Global Population Distribution
a continuum of modeling methods

Greg Yetman

Kytt MacManus

CIESIN, Columbia University

Center for International Earth
Science Information Network
EARTH INSTITUTE | COLUMBIA UNIVERSITY

Center for International Earth Science Information Network

EARTH INSTITUTE | COLUMBIA UNIVERSITY

Focus on applying *information technology*
to pressing *interdisciplinary data, information, and research problems*
related to *human interactions in the environment*

Applied Research in Global Gridded Population Mapping

- Greater demand for detailed mapping for multiple applications across country boundaries
  - Health, natural hazards, SDGs, communications infrastructure planning, many others
- Increased resolution in time and space is required
- Additional variables being used
 - For example, age structure in health and hazard assessments

Modeled Population Surfaces

- "Explosion" in projects: GPW, Landscan, WorldPop, GHSL, GUF, Esri, Facebook Connectivity Lab

- Different models with varying levels of complexity

- All use statistical units as a base and distribution population on a surface

Model Approaches

- Proportional allocation: equal distribution
  - Simple, uses few inputs
- Dasymetric (cartographic/map modeling)
  - Distributes population based on weights from multiple inputs
- Statistical/machine learning—random forest, neural networks
  - "Big Data" approaches to spatial modeling, can be complex to understand
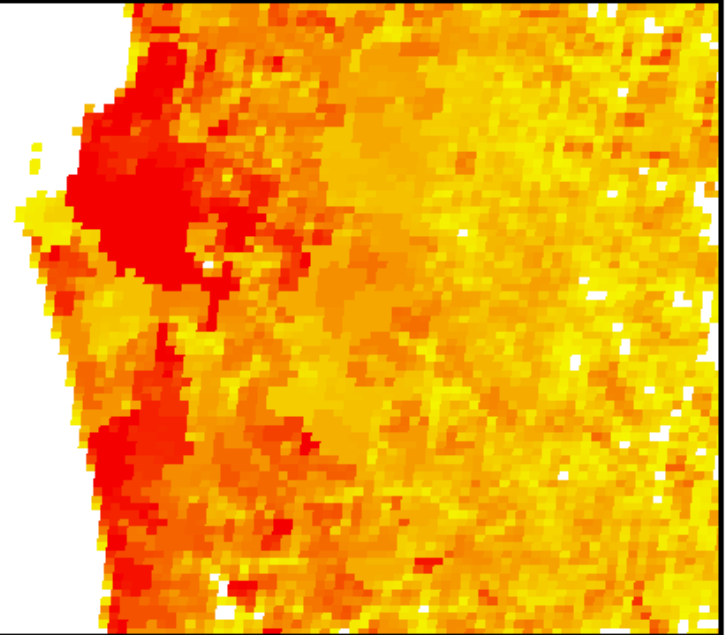 - Random forest model chooses most applicable input variables

Examples from Sri Lanka
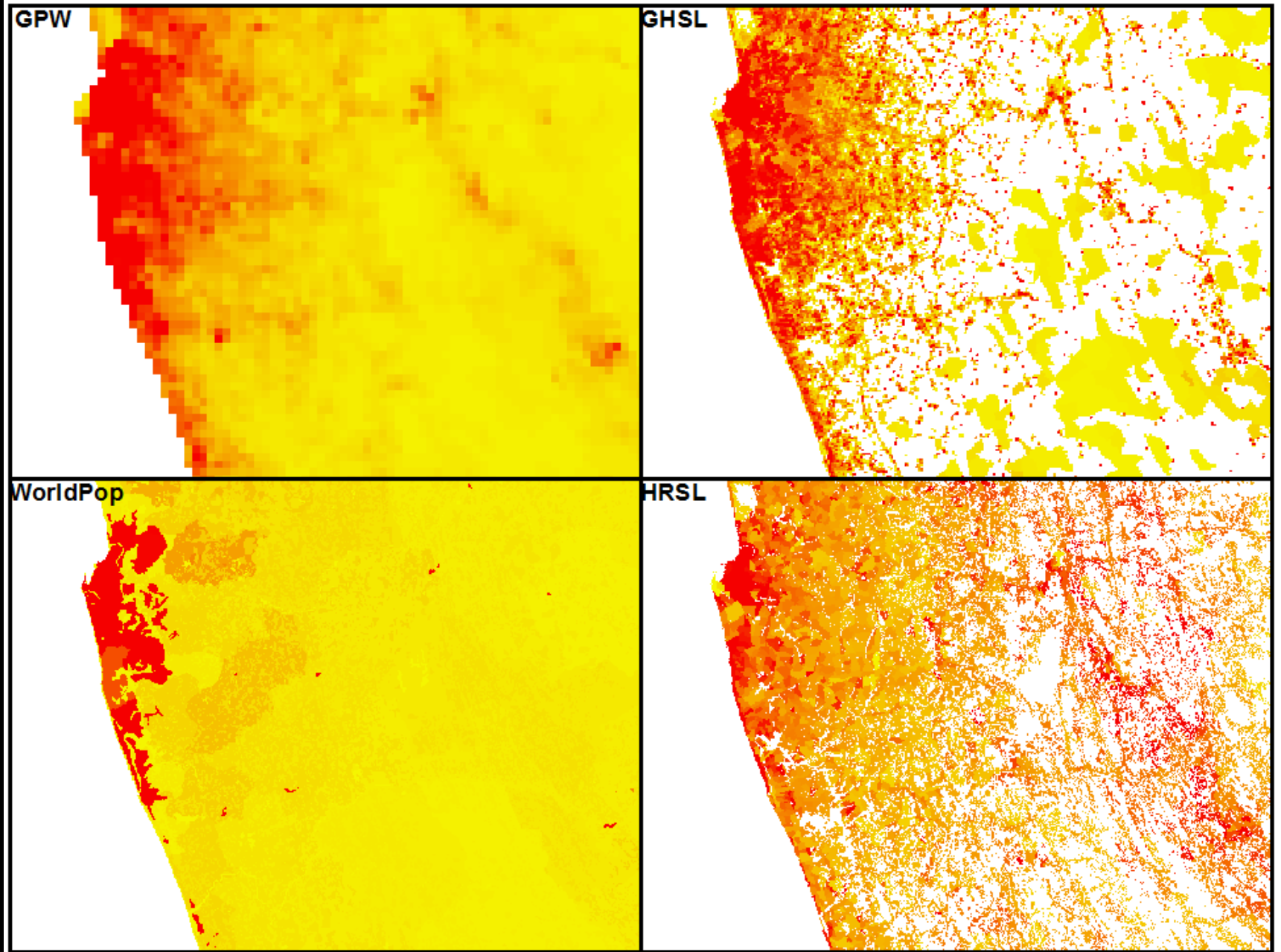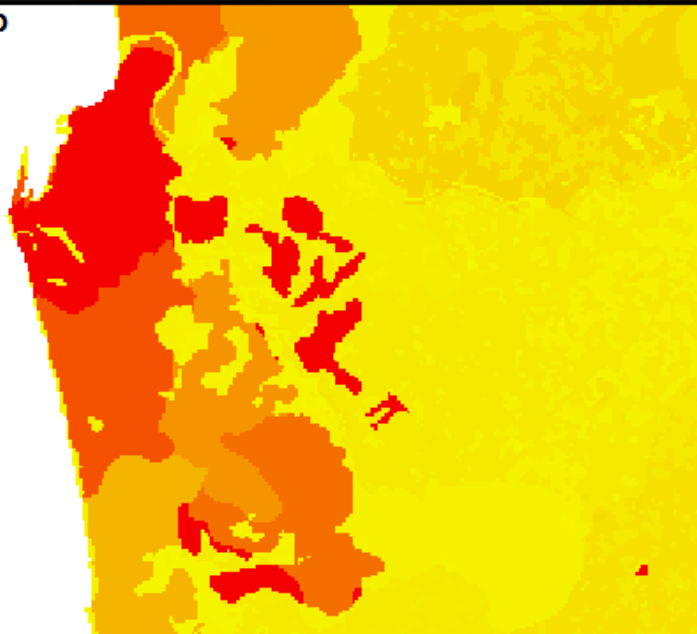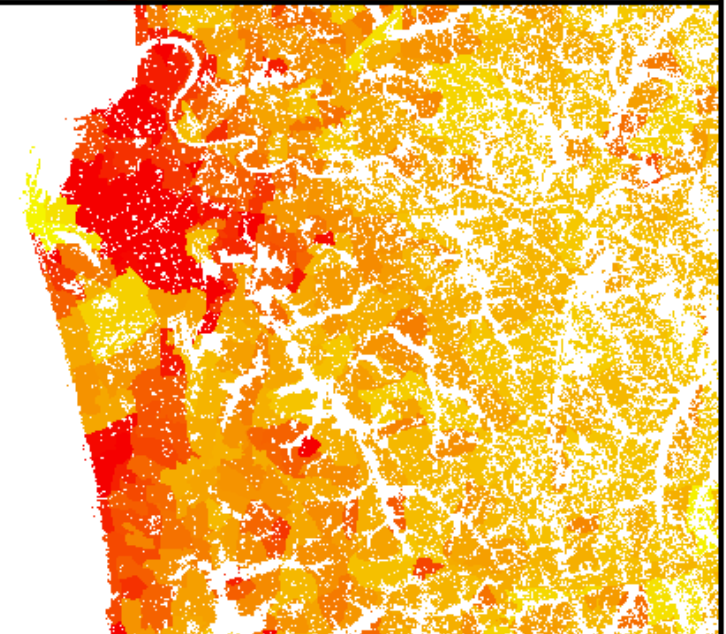
Examples from Sri Lanka (Colombo)
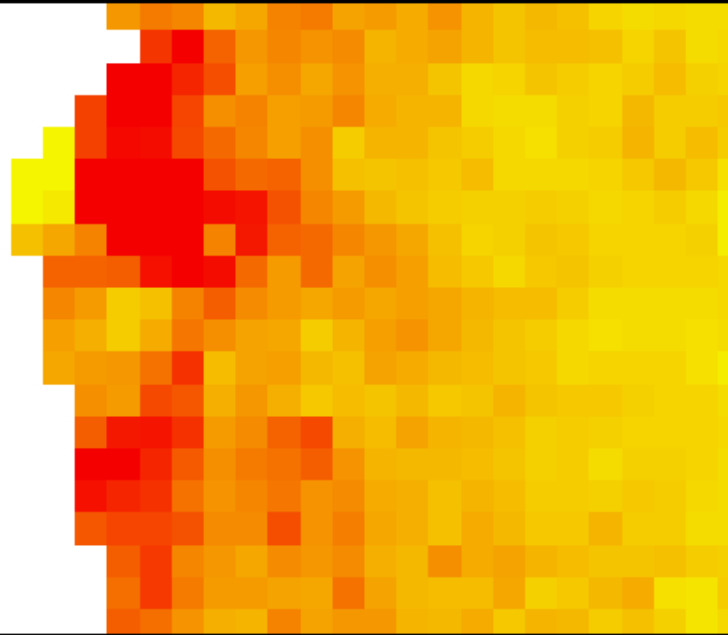
Model Inputs

Project	Census	Water Bodies	Roads	Distance to coast	Land Use / Cover	Topography measures	Imagery / spectral data	RADAR	Buildings /settlements
GPW	✔	✔							
Landscan	✔	✔	✔		✔	✔	✔		✔
WorldPop	✔	✔	✔✳	✔✳	✔✳	✔✳			✔✳
GHSL	✔	✔			✔✔		✔✔		
GUF	✔							✔	
Esri	✔	✔	✔		✔	✔			
HRSL	✔						✔		

✔ Fixed in time   ✔✔ Time series   ✔✳ Use varies based on model

Model Type and Outputs

Project	Proportional Allocation	Dasymetric	Statistical / machine learning	Time Series	Explicit (per pixel) quality measures	General quality measures
GPW	✔			✔	✔	✔
Landscan		✔		✔		✔
WorldPop			✔	✔ *	✔	✔
GHSL			✔	✔	✔	✔
GUF			✔			✔
Esri	✔					✔
HRSL			✔			✔

* Exists for some countries, planned for WorldPop Global

Implications for Data Use

- Additional input variables limits usefulness of population as an independent variable
  - Associations with slope, land cover, water bodies, etc. can be problematic
- Administrative boundary data cause problems in some types of analysis
  - Population in/near protected areas
  - Underrepresented or poorly sampled populations

Implications for Data Use

- Dizzying array of population products to choose from with little guidance
  - Which one is "best"?
- Spatially explicit quality measures (where available) seldom used in analysis
 - Examples and guidance on techniques for use may help

New Data Sources

- Social media and mobile phone data
  - Georeferenced tweets, photos, Flowminder data
- New and more frequent satellite observations
  - Micro satellites such as Terra Bella provide frequent updates
- Volunteered Geographic Information
  - OpenStreetMap, Ushahidi, commercial equivalents

Big Data have Problems

- Selection bias
- Heterogeneous measurements—data rich in large urban areas vs. rural coverage gaps
  - phone/Internet coverage gaps
  - little volunteer effort for mapping small communities

Twitter Density from [Mapbox](#)
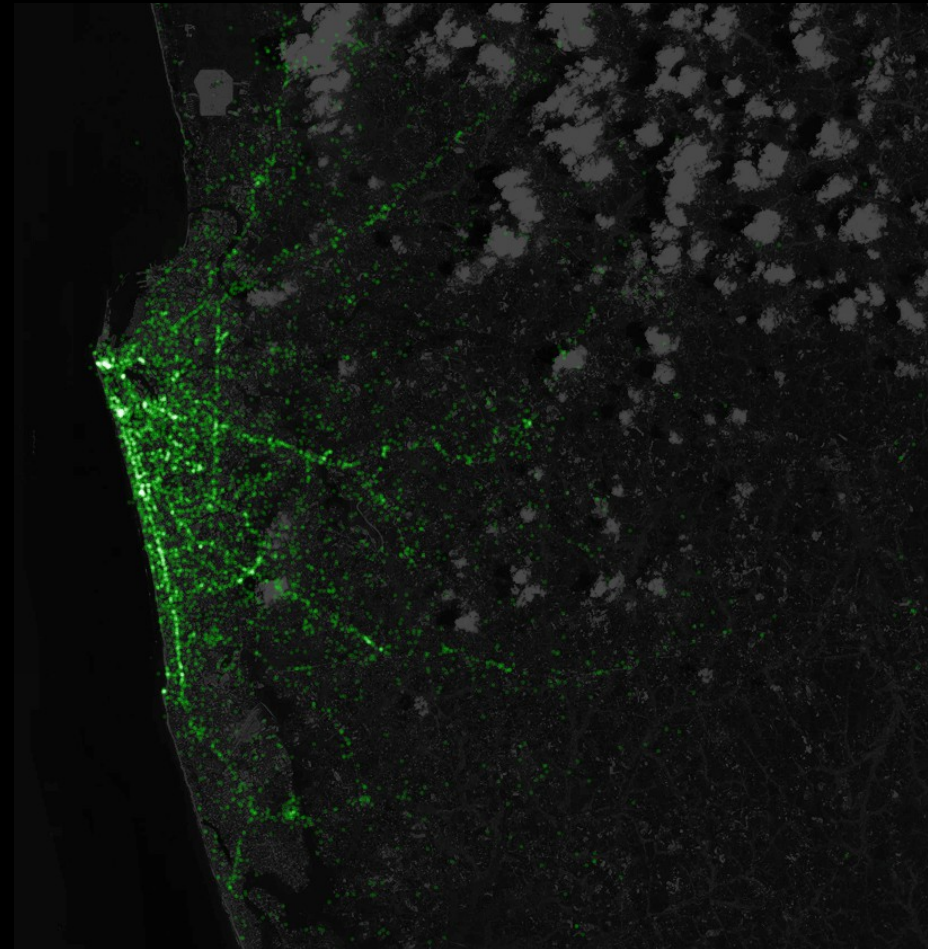
Combining "old data" with new

- Reliance on census data is already problematic for some under-represented populations and in areas with only older data
  - Survey data and estimates are traditional supplements
  - Many recent and planned census enumerations are more amenable to gridding
- Producing a time series of census data at detailed levels is very labor-intensive
- Adding new data streams and models to estimate population may be the only feasible way to increase resolution in space and time with acceptable accuracy

Thanks!

Questions?

Greg Yetman

gyetman@ciesin.columbia.edu

Kytt MacManus

kmacmanu@ciesin.columbia.edu