

Assessing the accuracy of census-independent small area modelled population datasets

Heather R. Chamberlain

h.chamberlain@soton.ac.uk

WorldPop, University of Southampton <https://orcid.org/0000-0003-0828-6974>

Claire A. Dooley

University College London <https://orcid.org/0000-0003-2415-9895>

Frank Kakungu

Zambia Statistics Agency

Andrew J. Tatem

WorldPop, University of Southampton <https://orcid.org/0000-0002-7270-941X>

Research Article

Keywords: Census, Demography, Building Footprints, Remote Sensing, Bayesian Statistics

Posted Date: October 21st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9558335/v2>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

1 Assessing the accuracy of census-independent small area modelled
2 population datasets

3 Heather R. Chamberlain*¹, Claire A. Dooley², Frank Kakungu³, Andrew J. Tatem¹

- 4 1. WorldPop, School of Geography and Environmental Science, University of
5 Southampton, Southampton, SO17 1BJ, UK
6 2. The Centre for Advanced Spatial Analysis (CASA), The Bartlett Faculty of the Built
7 Environment, University College London, 90 Tottenham Court Road, London, W1T 4TJ,
8 UK
9 3. Zambia Statistics Agency, Lusaka, Zambia

10
11 *Corresponding author: Heather R. Chamberlain

12 **Abstract**

13 Small area population estimates are fundamental for effective decision making across government,
14 humanitarian, development and commercial sectors. Many countries face growing uncertainty in their
15 population figures due to disrupted, delayed or incomplete censuses – an issue highlighted during the
16 2020 round, with many censuses delayed and 15% of countries ultimately not undertaking one. To
17 address this gap, methods to provide census-independent or “bottom-up” modelled population
18 estimates have been developed that combine geospatial data derived from satellite imagery with
19 population data samples within a statistical modelling framework. Such methods have been applied
20 in collaboration with governments in more than ten countries to date, typically in data-scarce settings
21 where opportunities for comparison against independent population data are limited. An exception is
22 Zambia, where census-independent modelled population estimates were developed in 2020, and a
23 national population census subsequently conducted in 2022. Using data from the Zambia 2022
24 census, this study presents the first national comparison of census-independent modelled estimates
25 with recent census enumeration. We directly compare the modelled estimates with census counts
26 across geographic scales and explore subsets of estimates considered to most plausibly represent
27 residential populations. Our findings show that the modelled estimates are strongly correlated with
28 the 2022 census figures ($R \geq 0.95$ for wards and districts) but tend to exceed census counts. Accounting
29 for the locations of residential census households reduces these disparities considerably,
30 highlighting the need for improved stratification of built settlement types in model input data. This
31 study strengthens the evidence base for bottom-up population modelling as an option in situations
32 where there are barriers to conducting a population census, whilst also identifying priorities for further
33 methodological development.

34
35
36 **Main Text**

37 **1. Introduction**

38 Reliable population figures are foundational for a wide range of critical activities, from planning and
39 delivering public services, to identifying vulnerable groups during emergencies and providing
40 denominators for metrics on disease prevalence or development indicators. As these data underpin
41 decision making for governments, humanitarian organisations, commercial actors and non-profit
42 organisations alike, accurate and up-to-date population figures are essential. Population data are
43 traditionally collected through population and housing censuses by national statistical offices,

typically on a decadal basis. Between censuses, population figures are projected, considering rates of fertility, mortality and migration.

As the time elapsed since the last census increases, the uncertainty in projected population figures grows, particularly at local levels. Conducting a census is a massive logistical exercise with high financial costs. For countries experiencing conflict, political instability or disasters associated with natural hazards, censuses can be disrupted, delayed or cancelled, resulting in increasingly uncertain population figures. Around 15% of countries globally did not undertake a census in the 2020 round (2015-2024) (United Nations Statistics Division, 2025), including some of the world's most populous countries. India's planned 2021 census was initially delayed due to COVID-19 and has yet to take place, such that current population figures are projections grounded in enumeration from the 2011 census (Mallapaty, 2024), whilst Nigeria's last national census was in 2006 (Makama, 2007). For some countries, the time elapsed since the last census can stretch to many decades; the last national census in Democratic Republic of the Congo (DRC) was conducted in 1984 (Marivoet & De Herdt, 2018), and in Afghanistan in 1979 (Pinney, 2012).

In response to these challenges, alternative methods of estimating population sizes at local levels have been developed. Over the past decade, geospatial datasets derived from satellite imagery have been used to both spatially disaggregate census population counts (Leyk et al., 2019; Stevens et al., 2015), and to independently estimate population sizes (Boo et al., 2022; Leasure et al., 2020; Neal et al., 2022; Weber et al., 2018). The latter approach, termed "bottom-up" (Wardrop et al., 2018) or "census-independent", utilises enumerated population samples together with comprehensively mapped geospatial datasets in a statistical modelling framework. Geospatial datasets of factors associated with population density and spatial distribution, are used as model covariates, enabling prediction of population in locations without enumerated counts (Wardrop et al., 2018).

Mapped data on buildings, frequently derived from satellite imagery, have been fundamental in advancing predictive modelling of populations (Lazar et al., 2025), given the close relationship between the spatial distributions of population and buildings (Nieves et al., 2021; Rubinyi et al., 2021). To date, the samples of population data used in bottom-up modelling have primarily comprised household listings from surveys or other non-census forms of population enumeration. Census-independent or bottom-up modelled estimates have now been developed for countries including Afghanistan, Nigeria (Leasure et al., 2020; Weber et al., 2018), DRC (Boo et al., 2025; Boo et al., 2022), Cameroon (Nnanatu et al., 2025a), Papua New Guinea (Nnanatu et al., 2025b), South Sudan, Colombia (Sanchez-Cespedes et al., 2024), Mozambique (Neal et al., 2022) and Zambia (Dooley et al., 2021), and is an area of active research for national statistical offices (Office for National Statistics, 2024).

The contexts in which modelled bottom-up population estimates have been developed are generally settings where up-to-date, reliable, population data are scarce, inherently resulting in limited opportunities for comparison against independent sources of population data (Lazar et al., 2025). The accuracy of bottom-up estimates is typically assessed as part of the modelling process through out-of-sample cross-validation, with the input population samples split into training and testing sets (Leasure et al., 2020), but sample sizes are typically small. Opportunities for independent assessment against other population data are rare, although have long been identified as a priority (Wardrop et al., 2018). Recent work by Darin et al. (2025) explored bottom-up model design using data from the 2018 Colombian census. However, of the countries where bottom-up modelled population estimates have been developed from survey data, very few have subsequently conducted a census or extensive population enumeration. Zambia is an exception - bottom-up modelled population estimates were

developed and published openly in 2020, with the 6th National Population and Housing Census subsequently conducted in 2022.

In this study, we compare the 2019 Zambia bottom-up (census-independent) modelled population estimates (Dooley et al., 2021; WorldPop, 2020), against population enumeration from the 2022 census (Zambia Statistics Agency, 2024). Our analyses are structured into three parts (Figure 1). Firstly, we directly compare the modelled estimates and census data at a range of geographic scales. Secondly, given the lack of data available on building type or use when the population modelling was conducted, we use census data on residential household locations to explore the mismatch between building footprints and residential populations. Building on this analysis, we finally estimate an adjusted census population against which to compare the modelled estimates, accounting for locations with residential populations in the census but without modelled population estimates.

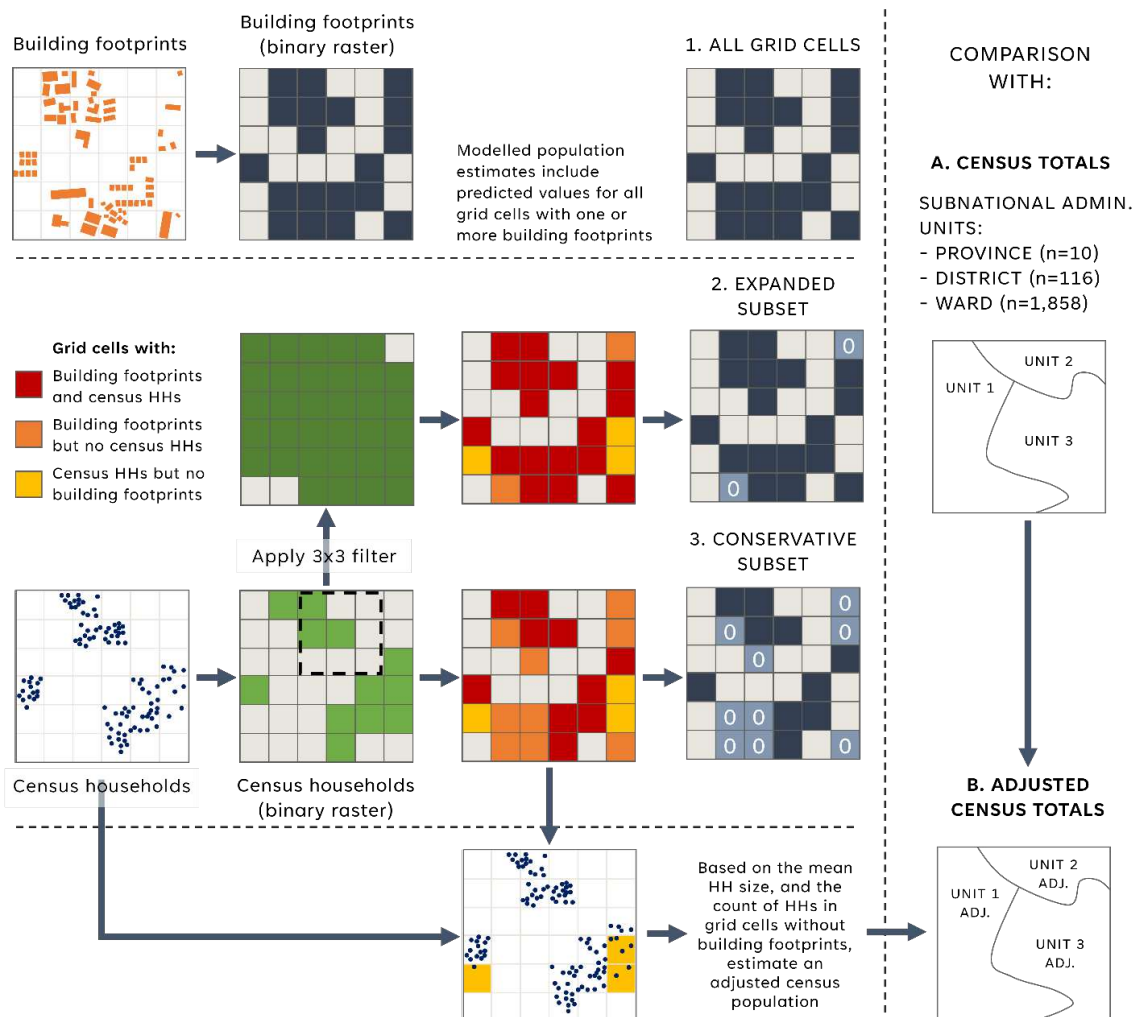


Figure 1. Overview of the comparative analysis, showing how the conservative and expanded subsets of grid cells with modelled population estimates were identified, based on the locations of census residential households. Grid cells with households but no building footprints were identified to calculate a set of adjusted census comparators.

2. Methods and Data

This section provides details of (i) the modelled population estimates dataset, (ii) the census-enumerated population counts, (iii) the grid cell subsets based on stratification of built settlement, (iv) the calculation of an adjusted census comparator and (v) the comparative analysis methods. The modelled estimates were openly published in 2020, prior to the 2022 census being conducted, and are based on geographically limited household listings with model covariates related to building size, count and density (Dooley et al., 2021). The multiple input data sources for the population modelling covered a range of timepoints, but are considered to be most representative of early 2019 (WorldPop, 2020). The full set of modelled population estimates at 3 arc second resolution, consist of 10,000 samples from the predicted posterior distribution for each grid cell, from which we calculated the mean estimate and upper and lower 95% credible intervals (CIs), for different geographic units. Throughout the analyses, we treat the 2022 census counts as a definitive and accurate representation of the population at the time of the census, and consider it as a benchmark against which modelled estimates are evaluated.

2.1. Modelled population estimates

Previous work by Dooley et al. (2021) estimated counts of population at a spatial resolution of 3 arc seconds (0.00083333 decimal degrees, approximately 100m), using a bottom-up modelling approach. This utilised population counts from routinely collected survey data and pilot census cartography, and a range of ancillary datasets as covariates, with similar approaches since implemented for population modelling in other countries (e.g. Nnanatu et al., 2025a; Nnanatu et al., 2025b). Population counts came from three sources: the Saving Mothers, Giving Life (SMGL) Maternal Mortality Endline Census conducted in 2017 (Central Statistical Office, University of Zambia Department of Population Studies, and ICF International, 2018), the Livestock Census (LSC) survey conducted in 2018 (Ministry of Fisheries and Livestock, and Central Statistical Office, 2019), and the pre-census pilot cartography, conducted in 2019. For each source, households had been enumerated and the household location recorded using a GPS-enabled device. From these household point locations, cluster boundaries were derived and population density per cluster calculated. For any locations where clusters from different surveys overlapped, only the data from the most recent survey was kept when the clusters from the three sources were combined into a single dataset. In the combined dataset, 97% (112/116) of districts nationally, had one or more survey cluster.

All ancillary datasets were prepared in gridded (raster) format, with the same grid cell dimensions and alignment. These included a range of metrics related to the count, size and density of buildings, as well as classification of settlement types (rural, small urban or large urban), based on the count of contiguous grid cells with one or more buildings. Metrics were calculated from building footprint data (2D outlines of buildings, extracted from satellite imagery), using building footprints produced by Ecopia with satellite imagery from Maxar through the DigitizeAfrica project (Price & Hallas, 2019). Nationally, 93% of building footprints were extracted from satellite imagery dating from 2017-2019 (Supplementary Figure A.3). Any building footprint that had an area greater than 750 m² was excluded in the calculation of metrics, as it was considered that these buildings were most likely non-residential.

Population density was calculated as population per total building footprint area, and the ancillary datasets were used as covariates to model population density. The statistical modelling approach consisted of a core regression model to predict population density (with model weights used to account for biases in the population counts associated with the survey sampling design) and an observation error sub-model (to account for uncertainty in the population counts due to missing

households). A hierarchical random intercept was included in the regression model to account for spatial similarity that was not explained by the covariates; districts were nested within provinces, which were nested within settlement type. Using the parameterised model, the probability distribution (posterior predictions) of population counts was calculated for every grid cell at 3 arc second resolution. The resulting dataset of census-independent modelled population estimates is openly published (WorldPop, 2020). The gridded nature of the dataset provides flexibility in terms of aggregation of grid cells to calculate estimates for areas of interest. Further details of the input data preparation and modelling methods are described in Dooley et al. (2021).

2.2. Enumerated population counts from 2022 census

Censuses are conducted in Zambia on a decadal basis, implemented by the Zambia Statistics Agency (ZamStats). The Fifth National Population and Housing Census was conducted in 2010, following previous censuses in 1969, 1980, 1990 and 2000. The Sixth National Population and Housing Census was scheduled for 2020, but was delayed to 2022, due to the COVID-19 pandemic and funding constraints. The census was ultimately conducted between August 18th and September 14th 2022, and was the first Zambian census to collect geolocated household-level data, recorded using GPS-enabled devices (Zambia Statistics Agency, 2023). The collected data was then processed by ZamStats to calculate demographic variables for different administrative units, with aggregated population counts published for wards (n=1,858), districts (n=116) and provinces (n=10). The final census results showed a population count of 19,693,423 and 4,087,031 households nationally (Zambia Statistics Agency, 2024).

2.3. Stratification of built settlement

Aside from surveyed population counts, the other main source of input data for the modelled population estimates, was building footprints, and the associated derived metrics. The building metrics have a major role in the estimation of population density, as well as the presence of building footprints also determining the grid cells for which non-zero population counts were predicted. However, a key limitation of building footprint datasets concerns the lack of attributes on building type or use (Chamberlain et al., 2024). In the context of bottom-up modelling, not being able to distinguish residential and non-residential buildings could potentially lead to over- or under-estimation of population. This is of particular concern in areas with predominantly non-residential buildings, and/or if the input survey data does not include areas with non-residential buildings, as population counts may be over-estimated as a result. To understand the influence that the inclusion of non-residential building footprints in the bottom-up population modelling has, comparative analyses were also conducted for subsets of grid cells identified as including residential households according to the census.

Two subsets, which we term “conservative” and “expanded” were identified (Figure 1). For the conservative subset, we included all grid cells that have one or more residential household points recorded in the 2022 census (n=1,145,112; Supplementary Table A.8). Grid cells with one or more residential households were based on the mastergrid file associated with gridded census-derived population estimates (WorldPop and Zambia Statistics Agency, 2024). In the expanded subset, all grid cells with one or more residential household points, plus the eight grid cells surrounding these, are included (n=1,747,697). The conservative subset likely omits some grid cells with residential building footprints, given differences in location between GPS-coordinates recorded for households and building footprint centroids. Conversely, the expanded subset will include some grid cells without residential buildings. The most realistic distribution of grid cells with residential buildings as of the 2022 census, likely falls between the two subsets.

2.4. Calculation of adjusted census comparator

Importantly, there are grid cells with residential census households recorded but with no building footprints (and hence no modelled estimates of population). This may be due to temporal differences between the household enumeration and the satellite imagery used in the building footprint extraction (Supplementary Figure A.3), or errors in the location of recorded residential households or building footprint data. Relatively consistent patterns in this mismatch are observed across provinces (Supplementary Figures A.4 and A.5). To account for this mismatch, we estimated a set of adjusted subnational census totals at ward-level.

For each ward, the count of census households in the subset of grid cells without building footprints was multiplied by the mean household size (census population count divided by count of census households). This estimate of the portion of the census population resident in grid cells without building footprints, was then subtracted from the ward census population, to provide an adjusted census total. Across the 380,754 grid cells nationally with residential census households but no modelled population estimates, we estimated the census population resident in those grid cells to be 2,311,349, which when subtracted from the national census population total, gives an adjusted national census population of 17,382,074 (Supplementary Table A.8).

2.5. Comparative analyses

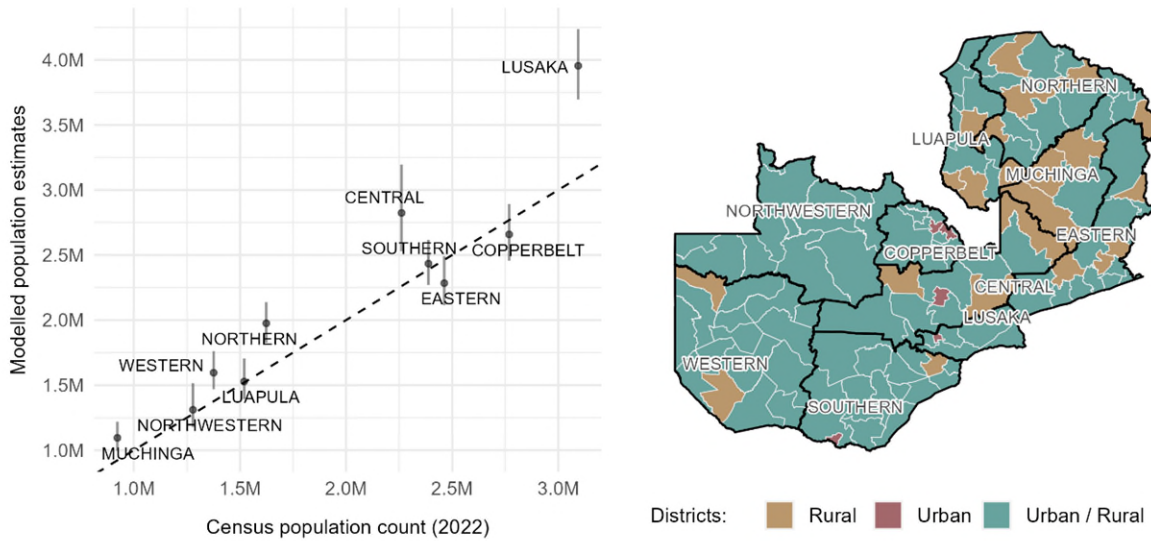
Our analyses compare the bottom-up modelled population estimates with the census enumerated counts, at a range of geographic scales. The modelled estimates were spatially aggregated to calculate estimates for all wards, districts and provinces, with estimates corresponding to the (i) mean and (ii) 95% credible intervals, calculated for each administrative unit from the posterior predictions. Ward, district, province and national estimates, based on all grid cells and the two grid cell subsets are provided in accompanying Supplementary dataset S1, alongside the corresponding census counts and adjusted census comparators.

For wards, and districts, the following model evaluation statistics were calculated based on the mean value of the modelled estimates and the census counts: Pearson correlation coefficient (R), Root Mean Square Error (RMSE; square root of the mean squared difference), Mean Absolute Error (MAE; the mean absolute difference), Mean Absolute Percentage Error (MAPE; the mean absolute percentage difference), Median Absolute Error (MedAE; the median absolute difference), bias (mean of the difference), and imprecision (standard deviation of the difference). We firstly undertake this comparative analysis with the full dataset of the modelled estimates (all grid cells) and subsequently focus on two grid cell subsets. The identification of the two subsets, based on residential household locations recorded as part of the census process, are described in section 2.3. The final part of our analysis compares the subsets of modelled estimates with adjusted census totals, to account for locations with residential households recorded in the census but no modelled population estimates.

3. Results

The results are structured in three parts. Section 3.1 has results of the direct comparison of modelled population estimates and census counts. The modelled estimates calculated from subsets of grid cells with residential households, are then compared with the census counts in Section 3.2, and adjusted census comparators in Section 3.3.

A Province



B District

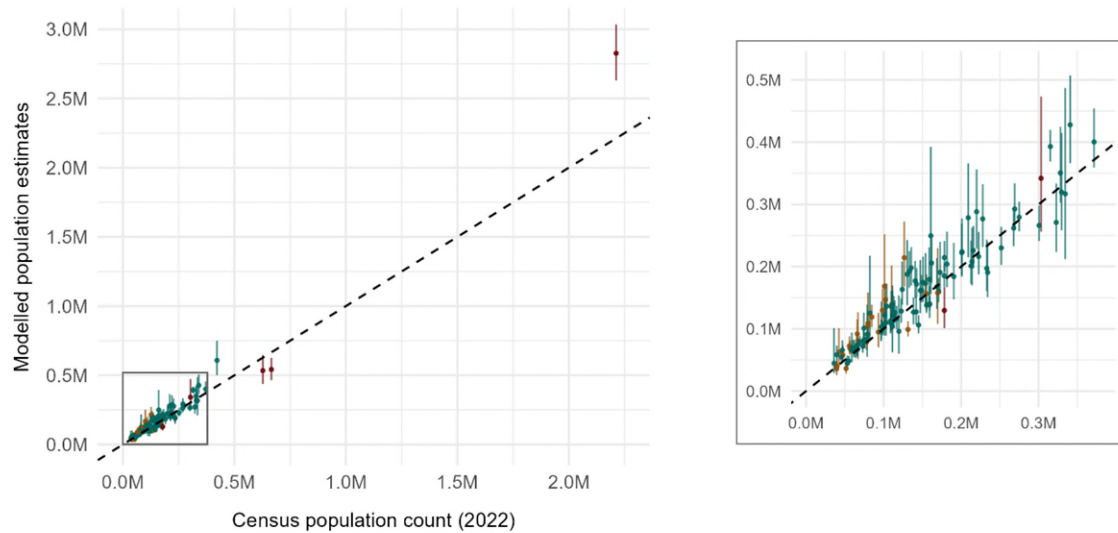


Figure 2. Modelled population estimates (~2019) with 95% credible intervals (CIs) plotted against census counts (2022) for each (a) province (n=10) and (b) district (n=116) nationally. The diagonal dashed line ($y=x$) on each plot shows a one-to-one relationship between the modelled and census-enumerated counts. Districts are stratified as urban, rural or both urban and rural, based on ZamStat's binary rural/urban classification of wards.

3.1. Comparison of modelled estimates with census counts

The bottom-up modelled population estimates (~2019), correlate well with the enumerated counts from the 2022 census. This is shown visually in Figure 2 for (a) provinces and (b) districts and in Figure 3 for wards, with the mean value of the modelled estimates (y-axis) generally well correlated with the

census enumerated counts (x-axis). Alongside the modelled mean values, the 95% CIs of the modelled estimates are also shown, which for half of provinces (5/10), and the majority of districts (74/116, 64%) and wards (1,001/1,858, 54%), span the $y=x$ line (Figure 4).

At the province-level, the modelled population mean estimates exceed the 2022 census enumerated counts for all provinces, except Copperbelt and Eastern (Figure 2a). The mean estimates of population for Muchinga, Western, Northern, Central, and Lusaka provinces are considerably higher, and the 95% CIs do not span the $y=x$ line. At the district-level (Figure 2b), the modelled population mean estimates exceed the 2022 census enumerated counts for 69% of districts ($n=80/116$), of which 46 have 95% CIs that span the $y=x$ line. Of the 36 districts in which the modelled mean population estimates are lower than the 2022 census counts, 28 have 95% CIs that span the $y=x$ line. Across all districts, this bias results in the mean value of the modelled population estimates exceeding the census count by an average of 16,929 people per district (Table 1), but this is reduced to 10,028 when the district of Lusaka is excluded.

Table 1. Model evaluation statistics, comparing the mean value of the modelled estimates against census enumeration for districts and wards.

Statistic	District-level			Ward-level		
	All grid cells	Expanded subset	Conservative subset	All grid cells	Expanded subset	Conservative subset
R (Pearson)	0.984	0.985	0.981	0.955	0.961	0.958
RMSE	68,056.2	63,561.7	66,540.1	4,294.9	3,804.3	4,128.8
MAE	30,125.3	25,100.5	36,810.0	2,427.7	1,985.6	2,537.3
MAPE [%]	17.8	12.6	21.3	25.4	18.0	24.1
MedAE	17,748.9	13,034.9	23,091.9	1,314.5	972.6	1,549.0
Bias	16,928.5	-2,246.2	-26,808.3	1,056.9	-140.2	-1,673.7
Imprecision	66,203.1	63,797.6	61,165.0	4,164.0	3,802.7	3,775.4

At ward-level, modelled mean population estimates and census counts are still strongly correlated ($R = 0.96$), but compared to district-level we expectedly see more variation in the differences between modelled estimates and 2022 enumerated population counts (Table 1 and Figure 3b). This is reflected in a MAPE value of 25.4% for ward estimates, compared to 17.8% for districts. When model evaluation statistics for wards are stratified by province (Supplementary Table S1), values of MAPE range from 17.4% (Eastern) to 37.5% (Lusaka), and bias from -868.7 (Eastern) to 7,293.4 (Lusaka). This variability is further explored in Supplementary Figure A.1, with scatterplots of ward-level population stratified by province. The proportion of wards for which the census count is within the 95% CI of the modelled estimates varies between provinces, ranging from 33% (Lusaka) to 69% (Eastern), with the proportion of wards in Central, Luapula and Northwestern all exceeding 60% (Figure 4).

For Lusaka province, ward-level modelled estimates are very strongly correlated ($R = 0.99$; Supplementary Table A.1), but generally higher than the enumerated counts in terms of both mean values and 95% CIs (Figures 3a and 4). For Copperbelt province, ward-level modelled estimates are slightly less strongly correlated ($R = 0.92$), with mean estimates lower than census counts by 434 people on average (Supplementary Table A.1). The census count is within the 95% CIs for 44% of wards in Copperbelt province, with estimates for the upper 95% CI less than the census count in 27% of wards: the highest for any province (Figure 4). Within Copperbelt, modelled estimates are generally

lower than census counts in urban, more populous wards, and higher than census counts in rural, less populous wards (Figure 3a and Supplementary Figure A.1).

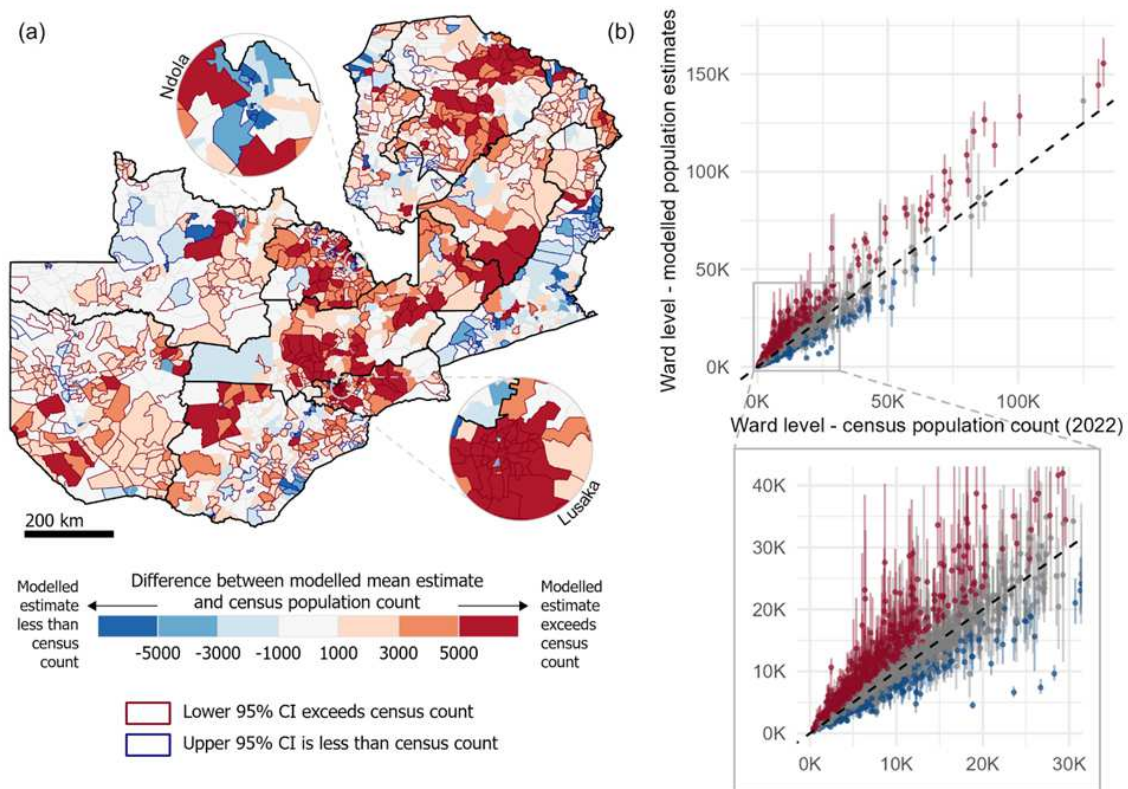


Figure 3. The difference between modelled population estimates (mean value) and census enumerated counts is shown for all wards nationally ($n=1858$). Part (a) shows this difference mapped at ward-level, with (b) showing the absolute counts as a scatterplot. In both (a) and (b), the census count is also compared against the 95% CIs of the modelled estimates, shown in terms of boundary outline colour in part (a) and line colour in part (b). Wards in which the lower 95% CI exceeds the census count, are shown with a red outline/line, whilst wards in which the upper 95% CI is less than the census count are outlined in blue.

3.2. Comparison with stratified subsets

The results so far have focussed on comparing the bottom-up modelled estimates against census-enumerated counts for different administrative units. Administrative unit totals have been calculated based on posterior distributions of estimated population counts for all grid cells that have their centroid within an administrative unit. At the national level, there are 2,273,498 grid cells summing to give a mean estimated population of 21,657,131, with lower and upper 95% CIs of 20,912,731 and 22,474,310 respectively. With modelled population estimates predicted for all grid cells with one or more building footprints (which have no attributes on building function or use), any stratification of

these grid cells and calculation of aggregate values of the estimates from a subset, will inevitably result in reduced population estimates.

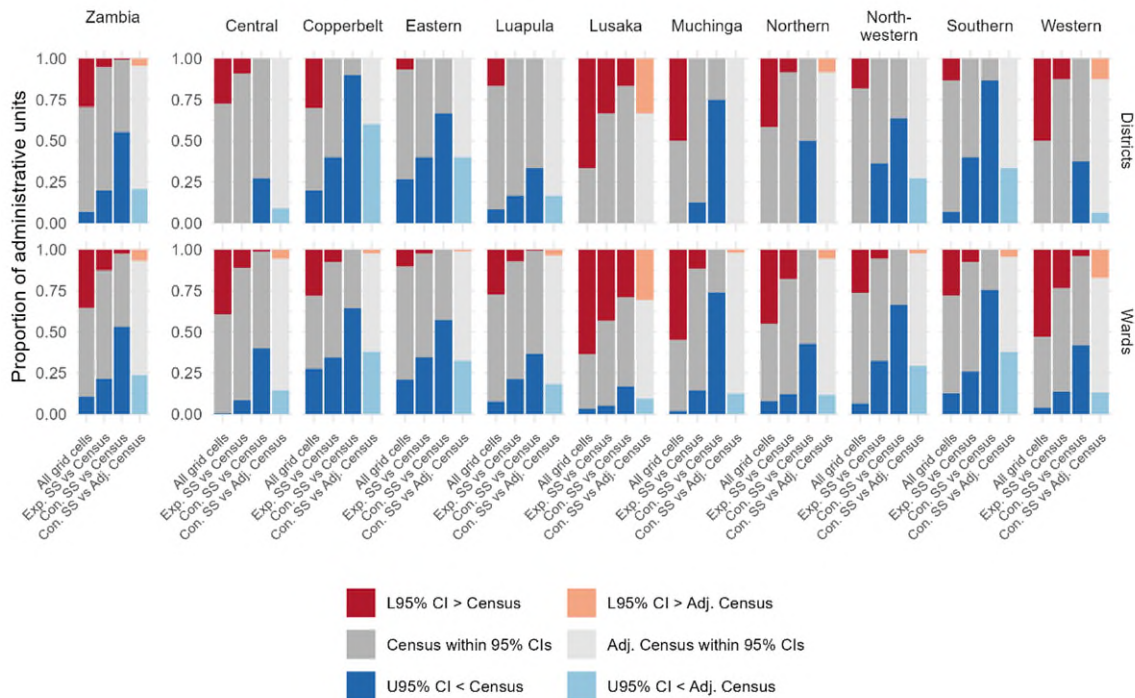


Figure 4. The proportion of administrative units for which the census totals (left three bars) or adjusted census totals (right bar) are (1) within the 95% CIs of the modelled estimate (grey), (2) the lower 95% CI exceeds the census total (red), or (3) census total exceeds the upper 95% CI of the modelled estimate (blue). Results are shown for all administrative units in Zambia and also stratified by province, with proportion of districts shown in the upper row and proportion of wards shown in the lower row. Tabulated values are provided in Supplementary Tables A.2, A.3, A.6 and A.7.

Figures 5 and 6, show administrative unit totals calculated from stratified subsets of the modelled population estimates, which we refer to as “conservative” and “expanded” subsets. Specifically, these are subsets of grid cells considered to reflect the spatial distribution of residential households at the time of the 2022 census (Figure 1). The conservative subset consists of the grid cells with modelled population estimates that contain one or more residential census households. The expanded subset relaxes this definition to also include the eight surrounding grid cells (to account for potential differences between building footprint centroids and GPS-recorded household locations). The conservative subset consists of 1,145,112 grid cells (approximately half the number of grid cells in the full set of modelled estimates), whereas the expanded subset includes 1,747,697 grid cells (Supplementary Figure A.2). Summing the values of the conservative subset gives a mean national value of 16,583,699, with 95% CIs of 16,002,549 - 17,231,084. The expanded subset sums to a mean national value of 19,432,901, with 95% CIs of 18,775,278 - 20,165,145.

In Figure 5A, the difference between modelled estimates and census counts, considering all grid cells and the two subsets, are shown for provinces, districts and wards. At province-level, the mean value of the modelled estimates with the expanded subset exceeds the census count in four provinces, but the census count is within the 95% CIs for six provinces. With the conservative subset, the province

estimates (mean value and upper 95% CI) are less than the census count for all provinces except Lusaka.

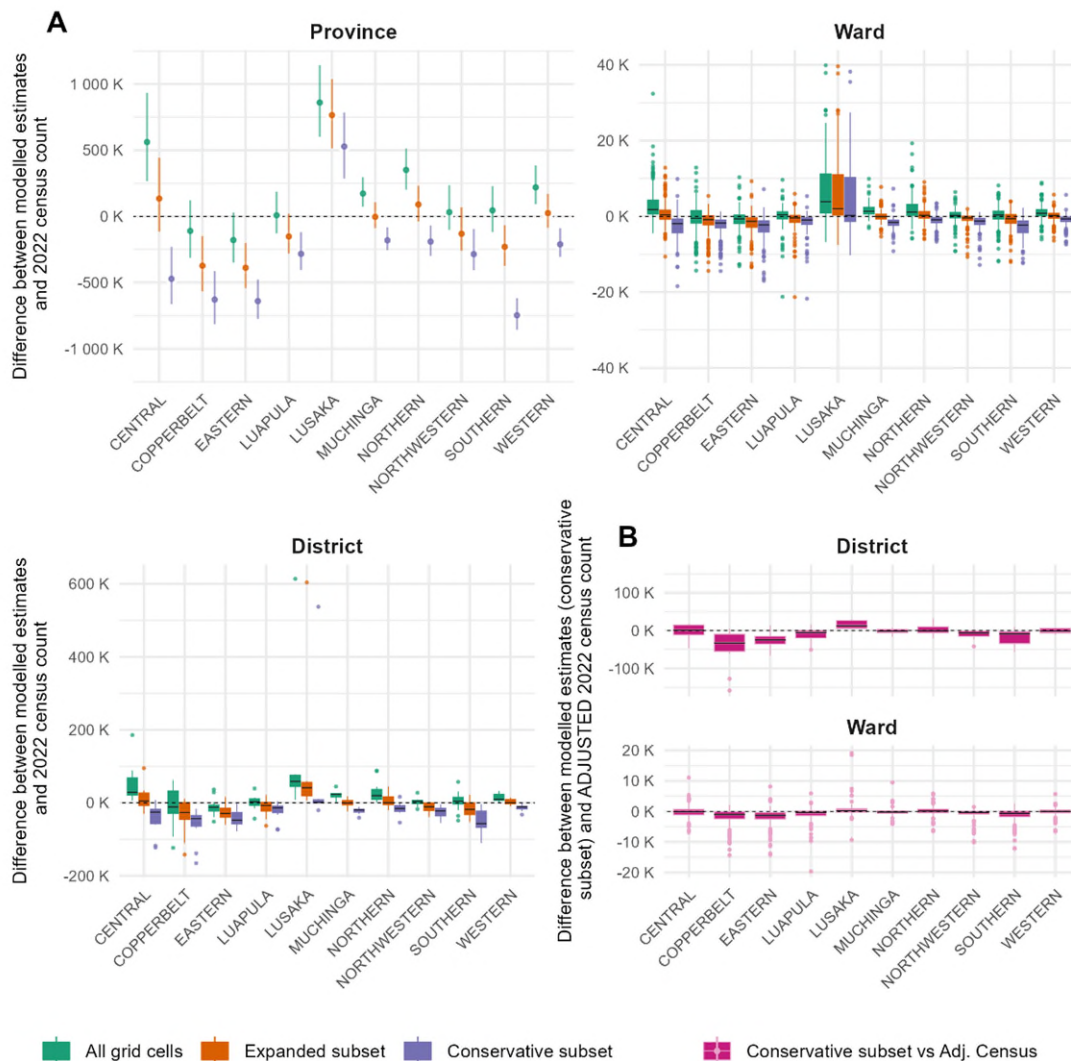


Figure 5. (A) The difference between the bottom-up modelled estimates of population and the 2022 census enumeration, for provinces (top left), districts (bottom left) and wards (top right), with values stratified by province (see map in Figure 2). Differences are quantified in terms of the mean modelled value, with aggregated estimates calculated from all grid cells, as well as the “expanded” and “conservative” subsets of grid cells. For wards and districts, the distribution of values is represented as a boxplot, whilst for provinces, the difference between the modelled province estimate and census count is shown, alongside the difference between the 95% CIs and census count. (B) The plots in the bottom right show the difference between the modelled estimates, based on the conservative subset, and the adjusted census totals. Note that Lusaka district is excluded from plot B.

At the district-level, for the expanded subset of grid cells, the mean value of the modelled estimates exceeds the census count in 40% of districts (46/116), with the census count being within the 95% CIs

of the modelled estimates for 75% of districts (Figure 4). Of the remaining districts, the upper 95% CI is less than the census count for most districts (20%; Supplementary Table A.2). When calculated from the conservative subset of grid cells, the modelled estimates are lower than the census counts for most districts, with the mean value less than the census count in 93% of districts (108/116). The proportion of districts for which the modelled estimates are within the 95% CIs is reduced to 44%, with the upper 95% CI being less than the census count for all remaining districts except those in Lusaka province (Figure 4). The district of Lusaka has the largest difference between the modelled estimates and census count (Figure 5), irrespective of whether calculated from the full set or subset of estimates (all grid cells: 2,826,439 (2,630,819 – 3,034,425), expanded subset: 2,816,522 (2,621,134 – 3,024,264), conservative subset: 2,749,855 (2,555,681 – 2,955,915) and census count: 2,212,301).

In terms of ward-level estimates, the mean value of the modelled population estimates exceeds the census count in 65% (1,208/1,858) of wards (Figures 3 and 6). Calculating ward-level values from the modelled estimates with the expanded subset of grid cells reduces this to 44% (816/1,858) of wards, and with the conservative subset of grid cells to 13% (239/1,858) of wards (Figures 5 and 6). Considering also the upper and lower 95% CIs, the census counts are within the 95% CIs for 66% and 44% of wards with the expanded and conservative subsets respectively, compared to 54% of wards, when estimates are calculated from the full modelled estimates dataset (Figure 4). The percentage of wards for which the lower 95% CI of the modelled estimates exceeds the census count is 35% (657/1,858) when calculated from all grid cells, compared to 12% (232/1,858) and 2.5% (46/1,858) for the expanded and conservative subsets respectively (Figure 4). Thus, the proportion of wards nationally for which the census count exceeds the upper 95% CI of the modelled estimates increases from 11% of wards with the full dataset, to 53% with the conservative subset. This varies sub-nationally though, with census counts exceeding the upper 95% CI of the modelled estimates based on the conservative subset in over 70% of wards in Muchinga and Southern provinces, but only 17% of wards in Lusaka province (Figure 4).

The model evaluation statistics in Table 1, show an improvement when comparing metrics for all grid cells against the expanded subset, for both ward- and district-level. The magnitude of the change does vary by province, but reductions in MAE, MAPE, MedAE and RMSE at ward-level with the expanded subset are observed for all provinces, except Eastern (with a negligible increase in RMSE for wards in Luapula; Supplementary Table A.1). Changes in model evaluation statistics with the conservative subset are more complicated, with some metrics improving and others worsening (Table 1). In terms of bias, ward-level estimates have a negative bias with both the expanded and conservative subsets of grid cells, with an average under prediction across all wards nationally of 140 and 1,673 people respectively compared to the census counts (Table 1). Bias (and other metrics) vary considerably between provinces however, with the large differences observed for ward-level estimates in Lusaka province (Supplementary Table A.1) skewing the national average. For the conservative subset, ward-level estimates are negatively biased for all provinces except Lusaka (range of bias values per province: -3,107 to -928, Lusaka = 4,480). When the wards in Lusaka province are excluded, the magnitude of the negative bias in the ward-level mean estimates increases, with modelled estimates on average 590 (expanded subset) or 2,091 people (conservative subset) less than the census counts.

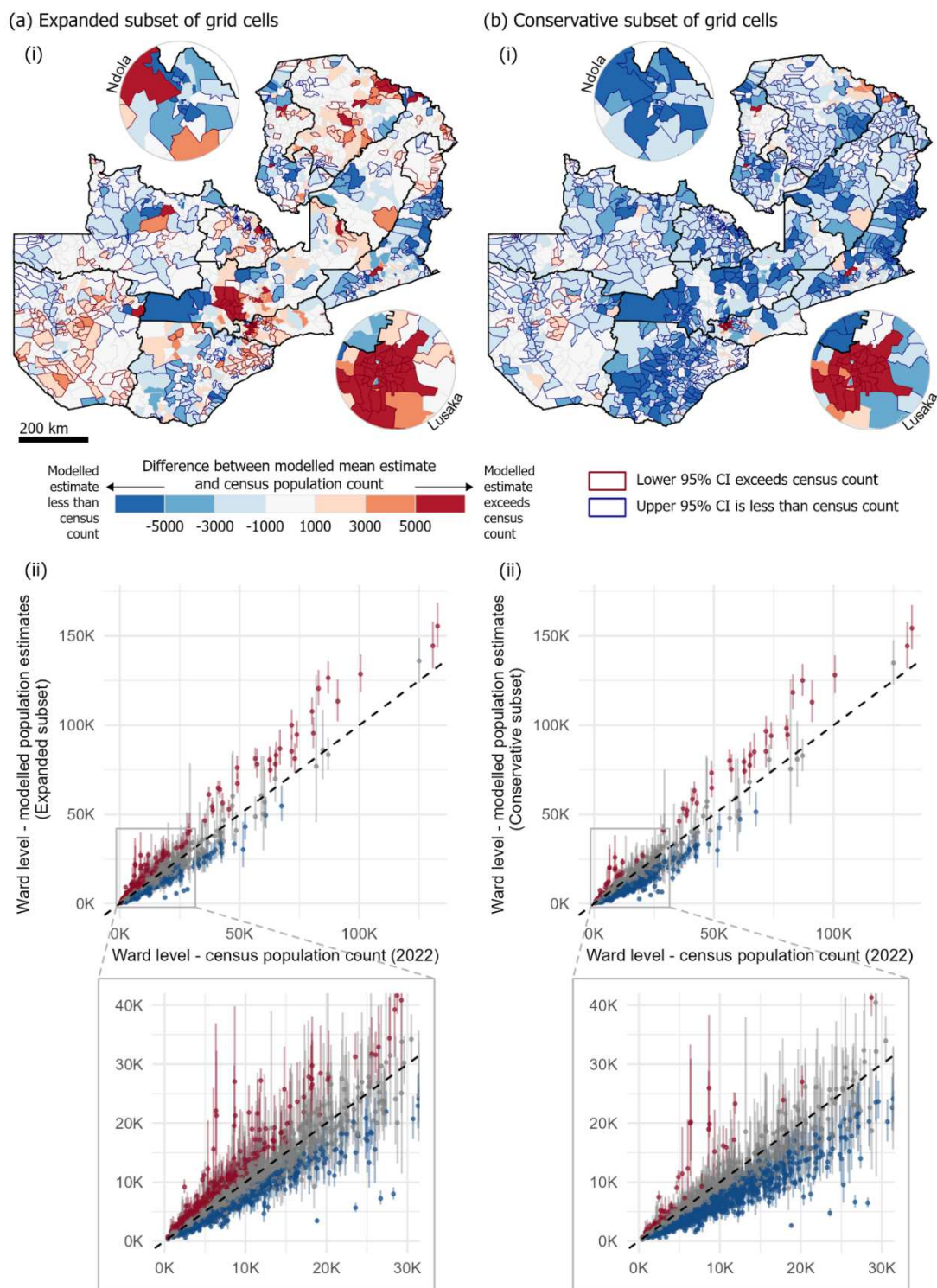


Figure 6. The difference between the modelled estimates aggregated at ward-level and census enumerated counts for the (a) expanded and (b) conservative subsets of grid cells. Absolute differences in terms of the mean population estimate are mapped in panel (i) and the ward-level population values shown on the scatterplot in panel (ii). In both (i) and (ii), the 95% CIs of the modelled estimates are compared against the census counts, with differences indicated by the boundary

outline colour in panel (i) and the line colour of the CIs in panel (ii). Wards in which the lower 95% CI exceeds the census count, are shown with a red outline/line, whilst wards in which the upper 95% CI is less than the census count are outlined in blue.

3.3. Comparison of stratified subsets with adjusted census counts

The comparisons based on stratified subsets described above, do not account for grid cells where residential households were recorded in the census process, but with no building footprints and therefore no modelled estimates of population (n=380,754 grid cells nationally). To provide a fairer comparison, we also calculated adjusted census totals, considering just the grid cells with both residential census households and modelled population estimates. The adjusted national census total of 17,382,074 (Supplementary Table A.8) compares favourably with the national total of the modelled estimates based on conservative subset (mean national value of 16,583,699, with 95% CIs of 16,002,549 - 17,231,084). Focussing subnationally, Figure 5B shows the difference between the modelled estimates for districts and wards based on the conservative subset and the adjusted census counts. The difference is of a smaller magnitude, compared to unadjusted census counts, with the mean estimate being less than the adjusted census total for 66% of districts and 63% of wards. The adjusted census population totals are within the 95% CIs of the modelled estimates, calculated from the conservative subset for 75% of districts and 69% of wards nationally, with the lower 95% CI exceeding the adjusted census total for 4.3% and 6.8% of districts and wards respectively (Figure 4). In Central, Muchinga and Northern provinces, adjusted census population totals are within the 95% CIs of the modelled estimates (conservative subset) for over 90% of districts (SI Appendix; Table A.7).

Figures A.6 and A.7 (SI Appendix) show the subsets of modelled estimates plotted against the adjusted census counts, at province, district and ward-level. Both subsets of modelled estimates are strongly correlated with the adjusted census totals ($R > 0.96$ at ward and district-level nationally; Supplementary Table A.4). Model evaluation statistics show the conservative subset of the modelled estimates and the adjusted census totals compare favourably for both districts and wards (Supplementary Table A.4), for example MAPE (12.3% and 16.2% respectively) and bias (-6,882.9 and -429.7 respectively). Table A.5 provides the same statistics for ward-level estimates, stratified by province; MAPE values were lowest for wards in Muchinga (10.7%) and highest in Lusaka (23.1%), with the smallest magnitude of bias for wards in Western (-76.5) and the greatest magnitude for wards in Lusaka (5,283.7).

4. Discussion

The need for small area population estimates is well recognised, but with growing challenges around accuracy and availability of demographic data, there is a real need to consider alternative approaches to censuses and registry-based sources (Espey et al., 2025). Census-independent or ‘bottom-up’ modelled estimates (Wardrop et al., 2018) have now been developed in partnership with governments for many countries. This study provides the first empirical comparison of bottom-up modelled population estimates with enumerated population counts from a subsequently conducted national census. Specifically, we focus on population count data in Zambia, comparing census-independent, modelled estimates (Dooley et al., 2021; WorldPop, 2020), against enumeration from the 2022 Population and Housing Census (Zambia Statistics Agency, 2024), across multiple spatial scales.

The modelled population estimates are highly correlated with the census-enumerated counts across all geographic scales that we studied (province, district and ward). A key consideration in this comparison is the temporal difference between the bottom-up modelled population estimates (circa 2019) and the census (2022). If the spatial distribution of population remained broadly stable between these timepoints (as would be expected with relatively high population growth rates nationally), 2022

census counts would typically be expected to exceed the 2019 modelled estimates. Our results show that the mean values of the bottom-up modelled population estimates exceed the census count in 80% of provinces, 69% of districts and 65% of wards. The census-enumerated counts are within the 95% CIs of the modelled estimates for half of provinces, 64% of districts and 54% of wards, with the lower 95% CI exceeding the census count in 50% of provinces, 29% of districts and 36% of wards. Overall, the modelled estimates tend to be higher than the census counts, despite representing a timepoint prior to the census.

To understand what may be driving these differences, we first focussed on the ancillary geospatial datasets that are used as predictive variables. The Zambia bottom-up modelling incorporated predictive variables related to building area and counts, alongside a simple three-strata classification of settlement type (Dooley et al., 2021). The building footprint data used to derive these variables did not include attributes to stratify residential and non-residential buildings. Comparing the locations of residential households recorded as part of the census process with the building footprints shows considerable differences (Supplementary Figures A.4 and A.5). The subset of grid cells with one or more residential households (conservative subset), reduces the number of grid cells with modelled population estimates nationally by half, corresponding to a 23% reduction in the mean national population estimate. The expanded subset of grid cells still results in a 23% reduction in the number of grid cells with an associated 10% reduction in the mean national population estimate. With both subsets, the difference between modelled estimates and the census count at the national-level is reduced (Figure 5), but the magnitude of the difference between modelled estimates and census counts varies sub-nationally, with modelled estimates based on subsets being much lower than the census counts in some locations (e.g. province estimates for Copperbelt, Eastern and Southern).

Comparisons based on the conservative and expanded subsets provide insights into differences between the building footprints and residential dwellings, but do not account for locations with residential census households but no building footprints, and hence with no modelled population estimates. To address this limitation, we separately calculated a set of adjusted census totals, to provide an adjusted comparator. The modelled estimates based on the conservative subset compare favourably with the adjusted census totals (mean national estimate of 16,583,699, (95% CIs: 16,002,549 - 17,231,084) versus adjusted census total of 17,382,074), with mean estimates exceeding the adjusted census total for 34% of districts and 37% of wards. The lower 95% CIs for the modelled estimates (conservative subset) exceeded the adjusted census totals for only 4.3% of districts and 6.8% of wards, meaning that adjusted census totals were within the 95% CIs, or greater than the upper CI, for 96% of districts and 93% of wards. Together these results show that much of the discrepancy between the modelled estimates and the census counts, can be explained by limitations with the ancillary data on building footprints, leading to over-estimation of population in locations without residential households. Once these are accounted for, the bottom-up modelled population estimates are broadly consistent with the adjusted census comparator.

The province for which we see the greatest difference between the bottom-up modelled population estimates and the census enumerated counts, is Lusaka. Within Lusaka province, the greatest difference occurs in Lusaka district, which encompasses most of the capital's urban area (Figure 3) and is the most populous district nationally (Figure 2). The 2022 census recorded a population of 2,212,301, whereas the modelled estimate was 2,826,439 (95% CI: 2,630,819 - 3,034,425). This substantial difference suggests that the bottom-up modelling overestimated the population of Lusaka considerably. Even after restricting the analysis to the expanded subset of grid cells, the lower bound of the 95% CI still exceeds the census count of in two of Lusaka province's six districts (Lusaka and Chongwe) and 43% of its wards (51/118) (Supplementary Tables A.2 and A.3). Notably, the districts of Lusaka and Chongwe, were the two districts where the 2019 census cartography pilot exercise was conducted, which included geolocated listing of households (Zambia Statistics Agency, 2023). As the

population count data collected during the pilot census cartography was one of three input data sources for the bottom-up population modelling and was the most spatially extensive of these (Dooley et al., 2021), we would expect that the modelled population estimates for these two districts to be comparatively accurate. The large difference between the modelled estimates and census counts is therefore surprising. Interestingly, sub-national population projections based on the 2010 census, provide mid-year population figures for Lusaka district of 2,627,716 in 2019 and 2,946,566 in 2022 (medium variant (Central Statistical Office, 2013b) – substantially higher than the 2022 census count of 2,212,301.

In comparing the modelled estimates and census counts, there is the complicating factor of the COVID-19 global pandemic which occurred between the collection of model input population data (2017-2019) and the census in 2022. Alongside a heavy burden of morbidity and mortality, the COVID-19 pandemic caused widespread changes in patterns of everyday life, including population movements, with out-migration from urban areas observed in some countries (Duan et al., 2026; González-Leonardo et al., 2022; Rowe et al., 2023). It is therefore conceivable that there were considerable changes in the geographic distribution of population between these timepoints, including potentially the urban population within the city of Lusaka between 2019 and 2022.

In this work, we compared the modelled estimates against counts from the 2022 Population and Housing Census, considering the census as a definitive and accurate representation of the population. Whilst censuses remain the primary source of population data for most countries, they rarely achieve full coverage, and thus adjustments to address undercounting are necessary (Pelletier et al., 2020). Common methods to adjust for undercounting include Dual Systems Estimation based on a post-enumeration survey (O'Hare, 2017). The undercount in the Zambia 2010 Census was estimated to be 7.3% nationally, with higher rates in rural areas (9.5%) and lower rates in urban areas (3.8%) (Central Statistical Office, 2013a). For the 2022 census, no post-enumeration survey results have yet been published, meaning that the extent of under-counting cannot be assessed. This uncertainty limits our ability to fully interpret the differences observed between the bottom-up modelled estimates and the census counts. Possible other contributing factors include potential over-enumeration in the household listing data used for model input (including pilot census cartography), errors in the geolocation of households and/or errors in the imputation of missing households used for modelling. Further work should explore these factors in more detail.

The considerable differences in the population estimates calculated from the two subsets of grid cells at all spatial scales, underscores the critical influence of building footprints as ancillary datasets in bottom-up modelling. We considered two subsets of grid cells, based on a conservative and expanded definition applied to the location of residential households recorded in the census. Whilst this binary classification is a simplification of the gradient across residential and non-residential building types, it nonetheless provides valuable insights into the influence of building footprint data in bottom-up population modelling. It also highlights the need for more refined approaches to more accurately map the gradient of residential and non-residential building types. Since the bottom-up modelled estimates for Zambia were developed in 2020, the availability and diversity of building footprint datasets has increased rapidly (Chamberlain et al., 2024; Minghini et al., 2024). Emerging near-global datasets that could be used in future work as predictive variables, include temporally explicit building footprints (Sirko et al., 2023), and data to stratify residential and non-residential built settlement or estimate building heights (Che et al., 2024; Esch et al., 2022; Oostwegel et al., 2025; Pesaresi et al., 2024).

As planning for the 2030 census round commences, national statistical offices (NSOs) are facing mounting challenges including declining response rates and funding cuts (Espey et al., 2025). With demographic data gaps almost inevitable in this context, there is a potential role for modelling

approaches to estimate population counts, with the caveat that modelled estimates are no replacement for the much wider range of sociodemographic variables collected through a census. To date, most bottom-up population modelling efforts have been led by academic institutions, collaborating with NSOs. If we are to see more widespread adoption of such approaches, investment to support sustained strengthening of statistical modelling capacity within NSOs is needed (Lazar et al., 2025). Further research is also needed to assess the accuracy of population estimates from bottom-up models in diverse contexts, and to guide methodological advancements in terms of input data, model form and appropriate spatial scales. Our comparative analysis of modelled estimates and census counts for Zambia, demonstrates that modelling approaches can produce relatively accurate small area estimates, but highlights the need for temporally aligned data on building type and use. Ongoing developments in satellite-derived data offer promising opportunities to better capture building form and function, which should further improve the accuracy of modelled population estimates.

Data Availability

All data in the paper are available to readers. The bottom-up modelled population estimates for Zambia are available to download (WorldPop, 2020) via the WorldPop Open Population Repository (<https://doi.org/10.5258/SOTON/WP00662>). The Zambia 2022 Census population counts for wards, districts and provinces are published in Table 5.2 in the 2022 Census of Population and Housing Summary Report: Volume Two (Zambia Statistics Agency, 2024). The census totals, as well as the modelled estimates (mean and 95% CIs) calculated from all grid cells and the expanded and conservative subsets, are provided for all provinces, districts and wards nationally in Supplementary Dataset S1.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the authors used Microsoft Copilot selectively to refine some sections of the manuscript text. All AI-suggested content was reviewed and edited by the authors who take full responsibility for the content of the published article.

References

- Boo, G., Darin, E., Chamberlain, H. R., Hosner, R., Akilimali, P. K., Kazadi, H. M., Nnanatu, C. C., Lázár, A. N., & Tatem, A. J. (2025). Tackling public health data gaps through Bayesian high-resolution population estimation: A case study of Kasai-Oriental, Democratic Republic of the Congo. *PLOS Global Public Health*, 5(9), e0005072.
- Boo, G., Darin, E., Leasure, D. R., Dooley, C. A., Chamberlain, H. R., Lázár, A. N., Tschirhart, K., Sinai, C., Hoff, N. A., Fuller, T., Musene, K., Batumbo, A., Rimoin, A. W., & Tatem, A. J. (2022). High-resolution population estimation using household survey data and building footprints. *Nature Communications*, 13(1), 1330. <https://doi.org/10.1038/s41467-022-29094-x>
- Central Statistical Office, University of Zambia Department of Population Studies, and ICF International. (2018). Saving Mothers, Giving Life: 2017 Zambia Maternal Mortality Endline Census in Selected Districts. <https://medbox.org/dl/5e148832db60a2044c2d5abb>

614 Central Statistical Office. (2013a). 2010 Census of Population and Housing Post Enumeration Survey
615 (PES). Central Statistical Office. [https://www.zamstats.gov.zm/wp-content/uploads/2023/12/2010-](https://www.zamstats.gov.zm/wp-content/uploads/2023/12/2010-Census-Post-Enumeration-Report-Final.pdf)
616 [Census-Post-Enumeration-Report-Final.pdf](https://www.zamstats.gov.zm/wp-content/uploads/2023/12/2010-Census-Post-Enumeration-Report-Final.pdf)

617 Central Statistical Office. (2013b). 2010 Census of Population and Housing: Population and
618 Demographic Projections 2011-2035. Central Statistical Office. [https://www.zamstats.gov.zm/wp-](https://www.zamstats.gov.zm/wp-content/uploads/2023/12/Zambia-Census-Projection-2011-2035-Final.pdf)
619 [content/uploads/2023/12/Zambia-Census-Projection-2011-2035-Final.pdf](https://www.zamstats.gov.zm/wp-content/uploads/2023/12/Zambia-Census-Projection-2011-2035-Final.pdf)

620 Chamberlain, H. R., Darin, E., Adewole, W. A., Jochem, W. C., Lazar, A. N., & Tatem, A. J. (2024).
621 Building footprint data for countries in Africa: to what extent are existing data products comparable?
622 Computers, Environment and Urban Systems, 110, 102104.

623 Che, Y., Li, X., Liu, X., Wang, Y., Liao, W., Zheng, X., Zhang, X., Xu, X., Shi, Q., & Zhu, J. (2024). 3D-
624 GloBFP: The first global three-dimensional building footprint dataset. Earth System Science Data,
625 16(11), 5357-5374.

626 Darin, E., Kashyap, R., & Leasure, D. R. (2025). How accurate are high resolution settlement maps at
627 predicting population counts in data scarce settings? Population, Space and Place, 31(6), e70083.

628 Dooley, C., Chamberlain, H., Leasure, D., Membele, G., Lazar, A., & Tatem, A. (2021). Description of
629 methods for the Zambia modelled population estimates from multiple routinely collected and
630 geolocated survey data, version 1.0. <https://doi.org/10.5258/SOTON/WP00707>

631 Duan, Q., Lai, S., Sorichetta, A., Tatem, A. J., Steele, J., & Eigenbrod, F. (2026). COVID-19 and urban
632 exodus: diverging population redistribution patterns across countries from 2020 to 2022. npj Urban
633 Sustainability. 6(59).

634 Esch, T., Brzoska, E., Dech, S., Leutner, B., Palacios-Lopez, D., Metz-Marconcini, A., Marconcini, M.,
635 Roth, A., & Zeidler, J. (2022). World Settlement Footprint 3D-A first three-dimensional survey of the
636 global building stock. Remote Sensing of Environment, 270, 112877.

637 Espey, J. M., Tatem, A. J., & Thomson, D. R. (2025). Disappearing people: A global demographic data
638 crisis threatens public policy. Science, 388(6753), 1277-1280.

639 González-Leonardo, M., López-Gay, A., Newsham, N., Recaño, J., & Rowe, F. (2022). Understanding
640 patterns of internal migration during the COVID-19 pandemic in Spain. Population, Space and Place,
641 28(6), e2578.

642 Lazar, A. N., Boo, G., Chamberlain, H. R., Nnanatu, C. C., Darin, E., Leasure, D. R., Yankey, O.,
643 Gadiaga, A., Juran, S., & de la Rua, L. (2025). Advances in Small Area Population Estimation in the
644 Absence of National Census Data. Preprints, <https://doi.org/10.20944/preprints202506.1625.v2>
645 [PREPRINT]

646 Leasure, D. R., Jochem, W. C., Weber, E. M., Seaman, V., & Tatem, A. J. (2020). National population
647 mapping from sparse survey data: A hierarchical Bayesian modeling framework to account for
648 uncertainty. Proceedings of the National Academy of Sciences, 117(39), 24173-24179.
649 <https://doi.org/10.1073/pnas.1913050117>

650 Leyk, S., Gaughan, A. E., Adamo, S. B., de Sherbinin, A., Balk, D., Freire, S., Rose, A., Stevens, F. R.,
651 Blankespoor, B., Frye, C., Comenetz, J., Sorichetta, A., MacManus, K., Pistoletti, L., Levy, M., Tatem,
652 A. J., & Pesaresi, M. (2019). The spatial allocation of population: a review of large-scale gridded
653 population data products and their fitness for use. Earth Syst. Sci. Data, 11(3), 1385-1409.
654 <https://doi.org/10.5194/essd-11-1385-2019>

655 Makama, S. i. D. (2007). Report of Nigeria's National Population Commission on the 2006 census.
656 Population and development Review, 33(1), 206-211.

657 Mallapaty, S. (2024). Three-year delay for Indian census frustrates scientists. Nature, 631, 718-719.

658 Marivoet, W., & De Herdt, T. (2018). Tracing down real socio-economic trends from household data
659 with erratic sampling frames: the case of the Democratic Republic of the Congo. Journal of Asian
660 and African Studies, 53(4), 532-552.

661 Minghini, M., Thabit Gonzalez, S., & Gabrielli, L. (2024). Pan-European open building footprints:
662 analysis and comparison in selected countries. The International Archives of the Photogrammetry,
663 Remote Sensing and Spatial Information Sciences, 48, 97-103.

664 Ministry of Fisheries and Livestock, and Central Statistical Office. (2019). The 2017/18 Livestock and
665 Aquaculture Census: Summary Report. [https://www.zamstats.gov.zm/wp-](https://www.zamstats.gov.zm/wp-content/uploads/2023/12/The-2017-18-Livestock-Aquaculture-Census-Summary-Report-1.pdf)
666 [content/uploads/2023/12/The-2017-18-Livestock-Aquaculture-Census-Summary-Report-1.pdf](https://www.zamstats.gov.zm/wp-content/uploads/2023/12/The-2017-18-Livestock-Aquaculture-Census-Summary-Report-1.pdf)

667 Neal, I., Seth, S., Watmough, G., & Diallo, M. S. (2022). Census-independent population estimation
668 using representation learning. Scientific Reports, 12(1), 5185.

669 Nieves, J. J., Bondarenko, M., Kerr, D., Ves, N., Yetman, G., Sinha, P., Clarke, D. J., Sorichetta, A.,
670 Stevens, F. R., & Gaughan, A. E. (2021). Measuring the contribution of built-settlement data to global
671 population mapping. Social Sciences & Humanities Open, 3(1), 100102.

672 Nnanatu, C. C., Bonnie, A., Joseph, J., Yankey, O., Cihan, D., Gadiaga, A., Voepel, H., Abbott, T.,
673 Chamberlain, H. R., & Tia, M. (2025b). Estimating small area population from health intervention
674 campaign surveys and partially observed settlement data. Nature Communications, 16(1), 4951.

675 Nnanatu, C. C., Yankey, O., Dzossa, A. D., Abbott, T., Gadiaga, A., Lazar, A., & Tatem, A. (2025a).
676 Efficient Bayesian Hierarchical Small Area Population Estimation Using INLA-SPDE: Integrating
677 Multiple Data Sources and Spatial-Autocorrelation. Preprints, 2025010588.
678 <https://doi.org/10.20944/preprints202501.0588.v1> [PREPRINT]

679 Office for National Statistics. (2024). ONS working paper series number 37 - Geospatial methods for
680 Small Area Population Estimates: proof of concept. ONS website.
681 <https://www.ons.gov.uk/methodology/methodologicalpublications/generalmethodology/onsworkin>
682 [gpaperseries/geospatialmethodsforsmallareapopulationestimatesproofofconcept](https://www.ons.gov.uk/methodology/methodologicalpublications/generalmethodology/onsworkin)

683 O'Hare, W. P. (2017). An international perspective on the undercount of young children in the US
684 Census. Statistical Journal of the IAOS, 33(2), 289-304.

685 Oostwegel, L. J., Schorlemmer, D., & Guéguen, P. (2025). From Footprints to Functions: A
686 Comprehensive Global and Semantic Building Footprint Dataset. Scientific Data, 12(1), 1699.

687 Pelletier, F., Hertog, S., Hovy, B., Schmid, K., Swiaczny, F., Wilmoth, J., & Zhang, J. G. (2020). Census
688 counts, undercounts and population estimates: The importance of data quality evaluation.
689 Population Division, United Nations Department of Economic and Social Affairs. Technical Paper
690 No. 2.

691 Pesaresi, M., Schiavina, M., Politis, P., Freire, S., Krasnodębska, K., Uhl, J. H., Carioli, A., Corbane,
692 C., Dijkstra, L., & Florio, P. (2024). Advances on the Global Human Settlement Layer by joint
693 assessment of Earth Observation and population survey data. International Journal of Digital Earth,
694 17(1), 2390454.

695 Pinney, A. (2012). An Afghan population estimation. Snapshots of an Intervention. The Unlearned
696 Lessons of Afghanistan's Decade of Assistance, 11.

697 Price, R., & Hallas, M. (2019). Mapping every building and road in sub-Saharan Africa. AGU fall
698 meeting abstracts,

699 Rowe, F., Calafiore, A., Arribas-Bel, D., Samardzhiev, K., & Fleischmann, M. (2023). Urban exodus?
700 Understanding human mobility in Britain during the COVID-19 pandemic using Meta-Facebook data.
701 Population, Space and Place, 29(1), e2637.

702 Rubinyi, S., Blankespoor, B., & Hall, J. W. (2021). The utility of built environment geospatial data for
703 high-resolution dasymetric global population modeling. Computers, Environment and Urban
704 Systems, 86, 101594.

705 Sanchez-Cespedes, L. M., Leasure, D. R., Tejedor-Garavito, N., Amaya Cruz, G. H., Garcia Velez, G.
706 A., Mendoza, A. E., Marín Salazar, Y. A., Esch, T., Tatem, A. J., & Ospina Bohórquez, M. (2024). Social
707 cartography and satellite-derived building coverage for post-census population estimates in
708 difficult-to-access regions of Colombia. Population Studies, 78(1), 3-20.

709 Sirko, W., Brempong, E. A., Marcos, J. T., Annkah, A., Korme, A., Hassen, M. A., Sapkota, K., Shekel,
710 T., Diack, A., & Nevo, S. (2023). High-resolution building and road detection from Sentinel-2. arXiv
711 preprint arXiv:2310.11622.

712 Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015). Disaggregating Census Data for
713 Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. PLOS ONE,
714 10(2), e0107042. <https://doi.org/10.1371/journal.pone.0107042>

715 UN Statistics Division. (2025). Results from the Third UNSD Survey on the 2020 Round of Population
716 and Housing Censuses and Review of Census Topics Collected in the 2020 Round.
717 [https://unstats.un.org/UNSDWebsite/statcom/session_56/documents/BG-3b-](https://unstats.un.org/UNSDWebsite/statcom/session_56/documents/BG-3b-Survey_Report_and_Analysis_of_Topics-E.pdf)
718 [Survey_Report_and_Analysis_of_Topics-E.pdf](https://unstats.un.org/UNSDWebsite/statcom/session_56/documents/BG-3b-Survey_Report_and_Analysis_of_Topics-E.pdf)

719 Wardrop, N. A., Jochem, W. C., Bird, T. J., Chamberlain, H. R., Clarke, D., Kerr, D., Bengtsson, L.,
720 Juran, S., Seaman, V., & Tatem, A. J. (2018). Spatially disaggregated population estimates in the
721 absence of national population and housing census data. Proceedings of the National Academy of
722 Sciences, 115(14), 3529-3537. <https://doi.org/10.1073/pnas.1715305115>

723 Weber, E. M., Seaman, V. Y., Stewart, R. N., Bird, T. J., Tatem, A. J., McKee, J. J., Bhaduri, B. L., Moehl,
724 J. J., & Reith, A. E. (2018). Census-independent population mapping in northern Nigeria. Remote
725 Sensing of Environment, 204, 786-798.

726 WorldPop. (2020). Bottom-up gridded population estimates for Zambia, version 1.0.
727 <https://doi.org/10.5258/SOTON/WP00662>

728 Zambia Statistics Agency. (2023). Zambia 2022 Census of Population and Housing Preliminary
729 Report. [https://www.zamstats.gov.zm/wp-content/uploads/2023/12/2022-Census-of-Population-](https://www.zamstats.gov.zm/wp-content/uploads/2023/12/2022-Census-of-Population-and-Housing-Preliminary.pdf)
730 [and-Housing-Preliminary.pdf](https://www.zamstats.gov.zm/wp-content/uploads/2023/12/2022-Census-of-Population-and-Housing-Preliminary.pdf)

731 WorldPop and Zambia Statistics Agency. (2024). Census disaggregated gridded population
732 estimates for Zambia, based on the 2022 Census of Population and Housing preliminary results,
733 version 2.0. <https://doi.org/https://doi.org/10.5258/SOTON/WP00783>

734 Zambia Statistics Agency. (2024). Zambia 2022 Census of Population and Housing Summary Report:
735 Volume Two (Revised).
736 [https://www.zamstats.gov.zm/Publications/Revised%202022%20Census%20of%20Population%20](https://www.zamstats.gov.zm/Publications/Revised%202022%20Census%20of%20Population%20and%20Housing%20Summary%20Report%20Volume%20II.pdf)
737 [and%20Housing%20Summary%20Report%20Volume%20II.pdf](https://www.zamstats.gov.zm/Publications/Revised%202022%20Census%20of%20Population%20and%20Housing%20Summary%20Report%20Volume%20II.pdf)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AppendixASIZMBpaperSEPT2026.pdf](#)