Use of active and passive VGI data for population distribution modelling: experience from the WorldPop project

Andrew J. Tatem
Department of Geography and Env.
University of Southampton

Forrest R. Stevens, Andrea E. Gaughan
Department of Geography and Geosciences
University of Louisville

Nirav Patel
Department of Geography and Geoinformation Science,
George Mason University

Zhuojie Huang
Center For Infectious Disease Dynamics
Penn State University

Catherine Linard
Biological Control & Spatial Ecology
Université Libre de Bruxelles

September 2014
Role of Volunteered Geographic Information in Advancing Science: Effective Utilization
Open access archive of spatial demographic datasets

Current focus: Central and South America, Africa and Asia

Transparent and shareable methods

Support development and health applications

www.worldpop.org.uk
The WorldPop project was initiated in October 2013 to combine the AfriPop, AsiaPop and AmeriPop population mapping projects. It aims to provide an open access archive of spatial demographic datasets for Central and South America, Africa and Asia to support development and health applications. The methods used are designed with full open access and operational application in mind, using transparent, fully documented and shareable methods to produce easily updatable maps with accompanying metadata.

Latest Map

Affiliation: WorldPop Project
Intro to gridded population data

Census data linked to GIS administrative boundaries

Ancillary data e.g. settlements, roads

Spatial modelling rules to disaggregate census counts

Estimates of number of people in each grid cell

Population Density (pph)

Cambodia Population (people per hectare)
Intro to gridded population data

- Census data linked to GIS administrative boundaries
- Ancillary data e.g. Settlements, roads
- Spatial modelling rules to disaggregate census counts
- Estimates of number of people in each grid cell

Use as ‘synthetic’ populations for spatial models, or to derive populations at risk metrics
Flexible ‘random forest’ method

Net primary productivity

Health facilities

Roads

Land cover

Net primary productivity

Health facilities

Roads

Land cover
Predicted population distribution in Kenya

Kenya, Final Map Output

People Per Pixel

High: 5
Low: 0
1:5,000,000
Active VGI: OpenStreetMap

**Vietnam Predicted Population Density**

- **High**: 25
- **Low**: 0
Passive VGI: Twitter

Indonesia Predicted Population Density

- High: 50
- Low: 0
Q1. How is the spatial component, such as scale and accuracy, of VGI being handled in scientific or operational applications?

- As any other spatial dataset:
  - Applications must be adapted to the scale of data
  - Accuracy must be evaluated

- Specificity of VGI data:
  - Heterogeneous level of spatial detail and accuracy
Q2. How are the limitations of VGI being dealt with in an applied spatial modeling environment?

- Active VGI
  - Data of obvious insufficient quality/uniformity are removed
Q2. How are the limitations of VGI being dealt with in an applied spatial modeling environment?

- **Active VGI**
  - Data of obvious insufficient quality/uniformity are removed
  - The performance of covariates to predict the spatial distribution of human populations is evaluated in the modelling process
Q2. How are the limitations of VGI being dealt with in an applied spatial modeling environment?

- Passive VGI (GeoTweets)
  - Privacy, confidentiality
    - Use of aggregated data
  - Biases
    - More detailed analyses and data calibration needed
Q3. What would be one specific aspect that you consider missing or under-represented in the current 'VGI landscape'?

- Better inform on temporal changes:
  - include year/date of data encoding?

- A way to evaluate the representativeness:
  - A map of contributors’ distribution?
Q4. Does VGI have a potential to impact governance?

- **YES**
  - Integration of data provides richer decision-making resources
  - Especially:
    - In developing countries where information can be incomplete or unreliable
    - At the local level
    - Collaborative governance

- **But...**
  - Control needed
  - Integration of authoritative and VGI data