Exploring Geospatial-based Approaches to Develop a Pre-Census National Sampling Frame in Armenia

Saida Ismailakhunova[1]; Avralt-Od Purevjav[1]; Tsenguunjav Byambasuren[1]; Sarchil Qader[2]

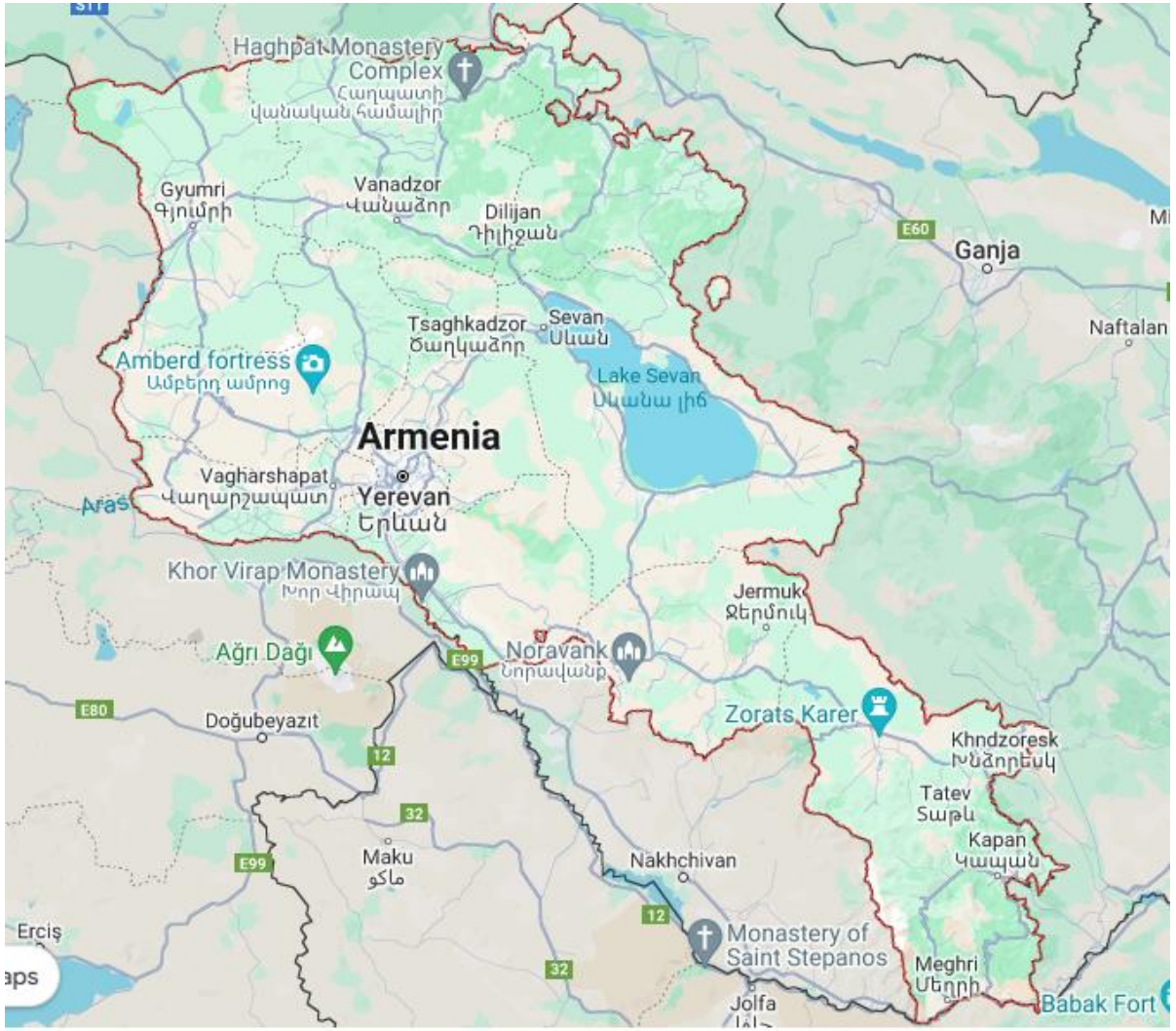
[1]World Bank; [2]WorldPop/University of Southampton/UK

ECA Poverty Meeting
30th of October 2024

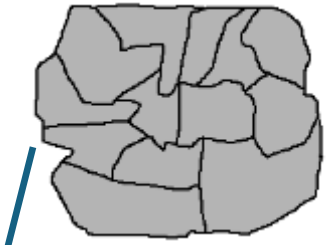
THE WORLD BANK

Background

- The last census completed in Armenia was conducted in 2011
- No up-to-date digitised enumeration areas or national sampling frames for representative socioeconomic surveys
- Outdated sample frame can lead to biased surveys and statistical invisibility



Challenge with existing sampling frames

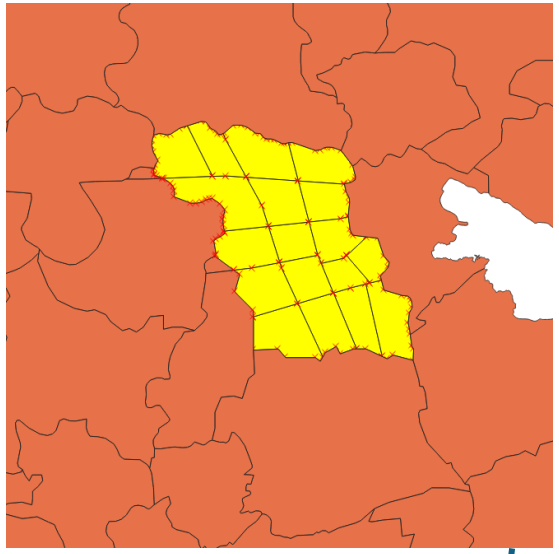


Electoral Precincts

- Boundary information is often unavailable
- Cover massive geographic areas and population
- Substantial resource is required for sampling
- Missing under 18 years old

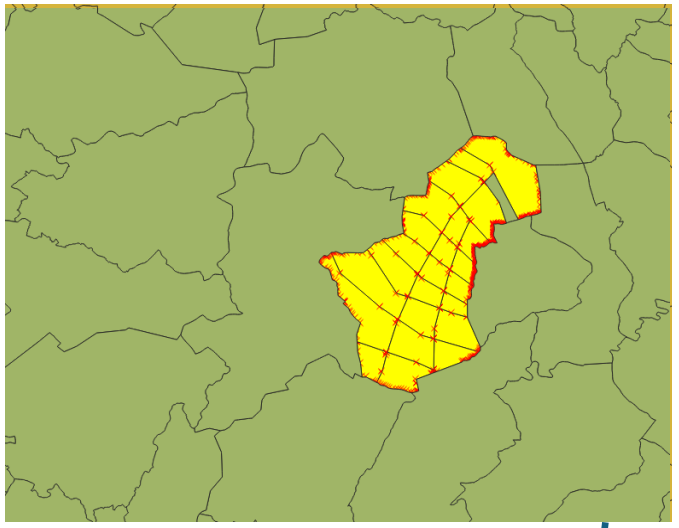
Registration data

Not available or outdated



Administrative boundary

- Cover massive geographic areas and population
- Substantial resource is required for further sampling



Enumeration Area

- Unavailable or outdated or incomplete
- Cover massive geographic areas and population
- Substantial resource is required for further sampling

Gridded Population Sampling

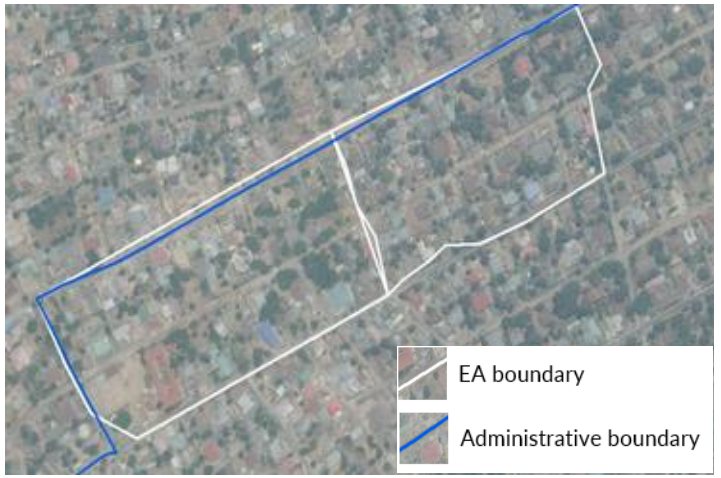
- High uncertainty at grid level
- Cut buildings and structures

Manual EA delineation

- Many countries are still digitising their EAs only through manual approaches. However, relying solely on a manual approach could lead to the following challenges:
  - Significant labour and resource requirements
  - Delays to the survey and census process
  - Various geometric errors such as overlap, gaps, disjoins, etc.
  - Heterogeneity among EAs nationwide in terms of total household/buildings and geographic area

Solution:

We think a combination of manual and automatic approach could be a viable solution to advance the census mapping.



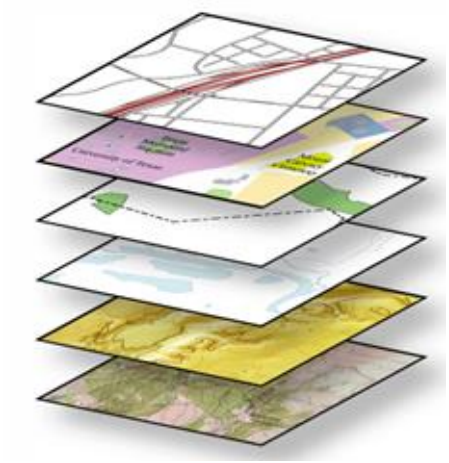
Innovation in the automatic creation of national sampling frame and Enumeration Areas

Input data

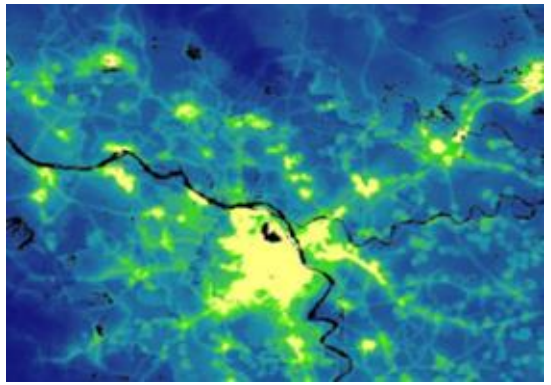
Splitting process

Merging process

Georeferenced layers

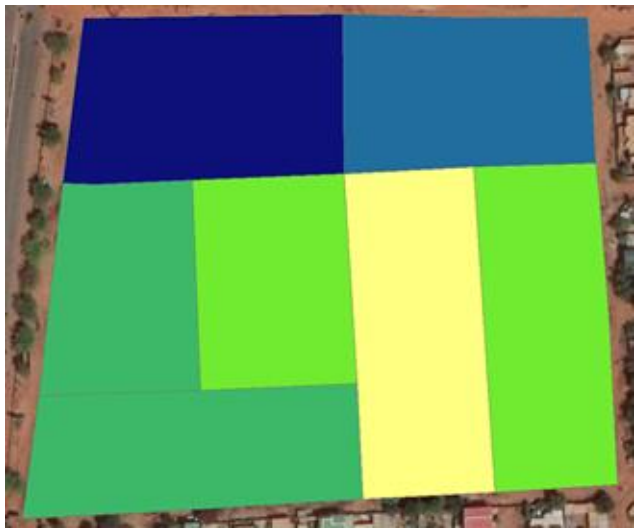


High-resolution gridded population/building data/GPS HH



Building blocks

Building blocks with total population



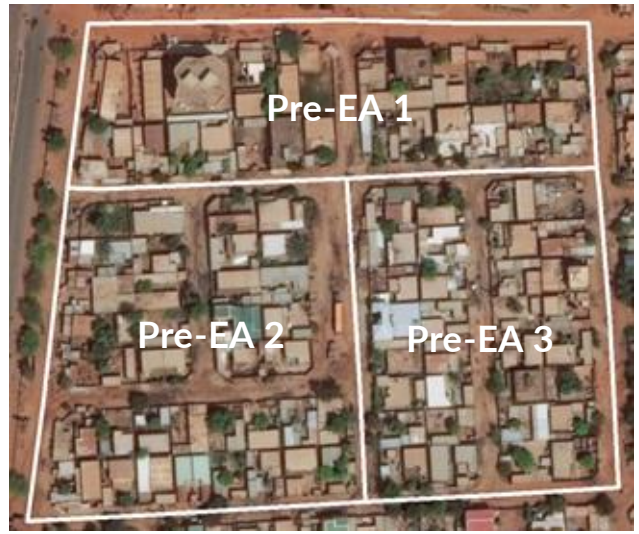
Hard Constraints:

- Population (Max; Min)
- Area (Max; Min)
- Uncrossable Border
- Adm boundary

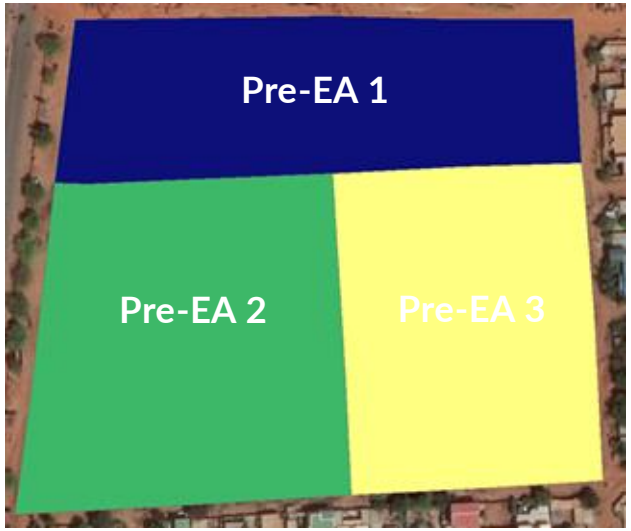
Soft Constraints:

- Shape (Compactness)
- Target (Pop & Area)
- Homogeneity (Health and Socio-economic variables)

Pre-Enumeration Area (EA) outline

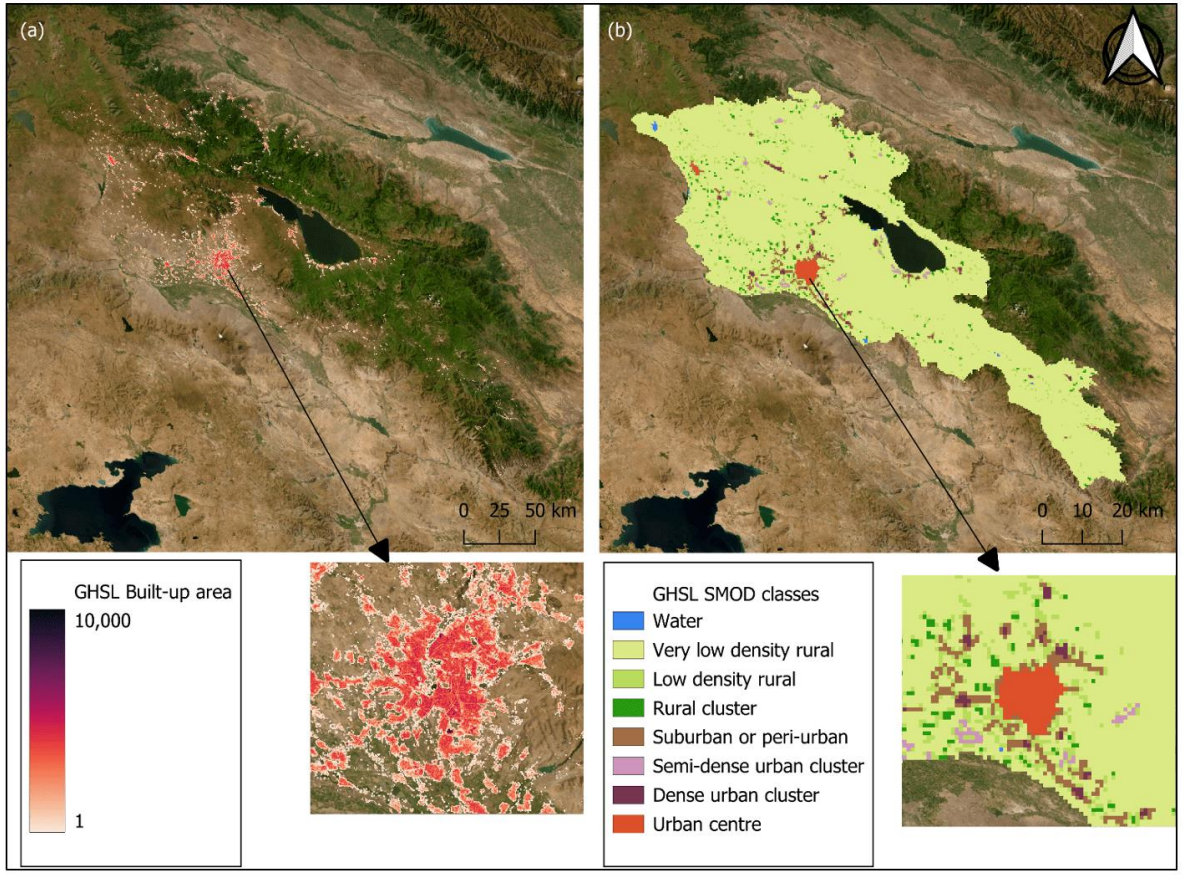


Pre-EAs with total population

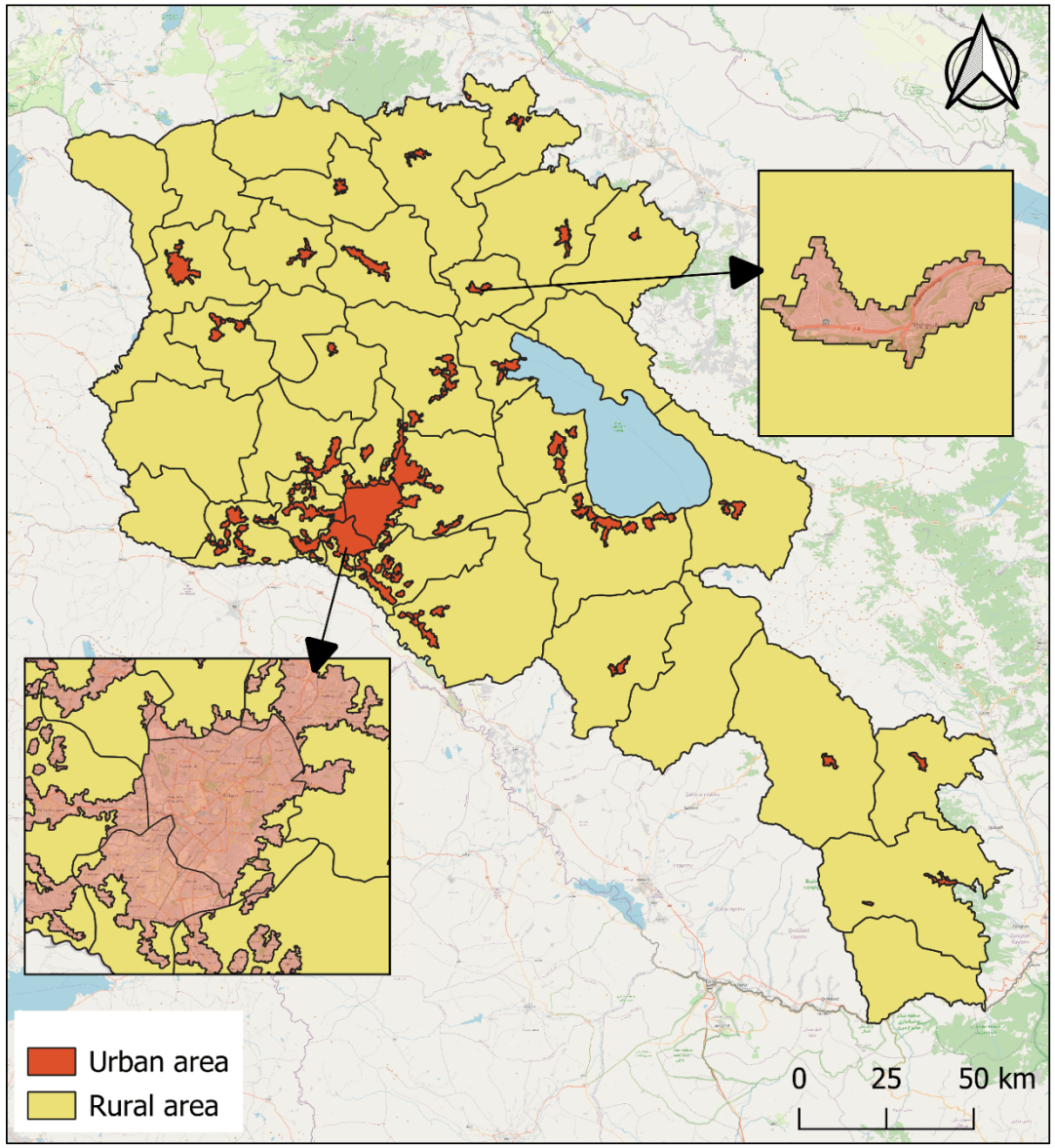


How the PreEA tool works

Urban and Rural Stratification



Panel (a) presents the built-up area, and panel (b) shows the SMOD classes.

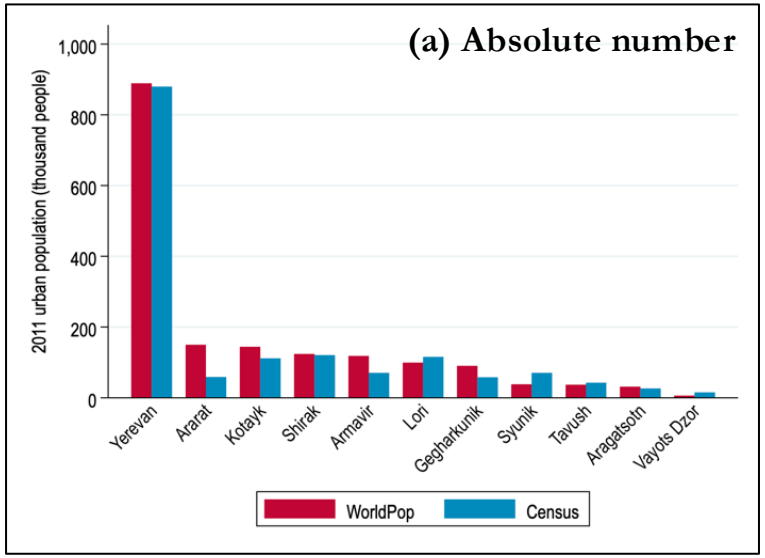


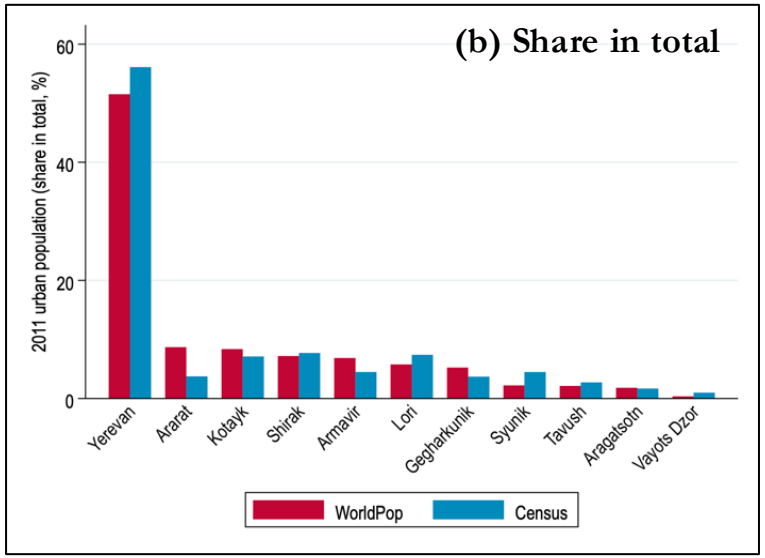
Generated urban and rural areas in Armenia

Urban and Rural Stratification

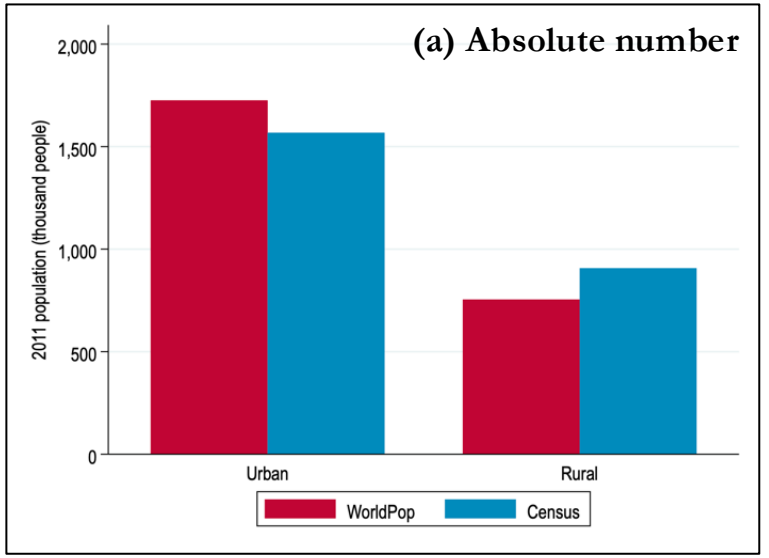
Comparison of Armenia's 2011 census population and WorldPop gridded population estimates for urban-rural classification using census and our geospatial method

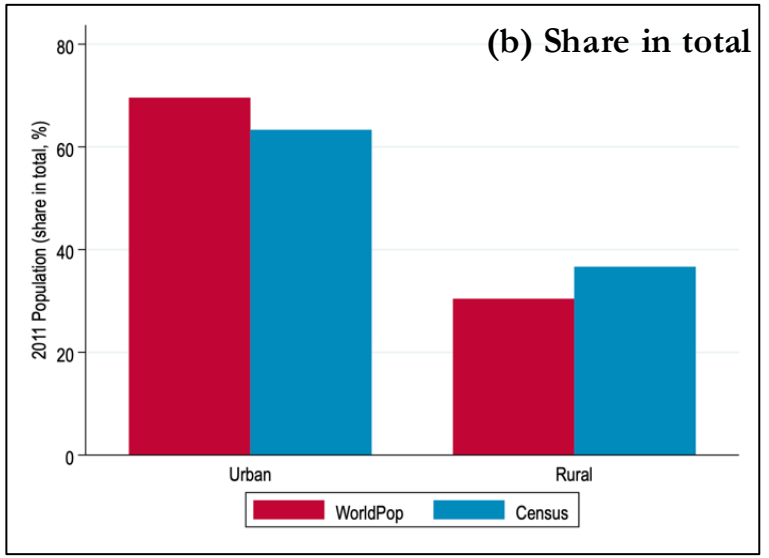
Urban population in *Marzes*, 2011





Urban and Rural Population, 2011

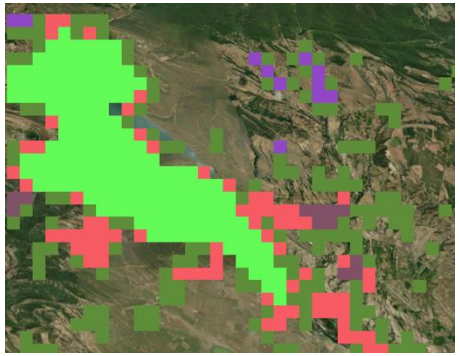


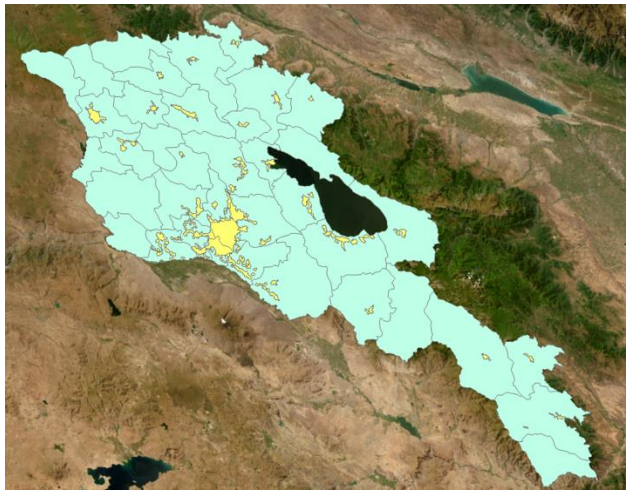


Creating national sampling frame for World Bank Household Budget Survey in Armenia

S1

Degree of urbanization

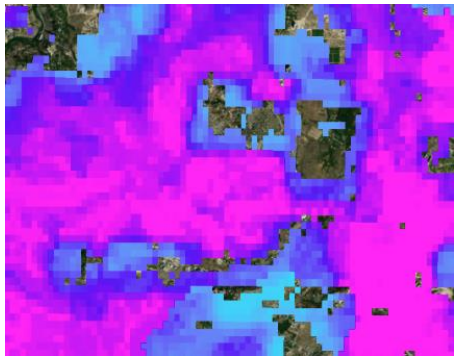


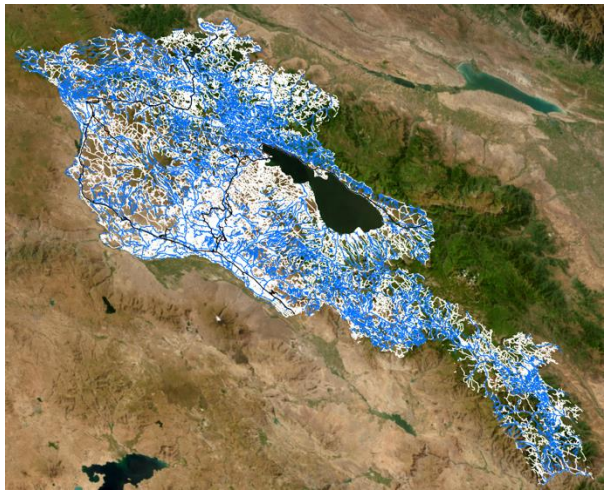


Using the degree of urbanization (GHSL) and various geospatial techniques to define urban and rural strata

S2

Digital boundary and population data





Pre-process and harmonies digitized natural and man-made features for the country. Gridded population estimates (100x100m; e.g. WorldPop) in the absence of recent census data

S3

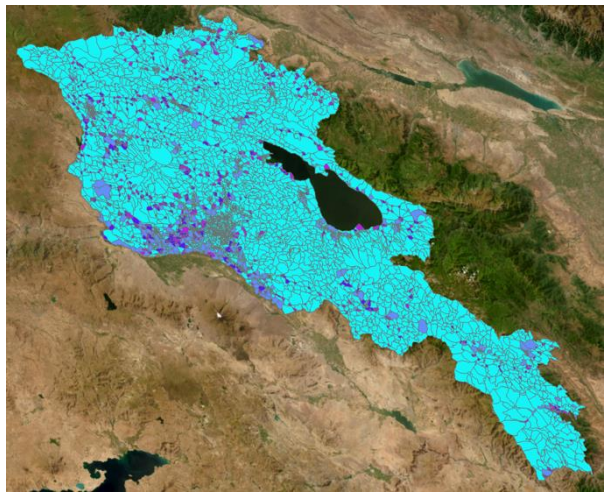
PreEA tool

Tool Parameters: Population; Area; Uncrossable features; Administrative boundary; shape factors; homogeneity, etc.

WorldPop developed a package called the preEA tool that includes several user-friendly tools to support census and national surveys. The national sampling frame can be automatically created by the program based on a variety of parameters.

S4

National Sampling Frame



The output is a shapefile, each preEA has a total estimate of population and geographic area with a unique ID. The preEA outlines are nested within administrative boundaries.

Conclusion

1. The developed approach has been shown to be cost-effective
2. The method employed different innovative geospatial techniques and datasets to classify urban and rural areas and generate a digitised national sampling frame automatically
3. The approach only used publicly available geospatial data, while users may optionally provide their data.
4. The method is simply transferable to other countries
5. The results were free of common geometric issues that occur when manual methods are used.

Thank you