The bottom-up modelling approach

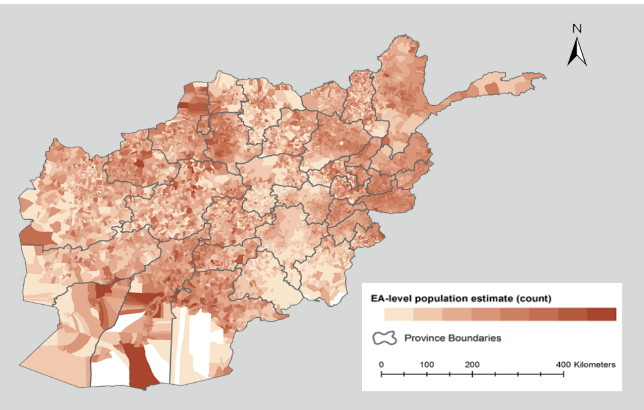
world pop
FLOWMINDER.ORG

Reliable data on population distributions and demographics ensures visibility, effectiveness, and robust decision making. However, in many countries, this data may be outdated or incomplete with long delays between updates due to lack of resources, hard to reach areas, security issues, etc.

How the model works: The bottom-up modelling approach

To produce accurate, timely, high-resolution population data, WorldPop at the University of Southampton has developed a novel **bottom-up modelling approach** to estimate populations (Wardrop et al., 2018).

This new method is based on a hierarchical Bayesian spatial statistical model which combines existing population data with satellite imagery to create **100 x 100m gridded population estimates and measures of uncertainty** for the whole country or targeted areas.



Working closely with the Afghan National Statistics and Information Authority (NSIA) and UNFPA, Flowminder and WorldPop at the University of Southampton applied a hybrid-census approach. Using the Bayesian bottom-up modelling method described here, we produced robust population counts - together with measures of uncertainty - for the national, province, district, and census enumeration area levels, and as 100m x 100m grids.

Data inputs

Population data

Sourced from:
Population Census Surveys with demographic, health, and household income data; 'Microcensus' conducted for the purpose of modelling.

Settlement data

Examples:
Settlement types
Individual building shapes
Point locations
Associated characteristics

Other geospatial data

Examples:
Elevation
Slope
Vegetation types
Accessibility to major cities

↓

Statistical model

↓

Data outputs

Gridded population estimates, with associated measures of uncertainty

Understanding the input data

Population data

The Bayesian modelling method requires recent population data from the country or area of interest **as a foundation to predict population**. This population data can come from an existing, partially complete national population census, other data sets like the Demographic and Health Surveys (DHS), Household Income and Expenditure Surveys (HIIES), or from a "microcensus" conducted for the purpose of modelling.

A microcensus is a set of complete population counts collected within small areas that are sampled from across a country. By sampling different housing types, socioeconomic and physical settings, using them as model inputs, we can better predict the geographical variations in population and demographic characteristics.

Settlement data

Settlement data is vital to provide **information on locations of potentially inhabited structures** for accurate high-resolution population estimates. The settlement data can be derived from high-resolution aerial or satellite imagery or other building maps and can take different forms: settlement boundaries, individual building shapes, or point locations.

Additional information on the characteristics associated with settlements or buildings (urban versus rural, industrial vs residential for example) can add information to produce more accurate population estimates (Weber et al., 2018).

Geospatial covariate data

Geospatial covariate data refers to a range of other information types available which can help predict population **based on the local context** (for example: elevation, slope, vegetation types, accessibility to major cities or other socioeconomic or physical measures).

Relationships between these covariates and population density and demographics can be quantified for areas where population survey data has been collected, and used to estimate population for areas with no population data.

Results and Outputs

Our modelling produces **population estimates at a resolution of 100m x 100m** (which we call 'grid cells'). Gridded population estimates give greater flexibility to group the data into larger administrative units or service catchment areas, based on the data users' needs. For each grid cell, the model predicts a range of potential population densities based on the input data. The mean value of those densities is the best estimate of the population for that area.

We also produce **information on the uncertainty attached to each estimate**. These uncertainty measures provide upper and lower figures around the mean estimate and indicate the degree of confidence in each estimate. The uncertainty measures can be helpful in informing decisions as they can make logistics and interventions more appropriate and/or cost effective.

References

Blangiardo, M., and Cameletti, M. (2015)
Wardrop, et al. (2018)
Weber, E. M., et al. (2018)

For more information about uncertainty measures and how to use them, please read our fact file on "The value of uncertainty measures in population estimation".