WorldPop | INSTITUT NATIONAL DE LA STATISTIQUE / NATIONAL INSTITUTE OF STATISTICS | GRID3 | University of Southampton

Efficient Bayesian Hierarchical Population Estimation Modelling using INLA-SPDE: Integrating Multiple Data Sources and Spatial Autocorrelation

Chibuzor Christopher Nnanatu[1*], Ortis Yankey[1], Anaclet Désiré Dzossa[2], Thomas Abbott[1], Assane Gadiaga[1], Attila Lazar[1], Andrew J Tatem[1]

1.WorldPop, School of Geography and Environmental Science, University of Southampton, SO17 1BJ, UK; 2. National Institute of Statistics (NIS)- Cameroon

BACKGROUND

Modelled population estimates provide small area population data that can support governance, equitable resource allocation, and delivery of aid, especially in countries where census data are outdated or incomplete. Bottom-up population modelling approaches [e.g. 1-2] use advanced statistical techniques to combine sample demographic data (counts of people), with geospatial covariates and satellite-derived settlement data, to produce high-resolution population estimates (usually at ~100m regular grid cells). However, while the existing methods have mostly implied spatial dependence of the observations through the geospatial covariates, the latent effects of the potential spatial autocorrelation on the observed data is still not very clear. Here, we produced small area estimates of population in Cameroon (along with uncertainty metrics) using the integrated nested Laplace approximation/stochastic partial differential equation (INLA/SPDE) framework [3], which allowed us to explicitly integrate spatial autocorrelation with five nationally representative household listing datasets within a bottom-up population modelling framework. Our methodology represents an important development within population modelling contexts and has facilitated the development of several other INLA/SPDE-based population modelling techniques developed to address different data challenges across the World.

DATA:

- ***Demographic data***: Five geolocated cluster household listing datasets from nationally-representative household surveys
- ***Geospatial covariates***: E.g., Nighttime lights, distance to local roads, distance to market places
- ***Settlement data***: Building footprints, settlement type classifications

METHOD:

- The datasets were collated, processed and cleaned.
- Geospatial covariates were stacked and tested for model fit.
- Spatial autocorrelation was implied through the triangulation of the entire country using non-convex hull mesh.
- Trained model parameters were used to make predictions at 100x100m grid cells.

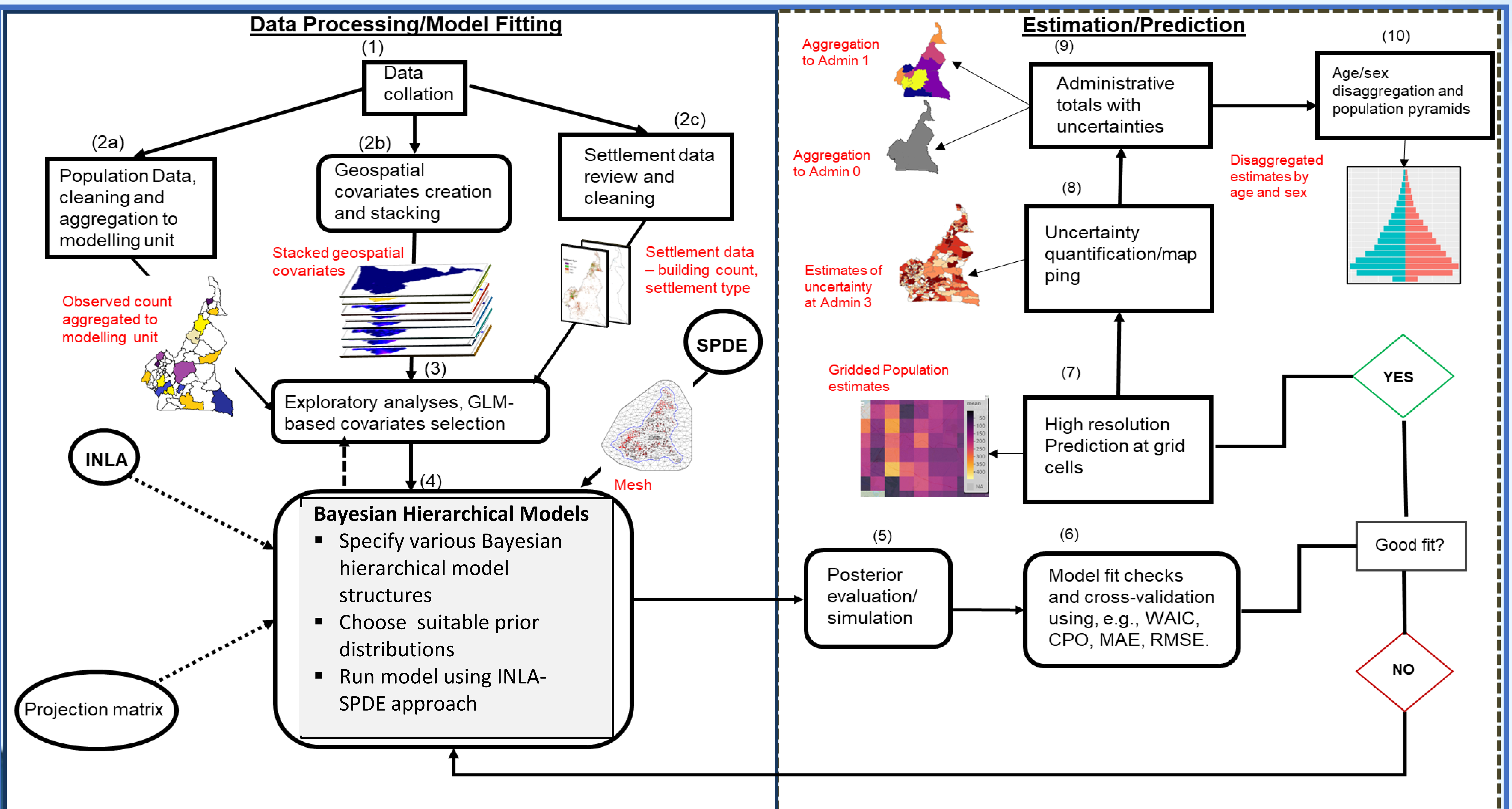

Figure 1. Modelling workflow. WAIC – Widely Acceptable Information Criterion; CPO – Conditional Probability Ordinate; MAE – Mean Absolute Error; RMSE – Root Mean Square Error

Table 1. Model Cross-Validation

Sample	MAE	RMSE	Absolute BIAS	CORR
In-Sample	58.745	82.527	3.997	0.997
Out-of-sample	63.327	93.278	3.212	0.996

Table 2. Estimated total national population

Total	Lower bound	Upper bound
28,663,487	27,147,814	30,612,947

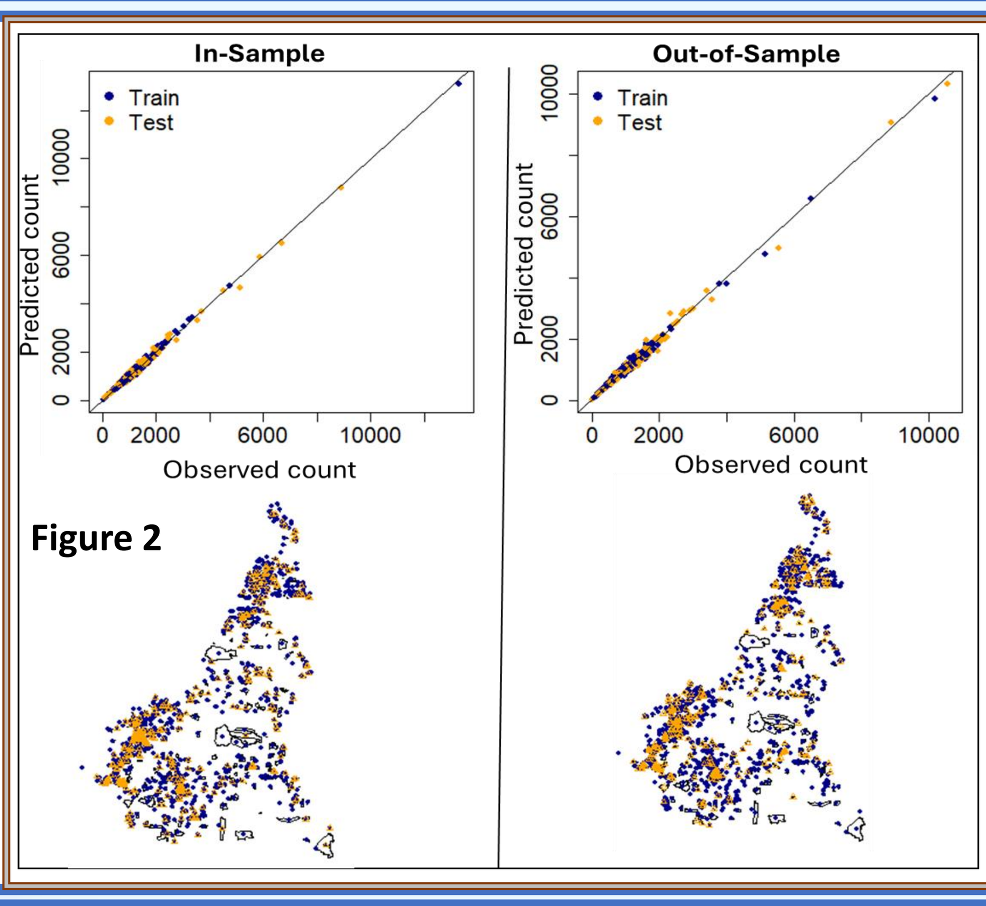

Figure 2

RESULTS:

- Model cross-validation using both in-sample and out-of-sample cross-validation techniques indicated good predictive ability for the best fit model (Table 1 & Figure 2).
- There is a probability of 95% that the 'true' total national population lies between 27.15M and 30.61M (Table 2)
- The datasets are now published and freely available in the WorldPop data repository

SCAN THE QR CODE TO ACCESS THE POPULATION DATA

REFERENCES

[1] Leasure, D. R., W. C. Jochem, E. M. Weber, V. Seaman and A. J. Tatem (2020). "National population mapping from sparse survey data: A hierarchical Bayesian modeling framework to account for uncertainty." *PNAS*: 201913050. DOI: 10.1073/pnas.1913050117. https://www.pnas.org/doi/pdf/10.1073/pnas.1913050117

[2] Boo, G., Darin, E., Leasure, D. R., …, Tatem, A. J. (2022). "High-resolution population estimation using household survey data and building footprints." *Nature Communications*, 13, 1330. https://doi.org/10.1038/s41467-022-29094-x

[3] Rue, Havard, Sara Martino, and Nicolas Chopin. 2009. "Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximations. Journal of the Royal Statistical Society, Series B 71 (2): 319–92. https://www.unfpa.org/resources/value-modelled-population-estimates-census-planning-and-preparation

ACKNOWLEDGEMENTS

This work is part of the GRID3 project funded by the Bill & Melinda Gates Foundation (BMGF) and the United Kingdom Foreign, Commonwealth & Development Office (OPP1182425). Project partners include WorldPop at the University of Southampton, the United Nations Population Fund (UNFPA), Center for International Earth Science Information Network in the Columbia Climate School at Columbia University, and the Flowminder Foundation. We acknowledge the Cameroon National Institute of Statistics (NIS) for providing the household listing datasets, and the projected population estimates used in this research.