Global Subnational Population Projection (2015-2030)

Duygu Cihan[1], Jason Hilton[2], Wenbin Zhang[1], Maksym Bondarenko[1], Dorothea Woods[1], Tom McKeen[1], Alexander Cunningham[1], Rhorom Priyatikanto[1], Alessandro Sorichetta[3], Thomas Brinkhoff[4], Andrew Tatem[1]

1.WorldPop, School of Geography and Environmental Science, University of Southampton, SO17 1BJ, UK;
2.WorldPop, School of Economic, Social and Political Sciences, University of Southampton, SO17 1BJ, UK;
3. Department of Earth Sciences, Universita Degli Studi Di Milano, Via Mangiagalli, 34 20133 MILANO (MI);
4. Institute of Applied Photogrammetry and Geoinformatics (IAPG), Jade University, Hansa-Ring 10, D-26133 Oldenburg (DEU)

Background & Aim

In recent years, the demand for subnational population projections has significantly increased. These projections are crucial for policy-making, public health, humanitarian efforts, and development planning, as they provide detailed population estimates at smaller geographic scales, facilitating more precise future planning. Previously, WorldPop produced subnational population estimates covering period 2010 to 2020 as part of a Global Demographic data project funded by Gates Foundation. These datasets have been extensively utilized by United Nations agencies, international development organizations (such as the IMF, USAID, and FCDO), public health organizations (including the CDC and GAVI), national statistics offices, ministries of health, environmental and urban planning departments, humanitarian organizations, and non-governmental organizations, among others. Due to a high demand for updated projections, a successor project (Global 2) was launched in 2022 to generate high-resolution subnational population estimates for the period from 2015 to 2030. This initiative involved the creation of a unique dataset by assembling subnational age- and sex-structured population data along with corresponding administrative boundaries, while simultaneously developing an improved methodology for generating these projections.

Data Collection

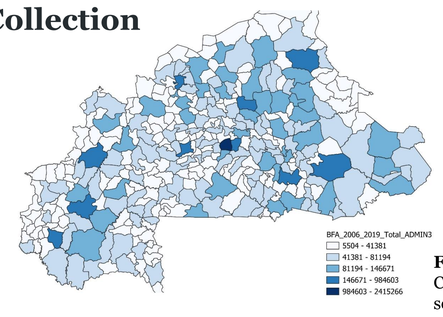

Figure 1: Burkina Faso 2019 Census- Admin 3 population, sourced from CityPopulation

Subnational population data matched to associated digital boundaries from a variety of different sources, including Gridded Population of the World (GPW.v4), CityPopulation, UNFPA, and the US Census Bureau, were obtained for 242 countries and territories aiming to provide complete coverage of the global population. Preference was given to census or population register data, and to age-and-sex specific data at the lowest possible administrative geography. Where census data wasn't available, the best quality alternative source of subnational population data was chosen.

To allow demographic change over time to be captured, data for two time points were obtained for each country, corresponding in most cases to census from the 2010 and 2020 census rounds, except in a small number of cases where only one time point could be found. In order to provide estimates that are consistent with numbers used elsewhere by international agencies, the United Nations World Population Prospects (UNWPP) 2024 data were utilized in the projection process.

Methodology

To support this global project, we collected data from a total of 242 countries, with a focus on prioritizing population census and registration data. The most recent census data were obtained, and for each country, the two latest censuses were used in the analysis. In instances where census data were not available, alternative data sources, such as official population estimates, were employed. The data processing phase involved reading and handling input files, addressing various issues, and extracting relevant information to organize the data in a structured format.

Following data processing, we standardized the age and sex structures of each dataset using the Penalized Composite Link Model (PCLM). This standardization step was essential to ensure consistency and comparability across datasets from different countries. The standardized data were then used for population projections, employing a novel approach tailored to subnational projection methods. This approach consists of three models, each based on the growth of the population share and the total population (see Figure 2):

1.Model 1 (Linear Model): Calculates the change in population share for each administrative unit during the intercensal period.
2.Model 2 (Exponential Growth Model): Utilizes a non-linear time term to model exponential growth in population share.
3.Model 3 (Hybrid Model): Combines elements of both the linear and exponential models to accommodate varying growth dynamics.

The choice of model depends on the following condition:

$$\boldsymbol{I\big((\pi_{it}-\pi_{i0})(P_{it}-P_{i0})>0\big)}$$

This condition determines the appropriate model based on the simultaneous growth of the population share and the total population, considering four scenarios: (1) both population share, and population increase, (2) both decrease, (3) population increases with an initial zero population share, and (4) population share, and total population change in opposite directions. When the population share and total population increase or decrease together, the exponential growth model (Model 2) is selected. When they move in opposite directions, the linear model (Model 1) is applied. This selection process helps prevent overshooting, ensures realistic projections, and aligns the projections with UN constraints. After completing the projection process, we disaggregated the total population projections by age and sex using Iterative Proportional Fitting (IPF). In this step, the UN World Population Prospects 2024 (UNWPP2024) age-sex structures were used as margins, and our data were adjusted accordingly. The final output provides age- and sex-disaggregated subnational population projections for the period from 2015 to 2030.

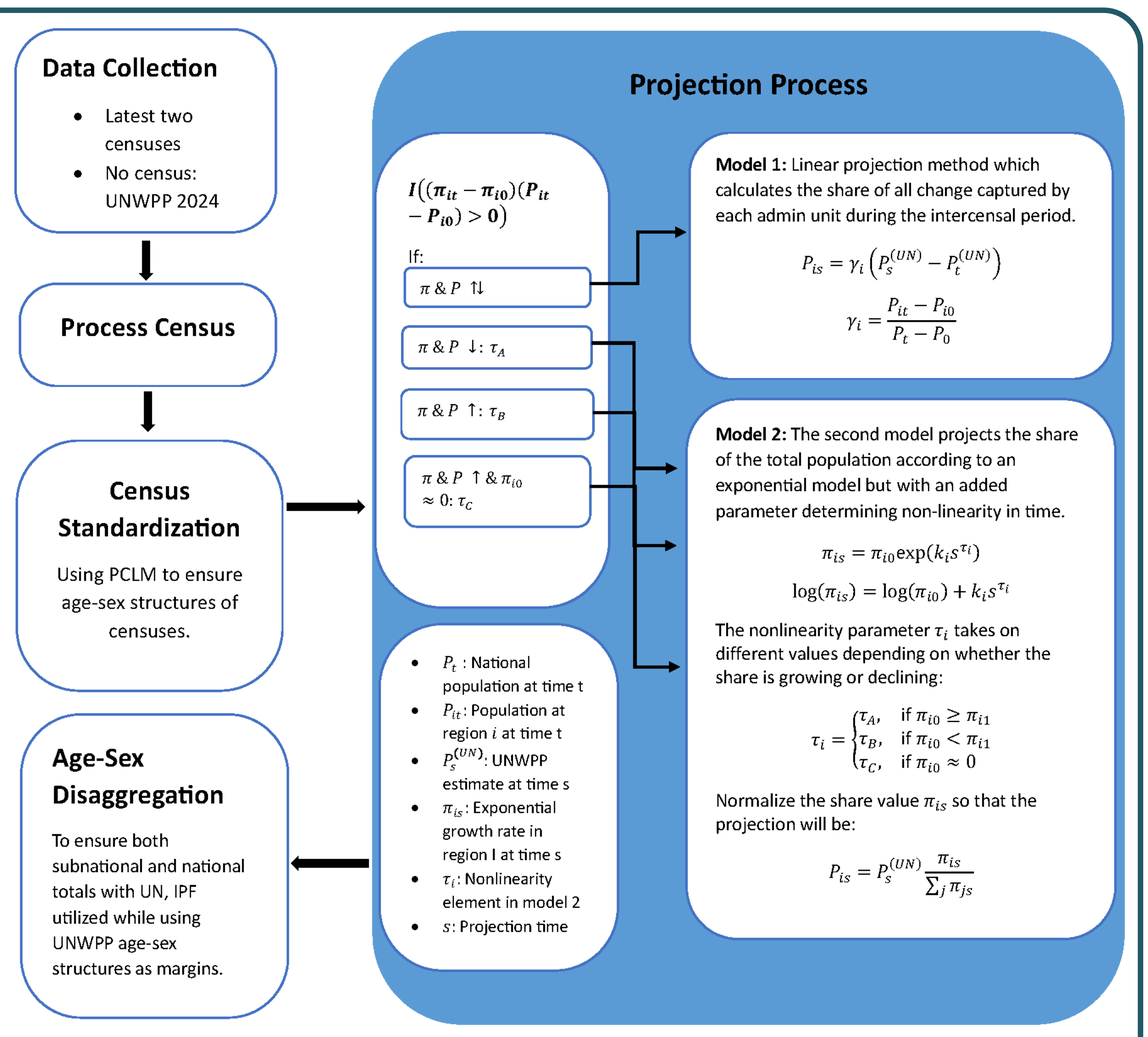

Figure 2: Flow Chart of the project processes and Methodology

Results

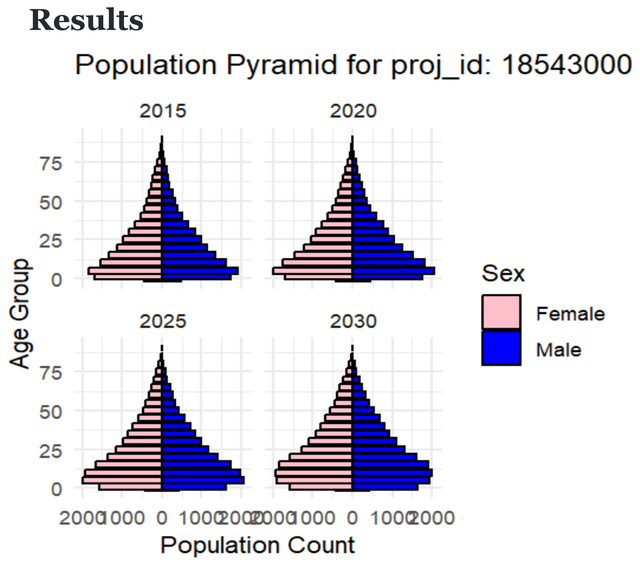

Figure 3: Population Pyramid for a randomly selected ADM3 region for the years 2015, 2020, 2025 and 2030

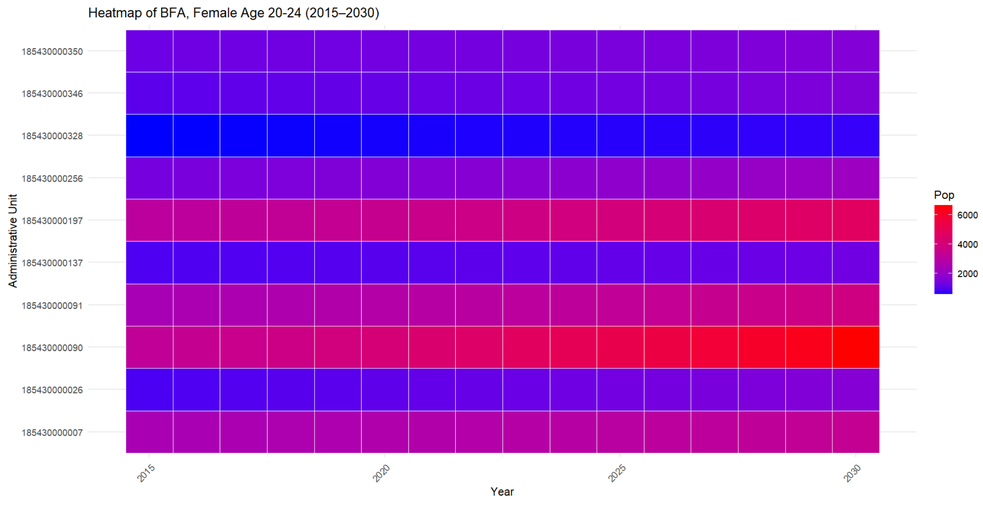

Figure 4: Heatmap for randomly selected 10 ADM3 region for females aged 20-24, in the period 2015 to 2030

With this project, we developed a novel method for subnational population projection and produced a unique global dataset that provides age and sex-disaggregated population estimates for 242 countries and approximately 716,000 subnational units. The dataset covers custom age groups, including 0, 1, 5-year intervals up to 85, and 90+. Furthermore, the results are fully aligned at the national level with the United Nations World Population Prospects (UNWPP) 2024 data.

References

[1] Hyndman, R.J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice* (2nd ed.). OTexts: Melbourne, Australia

[2] Robert M. Leddy, Jr. 2013. "METHODS FOR CALCULATING 5-YEAR AGE GROUP POPULATION ESTIMATES BY SEX FOR SUBNATIONAL AREAS". Geographic Studies Branch Population Division U.S. Census Bureau.

[3] George, M.V., Stanley K. Smith, David A. Swanson, and Jeff Tayman. 2004. "Population Projections." Chapter 21 in The Methods and Materials of Demography. Second Edition. Edited by Jacob S. Siegel and David A. Swanson. San Diego, CA: Elsevier Academic Press.

[4] Judson, D.H. and Carole L. Popoff. 2004. "Selected General Methods." Appendix C in The Methods and Materials of Demography. Second Edition. Edited by Jacob S. Siegel and David A. Swanson. San Diego, CA: Elsevier Academic Press