Micro-estimates of Multidimensional Child Poverty in sub-Saharan Africa

Marina Vicini, Oliver Fiala

February 3, 2023

In the world, 1.2 billion children suffer from child poverty (1). Child poverty, defined as the lack of realization of rights constitutive of poverty, has negative long-lasting effects on children. The dimensions included in the internationally comparable multidimensional index analyzed in this work are sanitation, water, housing, health, nutrition, and education. All dimensions are equally weighted since all rights are equally important (2; 3).

National statistics summarize important information but they cannot capture the intricate distribution of where poor children are located. In this study, we aim to provide finely-grained maps of prevalence, depth, and the other deprivations. We build our predictions on a hexagonal grid, developed by Uber (4), where each hexagon has an average area of $5.16\,\mathrm{km}^2$. The hexagonal shape is chosen due to the uniformity of neighbors and to reduce sampling bias from edge effects, which is attributed to a high perimeter-area ratio. From DHS surveys, we derive the ground truth data, mapping each child to a hexagon and taking the average of children in the same hexagon. We select a threshold of at least 30 children per hexagon. Of the 48 countries in sub-Saharan Africa, 25 have recent DHS surveys. To train the model, we consider alternative georeferenced data sources: Google Earth Engine, Uppsala Conflict Data Program, Open Street Map, Ookla Open Data, OpenCellID, Meta's Data for Good repository, WorldPop. The variables extracted from these sources have been aggregated at the hexagonal level, and can be observed in Table 1. To deal with the dislocation of GPS coordinates of DHS surveys, we copy the output to 1-ring neighboring cells in urban areas and 2-ring in rural areas. This "neighboring" approach has been implemented as a way to introduce smoothness, augment the data and account for DHS location displacement

Data	Source
Conflict Zones	Uppsala Conflict
Road Density	Open Street Map
Critical Infractures	Open Street Map
Connectivity Speed	Ookla
Cell towers	OpenCellID
GDP	Aalto University
Wealth	Meta for Good
Commuting Zones	Meta for Good
Elevation	Google Earth Engine
Vegetation, Water	Google Earth Engine
Precipitation	Google Earth Engine
Human Settlement	Google Earth Engine
Travel time to hospital	Google Earth Engine
Night light intensity	Google Earth Engine
Pollution	Google Earth Engine
Population	World Pop

Table 1: Input Data and respective source

(5).

The model used is XGBoost, a gradient boosting decision tree method. The metrics considered are mean square error during training and R^2 to evaluate. Moreover, since we are dealing with geographical data, to have a more generalizable model and avoid overfitting, we employ spatial cross validation.

Hence, two experiments have been conducted.

1. In the first, we predict multidimensionally child poverty only in the 25 countries that have a recent DHS survey. The data have been processed with robust scaling and KNN imputer.

2. In the second, we build a model to be able to generalize on the countries that do not have DHS information. We randomly split the countries for which we have ground truth data in training and test, and we start modeling using XGBoost. Here we do not use the neighboring approach to have a more generalizable model, and we process the data with robust scaling and median imputer.

The performance in the first experiment is higher than the second one, but models specific to one country cannot generalize well, while the models of the second experiment can generalize even on countries without DHS data.

To facilitate responsible downstream use of the predictions, we include prediction intervals to assess uncertainty. A prediction interval is an estimate of an interval in which a future observation will fall, with a certain probability, given what has already been observed. Prediction intervals are not the same as confidence intervals, and they are wider. In this work, we consider the 95% prediction intervals for each data point. We use a Model Agnostic Prediction Interval Estimator (MAPIE) to compute them (6).

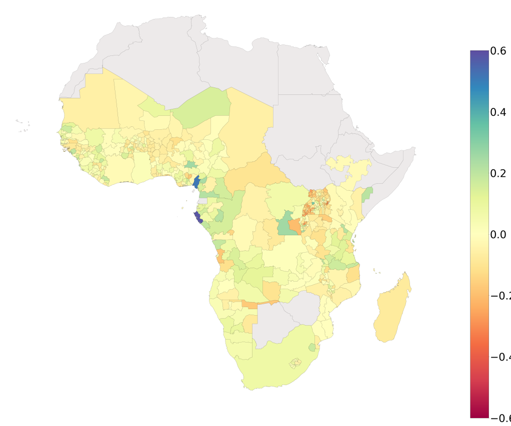

Figure 1: Difference between predicted values (weighted on child population) and weighted DHS / MICS measurements.

To have a more transparent model, we interpret the results in terms of overall importance of each feature through SHAP (SHaply Additive exPlanations) values (7), the feature importance is measured averaging the marginal contributions of the predictions across all permutations. The most predictive feature for prevalence has been found to be nighttime light intensity, that is in concordance with other results found in literature (8).

Lastly, we compare the aggregated predictions aggregated with the DHS and MICS subnational and national values, and their difference can be observed in Figure 1. In conclusion, this methodology can provide finely grained prediction of multidimensional child poverty at a $5.16\,\text{km}^2$ resolution, estimating prevalence, depth and other dimensions.

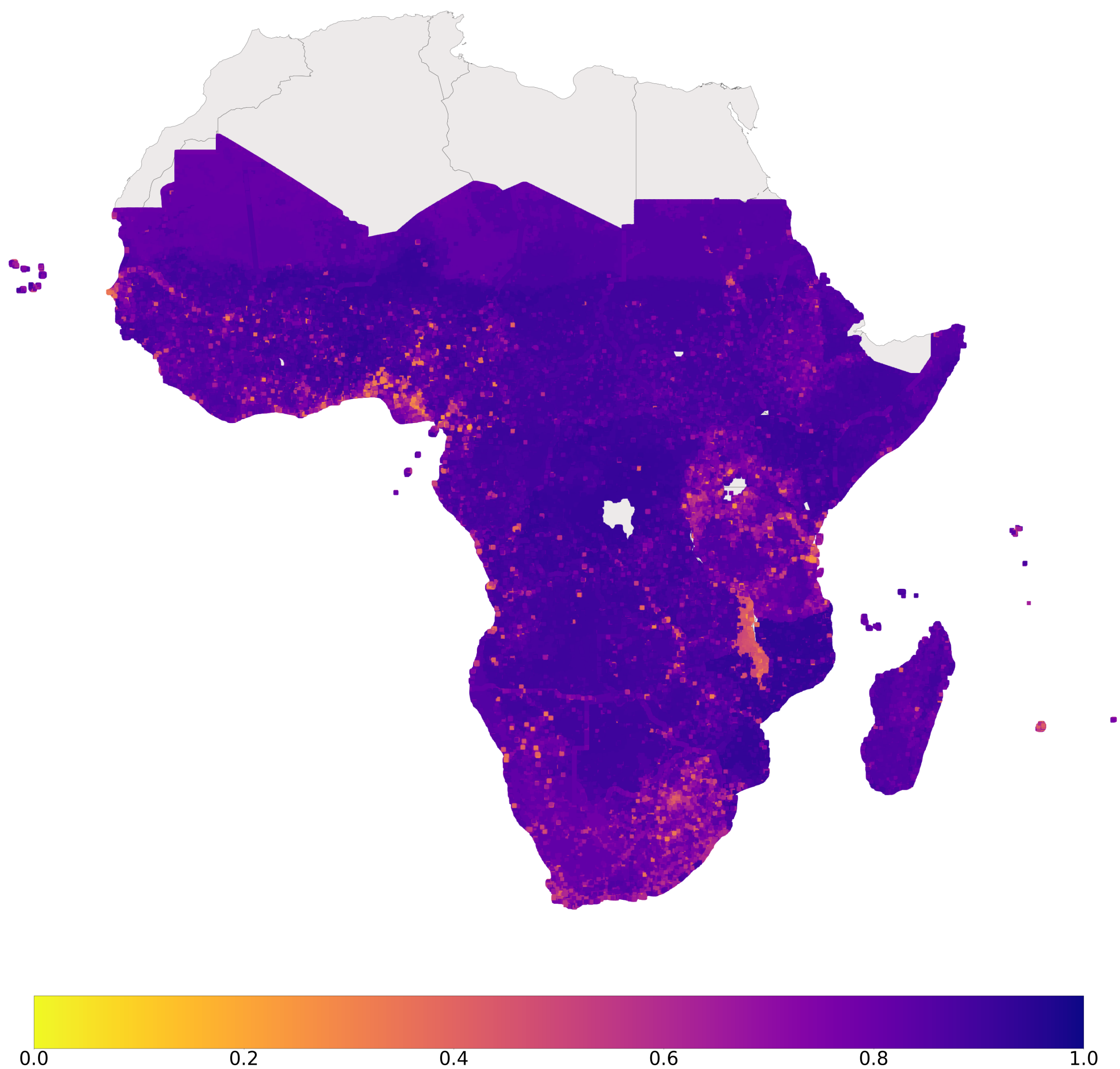

Figure 2: Distribution of prevalence in sub-Saharan Africa.

References

[1] UNICEF and Save the Children, "Impact of covid-19 on children living in poverty: A technical note," tech. rep., 2020.

[2] UN Committee on the Rights of the Child, "General guidelines regarding the form and content of periodic reports to be submitted by states parties under article 44, paragraph 1 (b), of the convention," *UN Doc. CRC/C/58/Rev.1*, 2005.

[3] M. R. Hagerty and K. C. Land, "Constructing summary indices of quality of life: A model for the effect of heterogeneous importance weights," *Sociological Methods & Research*, vol. 35, no. 4, pp. 455–496, 2007.

[4] Uber Technologies Inc., "H3: Uber's hexagonal hierarchical spatial index."

[5] G. Chi, H. Fang, S. Chatterjee, and J. E. Blumenstock, "Microestimates of wealth for all low- and middle-income countries," *Proceedings of the National Academy of Sciences*, vol. 119, no. 3, p. e2113658119, 2022.

[6] V. Taquet, V. Blot, T. Morzadec, L. Lacombe, and N. Brunel, "MAPIE: an open-source library for distribution-free uncertainty quantification," *CoRR*, vol. abs/2207.12274, 2022.

[7] S. M. Lundberg and S. Lee, "A unified approach to interpreting model predictions," pp. 4765–4774, 2017.

[8] N. Jean, M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon, "Combining satellite imagery and machine learning to predict poverty," *Science*, vol. 353, no. 6301, pp. 790–794, 2016.