **89**, 101681 (2021).
95. UNHCR (2024) Global Trends: Forced displacement in 2023. ed U. N. H. C. f. Refugees (United Nations High Commissioner for Refugees, Copenhagen, Denmark).
96. 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).
97. 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).
98. E. Darin *et al.*, Mapping refugee populations at high resolution by unlocking humanitarian administrative data. *Journal of International Humanitarian Action* **9**, 14 (2024).
99. 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).
100. M. Sinclair *et al.*, Assessing the socio-demographic representativeness of mobile phone application data. *Applied Geography* **158**, 102997 (2023).
101. D. R. Leasure *et al.*, Nowcasting Daily Population Displacement in Ukraine through Social Media Advertising Data. *Population and Development Review* **49**, 231-254 (2023).
102. 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).
103. 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*.

104. 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).
105. K. D. Mets, D. Armenteras, L. M. Dávalos, Spatial autocorrelation reduces model precision and predictive power in deforestation analyses. *Ecosphere* **8**, e01824 (2017).
106. S. Openshaw, The modifiable areal unit problem. *Concepts and techniques in modern geography* (1984).
107. S. A. Lee, T. Economou, R. Lowe, A Bayesian modelling framework to quantify multiple sources of spatial variation for disease mapping. *Journal of The Royal Society Interface* **19**, 20220440 (2022).
108. 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).
109. J. L. Dungan *et al.*, A balanced view of scale in spatial statistical analysis. *Ecography* **25**, 626-640 (2002).
110. P. M. Atkinson, A. Stein, C. Jeganathan, Spatial sampling, data models, spatial scale and ontologies: Interpreting spatial statistics and machine learning applied to satellite optical remote sensing. *Spatial Statistics* **50**, 100646 (2022).
111. 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).
112. Project on Government Oversight (2023) Dollars and Demographics: How Census Data Shapes Federal Funding Distribution. (POGO: Washington DC).
113. A. Chatzky, A. Cheatham (2021) Why Does the Census Matter? (Backgrounder, Council on Foreign Relations: New York).
114. S. Desmon (2024) CCP Part of $236 Million Contract that Conducts Key Health Surveys Worldwide. (Commentary, Johns Hopkins Center for Communication Programs: Baltimore).
115. T. Krieger, D. Meierrieks, Population size and the size of government. *European Journal of Political Economy* **61**, 101837 (2020).
116. C. Tuholske *et al.*, Implications for Tracking SDG Indicator Metrics with Gridded Population Data. *Sustainability* **13**, 7329 (2021).
117. B. Fischhoff, A. L. Davis, Communicating scientific uncertainty. *Proceedings of the National Academy of Sciences* **111**, 13664-13671 (2014).
118. 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).
119. S. MacFeely, N. Barnat, Statistical capacity building for sustainable development: Developing the fundamental pillars necessary for modern national statistical systems1. *Statistical Journal of the IAOS* **33**, 895-909 (2017).
120. 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).
121. B. Harrell-Bond, E. Voutira, M. Leopold, Counting the Refugees: Gifts, Givers, Patrons and Clients *. *Journal of Refugee Studies* **5**, 205-225 (1992).
122. 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).
123. 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; https://www.insd.bf/fr/resultats
.
124. 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.
125. PNG National Statistical Office (Population Estimates 2021. https://www.nso.gov.pg/statistics/population/.
126. 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/.
127. UNFPA (2019) Hybrid Census. Technical Brief. (United Nations Population Fund (UNFPA), https://www.unfpa.org/resources/new-methodology-hybrid-census-generate-spatially-disaggregated-population-estimates).

 doi:10.20944/preprints202506.1625.v1

128. 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).
129. WHO, UNICEF (2023) Geo-Enabled Microplanning Handbook: A product of the WHO-UNICEF COVAX GIS Working Group.
130. UNICEF (2025) Reach the Unreached - Geospatial modelling mapping methods. (https://github.com/unicef-drp/reach-the-unreached?tab=readme-ov-file).
131. 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).
132. 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.