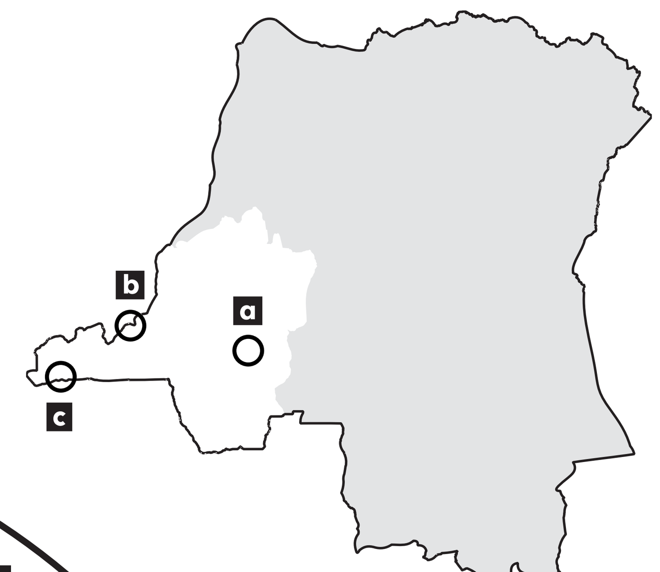
HIGH-RESOLUTION POPULATION MAPPING AND ESTIMATION

IN THE WESTERN PART OF THE DEMOCRATIC REPUBLIC OF CONGO

Gianluca Boo [1], **Édith Darin** [1], **Douglas R Leasure** [1], **Andrew J Tatem** [1,2]

[1] WorldPop Research Group, Department of Geography and Environmental Science, University of Southampton, UK

[2] Flowminder Foundation, Stockholm, Sweden



INTRODUCTION

According to the United Nations, the Democratic Republic of Congo (DRC) is the fourth most populous country in Africa [1]. Still, accurate demographic figures are currently unavailable because **the last national census was conducted over three decades ago**. This situation has shown to be critical to policymaking and intervention planning in the domain of public health. In the western part of the country, for instance, African trypanosomiasis (i.e., sleeping sickness) is highly endemic but **accurate sub-national population figures are required for effective disease mapping** [2].

This work aims at producing gridded population estimates with a 100 meters spatial resolution for the five most western provinces of DRC. In doing so, we adopt a **bottom-up modeling approach incorporating demographic data** from recent micro-census surveys and various geospatial datasets in a Bayesian hierarchical framework [3]. The resulting **high-resolution population estimates and associated uncertainty measures** are meant to inform spatial epidemiological applications in the region and foster the use of similar modeling approaches in other parts of the country.

DATA AND METHODS

Demographic data have been collected through **926 micro-census surveys conducted across the Kongo Central, Kinshasa, Mai-Ndombe, Kwilu and Kwango provinces** in 2017 and 2018, in an effort lead by the UCLA-DRC Health Research and Training Program. The surveys enable deriving complete enumerations (N) within well-defined areas (i) of approximately three settled hectares (A). Additional key datasets are the **settlement map** produced by the Oak Ridge National Laboratory in 2016 that identifies and classifies settled land, and the **administrative boundaries** provided by the Congolese Bureau Central du Recensement.

The proposed bottom-up model consists of a **hierarchical Bayesian regression** within the family of Poisson generalized linear mixed models **[eq. 1]**, where a log-normal distribution provides overdispersion **[eq. 2]** [4]. The hierarchical form allows for submodels to estimate average relationships between **population densities** (D) **and geospatial covariates** (x) for different **settlement types** (t), **administrative provinces** (p), and **local administrative areas** (l) **[eq. 3]**. This model set-up enables producing high-resolution population estimates and relative uncertainty measures across the settled grid cells of the targeted provinces.

$$N_i \sim Poisson(D_i A_i) \qquad \textbf{[eq. 1]}$$

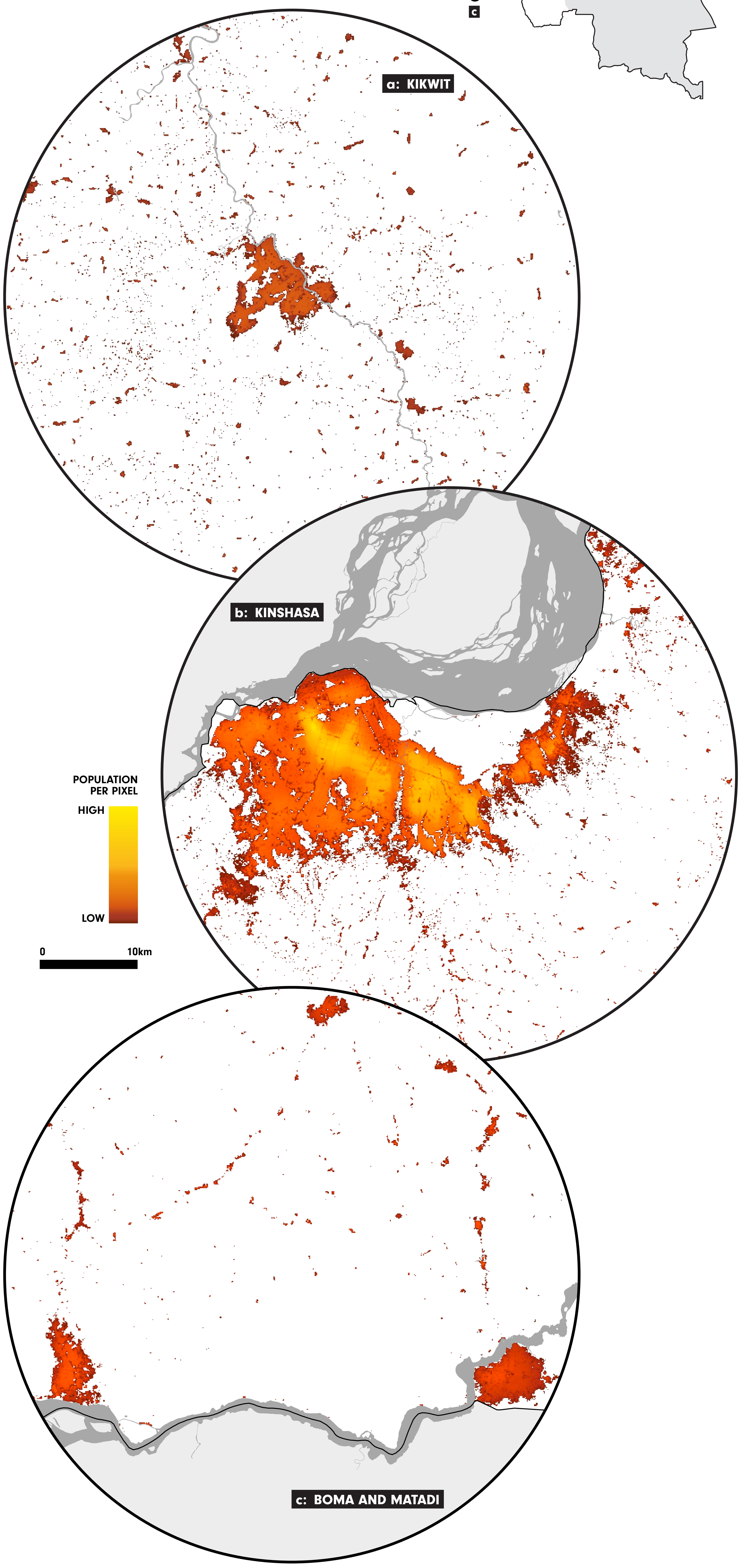
$$D_i \sim LogNormal(\bar{D}_i, \sigma_{t,p,l}) \qquad \textbf{[eq. 2]}$$

$$\bar{D}_i = \alpha_{t,p,l} + \sum_{k=1}^{K} \beta_{k,p} x_{k,i} \qquad \textbf{[eq. 3]}$$

RESULTS

The **hierarchical structure is crucial to the model estimation** because, alone, it already provides a substantial goodness-of-fit [R2=0.43]. When including **four geospatial covariates**, namely, travel distance to cities, density of residential roads, density of tertiary sector activities, and spatially disaggregated population counts based on the previous census, the goodness of fit slightly improves [R2=0.59]. These results are **supported by tenfold cross-validation** as the average goodness-of-fit across 1,000 replications remains similar [R2=0.55].

The mean population estimates enable for **high-resolution mapping of population** within the main cities in the western part of DRC — **KINSHASA**, **BOMA AND MATADI**, and **KIKWIT** — and the vast rural hinterland. The associated uncertainty measures also allow for the identification of **locations where the population estimates are potentially less accurate**. The results of the model iterations can be **aggregated within geographic boundaries** to provide population totals and uncertainty measures.



CONCLUSIONS

The proposed bottom-up modeling approach enables **overcoming the lack of an actionable national census** by producing high-resolution population estimates based on alternative demographic data sources. Implementing micro-census surveys has the advantage of being a **cost-effective way to produce up-to-date estimates for spatial epidemiological applications**. However, these estimates **can not provide the same richness of demographic information** provided in the census.

Future work will aim at **improving the accuracy of population mapping**, which is currently based on a settlement map derived from satellite imagery retrieved during the period 2008-2016. An outdated settlement map involves the risk of **not capturing the most recent settlement expansions and consequent population underestimation**. For this purpose, we will explore a model-based update of the settlement map through a measurement error submodel that will include additional covariates related to settlement expansion.

REFERENCES

[1] United Nations, Department of Economic and Social Affairs, Population Division. 2019. "World Population Prospects 2019: Data Booket".

[2] Lumbala C. et al. 2015. "Human African Trypanosomiasis in the Democratic Republic of the Congo: Disease Distribution and Risk." International Journal of Health Geographics 14 (1).

[3] Wardrop NA. et al. 2018. "Spatially Disaggregated Population Estimates in the Absence of National Population and Housing Census Data." Proceedings of the National Academy of Sciences 115 (14).

[4] Plummer M. 2003. "JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling." Proceedings of the 3rd International Workshop on Distributed Statistical Computing 124 (125).

AKNOWLEDGEMENTS

This work was funded by the Bill and Melinda Gates Foundation and the United Kingdom Department for International Development (OPP1182408).
The authors acknowledge the use of the IRIDIS 5 High-Performance Computing Facility, and associated support services at the University of Southampton, in the completion of this work.

gianluca.boo@soton.ac.uk @gianluca_boo www.worldpop.org