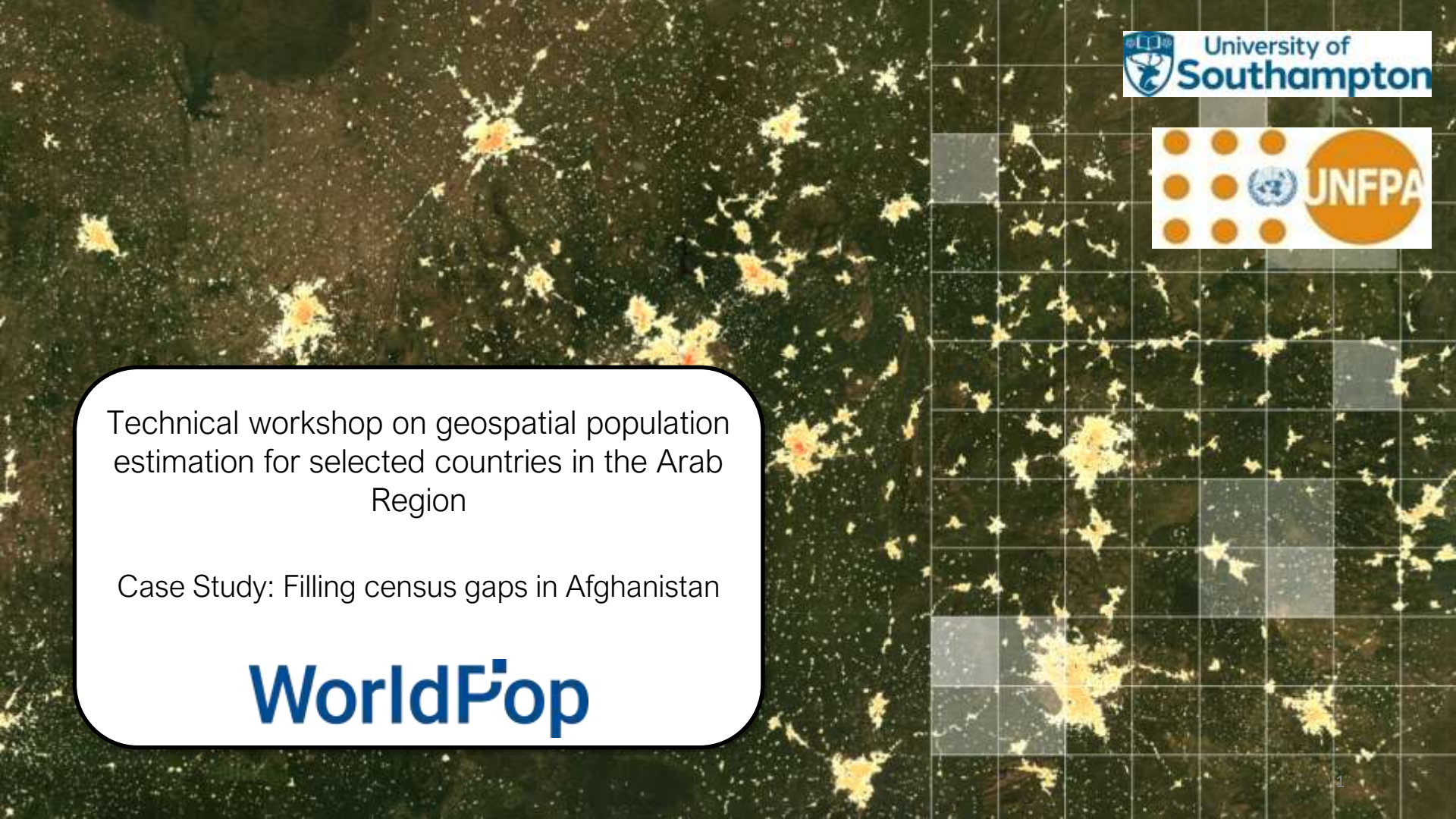


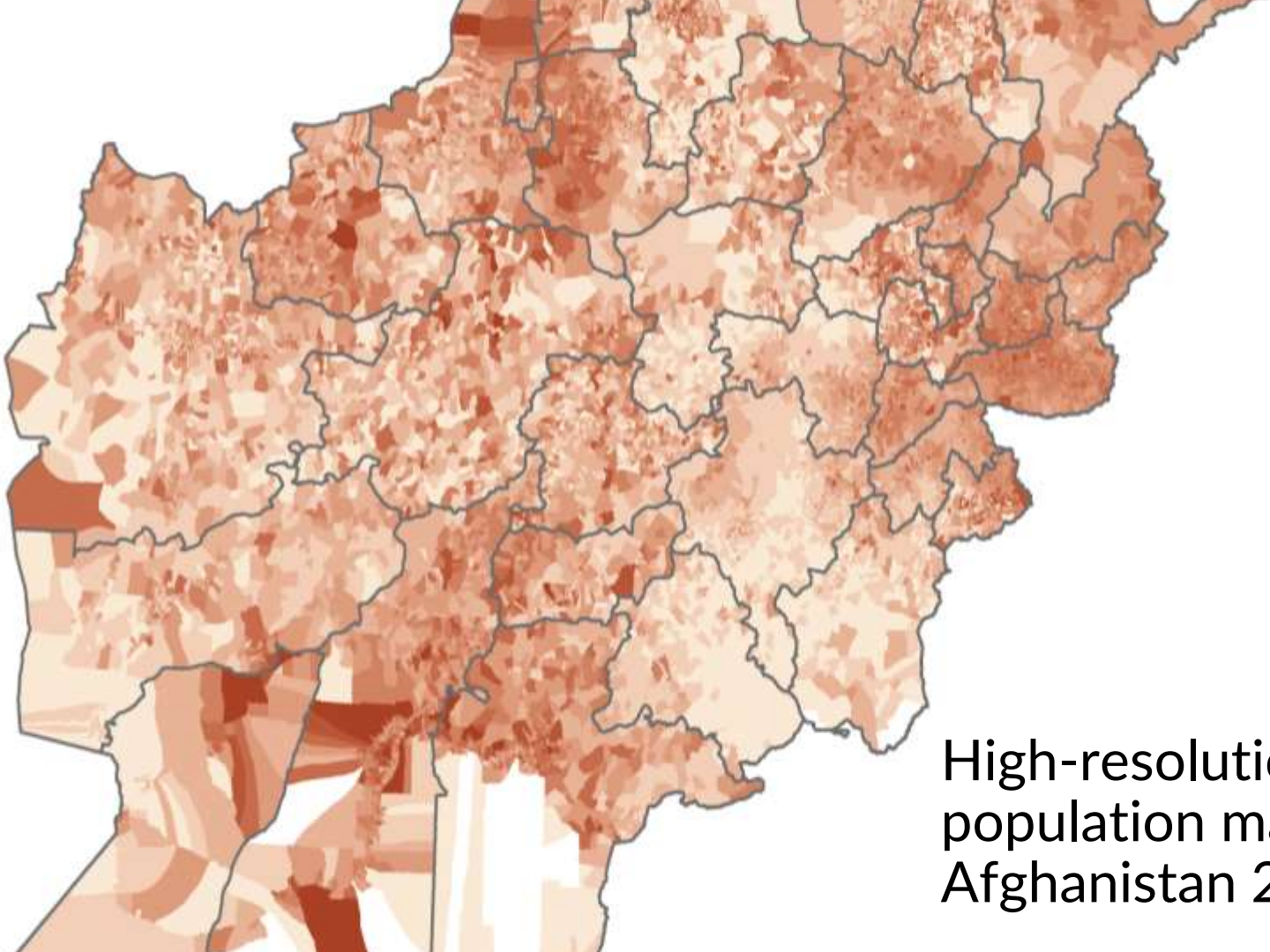
University of Southampton

UNFPA

Technical workshop on geospatial population estimation for selected countries in the Arab Region

Case Study: Filling census gaps in Afghanistan

WorldPop

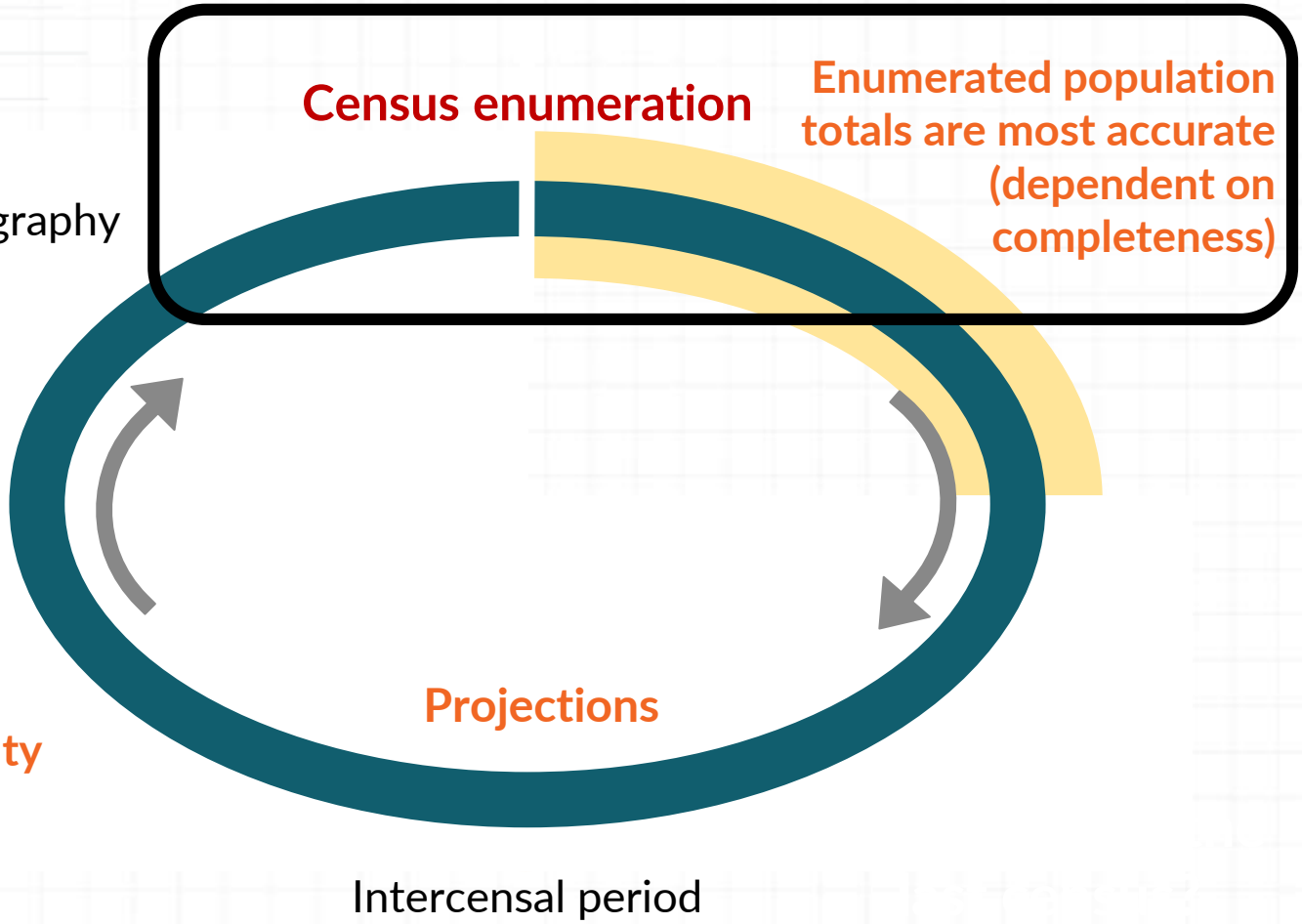


High-resolution population mapping in Afghanistan 2017/18

Background

- Last census conducted in 1979; insecurity prevented a recent one; existing estimates based on 1979 baseline population;
- From 2011-2017 the CSO conducted a form of rolling census, the Socio-Demographic and Economic Survey (SDES) which included enumeration for 50% of households (the survey covered 12 of 34 provinces);
- Challenge: Lack of reliable current disaggregated population data at provincial, district level;
- President Ghani requested UN to assist CSO in estimating spatially disaggregated population data - Collaborative partnership including Government, UNFPA, WorldPop

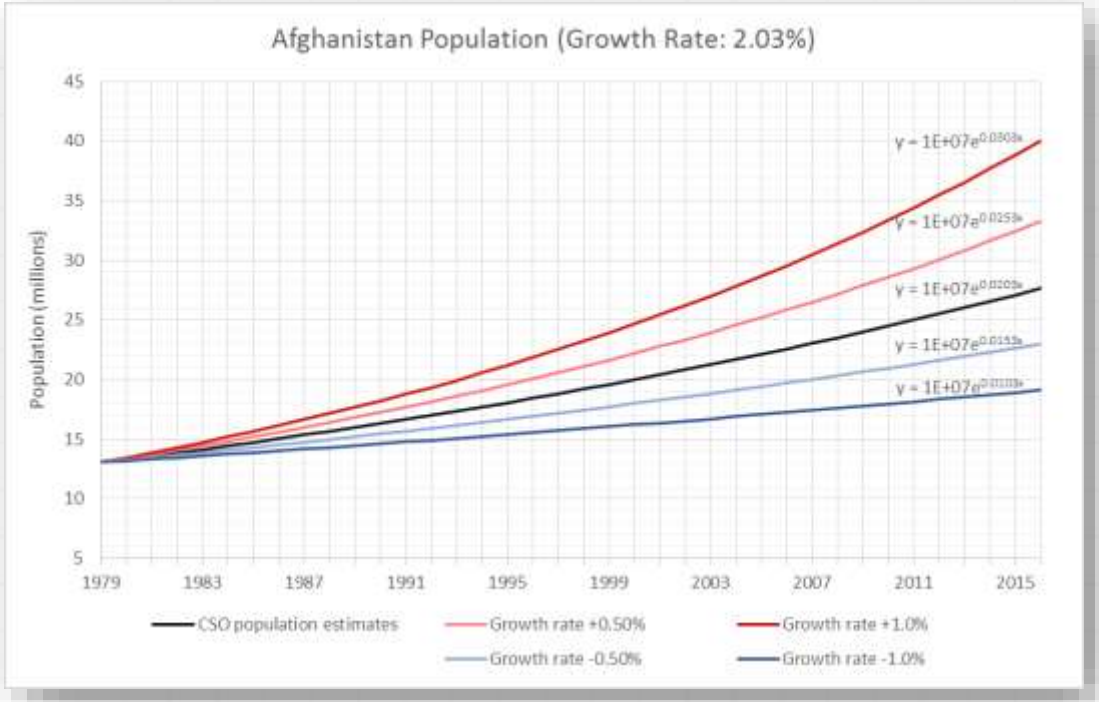
Stages of the census cycle



As time from the census elapses, increasing uncertainty in projections

Existing CSO estimates

- Existing estimates based on 2.03% growth rate from 1979 baseline
- Significant differences with recently gathered SDES numbers (typically underestimates)
- Analyses of newer data suggest that recent growth rates are more likely to be 3% or higher



Afghanistan: Project aims

- Estimate population counts at enumeration area (EA), district, province and national levels and associated measures of uncertainty;
- 100m x 100m gridded population estimates;
- Estimates of the age/sex structure of the population and associated measures of uncertainty
- Co-development with CSO and capacity strengthening

How does the population estimation process work?

Population estimation & mapping process

1. Enumerate population with ground-based surveys in defined areas

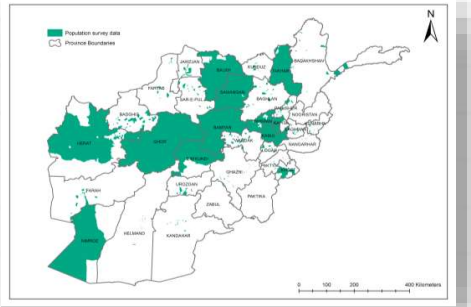
2. Predict population in un-surveyed areas

3. Spatial disaggregation of population estimates

WorldPop

We want to….

- *Estimate population density and total in each EA (with confidence intervals)*
- We have population density and total data for some EAs, but not all
- We do have data on factors related to populations in all EAs from satellites and other geospatial data



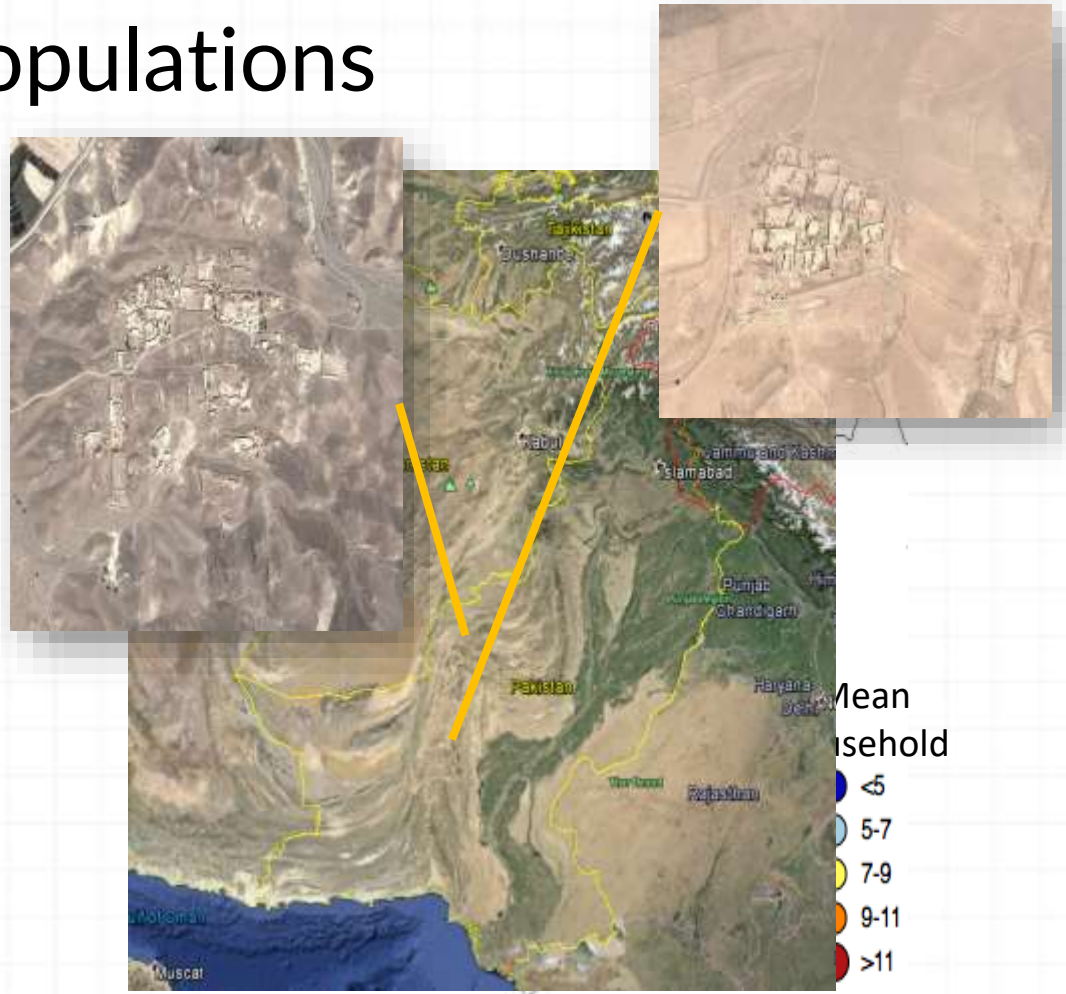
Mapping populations

Mapping buildings/settlements is a valuable initial step

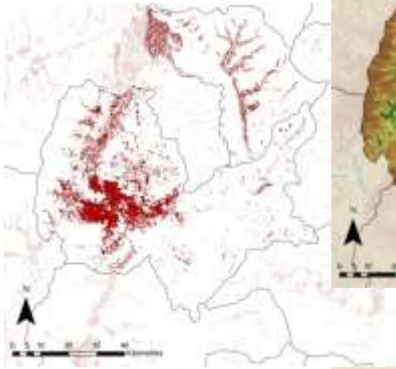
But, buildings do not directly translate to people

Multiple factors drive differences in distributions and demographics

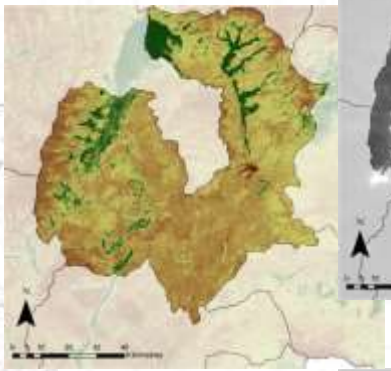
Importance of *integrating multiple data sources*



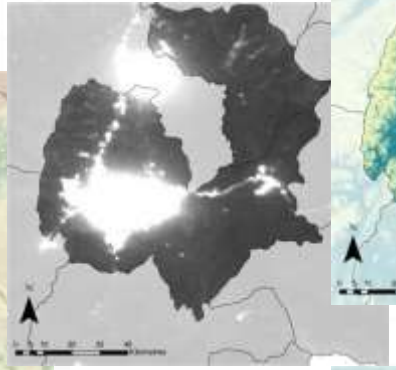
Example geospatial datasets



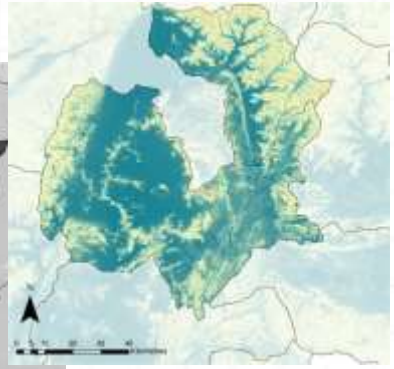
Number of compounds & area settled



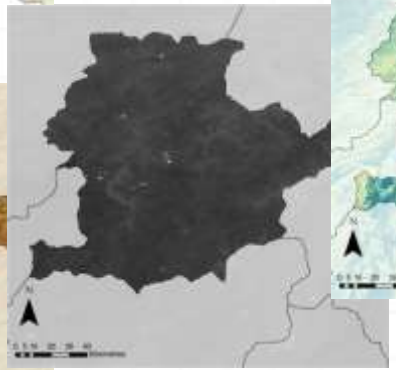
Vegetation index

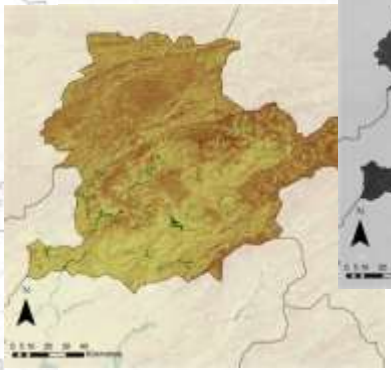


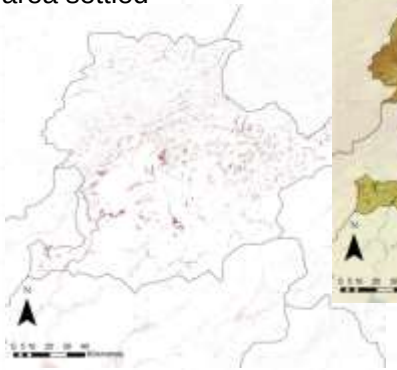
Night-time lights

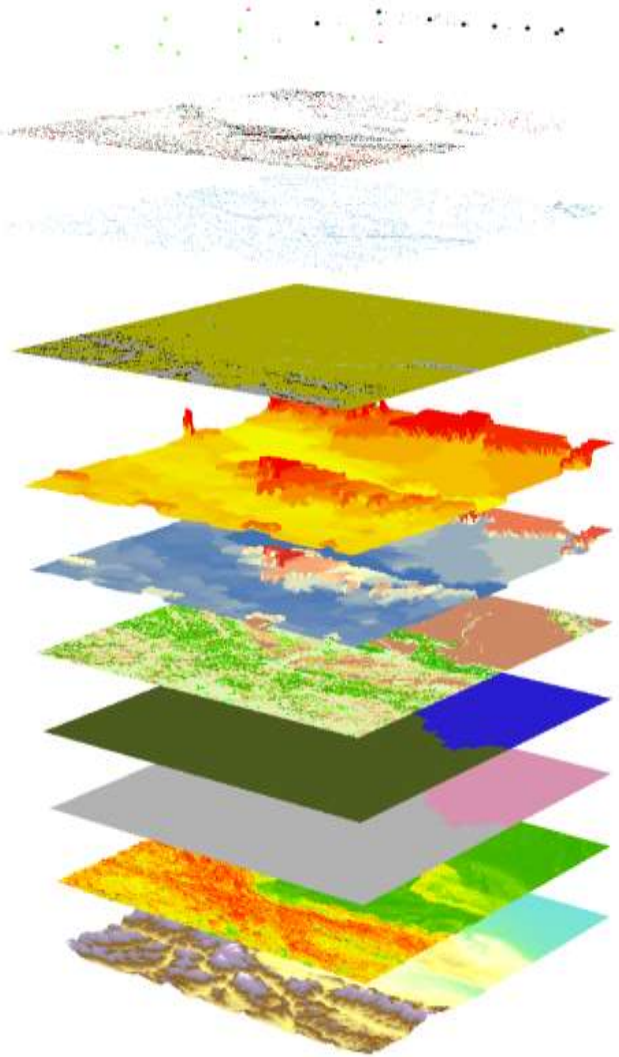


Slope

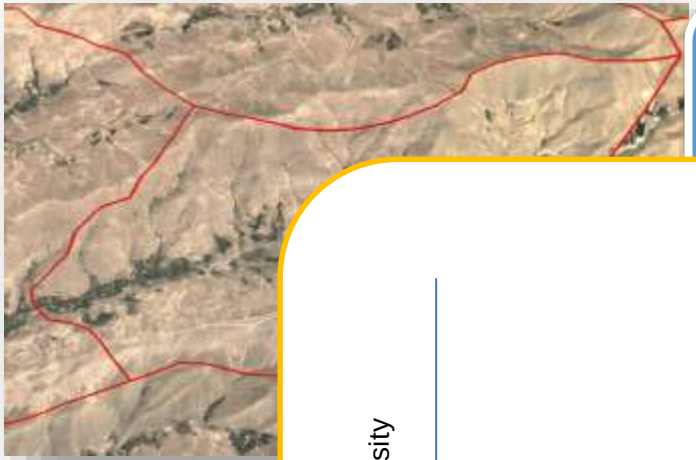






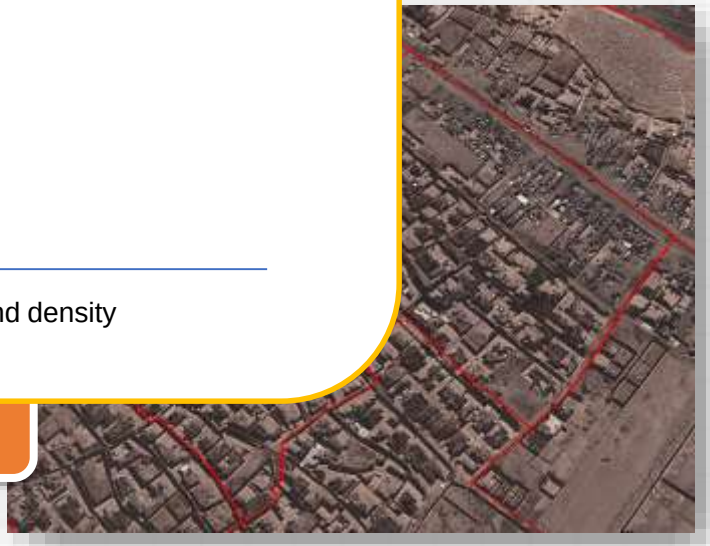


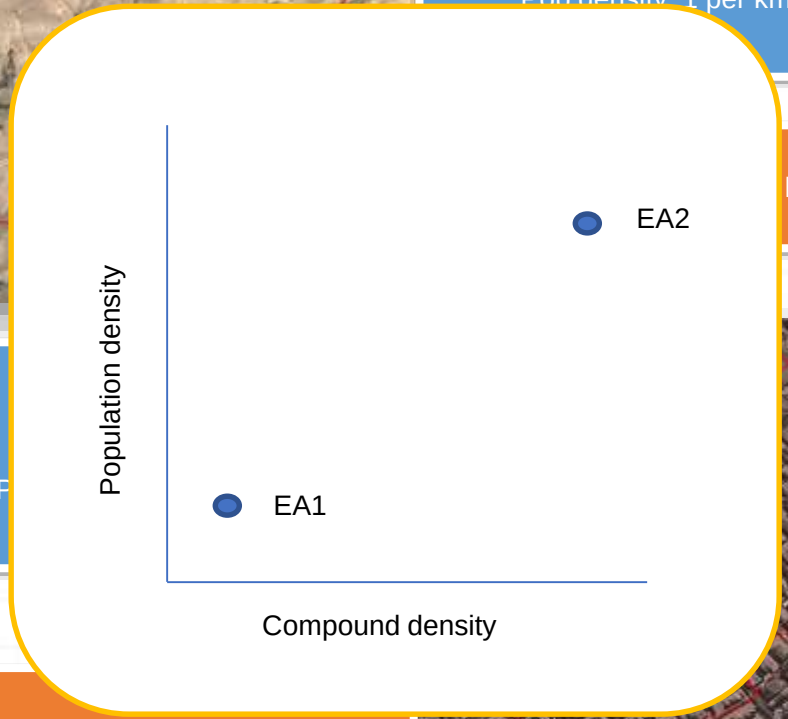
- >100 geospatial layers
- Sourced from across ministries, agencies and companies
- Complete national coverage
- Recent acquisition
- Significant effort in checking, cleaning, harmonizing
- Processed to extract indicators likely to be predictive of population distributions and densities



EA 1
Pop count: 1,000
Pop density: 1 per kmsq

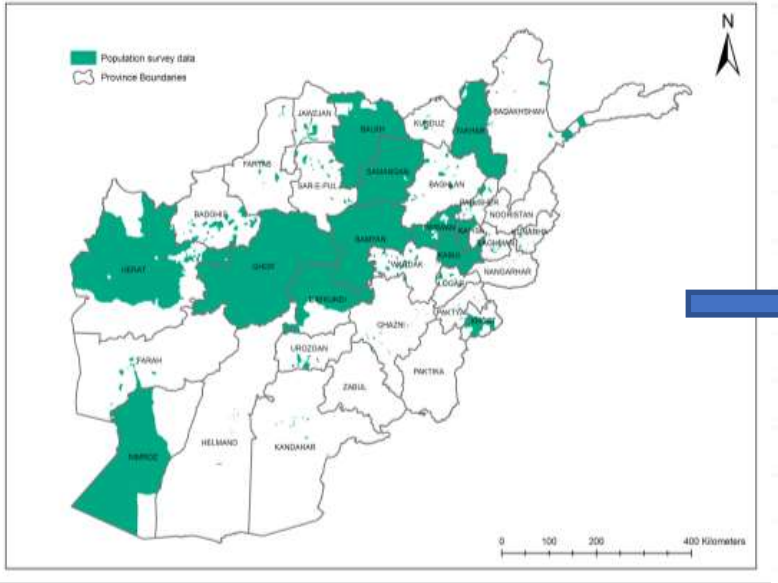
low

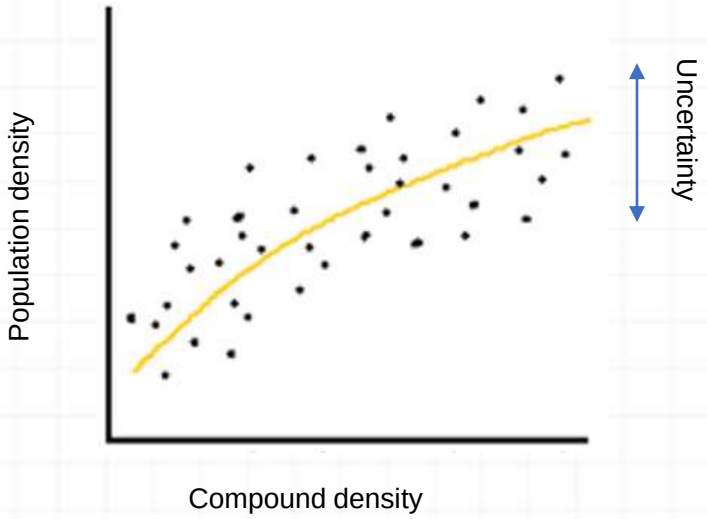




Compound density = high

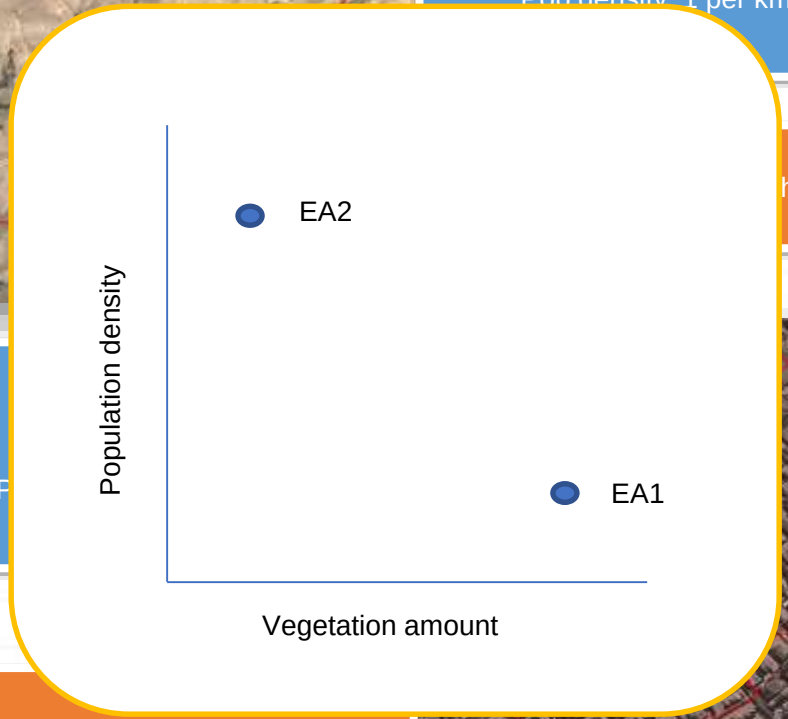
What about other EAs where we have population data?





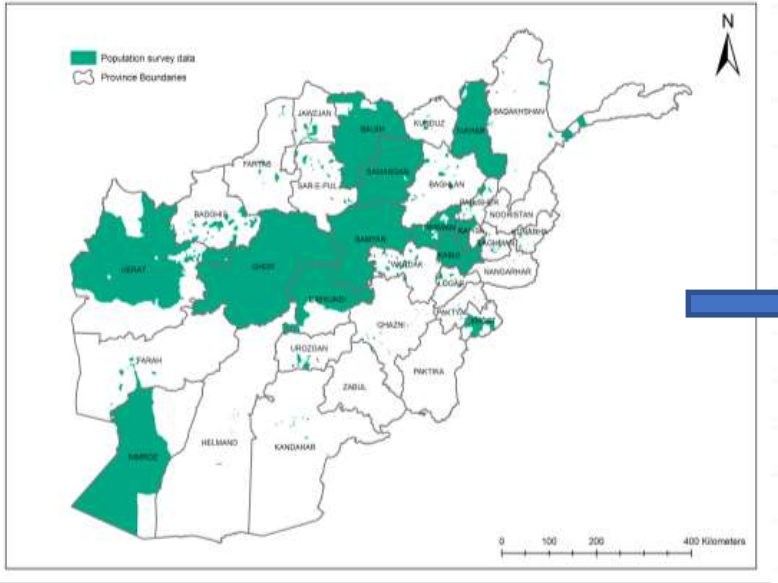
EA 1
Pop count: 1,000
Pop density: 1 per kmsq

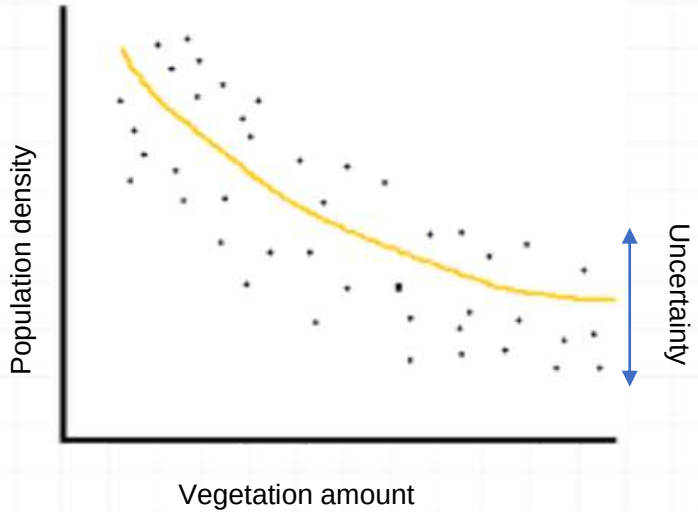
high



Vegetation amount =low

What about other EAs where we have population data?





Distance from urban area = high

Compound density = low

Mean slope gradient = high

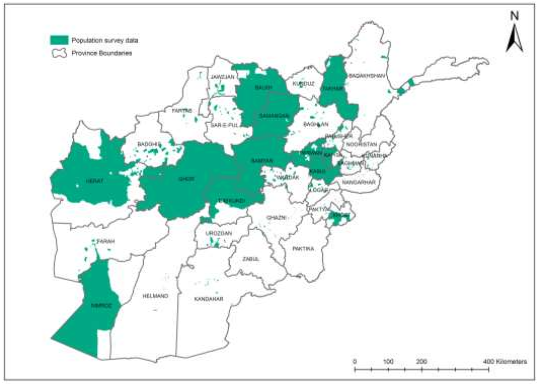
Nightlights brightness = low

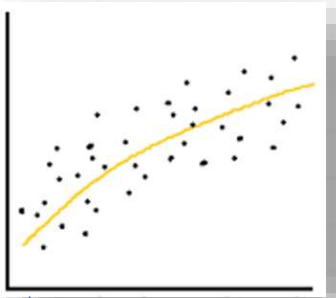
Distance from urban area = low

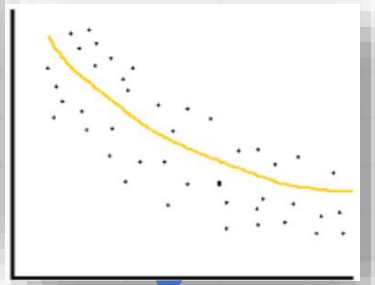
Compound density = high

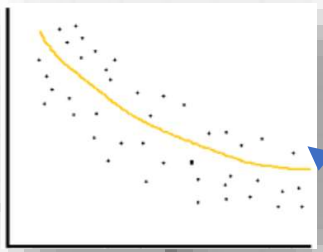
Mean slope gradient = low

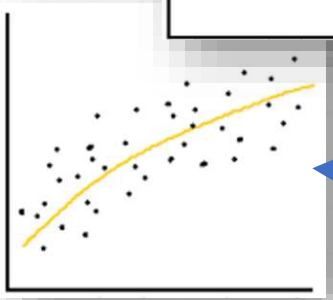
Nightlights brightness = high











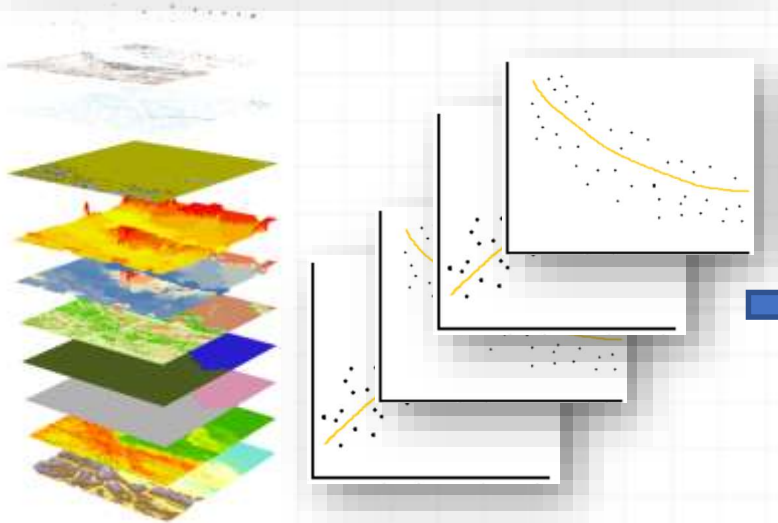
Distance from urban area

Compound density

Mean slope gradient

Nightlights brightness

EA 3
Pop count: ?
Pop density: ? per kmsq

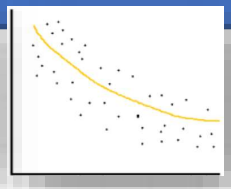


EA 3
Pop count: 350 (95% range = 300-400)
Pop density: 10 per kmsq (95% range = 7-13)

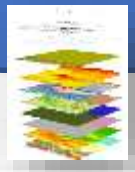
Population estimation steps

Assemble existing recent population data

Explore and test relationships between population data and geospatial datasets



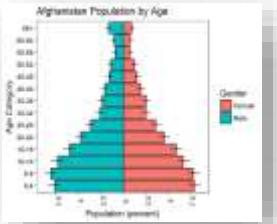
Select most predictive subset of geospatial datasets



Build model, make predictions of populations per EA in unsampled locations, cross-validation

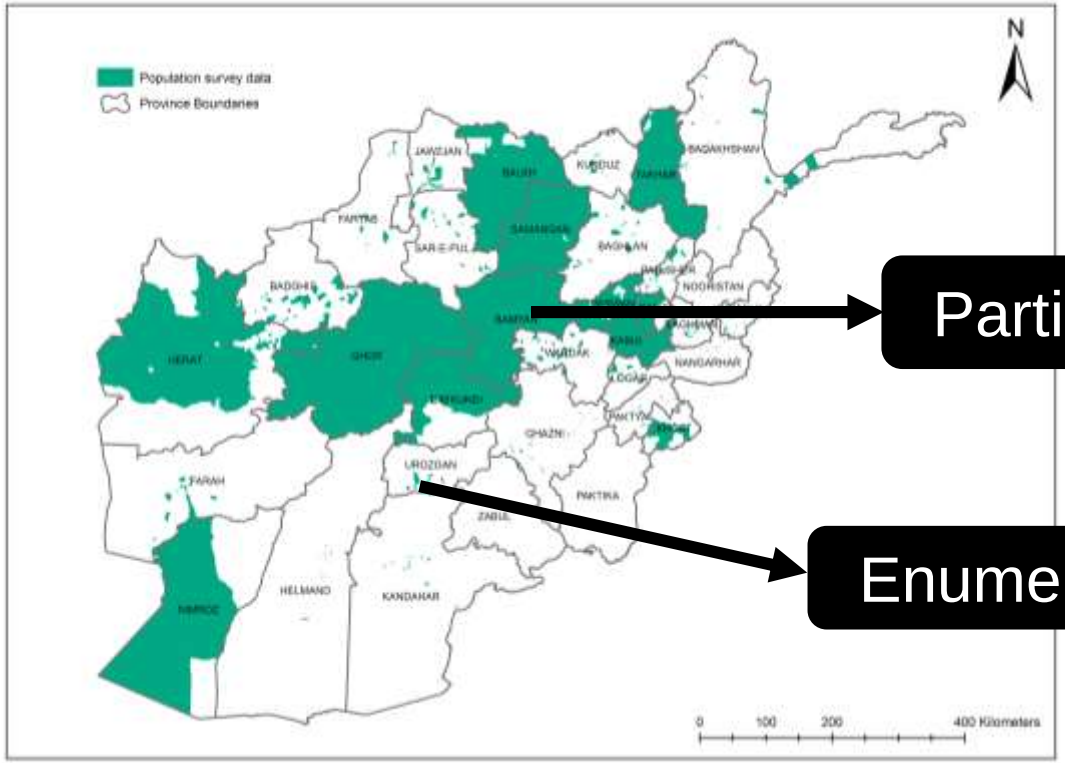
Disaggregate to 100x100m grid cells

Repeat for age/sex group proportions



Implementation for Afghanistan

Existing recent data



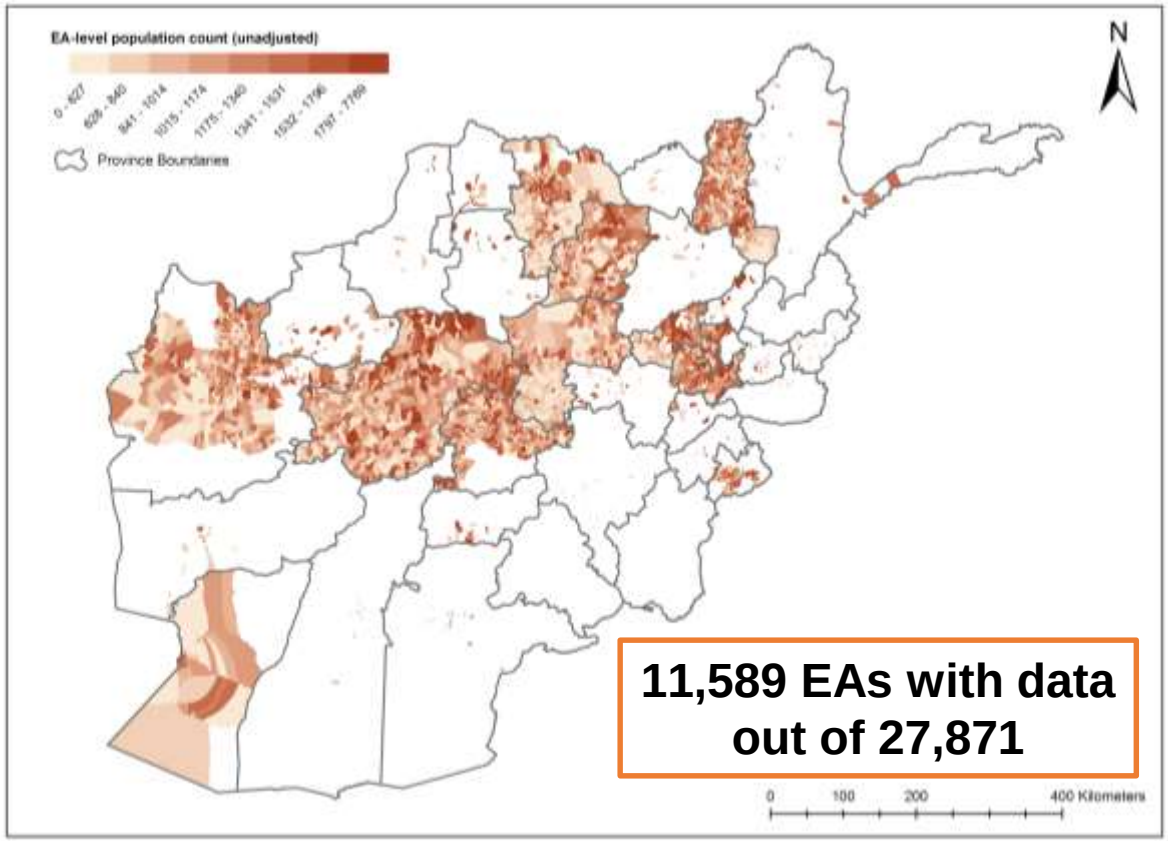
Partial census

+

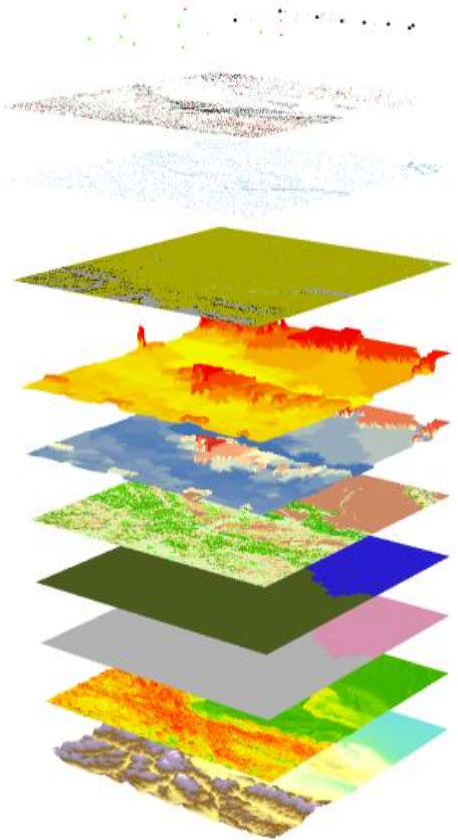
Enumeration surveys

11,589 EAs with data out of 27,871

Existing recent data



Finding the strongest predictors



Selected geospatial variables
Number of residential compounds
Compound density
Distance from edge of urban areas
Mean slope
Mean average distance between compounds
Minimum distance from a village
Maximum nightlight brightness
Fragmentation of settled area
Area of the EA
Vegetation amount

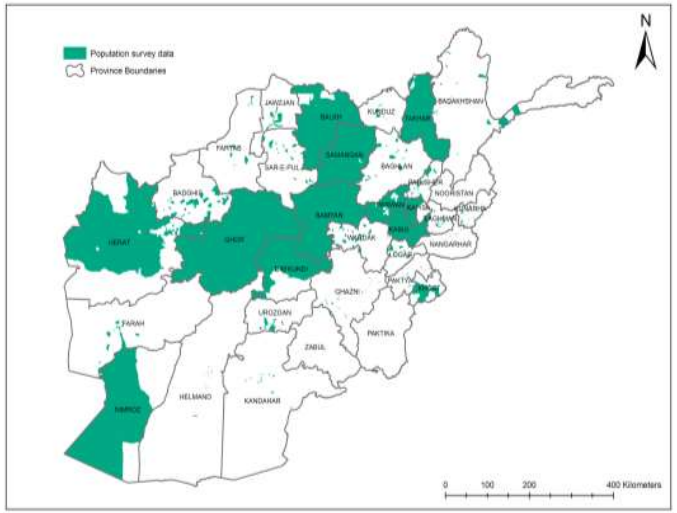
Training and testing

- Multiple model types and forms tested
- Selected the type and set of covariates that optimised predictive ability

Bayesian hierarchical spatial model:

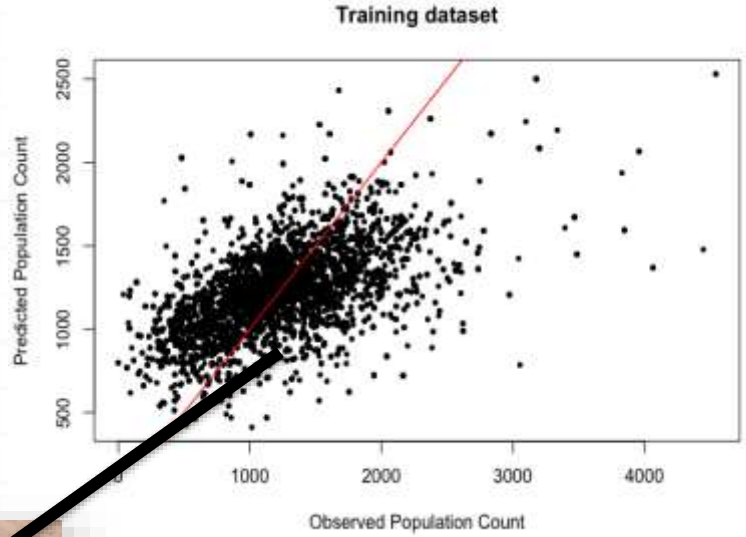
-multiple model forms
-accounts for spatial relationships
-measures of uncertainty

- **10-fold cross-validation**
- 11,589 EAs with data
- Split randomly into 10 groups
- Further subdivided into training and testing sets
- Mean observed vs prediction stats



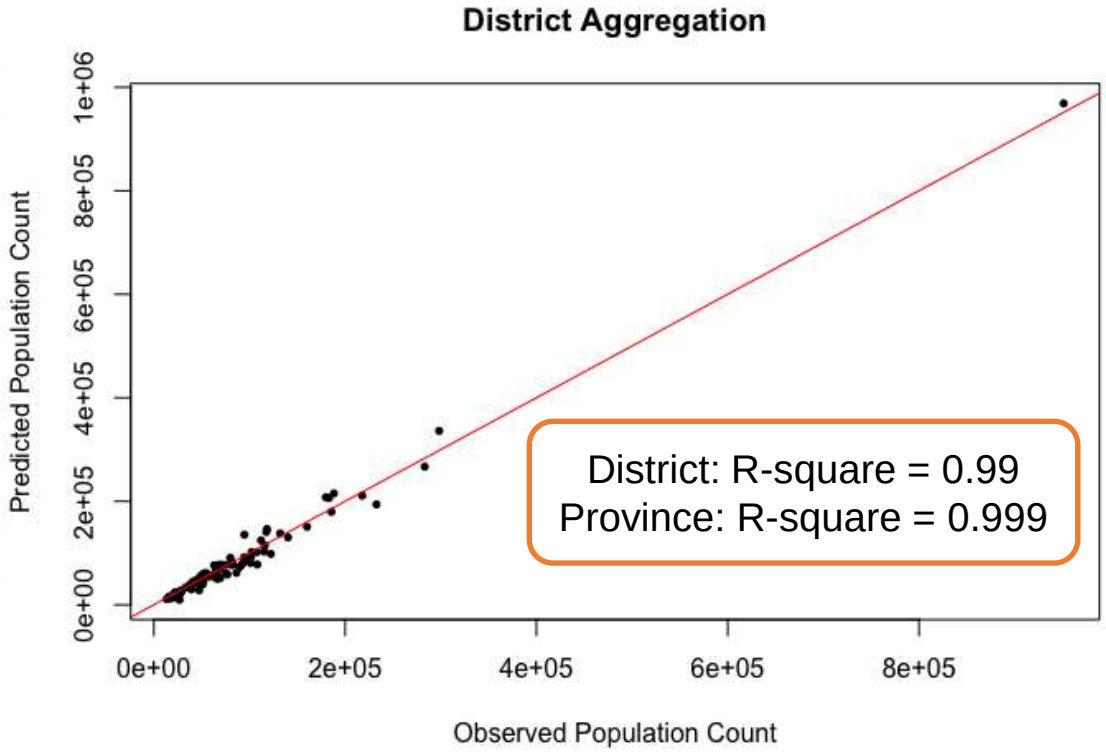
Cross-validation: EA level

Correlation: 0.61

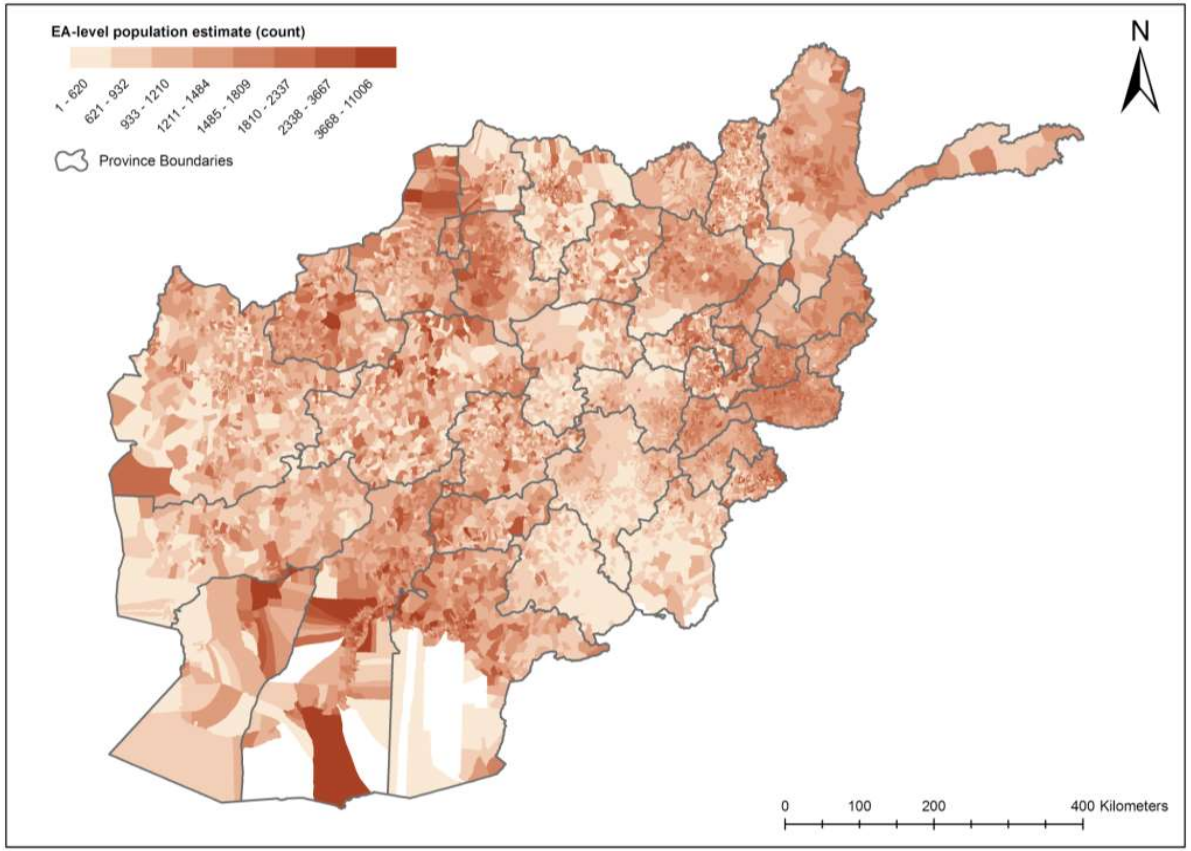


Mean population weighted error 22%

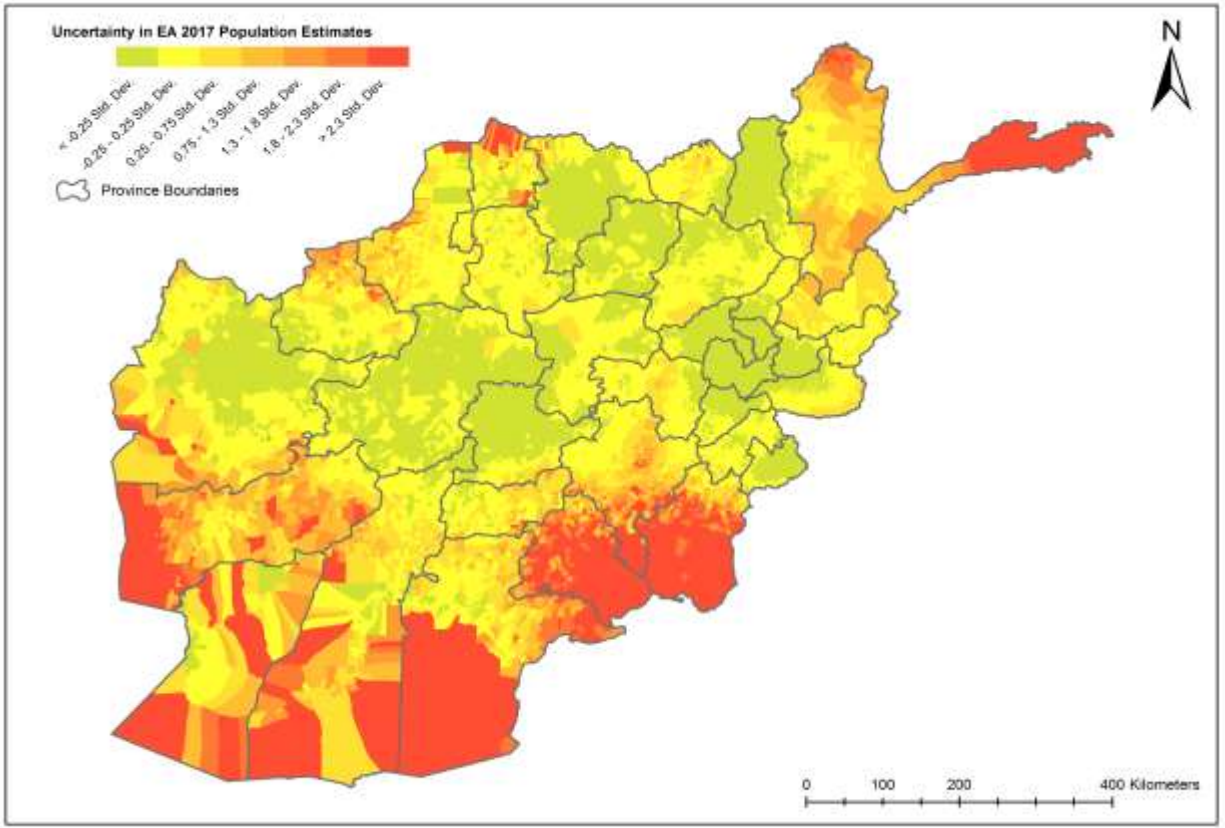
Accuracy assessment: District and Province level



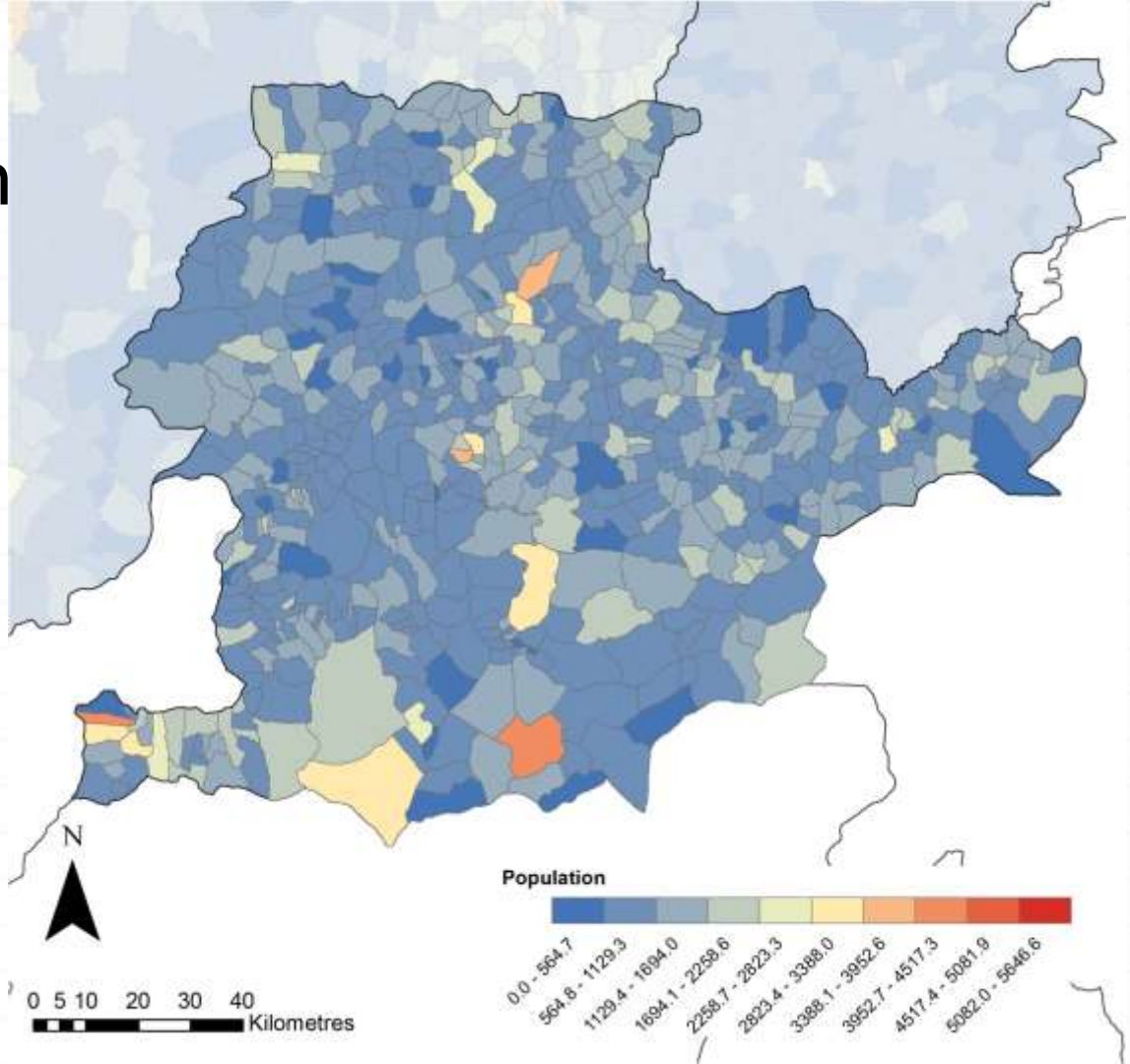
Outputs: EA counts and estimates



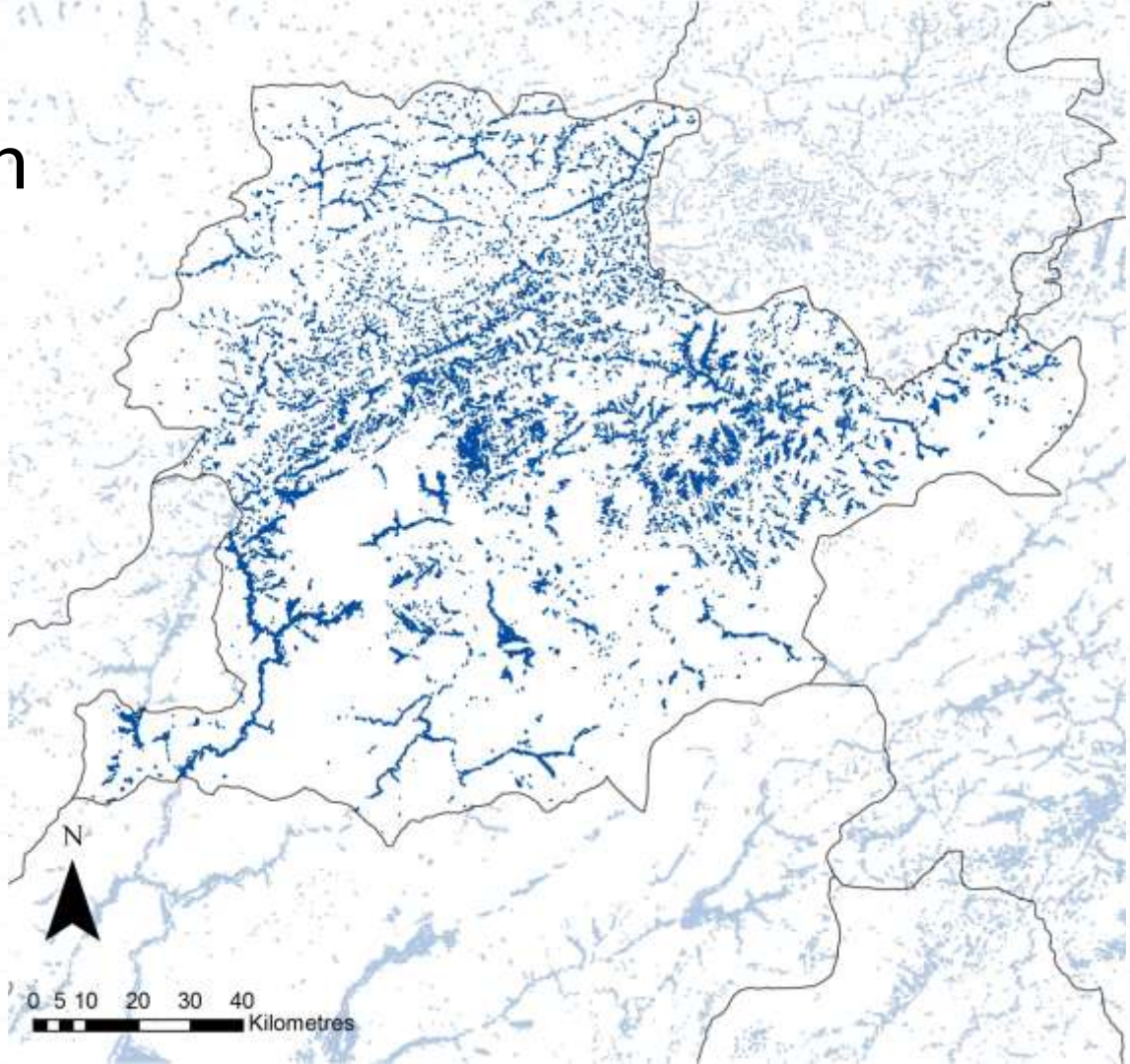
Outputs: EA level uncertainties



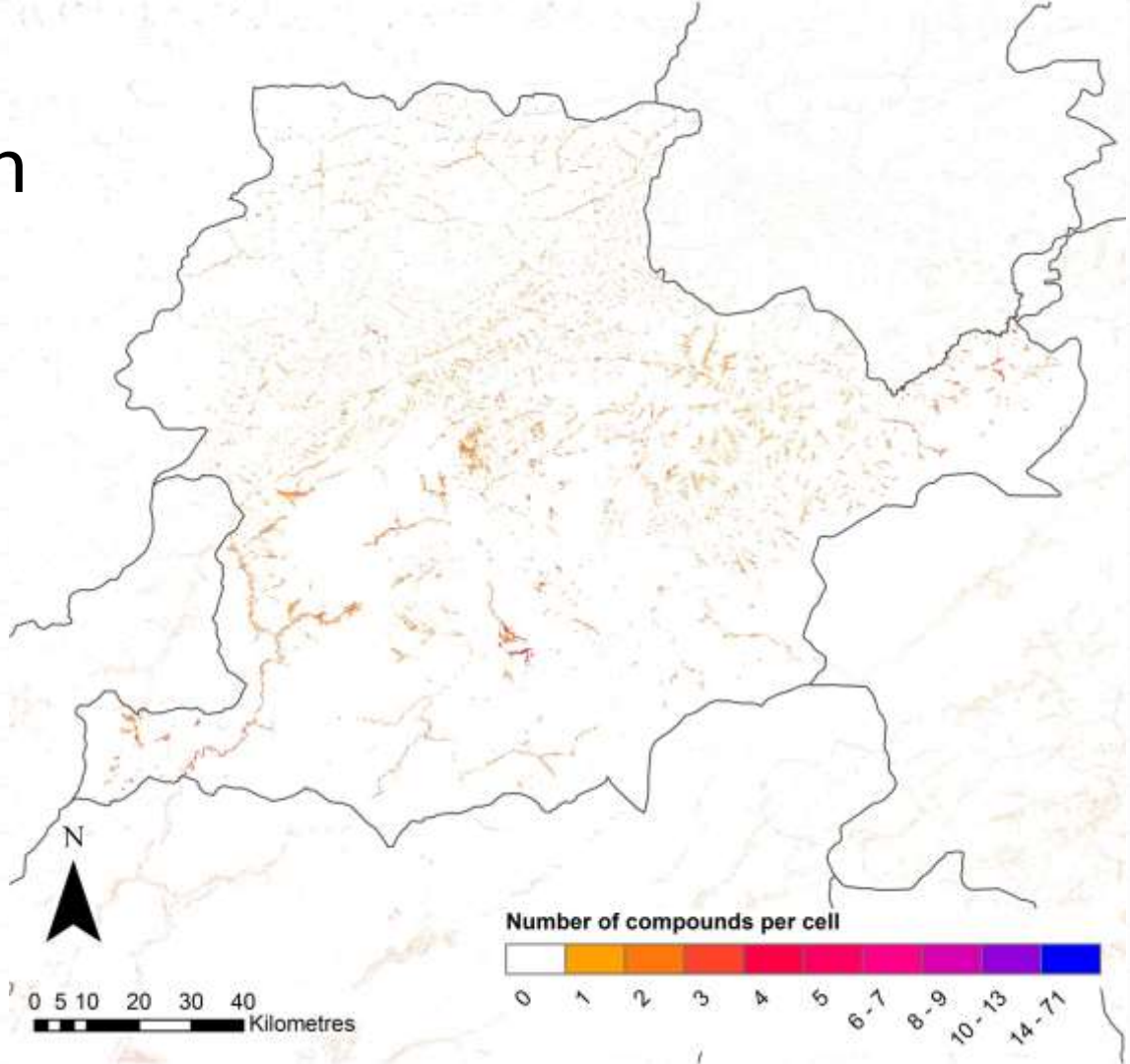
Spatial disaggregation to 100x100m cells



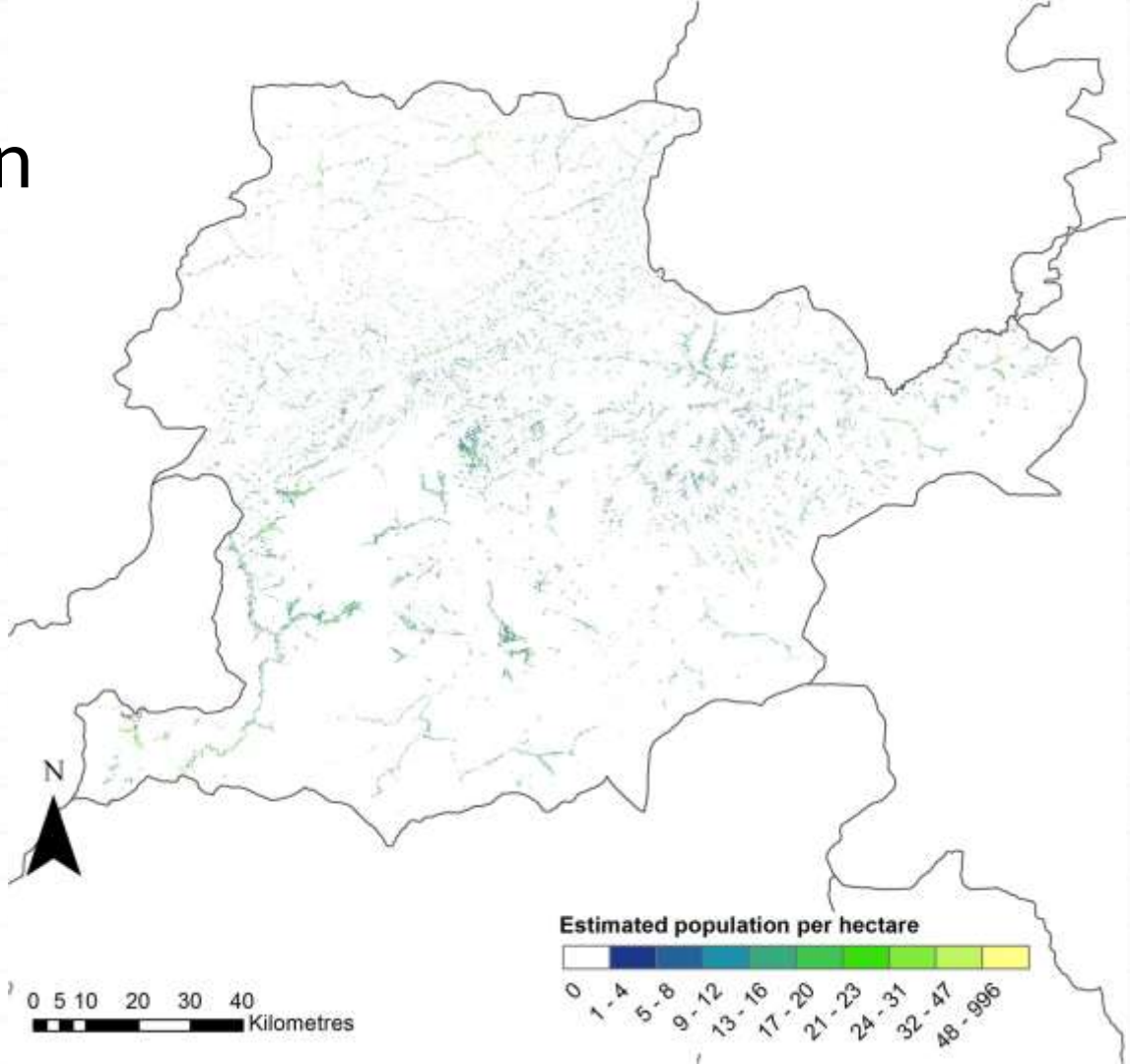
Spatial disaggregation to 100x100m cells

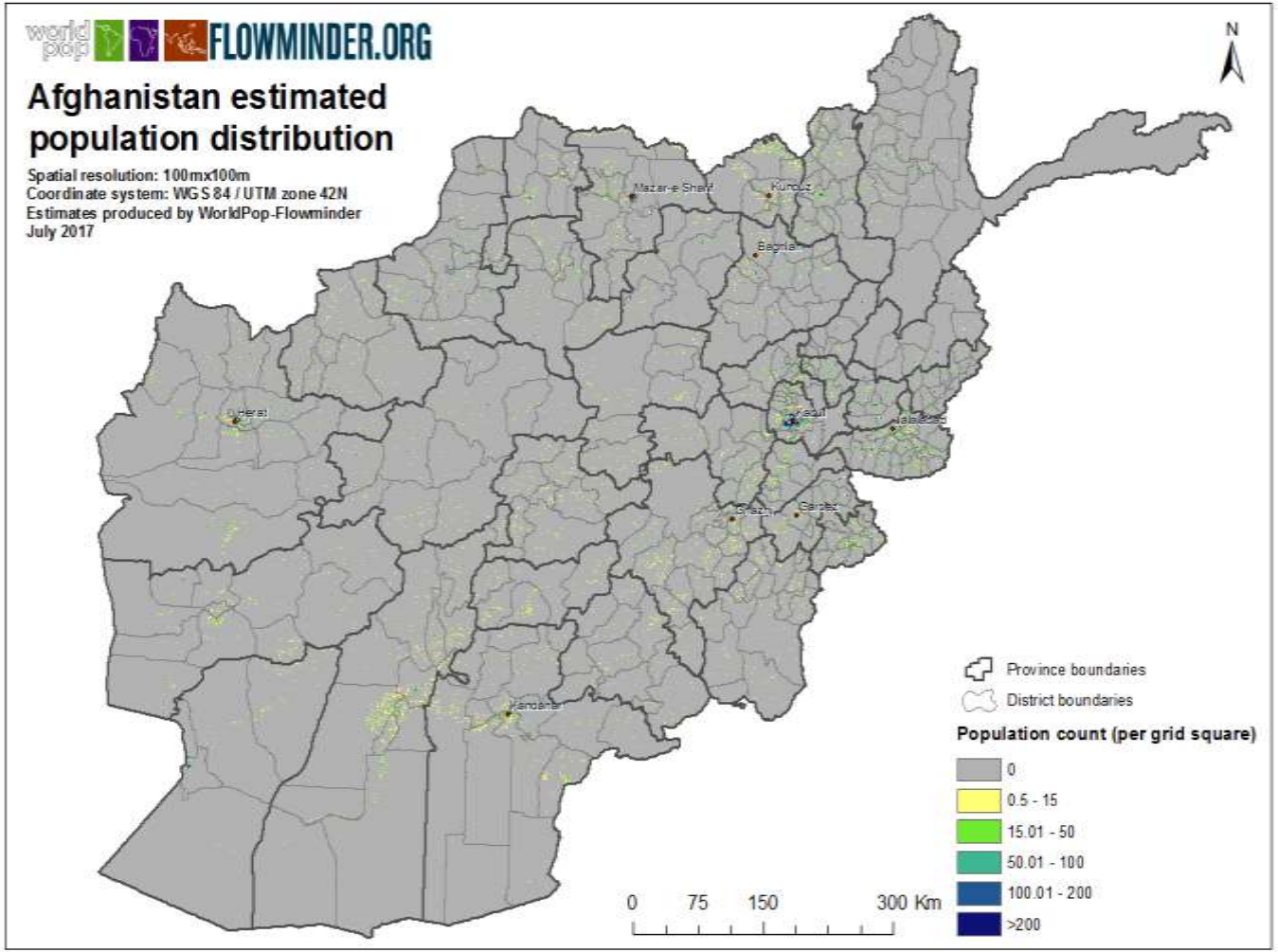


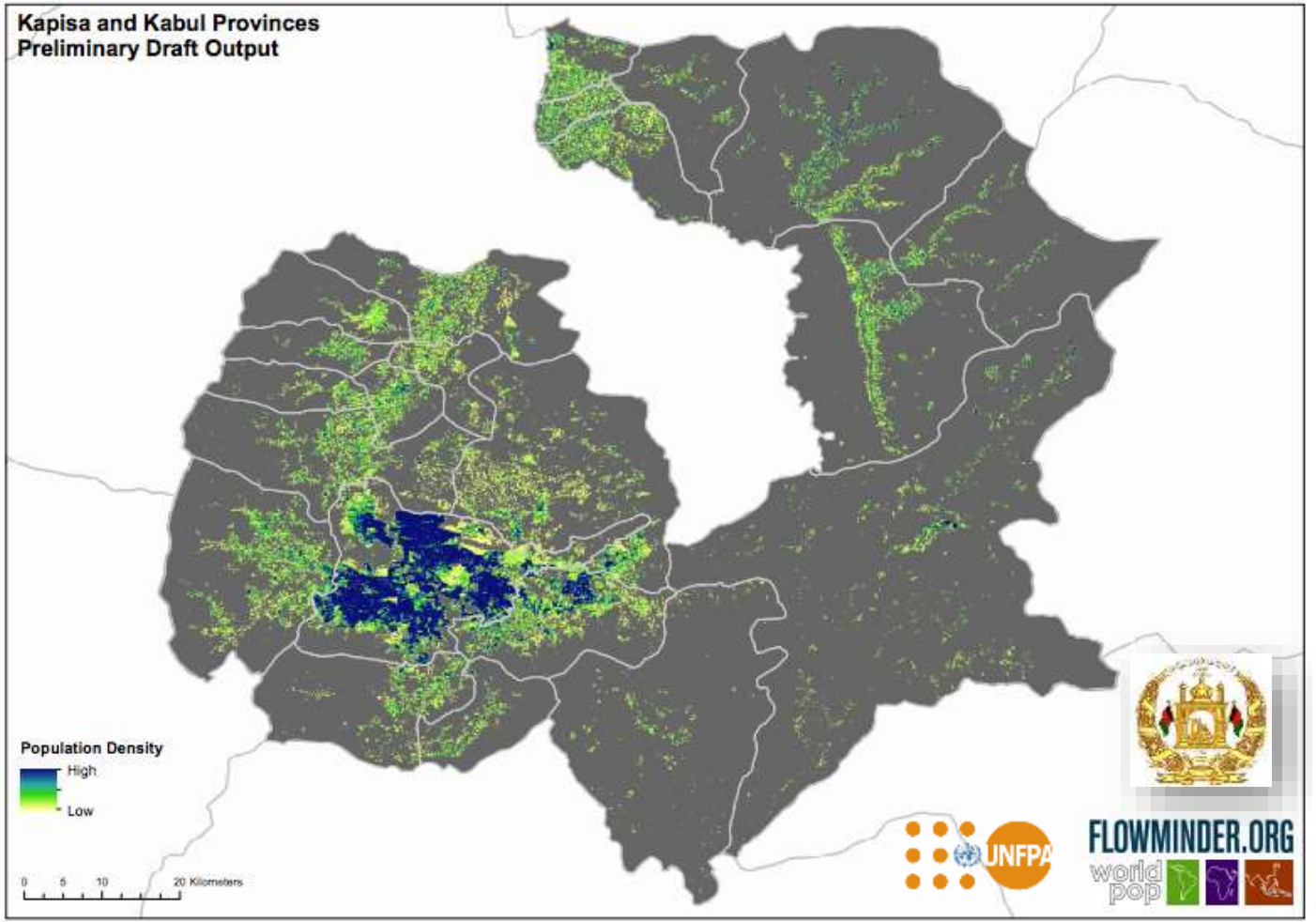
Spatial disaggregation to 100x100m cells



Spatial disaggregation to 100x100m cells

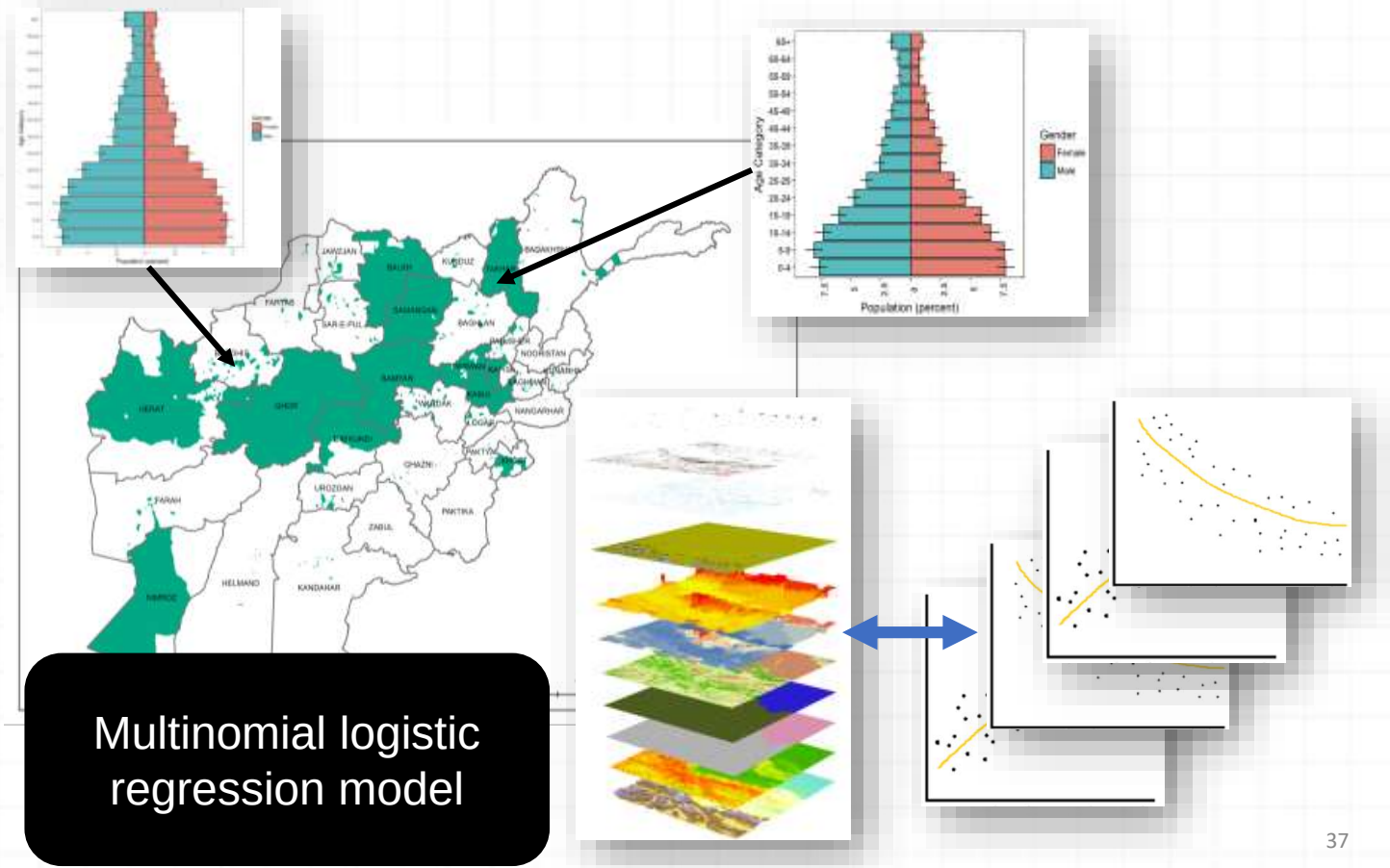



FLOWMINDER.ORG
Afghanistan estimated population distribution
Spatial resolution: 100mx100m
Coordinate system: WGS 84 / UTM zone 42N
Estimates produced by WorldPop-Flowminder
July 2017
N
Province boundaries
District boundaries
Population count (per grid square)
0
0.5 - 15
15.01 - 50
50.01 - 100
100.01 - 200
>200
0
75
150
300 Km

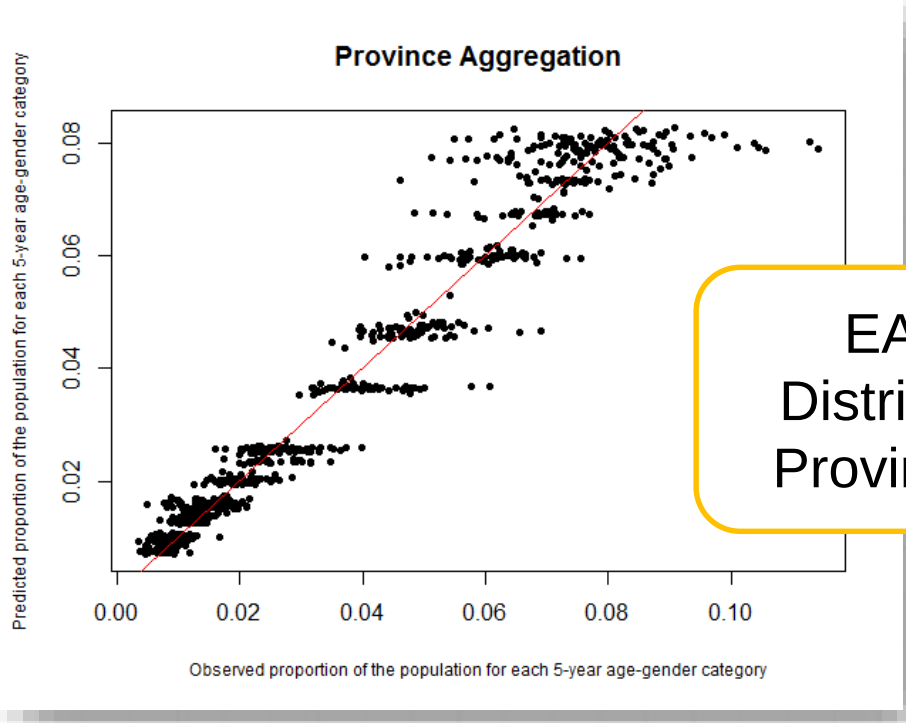

Kapisa and Kabul Provinces
Preliminary Draft Output
Population Density
High
Low
0 5 10 20 Kilometers
UNFPA
FLOWMINDER.ORG
world pop

Modelling population pyramids

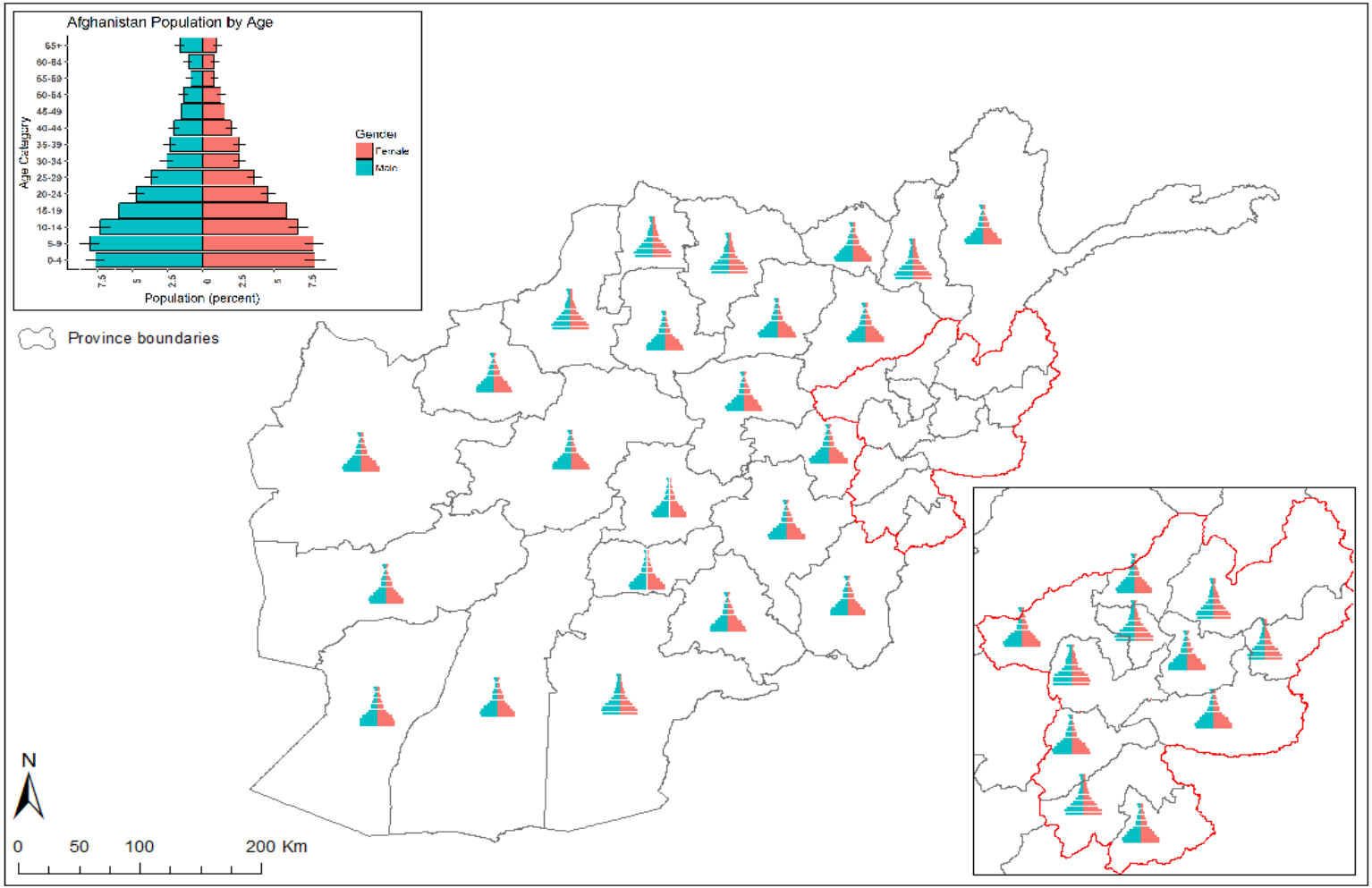
Estimating age/sex structures



Age/sex structure cross-validation

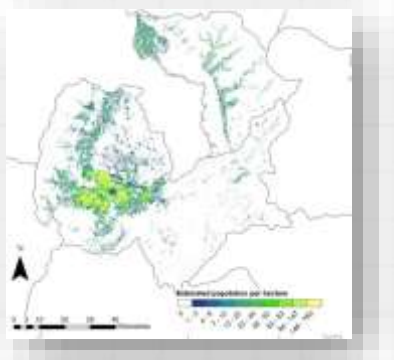


EA: R-square = 0.80
District: R-=square = 0.89
Province: R-square = 0.97


Afghanistan Population by Age
65+
60-64
55-59
50-54
45-49
40-44
35-39
30-34
25-29
20-24
15-19
10-14
5-9
0-4
Age Category
7.5
5
2.5
0
2.5
5
7.5
Population (percent)
Gender
Female
Male
Province boundaries
N
0
50
100
200 Km

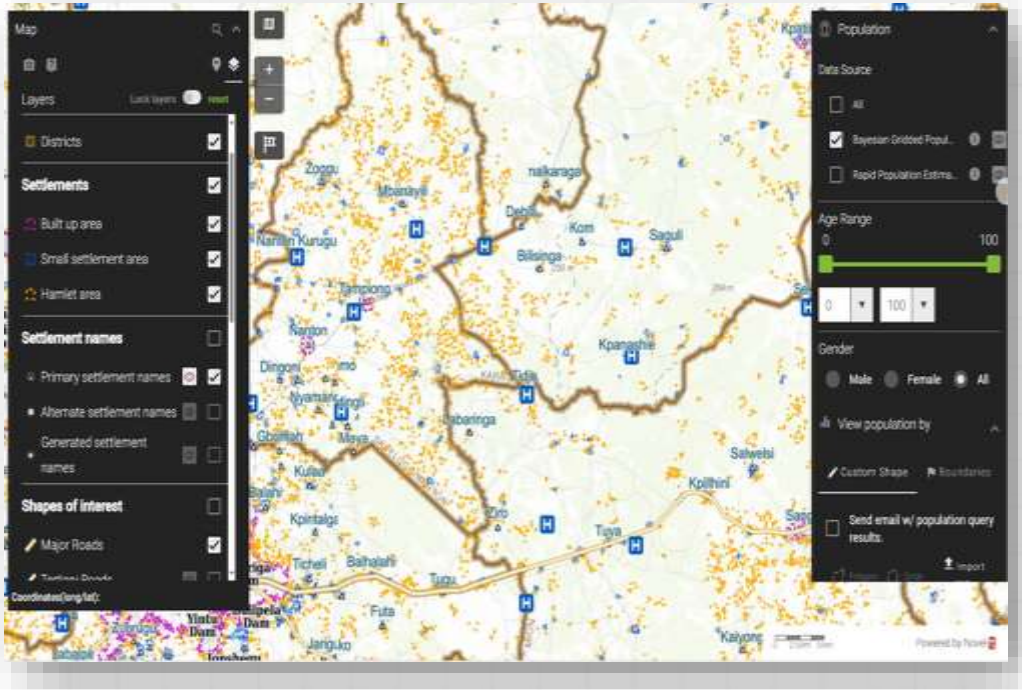
Key considerations

- Uncertainties are substantially higher in unsampled locations
- Uncertainties are higher for predictions made at fine spatial scales
- Mobile populations are not explicitly accounted for
- Estimates can be improved further with new survey and geospatial data

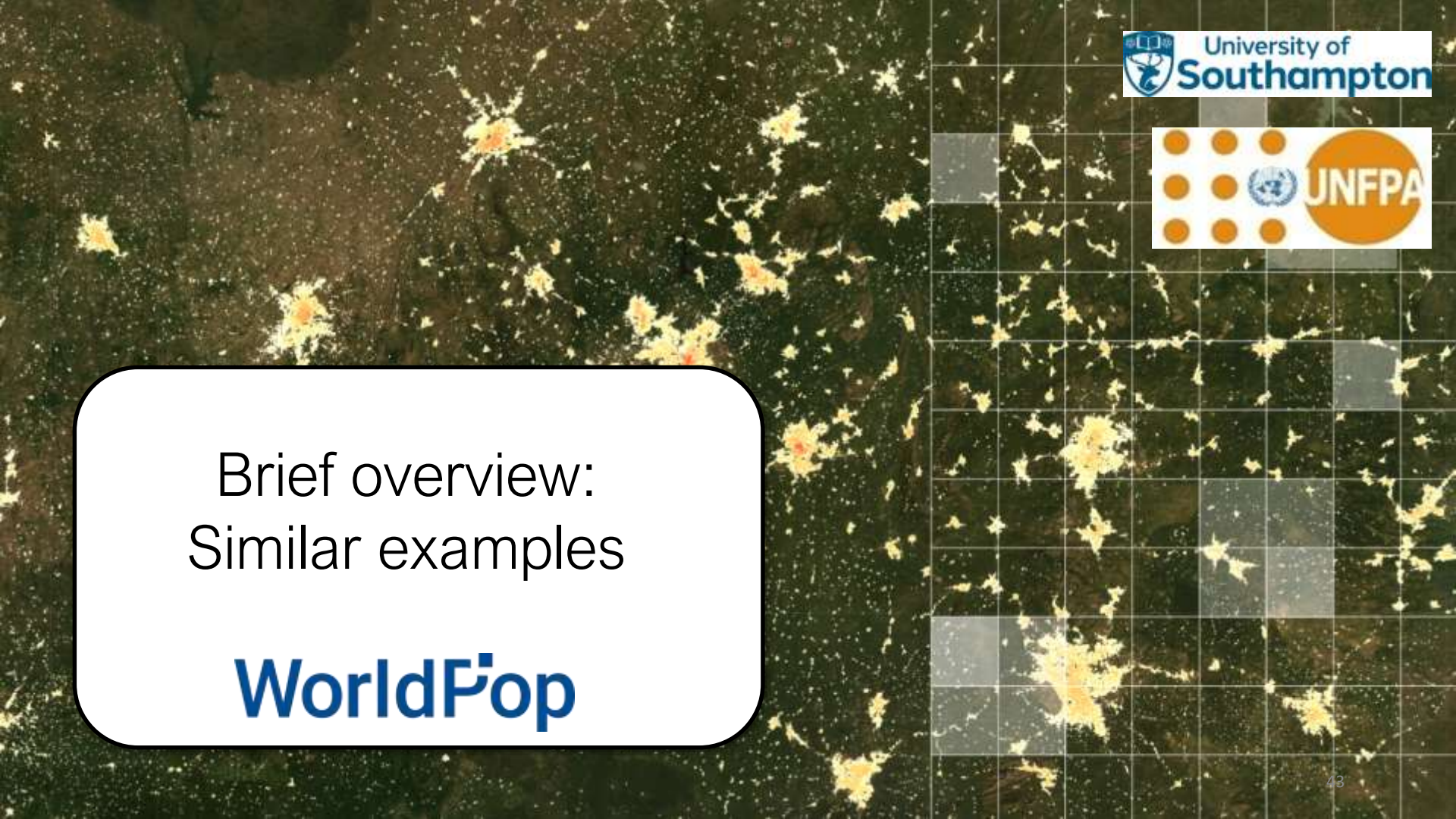


Usage since 2017

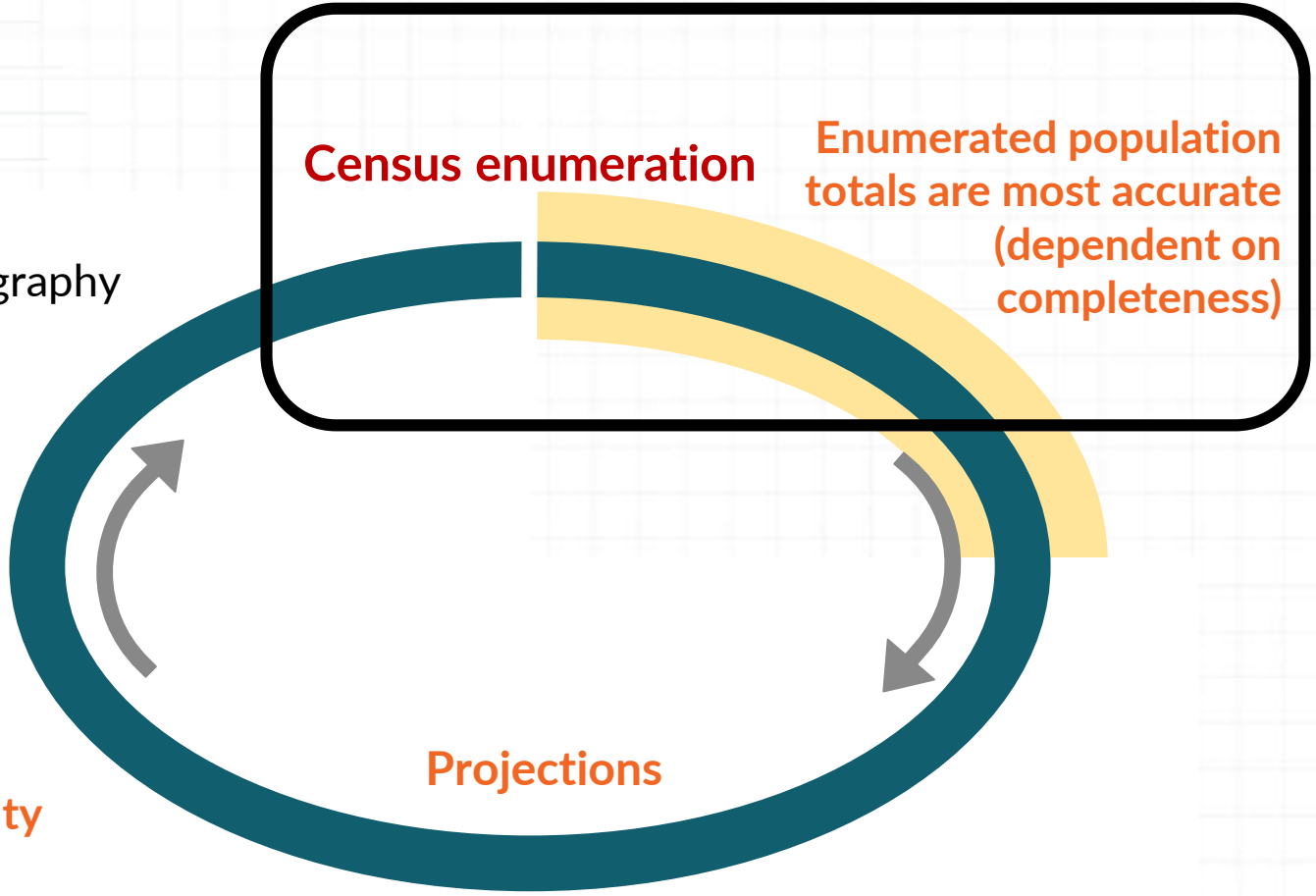
- Child vaccination microplanning
- Sample frame for designing new household surveys
- Health financing
- Modelling updates with WHO planned for 2024/25



GeoPode polio vaccination portal


University of Southampton
UNFPA
Brief overview:
Similar examples
WorldPop

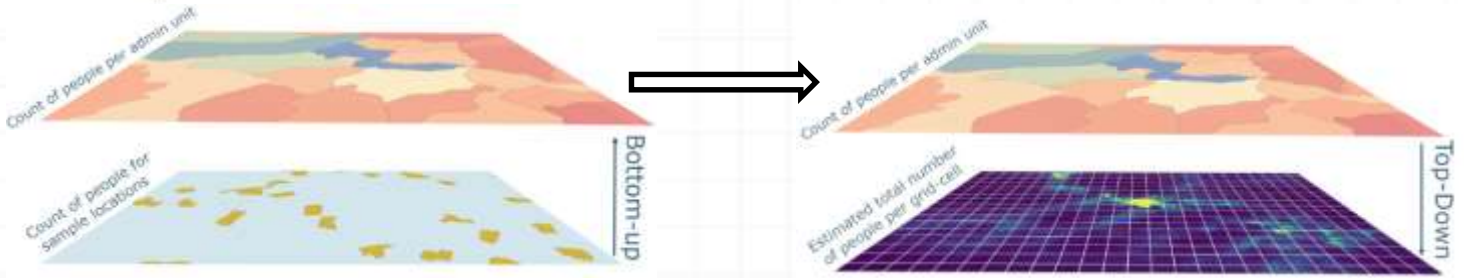
Stages of the census cycle



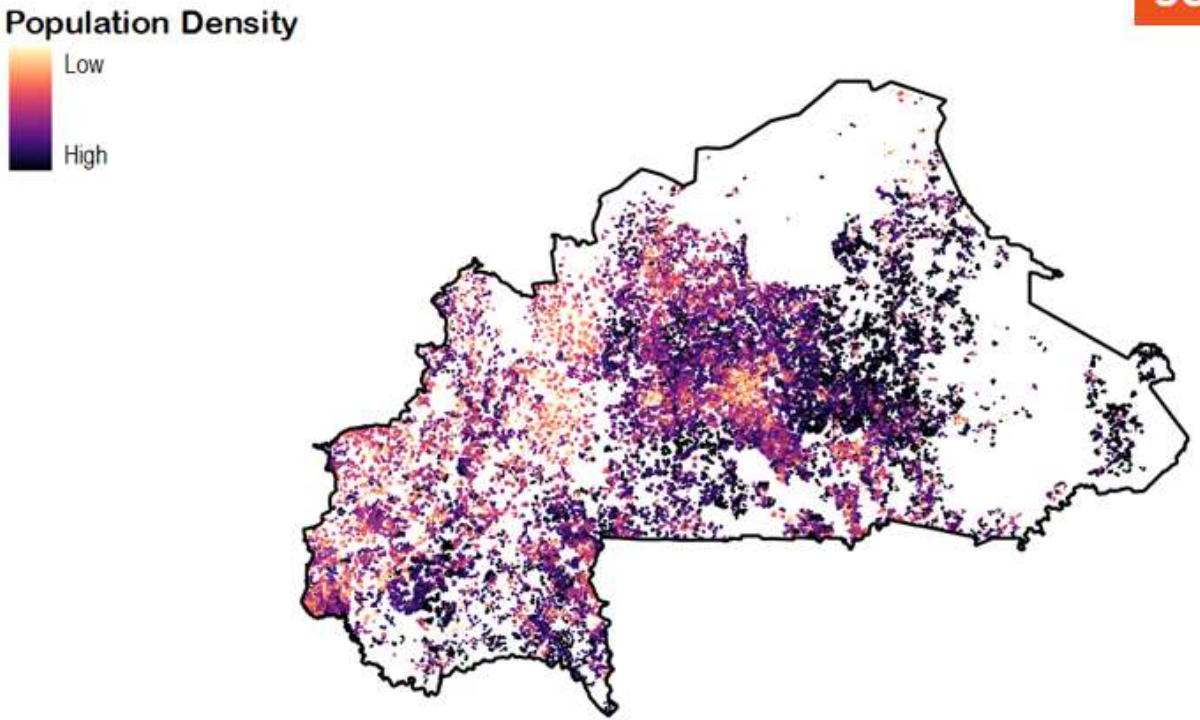
As time from the census elapses, increasing uncertainty in projections

Census under-enumeration

- Unable to enumerate in hard-to-access regions
- Bottom-up estimation for under-enumerated units, followed by top-down disaggregation to grid squares



5e RGPH Burkina Faso



Burkina Faso: 2019 census

Census 2019 database:
Non-enumerated EAs

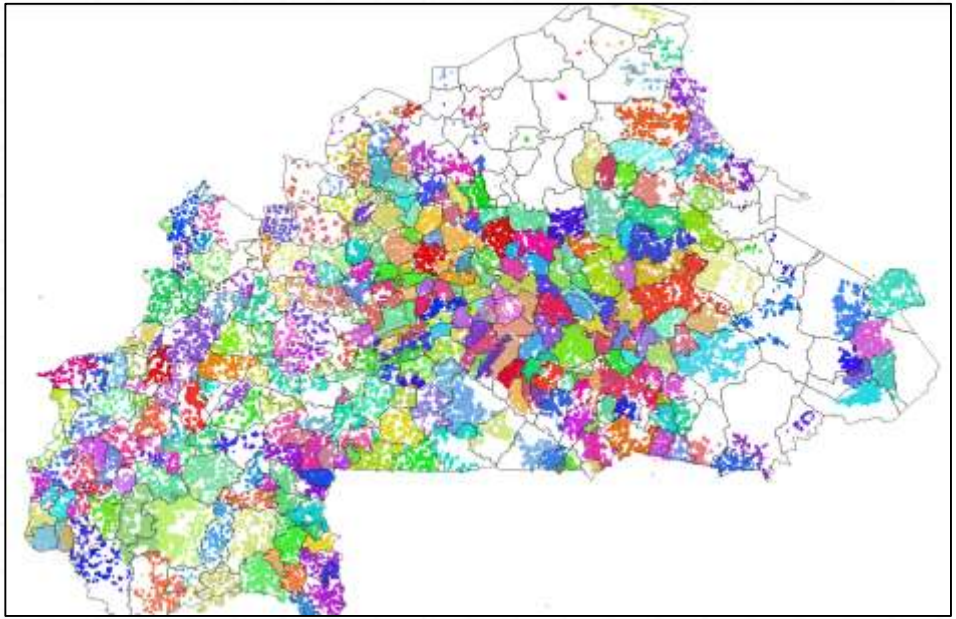
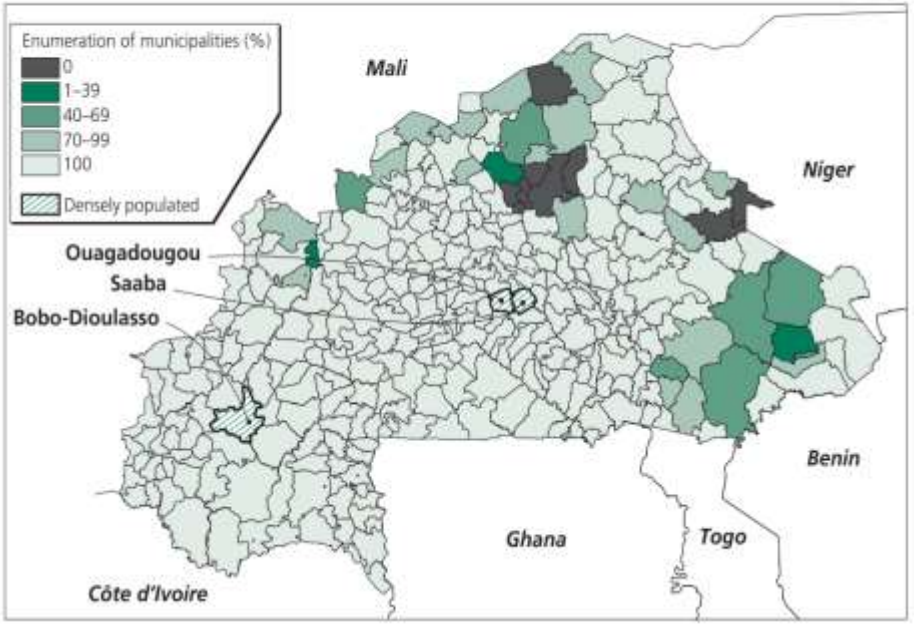
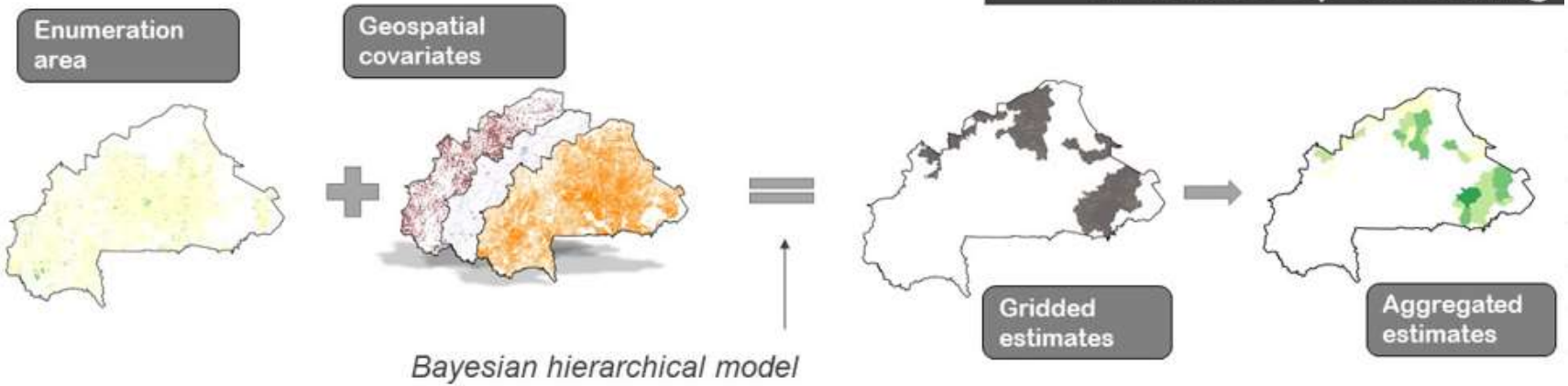


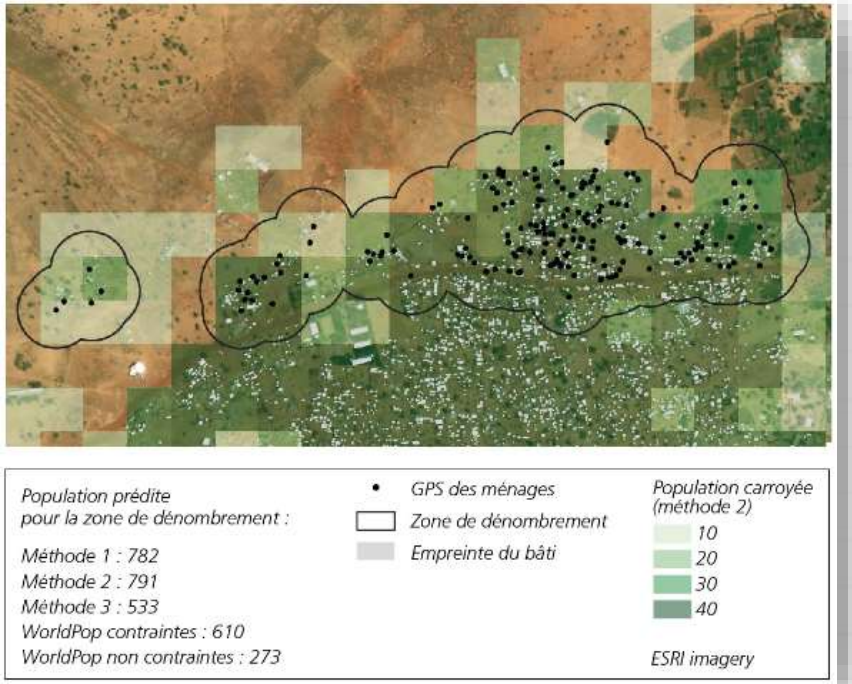
Figure 1. Coverage of the 351 municipalities (communes) in the 2019 census of Burkina Faso

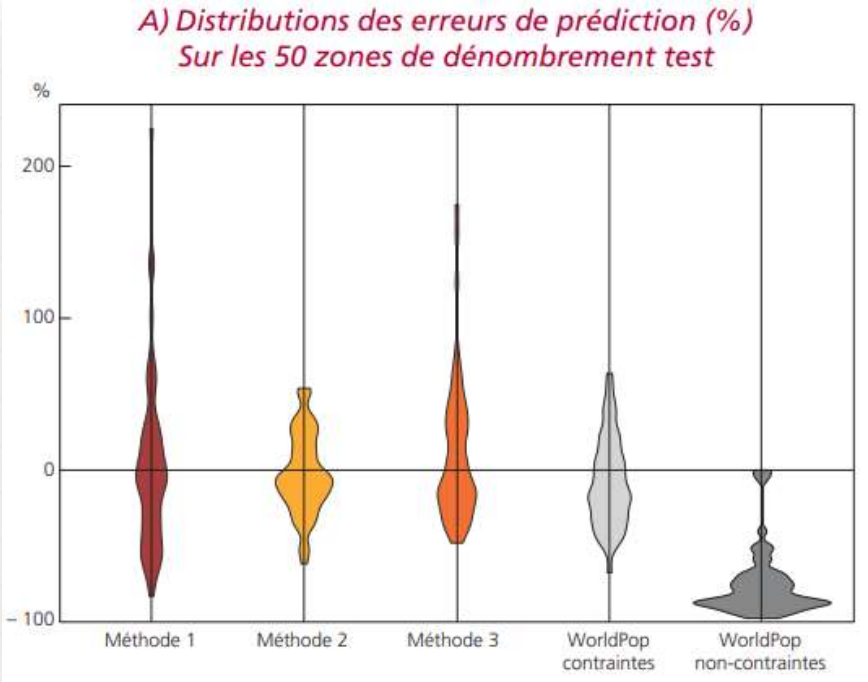


1. Bottom-up modelling



Post-census validation





https://wopr.worldpop.org/?/Population ;

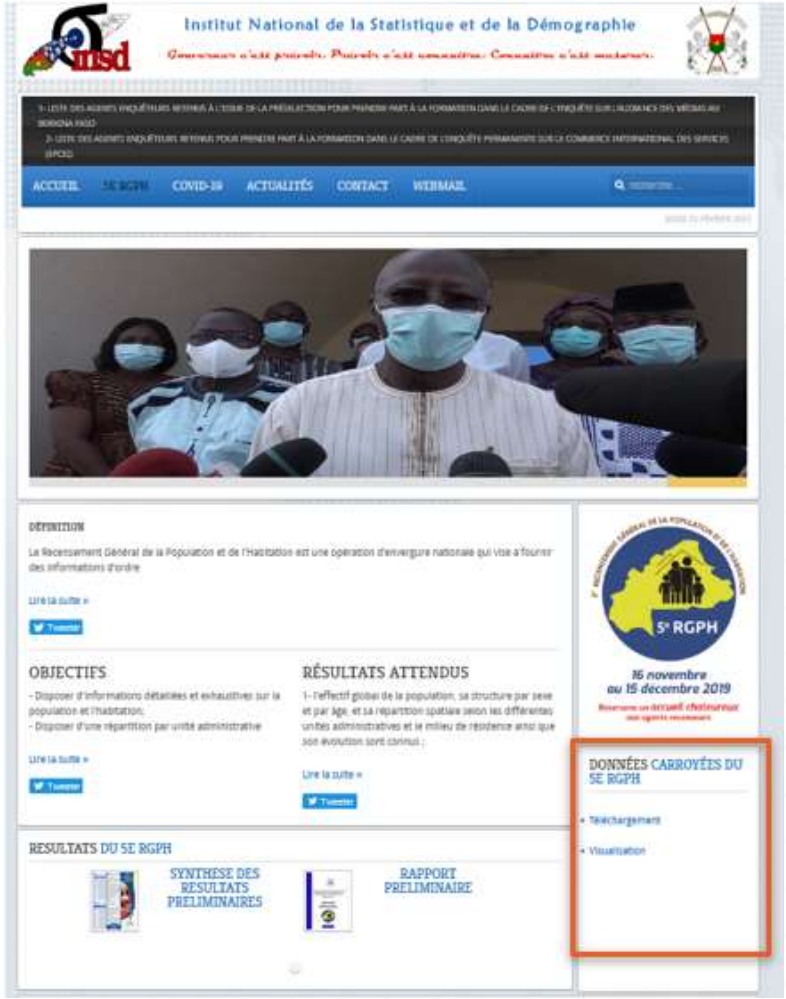
Darin et al (2022) Population

Cinquième Recensement Général de la Population et de l'Habitation du Burkina Faso

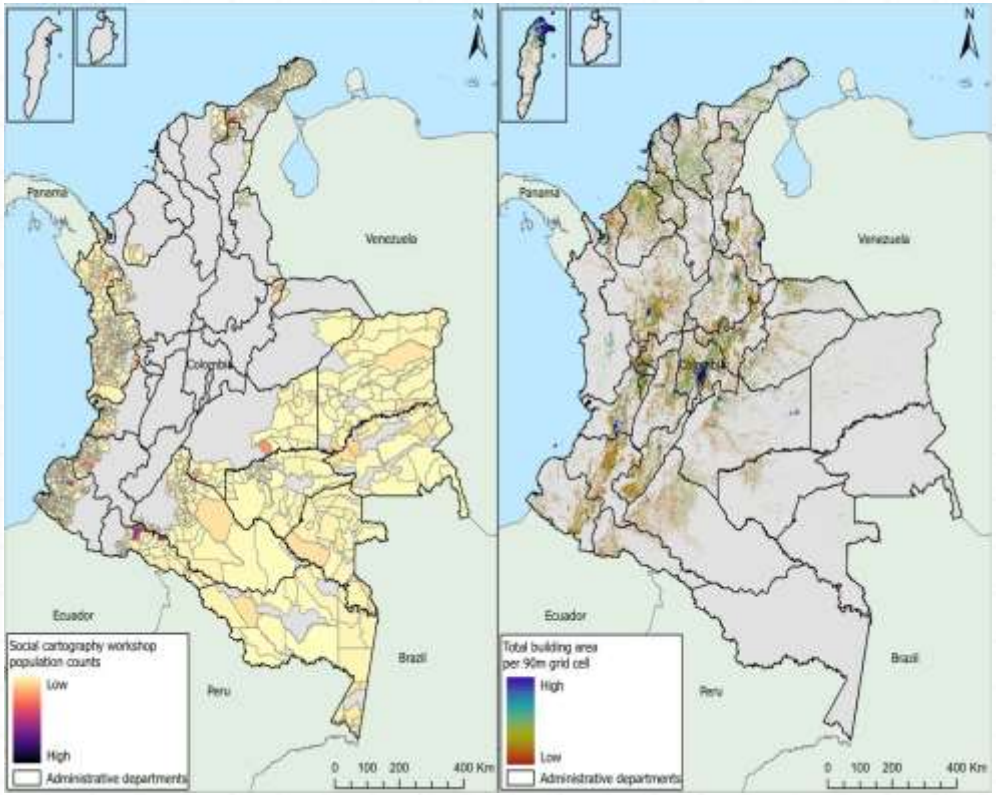
Résultats Préliminaires

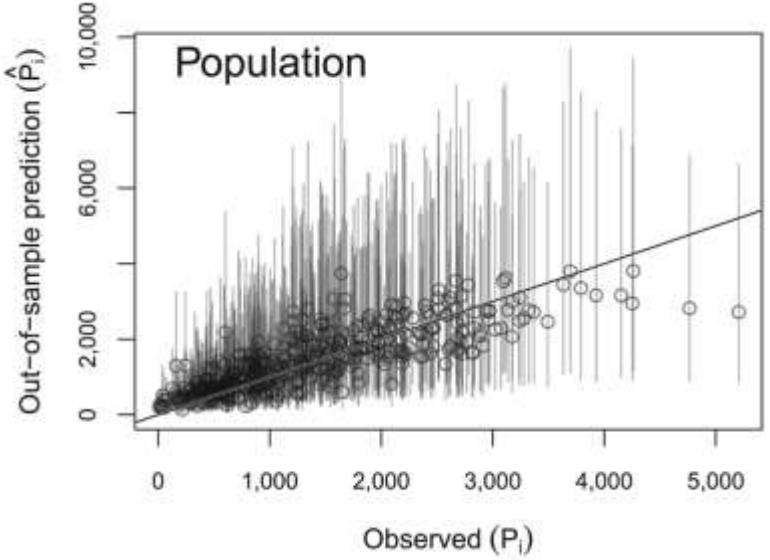
5e RGPH

Septembre 2020



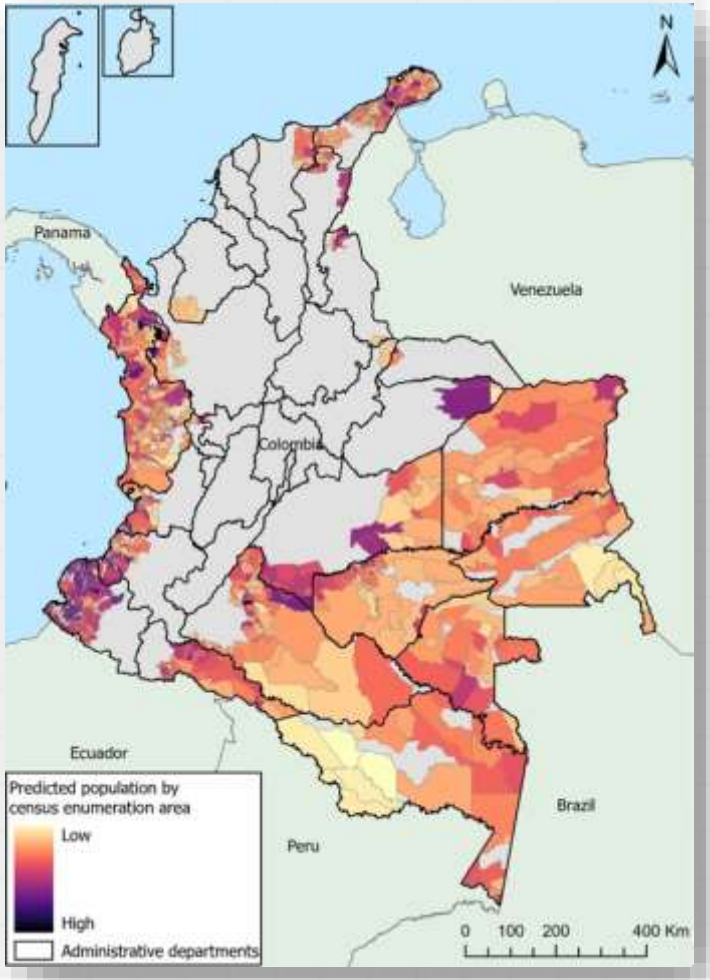
Filling census gaps using satellite and “social cartography”: Colombia

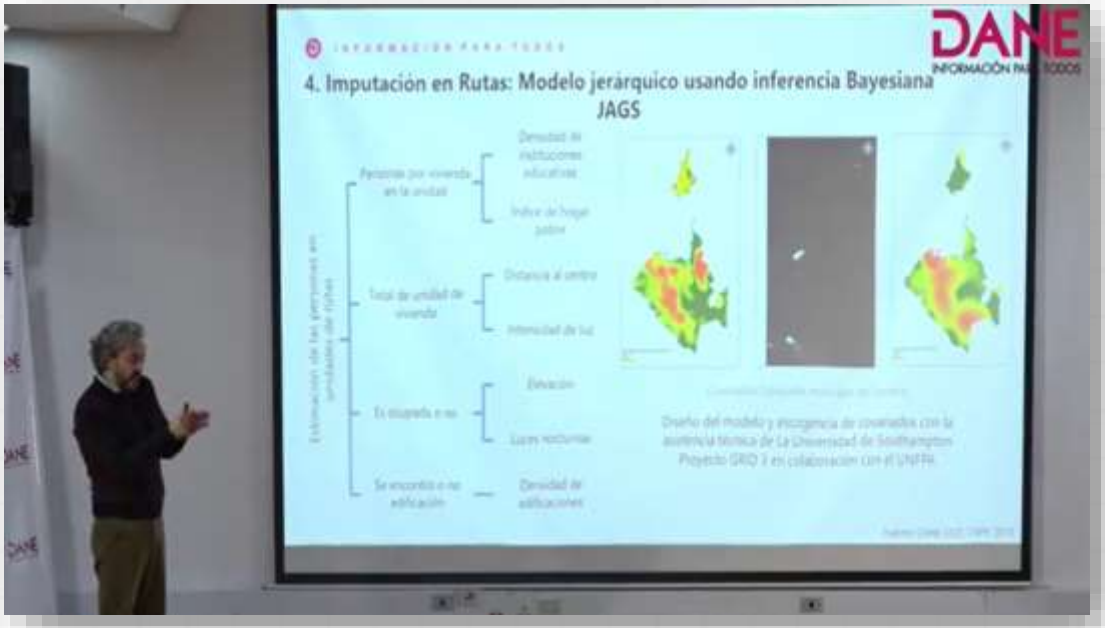




Sanchez-Cespedes et al (2023) Population Studies

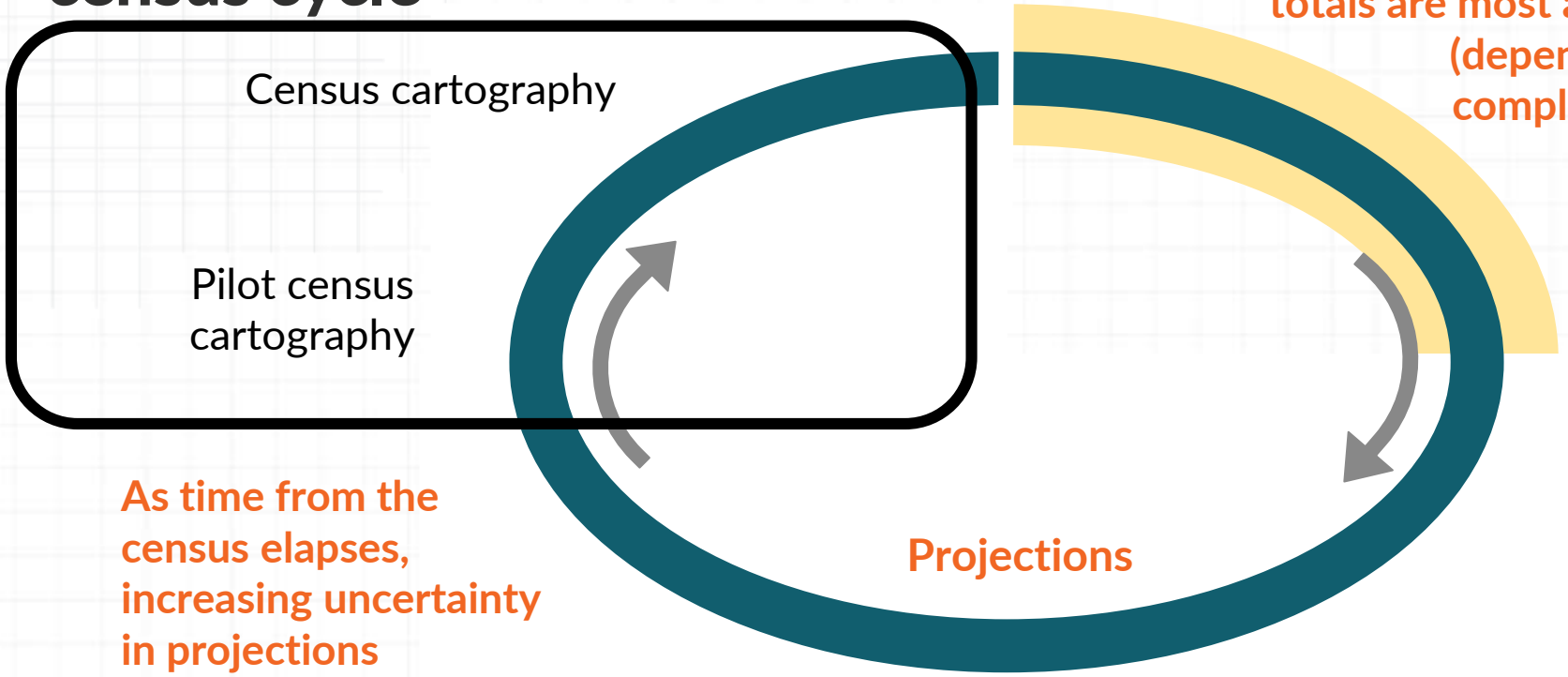
Filling census gaps using satellite and social cartography: Colombia





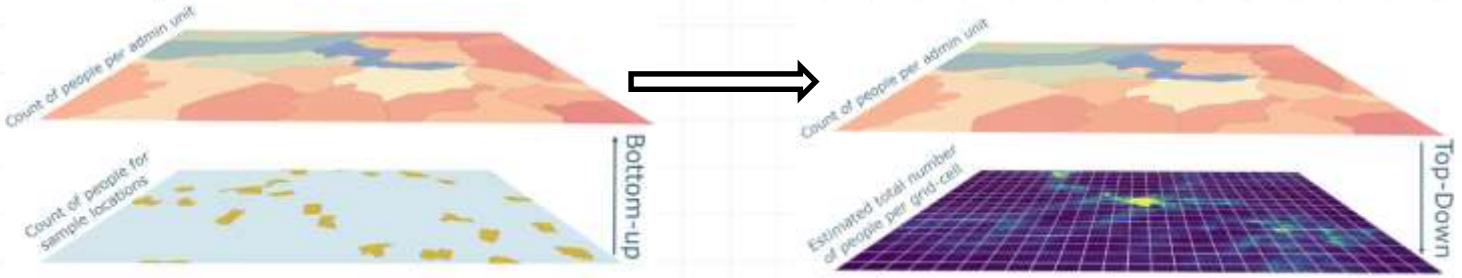
Sanchez-Cespedes et al (2023) Population Studies

Stages of the census cycle



Census process support: Use of census cartography

- Population counts are collected in census preparation phases – e.g. cartography – but can often be incomplete
- Value in having full coverage of recent estimated population for planning the full census
- Bottom-up estimation for under-enumerated units, followed by top-down disaggregation to grid squares



Census cartography: Mali

- Bottom-up model fitted using the population counts of the 25,466 enumeration areas fully covered by the census cartography
- Top-down model used to map enumeration area total to 100x100m grid square estimates

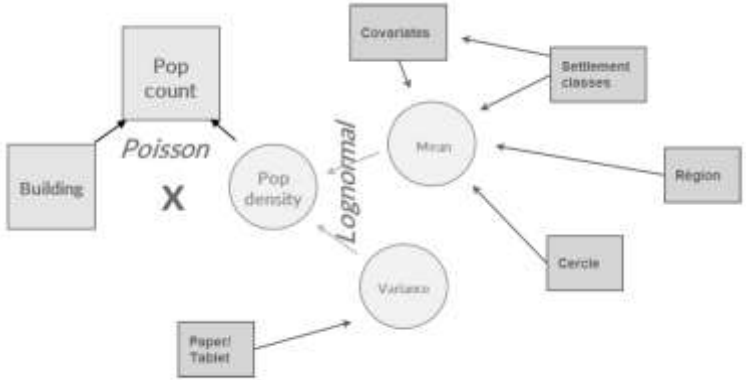


Figure 2 Schema of population model

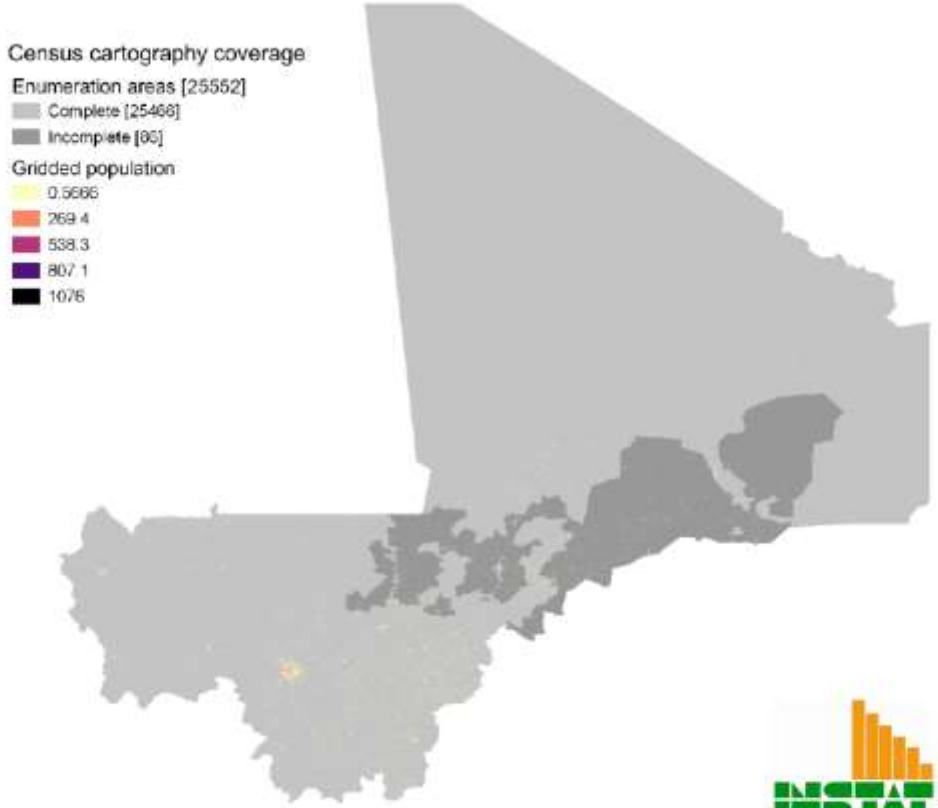


Figure 1 Census cartography coverage

WorldPop and INSTAT (2022) 10.5258/SOTON/WP00745

https://wopr.worldpop.org/?/Population

Questions?

To think about (as you listen to case studies):

-Do the case studies presented have similarities to your own situations?

-If you see a need for it in your country, what types of population data might be available to support bottom-up estimation modelling?