Release Statement

Modelled gridded population estimates for Tshopo Province in the Democratic Republic of Congo version 4.2.

13 March 2025

Abstract

This data release provides gridded population estimates (spatial resolution of 3 arc-seconds, approximately 100-metre grid cells) for Tshopo Province in the Democratic Republic of Congo (DRC), along with estimates of the number of people belonging to various age-sex groups. The project team used the Pre-Distribution Registration Survey (PDRS) data from the National Malaria Control Programme (PNLP) collected as part of anti-malarial campaigns in the DRC for 2023, settlement extents and geospatial covariates to model and estimate population numbers at grid cell level using a Bayesian statistical hierarchical modelling framework. The approach facilitated simultaneous accounting for the multiple levels of variability within the data. It also allowed the quantification of uncertainties in parameter estimates. These model-based population estimates can be considered as most accurately representing the year 2023. This time period corresponds to the PDRS survey date for Tshopo. Although the methods were robust enough to explicitly account for key random biases within the datasets, it is noted that systematic biases, which may arise from sources other than random errors within the observed data collection process, are most likely to remain.

These data were produced by the WorldPop Research Group at the University of Southampton. This work was part of the GRID3 – Phase 2 Scaling project, with funding from the Gates Foundation (INV-044979). Project partners included 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 Mohamed Megheib with support from Ortis Yankey. Data processing was done by Mohamed Megheib and Tom Abbott with additional support from Heather Chamberlain. Project oversight was done by Attila Lazar, Chris Nnanatu and Andy Tatem. The PDRS data from the malaria insecticide treated net (ITN) distribution campaigns were collected, processed, anonymised and shared by the PNLP and its implementing partners. The settlement extent data was prepared and shared by CIESIN (2024). The data has been clipped to Grid3-CIESIN health area extent (CIESIN, 2025)

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.

WorldPop, University of Southampton, and their sponsors offer these data on a "where is, as is" basis; do not offer an express or implied warranty of any kind; do not guarantee the quality, applicability, accuracy, reliability or completeness of any data provided; and 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. COD_Tshopo_province_population_v4.2_gridded.zip
2. COD_ Tshopo_province_population_v4.2_agesex.zip

LICENSE

These data may be redistributed following the terms of a Creative Commons Attribution 4.1 International (CC BY 4.1) license.

SUGGESTED CITATION

Megheib M, Yankey O., Nnanatu C., Abbott T. J., Chamberlain H., Lazar A. N., Tatem A. J. 2025. Bottom-up gridded population estimates for Tshopo Province in the Democratic Republic of Congo (2023), version 4.2. *WorldPop, University of Southampton*. doi: https://dx.doi.org/ 10.5258/SOTON/WP00786

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.

COD_Tshopo_province_population_v4_2_gridded.tif

This geotiff raster contains estimates of total population size for each approximately 100-metre grid cell (0.0008333 decimal degrees grid) across Tshopo Province. The values are the mean of the posterior probability distribution for the predicted population size in each grid cell. Grid cells with values of 0 represent areas that were mapped as unsettled according to building footprints data.

COD_ Tshopo_province_population_v4_2_lower.tif

This geotiff raster contains estimates of the lower bound credible interval (2.5% CI) for each grid cell across Tshopo. The values are the 2.5% posterior probability distribution for the predicted population size in each grid cell. The lower bound estimates cannot be summed across grid cells to produce a lower credible interval measure for a multi-cell area. Grid cells with values of 0 represent areas that were mapped as unsettled according to building footprints data.

COD_ Tshopo_province_population_v4_2_upper.tif

This geotiff raster contains estimates of the upper bound credible interval (97.5% CI) for each grid cell across Tshopo. The values are the 97.5% posterior probability distribution for the predicted population size in each grid cell. The upper bound estimates cannot be summed across grid cells to produce an upper bound credible interval measure for a multi-cell area. Grid cells with values of 0 represent areas that were mapped as unsettled according to building footprints data.

COD_ Tshopo_province_population_v4_2_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 Tshopo. 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 older 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). These data were produced using age-sex proportions from the 2024 WorldPop Global subnational population pyramids for the DRC. The age-sex proportions are available per a given province. Hence we applied the age-sex proportions for Tshopo to the gridded population estimates (COD_ Tshopo_Province _population_v4_2_gridded.tif) to allocate the population to the different age-sex classes.

RELEASE HISTORY

Version 4.2 (13 March 2025)

- This is the original release of the data for Tshopo Province [doi: 10.5258/SOTON/WP00786] (as described in this release statement).
- This data release utilizes operational National Malaria Control Programme data, composite openly accessible building footprint datasets and a new mastergrid.
- This data is released as part of a collection of population estimates for 11 DRC provinces: https://wopr.worldpop.org/?COD/Population/v4.2

ASSUMPTIONS AND LIMITATIONS

These population estimates most likely represent the 2023 time, but because of the different ages of the input data used to build the model, a precise time point cannot be allocated. The PDRS data that was used as the response variable was collected in 2023, while geospatial covariates data were collected from different time periods between 2020 and 2023. Similarly, the CIESIN settlement layers were produced in 2024. The inherent heterogeneity in the temporal alignment of these datasets used in the modelling may introduce uncertainties and potential inaccuracies in the modelling process.

Data on population per household (household size), collected during ITN distribution campaigns, was aggregated to calculate total population count for a given spatial unit. Given that the number of ITNs received by a household is proportional to the household size, there is an incentive for respondents to potentially inflate counts of population per household. The presence of inflated household sizes in the input population data would likely introduce systematic biases in the modelled estimates.

The statistical model predicted unrealistically high population estimates for some grid cells relative to the number of building counts. These grid cells are mostly in rural and remote areas and include those with the following coordinates: (28°47'22.995"E 2°23'7.892"S), (28°34'20.137"E 2°41'44.222"S), (28°34'20.49"E 2°41'44.223"S), (27°41'47.035"E 3°32'19.882"S), and (28°26'43.074"E 2°16'7.684"S). These seemingly overinflated population numbers might be a function of residual errors not accounted for by the model, and it may also be that the PNLP data used as input in these locations could be highly inflated. One specific case of concern is the pixel located at (0°33'01.5"N, 25°09'25.5"E), which shows an estimated population of 755.11, despite the settlement layer indicating only one building. Upon further inspection of OpenStreetMap data, this location appears to house a large complex, possibly a factory with residential

accommodations for employees. This suggests that a substantial number of individuals may occupy this structure, which could explain the seemingly inflated population estimate. The model does not account for external factors such as migration, displacement, or sudden demographic changes, which could significantly influence population dynamics. Consequently, the estimates may not fully reflect dynamic population shifts occurring beyond the scope of the input data.

Grid cell alignment is based on a mastergrid. Note that this version's (v4.2) mastergrid aligns with version 4.1 but does not align with previous DRC gridded population layers, namely versions v1.0, v2.0, v3.0. We updated the mastergrid in 2024 to ensure grid cell alignment across all new WorldPop data products.

SOURCE DATA

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

PDRS Data

The input population dataset used for the population modelling for Tshopo Province was the PDRS malaria bednet campaign data. The PDRS dataset, which was collected in 2023, provided detailed information on a given household for which a bednet was issued, such as the household size, the number of bednets issued, the number of children in the household, the number of males, and the number of females, among others.

Although the malaria bednet campaign was designed to distribute bednet to every household within the province, a preliminary exploratory data analysis carried out on the PDRS data indicated that some households were not visited during the campaign, while others were not completely covered.

The GPS points of all households within the province were provided in the PDRS data. We implemented population modelling for small spatial units, utilising unofficial boundaries similar to census Enumeration Areas ("pre-EAs"; Qader et al., 2024). The household-level data on population counts was spatially aggregated to these spatial units, by summing the household size data for all GPS points within each pre-EA boundary.

Settlement Data

Settlement data was provided by CIESIN in the form of raster files (CIESIN, 2024). We obtained two different settlement products, namely (i) settlement area, which indicates the area of all buildings whose centroid falls within a given cell, and (ii) building count, which is the number of building centroids within a given cell. Each of these settlement layers was used in separate analyses together with the observed population count and ancillary geospatial data in robust statistical modeling. After using each settlement layer in the analysis, we compared model metrics and the gridded population layer from both

layers. Settlement building count provided more realistic population numbers at the gridcell level and hence was retained for the final population predictions.

Geospatial Covariates

A wide variety of geospatial covariates, which are related to population distribution, were considered in the modelling. These geospatial covariates include land uses and land cover data, climate variables such as temperature and rainfall, physical features and infrastructure such as roads and schools, and conflict data. Population model covariates were selected using a generalized linear model (GLM) based stepwise selection method. The selected covariates were further assessed for multi-collinearity and statistical significance. Eventually, of the 85 geospatial covariates initially tested, 14 were retained as the best fit covariates with variance inflation factor (VIF) of less than 5. The descriptions of these final geospatial covariates are presented in Table 1 below.

Table 1. Selected geospatial covariates for the modelling.

Description	**Source**	**Link/Reference**
Coefficient of variation – building area	WorldPop	Woods et al (2024)
Coefficient of variation – building count	WorldPop	Woods et al (2024)
Euclidean distance to GRID3 roads	GRID3	https://data.grid3.org/datasets/8a8d510bd9404212864348010112212b_0/explore
Walking – travel time to the nearest health facility	GRID3	https://data.grid3.org/datasets/8a8d510bd9404212864348010112212b_0/explore
Euclidean distance to tree covers 2020	WorldPop	Woods et al (2024)
Mean – Normalized vegetation Index 2021	Copernicus	https://land.copernicus.eu/global/products/ba
Coefficient of variation – Microsoft building length	Microsoft	https://github.com/microsoft/RoadDetections
Mean – Microsoft building length	Microsoft	https://github.com/microsoft/GlobalMLBuildingFootprints
Road density	Microsoft	https://github.com/microsoft/RoadDetections
Euclidean distance to cities 2020	Malaria Atlas Project	https://malariaatlas.org/explorer/#/

Euclidean distance to ACLED conflict locations 2022	ACLED	https://acleddata.com/
Euclidean distance to OSM places of worship	OSM	https://www.openstreetmap.org
Euclidean distance to OSM local Roads 2023	OSM	https://www.openstreetmap.org
Euclidean distance to OSM main roads 2023	OSM	https://www.openstreetmap.org

Age-Sex Proportions

We used the 2024 WorldPop Global subnational population pyramids (Bondarenko et al 2025) to calculate the age-sex proportions for Sud_Kivu. We multiplied our gridded population estimates (COD_ Tshopo_province_population_v4_2_gridded.tif) by the age-sex proportions(grouping) to produce COD_ Tshopo _province_population_v4.2_agesex.zip.

METHODS OVERVIEW

The key steps of our approach were as follows:

- Cleaning and summarizing the household sizes from the PDRS dataset to get the total population at the pre- Enumeration Area (pre-EA) level (Qader at al. 2024). PDRS data points with household sizes above 500 people per household were signalled as clear outliers and as such we imputed these household sizes with the median household size. Similarly, PDRS data point with household sizes of 0 were also imputed using the median household size.
- Geospatial covariates were subjected to robust covariate selection for model training and parameter estimation.
- We developed a hierarchical Bayesian statistical model using the INLA-SPDE approach (Lindgren et al. 2011) to fit and predict the population count.
- Population estimates were predicted at grid cell level using the grid cell values of the covariates selected at the model training level.

Statistical Modelling

We approached the population modelling using two complementary methods. The first method assumes that the PDRS data provides unbiased estimates of the population counts and does not require any form of bias correction (here after known as *unscaled model approach).* Whereas the second method acknowledges the fact that the PDRS data, which was our population model's primary input, could be systematically biased and requires bias adjustment to avoid under- or over-estimation of population counts. Specifically, for the second method, we took advantage of an existing recent Microcensus data (Flowminder Foundation et. al., 2021; UCLA and Kinshasa School of Public Health, 2018) that overlapped with the PDRS data at cluster unit levels across six provinces in the DRC (Kongo Central, Kinshasa, Kwilu, Haut-Lomami, Sud-Kivu, and Ituri). Thus, the second approach which is also known as the *scaled model approach* uses a scaling factor to adjust for the potential bias in the population count. The model result presented here are based on the *scaled model approach*. Further details of the models' specifications are provided below:

Unscaled Model Approach

In general, within the context of bottom-up population modelling (Leasure et al. 2022, Boo et al., 2022; Darin et al., 2022, Nnanatu et al. 2022), the observed population count at area unit k, y_k, is a Poisson distributed random variable with mean parameter $\lambda_k = \bar{d}_k B_k$ where k is the estimation unit (e.g., enumeration area), while $\bar{d}_k$ and B_k are the mean parameter of the corresponding population density and the number of buildings/settled area, respectively. That is,

$$y_k \sim \text{Poisson}(\bar{d}_k B_k) \qquad (1)$$

Then, the transformed mean population density $\bar{d}_k$ is assumed to be linked to a set of geospatial covariates with log-link function:

$$log(\bar{d}_k) = \mu + \sum_{j=1}^{J} \beta_j x_{kj} + \sum_{l=1}^{L} f_l(z_{kl}) \qquad (2)$$

where μ is the intercept parameter, $\boldsymbol{\beta} = (\beta_1, \ldots \beta_J)$ is a vector of fixed effects coefficients of the $(x_1, \ldots, x_J)$ geospatial covariates; $f_l(.)$ is a function of L random effects covariates including those that capture variability in the population estimates due to settlement type, cluster location and spatial autocorrelations. The population density (defined as people per building or people per settled area) is assumed to be a Gamma distributed random variable with parameters α and γ with mean and variance given by $\bar{d}_k = \alpha/\gamma$ and $\sigma_d^2 = \alpha/\gamma^2$, respectively.

The inclusion of spatial autocorrelation requires the use of computationally efficient statistical modelling software. Thus, the integrated nested Laplace approximation (INLA; Rue et al 2009; Lindgren et al., 2011) is used via the R-INLA statistical package. Note

that the method described above predicts population count at regular grid cells using the parameter values trained at the cluster/pre-EA level by calculating the predicted grid-cell level population density as

$$\hat{d}_g = \exp\left(\hat{\mu} + \sum_{j=1}^{J} \hat{\beta}_j x_{gj} + \sum_{l=1}^{L} \hat{f}_l\left(z_{g_l}\right)\right) \qquad (3)$$

where $\{x_g\}_{g=1}^{G}$ are the corresponding grid cell level values of the geospatial covariates used in training the model at the cluster level, so that the overall predicted population count across the G 100m by 100m grid cells is given by

$$\widehat{pop} = \sum_{g=1}^{G} B_g \hat{d}_g \qquad (4)$$

where B_g is the corresponding building count or the size of settled area in grid g. We assumed default INLA priors for each of the parameter estimates which have been found to be robust.

Scaled Model Approach

For the second method, a scale factor is calculated for each of the overlapping clusters across the 6 provinces according to equation (5):

$$c_k = \frac{m_k}{p_k} \qquad (5)$$

where c_k, m_k and p_k are the scale/adjustment factor, Microcensus population count, and PDRS population count for cluster k, respectively. That is, when $c_k < 1$, the PDRS overestimates the population count by $100(1 - c_k)\%$, and underestimates by $100(c_k - 1)\%$ when $c_k > 1$. Thus, the first approach described above assumes that the PDRS data is as good as the Microcensus data, that is, $c_k = 1$ (or $m_k = p_k$).

We predicted estimates of the scale factor c_k at locations where Microcensus and PDRS datasets did not overlap using model parameters trained with the scale factor data $\mathbf{c} = (c_1, \ldots, c_n)^T$ obtained at the n Microcensus-PDRS overlapping clusters. A similar approach was also adopted to predict the scale factors at grid cell levels to better capture local variations. To do this, first we assume that the scale factors are realisations from Gamma distributed random variables, that is,

$$C_k \sim \text{Gamma}(\alpha_c, \gamma_c) \qquad (6)$$

with mean and variance given by $\bar{c}_k = \alpha_c/\gamma_c$ and $\sigma_c^2 = \alpha_c/\gamma_c^2$, respectively. So that the mean parameter is linked to a set of geospatial covariates as well as spatial autocorrelations and other random effects through a log-link function defined in equation (7):

$$log(\bar{c}_k) = \mu + \sum_{j=1}^{J} \beta_j x_{kj} + \sum_{l=1}^{L} f_l(z_{kl}) \qquad (7)$$

where μ is the intercept parameter, $\boldsymbol{\beta} = (\beta_1, \ldots \beta_J)$ is a vector of fixed effects coefficients of the $(x_1, \ldots, x_J)$ geospatial covariates (See Table 1 for the description of the final covariates selected through stepwise regression); $f_l(.)$ is a function of L random effects covariates including those that capture variability in the population estimates due to settlement type, cluster location and spatial autocorrelations. Then, grid cell predictions of the scale factors $\hat{c}_g$ are obtained using the grid cell values of the geospatial covariates and the estimates of the model parameters $\hat{\mu}, \hat{\beta}$ and $\hat{f}$ trained at the cluster level:

$$\hat{c}_g = \exp\left(\hat{\mu} + \sum_{j=1}^{J} \hat{\beta}_j x_{gj} + \sum_{l=1}^{L} \hat{f}_l\left(z_{g_l}\right)\right) \qquad (8)$$

Finally, the grid cell level scaled population estimates are obtained as

$$\widehat{pop}_g^{(scaled)} = \hat{c}_g \times \widehat{pop}_g \qquad (9)$$

where $\widehat{pop}_g = B_g \hat{d}_g$ is the corresponding unscaled grid cell level population estimate.

In this study, we utilized a scaled approach for population modelling, using building count to define population density. Within this framework, we tested three different model re-parameterizations. The first model, Model 1, included fixed effects for the geospatial covariates and a random effect for the Global Human Settlement Layer Degree of Urbanization classes (GHSL-SMOD). Model 2 extended this by incorporating an additional random effect at the cluster level. The final model, Model 3, further included a spatial random effect component in addition to the specifications in Model 2. We compared these three models with each other. Model 2 and model 3 provided similar metrics; therefore, we decided to use model 3 in the final gridded population because the predicted population from model 2 was a bit inflated at the pixel level compared to model 3, which gave optimum predictions.

Model fit checks.

Model fit checks and model selection of the three models described above relied primarily on a constellation of model fit metrics, including the absolute bias (BIAS), the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE) and the Pearson correlation coefficient (CORR). A lower value for the absolute bias, MAE and the RMSE indicates a better-fit model. A higher value for the Pearson correlation coefficient indicates a better-fit model. Table 2 below provides the model-fit metrics across the three models. Based on the model fit checks, model 3 has a lower MAE, which might make it better in terms of average absolute error and predicted values. Additionally, it provides the best fit, and the final population predictions at the grid cell level are based on this model.

Table 2. Model fit metrics.

Models	BIAS	RMSE	MAE	Corr
Model1	-6487.59	152936.21	6663.63	0.006
Model 2	23.29	414.30	159.45	0.93
Model 3	37.77	441.32	155.78	0.93

The novelty of the modelling approach utilised here is that it allows for the adjustment of potential systematic bias in the two settlement layers used as input in defining population density within a coherent Bayesian hierarchical population modelling framework while at the same time adjusting for spatial autocorrelation within the observed data.

All data processing and analysis was carried out using R (v.4.2.2) (R Core Team, 2023)

and INLA (v 22.05.07) (Rue et al. 2009). The concept of bottom-up population modelling for estimating population in the absence of recent census data was described by Leasure et al. (2020). Approaches similar to the one used here for Haut-Katanga have been carried out for Papua New Guinea (WorldPop and NSO PNG, 2022) and Cameroun (Nnanatu et al. 2022)

ACKNOWLEDGEMENTS

We thank the DRC PNLP and its implementing partners for providing access to the anonymised 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 Attila Lazar for reviewing the data and providing thoughtful suggestions prior to this release.

WORKS CITED

Bondarenko M., Priyatikanto R., Tejedor-Garavito N., Zhang W., McKeen T., Cunningham A., Woods T., Hilton J., Cihan D., Nosatiuk B., Brinkhoff T., Tatem A., Sorichetta A. (2025) Constrained estimates of 2015-2030 total number of people per grid square broken down by gender and age groupings at a resolution of 3 arc (approximately 100m at the equator) R2024B version v1. Global Demographic Data Project - Funded by The Bill and Melinda Gates Foundation (INV-045237). WorldPop - School of Geography and Environmental Science, University of Southampton. DOI:10.5258/SOTON/WP00805

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), Columbia University, Ministère de la Santé Publique, Hygiène et Prévention, Democratic Republic of the Congo, and GRID3. 2025. GRID3 COD - Health Areas v4.0. New York: Columbia University. https://doi.org/10.7916/nnew-da26. Accessed 16 March, 2025.

Center for International 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.

Flowminder Foundation, École de Santé Publique de Kinshasa (ESPK), WorldPop (University of Southampton), Bureau Central du Recensement (BCR). 2021. Microcensus survey in the provinces of Haut-Katanga, Haut-Katanga, Ituri, Kasaï, Kasaï-Oriental, Lomami, and Sud-Kivu (Democratic Republic of the Congo). Version 1.5. [Dataset].

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

Nnanatu C.C., Yankey O., Abbott T. J., 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

Qader S H, Batana Y. M., Kosmidou-Bradley W., Skoufias E., Tatem A. J. 2024. Automatic pre-Enumeration Areas (pre-EAs) delineation and national sampling frame for the Democratic Republic of Congo. Policy Research Working Paper; No. (under review). DRC - Automatic Pre-Enumeration Area Delineation for National Sample Frame Data Report | Data Catalog (worldbank.org)

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

UCLA-DRC Health Research and Training Program (University of California, Los Angeles) and Kinshasa School of Public Health. 2017 and 2018. Kinshasa, Kongo Central and former Bandundu microcensus survey data.

D. Woods, T. McKeen, A. Cunningham, R. Priyakanto, A. Soricheta , A.J. Tatem and M. Bondarenko. 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