Release Statement

Modelled gridded population estimates for Kwilu 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 Kwilu 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 2022 as well as settlement extents and geospatial covariates, to model and estimate population numbers at grid cell level using a Bayesian hierarchical statistical modelling framework. The approach facilitated accounting for the multiple levels of variability within the data while simultaneously quantifying for uncertainties in parameter estimates. These model-based population estimates can be considered as most accurately representing the year 2022, which is the period following the PDRS survey data collection for Kwilu. Although the methods were robust enough to explicitly account for key random biases and adjust for potential systematic biases within the observed datasets, it is important to note that some systematic biases arising from other sources may remain.

These data were produced by the WorldPop Research Group at the University of Southampton. The 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. A robust Bayesian joint (hurdle) population modelling approach was developed to estimate population density whilst at same time accounting for probability of settlement detection. The final statistical modelling was conceived, designed, and implemented by Chris Nnanatu. Data processing was done by Ortis Yankey, while project oversight was provided by Attila Lazar, 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 the 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_Kwilu_province_population_v4.2_gridded.zip
2. COD_Kwilu_province_population_v4.2_agesex.zip

LICENSE

SUGGESTED CITATION

Nnanatu C., Yankey O., Chaudhuri S., Chamberlain H., Lazar A. N., Tatem A. J. 2025. Bottom-up gridded population estimates for Kwilu Province in the Democratic Republic of Congo (2022), version 4.2. WorldPop, University of Southampton. doi: https://dx.doi.org/10.5258/SOTON/WP00779

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_Kwilu_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 Kwilu 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_Kwilu_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 Kwilu. The values are the 2.5% posterior probability distribution for the predicted population size in each grid cell obtained as part of the 95% credible interval of the posterior probability distribution. 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_Kwilu_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 Kwilu. The values are the 97.5% posterior probability distribution for the predicted population size in each grid cell obtained as part of the 95% credible interval of the posterior probability distribution. 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_Kwilu_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 Kwilu. 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 Kwilu to the gridded population estimates (COD_Kwilu_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 Kwilu Province [doi: 10.5258/SOTON/WP00779] (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

The key **assumptions** upon which the statistical modelling techniques developed and implemented here are based upon are:

1) The PDRS data provide complete counts of people within the Kwilu Province.
2) Missing households within the PDRS data were either missed completely at random (MCAR) or simply missing at random (MAR), thus median imputation was used to impute missing household sizes at household levels before aggregation to the modelling unit.
3) Sources of biases within the PDRS data are mostly random which could be accounted for by using appropriate random effects within the statistical models.
4) The presence of zero building counts within the settlement data could be because of imperfect detection within the settlement data which could be accounted for by jointly modelling the detection probability.
5) All geospatial covariates are accurately detected and processed.

Despite the robust statistical technique utilised here, below are some **limitations** that's worth highlighting:

1) These population estimates most likely represent the 2022 time, but because of the different ages of the input data used to build the model, a precise time point could not be allocated. The PDRS data that was used as the response variable was collected in 2022, 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 additional uncertainties in the parameter estimates.
2) Data on the number of people 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 per household is proportional to the household size, there is an incentive for the respondents to potentially provide incorrect counts of population per household in an attempt to either get more or less number of ITNs. The presence of incorrect household

sizes in the input population data may introduce systematic biases in the modelled estimates.

3) Where there are unrealistically high values of input population data and/or very low number of settlements (especially in rural areas), the tendency of unrealistically high predictions of population estimates across the affected grid cells is usually high. High input population number and/or low building count from the CIESIN settlement layer could be because of imperfect detection due to tree canopy or cloud cover, for example.
4) Although the model draws upon recent datasets and geospatial covariates which could reflect current population density and distribution, the model does not explicitly 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 and 4.2 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 Kwilu Province was the PDRS malaria bednet campaign data. The PDRS dataset, which was collected in 2022, 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 (Figure 1B), 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 density and distribution, were considered in the 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. Final 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 80 geospatial covariates initially tested, 3 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**
Euclidean distance to water bodies 2021(x1)	WorldPop	Woods et al (2024)
Euclidean distance to OSM educational facilities 2023 (x2)	OSM	https://www.openstreetmap.org/#map=3/68.59/70.05
Euclidean distance to OSM places of worship 2023(x3)	OSM	https://www.openstreetmap.org/#map=3/68.59/70.05

Age-Sex Proportions

We used the 2024 WorldPop Global subnational population pyramids (Bondarenko et al 2025) to calculate the age-sex proportions for Kwilu. We multiplied our gridded population estimates (COD_Kwilu_province_population_v4_2_gridded.tif) by the age-sex proportions(grouping) to produce COD_Kwilu_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).
- Household sizes from the PDRS data point ranged between 0 and 20. Out of 1196306 PDRS data points, 6714 data points had a household size of 0. Points with a household size of 0 were imputed using the median household size for the province.
- Geospatial covariates were subjected to robust covariate selection for model training and parameter estimation.
- An advanced Bayesian hierarchical joint (hurdle) modelling framework was developed and implemented using the INLA-SPDE approach (Lindgren et al. 2011) to easily address issues of settlement detection biases and readily quantify uncertainties.
- Datasets were used to train population density and detection probability models at EA levels which were eventually used to predict populations at grid cell level using the grid cell values of the covariates selected at the model training level as well as the corresponding building counts.
- Prior to model training, the input population data was aggregated to 14,283 EAs with values ranging from 1 to 29,435 per grid cell with a median value of 398 people per EA. However, there were 2,490 (~17%) EAs without population counts and these were set as NAs.
- The total building counts for the EAs ranged between 0 to 4687 with a median value of ~58 buildings per EA. There were no NAs, but 916 EAs had no buildings observed.
- The density variable calculated as people divided by the number of buildings per EA, returned infinite values wherever the denominator (building counts) has a value of zero. The infinite values were set to NAs so that their values could be estimated.
- After model training, covariate values found to contain high number of missing values or covariates with extreme values of z-scores (exceeding 4.5) were dropped. The models were then retrained using the final covariates retained before proceeding to the final predictions at the grid cells.
- Best fit covariates were selected for the density model, and these were also assumed to equally predict settlement detection probability. Thus, the same set of covariates were used for both models.

Statistical Modelling

Model set up

Usually, in bottom-up population modelling (Leasure et al. 2022, Boo et al., 2022; Darin et al., 2022, Nnanatu et al. 2022), a Poisson probability distribution is used to describe the distribution of the population count C_i with the mean parameters decomposed into building count B_i and average population density μ_i, that is,

$$C_i \sim Poisson(\mu_i B_i) \qquad (1)$$

instead of a single mean parameter $\lambda_i > 0$. The reason for this is because a Poisson distribution requires that the mean and the variance of the random variable be equal. However, this is rarely the case in most practical situations especially within the context of population modelling where issues of overdispersion are a common place. Besides, the use of multiple component parameters ensures that we could account for spatial aggregation error.

Typically, robust estimates of population are obtained by first modelling population density D_i and then estimate the corresponding population counts thereafter using the relationship $D_i = C_i/B_i$ or $C_i = D_i \times B_i$. Here, the population density is assigned a Gamma distribution with mean and variance of μ_i and ϕ, respectively. That is,

$$D_i \sim Gamma\left(\frac{\mu_i^2}{\phi}, \frac{\mu_i}{\phi}\right) \qquad (2)$$

However, when the building counts are susceptible to imperfect detection due to factors such as tree canopy and cloud covers, it makes sense to take this into account while developing the population modelling. Nnanatu et al. (2024) developed an approach for addressing such systematic bias in settlement data in the context of Papua New Guinea by explicitly integrating building count model within the bottom-up population modelling framework (Nnanatu et al., 2024). However, in the present context, other factors other than obscured satellite observations, e.g., human error, could lead to settlement data imperfect detection, thus, the integration of settlement data detection probability would be appropriate.

To do this, we assume that the value of the population density is governed by settlement detection binary variable z_i defined as

$$z_i = \begin{cases} 1, & if\ B_i > 0 \\ 0, & if\ B_i = 0 \end{cases} \qquad (3)$$

so that adjusted population density $\widetilde{D}_i = z_i D_i$ which is defined as

$$\widetilde{D}_i = \begin{cases} \widetilde{D}_i, & if\ B_i > 0 \\ NA, & if\ B_i = 0 \end{cases} \qquad (4)$$

follows a Gamma distribution

$$\widetilde{D}_i \sim Gamma\left(\frac{p_i^2\mu_i^2}{\phi}, \frac{p_i\mu_i}{\phi}\right) \quad (4)$$

with mean and variance of $p_i\mu_i$ and ϕ, respectively; where p_i is the probability that a given location or spatial unit i contains settlement, that is, the mean of the settlement detection variable z_i with a Bernoulli (or Binomial with number of trials of 1) probability distribution given by

$$z_i \sim Bernoulli(p_i) \quad (5)$$

So that,

$$p_i = \frac{\mathrm{e}^{\mathrm{\alpha+\beta X+fZ+\zeta_i}}}{1 + \mathrm{e}^{\mathrm{\alpha+\beta X+fZ+\zeta_i}}} \quad (6)$$

where $\alpha, \beta, X, f,$ and Z are the intercept parameter, a vector of fixed effects coefficients of the geospatial covariates contained in the design matrix X, the functions for modelling random effects (e.g., settlement class), and design matrix for the random effect variables, respectively. Also, ζ_i is spatial random effect which can be further decomposed into a spatially varying (ξ_i) and spatially independent (ϑ_i) random effects.

Similarly, the log of the mean density μ_i is modelled as a linear combination of the geospatial covariates and random effects as given below:

$$\mu_i = \mathrm{e}^{\mathrm{\alpha+\beta X+fZ+\xi_i}+\vartheta_i} \quad (7)$$

Then the predicted population count $\hat{C}_i$ is given by

$$\hat{C}_i = \hat{p}_i\widehat{D}_iB_i \quad (8)$$

where $\hat{p}_i$ is the predicted detection probability; $\widehat{D}_i$ is the predicted density; and the building count. Thus, equation (1) can be rewritten as

$$C_i \sim Poisson(p_i\mu_iB_i) \quad (9)$$

that is, $\lambda_i = p_i\mu_iB_i$.

Model fitting

In this study, we used building count to define population density by dividing the observed population count with the corresponding number of buildings. Then robust Bayesian hierarchical models were trained and tested separately for each of the population density and settlement detection probability. The three top competing models are shown below:

Model 1: *y ~ -1+ Intercept + x1 + x2 + x3 + f(eps, model="iid") + f(s, model=spde) + f(set_typ, model="iid")*

Model 2: y ~ -1 + Intercept + x1 + x2 + x3 + f(eps, model="iid") + f(set_typ, model="iid")

Model 3: y ~ -1 + Intercept + x1 + x2 + x3 + f(eps, model="iid") + f(s, model=spde)

Where *y is population density (or the settlement detection variable), x1, x2, and x3 are the fixed effect geospatial covariates defined in table 1, f(eps, model="iid") represent random effect term for the EAs, f(set_typ, model="iid") is a random effect component for settlement type classification using the Global Human Settlement Layer Degree of Urbanization classes (GHSL-SMOD) (Schiavina et al. 2023), and f(s, model=spde) represent a spatial random effect.*

Thus, the three models are nested with Model 1 serving as the full model.

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 Mean Absolute Error (MAE), the Root Mean Square Error (RMSE), the Deviance Information Criterion (DIC) and the Pearson correlation coefficient (CORR). A lower value for the absolute bias, MAE and the RMSE and the DIC 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 1 provided the best fit for both detection and density variables, and the final population predictions at the grid cell level were based on this model. Model selections were based on Deviance Information Criterion with models with lowest values being selected as the best fit.

Table 2. DIC values of the models fitted to each dataset

Data	**DIC**		
	Model 1	*Model 2*	*Model 3*
Detection	4304.48	5012.67	4601.98
Density	55685.92	57941.14	56206.84

The novelty of the modelling approach utilized here is that it allows for the adjustment of potential systematic bias due to imperfect settlement data observations.

All data processing and analysis was carried out using R (v.4.3.2) (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 (Nnanatu et al. 2024) and Cameroun (Nnanatu et al. 2022).

Model Fit Checks and Model Cross-validation (Adjusted)

We compare the model performance of the adjusted and unadjusted models based on model 1 above which was the best model. The adjusted population utilises the settlement detection probabilities to scale the predicted population, whereas the unadjusted population is the predicted population without any adjustment. That is, the adjusted population estimates are obtained from $\hat{C}_i = \hat{p}_i \widehat{D}_i B_i$, while the unadjusted estimates are from $\hat{C}_i = \widehat{D}_i B_i$ only. The detection probability-adjusted population provides better model fits than the unadjusted one (Table 3).

Table 3. Model fit metrics for the adjusted and the adjusted model

Data	MAE	RMSE	CORR
Unadjusted (Base)	3092.09	4947.38	0.59
Adjusted	2719.72	4543.27	0.65

Cross-validation allows us to test the predictive ability of our model by predicting values of response variables that were either part of the training samples and withheld for prediction (in-sample), or dividing the data into two of 20% test and 80% train sets. We used k-fold cross-validation with 9 folds to test the predictive ability of our methodology. Each of the 9 folds (subsamples) represents a test set. Thus, the test set was never part of the training sample for out-of-sample cross-validation. Results on Table 4, show similar values with high Pearson correlation coefficient of at least ~0.8 (see also Figure 1).

Table 4. Model cross-validation metrics

Dataset	MAE	RMSE	CC
In-Sample	269.64	519.92	0.80
Out-of-Sample	263.33	520.28	0.80

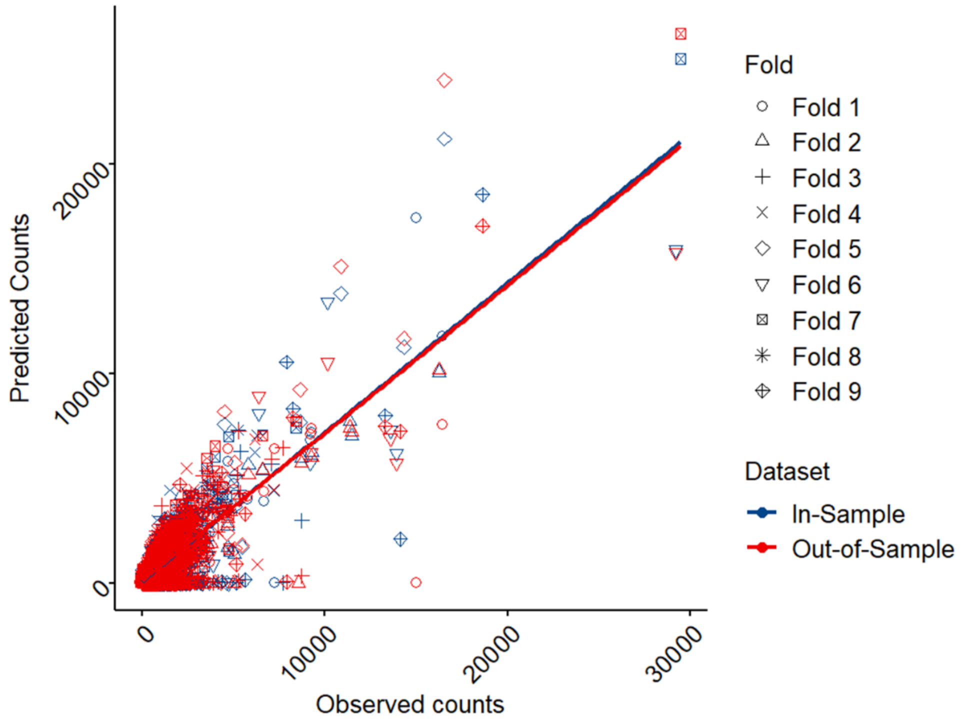

Figure 2. Scatter plots of model cross validations for in-sample and out-of-sample test datasets

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 the modelling team for the internal reviews of the earlier versions and for providing very useful 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 and Ministère de la Santé Publique, Hygiène et Prévention, Democratic Republic of the Congo. 2023. GRID3 COD - Health Areas v2.0. Unpublished.

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 Lomami(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

Nnanatu, C., Bonnie, A., Joseph, J., Yankey, O., Cihan, D., Gadiaga, A., ... & Tatem, A. (2024). Small area population estimation from health intervention campaign surveys and partially observed settlement data.

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

Schiavina M., Melchiorri M., & Pesaresi M. (2023). GHS-SMOD R2023A - GHS settlement layers, application of the Degree of Urbanisation methodology (stage I) to GHS-POP R2023A and GHS-BUILT-S R2023A, multitemporal (1975-2030). *European Commission, Joint Research Centre (JRC)*. doi:10.2905/A0DF7A6F-49DE-46EA-9BDE-563437A6E2BA.

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