Estimating population density from remote sensing and microcensus

Sohan Seth, Isaac Neal

School of Informatics

THE UNIVERSITY *of* EDINBURGH

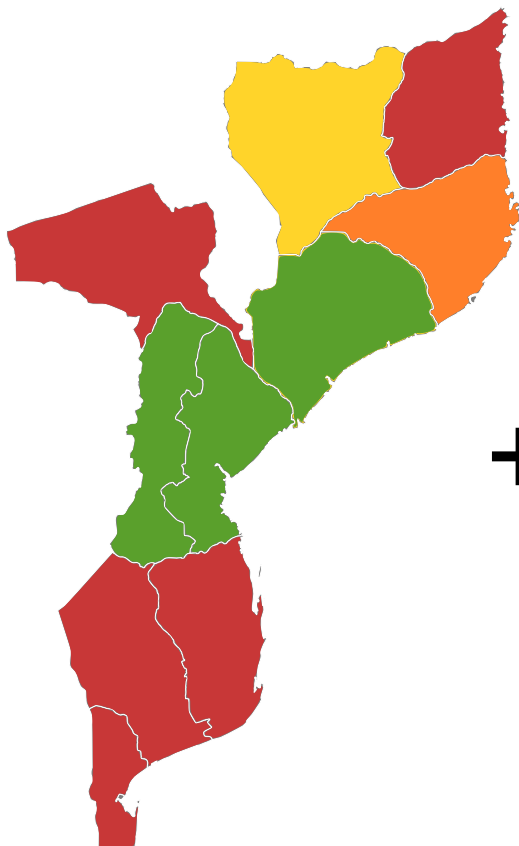# Population estimation

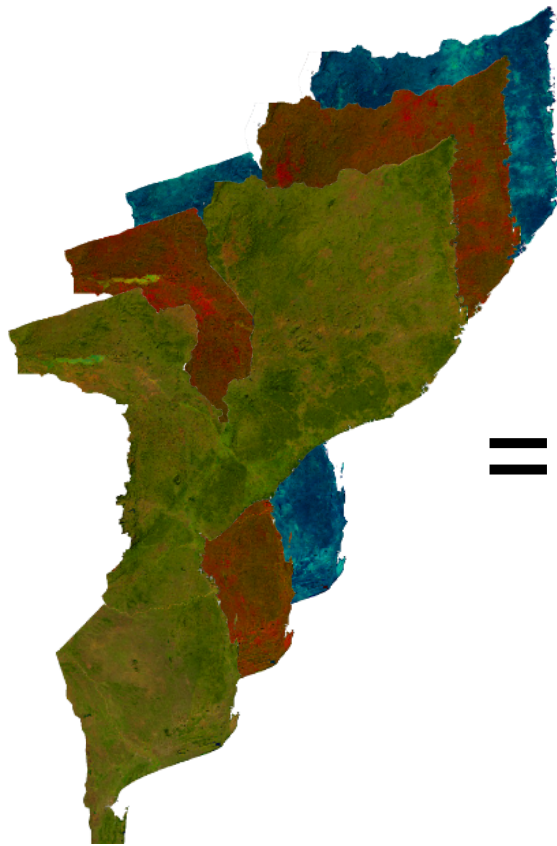# Population estimation

Top-down / population disaggregation

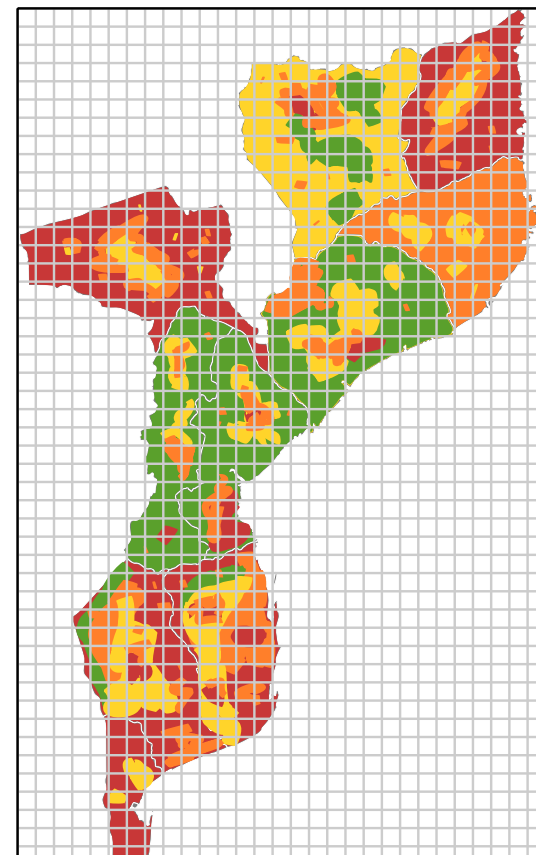**Top-down Estimation**

Regional Census Totals   Geospatial Covariates   Gridded Population Estimate
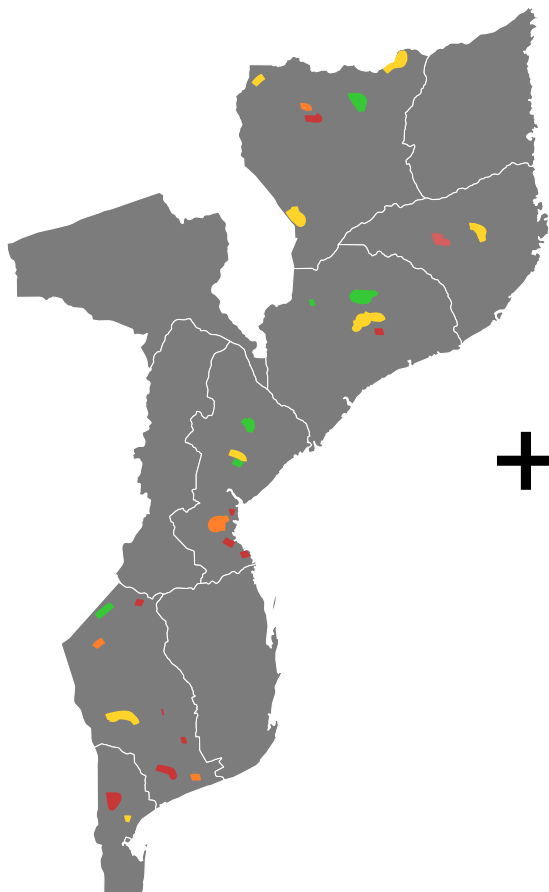

+

=

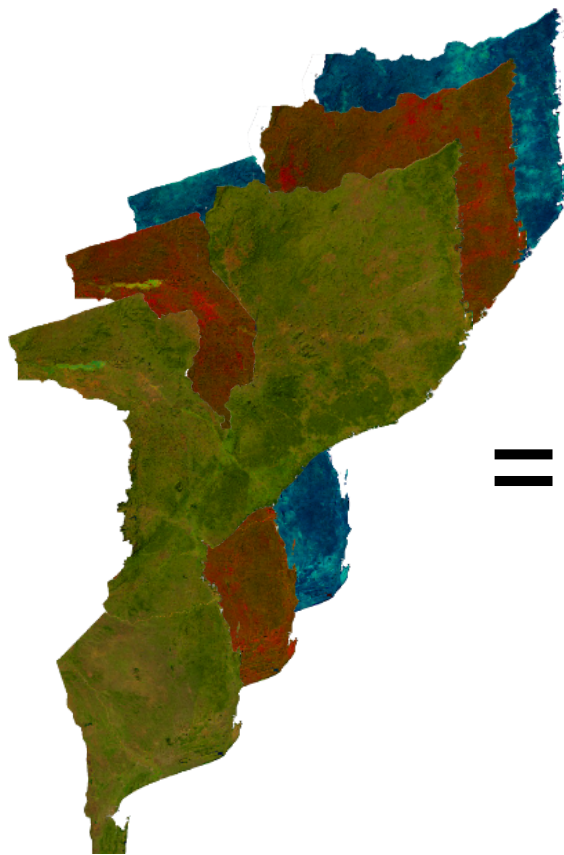# Bottom-up / census independent approach

Bottom-up Estimation
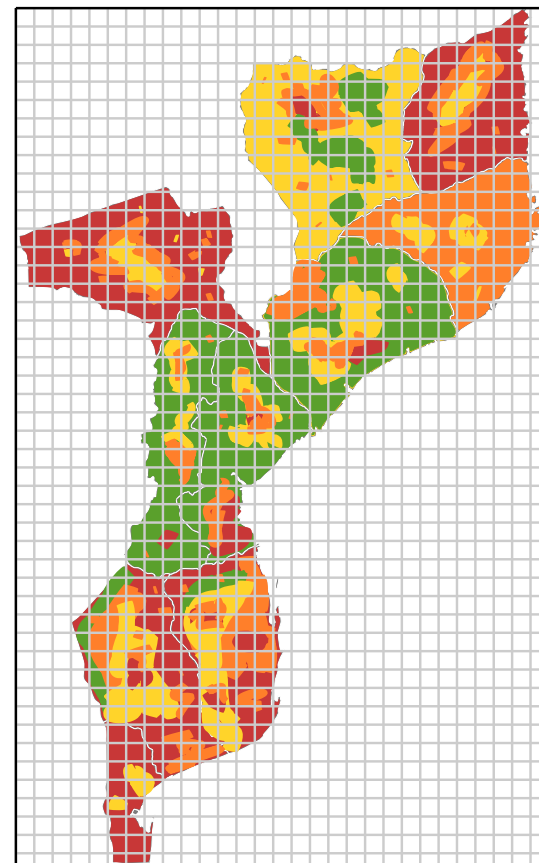
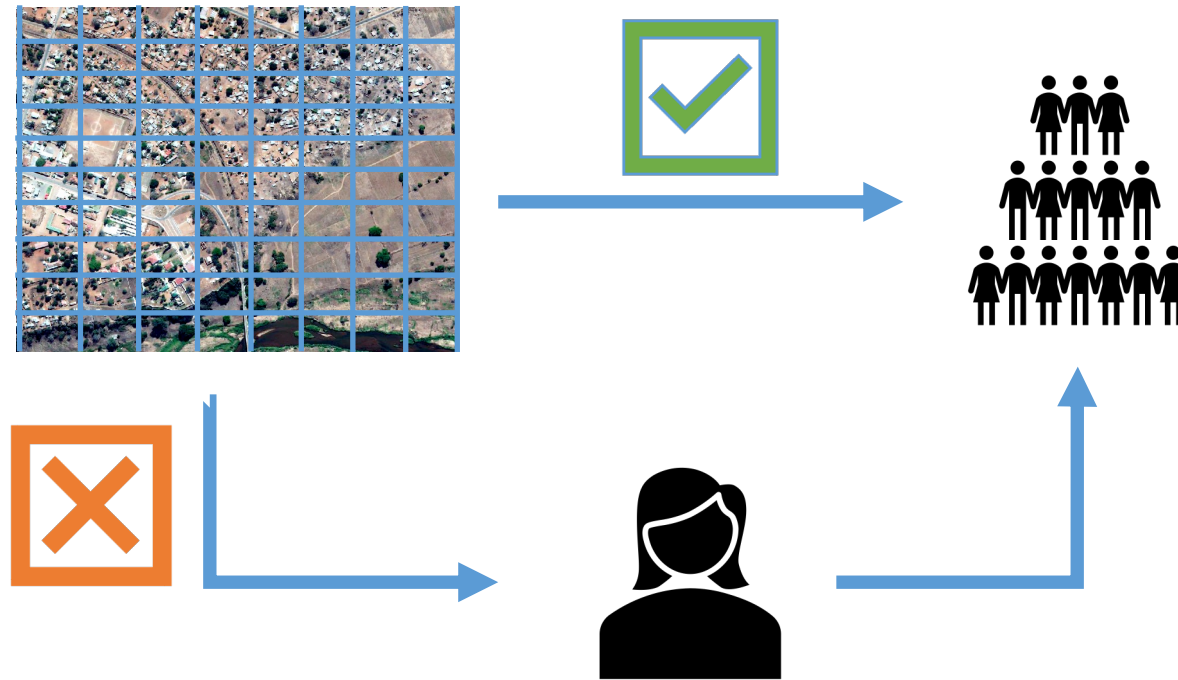Local Micro-census Totals   Geospatial Covariates   Gridded Population Estimate

+

=

Our focus

- Find 'sustainable' approaches, in terms of data
- Reduce human supervision, in terms of methods

Data

Microcensus:

- Two districts in Mozambique Boane and Magudo (288 grid cells)
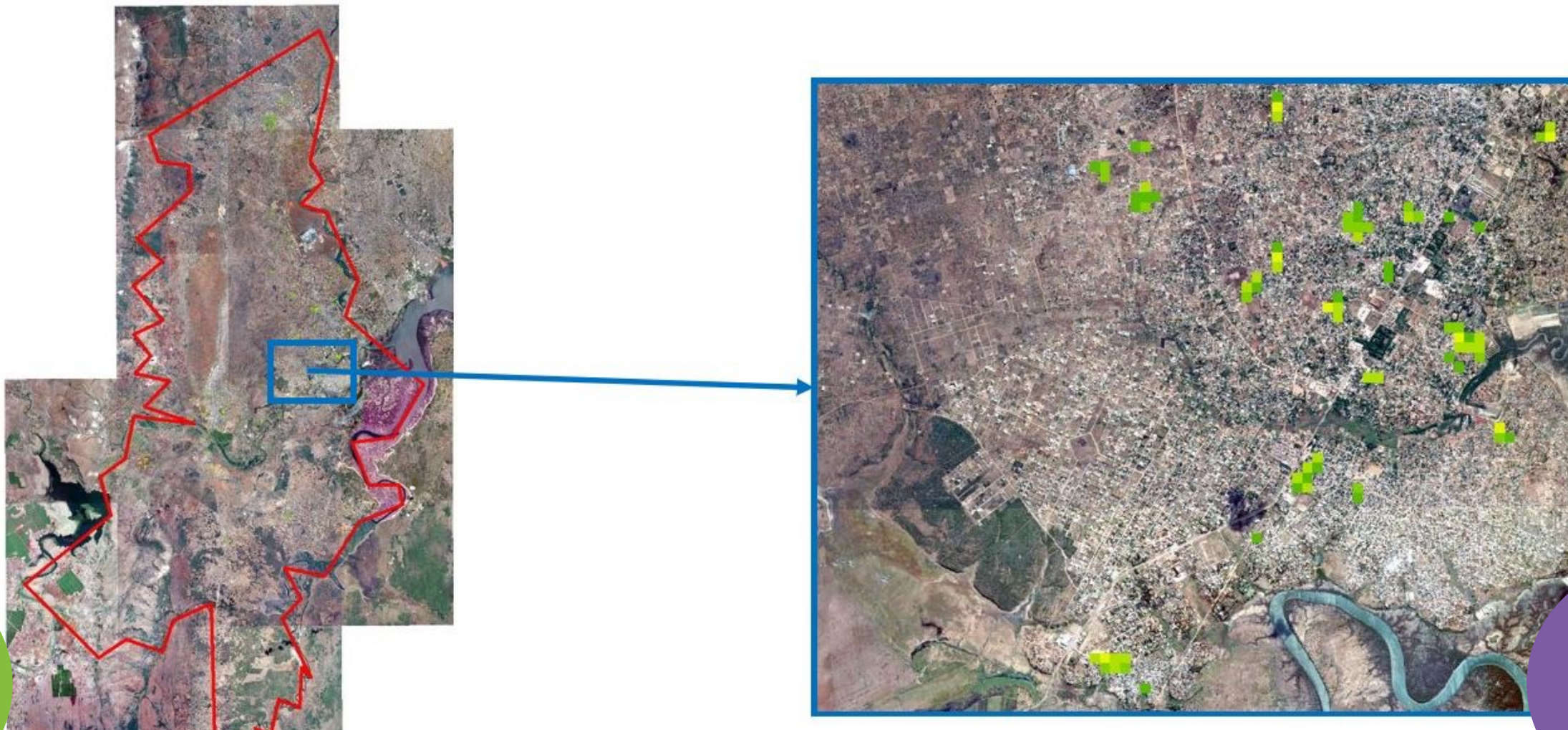
Data

Microcensus:

- Two districts in Mozambique Boane and Magudo (PSU -> 288 grid cells)

Publicly available:

- Road information from OSM (volunteer annotation)
- Land cover classification (100m resolution, updated yearly) [ESA]
- Night-time light information (750m resolution, updated daily)
- LandSat information [Bands, NDVI, NDWI] (30m resolution, every 8 days)
- High Resolution Settlement Layer (HRSL) binary settlement map (30m resolution, 2015) [Facebook]

Data

Microcensus:
- Two districts in Mozambique Boane and Magudo (PSU -> 288 grid cells)

Publicly available:
- Road information from OSM (volunteer annotation)
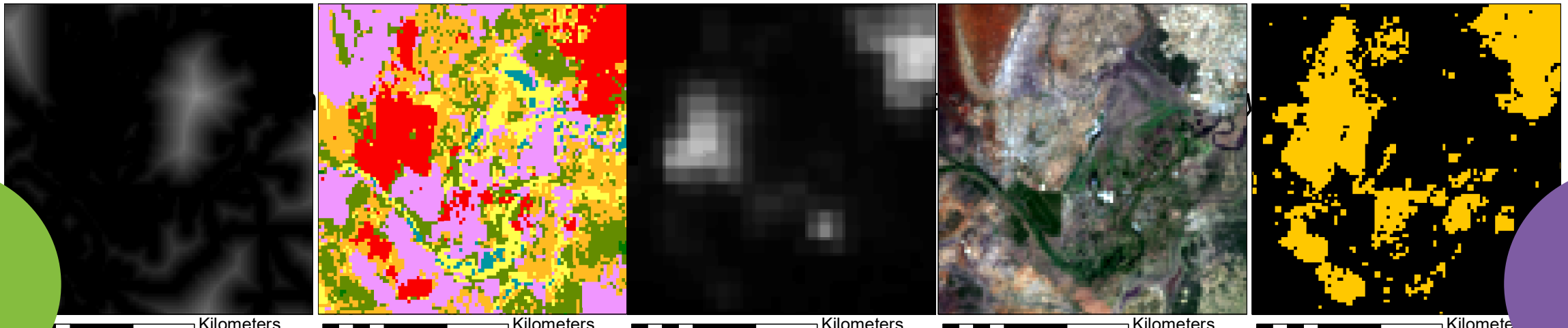- Land cover classification (100m resolution, updated yearly) [ESA]
- Night-time light information (750m resolution, updated daily)
- LandSat information [Bands, NDVI, NDWI] (30m resolution, every 8 days)
- High Resolution Settlement Layer (HRSL) binary settlement map (30m re

Proprietary:
- 0.5m resolution satellite images for extracting building footprints

Kilometers
0 1 2   4   6
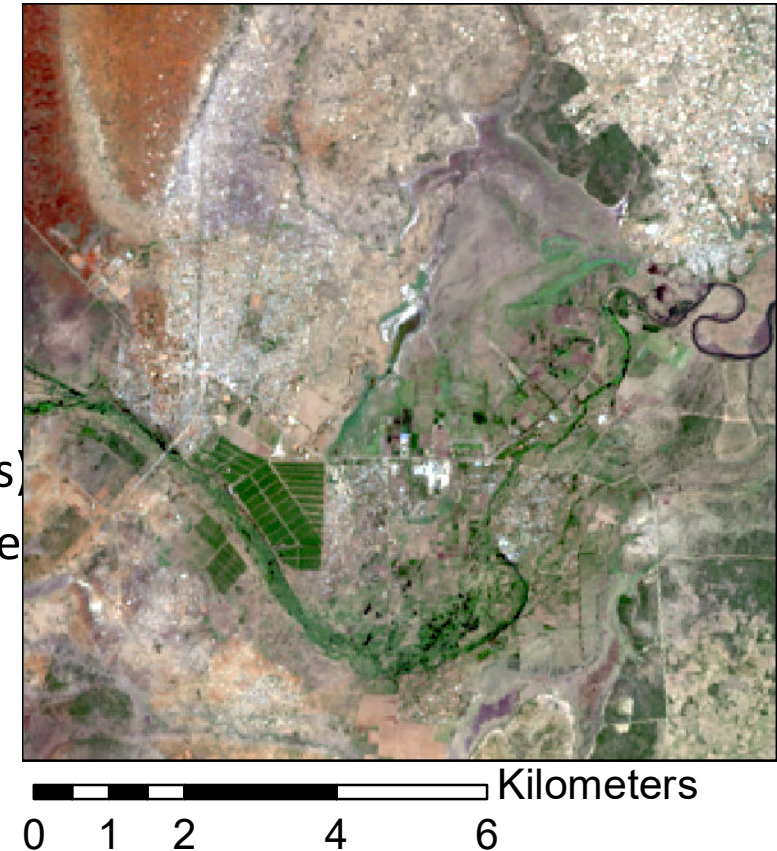
Building footprint

SpaceNet

SpaceNet

SpaceNet

SpaceNet +
SpaceSUR Fine Tuning

SpaceNet +
SpaceSUR Fine Tuning

SpaceNet +
SpaceSUR Fine Tuning
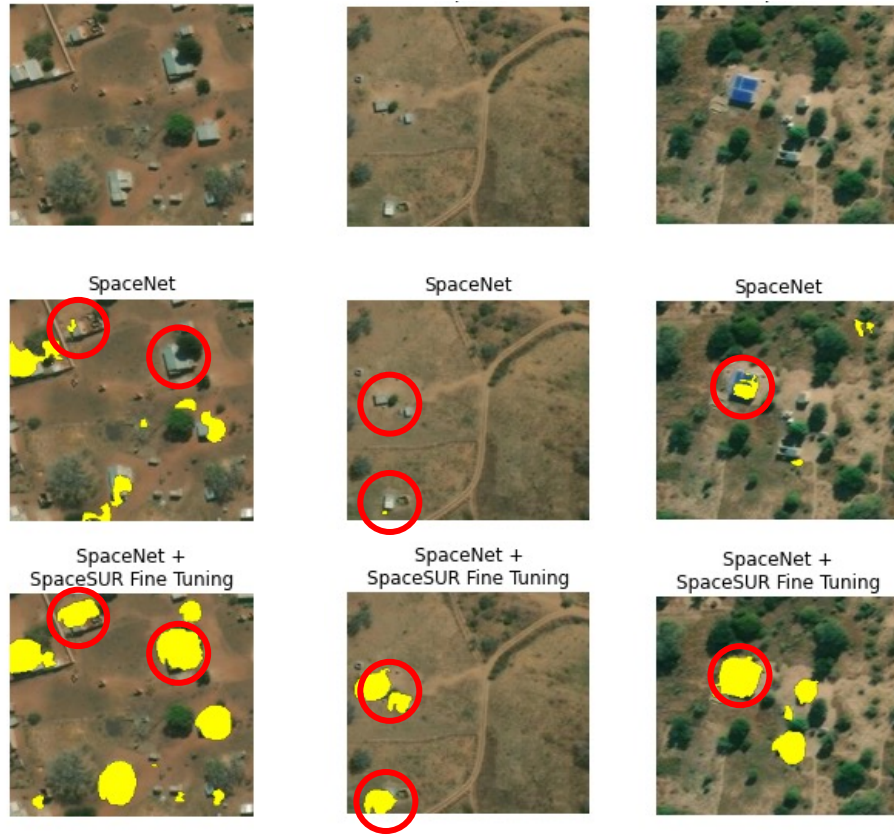
- SpaceNet is a publicly available building footprint dataset covering a large area in Brazil
- Pretrained model (U-Net) without additional human labelling in Mozambique using SpaceNet
- Retrained proof of concept model (1 epoch) on 'dot' labels (SpaceSUR and GroundWork)

- We don't need exact estimate: our goal is population not footprint

Results

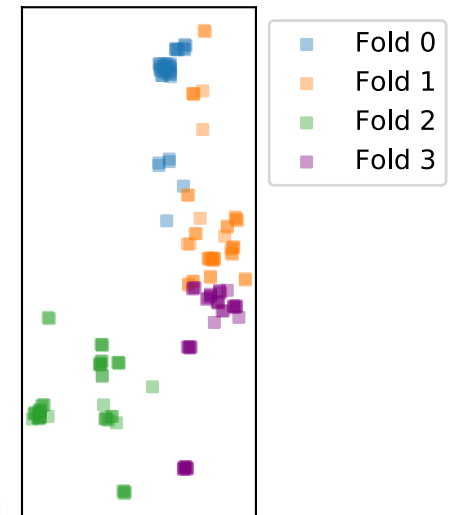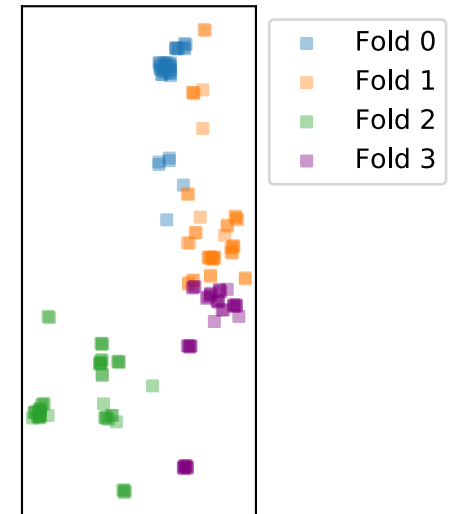| Features used	GLM (Poisson)				Random Forest			
	R^2	MeAPE	MeAE	AggPE	R^2	MeAPE	MeAE	AggPE
Outliers Retained								
All	0.0756	54.7%	4.65	2.4%	0.0217	50.3%	3.59	18.9%
SpaceSUR only[6]	0.127	53.5%	4.30	2.11%	-0.00272	49.7%	3.41	20.8%
SpaceNet only[7]	0.0486	63.3%	5.08	1.54%	-0.174	54.8%	3.85	25.3%
Public + SpaceNet	-0.0113	54.6%	4.94	4.42%	-0.0883	53.1%	4.06	18%
Public	0.0126	55.8%	4.89	2.79%	-0.108	53%	3.94	19%
Outliers Removed								
All	0.298	42.4%	3.71	4.34%	0.441	36.3%	3.09	10.6%
SpaceSUR only	0.468	39.9%	3.62	0.626%	0.399	35.6%	3.04	12.7%
SpaceNet only	-0.0118	58.7%	6.17	6.12%	-0.0814	49.8%	4.48	17.8%
Public + SpaceNet	0.135	46.1%	4.44	7.59%	0.232	40.8%	4.21	11.1%
Public	0.0988	48.1%	4.25	5.53%	0.163	44.2%	3.98	14.7%

Legend: Fold 0, Fold 1, Fold 2, Fold 3

- MeAPE: Median absolute percentage error
- MeAE: Median absolute error
- AggPE: Aggregated percentage error

Results

Features used	GLM (Poisson)				Random Forest			
	R^2	MEAPE	MEAE	AGGPE	R^2	MEAPE	MEAE	AGGPE
Outliers Retained								
All	0.0756	54.7%	4.65	2.4%	0.0217	50.3%	3.59	18.9%
SpaceSUR only[6]	0.127	53.5%	4.30	2.11%	-0.00272	49.7%	3.41	20.8%
SpaceNet only[7]	0.0486	63.3%	5.08	1.54%	-0.174	54.8%	3.85	25.3%
Public + SpaceNet	-0.0113	54.6%	4.94	4.42%	-0.0883	53.1%	4.06	18%
Public	0.0126	55.8%	4.89	2.79%	-0.108	53%	3.94	19%
Outliers Removed								
All	0.298	42.4%	3.71	4.34%	0.441	36.3%	3.09	10.6%
SpaceSUR only	0.468	39.9%	3.62	0.626%	0.399	35.6%	3.04	12.7%
SpaceNet only	-0.0118	58.7%	6.17	6.12%	-0.0814	49.8%	4.48	17.8%
Public + SpaceNet	0.135	46.1%	4.44	7.59%	0.232	40.8%	4.21	11.1%
Public	0.0988	48.1%	4.25	5.53%	0.163	44.2%	3.98	14.7%

- MeAPE: Median absolute percentage error

- MeAE: Median absolute error

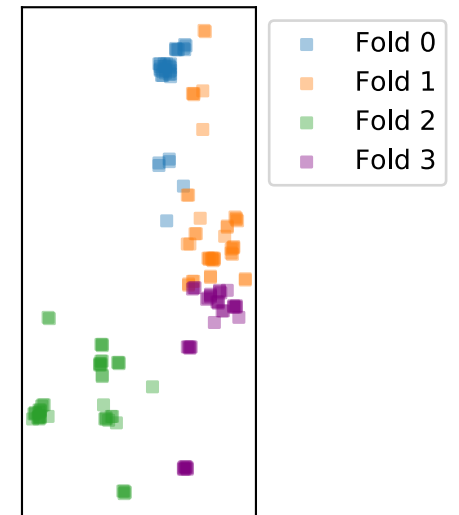- AggPE: Aggregated percentage error

Results

Features used	GLM (Poisson)				Random Forest			
	R^2	MeAPE	MeAE	AggPE	R^2	MeAPE	MeAE	AggPE
Outliers Retained								
All	0.0756	54.7%	4.65	2.4%	0.0217	50.3%	3.59	18.9%
SpaceSUR only[6]	0.127	53.5%	4.30	2.11%	-0.00272	49.7%	3.41	20.8%
SpaceNet only[7]	0.0486	63.3%	5.08	1.54%	-0.174	54.8%	3.85	25.3%
Public + SpaceNet	-0.0113	54.6%	4.94	4.42%	-0.0883	53.1%	4.06	18%
Public	0.0126	55.8%	4.89	2.79%	-0.108	53%	3.94	19%
Outliers Removed								
All	0.298	42.4%	3.71	4.34%	0.441	36.3%	3.09	10.6%
SpaceSUR only	0.468	39.9%	3.62	0.626%	0.399	35.6%	3.04	12.7%
SpaceNet only	-0.0118	58.7%	6.17	6.12%	-0.0814	49.8%	4.48	17.8%
Public + SpaceNet	0.135	46.1%	4.44	7.59%	0.232	40.8%	4.21	11.1%
Public	0.0988	48.1%	4.25	5.53%	0.163	44.2%	3.98	14.7%

Fold 0
Fold 1
Fold 2
Fold 3

- MeAPE: Median absolute percentage error
- MeAE: Median absolute error
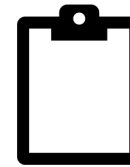- AggPE: Aggregated percentage error

Challenges

The study is limited by the number of samples we have. We discarded many samples as remote sensing and microcensus do not match.
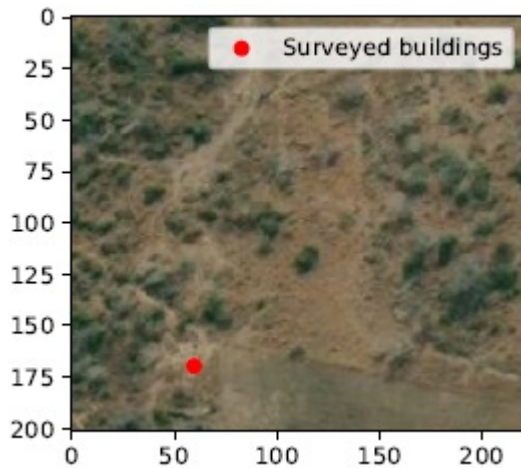
No buildings — Non-zero population

Low population — Built up area

Next

- More microcensus data to validate the findings.
- Improve pipeline for detecting building counts, individual areas, non-residential buildings, historic sites, buildings under construction and/or in ruins.
- Use spatial correlation of local population
- Better validation tools and evaluation metrics

Thanks

- Isaac Neal (School of Informatics)
- Gary Watmough (School of GeoSciences)
- Mamadou Saliou Diallo (UNICEF)
- Alex Hutchison (Data for Children Collaborative)