Release Statement

Modelled gridded population estimates for Nigeria 2025 (version 3.0)

29 August 2025

Abstract

This data release provides gridded population estimates (at spatial resolution of 3 arc-seconds, approximately 100-metre grid cells) for Nigeria, along with the estimates of the number of people belonging to various age and sex groups. Using robust Bayesian statistical hierarchical modelling framework, population modelling and estimation experts from WorldPop (www.worldpop.org) at the University of Southampton combined 'head count' (input population) datasets obtained from the 2022-23 National Malaria Elimination Program (NMEP) with settlement footprint and geospatial covariates to estimate population numbers at high-resolution grid cells. The approach facilitated accounting for the multiple levels of variability within the data, while simultaneously quantifying uncertainties in the parameter estimates. After capturing the spatial variability of population, the modelled estimates were scaled based on the UN WPP July 2025 median national population projections.

These data were produced by the WorldPop Research Group at the University of Southampton as part of the GRID3 – Phase 2 Scaling project, with funding from the Bill and Melinda Gates Foundation (INV-044979). Project partners included the GRID3 Inc., the Center for Integrated Earth System Information (CIESIN) within the Columbia Climate School at Columbia University, and WorldPop at the University of Southampton. The final statistical modelling was designed, developed, and implemented by Chris Nnanatu supported by Assane Gadiaga. Data processing was done by Assane Gadiaga with additional support from Attila Lazar, Tom Abbott and Heather Chamberlain. Project oversight was done by Attila Lazar and Andy Tatem. The NMEP shared household bednet distribution data along with the location of the households. The settlement footprint data was prepared and shared by CIESIN.

The authors followed rigorous procedures designed to ensure that the used data, the applied method and thus the results are appropriate and of reasonable quality. If users encounter apparent errors or misstatements, they should contact WorldPop at release@worldpop.org.

shall not be liable for incidental, consequential, or special damages arising out of the use of any data that they offer. These data are operational population estimates and are not official government statistics.

RELEASE CONTENT

1. NGA_population_v3_0_gridded.zip
2. NGA_population_v3_0_agesex.zip
3. NGA_population_v3_0_table.zip
4. NGA_population_v3_0_mastergrid.tif

LICENSE

SUGGESTED CITATIONS

Nnanatu C.C., Gadiaga A., Abbott T. J., Chamberlain H., Lazar A. N., Tatem A. J. (2025). Modelled gridded population estimates for Nigeria 2025 version 3.0. *WorldPop, University of Southampton*. (https://dx.doi.org/10.5258/SOTON/WP00782)

FILE DESCRIPTIONS

The projection for all GIS files is the geographic coordinate system WGS84 (World Geodetic System 1984). Kindly note that while this data represents population counts, values contain decimals, i.e. fractions of people. This is because both the input population data and age-sex proportions contain decimals. For this reason, it is advised to aggregate the rasters at a coarser scale. For example, if four grid cells next to each other have values of 0.25 this indicates that there is 1 person somewhere in those four grid cells.

NGA_population_v3_0_gridded.tif

This geotiff raster contains estimates of total population size for each approximately 100-metre grid cell (0.0008333 decimal degrees grid) across Nigeria. The values are the scaled mean of the posterior probability distribution for the predicted population size in each grid cell. Any NA values within the national boundary will represent areas that were mapped as unsettled according to building footprints data, while any other NA values represent areas mapped as being outside Nigerian boundary.

NGA_population_v3_0_agesex.zip

This zip file contains 40 geotiff rasters at a spatial resolution of 3 arc-seconds (approximately 100-metre grid cells). Each raster provides gridded population estimates for an age-sex group per grid cell across Nigeria. We provide 36 rasters for the commonly reported age-sex groupings of sequential age classes for males and females separately.

These are labelled with either an “m” (male) or an “f” (female) followed by the number of the first year of the age class represented by the data. “f0” and “m0” are population counts of under 1-year olds for females and males, respectively. “f1” and “m1” are population counts of 1- to 4-year-olds for females and males, respectively. Over 4 years old, the age groups are in five-year bins labelled with a “5”, “10”, etc. Eighty-year-olds and over are represented by the groups “f80” and “m80”. We provide four additional rasters that represent demographic groups often targeted by programmes and interventions. These are “under1” (all females and males under the age of 1), “under5” (all females and males under the age of 5), “under15” (all females and males under the age of 15) and “f15_49” (all females between the ages of 15 and 49, inclusive).

NGA_population_v3_0_table.zip

This contains three CSV files, namely lga_pop_total_scaled.csv, state_pop_total_scaled.csv and national_pop_total_scaled.csv. lga_pop_total_scaled.csv contains population totals at the local government area level. The state_pop_total_scaled.csv contains population totals for each state. The national_pop_total_scaled.csv contains total population size according to the UN WPP 2025 projections for Nigeria.

NGA_population_v3_0_mastergrid.tif

This geotiff raster, at a spatial resolution of 3 arc-seconds (approximately 100 m), contains 1s for each grid cell that is classified as settled based on the CIESIN settlement raster, whereas contains 0s for pixels that are identified as non-settled by CIESIN and within the national border. NAs represent grid cells considered to be outside Nigeria.

RELEASE HISTORY

Version 3.0 (29 August 2025, https://dx.doi.org/10.5258/SOTON/WP00782)

- This gridded population estimates were produced using recent NMEP household listing data and CIESIN settlement information. The scaled population estimate is adjusted to the July 2025 median UN WPP projection.

Version 2.1 (19 July 2023, https://dx.doi.org/10.5258/SOTON/WP00765)

- The population map was updated using amended settlement data due to a change in the feature extraction algorithm.
- Totals at ward level are the same as previous version therefore they are representative of the year 2019.

Version 2.0 (17 November 2021, https://dx.doi.org/10.5258/SOTON/WP00729)

- Refinement of gridded population estimates using more recent settlement data based on building footprints.

- Predictions of residential and non-residential buildings incorporated in the settlement map.
- A different regional boundary definition was used in the model, corresponding with Nigerian statistical regions.
- Representative of the year 2019.

Version 1.2 (15 September 2020)

- A peer-reviewed article (Leasure et al., 2020b) was added to describe the statistical method that were developed to produce the population estimates (https://doi.org/10.1073/pnas.1913050117).

Version 1.2 (20 May 2020)

- Gridded population estimates were added to NGA_population_v1_2_agesex.zip for the following demographic groups: children under 1, children under 5, children 15, and women 15 to 49 years of age.

Version 1.2 (26 mars 2020, https://dx.doi.org/10.5258/SOTON/WP00661)

- Gridded population estimates were added for individual age-sex groups (NGA_population_v1_2_agesex.zip).
- The SQL database "NGA_population_v1_2_sql.sql" that is used in WOPR applications was updated to remove unnecessary data (e.g. covariate values, names of administrative units).
- Population tiles were updated with a revised color palette. This file was renamed from "NGA_population_v1_2_tiles_population.zip" to "NGA_population_v1_2_tiles.zip".
- Uncertainty tiles "NGA_population_v1_2_tiles_uncertainty.zip" were removed because they were discontinued for use in WorldPop web applications (e.g. https://apps.worldpop.org/woprVision).

Version 1.2 (10 July 2019) [https://dx.doi.org/10.5258/SOTON/WP00655]

- The previous release contained a few grid cells with erroneously high population estimates that resulted from the way the statistical model was summarized (based on 1,000 samples from posterior predictions as opposed to 10,000 samples used here).
- This update changes the population estimates slightly in every grid cell. State and LGA totals have changed marginally but remain within 1% of previous estimates.
- Representative of the time period from 2016 to 2017.

Version 1.1 (22 February 2019) [https://dx.doi.org/10.5258/SOTON/WP00657]

- Updated to include floating-point rasters rather than integer rasters to resolve rounding errors when calculating population totals for larger areas (e.g. zonal sums)

Version 1.0 (11 November 2018) [https://dx.doi.org/10.5258/SOTON/WP00656]

- Original release of Nigeria population dataset

ASSUMPTIONS AND LIMITATIONS

The NMEP household listing datasets containing counts of people per household across nine (9) spatial units at Admin 1, namely, Adamawa, Delta, Kaduna, Kano, Katsina, Kwara, Niger, Osun, Taraba, were totally anonymised due to confidentiality issues. Then, the NMEP data were cleaned and aggregated to ward level (Admin 3) to address potential spatial inaccuracies and thus to ensure more accurate population estimates. The ward level formed the statistical model parameters training units. Other associated datasets were also aggregated to the ward level including the geospatial covariates and the human settlement datasets. The means of the gridded continuous covariates across all the grid cells within a given ward were obtained as the corresponding ward-level value, while building counts were aggregated as the sum of all buildings across all the grid cells within a given ward. Except the World Settlement Footprint-derived information on the evolution of settlement for the period 2005-2015, the geospatial covariates were collected from different sources at different time periods (see ‘Geospatial Covariates’ Section below for more details) between 2021 and 2023 to match the NMEP data collection period, and these were used in conjunction with the CIESIN settlement datasets that represent the year 2024 for the population modelling.

The final population model results are scaled to the UN WPP July 2025 population projection value; therefore, we assume that the bottom-up model results accurately represent the spatial distribution of the population, and the UN projection accurately represents the overall total population of Nigeria.

Age-sex structure data for the year 2022 that were obtained from the National Population Commission (NPC, 2020) were used to disaggregate the scaled gridded population totals into age groups and gender breakdown. Since 2022 is the last available subnational age/sex information from NPC, we assume that the state-level population pyramids have not changed between 2022 and 2025. Subnational age/sex projections from NPC were only available at state-level, therefore, the grid cells within individual states have identical age and sex proportions.

Because the estimation methodology locates people only in settled areas as determined primarily by satellite-based identification of human settlements, its accuracy may be susceptible to volatility in settlement patterns. Areas that have been settled or abandoned too recently to be reflected in the satellite imagery data used to map building footprints (e.g.., through displacement), will not be accurately estimated. Other sources of population dynamics such as seasonal migration are also not fully captured in the model. However, information on the evolution of settlement. and the recent demographic datasets utilised within the model provide an opportunity to capture recent changes in the population density and distribution thereby minimizing the potential impacts of the aforementioned settlement patterns volatility.

Grid Cell Alignment

At the start of population modelling, it is important to ensure that grid cells of the spatial extents of the raster files of the input and output datasets align with the extents of the study location of interest achieved through the use of the mastergrid of the study location. For the current project, please note that the mastergrid used for this version (v3.0) differs from the one used in the previous versions of gridded population estimates for NGA (v1.0, 1.1, 1.2, 2.0, and 2.1) and other previous WorldPop data products. The current mastergrid is based on the CIESIN settlement data extents that is aligned with the new WorldPop Global 2 mastergid for consistency that will be used for any future demographic maps by WorldPop.

We used the Large Scale International Boundaries (LSIB) national boundary for Nigeria. Because the extent of the CIESIN settlement layer contains a few kilometres buffer around the national boundary, it is thus slightly larger than the LSIB Nigeria boundary. There are 1024 settled pixels that were cut-off along the national boundary from the CIESIN settled raster. Furthermore, there are two Wards where the CIESIN settled raster does not show settled pixels, and therefore the total estimated population of the wards is zero. These wards are Okpokwu, in Obi LGA (Local Government Areas) of Benue state; and Akpankanya, in Bakassi LGA of Cross River state.

SOURCE DATA

The key datasets used to produce the modelled population estimates are:

NMEP Data

The 2022-23 NMEP data used for the population modelling was malaria bednet campaign data that covered nine states (see details in Table 1). The NMEP data provided information on the number of bednets that were issued per household as well as count of people living in each household. The year of the data collection as well as the average people per household for the nine states are listed in Table 1.

Table 1: Average people per household

States	Year	Average people per household
Adamawa	2023	5.69
Delta	2022	5.27
Kaduna	2022	4.99
Kano	2022	5.68
Katsina	2022	5.55
Kwara	2023	6.08
Niger	2022	5.02
Osun	2023	5.57
Taraba	2022	5.10

Source: NMEP, 2022 – 2023.

Geographic coordinates of the households were provided, but the datasets were anonymised, cleaned and aggregated at the ward level to overcome potential spatial inaccuracies related to the spatial precision of the household points on one hand, and achieve the granularity required for more accurate estimates, on the other hand. The NMEP data assumes complete enumeration of the individuals living in each of the states. However, exploratory preliminary analysis carried out on the NMEP data indicated that there were missing household points in some states, as well as household points with coordinates outside their respective states. As part of data cleaning processes, household points that fell outside their respective state and wards that showed potential spatial anomalies were excluded from the analysis to avoid biases and inflation of the input population data.

Settlement Data

The settlement data was provided by CIESIN in the form of raster files (CIESIN, 2024) with a reference year of 2024. We obtained three different settlement products, namely (i) building area, which indicates the area of the buildings whose centroids are within a grid cell; (ii) building count, which is the number of buildings within a given cell; and (iii) probability of settlement, which gives a probability value of a grid cell to be settled. Following preliminary modelling and analyses using the three gridded layers obtained from CIESIN, the building count settlement layer was selected as the best layer that best described population density and distribution in the context of Nigeria. Please note that the use of either settled area or building count in defining population density is often context specific.

Geospatial Covariates

A wide variety of other geospatial covariates, which are related to population density and distribution were considered for the statistical modelling. These geospatial covariates include land use and land cover data, climate variables such as temperature and rainfall, physical features and infrastructure such as roads and schools, and conflict data (Woods et al. 2024). All continuous geospatial covariates were scaled using z-score by dividing the difference between the covariate value x_k and the mean $\underline{x}_k$ by the corresponding standard deviation σ_x, i.e., $x_k^{(scaled)} = (x_k - \underline{x}_k)/\sigma_x$. This ensures that the regression coefficients of the continuous geospatial covariates which were obtained using different units of measurements are comparable. Population model covariates were selected using a generalized linear model (GLM) – based stepwise selection method (McCullagh and Nelder, 1989). The selected covariates were further accessed for multi-collinearity and statistical significance. Eventually, of the 55 geospatial covariates initially tested, 12 were retained as the best fit covariates with variance inflation factor (VIF) of less than 5. However, further model and prediction checks led to the retaining of the final six (6) covariates described in Table 2.

Table 2. List of the finally selected geospatial covariates for the final model training.

Description	**Source**	**Resolution**	**URL/DOI**
Distance to inland water	ESA CCI Land Cover v2.0.7 & v2.1.1	100m	DOI:10.5258/SOTON/WP00772
Distance to cropland, natural vegetation	ESA CCI Land Cover v2.0.7 & v2.1.1	100m	DOI:10.5258/SOTON/WP00772
Annual average precipitation	TerraClimate	100m	DOI:10.5258/SOTON/WP00772
Annual average temperature	Terra MODIS LST	100m	DOI:10.5258/SOTON/WP00772
Distance to IUCN strict nature reserve and wilderness area edges 2015-2022	UNEP-WCMC & IUCN (2023)	100m	DOI:10.5258/SOTON/WP00772
Settlement growth index 2005-2015	World Settlement footprint	100m	https://download.geoservice.dlr.de/WSF_EVO/

Notes: ESA: European Space Agency, CCI: Climate Change Initiative, MODIS: Moderate Resolution Image Spectroradiometer, LST: Land Surface Temperature, UNEP: United Nation Environmental Program, WCMC: World Conservation Monitoring Centre, IUCN: International Union for Conservation of Nature

UN WPP projection

We used the median July 2025 United Nations World Population Projection (UN WPP) for Nigeria (237,527,782) to scale the bottom-up population model results (https://population.un.org/wpp/downloads?folder=Standard%20Projections&group=Population).

Age-Sex Proportions

We used the National Population Commission (NPC) state population projections dataset by age and sex (NPC, 2020) to calculate the age-sex proportions for Nigeria. The latest available year was 2022.

LSIB national boundary

We used the Large Scale International Boundaries (LSIB) dataset by the U.S. Department of State to define the national boundary of Nigeria before the scaling was applied (https://geodata.state.gov/geonetwork/srv/eng/catalog.search#/home).

METHODS OVERVIEW

The key steps of our approach were as follows:

- Data collation, preparation, exploratory analyses and data cleaning. This includes summarizing the household sizes from the NMEP dataset to get the total population at the ward level.
- Geospatial covariates were subjected to robust covariate selection for model training and parameter estimation.
- Development of robust Bayesian hierarchical statistical models implemented within the integrated nested Laplace approximation techniques in conjunction with the stochastic partial differential equations strategies (INLA-SPDE; Rue et al, 2009; Lindgren et al. 2011). These were used in training model parameters.
- Population estimates were predicted at grid cell level using the corresponding grid cell values of the covariates of the best fit model produced using the training data sets.
- The modelled population estimates were scaled to the UN WPP median July 2025 projection
- The total scaled population is disaggregated to age and sex classes using the NPC 2022 subnational projections.

Statistical Modelling

All data processing, and statistical modelling and analyses were carried out using R version 4.4.2 (R Core Team, 2023) with the Bayesian hierarchical modelling implemented using INLA package version ‘24.12.11’ (Rue et al. 2009). Modelled estimates of population were produced using the of bottom-up population modelling framework (Wardop et al., 2018) which have been implemented using Markov chain Monte Carlo (MCMC) strategies (Leasure et al., 2020; Boo et al. 2022 ; Darin et al.,2022) as well as the optimised INLA-SPDE schemes recently used to produce small area population estimates for Papua New Guinea (WorldPop and NSO PNG, 2022; Nnanatu et al., 2025a), Cameroon (Nnanatu et al. 2022; Nnanatu et al.,2025b) and the Democratic Republic of Congo (e.g., Nnanatu et al., 2024).
Overall, there are four (4) key steps involved in the statistical modelling processes utilised here (Figure 1).

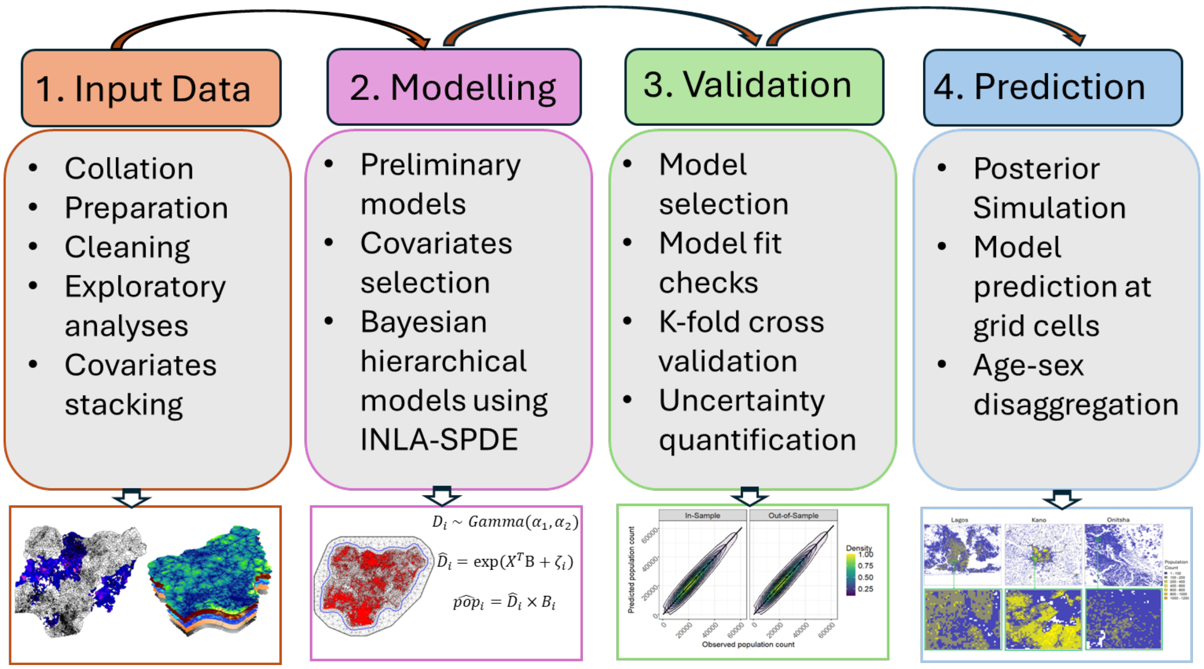

Figure 1. INLA-SPDE based small area population modelling and estimation workflow.

The first step involved data collation from the various sources including the input population data from NMEP, the human settlement data from CIESIN and a stack of geospatial covariates from WorldPop and other sources. These datasets were rigorously explored, cleaned and prepared for statistical modelling. The exploratory analyses included checks on the distribution of the datasets as well as for anomaly/outlier detection across all the input datasets. As part of the data cleaning processes, the NMEP household points data was converted into a raster file using the same spatial extents as the CIESIN settlement data at 100m-by-100m resolution. This provided a robust approach for the detection of inconsistent and unrealistic values within the datasets. Input NMEP population data was available for only 2,589 (~28%) wards from 9 states. A total of 477 wards out of 2,589, had significantly higher population counts compared to the estimated number of buildings, and were therefore excluded (dropped) from the modelling

In the second step, preliminary models were tested to identify the best predictors (geospatial covariates) of population density using GLM-based stepwise regression with ‘both’ forward and backward selection algorithms. Only statistically significant covariates with a variance inflation factor (VIF) of less than 5 were retained and used within the INLA-SPDE-based Bayesian hierarchical regression modelling. An additional layer of model selection was carried out after incorporating random effects (e.g., settlement type, spatial autocorrelation and spatially independent) to the models in the third step.

Model selection for the Bayesian hierarchical regression modelling relied on the deviance information criterion (DIC) values while the selected model predictive ability was examined using the mean absolute error (MAE), the root mean square error (RMSE) and the correlation coefficient (CC). Smaller values of DIC, MAE and RMSE indicate better fit and predictive ability while larger values of CC indicate better predictive ability. These model performance metrics were also used within k-fold cross-validation where the data was first divided into two with the model parameters trained with 80% of the data while the remaining 20% was used as a test to predict population density. This was repeated 10 times (10-fold) whilst ensuring that none of the test samples was repeated. The primary aim of the cross-validation is to test how well the best fit model parameters were able to predict population outside the observed locations. To more accurately capture the performance of the model, we carried out in-sample and out-of-sample cross validations. In the in-sample cross-validation, all the data points were used in the training set but 20% of the data points were used as a test set to predict their population density. Whereas in the out-of-sample cross validation, the 20% test set was excluded completely from the 80% training set.

Finally, further posterior inference and grid cell prediction are carried out in the fourth step. The prediction at the grid cell uses the model parameters of the best fit training set model to predict population counts at 100m-by-100m grid cells across Nigeria using the corresponding grid cell values of the geospatial covariates and the building counts. The predicted population counts are then further disaggregated by age and sex groups using Nigerian age-sex proportion table from the NPC (NPC, 2020).

Model Specification

In general, the population count pop_i at a given (ideally geolocated) area unit is assumed to be Poisson-distributed, such that $pop_i \sim Poisson(\lambda_i)$. However, in the context of small area population modelling (Leasure et al., 2020; Boo et al. 2022 ; Darin et al.,2022; Nnanatu et al., 2022; Nnanatu et al., 2025a; Nnanatu et al., 2025b; Nnanatu et al., 2024), a key assumption of the Poisson model which requires both mean and variance to be equal is often violated due to overdispersion in which case $mean(pop_i) \neq var(pop_i)$. For this reason, the mean parameter λ_i is usually expressed in terms of population density to account for spatial aggregation error (e.g., Leasure et al 2020, Nnanatu et al 2022). Typically, the mean parameter is given as $\lambda_i = \mu_i B_i$, where B_i is the total number of buildings within ward i and

$$D_i = \frac{pop_i}{B_i} \qquad (1)$$

is the population density defined as the number of people per settlement building with expected value equals μ_i, and which follows Gamma distribution given by

$$D_i \sim Gamma(\alpha_1, \alpha_2) \qquad (2)$$

where α_1 and α_2 are the shape and rate parameters with mean $\mu_i = \alpha_1/\alpha_2$ and variance $\phi = \alpha_1/\alpha_2^2$, respectively. The predicted population density $\widehat{D}_i$ for ward i is given by

$$\widehat{D}_i = exp\ exp\ (X_i^T \beta + Z_i^T \gamma + \xi(s_i) + \zeta_i) \qquad (3)$$

where X and Z are the design matrices of fixed effect covariates (e.g., average annual precipitation, average annual temperature, distance to crop land) and random effects (e.g., settlement type, state, LGA), respectively. Also, the terms $\beta \in R^{(K\times 1)}$ and γ are the vectors of fixed effects regression parameters and random effects variances, respectively. While the terms $\xi(s_i)$ and ζ_i are the spatially varying and spatially independent random effects accounting for spatial autocorrelations and dissimilarities between observations, respectively. Here, $\xi(s_i)$ at spatial location s_i follows a Gaussian Random Field (GRF)

$$\xi(s_i) \sim GRF(0, \Sigma) \qquad (4)$$

where Σ is a dense covariance matrix. Here, we evaluated Σ using the INLA-SPDE approach via Gaussian Markov Random Field (GMRF) and gained computational speed by first discretising the continuous spatial domain using mesh (Lindgren et al., 2011). Additionally, the term ζ_i is a zero-mean Gaussian noise specified by

$$\zeta_i \sim Normal\left(0, \sigma_\zeta^2\right) \qquad (5)$$

where $\sigma_\zeta^2 > 0$ is a variance parameter. Then, finally, the predicted population counts at grid cell g is obtained as

$$\widehat{pop}_g = \widehat{D}_g \times B_g \qquad (6)$$

where $\widehat{D}_g$ is the predicted population density in gride cell g using the corresponding grid cell covariate values and the model parameter values based on equation (3). Also, B_g is the corresponding building count for grid cell g $(g = 1, \dots, G)$. The prediction covariates included $G(= 7{,}185{,}917\ )$ grid cells at 100m-by-100m resolution, and population counts were predicted in each grid cell that contains values of building counts.

All models were implemented within the integrated nested Laplace approximation (INLA; Rue et al, 2009) in conjunction with the stochastic partial differential equation (SPDE Lindgren et al, 2011) frameworks. It allowed us to gain more computational advantage by discretizing the entire Nigerian continuous space into a Gaussian Markov random fields (GMRF) process. To ensure flexibility and better capture local variabilities within the data, we used the Penalized Complexity (PC) (Simpson et al., 2017) on the standard deviation parameters throughout, such that a small probability of 0.01 is assigned for the standard deviation σ being greater than 1, that is, $P(\sigma > 1) = 0.01$.

Model fit checks and model validation

Table 3 and Table 4 show the model fit metrics of the top three competing models and the cross-validation performance metrics of the best fit model, respectively. Table 3 provides the specifications and the DIC values of these nested models. It shows that Model 3 which accounted for only the spatially varying ($spatial$) and spatially independent

(IID) random effects (ward) along with the geospatial covariates provided the best fit with the lowest DIC value indicating that the variabilities across the states did not significantly influence population density in Nigeria. Instead, the spatial location of the wards matters, and the approach enabled us to borrow strength from neighboring wards/grid cells to predict population in nearby wards/grid cells with no population observations.

Table 3. Model fit metrics of the top 3 competing models

Model	Specification	DIC
Model 1	$log(\mu_i) \sim intercept + covariates + spatial + state$	6874.722
Model 2	$log(\mu_i) \sim intercept + covariates + state$	7493.952
Model 3	$log(\mu_i) \sim intercept + covariates + spatial + IID$	2590.333

Note. DIC – Deviance information criterion (the smaller, the better). The best fit model is Model 3 (highlighted in light green).

Additionally, as shown in Table 4, the 10-fold cross validation carried out further indicate good fit and high predictive ability of the best fit model at both in-sample and out-of-sample test versus training set samples. The values of the MAE and RMSE were close with high correlation coefficients of more than 80% in both scenarios.

Table 4. Model fit metrics for cross-validations

Data	**MAE**	**RMSE**	**CC (%)**
In-Sample	5212.302	9701.98	91.31
Out-of-Sample	7558.738	10625.97	84.56

Note. MAE – Mean Absolute Error; RMSE – Root Mean Square Error; CC – Correlation Coefficient.

The estimated total population using the bottom-up method above was 237,345,980 (95% CI; lower = 233,596,990 0.9131219; upper = 243,655,702).

Scaling

The scaling process involved normalising [0,1] the model-based predicted population counts to get their relative distribution per grid cell. Then, the normalised values were multiplied by the UN population totals to adjust the modelled gridded population estimates, but keep the spatial distribution. Using these adjusted gridded population counts, the predicted population totals at the different administrative levels were also updated.

Therefore, the final results at national level are matching the median July 2025 United Nations World Population Projection (UN WPP) for Nigeria (237,527,782).

Age-Sex disaggregation

We used the 2022 National Population Commission (NPC) subnational population projections dataset by age and sex (NPC, 2020) to calculate the age-sex proportions for each state across Nigeria. We multiplied our gridded population estimates (NGA_population_v3_0_gridded.tif) by the gridded age-sex proportions to produce NGA_population_v3_0_agesex.zip. Subnational age/sex projections from NPC were only available at state-level, therefore, the grid cells within individual states have identical age and sex proportions.

ACKNOWLEDGEMENTS

We thank the NPC and NMEP for providing access to projected population pyramid and anonymized household data collected during malaria ITN distribution campaigns, in accordance with the relevant data sharing agreements. The whole WorldPop group are acknowledged for overall project support. We thank the WorldPop's spatial statistics and population modelling (SSPM) and GRID3 teams for reviewing the data and providing thoughtful suggestions which helped to improve the modelled estimates.

References

Boo, G., Darin, E., Leasure, D. R., Dooley, C. A., Chamberlain, H. R., Lázár, A. N., ... & Tatem, A. J. (2022). High-resolution population estimation using household survey data and building footprints. *Nature communications*, *13*(1), 1330.

Center for Integrated Earth System Information (CIESIN). (2025). GRID3 NGA - Operational Wards - 20250220. Unpublished.

Center for Integrated Earth Science Information Network (CIESIN), Columbia University. (2024). GRID3 COD - Settlement Extents v3.0 alpha. Unpublished.

Darin, E., Kuépié, M., Bassinga, H., Boo, G., Tatem, A. J., & Reeve, P. (2022). The Population Seen from Space: When Satellite Images Come to the Rescue of the Census. *Population*, *77*(3), 437-464.

Leasure, D.R., Jochem, W.C., Weber, E.M., Seaman, V., & Tatem, A.J. (2020). High resolution population mapping with limited survey data: a hierarchical Bayesian modelling framework to account for uncertainty. *Proceedings of the National Academy of Sciences of the United States of America*, 117(39): 24173–24179.

Lindgren, F., Rue, H., & Lindström, J. (2011). 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(4), 423–498

McCullagh, Peter, and John A Nelder. (1989). *Generalized Linear Models*, 2nd Edition. Chapman; Hall/CRC

National Population Commission (2020). Nigeria population projections and demographic indicators. https://cdn.sanity.io/files/5otlgtiz/production/907db2f19eebad96152b17e9054584335642a33b.pdf; downloaded 4 December 2023

Nnanatu C.C., Yankey O., Abbott T. J., Assane, G., Lazar A. N., Darin E., Tatem A. J. (2022). *Bottom-up gridded population estimates for Cameroon (2022),* version 1.0. https://dx.doi.org/10.5258/SOTON/WP00784

Nnanatu et al. (2025a), Estimating small area population from health intervention campaign surveys and partially observed settlement data. *Nature Communications* 16, 4951. https://doi.org/10.1038/s41467-025-59862-4

Nnanatu C., Yankey O., Bonnie A., Abbott T. J., Chamberlain H., Lazar A. N., Tatem A. J. (2024). *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

Nnanatu, C. C., Yankey, O., Dzossa, A. D., Abbott, T., Gadiaga, A., Lazar, A., & Tatem, A. (2025b). Efficient Bayesian Hierarchical Small Area Population Estimation Using INLA-SPDE: Integrating Multiple Data Sources and Spatial-Autocorrelation. *Preprints.* https://doi.org/10.20944/preprints202501.0588.v1

R Core Team. (2023). R: *A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. https://www.R-project.org.

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the royal statistical society:Series b* (statistical methodology), 71(2), 319-392

Simpson, D. P., H. Rue, A. Riebler, T. G. Martins, and S. H. Sørbye. (2017). "Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors." *Statistical Science* 32 (1): 1–28.

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

Woods, D., McKeen, T., Cunningham, A., Priyatikanto, R., Sorichetta, A., Tatem, A. J., & Bondarenko., M., (2024). *WorldPop high resolution, harmonised annual global geospatial covariates. Version 1.0.* University of Southampton: Southampton, UK DOI:10.5258/SOTON/WP00772

WorldPop and National Statistical Office of Papua New Guinea. (2022). *Census-independent population estimates for Papua New Guinea (2020-21),* version 1.0. WorldPop, University of Southampton. DOI: 10.5258/SOTON/WP00763