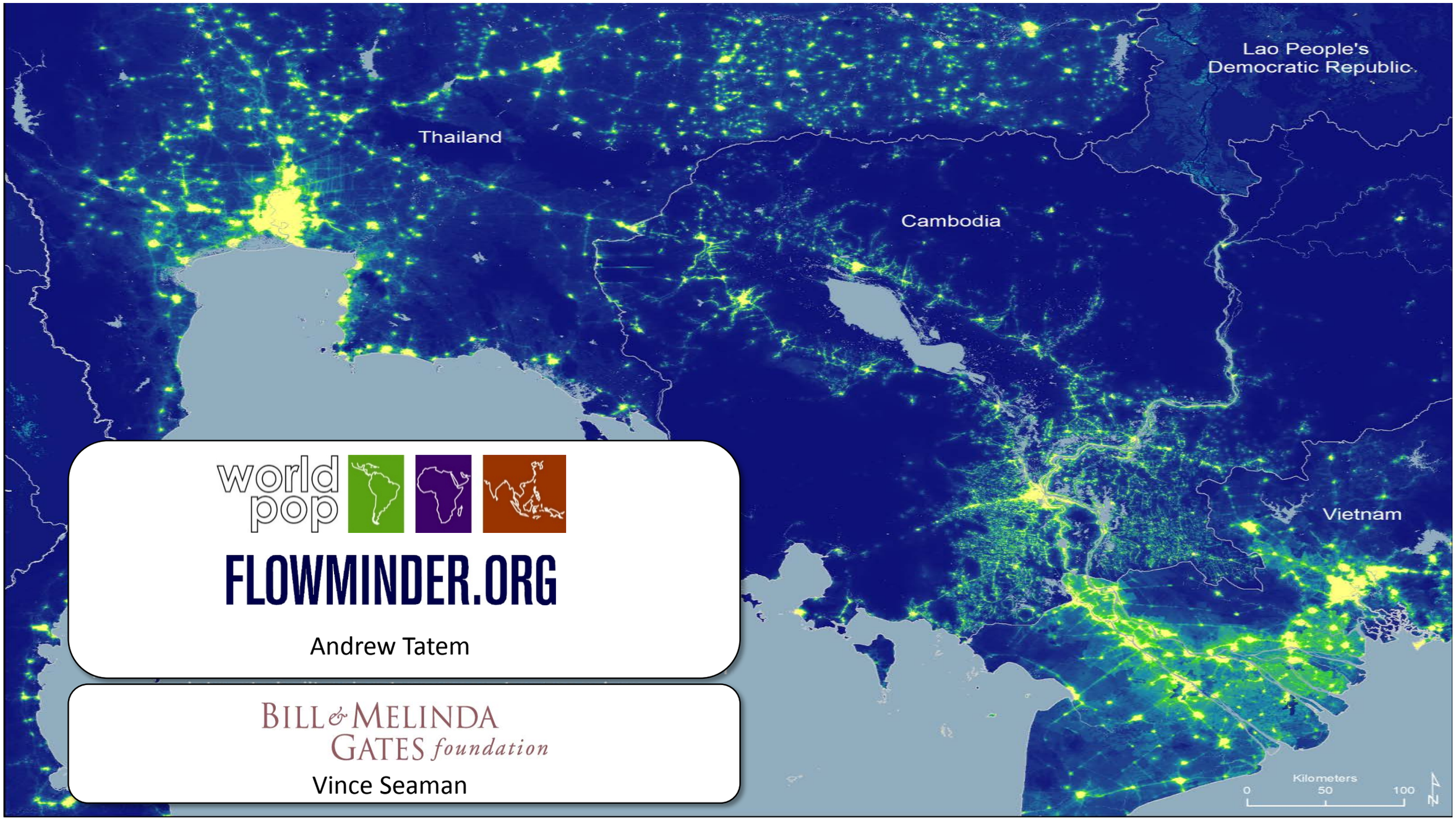



world pop

# FLOWMINDER.ORG

Andrew Tatem

BILL & MELINDA GATES foundation

Vince Seaman

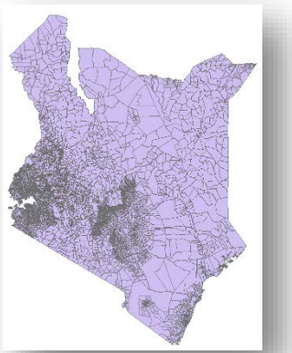
# Development goal denominators

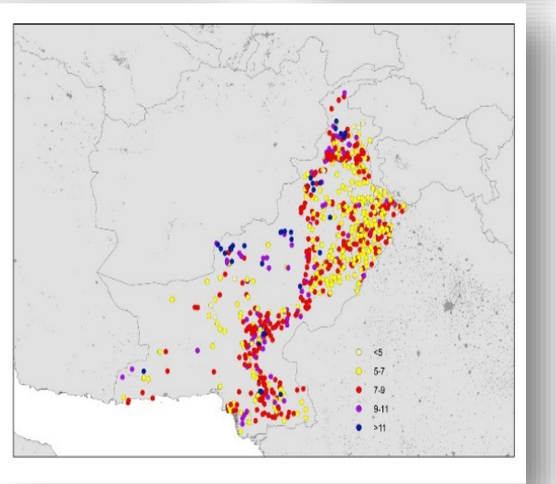
- All MDGs and SDGs are based on ensuring a certain percentage of the population has access to specific services or resources, or achieves a certain level of social, economic, or physical health.
- Improved understanding of *sub-national geographic variation and inequity* in health status, wealth, and access to resources within countries is increasingly recognized as central to meeting development goals.
- Requires a consistent, comparable and regularly updated understanding of not only how many people live in a country, but where the people are, and who they are.

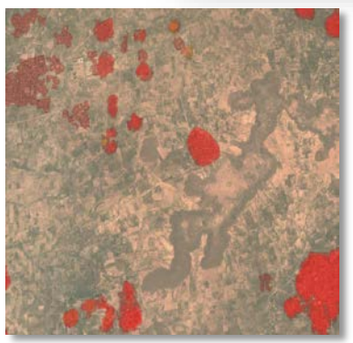
# WorldPop

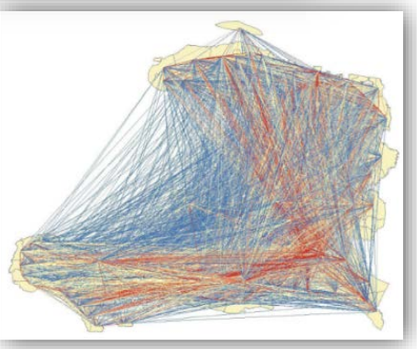
***Improve the spatial demographic evidence base for low and middle income countries***

- Work with statistical agencies, health depts to obtain finest detail, boundary-matched census data
- Develop scalable methods and models for integrating ancillary datasources to complement and fill data gaps in census
- Integrate new technologies
- Freely available high resolution data and fully documented, peer-reviewed methods





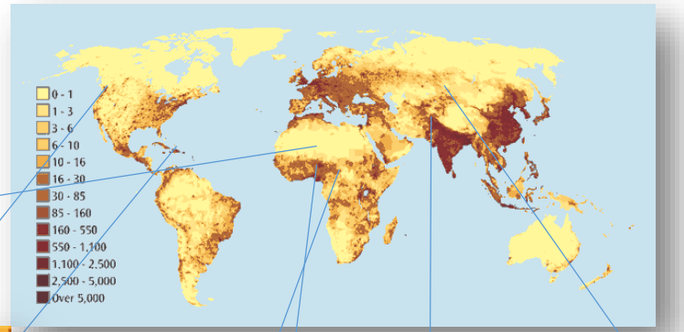



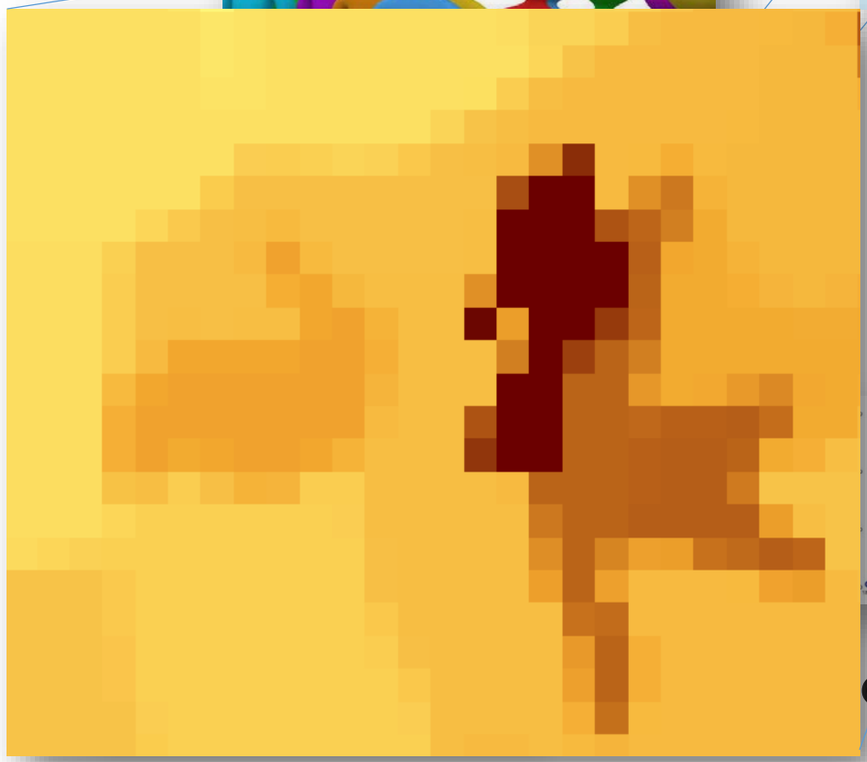


www.worldpop.org

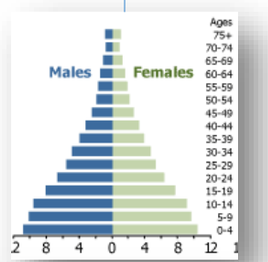
# What do we mean by spatial demographics here?

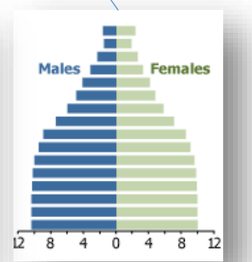
- Where is everybody?

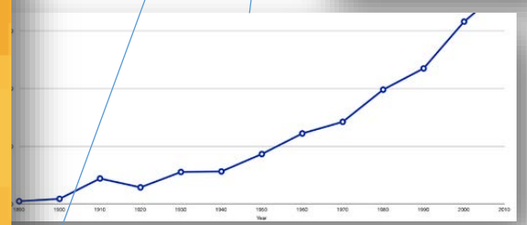


- Who is there?



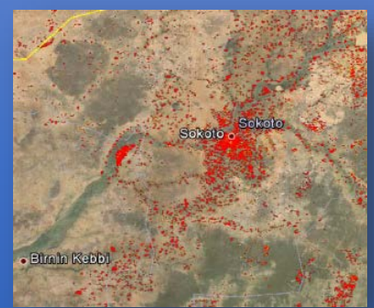



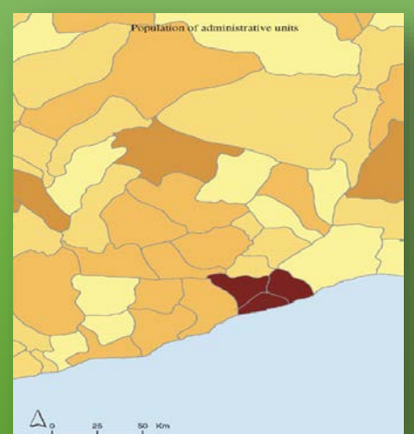
obile is the population?

# Modelling population distributions

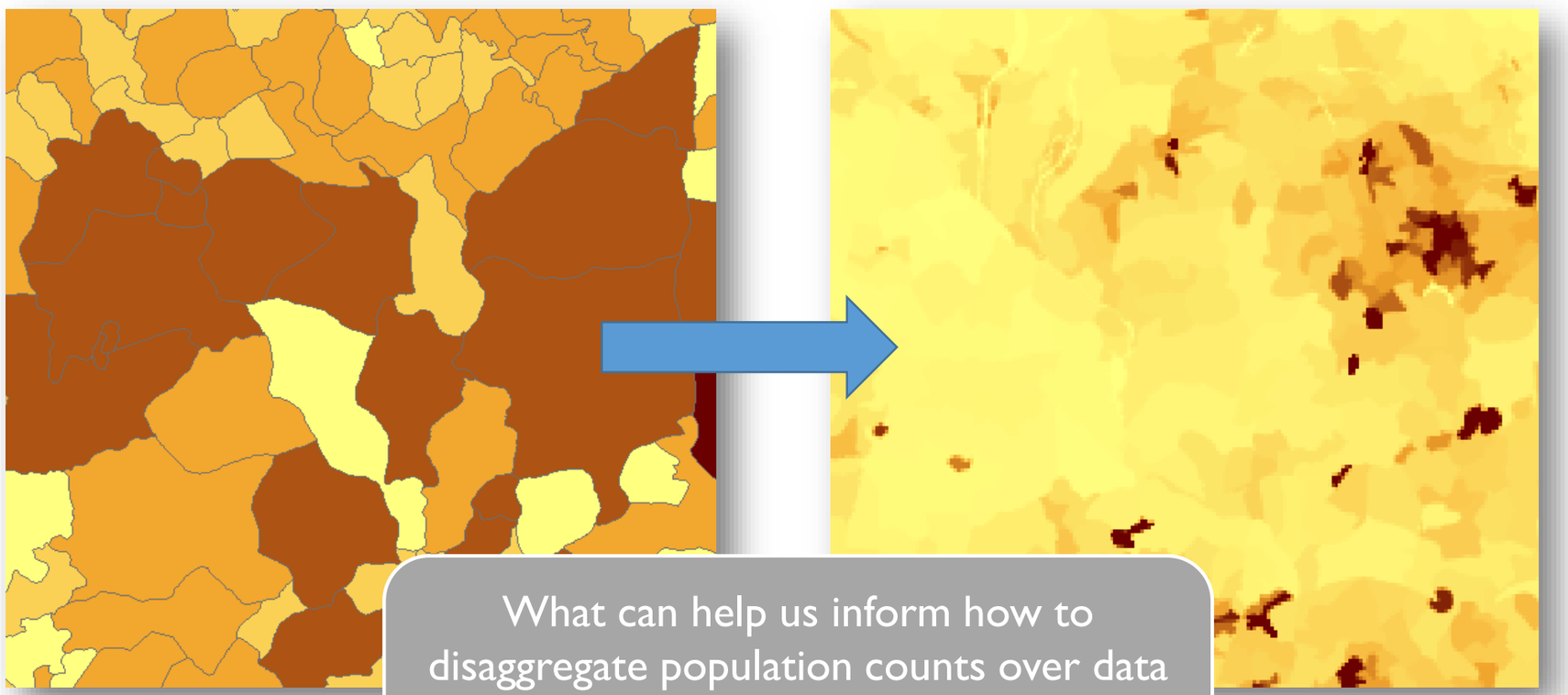
***Bottom-up:*** where census data are outdated/unreliable, integration of high resolution settlement mapping with survey data




***Top-down:*** disaggregation of admin-unit based census/official estimate counts

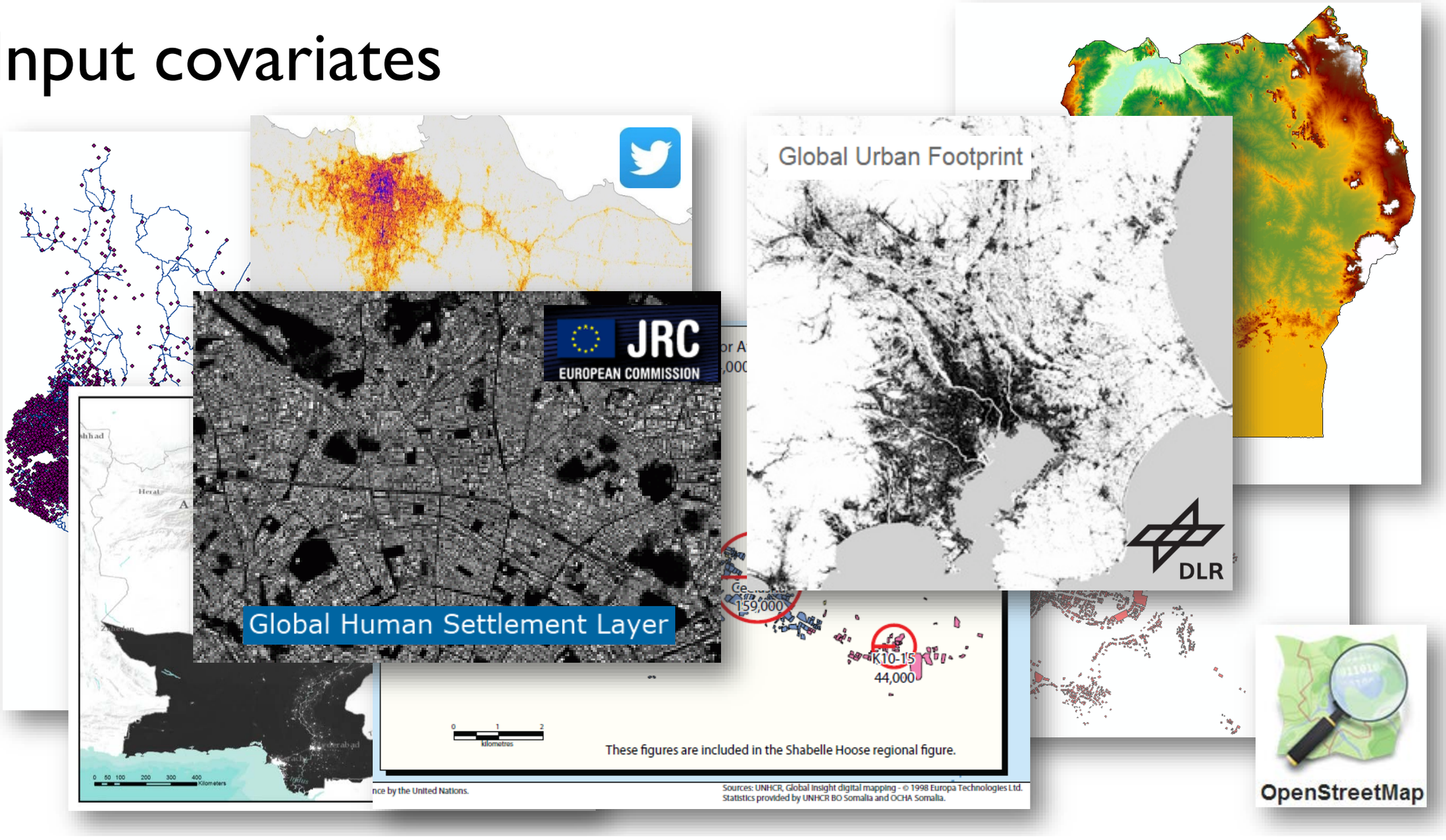


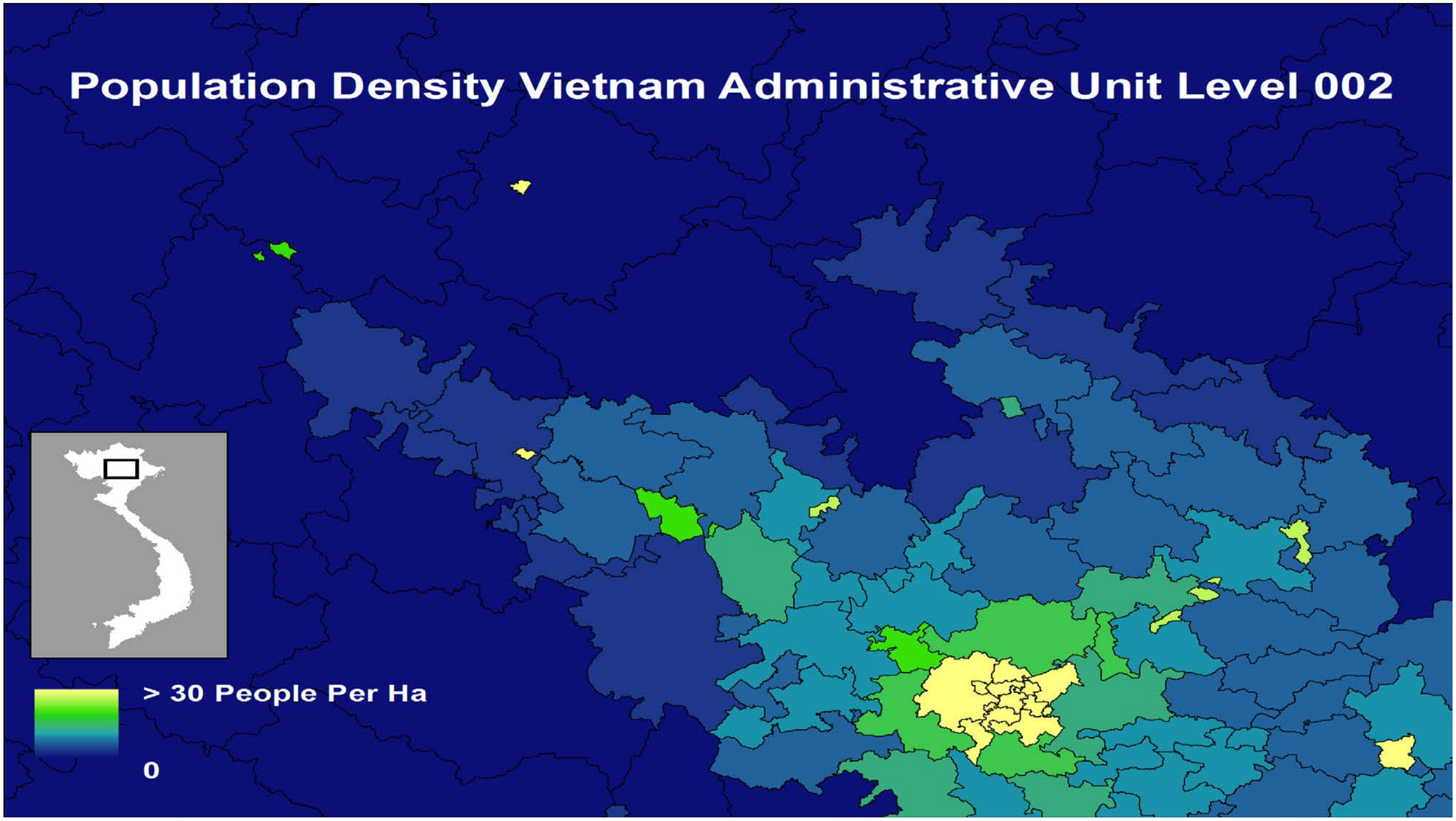
# Top-down: How can we go from aggregated counts to gridded surfaces?

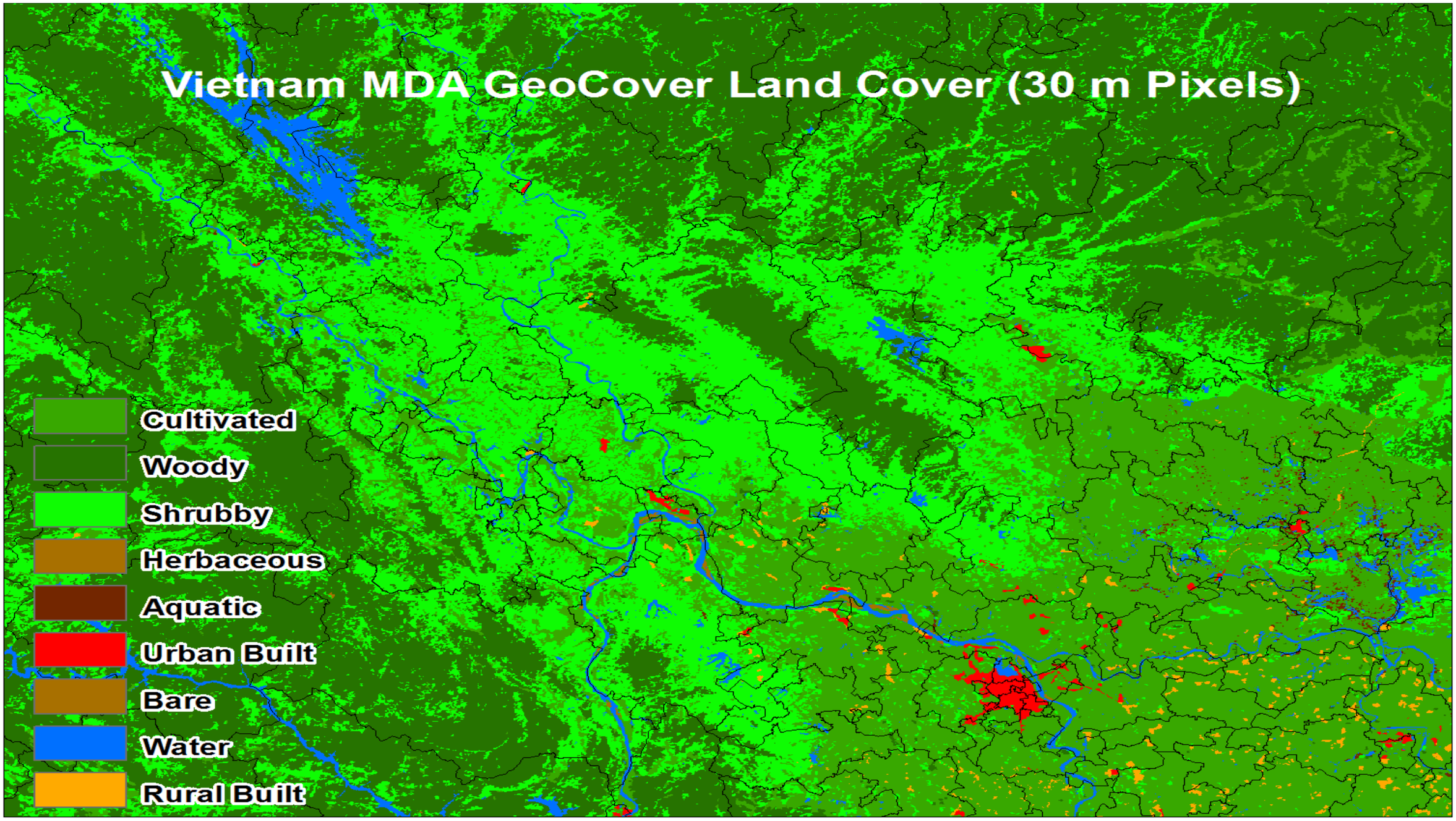


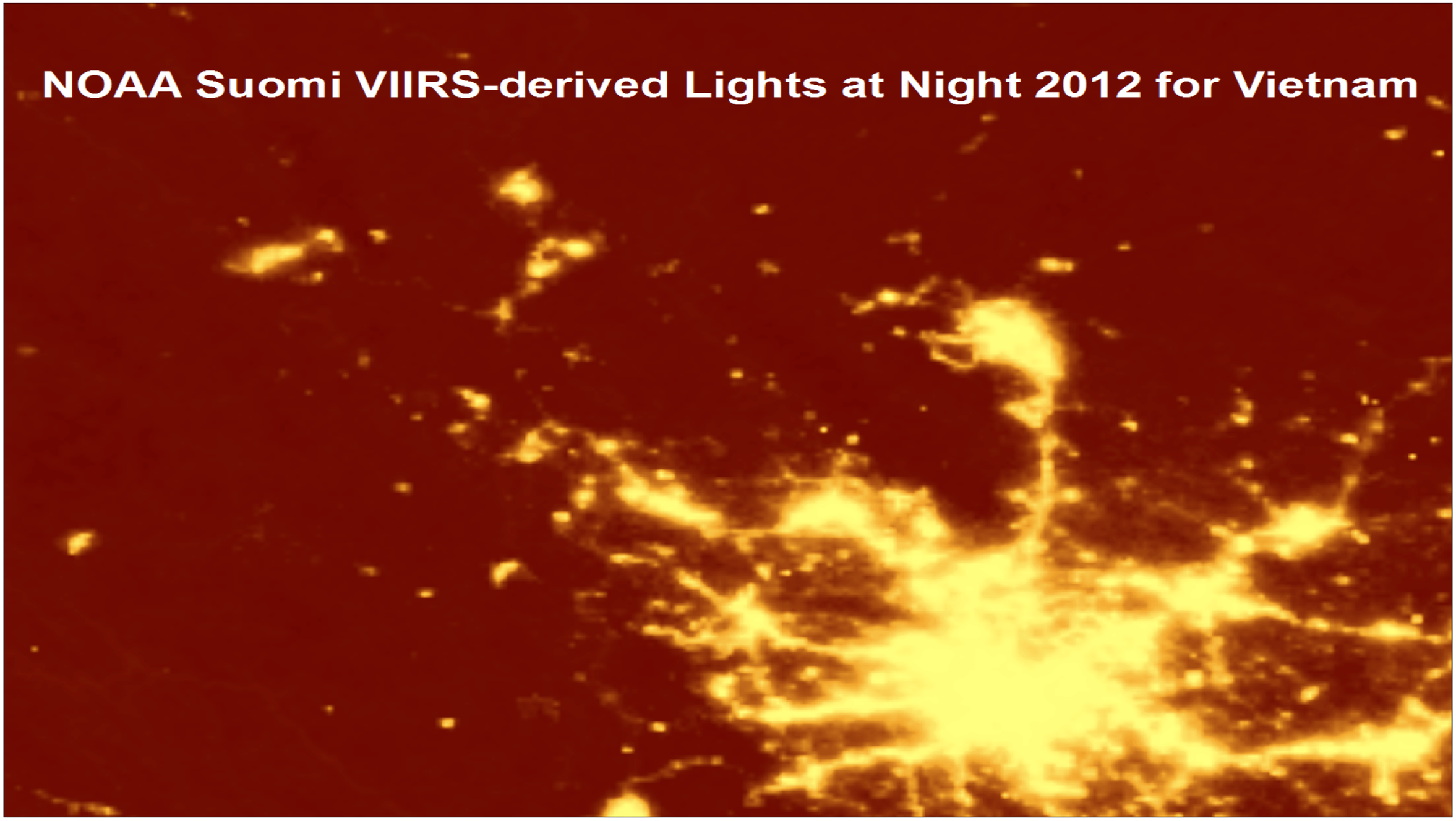
What can help us inform how to disaggregate population counts over data poor national/continental areas into small grid cells accurately?

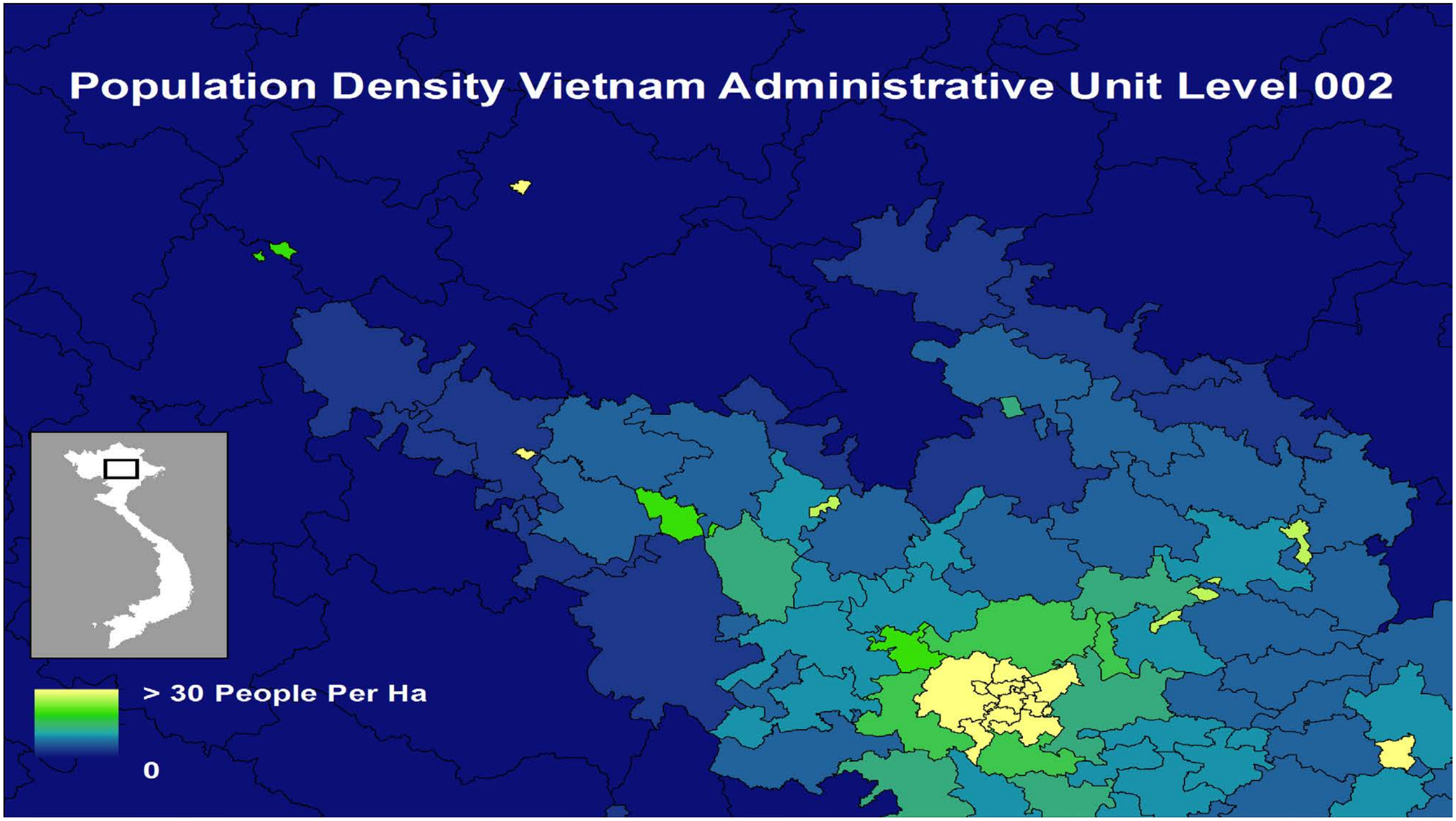
# Input covariates

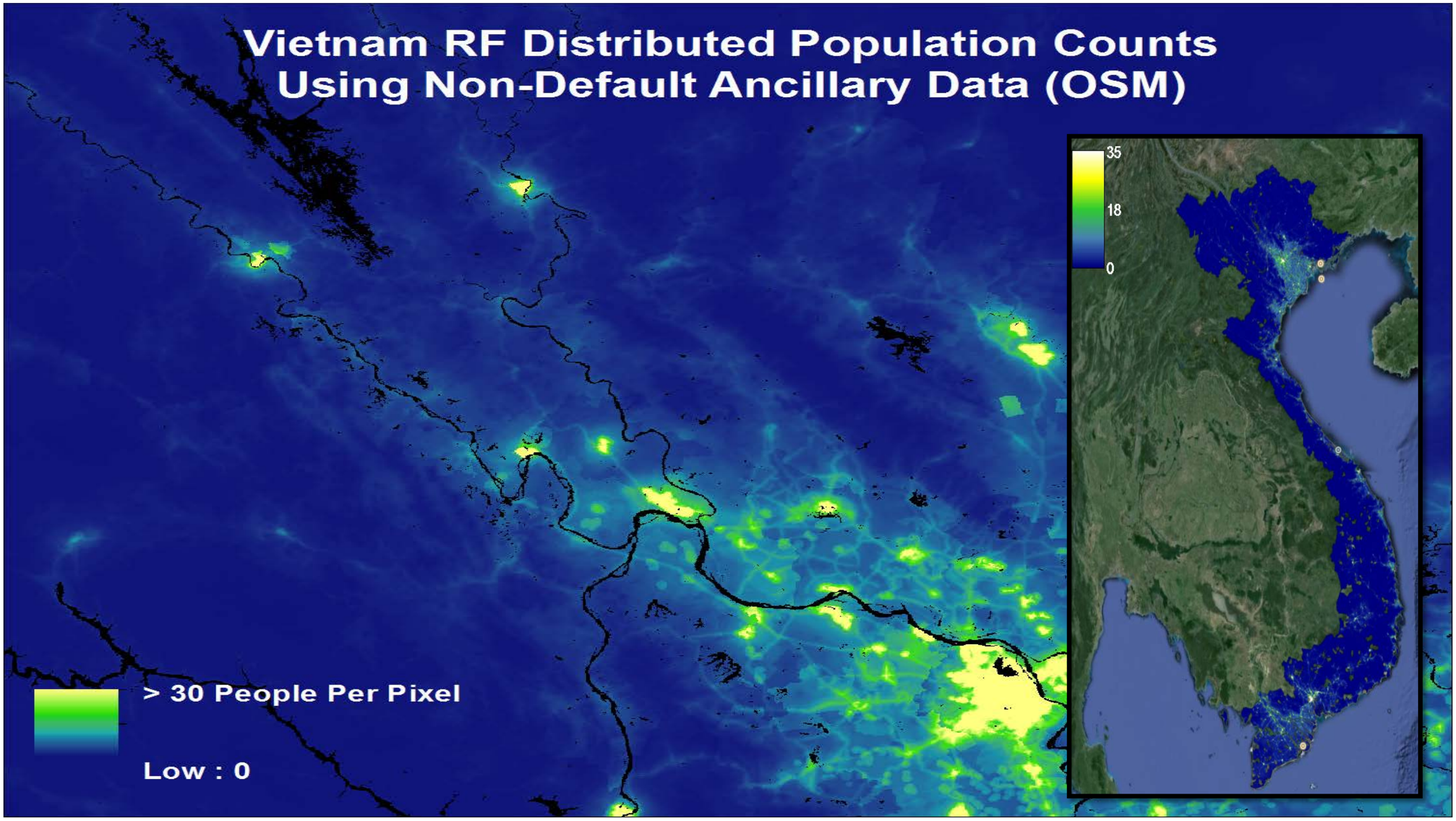



Population Density Vietnam Administrative Unit Level 002
> 30 People Per Ha
0

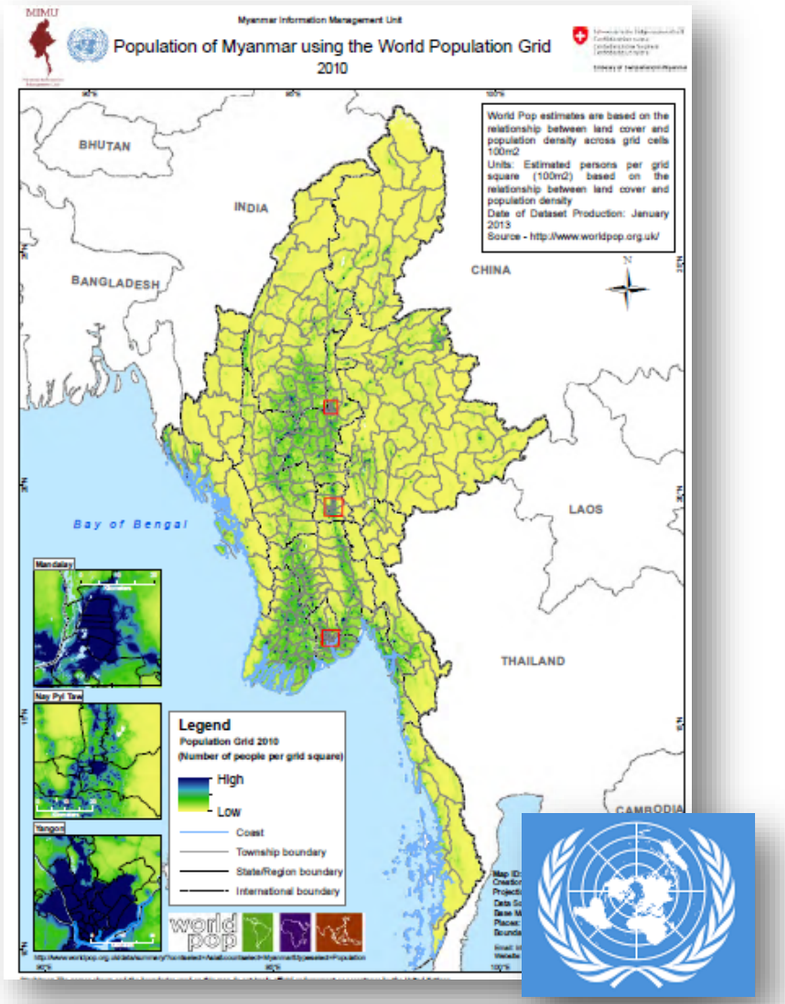

Vietnam MDA GeoCover Land Cover (30 m Pixels)
Cultivated
Woody
Shrubby
Herbaceous
Aquatic
Urban Built
Bare
Water
Rural Built

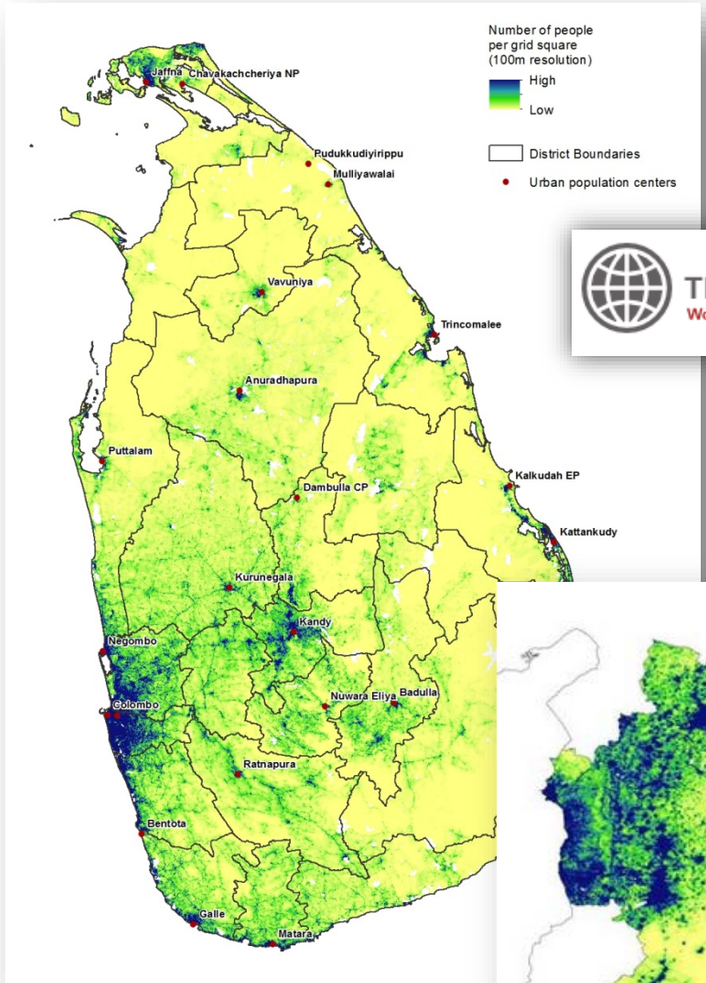

NOAA Suomi VIIRS-derived Lights at Night 2012 for Vietnam


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

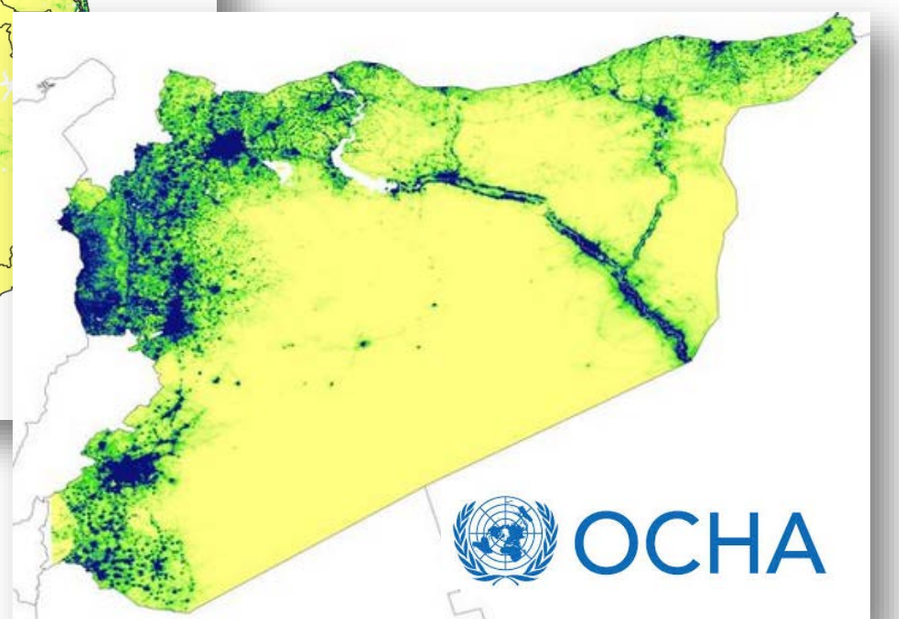

Vietnam RF Distributed Population Counts
Using Non-Default Ancillary Data (OSM)
35
18
0
> 30 People Per Pixel
Low : 0

www.worldpop.org



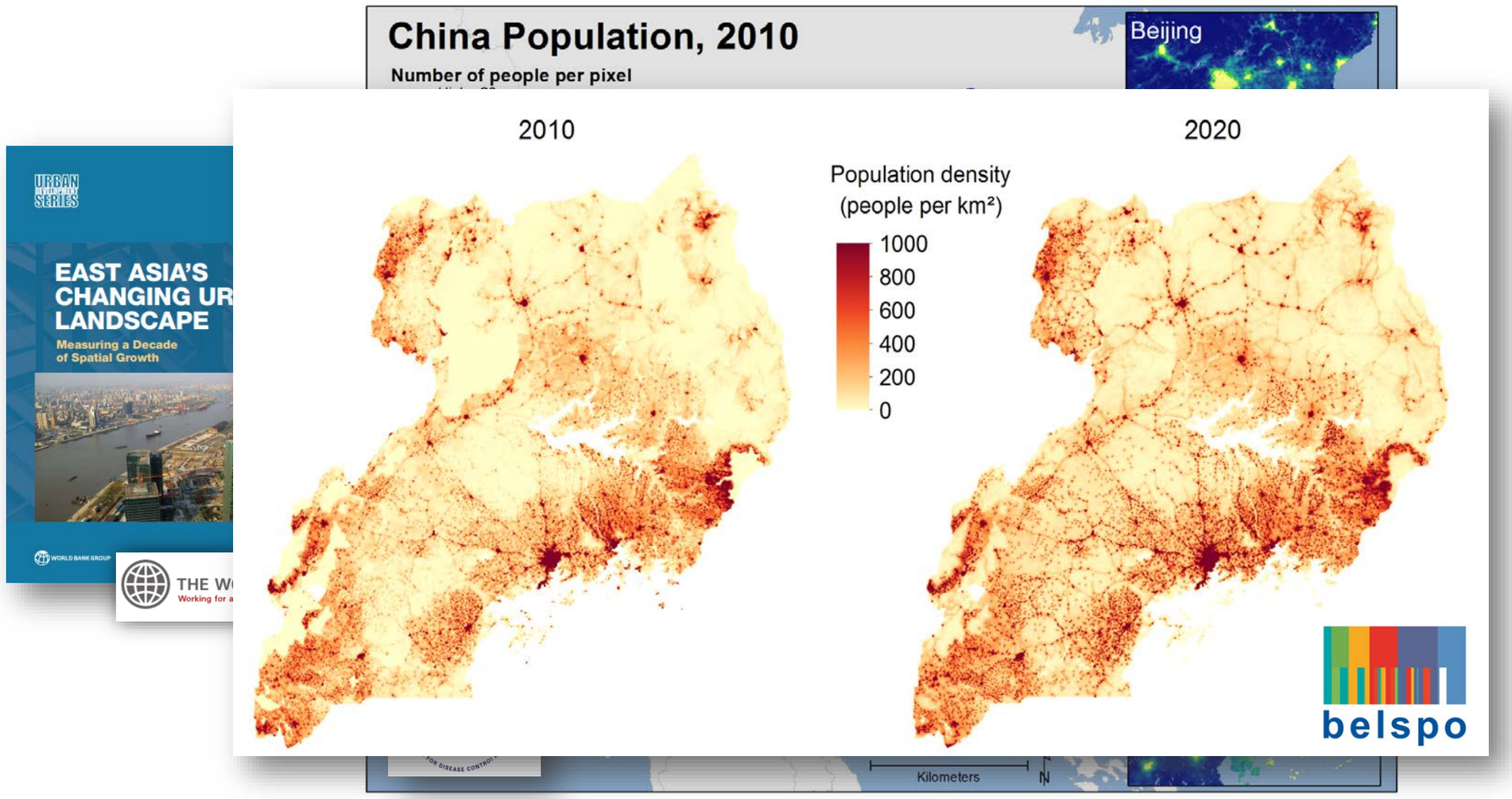




THE WORLD BANK
Working for a World Free of Poverty



OCHA

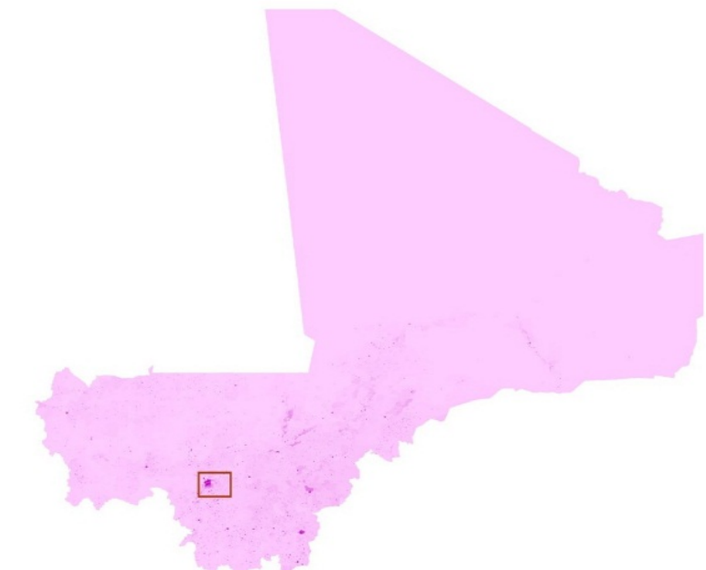
# Temporal change



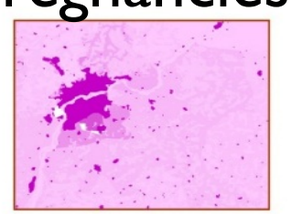

*Gaughan et al (2015) under review; Linard et al (2013) Applied Geography; Schneider et al (2015) Env Res Letters*

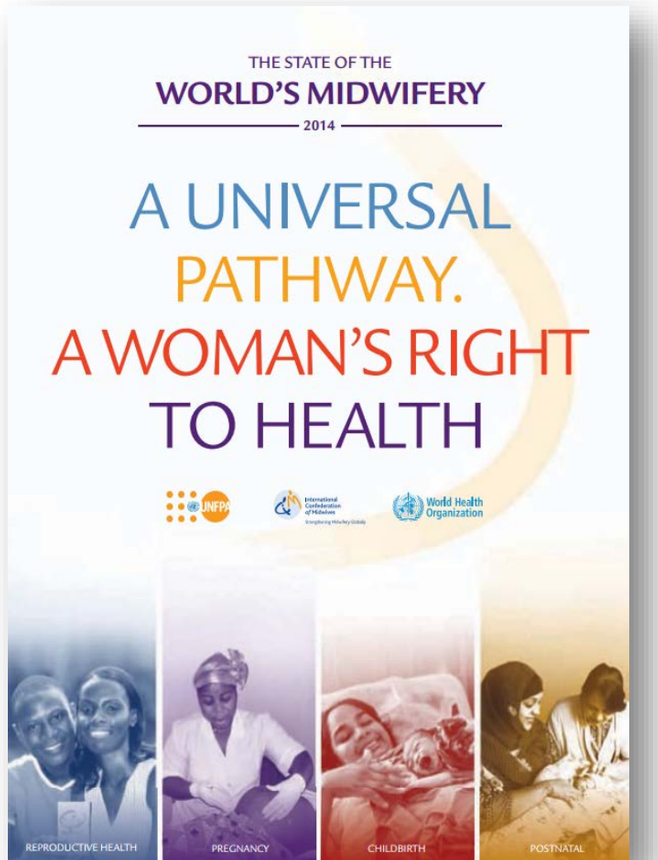
# Data integration: Maternal and newborn health

Mali



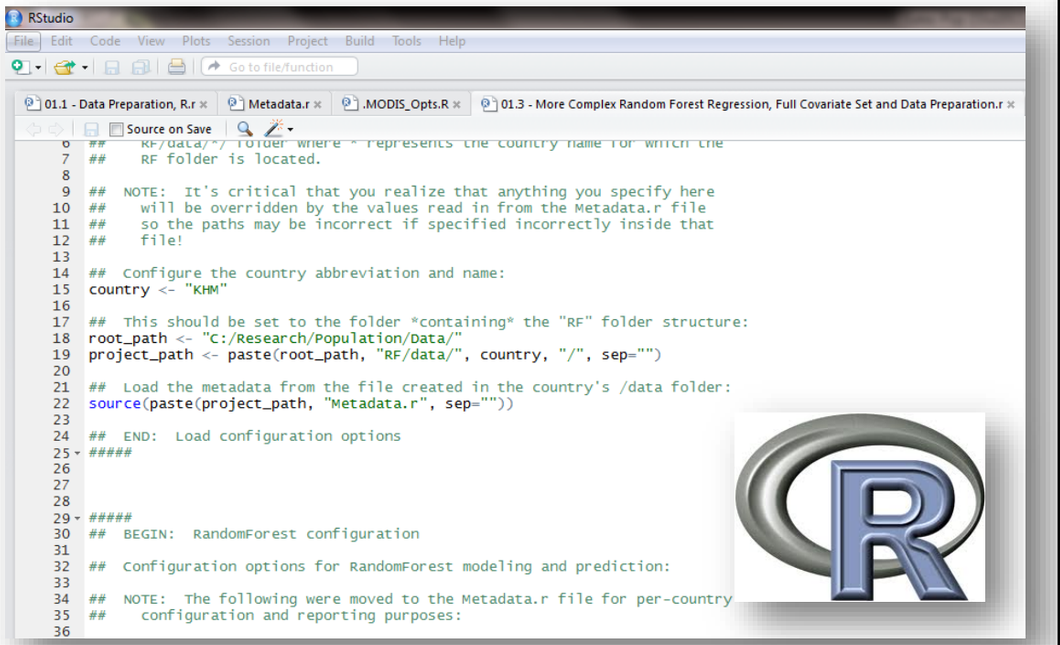

Pregnancies

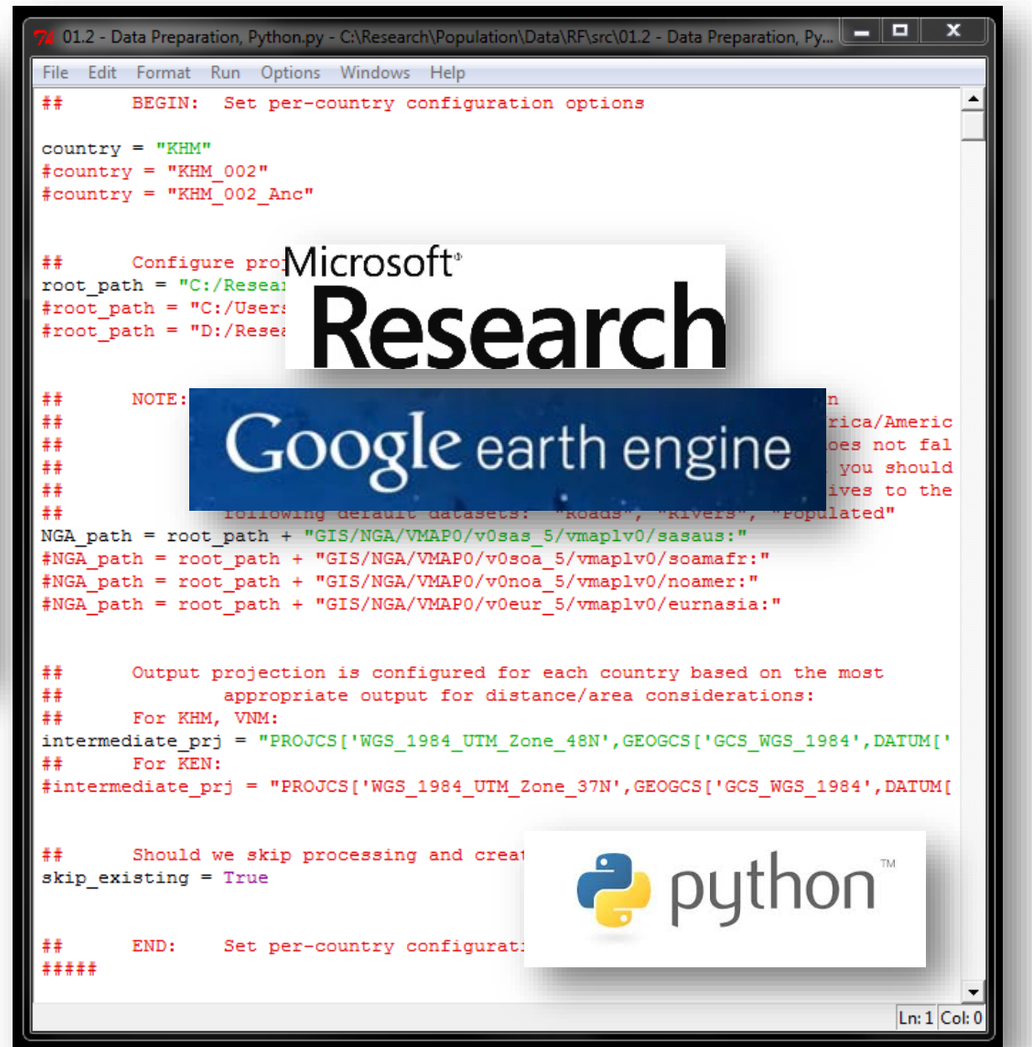





*Tatem et al (2014) IJHG, Ebener et al (2015) IJHG*

# Open code

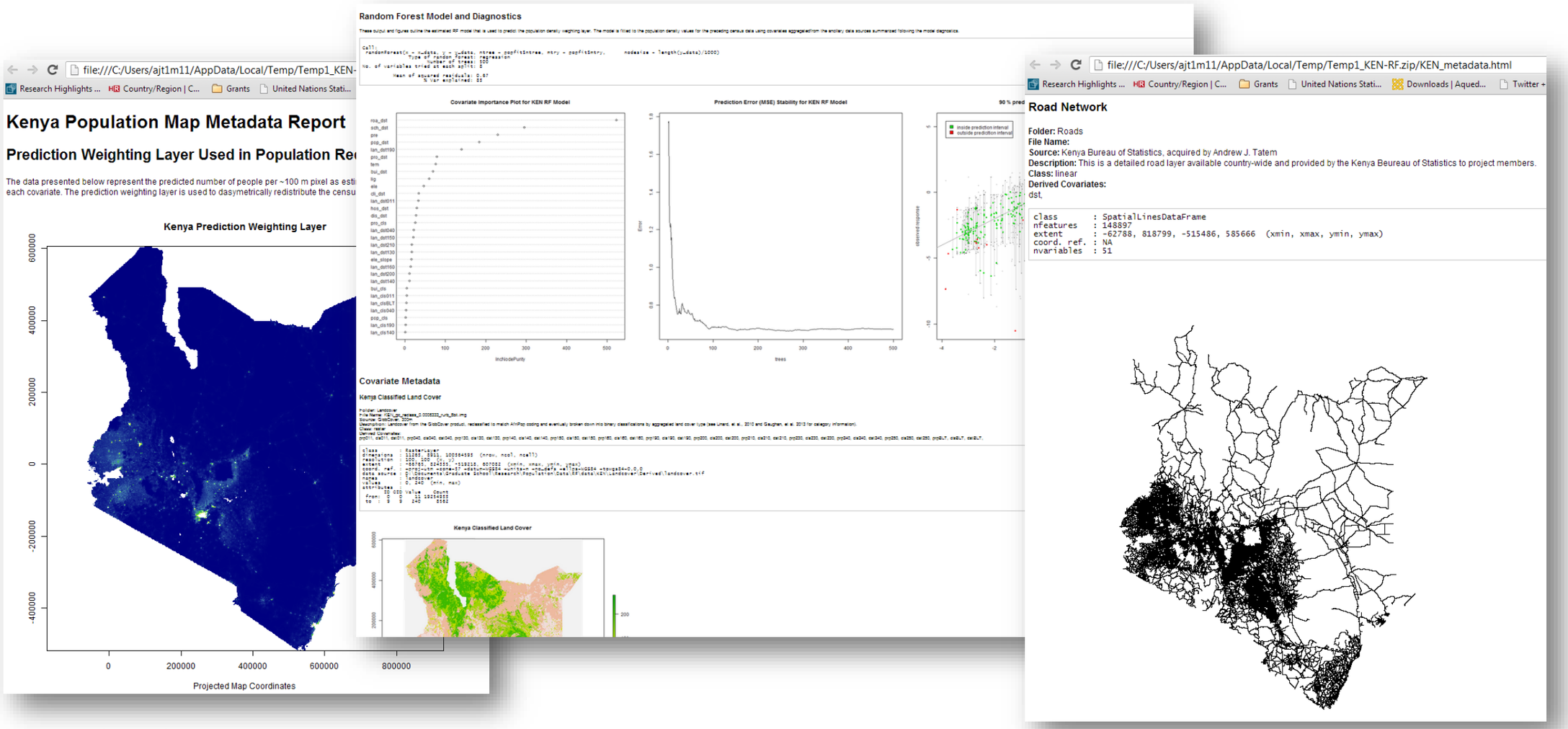



- Feedback & improvements
- Methodological clarity
- User-specific datasets
- Platform independent

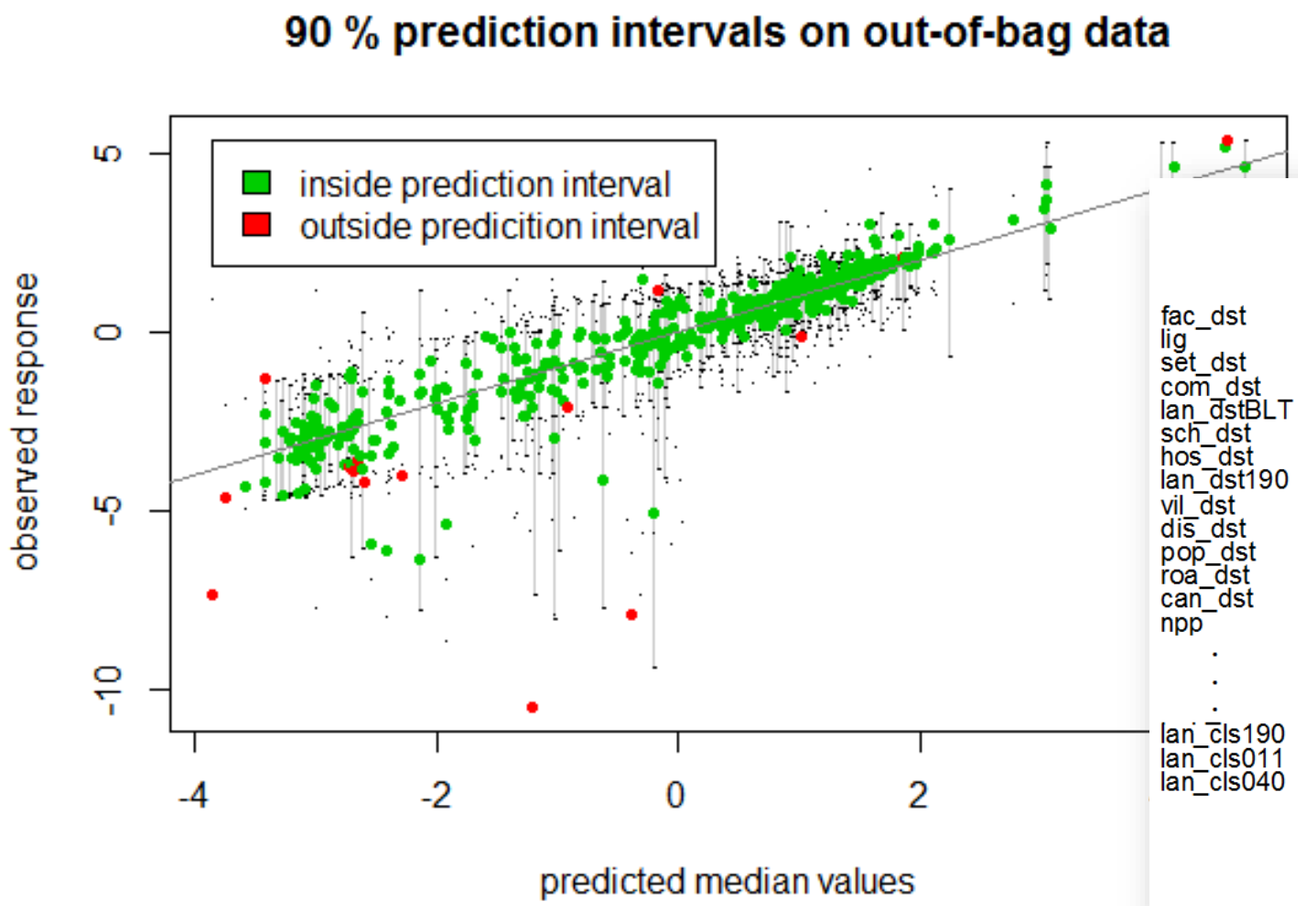
*Stevens et al (2015) PLoS ONE*

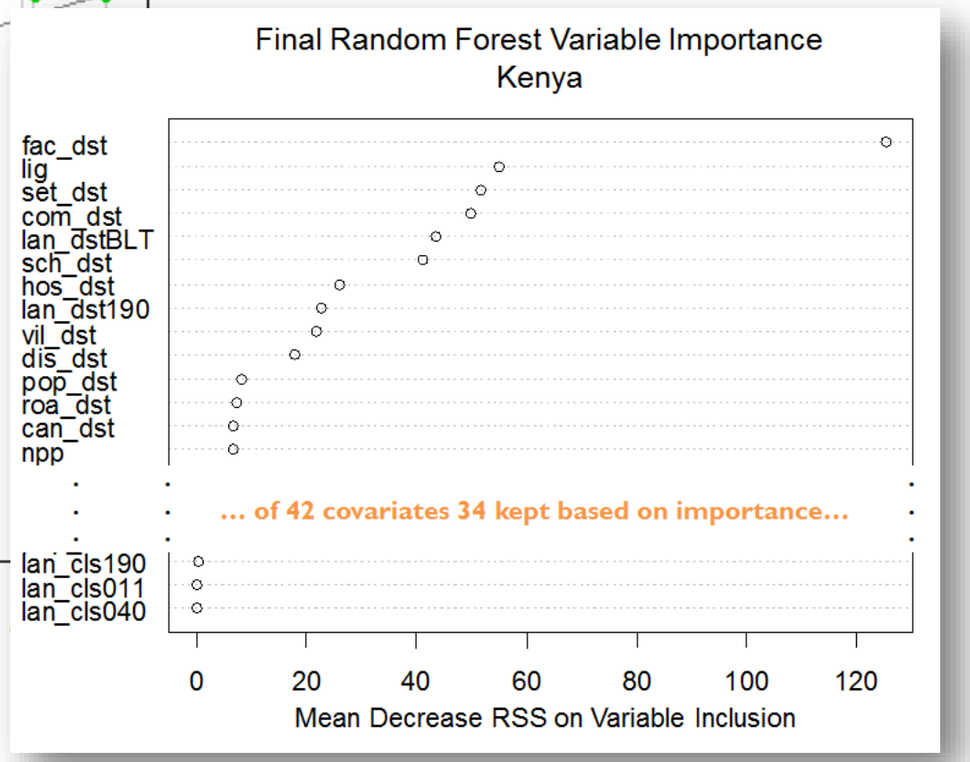
# Automated metadata




*Stevens et al (2015) PLoS ONE*

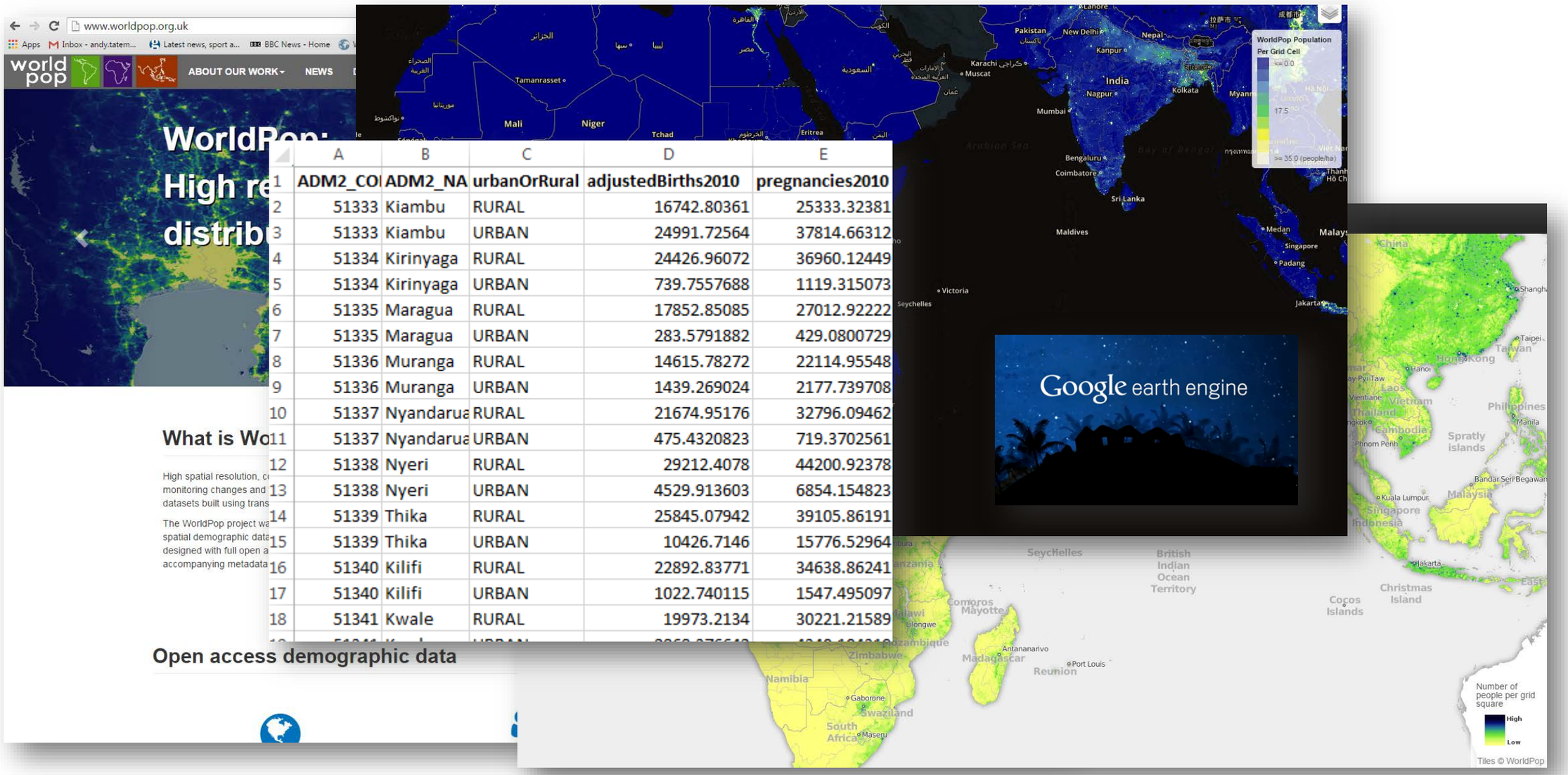
# Internal and external accuracy assessments




*Stevens et al (2015) PLoS ONE*

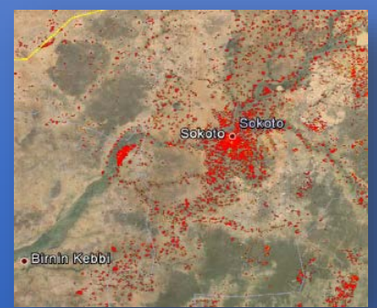
# www.worldpop.org



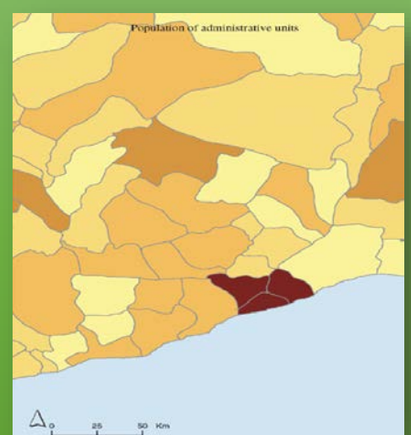

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | ADM2_CO | ADM2_NA | urbanOrRural | adjustedBirths2010 | pregnancies2010 |
| 2 | 51333 | Kiambu | RURAL | 16742.80361 | 25333.32381 |
| 3 | 51333 | Kiambu | URBAN | 24991.72564 | 37814.66312 |
| 4 | 51334 | Kirinyaga | RURAL | 24426.96072 | 36960.12449 |
| 5 | 51334 | Kirinyaga | URBAN | 739.7557688 | 1119.315073 |
| 6 | 51335 | Maragua | RURAL | 17852.85085 | 27012.92222 |
| 7 | 51335 | Maragua | URBAN | 283.5791882 | 429.0800729 |
| 8 | 51336 | Muranga | RURAL | 14615.78272 | 22114.95548 |
| 9 | 51336 | Muranga | URBAN | 1439.269024 | 2177.739708 |
| 10 | 51337 | Nyandarua | RURAL | 21674.95176 | 32796.09462 |
| 11 | 51337 | Nyandarua | URBAN | 475.4320823 | 719.3702561 |
| 12 | 51338 | Nyeri | RURAL | 29212.4078 | 44200.92378 |
| 13 | 51338 | Nyeri | URBAN | 4529.913603 | 6854.154823 |
| 14 | 51339 | Thika | RURAL | 25845.07942 | 39105.86191 |
| 15 | 51339 | Thika | URBAN | 10426.7146 | 15776.52964 |
| 16 | 51340 | Kilifi | RURAL | 22892.83771 | 34638.86241 |
| 17 | 51340 | Kilifi | URBAN | 1022.740115 | 1547.495097 |
| 18 | 51341 | Kwale | RURAL | 19973.2134 | 30221.21589 |

# Modelling population distributions

***Bottom-up:*** where census data are outdated/unreliable, integration of high resolution settlement mapping with survey data

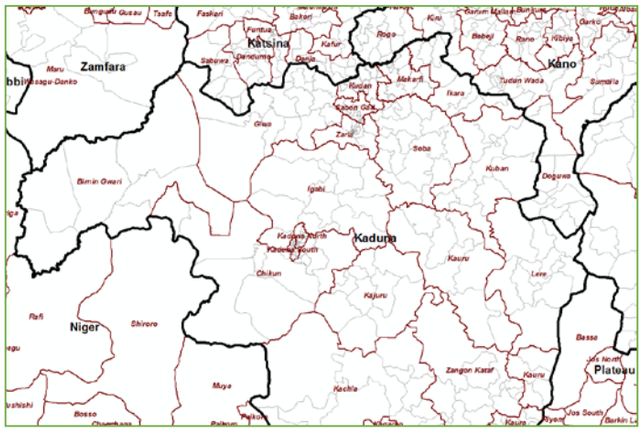



***Top-down:*** disaggregation of admin-unit based census/official estimate counts

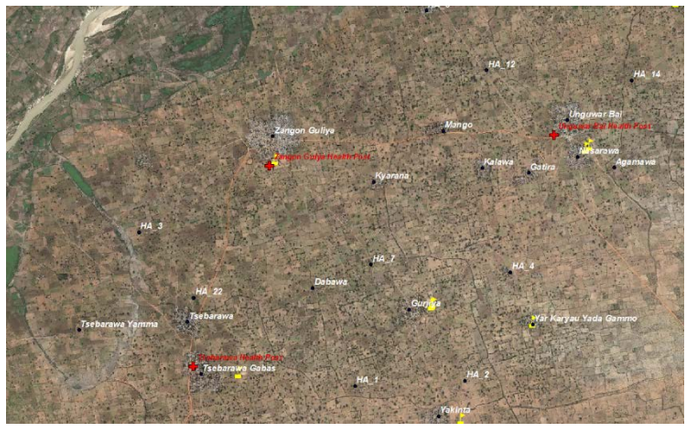


BILL & MELINDA GATES foundation

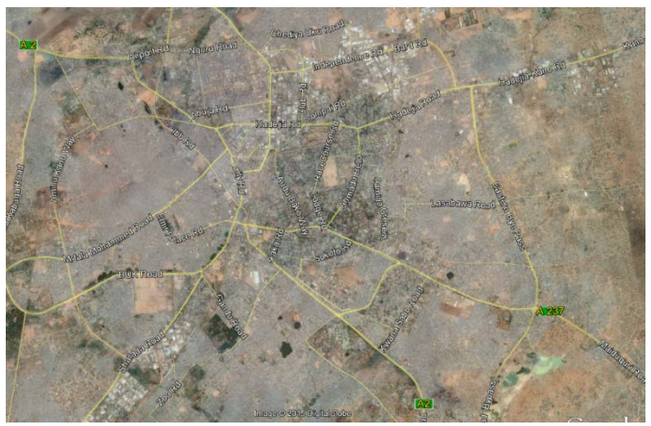
# Core Data Layers



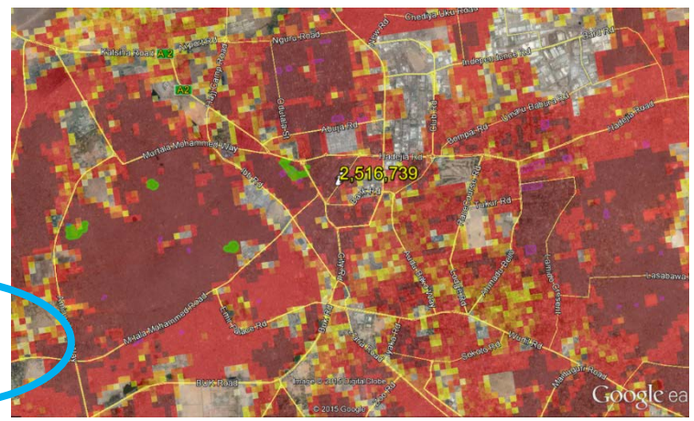

***Administrative Boundaries***




***Settlements and Points of Interest***



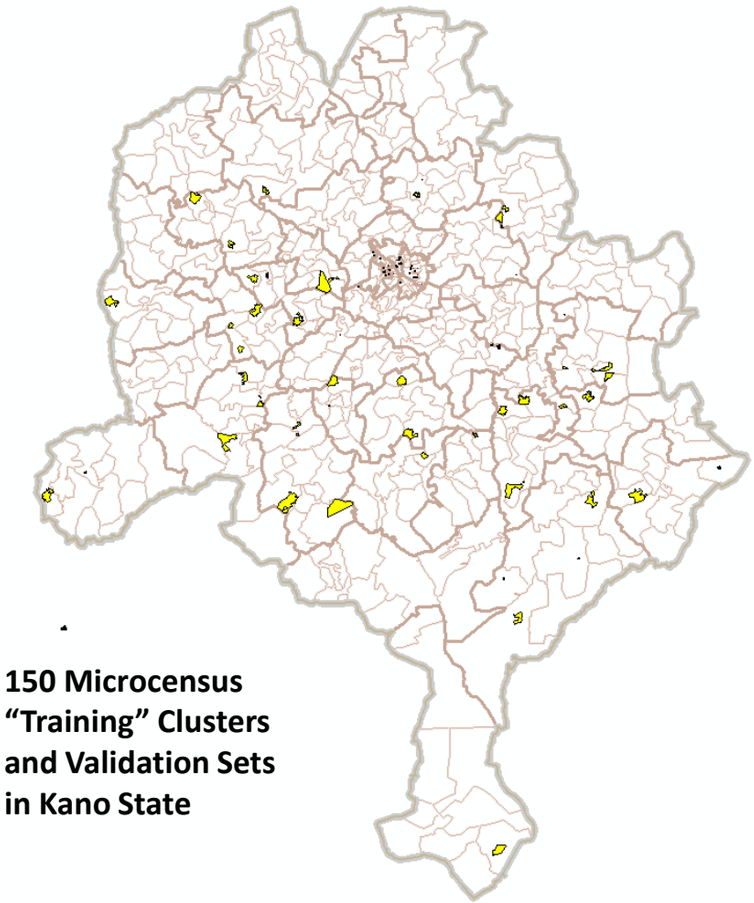
***Transportation Network***



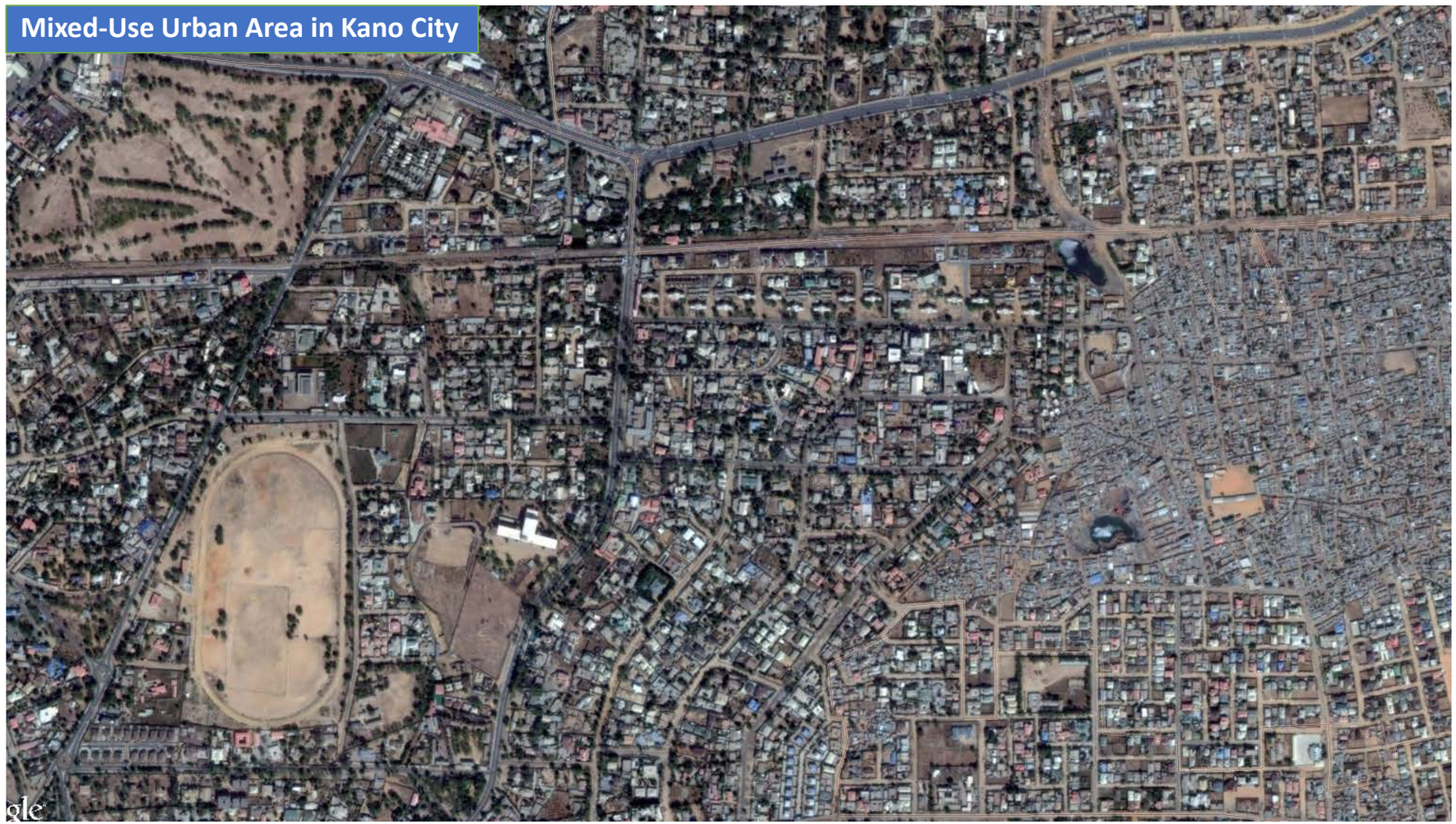

***Population / Demographics***

# Basis for Population Modeling with Satellite Imagery

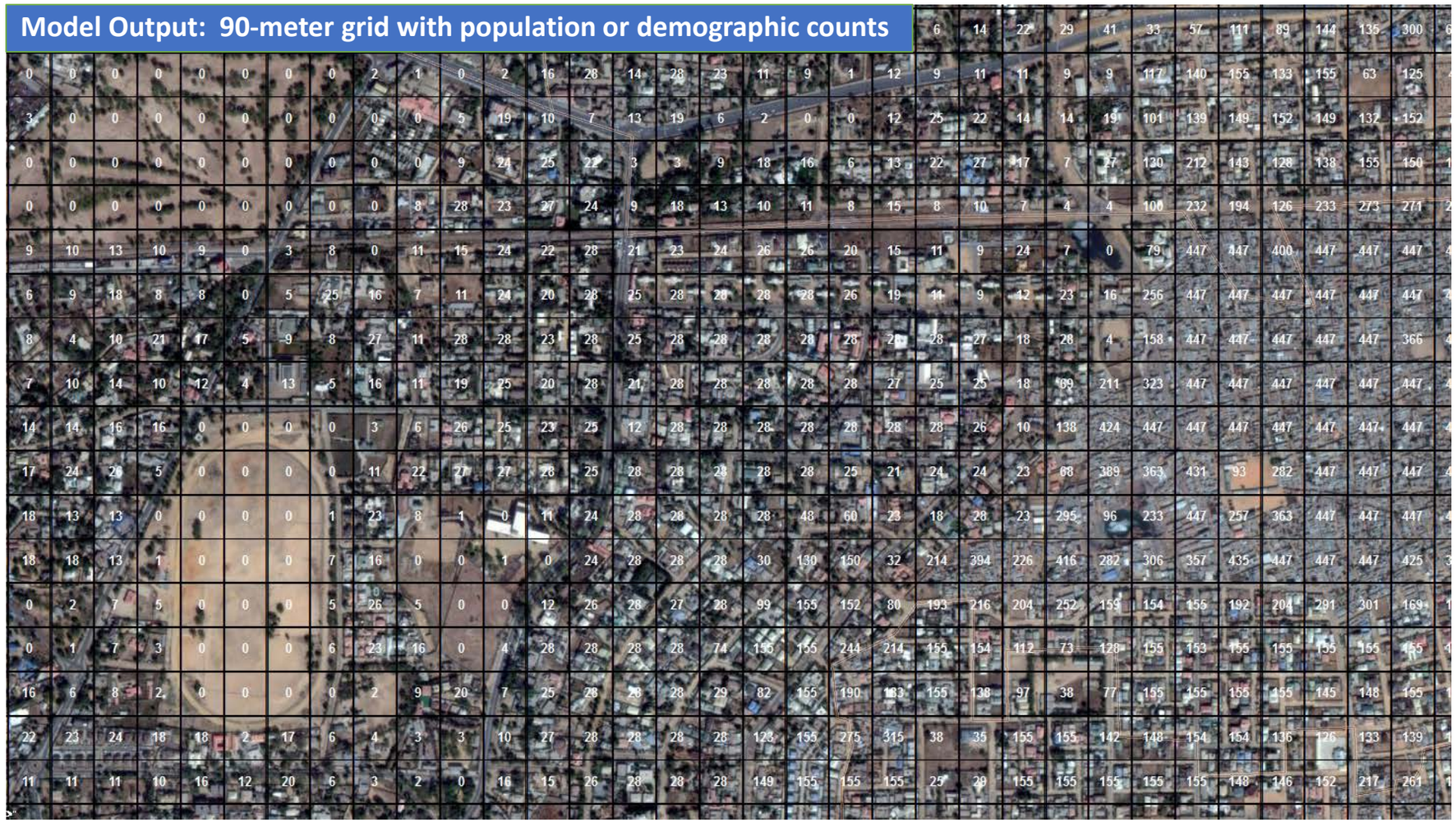
- **Obtain High-Resolution Satellite Imagery**
- **Extract Settlement & Building Features**
- **Identify Neighborhood categories, covariates**
- **Collect Microcensus/Validation Data**
- **Generate & Validate Initial Output**
- **If Needed, Collect & Validate Additional Microcensus Data.**
- **Repeat until Accuracy Requirements are met.**



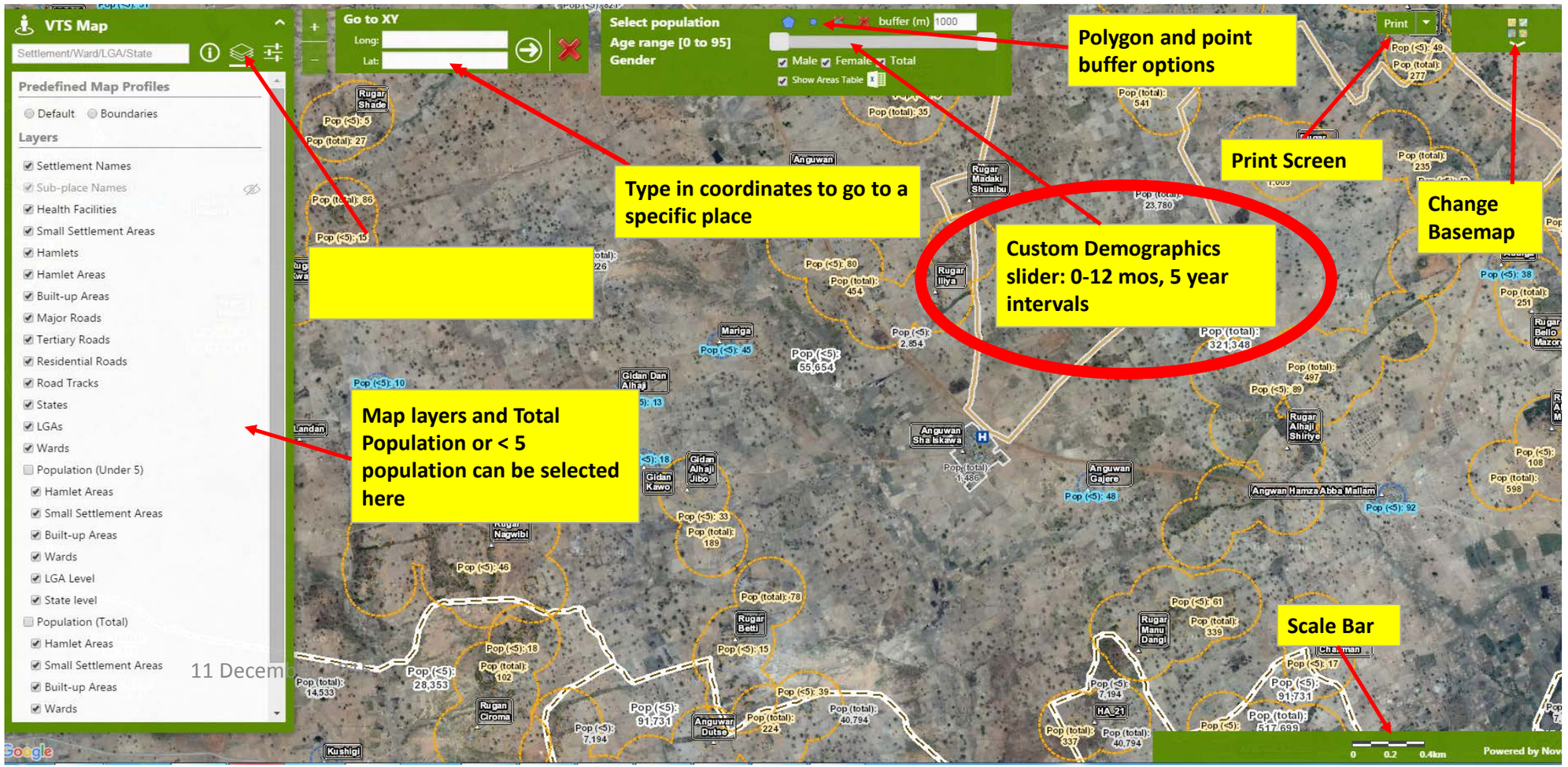
**150 Microcensus "Training" Clusters and Validation Sets in Kano State**


Mixed-Use Urban Area in Kano City

## Model Output: 90-meter grid with population or demographic counts

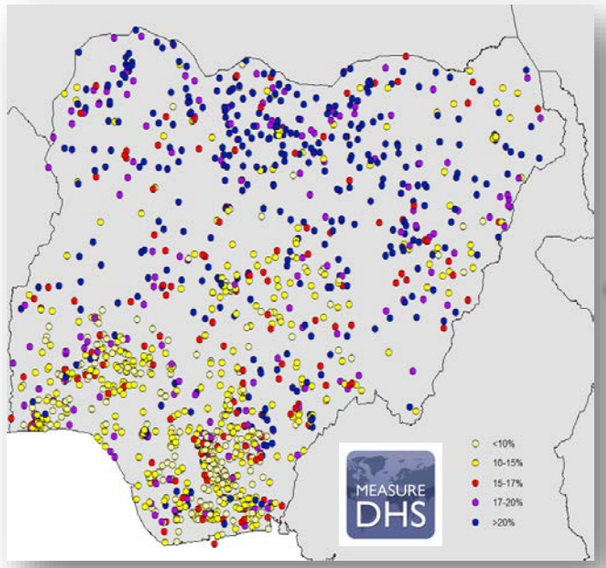


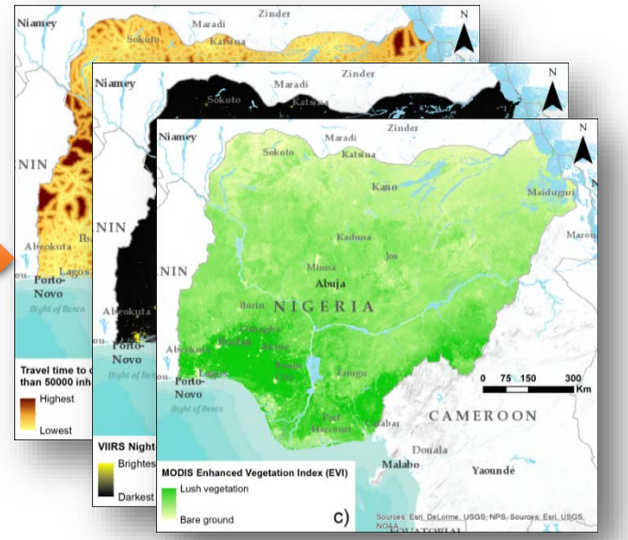
# POPULATION MODEL – USER INTERFACE OPTIONS

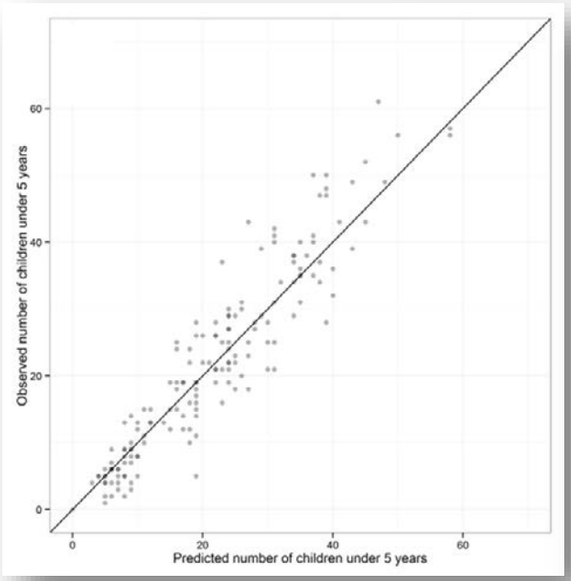


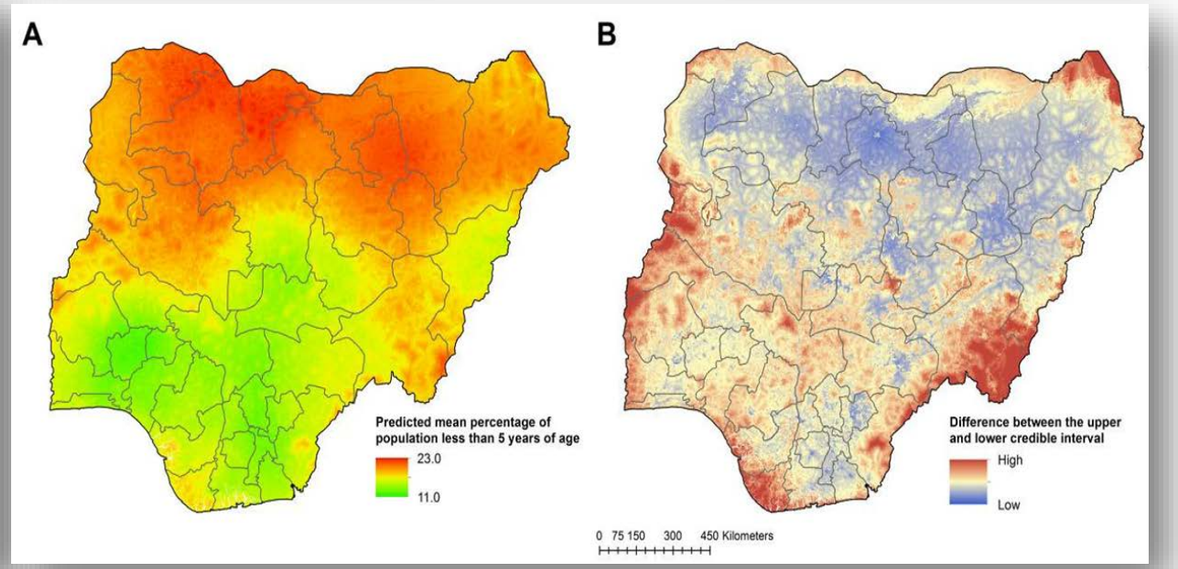

http://vts.eocng.org/

# Nigeria age structures: Bayesian model-based geostatistics

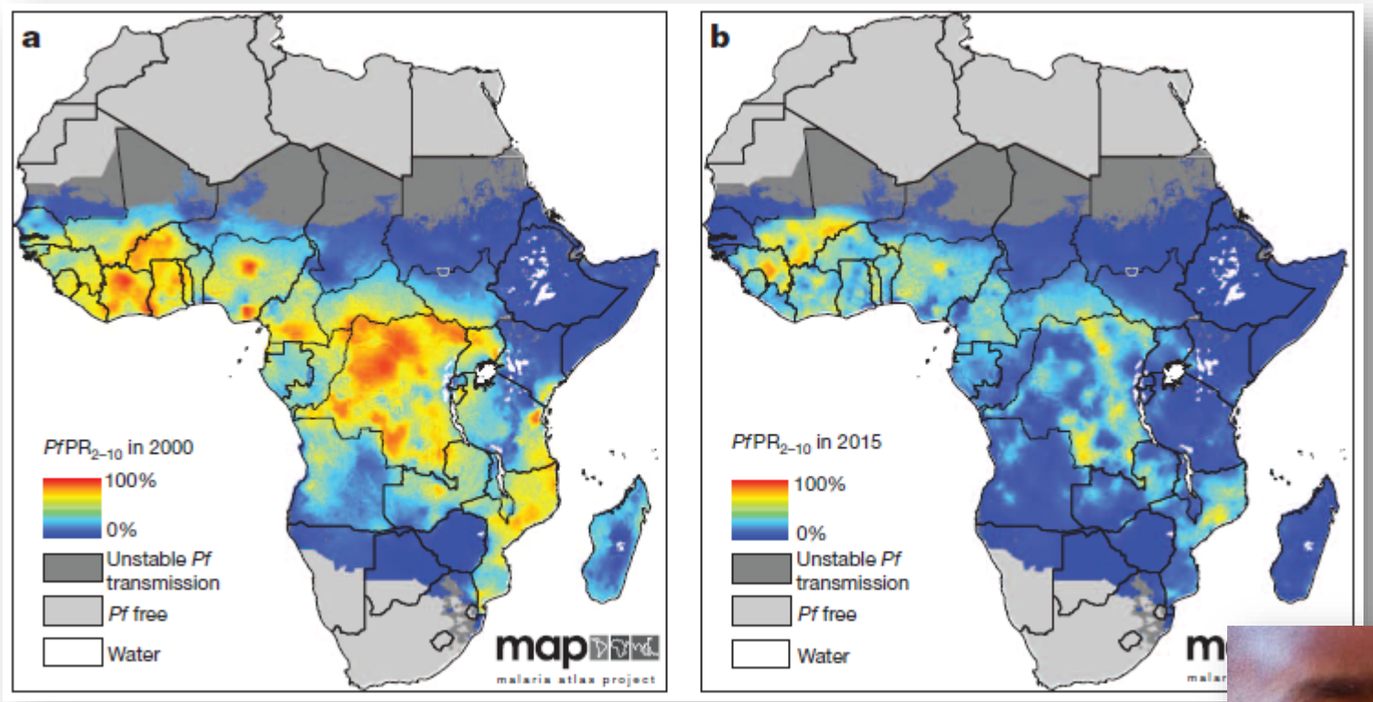


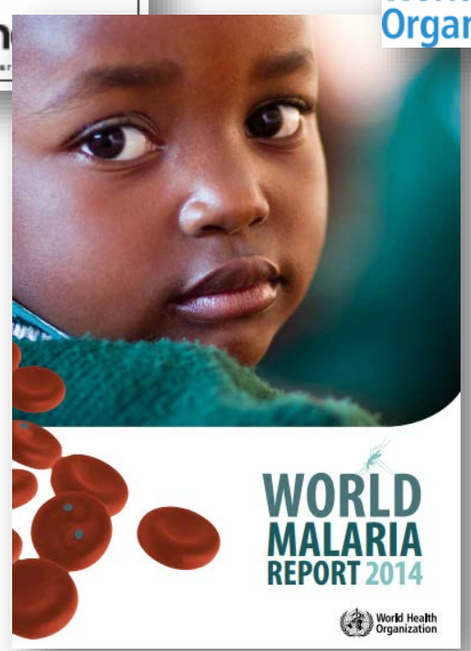

*Alegana et al (2015) Royal Society Interface*

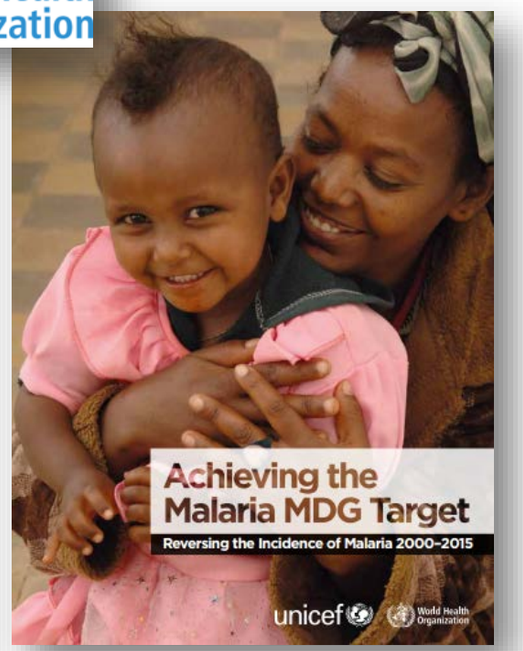



world pop

map

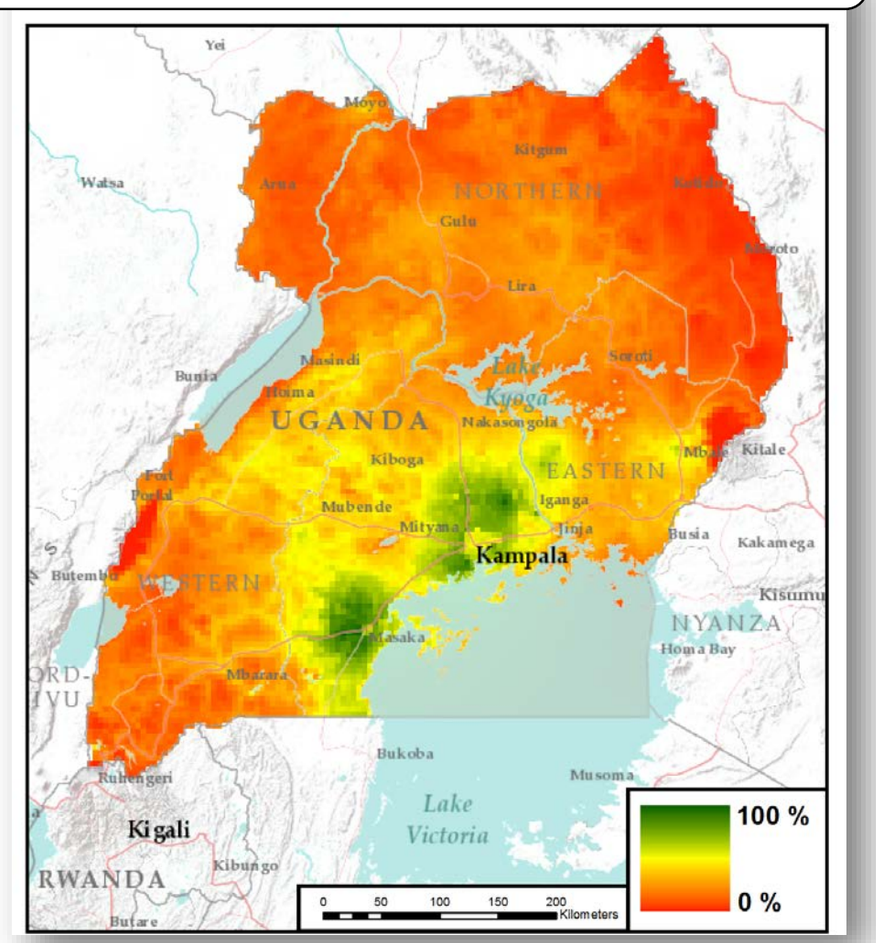
World Health Organization

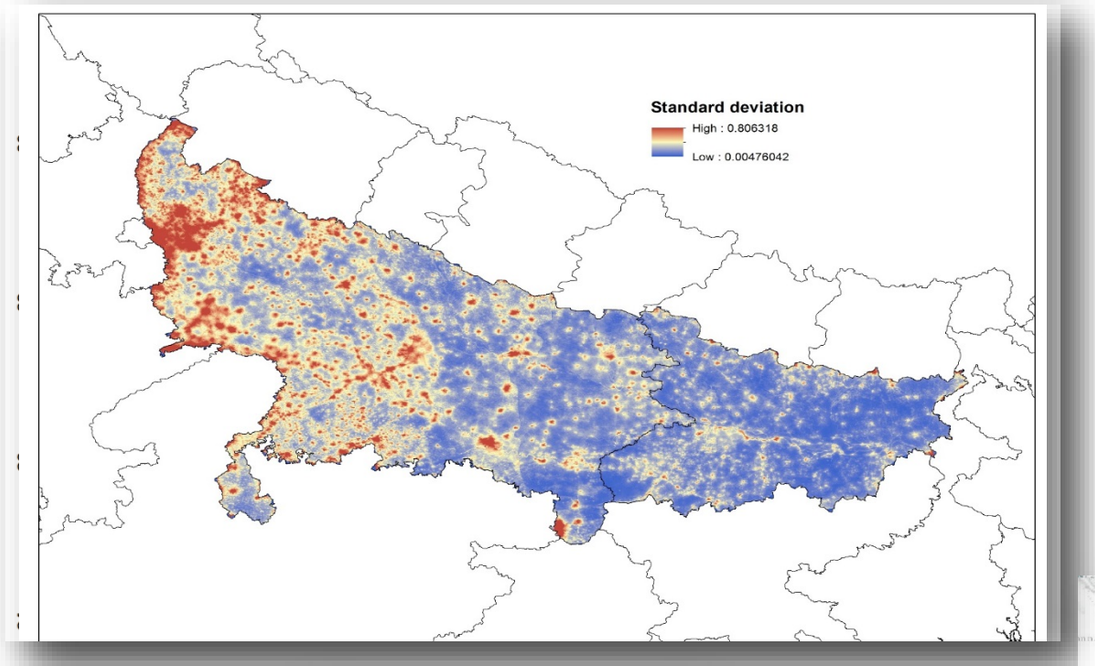


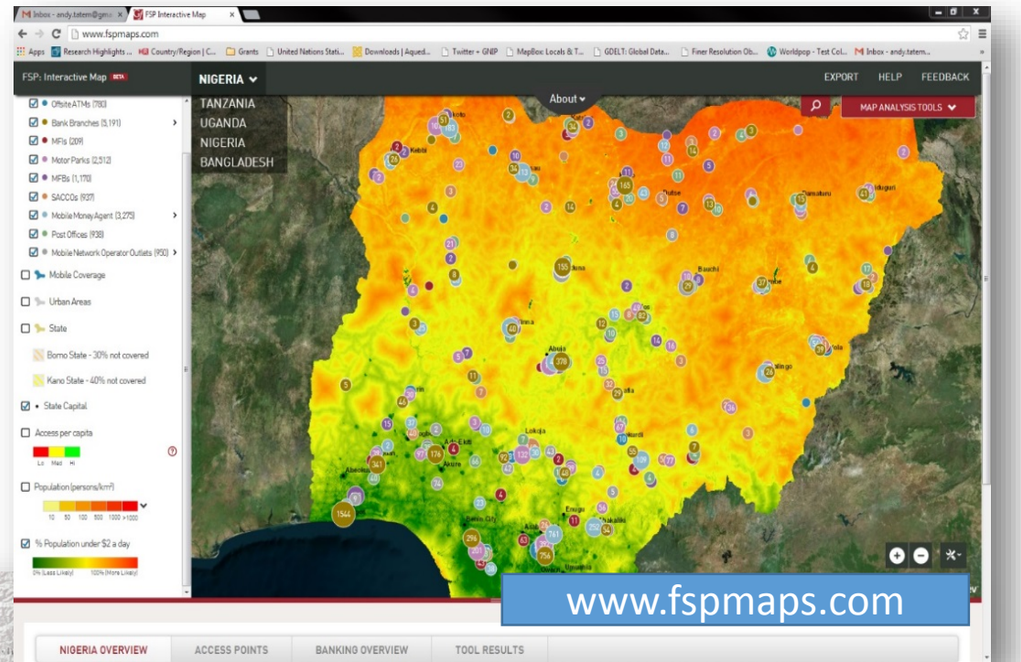



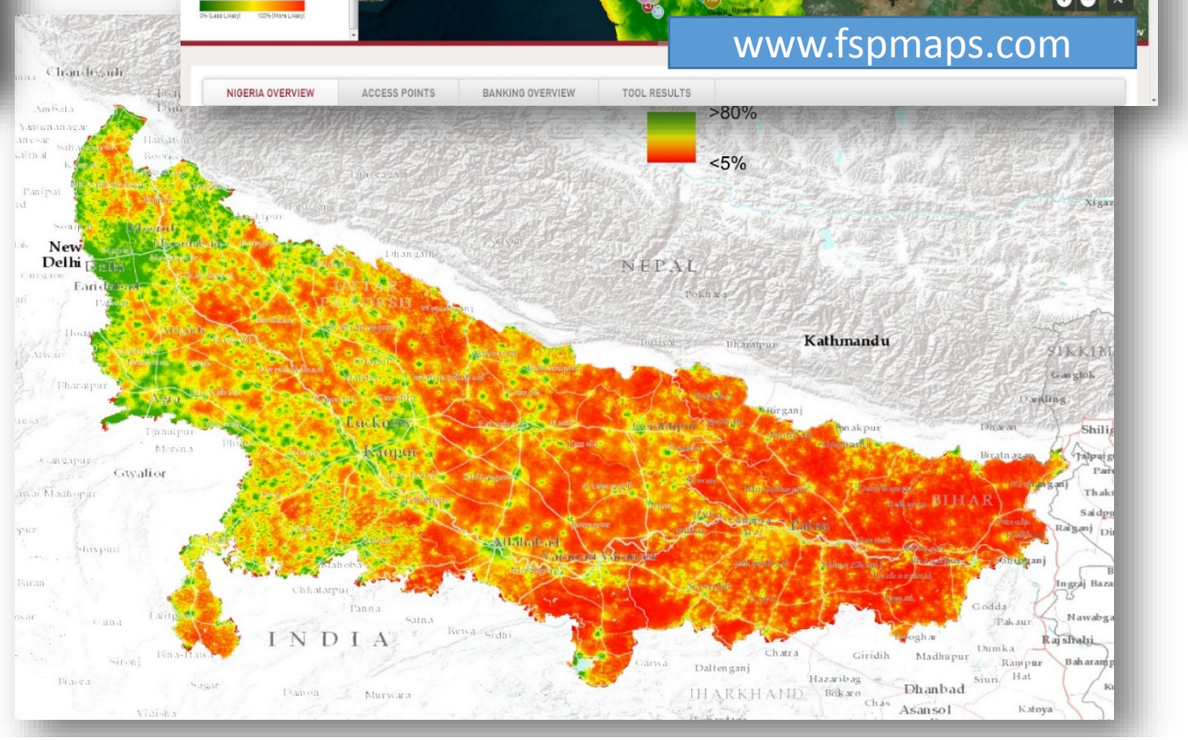

*Bhatt et al (2015) Nature*


Percentage of households with access to improved sanitation

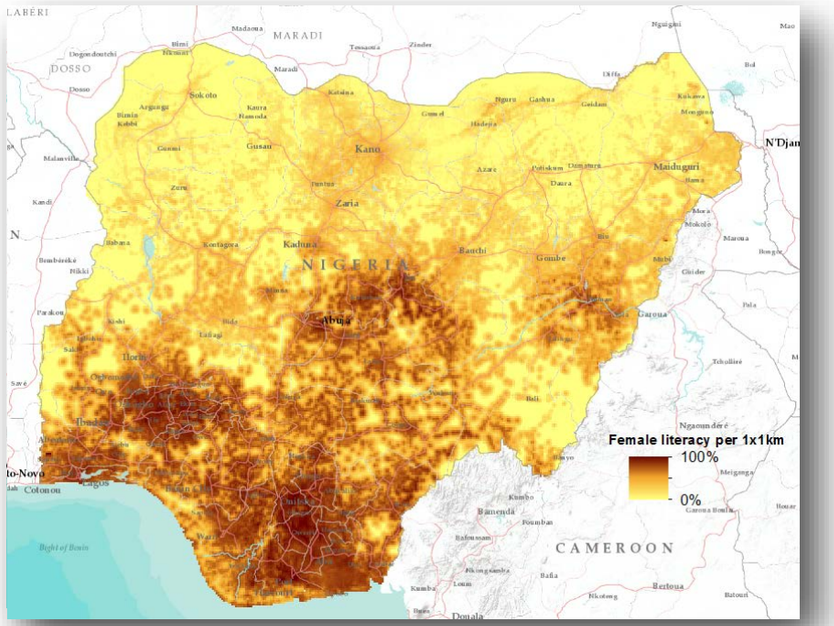


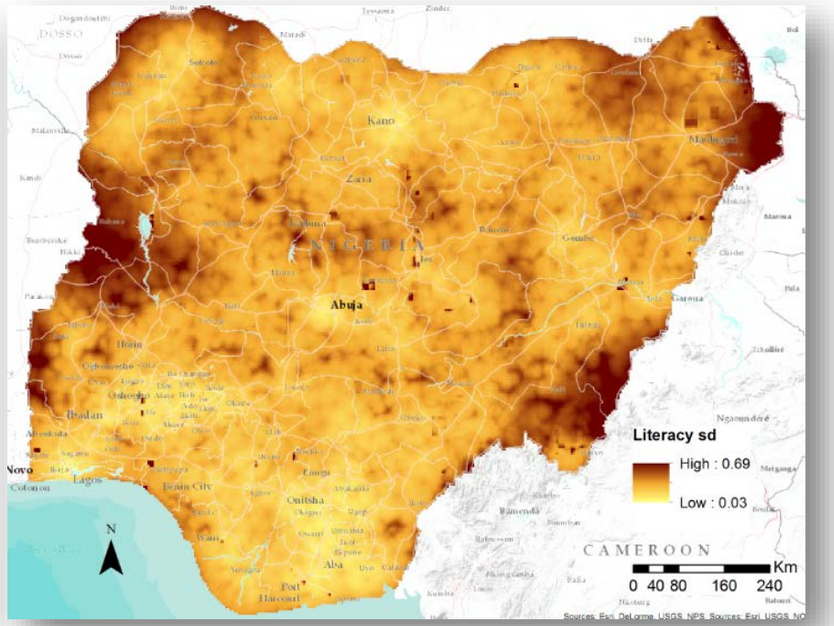


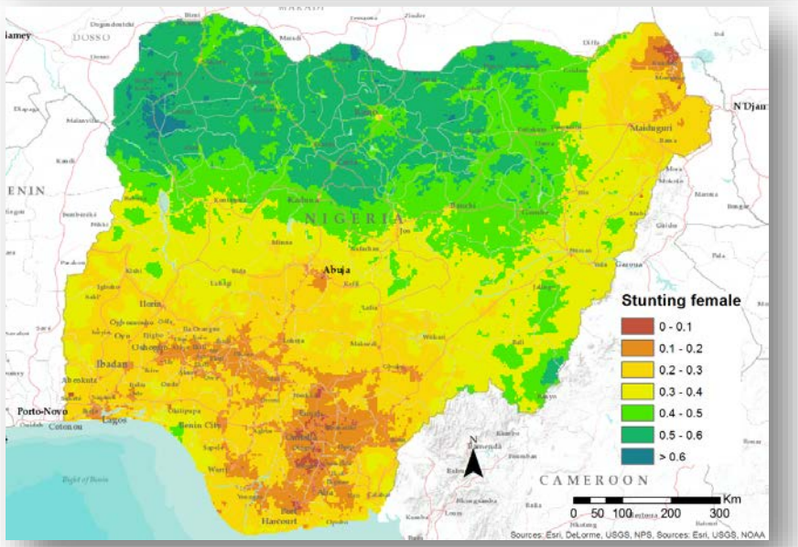



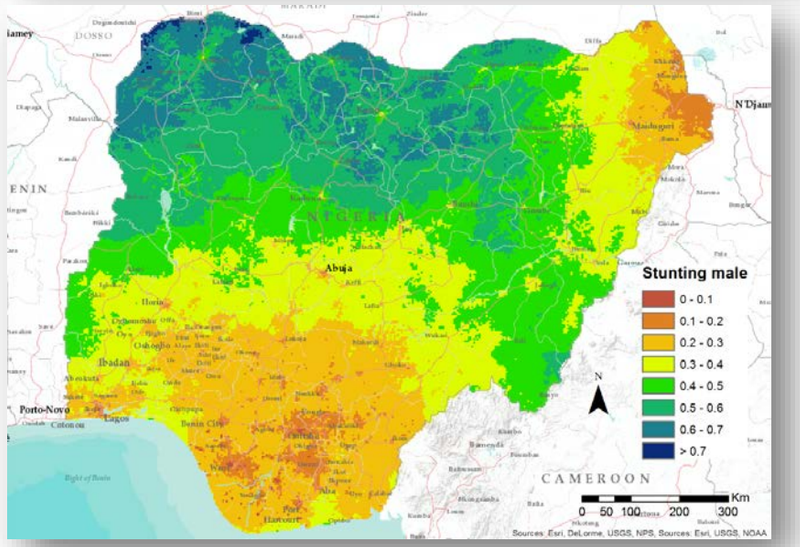




*Bird et al (2015) under review,*
*Tatem et al (2014) BMGF report*



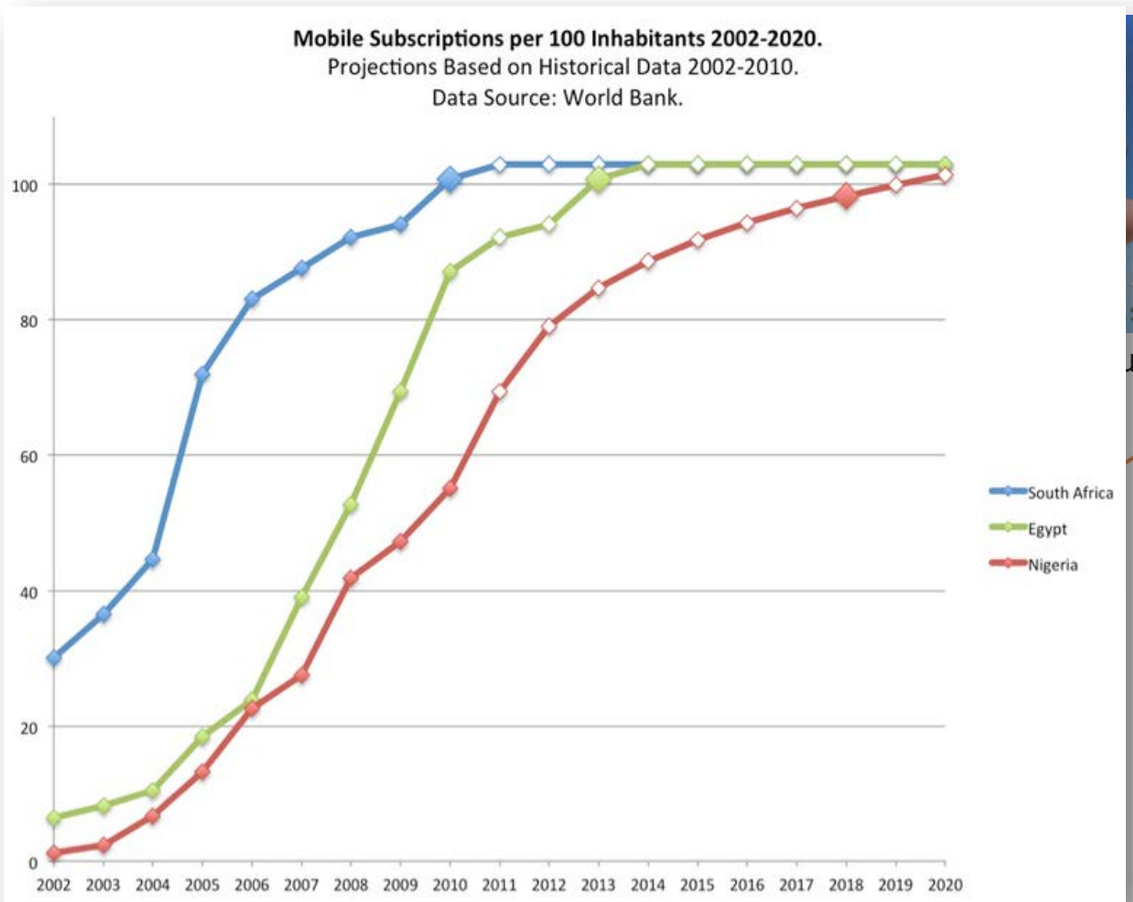

data2x
partnering for a gender data revolution

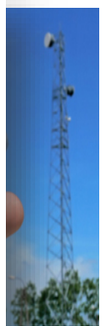
UNITED NATIONS FOUNDATION

world pop

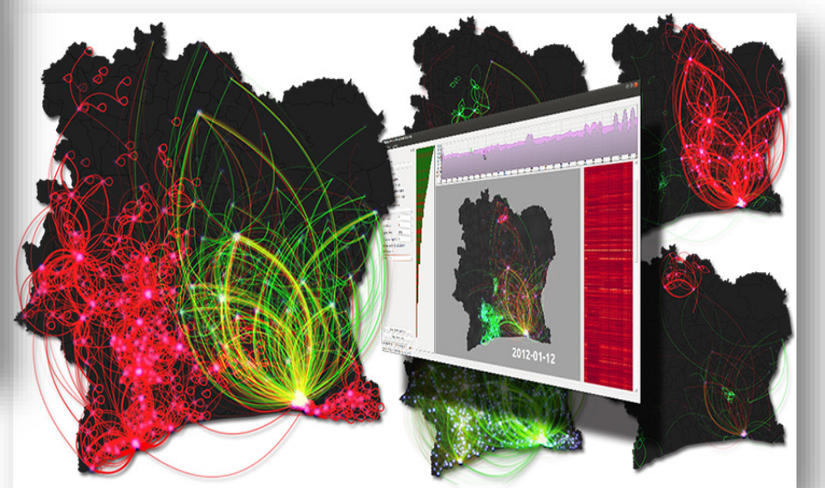
FLOWMINDER.ORG

# Cellphone call data records (CDRs)



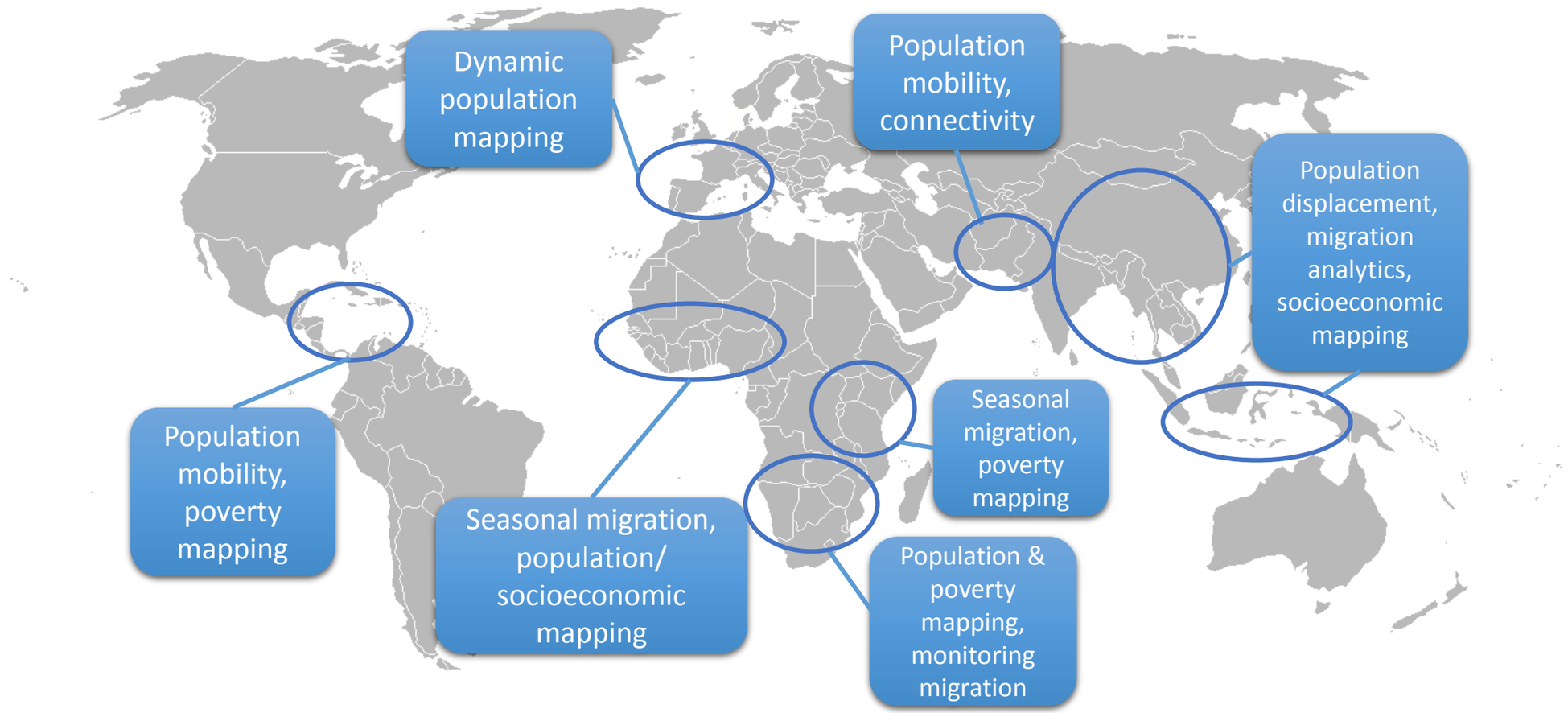
Network operator records time and tower of call for billing



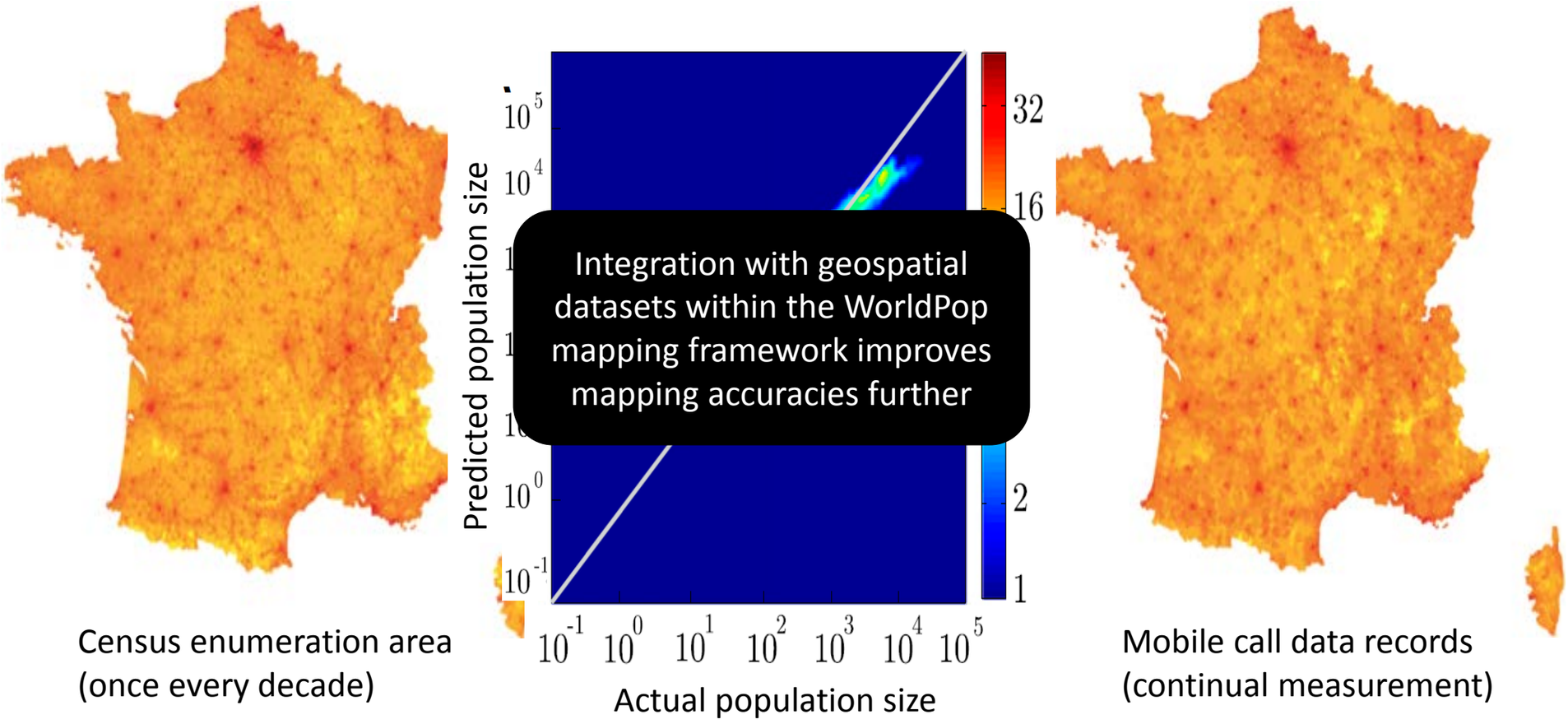
## Cellphone demographics work

FLOWMINDER.ORG

worldpop



# Dynamic population mapping

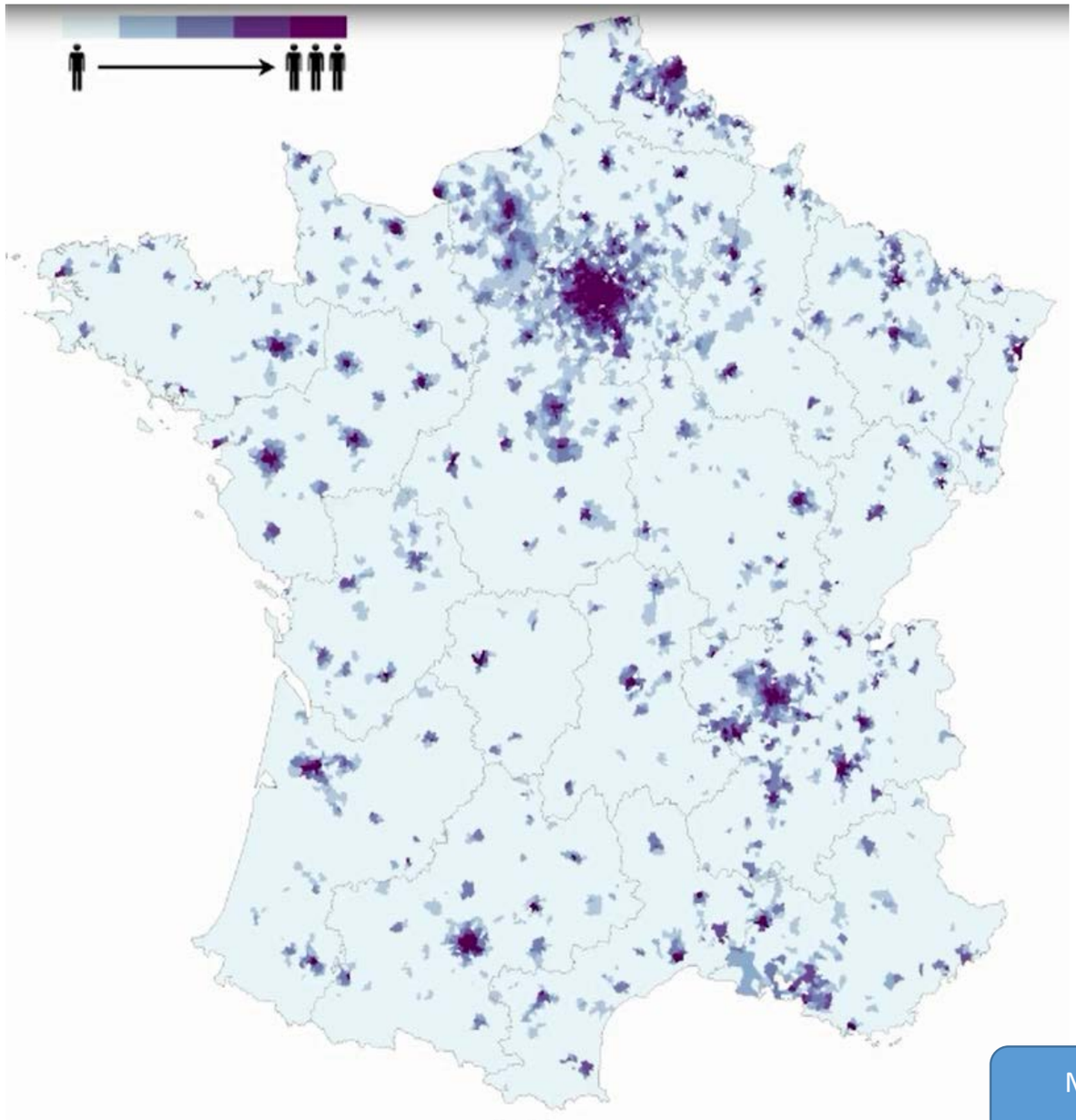



Census enumeration area (once every decade)

Mobile call data records (continual measurement)

*Deville et al (2014) PNAS*

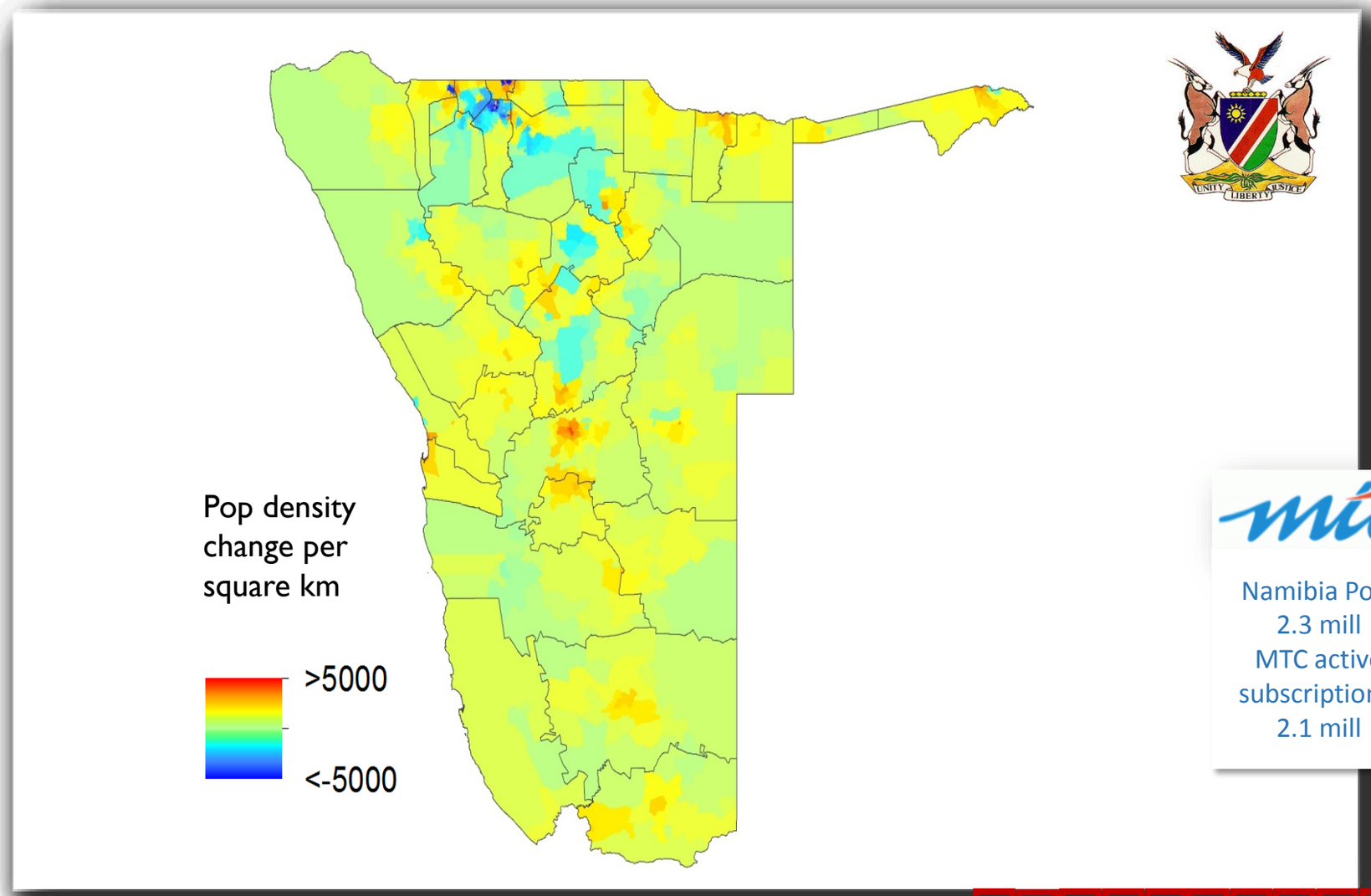
Wednesday
Thursday
Friday
Saturday
Sunday
Monday



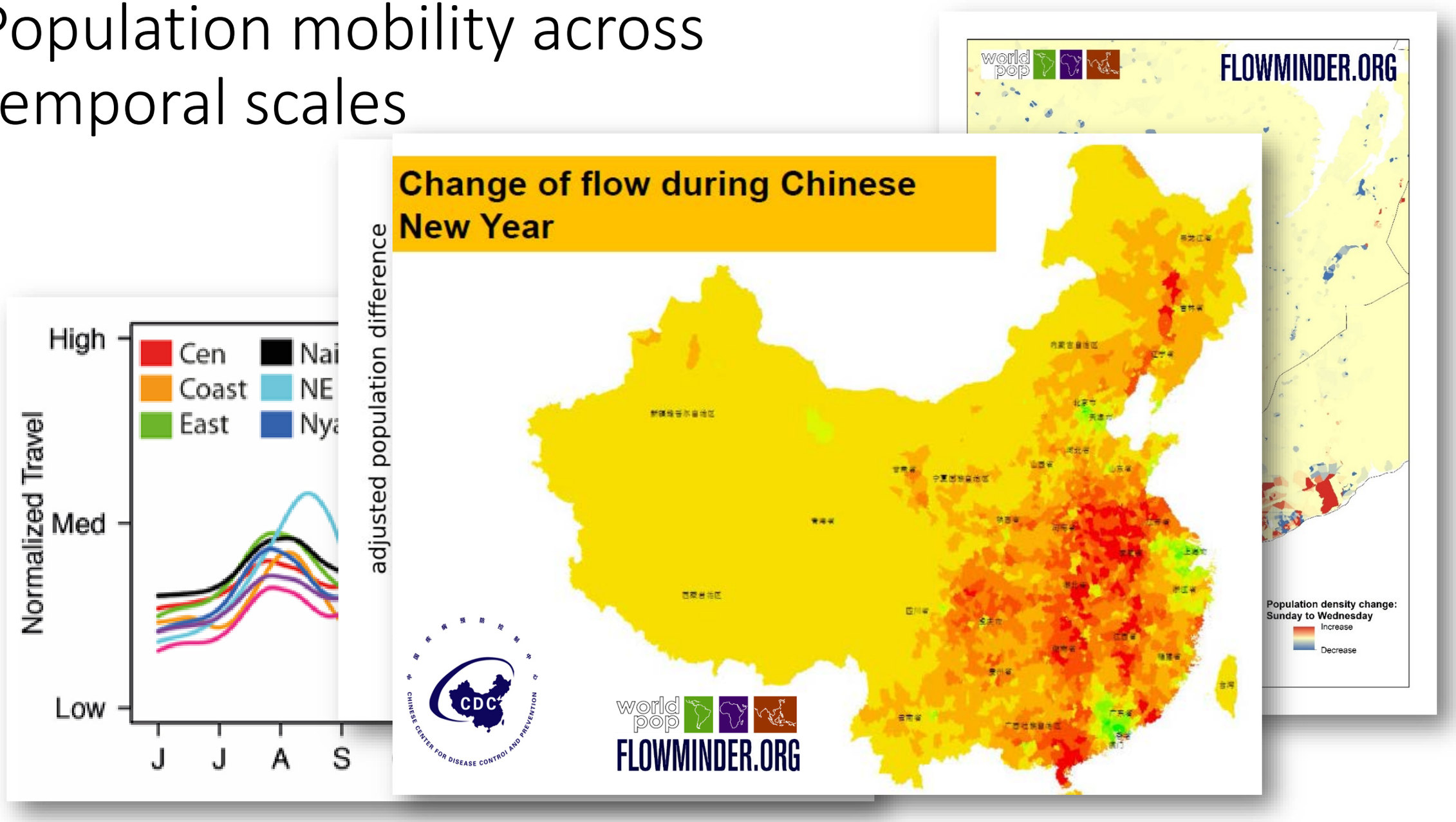
*Deville et al (2014) PNAS*

Watch the video at:
http://youtu.be/qsUDH5dUnvY

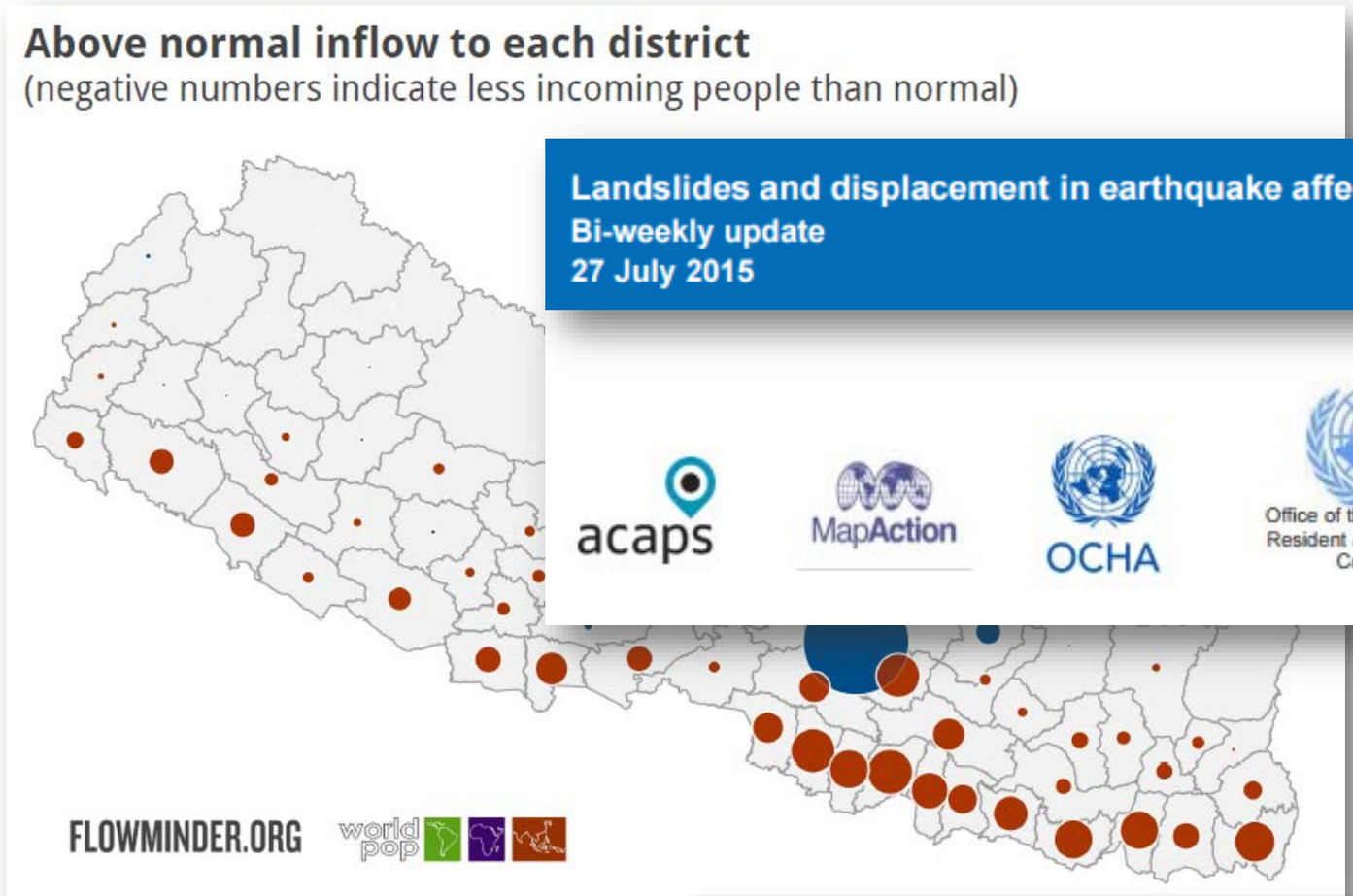
Monthly data available:
www.worldpop.org


Pop density
change per
square km
>5000
<-5000
UNITY
LIBERTY
JUSTICE
mic
Namibia Pop:
2.3 mill
MTC active
subscriptions:
2.1 mill
NOV_12
DEC_12
JAN_13
FEB_13
MAR_13
APR_13
MAY_13
JUN_13
JUL_13
AUG_13
SEP_13
world pop
FLOWMINDER.ORG

# Population mobility across temporal scales



# Population displacements




**Landslides and displacement in earthquake affected areas**
**Bi-weekly update**
**27 July 2015**

Nepal Earthquake Assessment Unit

acaps

MapAction

OCHA

Office of the United Nations Resident and Humanitarian Coordinator

UKaid

FLOWMINDER.ORG
world pop

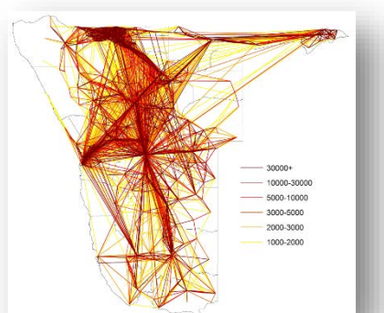
www.worldpop.org.uk/nepal

# Modelling connectivity for disease elimination


*Wesolowski et al (2014) PLoS Currents Outbreaks; Garcia et al (2014) Migration Studies; Sorichetta et al (2015) in review; Tatem et al (2009) Malaria Journal*

# Data integration: Targeting interventions

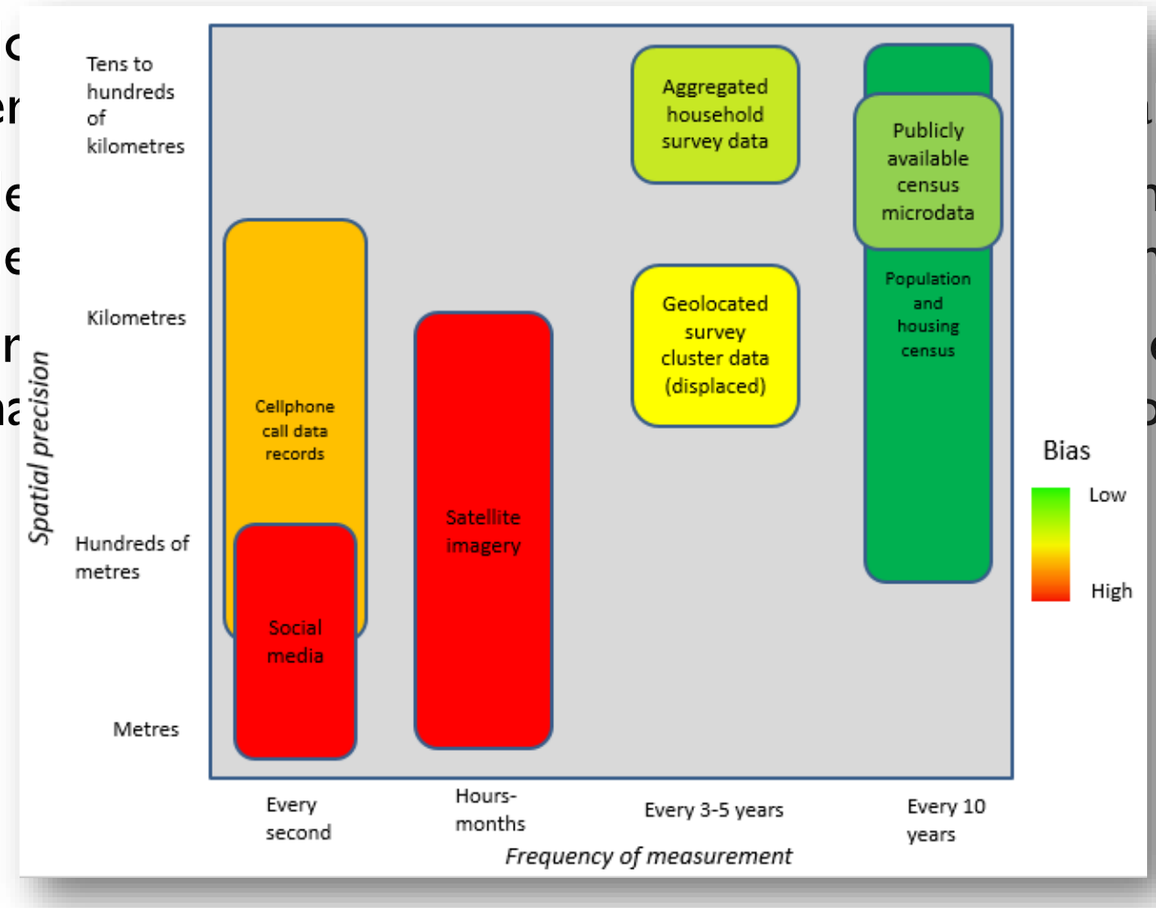
| Malaria risk zone | Zone map | Population residing in zone in 2011 |
|---|---|---|
| Pre-2013 intervention deployment |  | 1.29 million |




*Tatem et al (2014) Malaria Journal*

# Summary

- In producing estimates for different geographical scales and time periods, the integration of multiple types of data to compliment traditional sources is required
- Nc each has advantages over census data in ter l precision
- Me nties and providing clear me portant
- Gr es to build strong demographic databases tha proaches



# Acknowledgements

# Further information


www.worldpop.org

@WorldPopProject

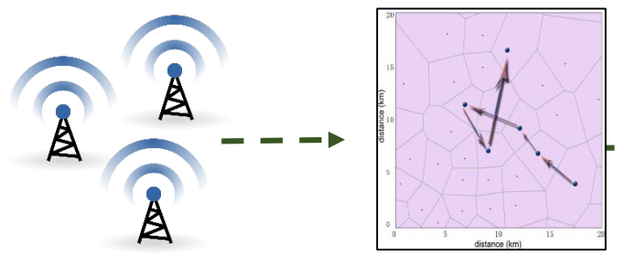
FLOWMINDER.ORG

www.flowminder.org

@Flowminder

# Extra slides

# Preserving confidentiality



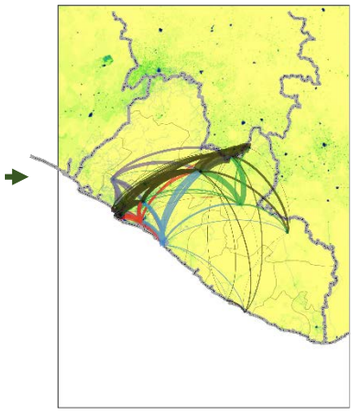



Operators Call Detail Records' (CDRs) including low-resolution location data (nearest tower location) anonymized on separate server hosted by operator.

We conduct analyses under operator supervision, anonymous raw data always behind operator firewall

Mobile operator firewall

FLOWMINDER.ORG



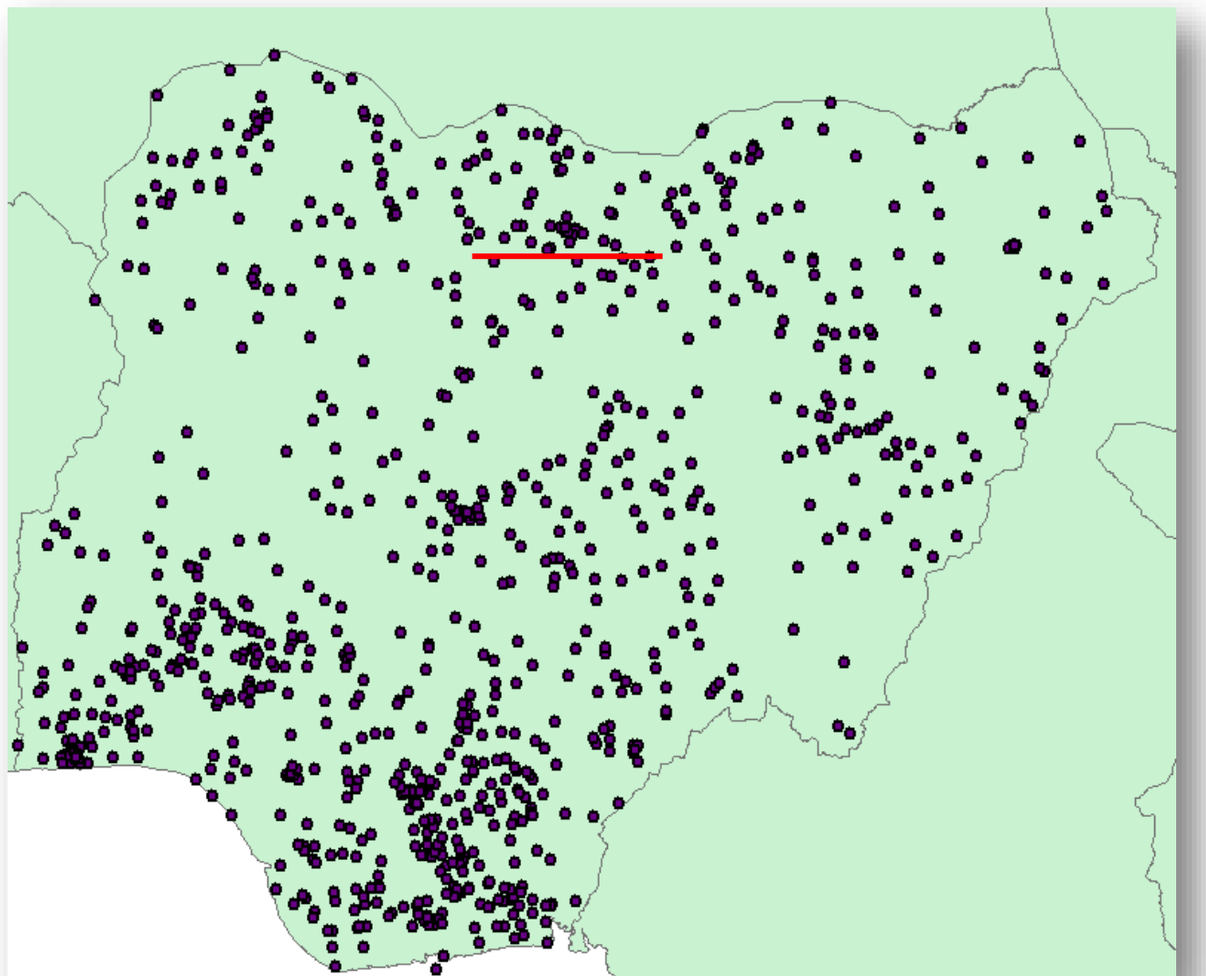
Aggregated mobility estimates are exported and made open access - can be used with other mobility estimates, epidemiological data

**<u>Compliance with GSMA data integrity guidelines</u>: Data never leaves mobile operator's system to avoid any privacy, commercial concerns.**

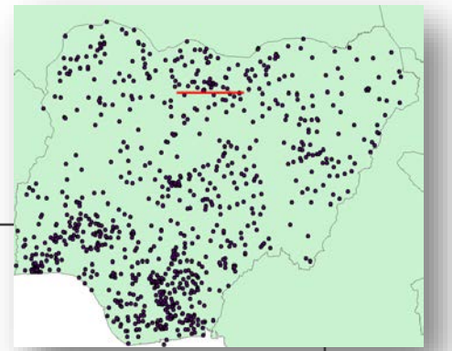
# Problem statement

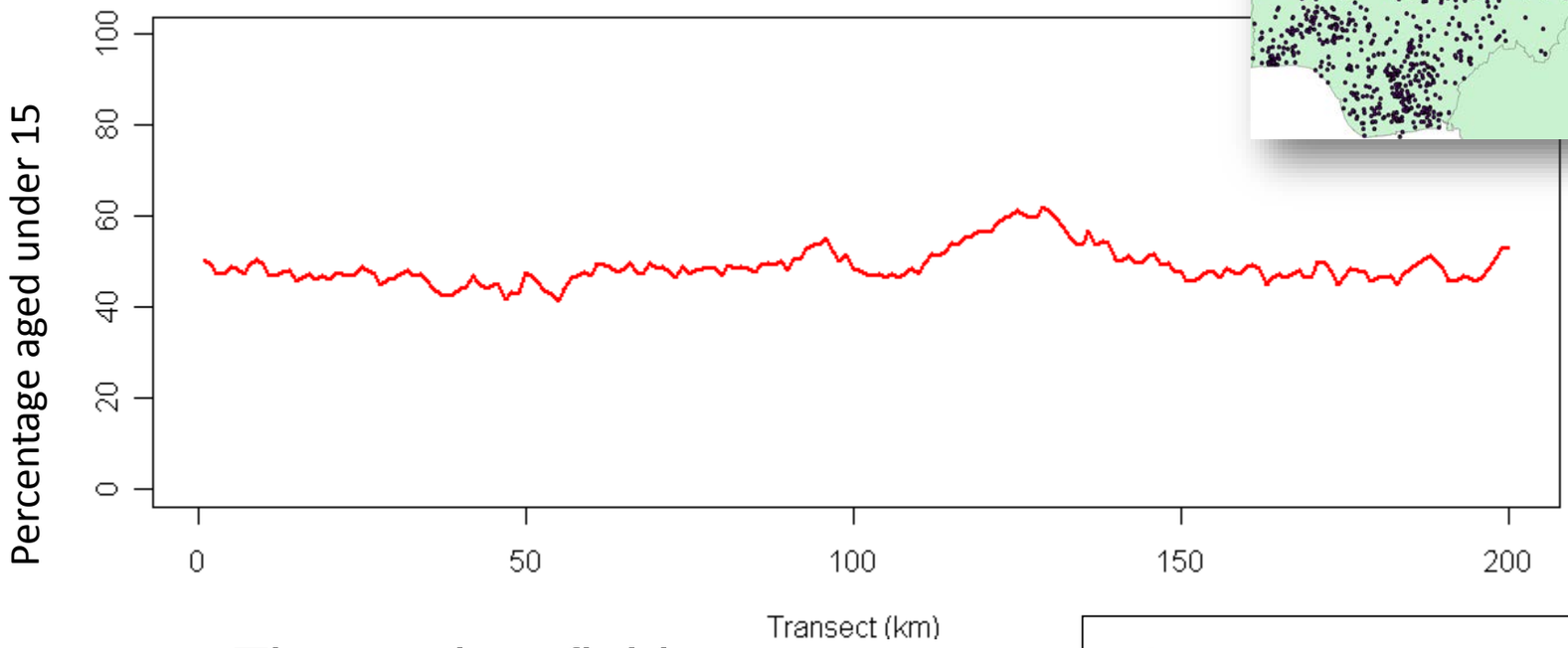
- We have
  - Hundreds of geolocated community surveys
  - Spread (unevenly) across areas of interest
  - Spread (unevenly) through time
- We want
  - Per-grid square estimates of age structure/household size etc
  - Must estimate these variables at locations with no data
  - Want to model uncertainty in our estimates
- Approach
  - Space-time model based geostatistics within a Bayesian framework

# Geostatistics: the random field



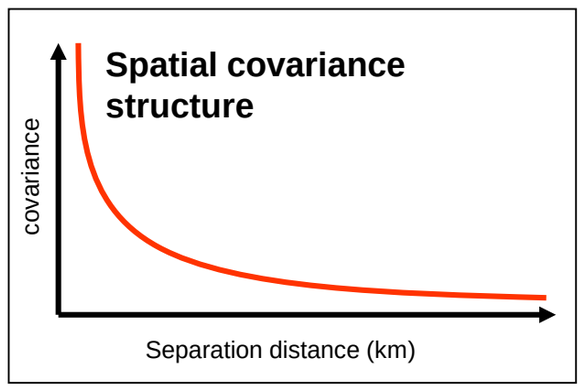
# Geostatistics: the random field



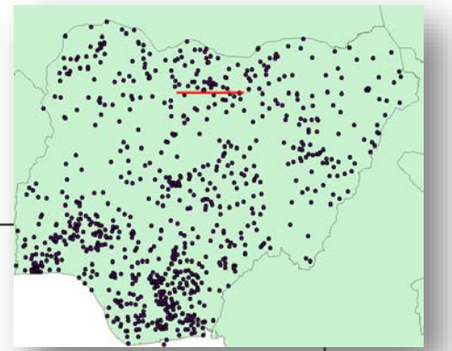


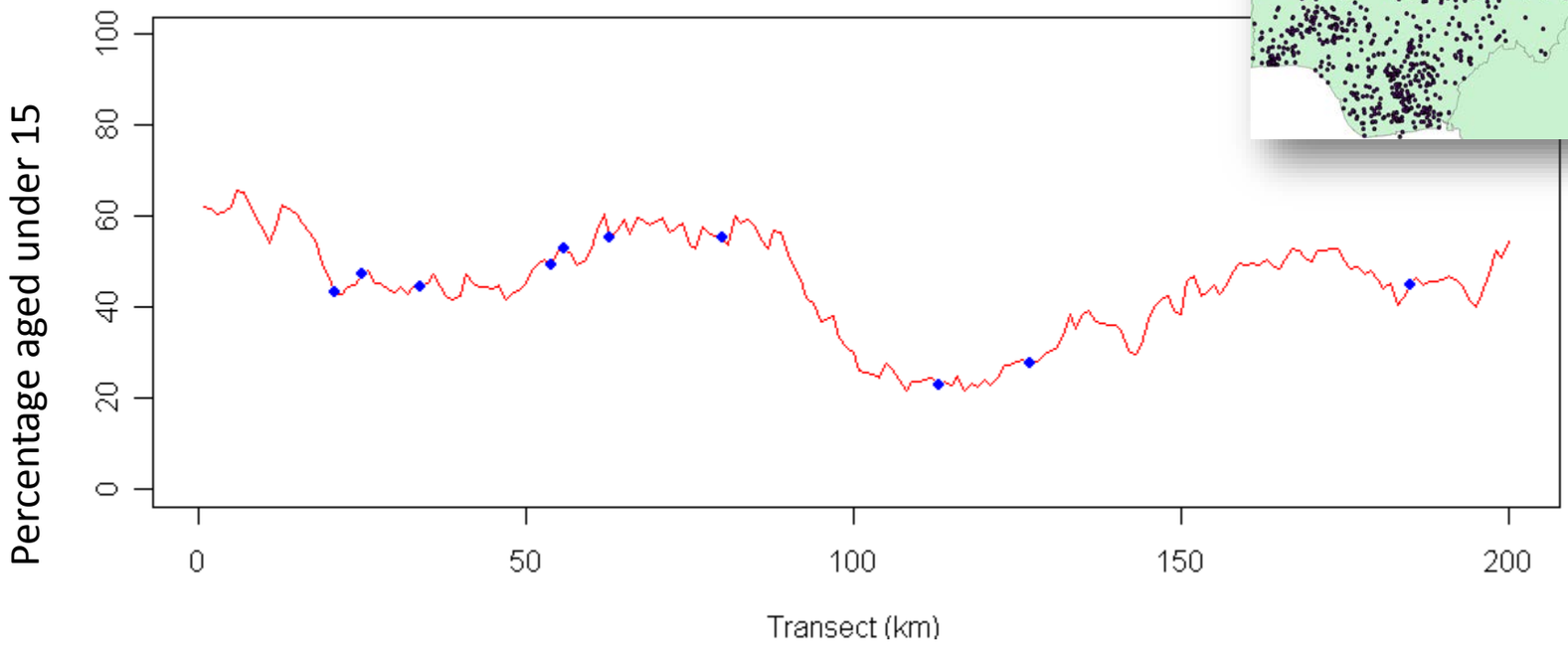

The random field:

- Mean
- Covariance structure
  1. Amplitude of variation
  2. Spatial range of correlation



# Geostatistics: the random field

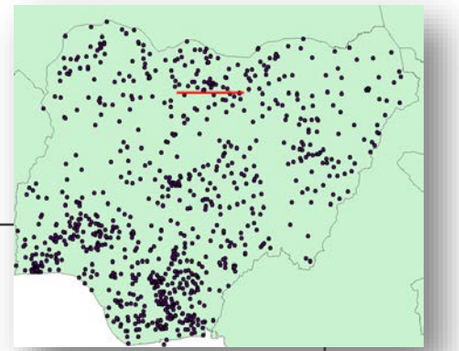


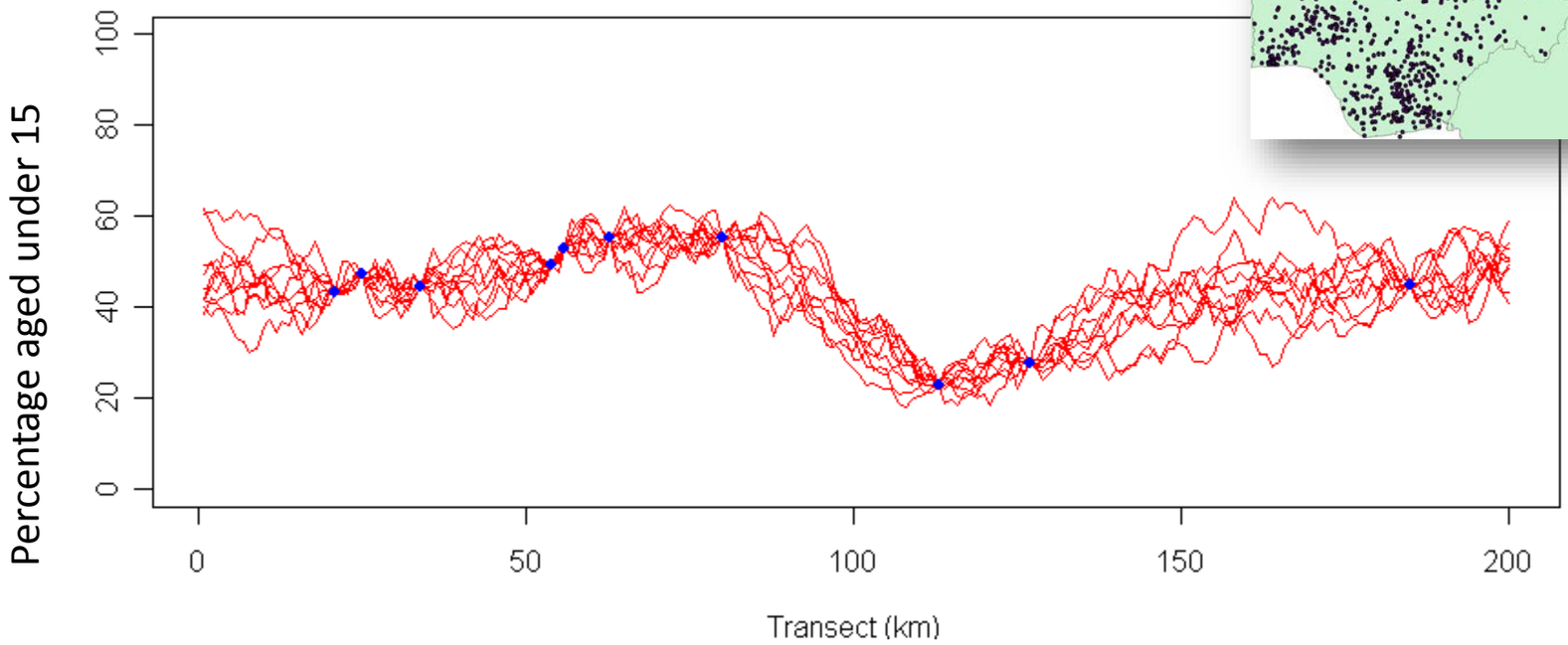



How do we use the data?

1. Parameterise mean and covariance function
2. ‘Condition’ on data values to predict everywhere

# Geostatistics: the random field

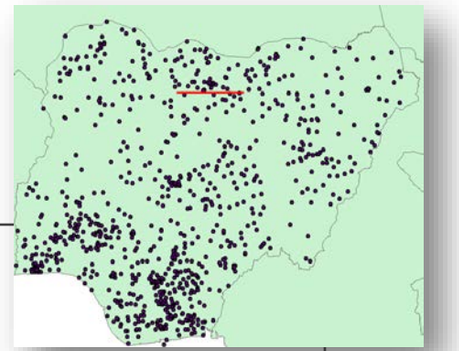


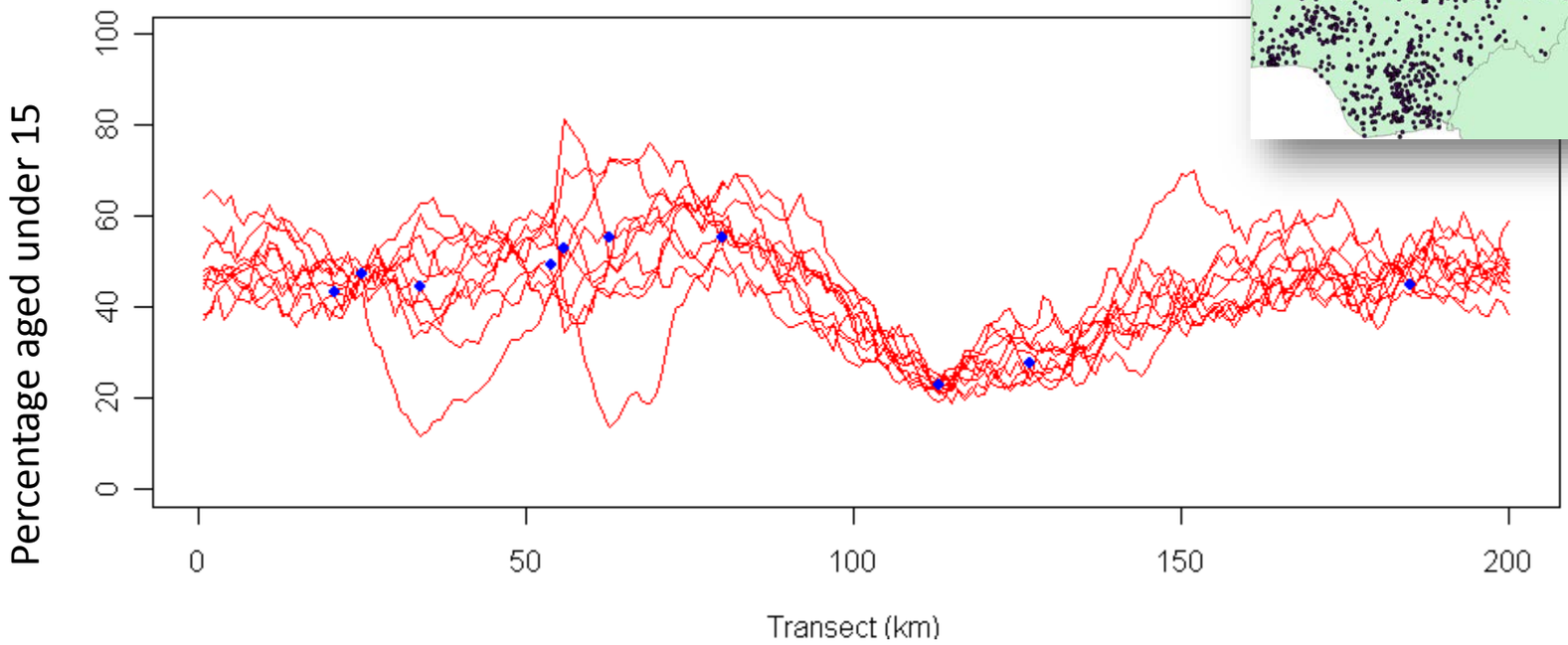



## Incorporating prediction uncertainty

- Represent using multiple realisations
- Uncertainty increases away from data
- Determined by amplitude / range

# Geostatistics: the random field



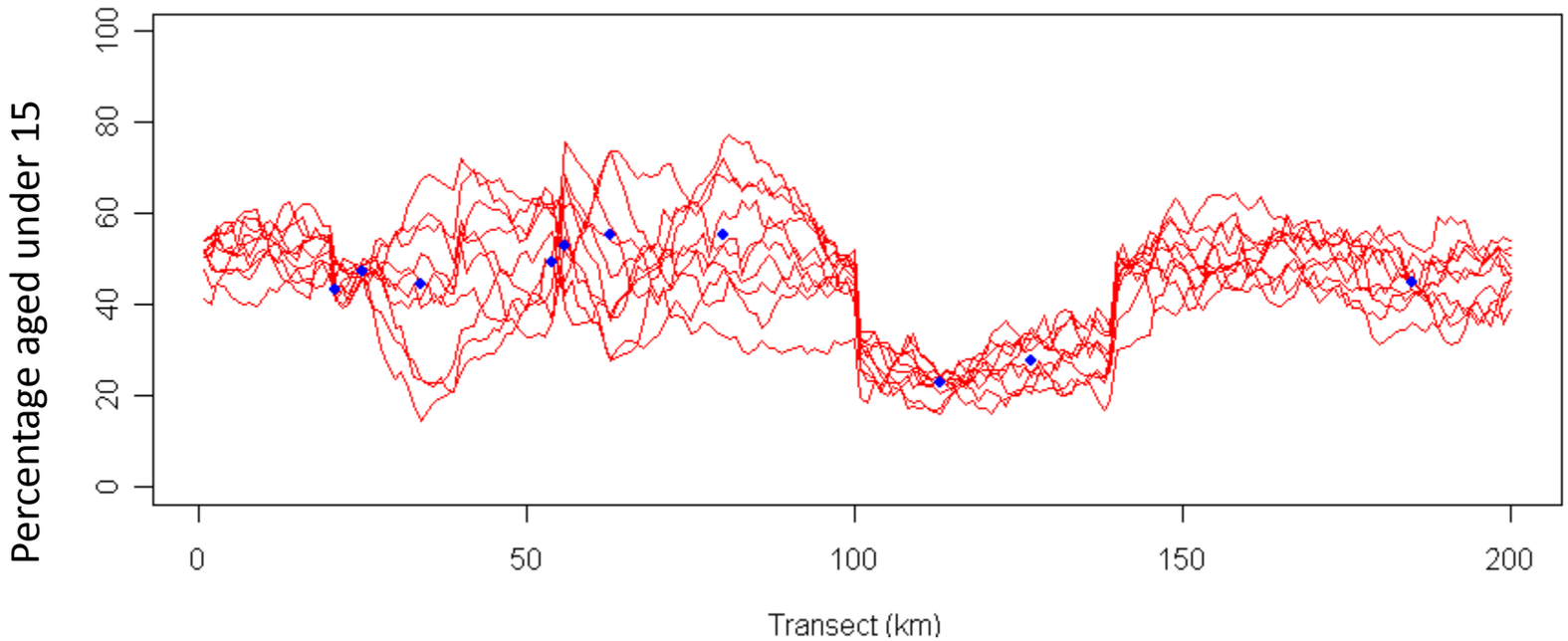



## Incorporating sampling variation

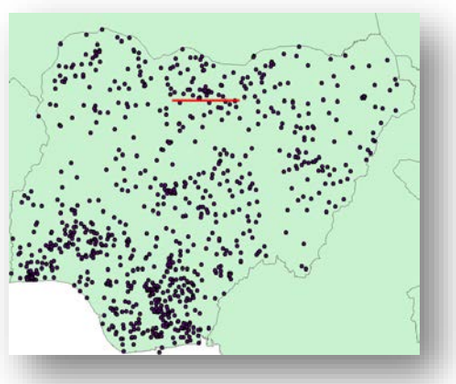
- Data are sampled with error (e.g. function of sample size)
- Adopt sampling model (e.g. binomial)
- Field is better defined where data are more certain

# Elaborations for demographic mapping

2. Incorporating environmental/infrastructure/socioeconomic covariates

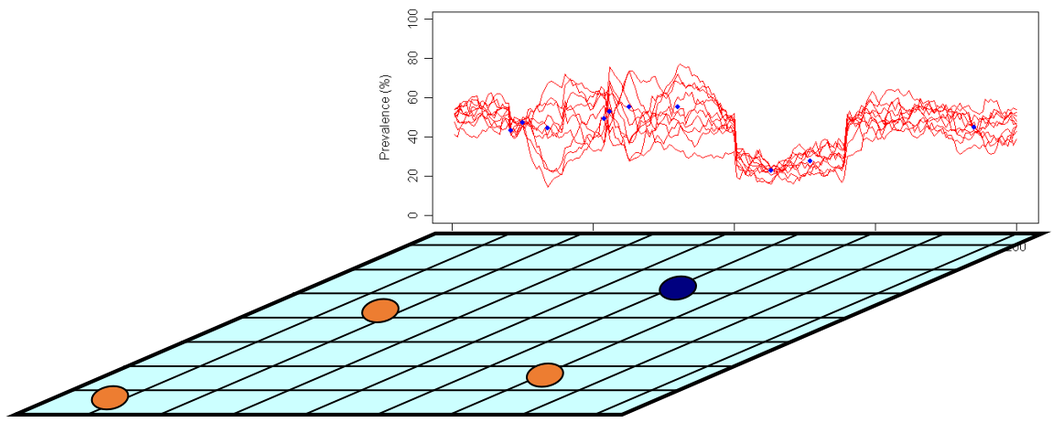


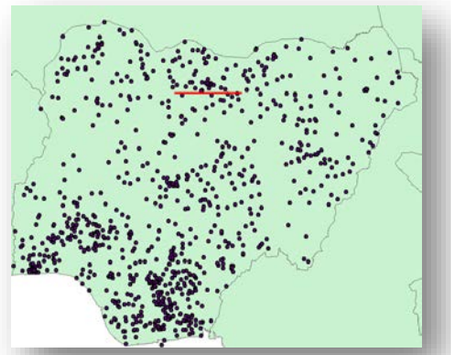

- E.g. Effect of urban vs rural areas
- R/U/Other status of data and predictions assigned from satellite data analyses
- Allowed underlying mean to vary between classes
- Multiple covariates selected in multivariate model



# Elaborations for demographic mapping

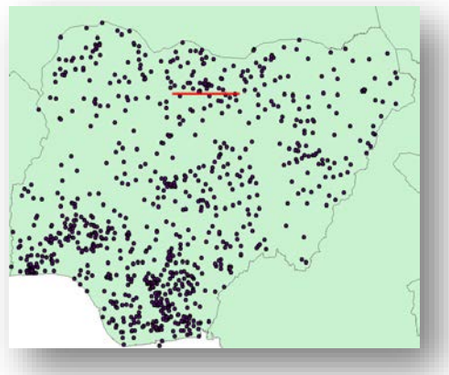
3. Incorporating spatial anisotropy

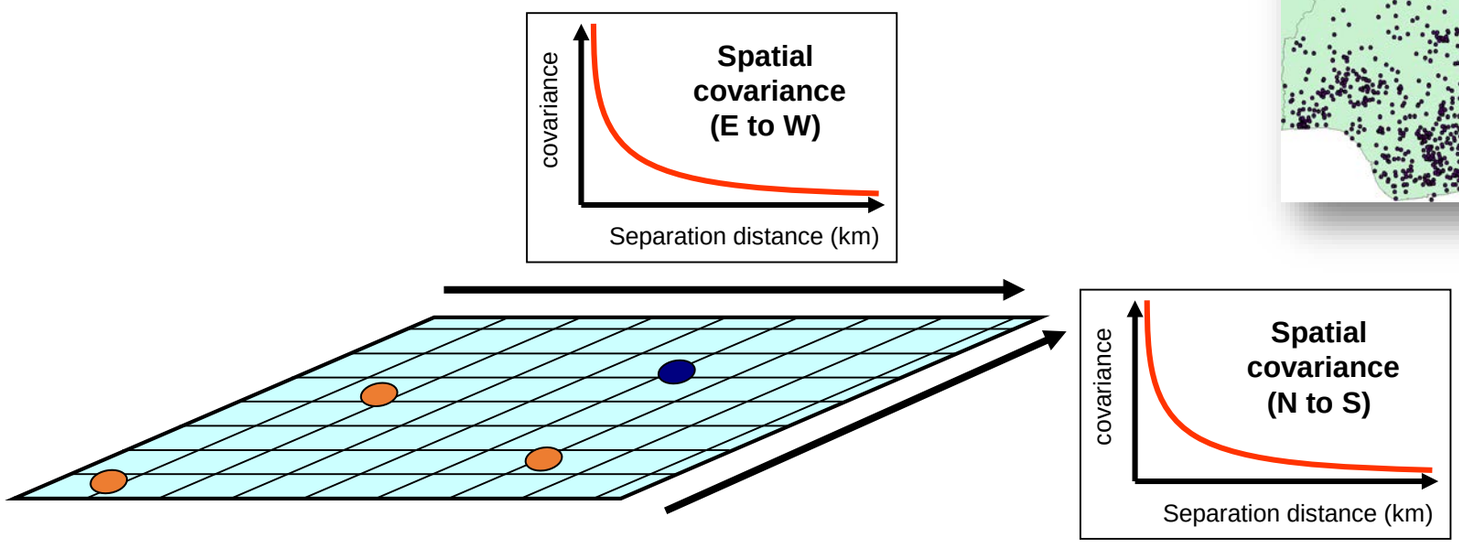


# Elaborations for demographic mapping

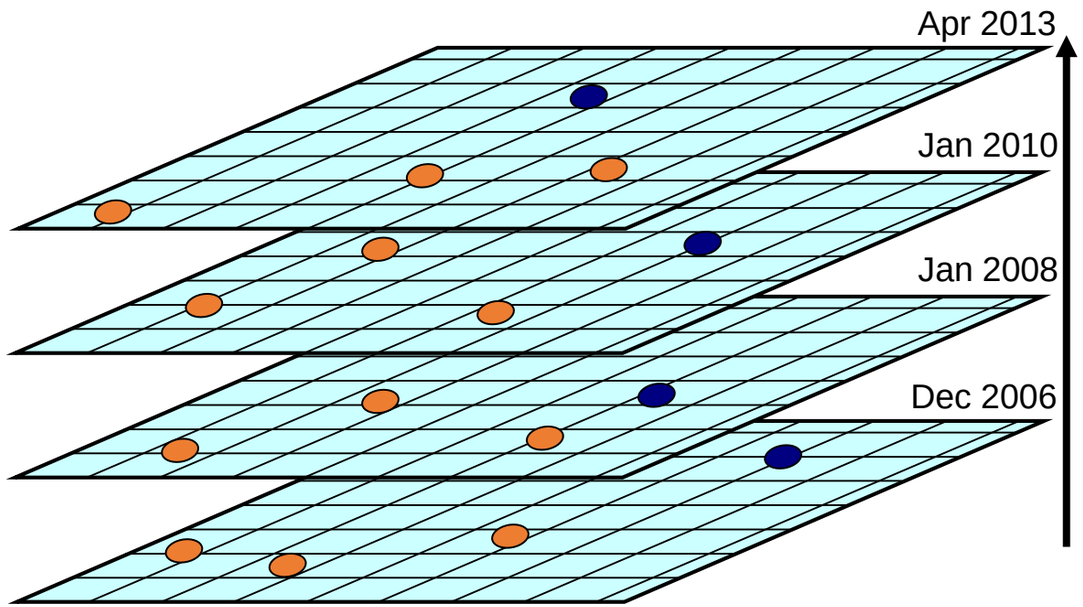
3. Incorporating spatial anisotropy

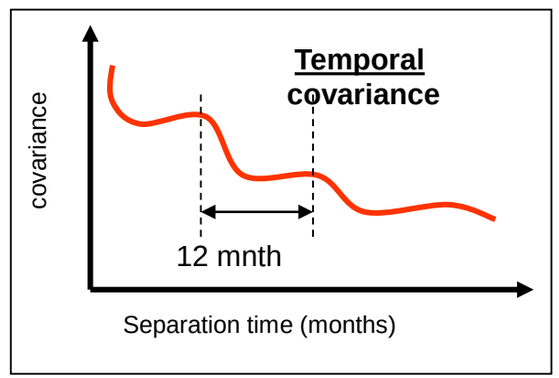




# Elaborations for demographic mapping

4. Full space-time dimensionality

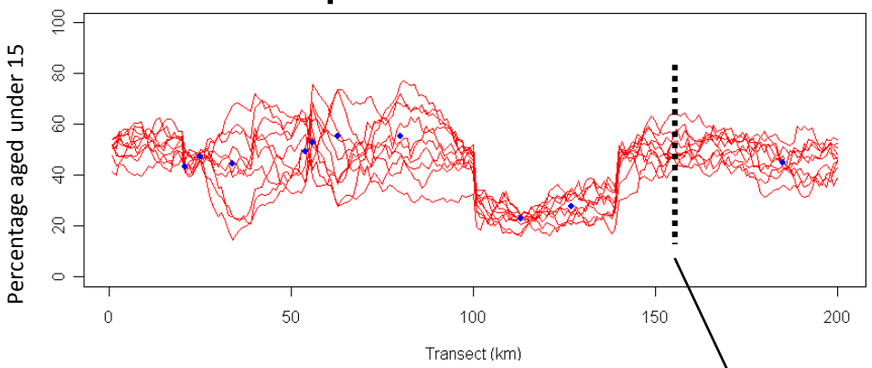


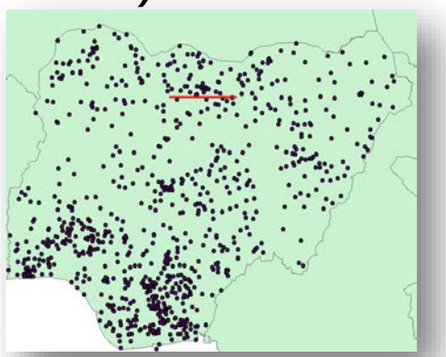




- 2d spatiotemporal covariance function: temporal axis
- Sinusoidal seasonal component…? (see mobility mapping later)

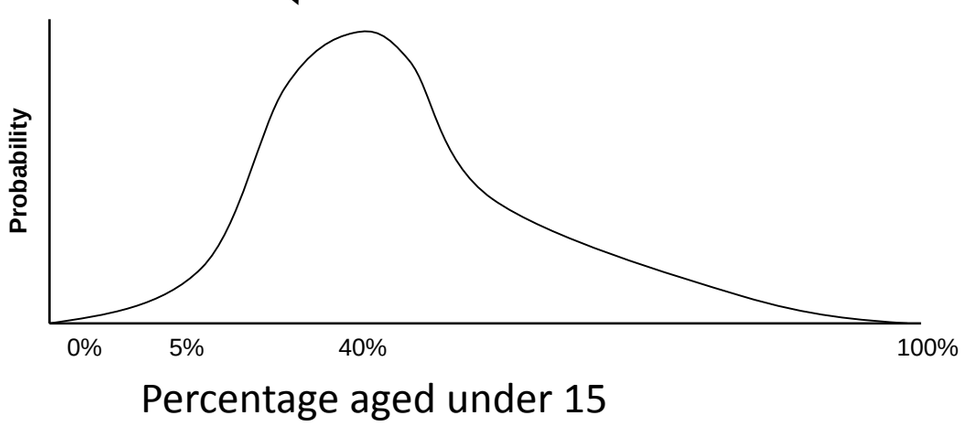
# Geostatistics: making maps

- Model output is set of realisations (e.g. $n$ = 10000)



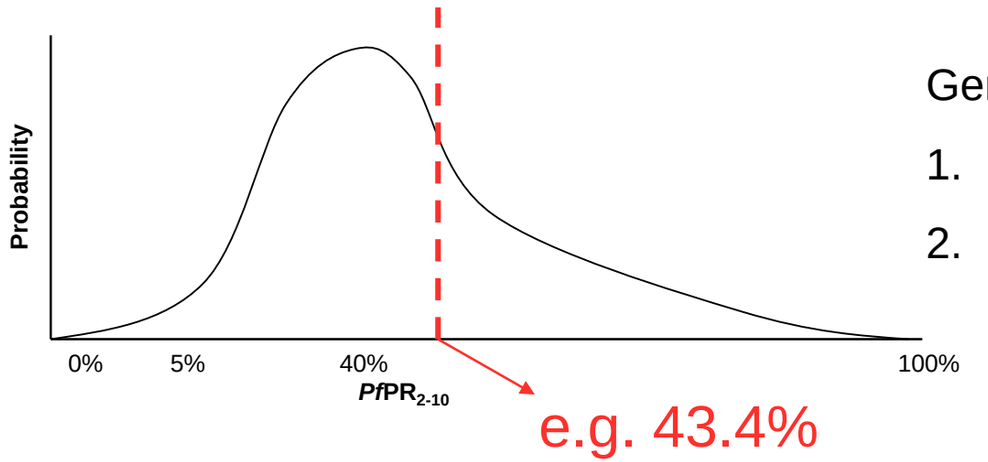



- Realisations provide posterior distribution for each location



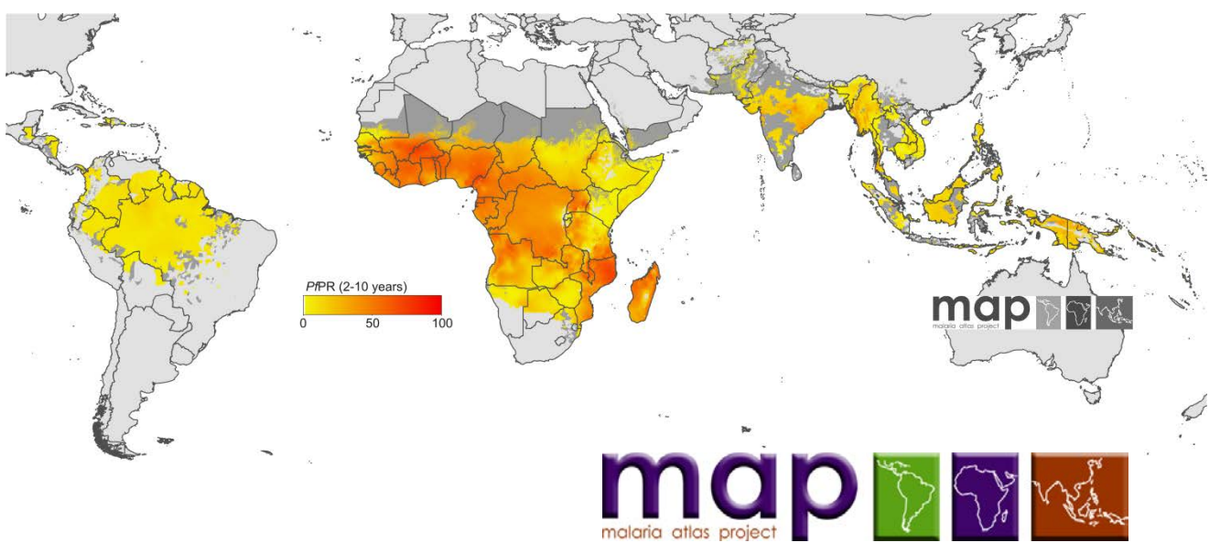
# Geostatistics: making maps

## Generating maps from posterior distributions of *Pf*PR$_{2\text{-}10}$




Generated various summaries:

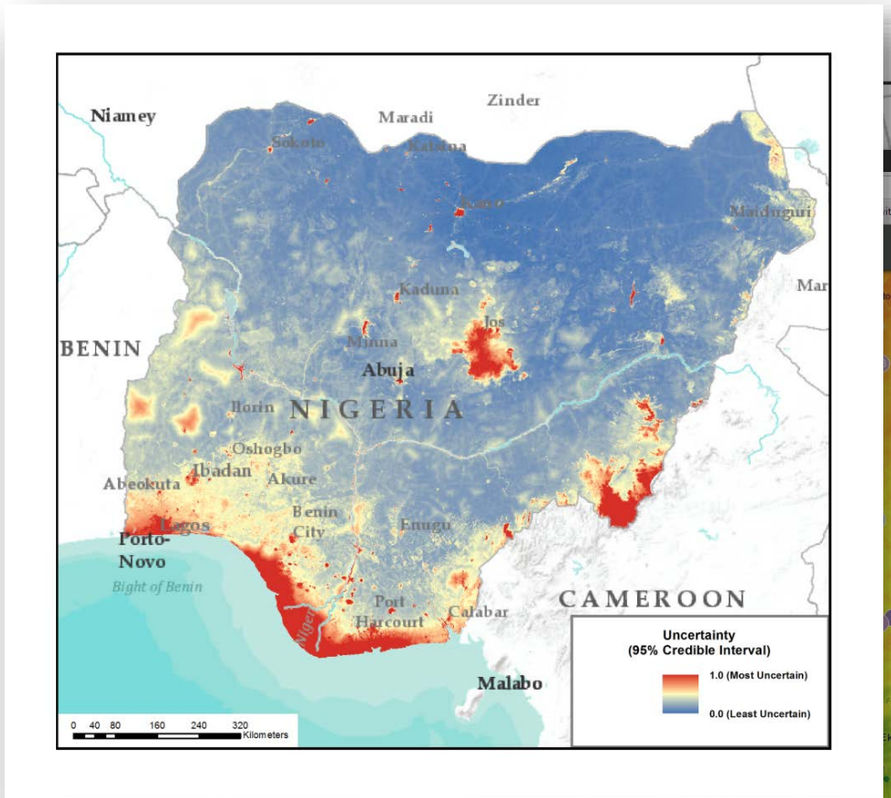
1. A ‘point estimate’ (mean)
2. A summary of uncertainty (e.g. St.deviation)

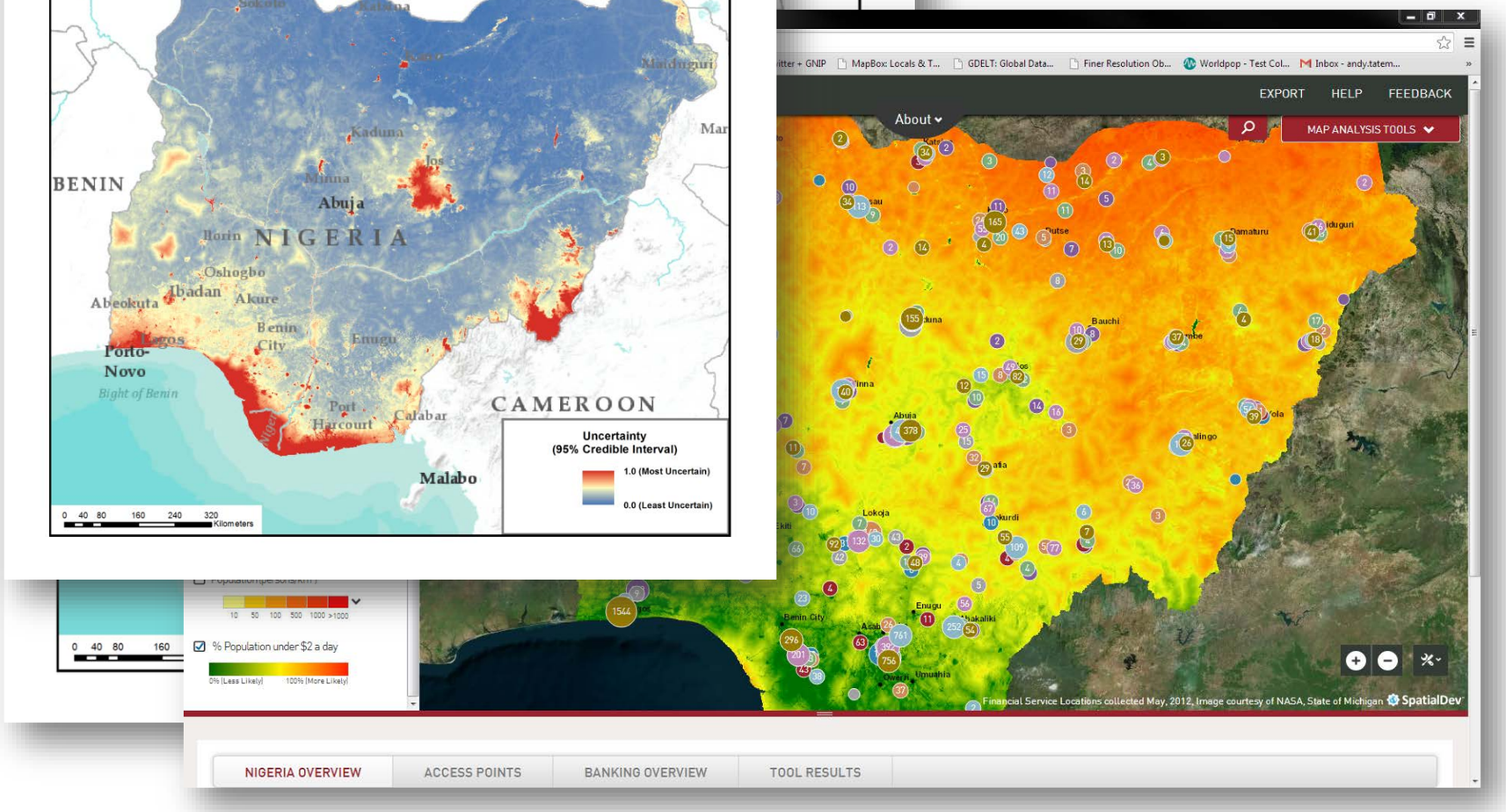


# Poverty mapping

THE WORLD BANK
Working for a World Free of Poverty

world pop



# Modeling Approach Overview