MAUPP

MODELLING INTRAURBAN POPULATION DISTRIBUTION USING FREELY AVAILABLE DATASETS

JE. Steele, JJ. Nieves, Y. Forget, M. Shimoni, AJ. Tatem and C. Linard

Advanced mapping of urban population patterns in sub-Saharan Africa: MAUPP Conference, Nairobi KEN 30.01.19

Why map populations?

Mapping the denominator

Governments are reliant on accurate and up-to-date dat
On population numbers and distributions for planning health services

-What is the catchment population of my health facility?
-How many people are at risk of malaria?

Targeting vulnerable populations

To improve the health of the poorest and most vulnerable an target interventions to lift them out of poverty, we need to know where they are

-How much vaccine is needed for this ward?
-Do the poorest have access to bednets

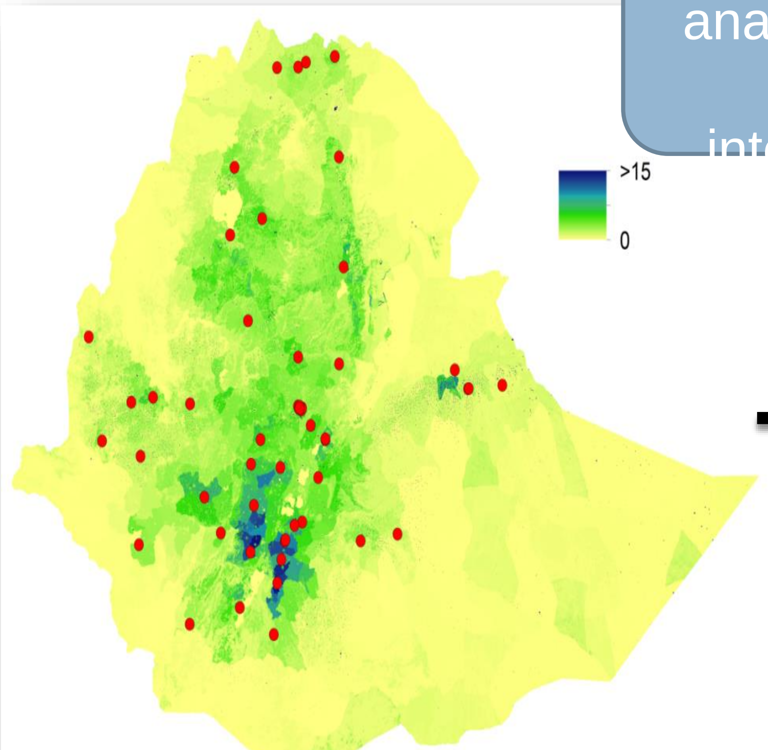
Emergency response

Effective response requires rapid and ongoing assessments of numbers of vulnerable people affected and future risks

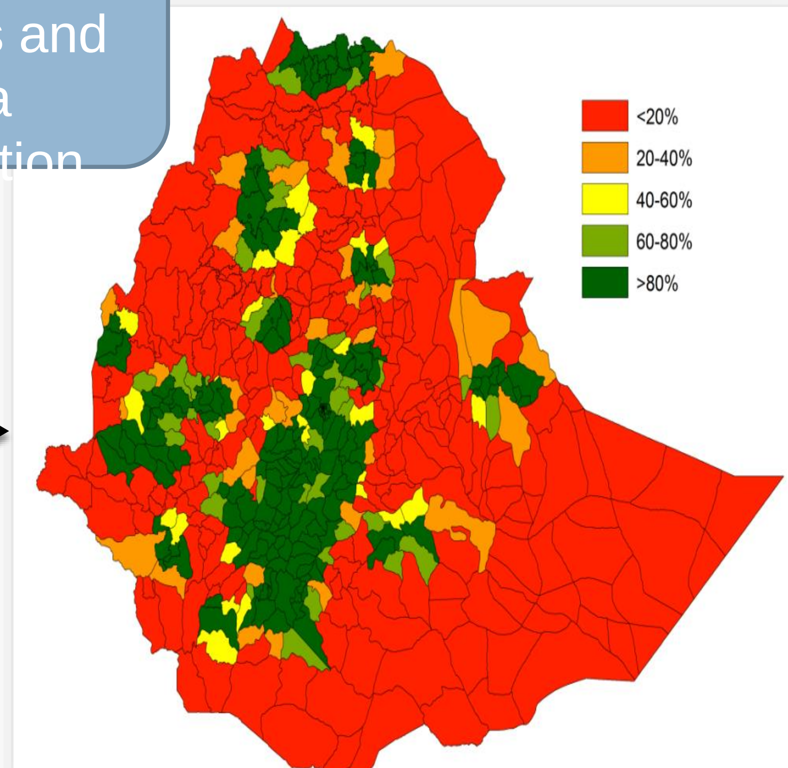
-How many people are in the outbreak area?
-How many vulnerable people were affected and displaced by the earthquake?

Benefits of ‘gridded’ demographic data

Grids: flexibility in analysis and data integration



Comprehensive Emergency Obstetric and Neonatal Care (CEmONC) Facilities overlaid on grid of women of childbearing age

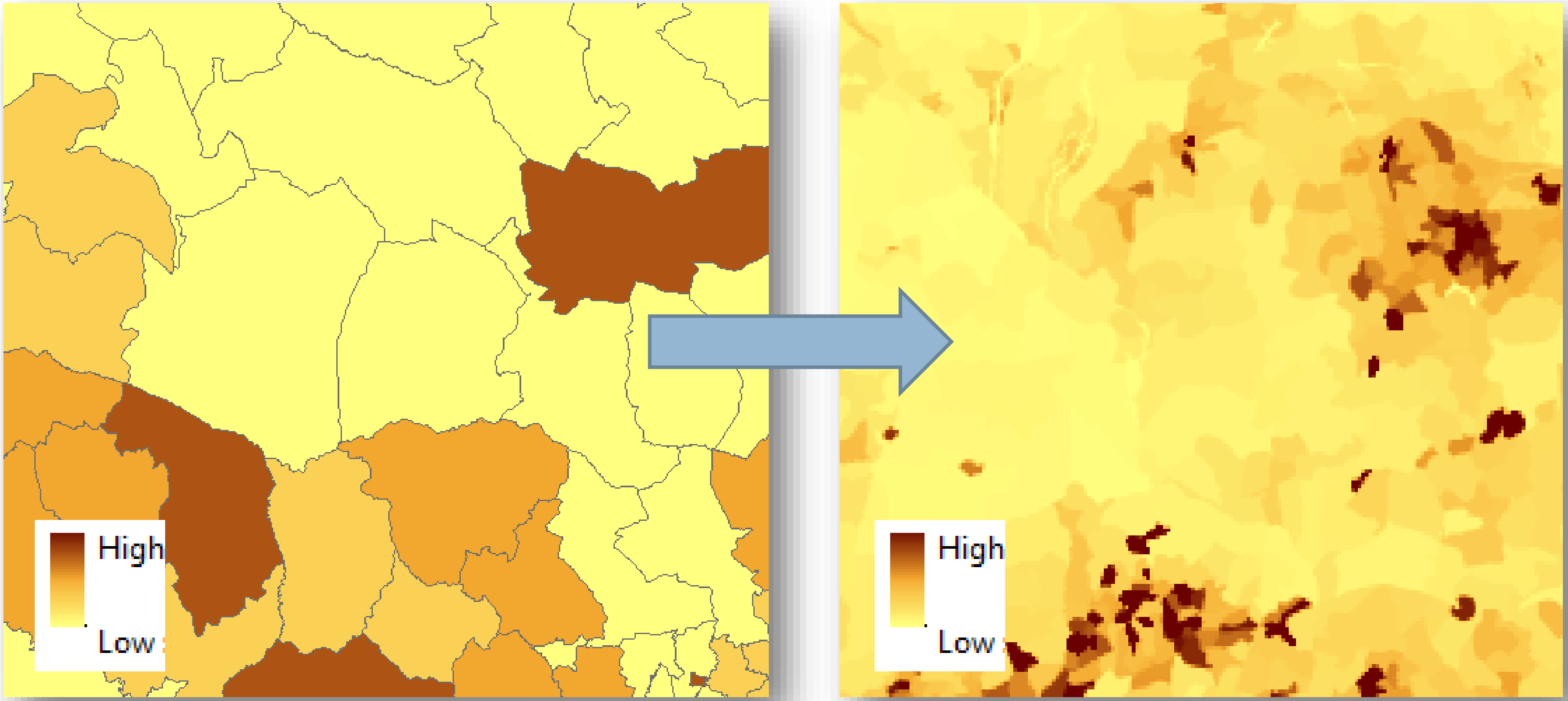


Percentage of women of childbearing age per woreda within 50km of a CEmONC

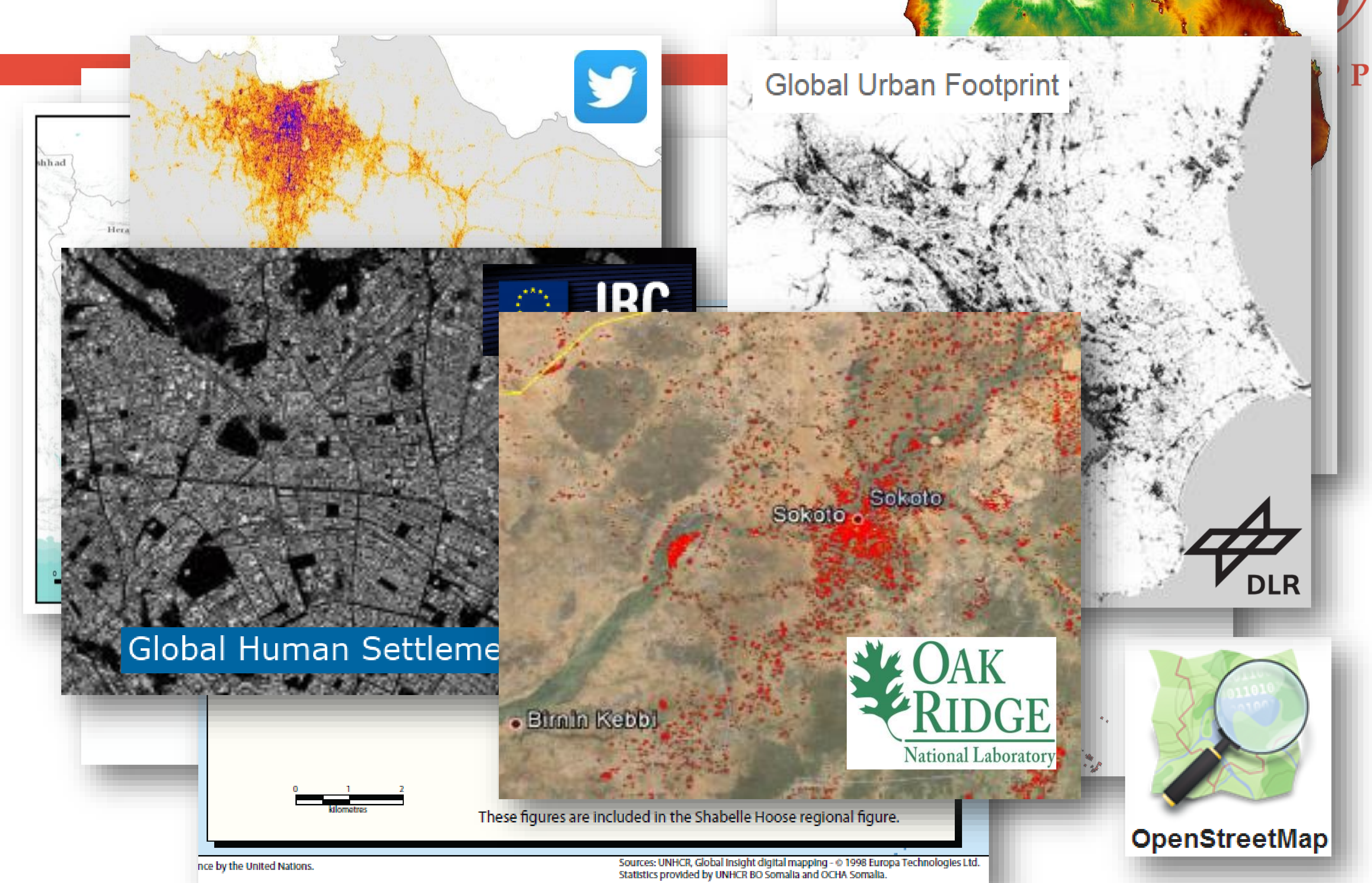
differing data types

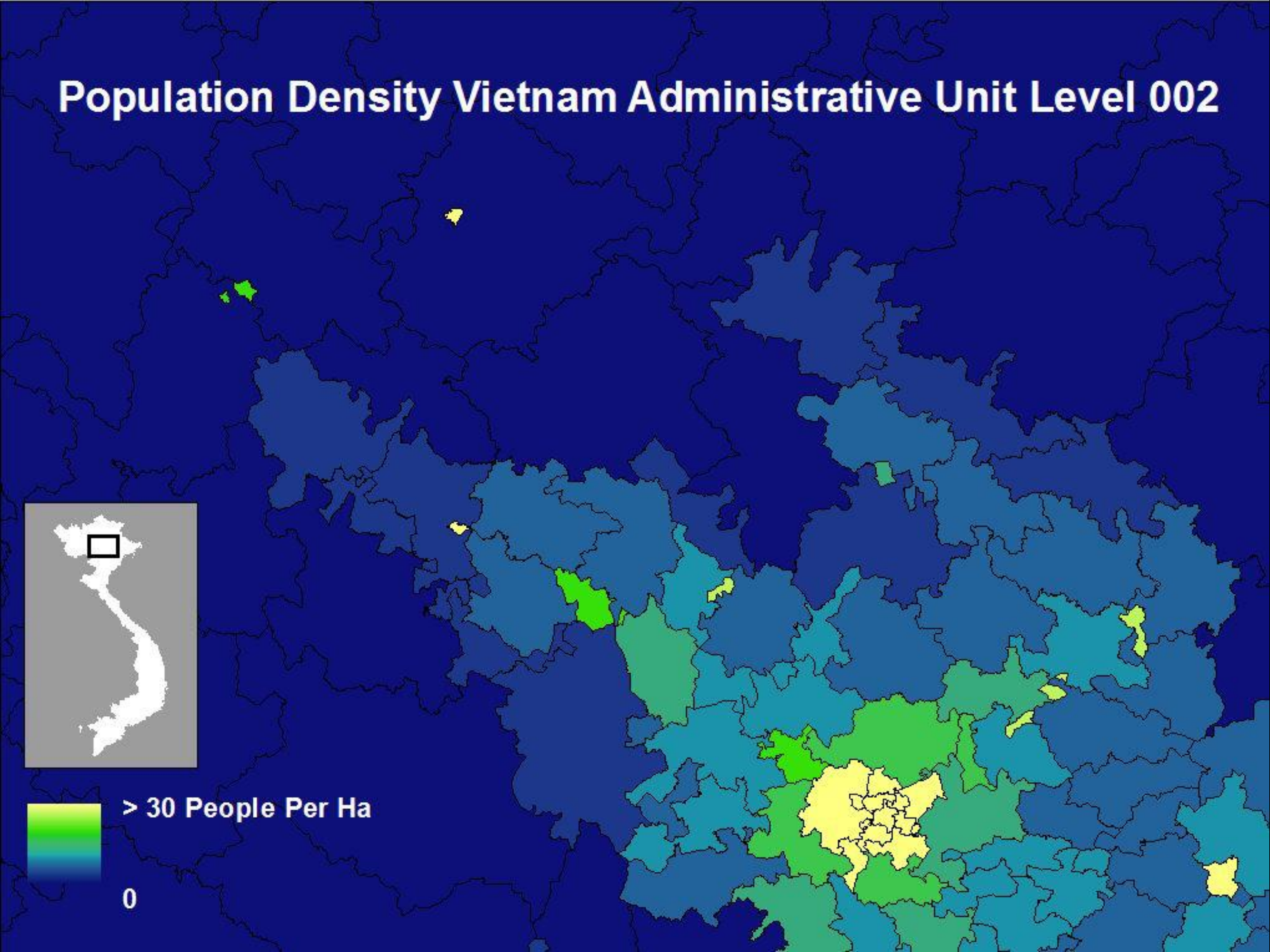
unit population totals

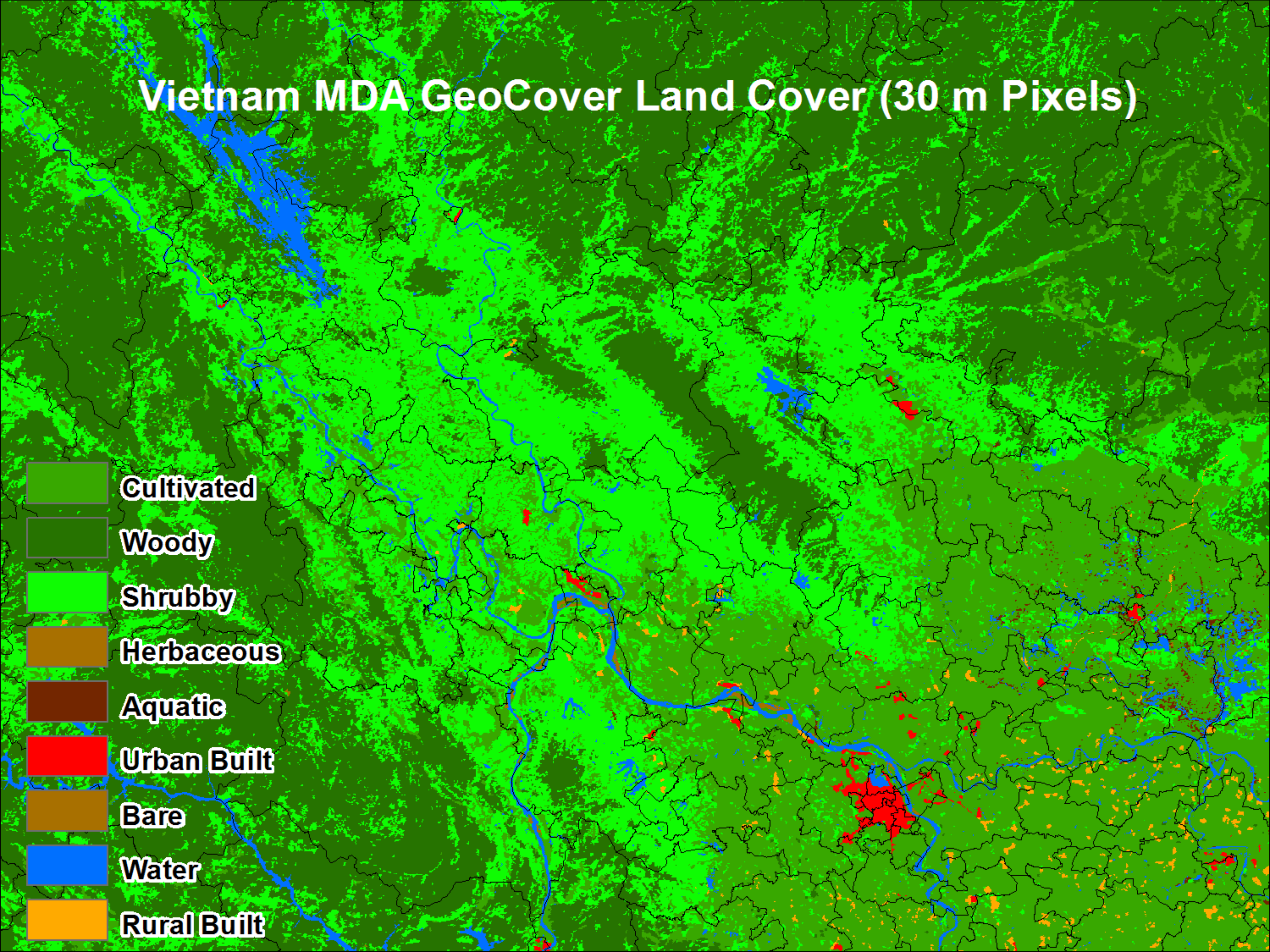
How can we go from aggregated counts of population to fine-scale gridded estimates?



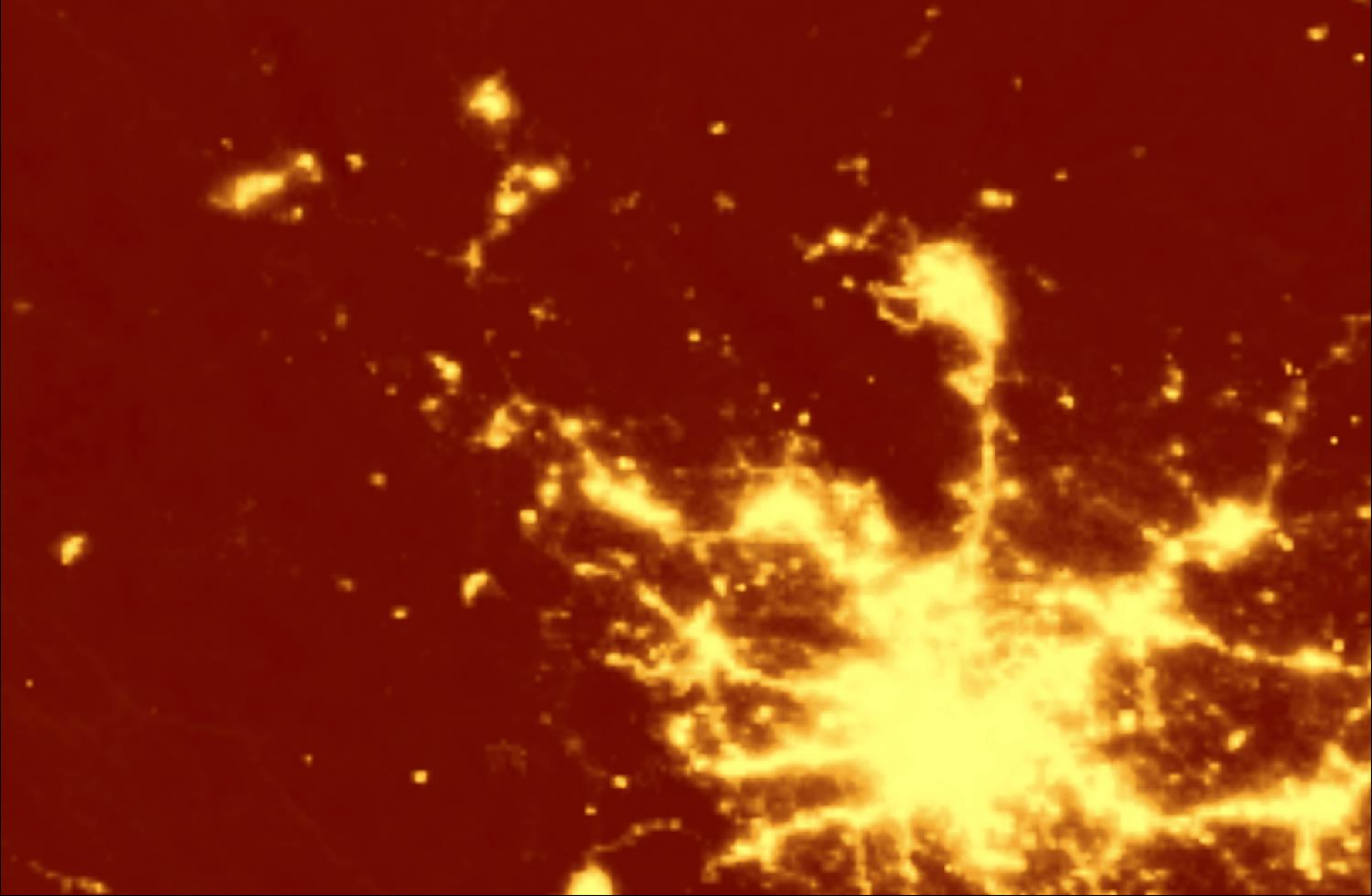
Useful data types



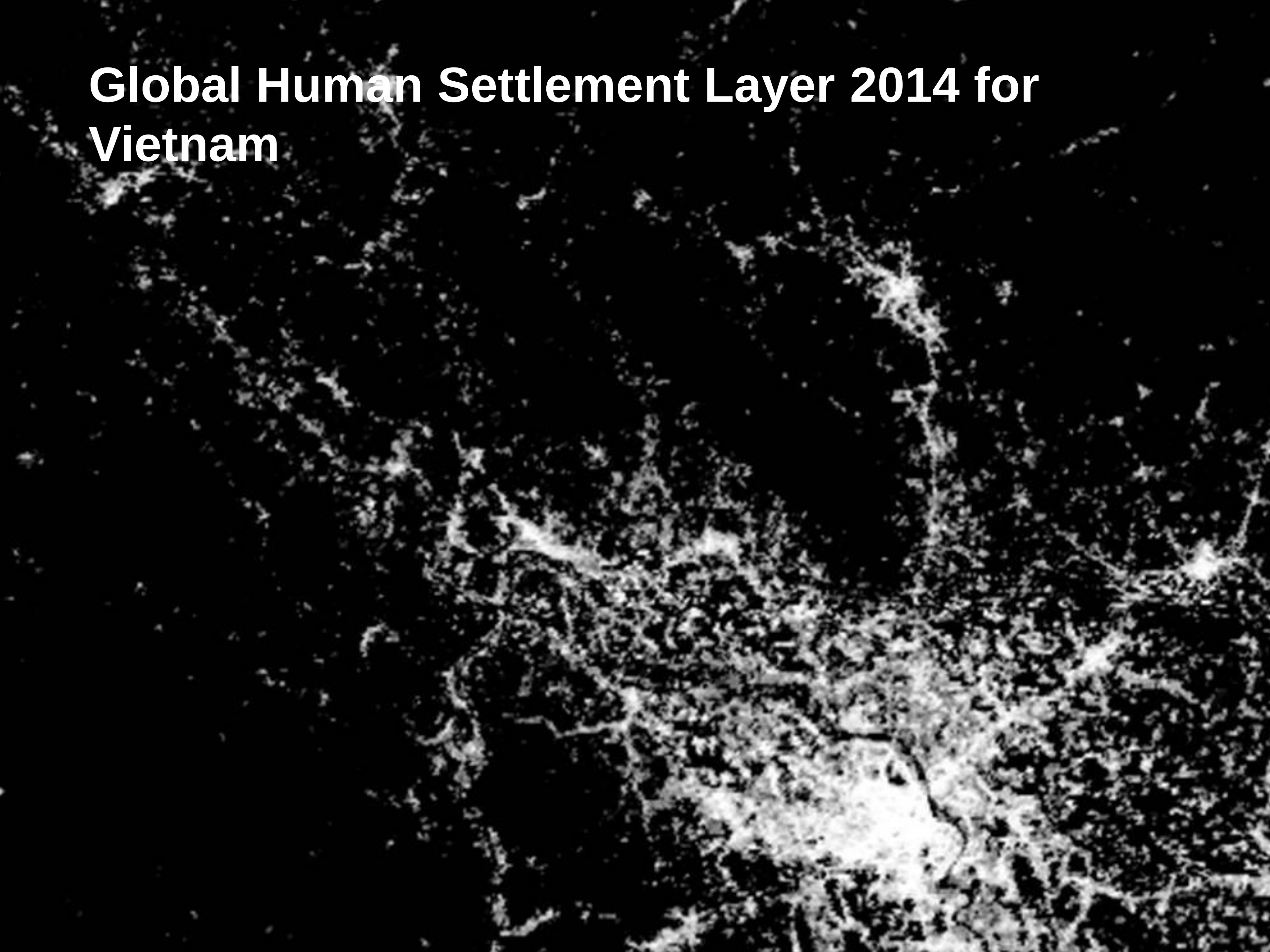

Population Density Vietnam Administrative Unit Level 002
> 30 People Per Ha
0

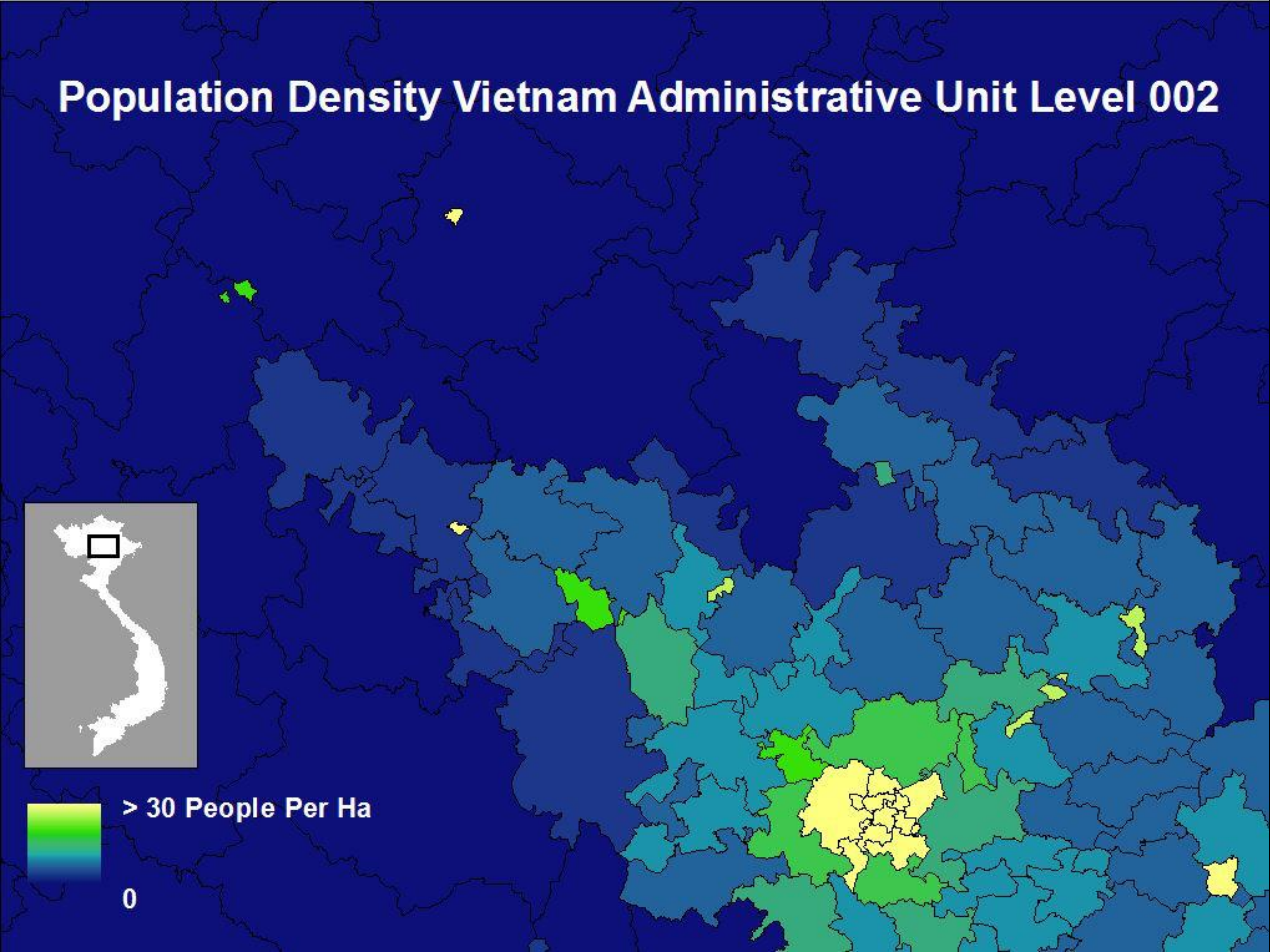

Vietnam MDA GeoCover Land Cover (30 m Pixels)
Cultivated
Woody
Shrubby
Herbaceous
Aquatic
Urban Built
Bare
Water
Rural Built

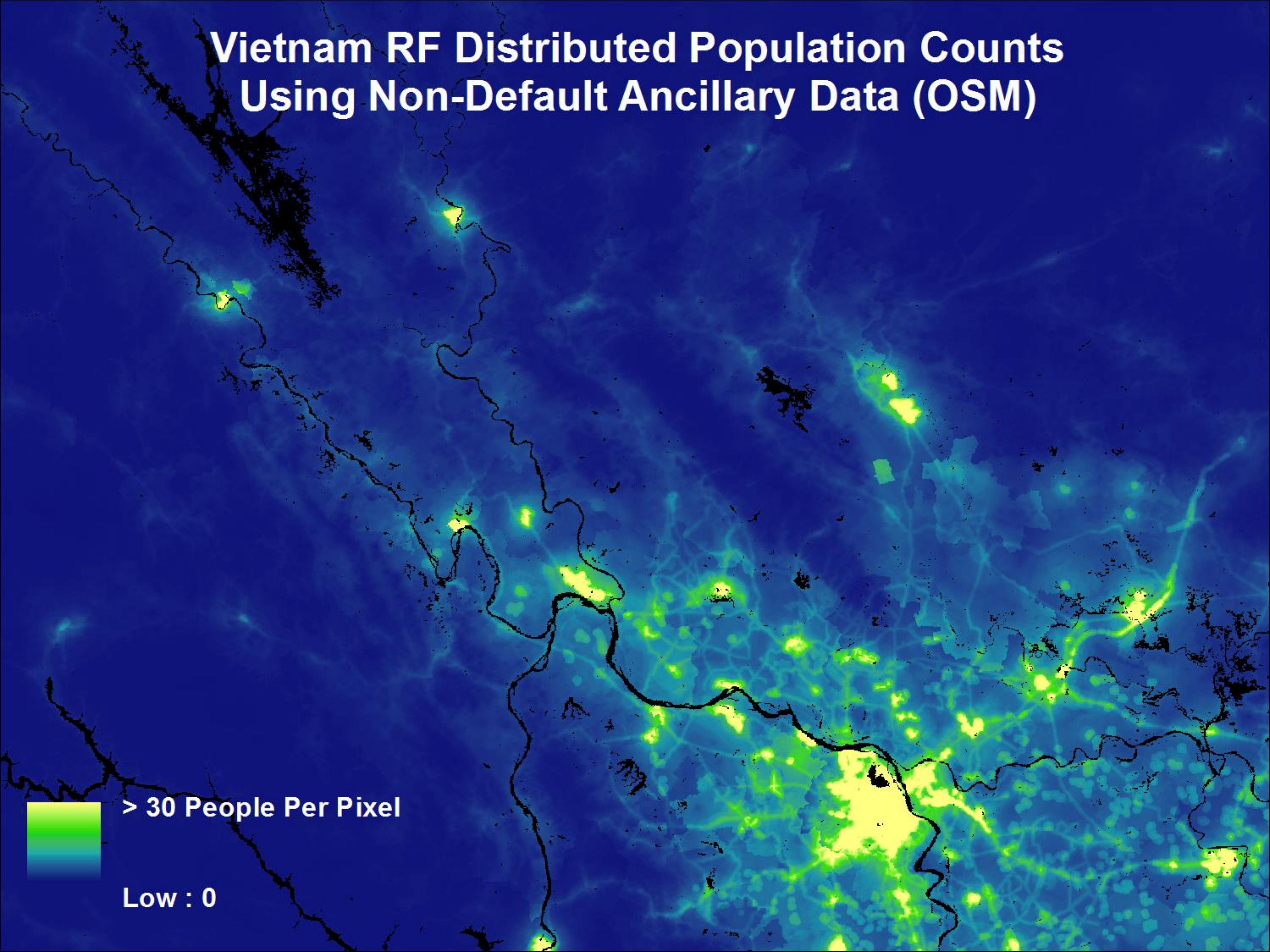
NOAA Suomi VIIRS-derived Lights at Night 2012 for Vietnam



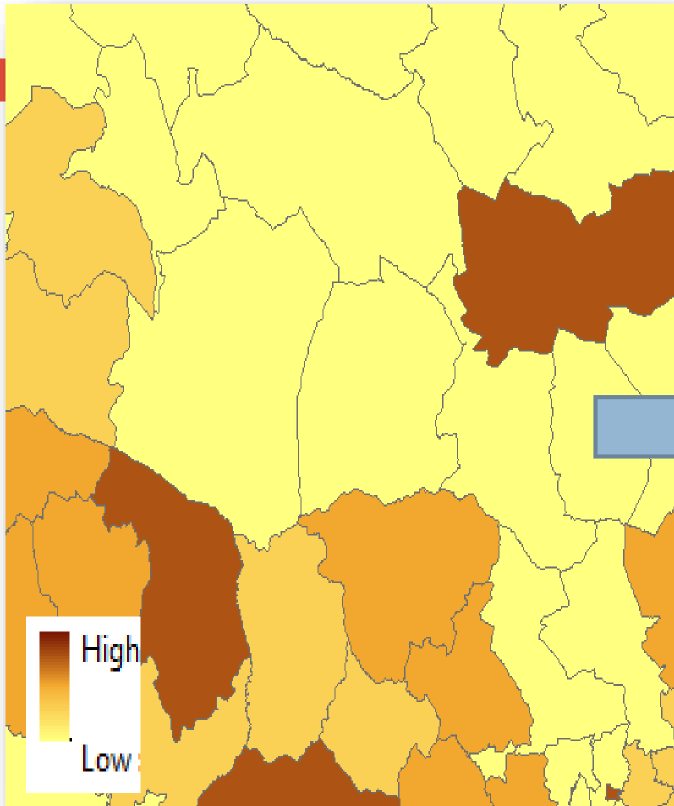
Global Human Settlement Layer 2014 for Vietnam

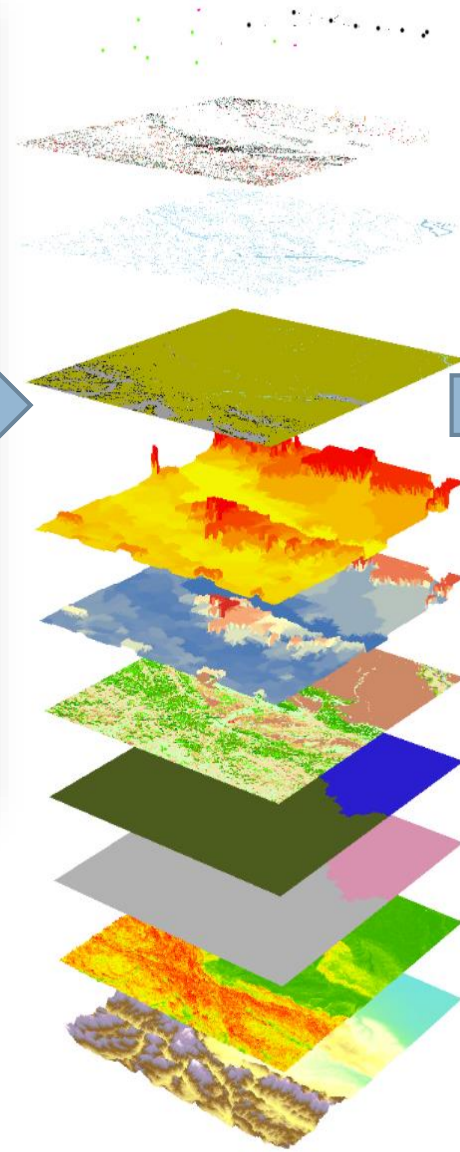


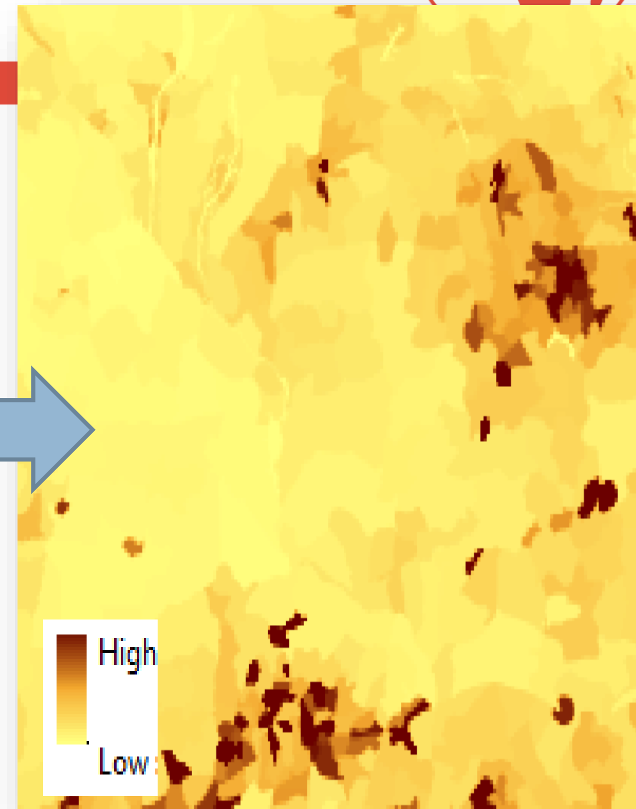

Population Density Vietnam Administrative Unit Level 002
> 30 People Per Ha
0


Vietnam RF Distributed Population Counts
Using Non-Default Ancillary Data (OSM)
> 30 People Per Pixel
Low : 0

Census data disaggregation



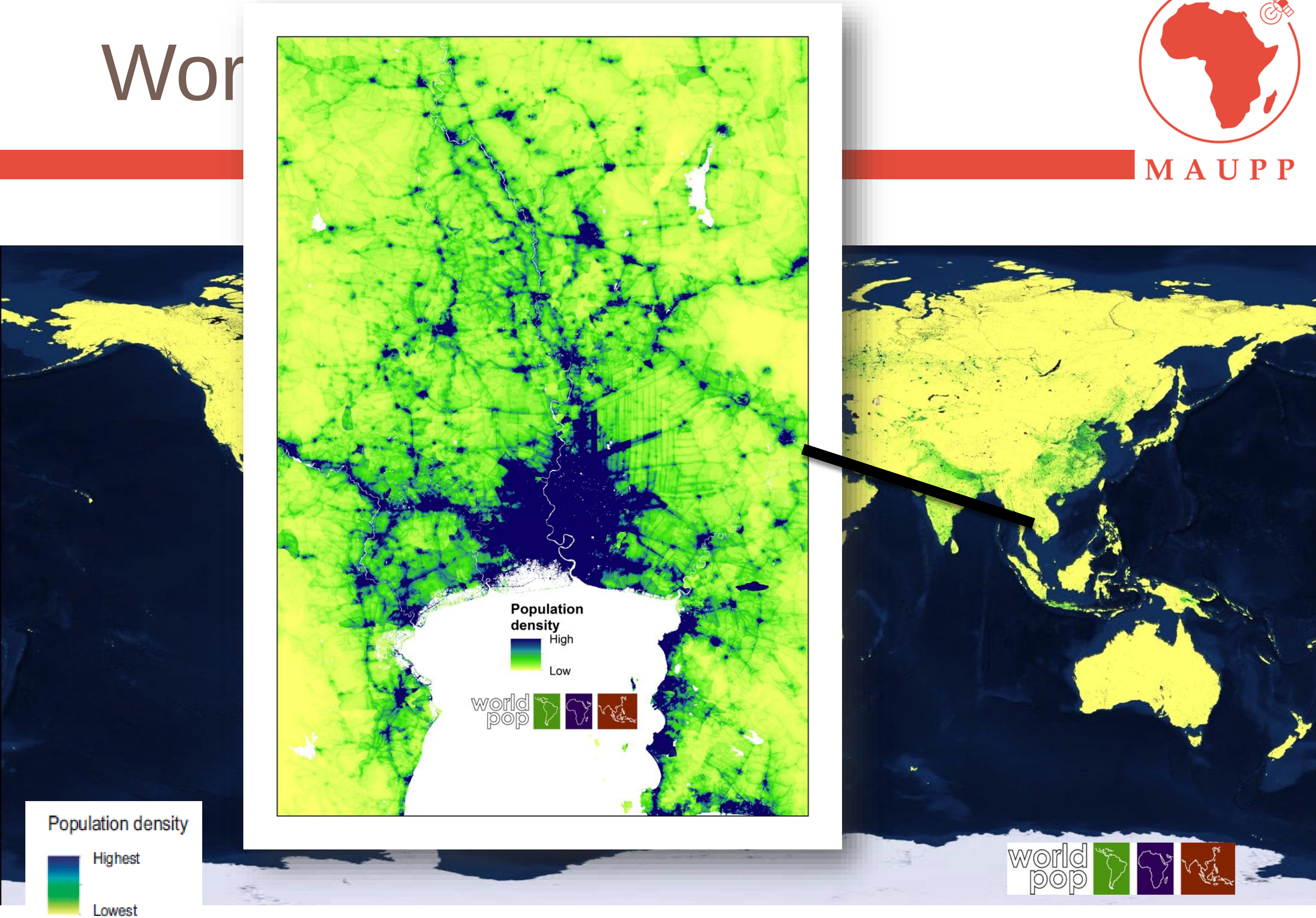




Census counts generally aggregated at coarse, irregular administrative unit level, making integration and comparisons with other data challenging

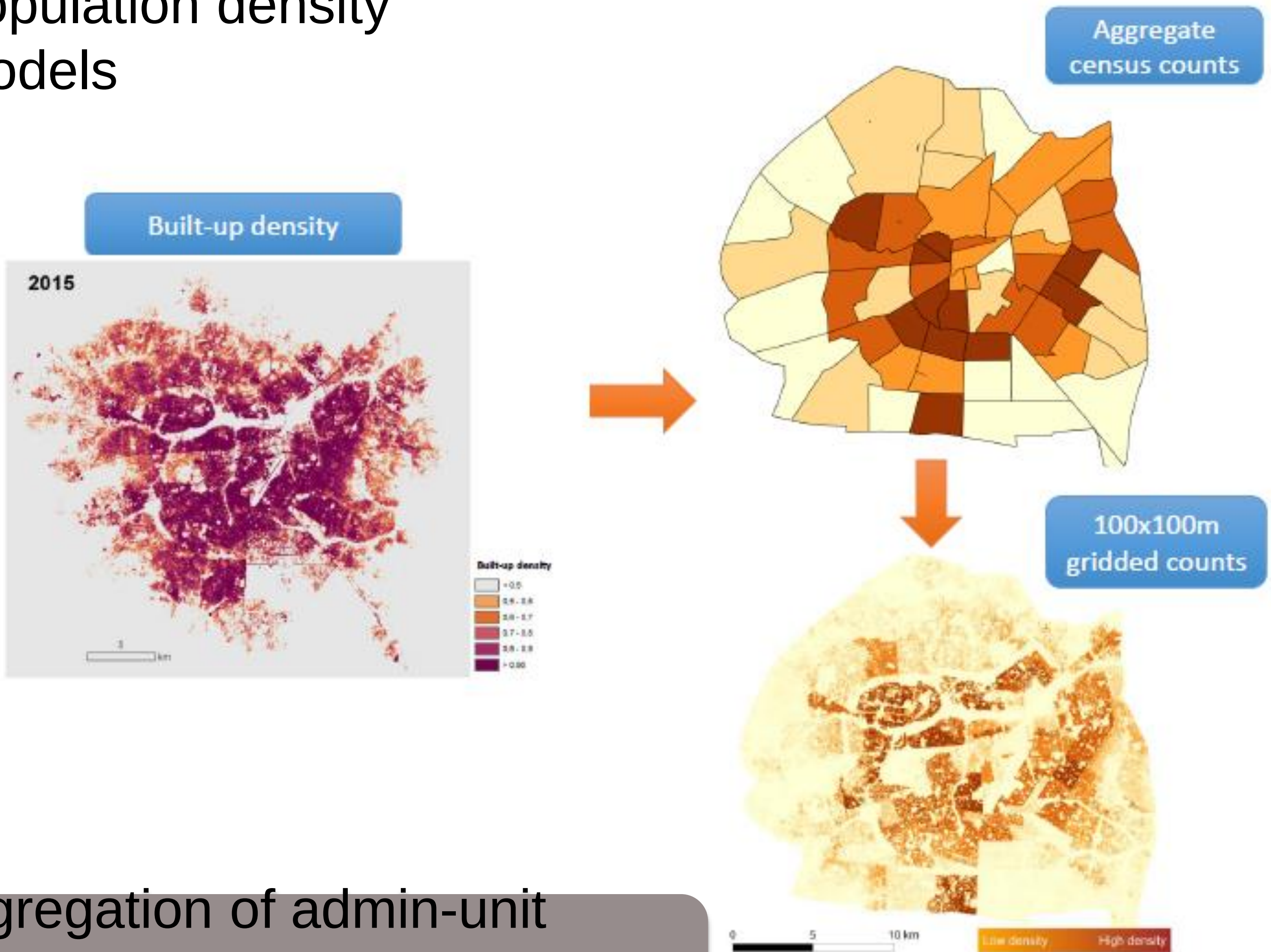
Integration with satellite/GIS data related to human population distribution patterns to disaggregate counts to regular grids using machine learning

Wor


Population density
Highest
Lowest
world pop
Population density
High
Low
world pop

MAUPP

Population density models



Disaggregation of admin-unit based census/official estimate counts

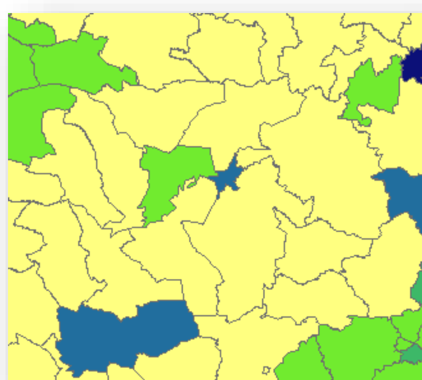
Urban population mapping in Africa

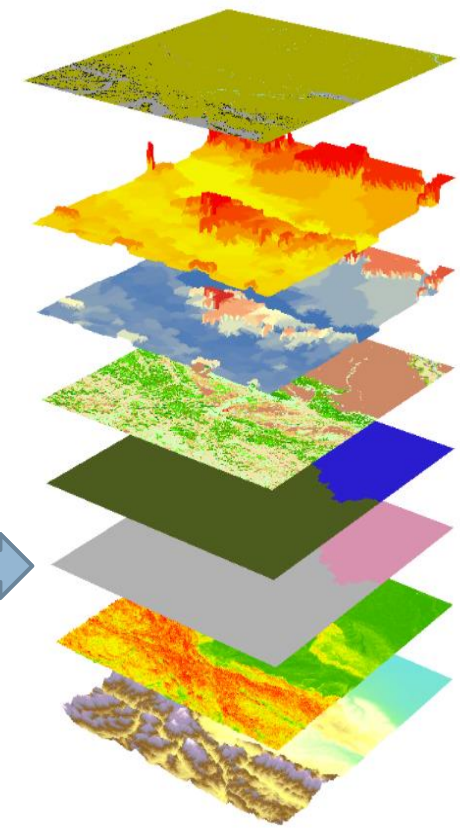
- Challenges
  - Heterogeneity of the build-up structures, and corresponding human population density
  - Lack of good quality training datasets

- Objective
  - Test the added value of the built-up probability/density layer produced by MAUPP to improve models of human population density within urban extents

Random Forest Classification and Regression

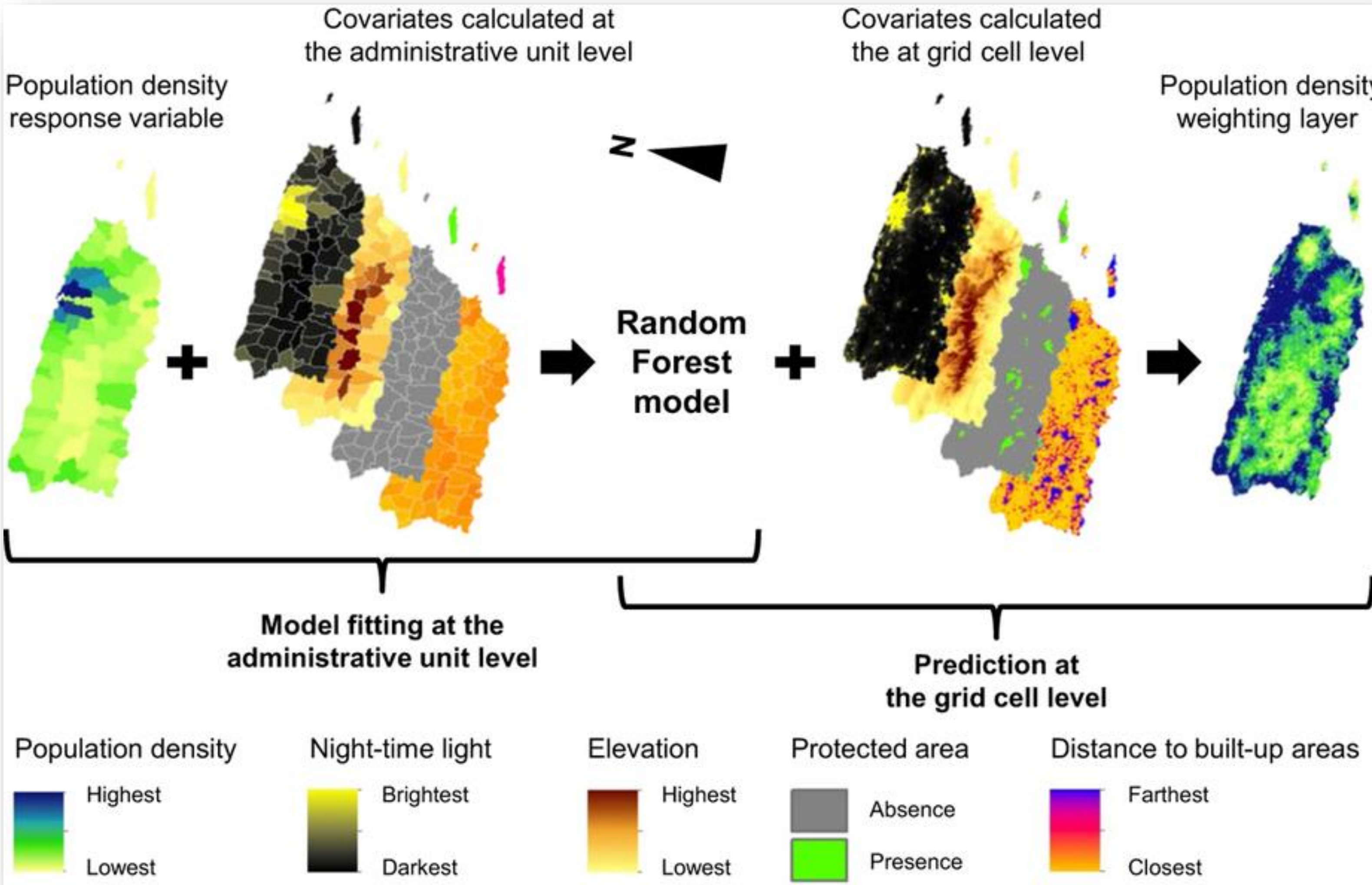
- ‘Machine Learning’ approach
- Create predictions of population density estimates at a pixel level using Random Forest models
- Robust to outliers and noise
- Provides useful internal estimates of error, strength, correlation and variable importance, built-in cross-validation





Breiman (2001) Machine Learning; Stevens et al (2015) PLoS One

City-specific models

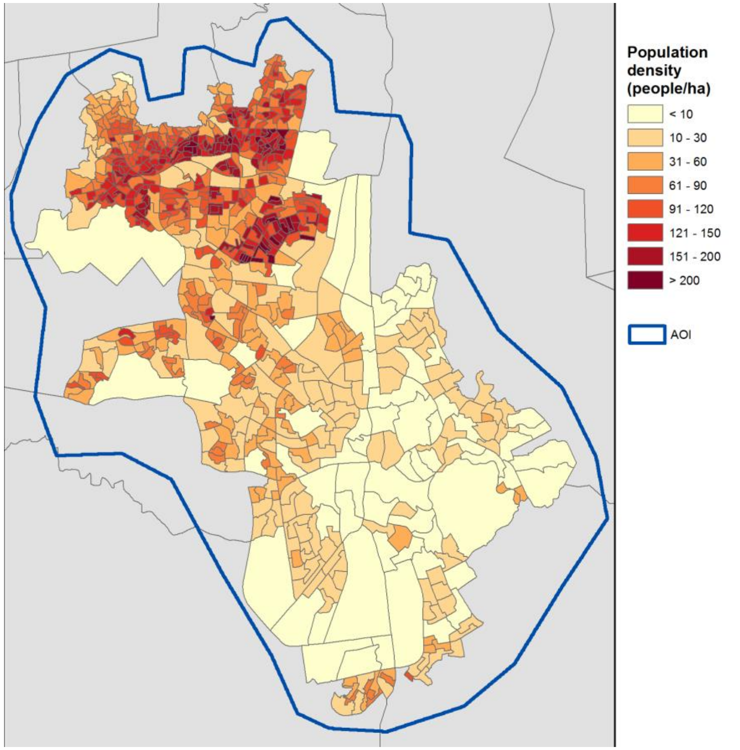


Stevens et al (2016) PLoS ONE; Sorichetta et al (2016) Nature Sci Data

Census population data: select African cities

CITY	COUNTRY	N. census units
Windhoek	NAM	743
Antananarivo	MDG	228
Iringa	TZA	147
Toamasina	MDG	138
Kampala	UGA	116
Mbeya	TZA	113
Nairobi	KEN	106
Dodoma	TZA	54
Ouagadougou	BFA	44
Dakar	SEN	43
Tulear	MDG	42
Tamale	GHA	24
Lusaka	ZMB	24
Kisumu	KEN	19
Kinshasa	COD	17

All others: < 10 (31 cities within only 1 unit)



Windhoek, Namibia

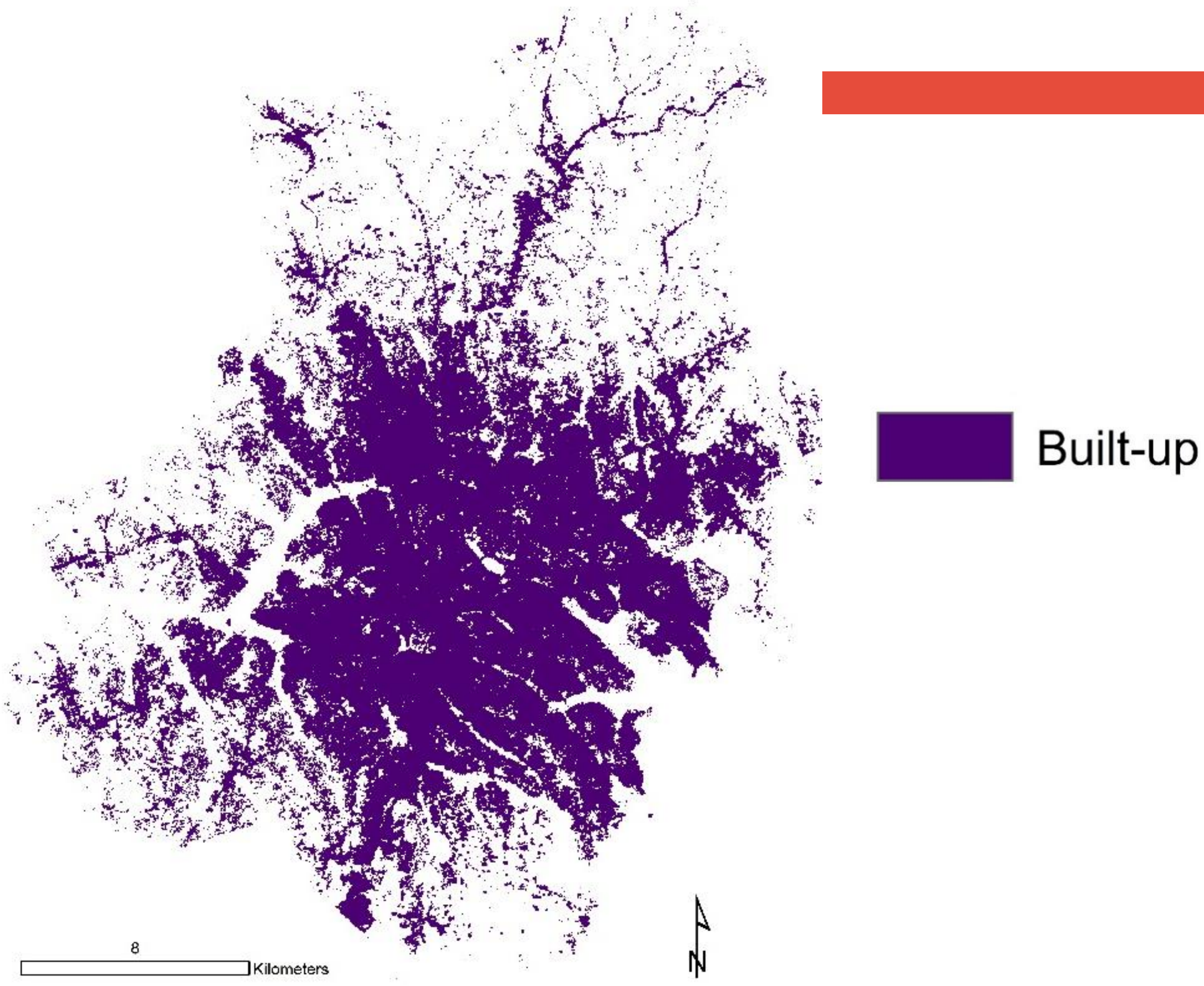
MAUPP Variables description

- **Built-up density**: derived from the binary output of the classification (built-up / non-built-up)
  - Spatial resolution: 12.5m binary raster aggregated to 100m (x8)
  - Values between 0 and 1

- **Built-up age**: Normalized age of built-up density
  - Spatial resolution: 12.5m binary raster aggregated to 100m (x8)
  - Values between 0 (recent) to 1 (old)

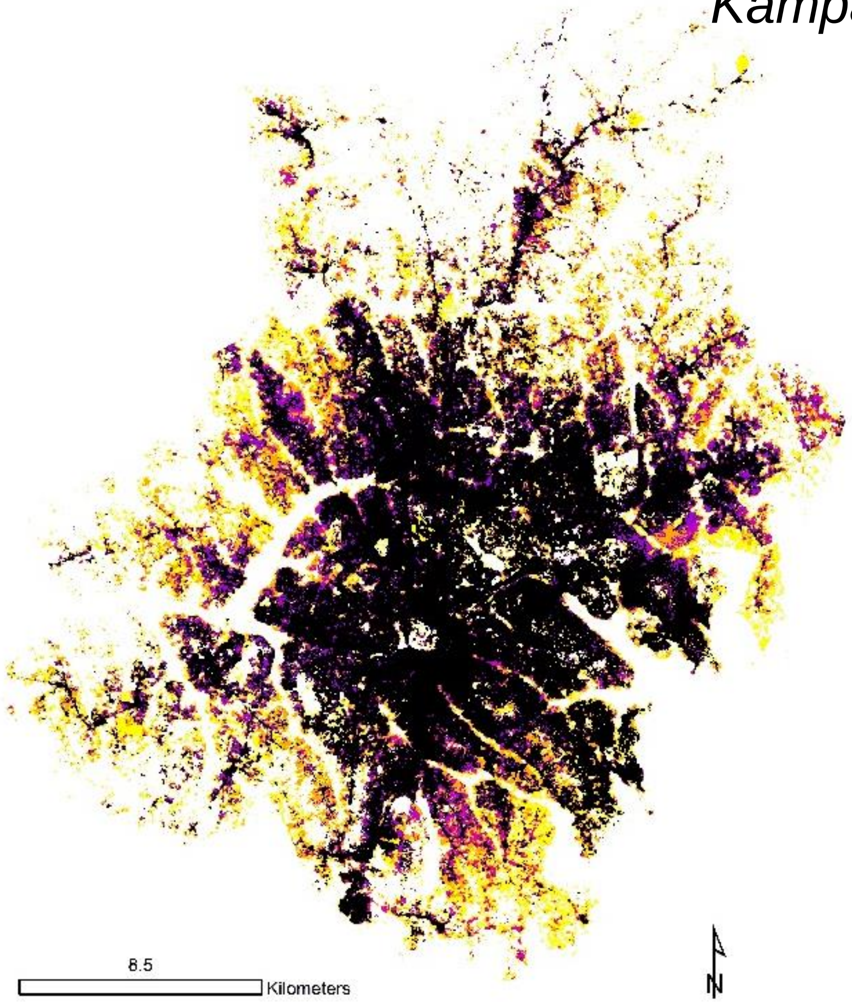
- **Population density**: derived from census data
  - Spatial resolution: census units
  - Units: people/ha

Urban built-up density
Kampala



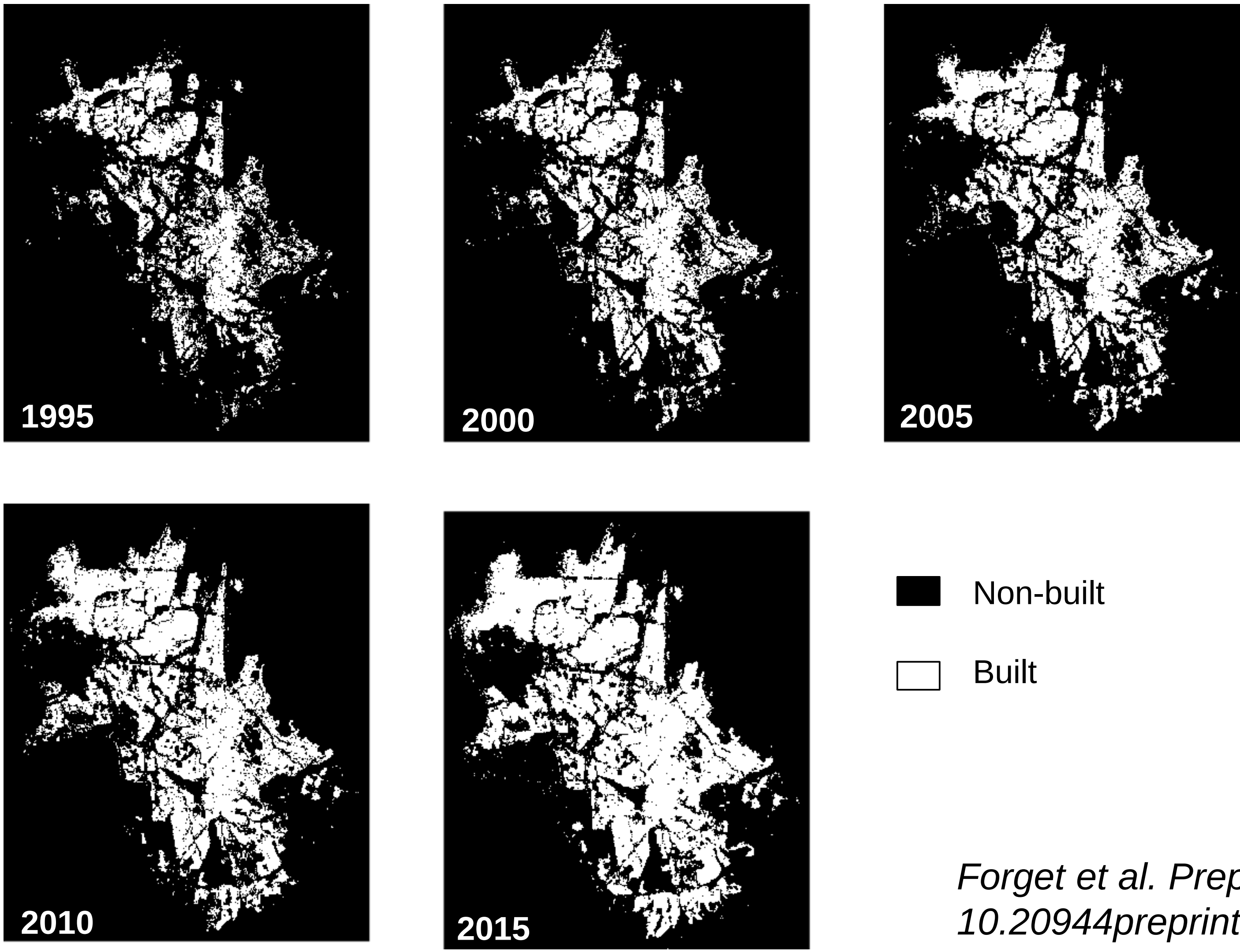
Forget et al. Preprints 2018,(doi: 10.20944preprints201810.0695.v1)

Normalized age of built-up
From 0 (past) to 1 (present)
Kampala



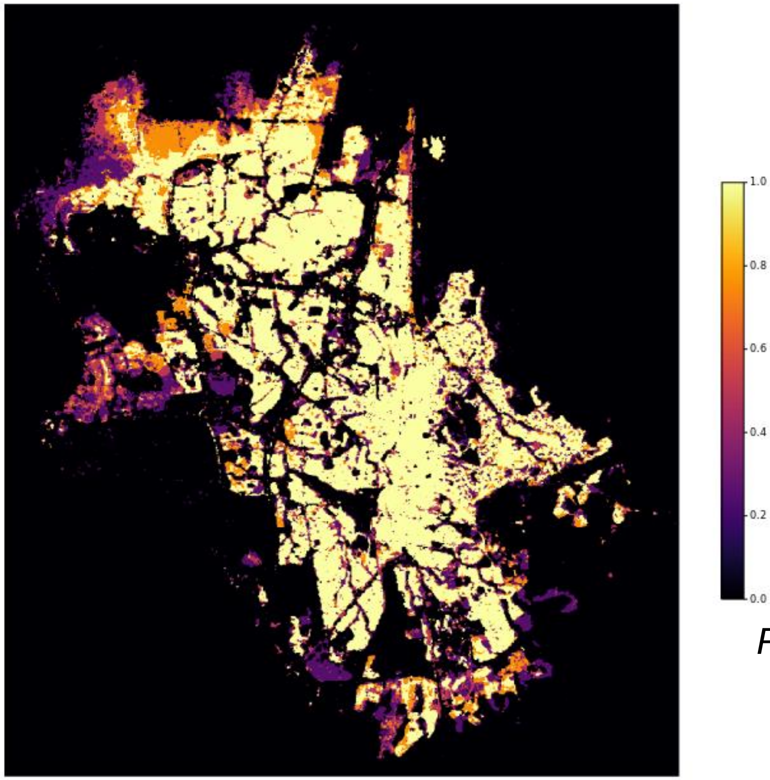
Forget et al. Preprints 2018,(doi: 10.20944preprints201810.0695.v1)

Urban change (Windhoek)



Forget et al. Preprints 2018,(doi: 10.20944preprints201810.0695.v1)

Urban change (Windhoek)



Age of built-up
From 0 (recent) to 1 (old)

Forget et al. Preprints 2018,(doi: 10.20944preprints201810.0695

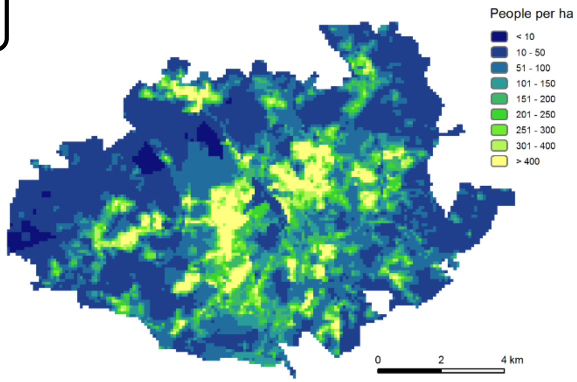
Final Model Data List		Source
Built-up density	t	MAUPP
	t-1	
	t-2	
Built-up age	Normalized age of built-up density	MAUPP
Landcover	Cultivated terrestrial lands	ESA-CCI
	Woody / Trees	
	Shrubs	
	Herbaceous	
	Other terrestrial vegetation	
	Aquatic vegetation	
	Water bodies	
	No data, cloud/shadow	
	Industrial area	
Nightlights		Suomi NPP-VIIRS, NOAA
Topography	Elevation	HydroSheds (3 s GRID: Void-filled DEM), WWF
	Slope	
Roads (distance-to)	Primary	OpenStreetMap
	Secondary	
	Others	
Railways (distance-to)		OpenStreetMap
Rivers (distance-to)		OpenStreetMap
Natural areas		OpenStreetMap
Industrial landcover class (distance-to)		OpenStreetMap
Points of interest		OpenStreetMap

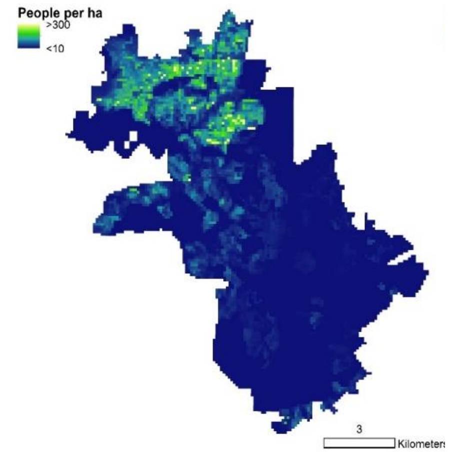
Results ++

Antananarivo

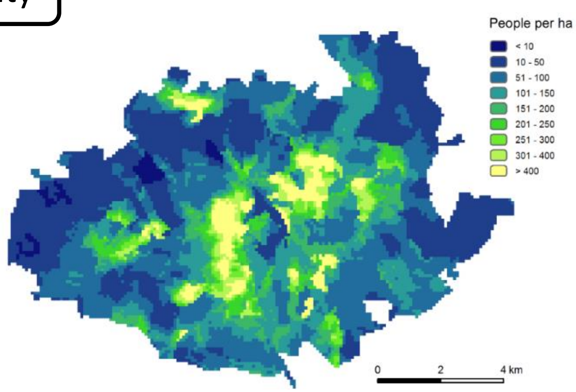
Windhoek

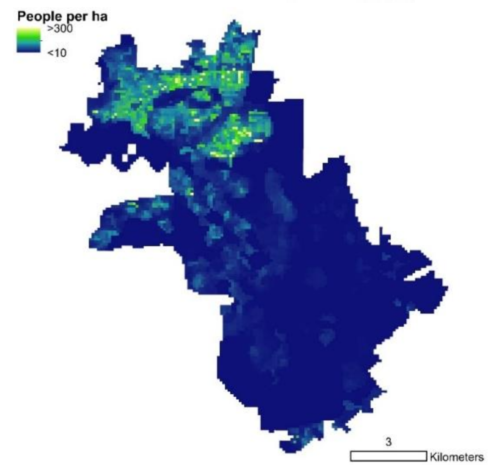
With built-up density





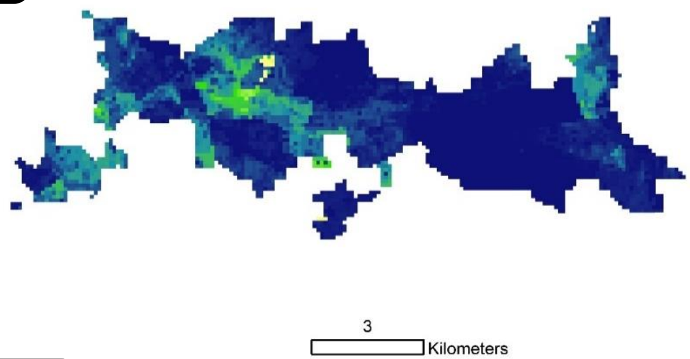
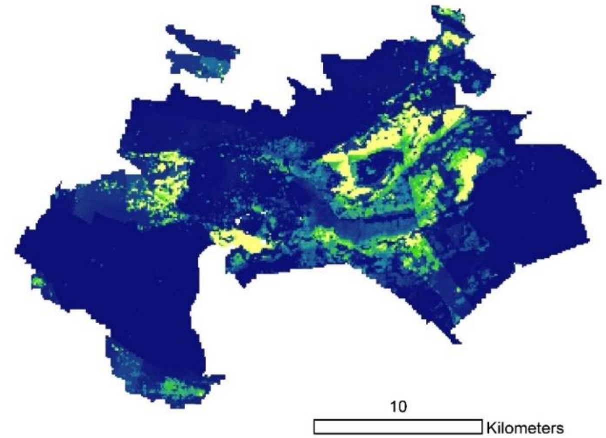
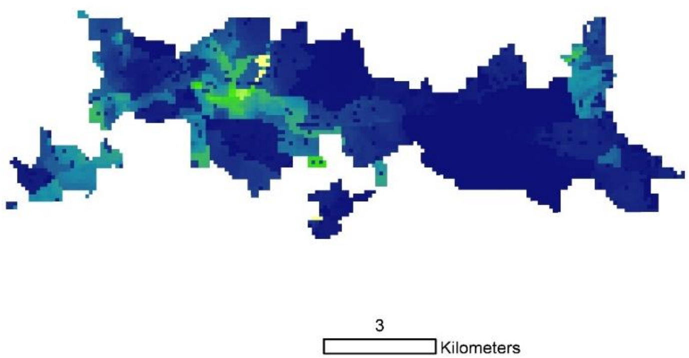
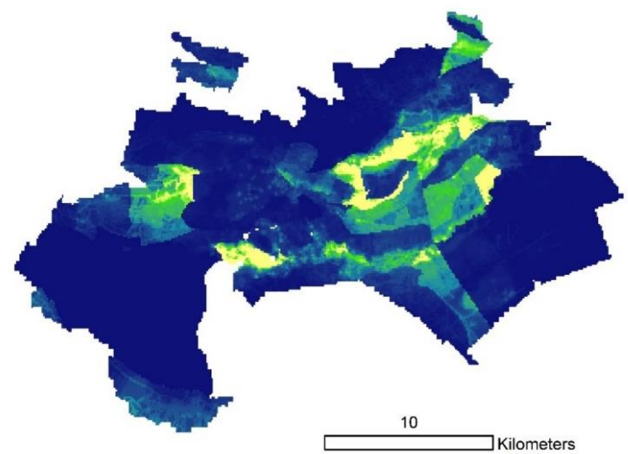
Without built-up density



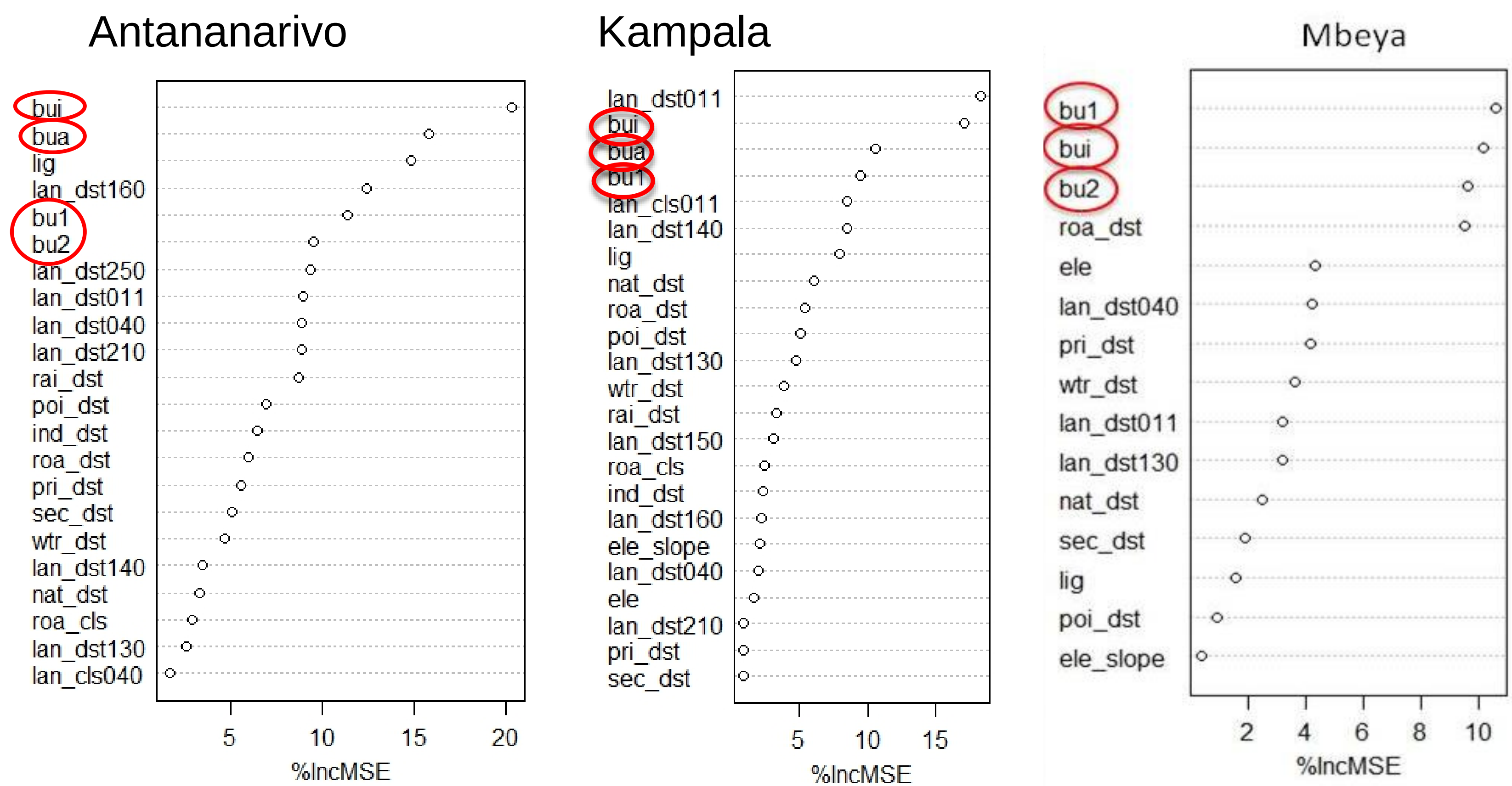


Results ++

MAUPP

	Mbeya	Nairobi
With built-up density		
Without built-up density		

Covariate importance

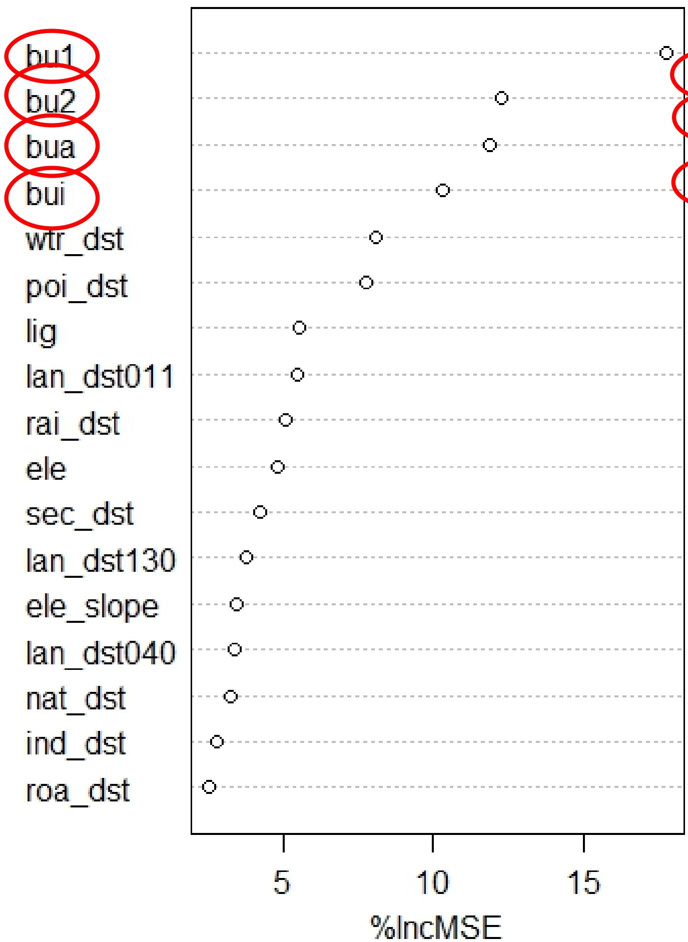


Covariate importance plots for the models including built-up density layers for Antananarivo, Kampala and Mbeya. Built-up density layers are highlighted in red (bui = built-up density for time t - time period closest to the census data -; bu1 = built-up density for t-1; bu2 = built-up density for t-2; bua = normalized built-up age)

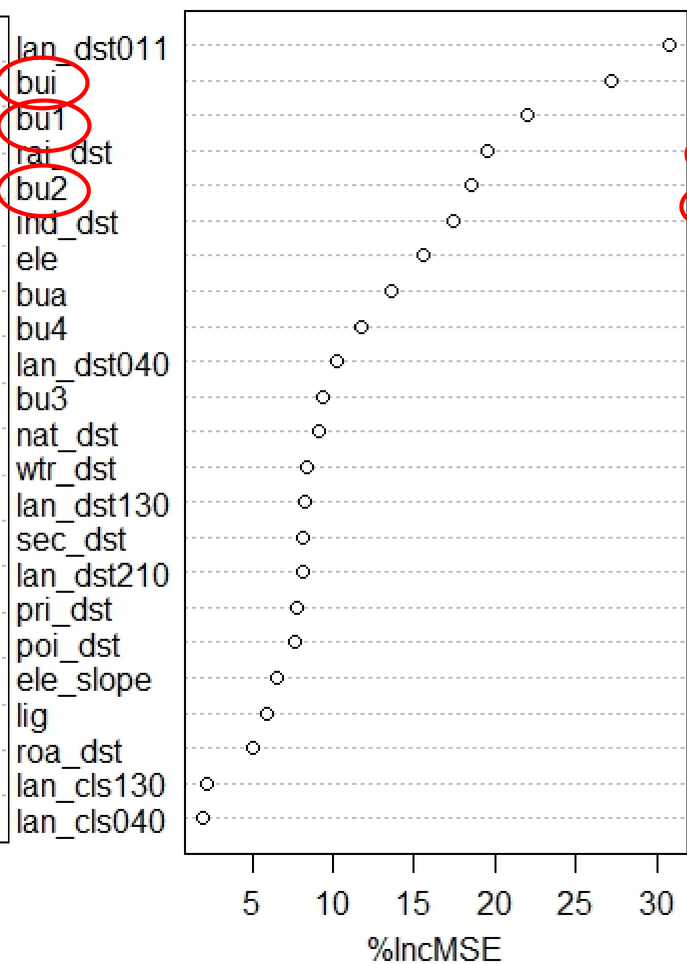
Covariate importance (continued)

MAUPP

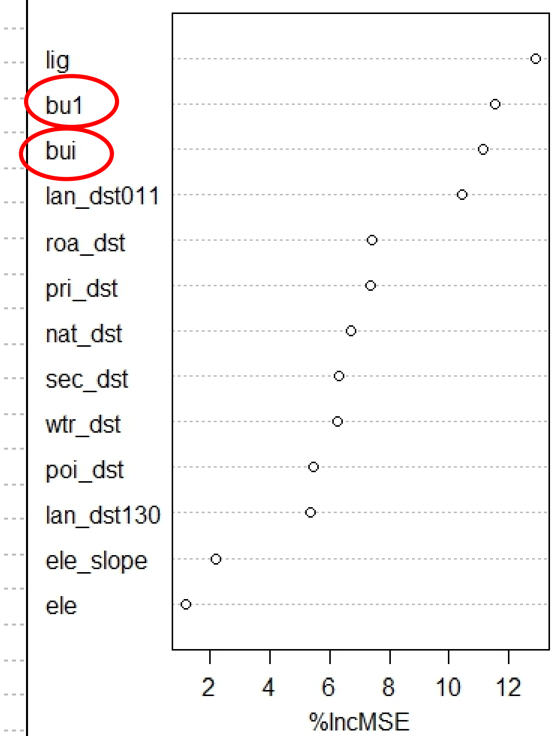
Nairobi



Windhoek



Iringa



Intra-urban population maps

City	Country ID	N. admin. units	% explained variance with built-up layers	% explained variance without built-up layers	
Windhoek	NAM	743	80	69	++
Antananarivo	MDG	228	64	56	++
Iringa	TZA	147	42	43	-
Toamasina	MDG	138	70	70	
Kampala	UGA	116	67	60	+
Mbeya	TZA	113	38	25	++
Nairobi	KEN	106	62	45	++

% Explained variance for population models with and without built-up density layers for the 7 cities with more than 100 administrative units.

Accuracy Assessment

MAUPP

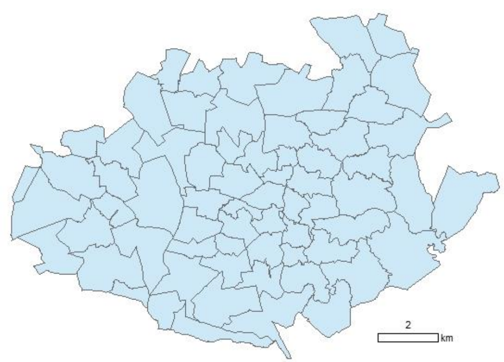
Aggregate census blocks to obtain one third of blocks → Aggregate covariates by census blocks and estimate RF model → Predict per-pixel population density with RF algorithm → Redistribute census block totals weighted by predictions

↓

Aggregate per-pixel population to finer level admin units and compare to original census counts

Antananarivo

Census blocks (N=228)



Validation blocks (N=68)

Accuracy Assessment statistics

MAUPP

City	Country ID	N. admin. units	N. admin units for validation	RF With built-up layers		RF Without built-up layers		
				%RMSE	MAE	%RMSE	MAE	
Windhoek	NAM	743	230	43.3	129	51.7	150	++
Mbeya	TZA	113	34	56.4	798	69.3	937	++
Nairobi	KEN	106	32	73.9	13761	93.2	17490	++
Antananarivo	MDG	228	68	42.9	1851	49.4	2125	++
Iringa	TZA	147	44	56.5	231	52.4	219	-
Toamasina	MDG	138	41	34.3	484	33.4	471	-
Kampala	UGA	116	35	35.4	3375	39.7	3687	+

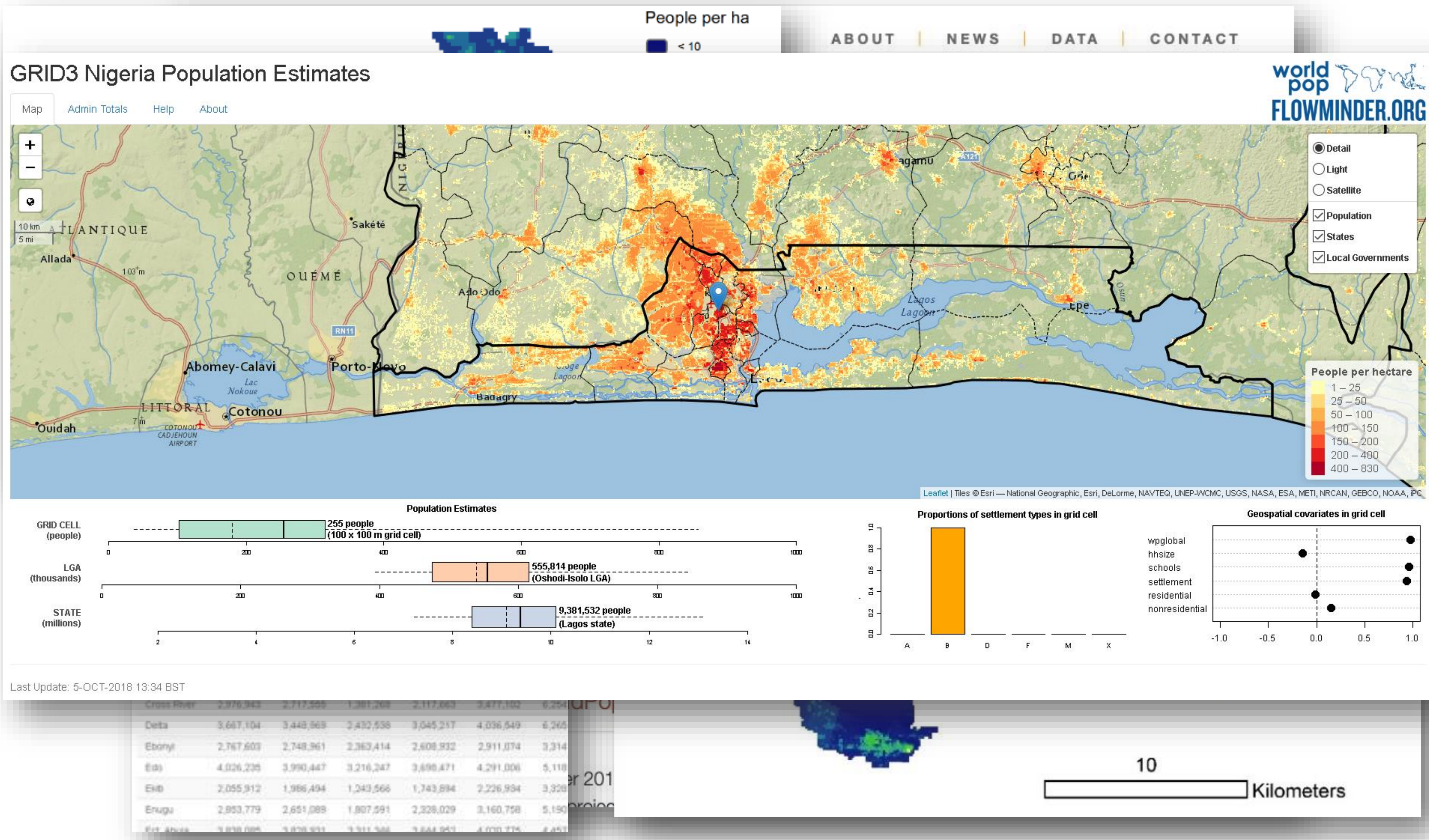
Summary

MAUPP

- Satellite imagery can be processed to map intraurban heterogeneities in African cities
- Including built-up density layers in urban population models produces clear improvements accuracy of model outputs
- Many useful freely available datasets exist for mapping populations and their characteristics
  - Proper statistical methodologies essential
  - Clear communication of accuracy assessment and limitations
- Next steps: Extending to all other MAUPP cities, documentation, open tools, open outputs, training

Data availability: www.worldpop.org

MAUPP



Further information

A.J.Tatem@soton.ac.uk

maupp.ulb.ac.be

www.worldpop.org

@WorldPopProject

Under the funding of BELSPO – the Belgian Science Policy