Tools for disaggregated data: geospatial analysis for mapping population demographics

*Dr Donna Clarke and Prof. Andy Tatem*
• **WorldPop**: Research program focused on methods for improving the demographic evidence base in low/middle income countries

• **Flowminder**: Non-profit foundation working with data providers and international/government agencies to operationalize and scale research in support of vulnerable populations and sustainable development

• Open data, open methods

• 60+ researchers, data scientists, PhD students, support staff based primarily in Southampton
2015-2030: 17 goals, 169 targets
LEAVE NO ONE BEHIND

“Everywhere”
Women who are literate in Nigeria

Legend:
- Orange: Up to 28.3
- Light Orange: 28.4 to 54.3
- Yellow: 54.4 to 82.0
- Red: 82.1 and higher
How can we produce subnational data?
Geographic Information System (GIS) Data
High Resolution Built-Up/Settlement Datasets

Global Human Settlement Layer

Facebook

European Commission

DLR
Mobile phone call detail records (CDRs)

![Unique subscribers in Asia Pacific](image1)

Source: GSMA The mobile economy Asia Pacific 2017
https://www.gsmaintelligence.com/research/?file=336a9db2abed95bc70e62bf7e867855&download

![Mobile phones](image2)

Source: Open Institute and the Asia Foundation 2015

![Quick facts about digital media in Thailand](image3)

Source: Nielsen. The digital media habits and attitudes of South East Asian Consumers 2011
https://wiki.smu.edu.sg/digitalmediaasia/Digital_Media_in_Thailand
Mobile phone call detail records (CDRs)

User makes a call from location X

Call routed through nearest tower

Network operator records time and tower of call for billing

User travels to Y and makes a call
Mobility: Changing densities, flows, seasonal/permanent migration

Social networks: Number of contacts, calling patterns

Consumption: Credit purchase frequencies, top-up amounts

Protecting confidentiality
- Aggregate summaries
- Regulator approval
- Raw data never leaves operator

FLOWMINDER.ORG
How can we use this data?
Gridded population data

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

Grids: flexibility in analysis and data integration

Percentage of women of childbearing age per woreda within 50km of a CEmONC population totals
Top-Down Vs Bottom-Up Approach

a) Top-down approach
Census population counts

Geospatial covariates

Population disaggregation
Spatial weighting layer created based on covariates, using dasymetric mapping

Gridded population

b) Bottom-up approach
Microcensus population counts

Geospatial covariates

Population estimation
Prediction of population in unsurveyed areas based on covariates, using statistical modelling

Gridded population

Wardrop et al., 2018 (PNAS)
Top-down disaggregation

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

Aggregate census counts

100x100m gridded counts

People per 1x1km 2017

www.worldpop.org

Random Forest-based Dasymetric Approach

- Used to disaggregate subnational census-based figures\(^\text{,}^\text{1,2}\)

\(^\text{1}\) Stevens et al, 2015; \(^\text{2}\) Sorichetta et al., 2015
Isn’t there a basic data problem in some countries?
What do we do in the absence of national population and census data?
High-resolution population mapping in Afghanistan
Afghanistan - Background

Last national population census was in 1979
Significant uncertainties in national and subnational estimates

1979 Census map

One-third of country covered by a rolling census (SDES), but insecurity preventing additional data collection

President Ghani requested exploration of new methods for obtaining subnational population numbers
Example geospatial datasets for population prediction

- Vegetation index
- Night-time lights
- Slope
- Number of compounds & area settled

>100 geospatial datasets
Model-based estimates
Our vision

“We envision a world where data analytics puts everyone on the map, ensuring that especially the most vulnerable count.”
GRID$^3$ provides support to low- and medium-income countries to collect, analyse, integrate, disseminate, and utilise high-resolution geo-referenced data for development and humanitarian decision making.
GRID³ helps produce three spatial data layers to meet critical development needs:

- **High-Resolution Population Maps**
- **Settlement Locations**
- **Subnational Boundaries**

*Images of each data layer are shown.*
GRID$^3$ is underway within five African countries:

- Nigeria
- Mozambique
- Zambia
- Tanzania
- Democratic Republic of the Congo
Parallel Session: Measuring and Monitoring the SDGs

Geospatial Innovations for Detecting and Mapping Infrastructure, Censuses and Modelled Population

Wednesday 21 November 2018
9:00 – 10:30
E303, Exhibition Centre