Gridded Population of the World Version 4 (GPWv4)

Greg Yetman
Chris Small
Erin Doxsey-Whitfield
Kytt MacManus

May 28, 2015
Global Urbanization Workshop
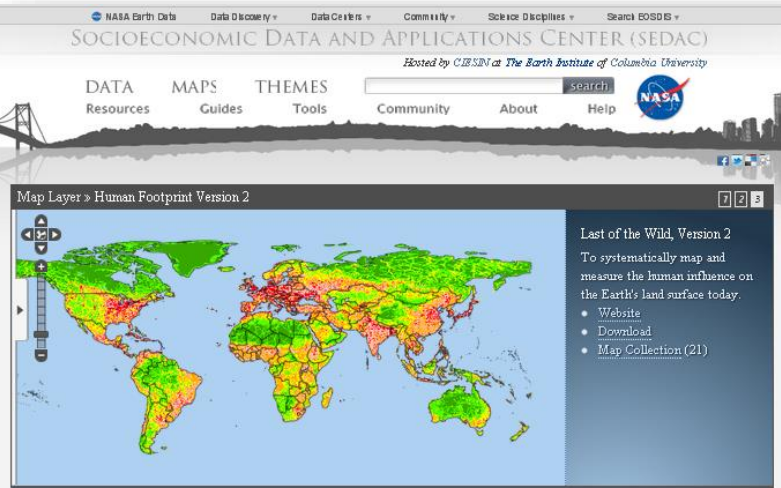
http://sedac.ciesin.columbia.edu

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*

Socioeconomic Data & Applications Center (SEDAC)

- Focus on human dimensions of environmental change

- Integration of social and earth science data, especially with remote sensing

- Direct support to scientists, applied and operational users, decision makers, and policy communities
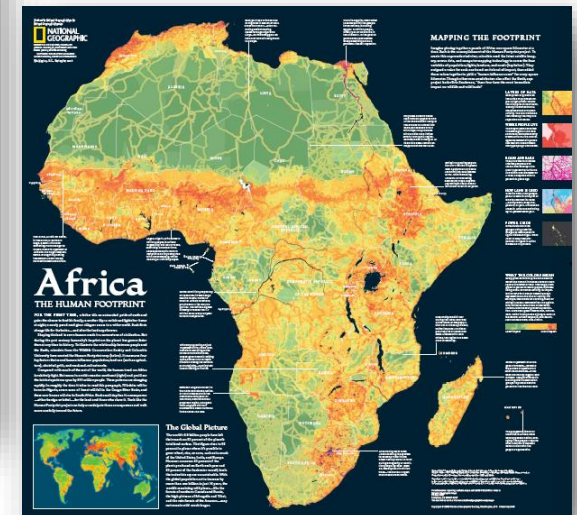
Gridded Population of the World

- Gridded (raster) data product developed to provide a **spatially-disaggregated population surface** that is compatible with data sets from social, economic, and Earth science fields.

- Census population data are transformed from their native spatial units to a global grid of quadrilateral latitude-longitude cells (Balk et al. 2010)

- Free and openly available
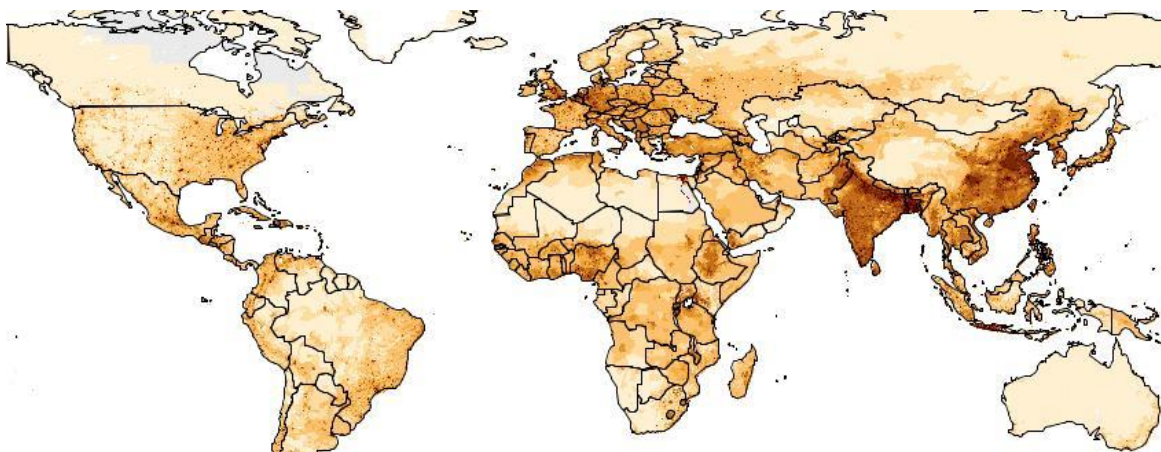
GPW version 3, 2000 population density
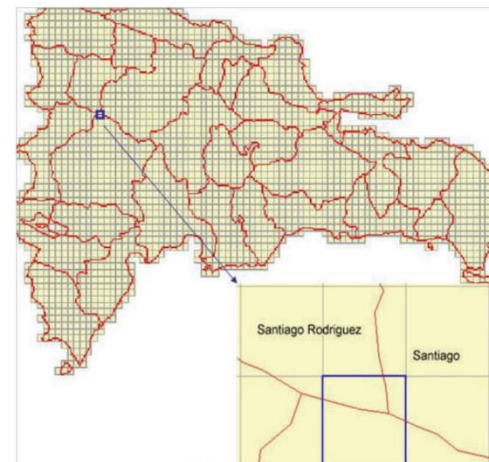
Transforming census units to a grid

History of GPW

- GPWv1 was an outgrowth of a Global Demography Workshop held at CIESIN in 1994

- Consensus that a consistent global database of population totals in raster format would be invaluable for interdisciplinary study (Deichmann et al., 2001)

- Produced by Waldo Tobler, Uwe Deichmann, Jon Gottsegen, and Kelly Maloy at the UC Santa Barbara

Development of GPW

	Publication Year	Years of Estimation	Grid Resolution	Number of Input Units (subnational geographic units)	Census variables	Population Density Grid
GPWv1	1995	1994	5 arc-minutes	19,000	Total Population	1994
GPWv2	2000	1990, 1995	2.5 arc-minutes	127,000	Total Population	1995
GPWv3	2005	1990, 1995, 2000	2.5 arc-minutes	~ 400,000	Total Population	2000
GPWv4	2015	2000, 2005, 2010, 2015, 2020	30 arc-seconds (1 km)	~ 12,500,000	Total Population, Sex, Age, Urban/Rural status	2010

Basic Workflow Diagram

Find tabular population counts and geographic boundaries

Match to geographic boundaries (census or administrative)

Adjust boundaries to global framework and mask inland water

Estimate the population for target years (2000, 2005, 2010, 2015, 2020)

Adjust estimates to UN World Population Prospects for target years

Proportionally allocate population to 1 km grids using an areal-weighting method

Census Statistics

- Although tabular census data is available in most countries, the format and structure of the data varies.

Primary method of census data dissemination	Percent of respondents (121 countries)
Paper publications	52%
Static web pages (html, excel, PDF)	28%
Interactive online databases	14%
CD/DVD	4%
Other	2%

Source: 2011/2012 survey for the review of the 2010 World Programme on Population and Housing Censuses; UN Statistics Division, 2013.

Acquisition Methods in GPW v4

1. Many countries were acquired through basic web research

2. Other countries required more intensive personal communications (email, phone)

3. Still others could not be acquired without the assistance of in-country contacts and networks

4. In some other cases the data were licensed through purchase agreements

Evaluation of Census Data Quality

- GPWv4 will include national level metadata and other qualitative indicators which will inform users of:

  1. The currency of the population data
  2. Whether the data is from estimate, preliminary, or final census
  3. Coverage errors evaluated in post-enumeration surveys

Boundary Matching

BRA Level 5 Boundaries From IGBE

GADMv2

Santana do Livramento

Rivera

URY Level 3 Boundaries from INE
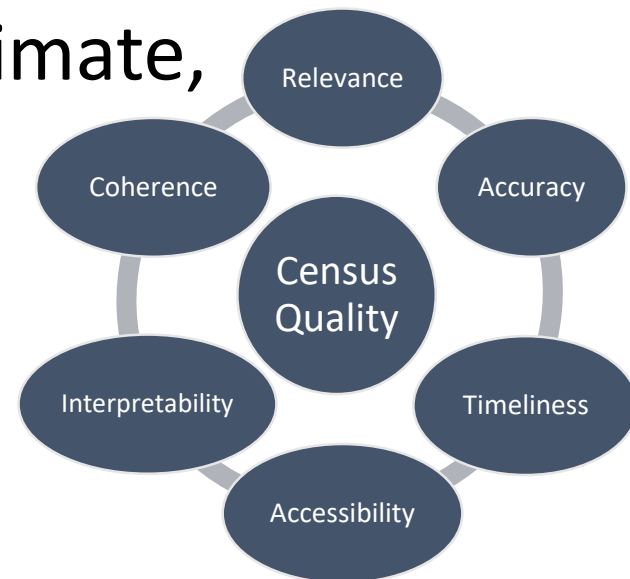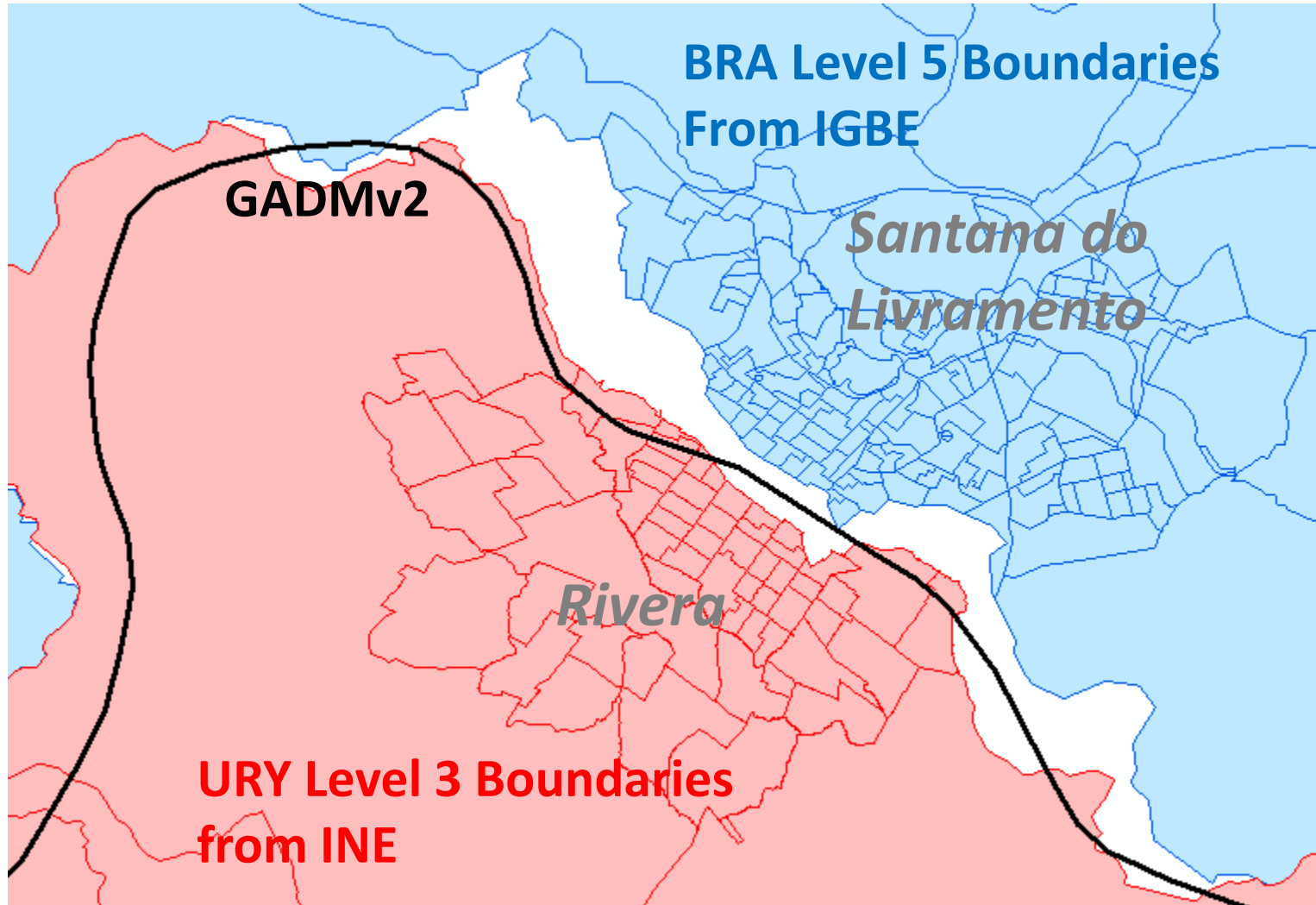
Data Quality Indicators: Water Mask

Winkel Triple Projection

Map Credit: CIESIN Columbia University, December 2014.

Water Mask — Total Water Pixels — Partial Water Pixels
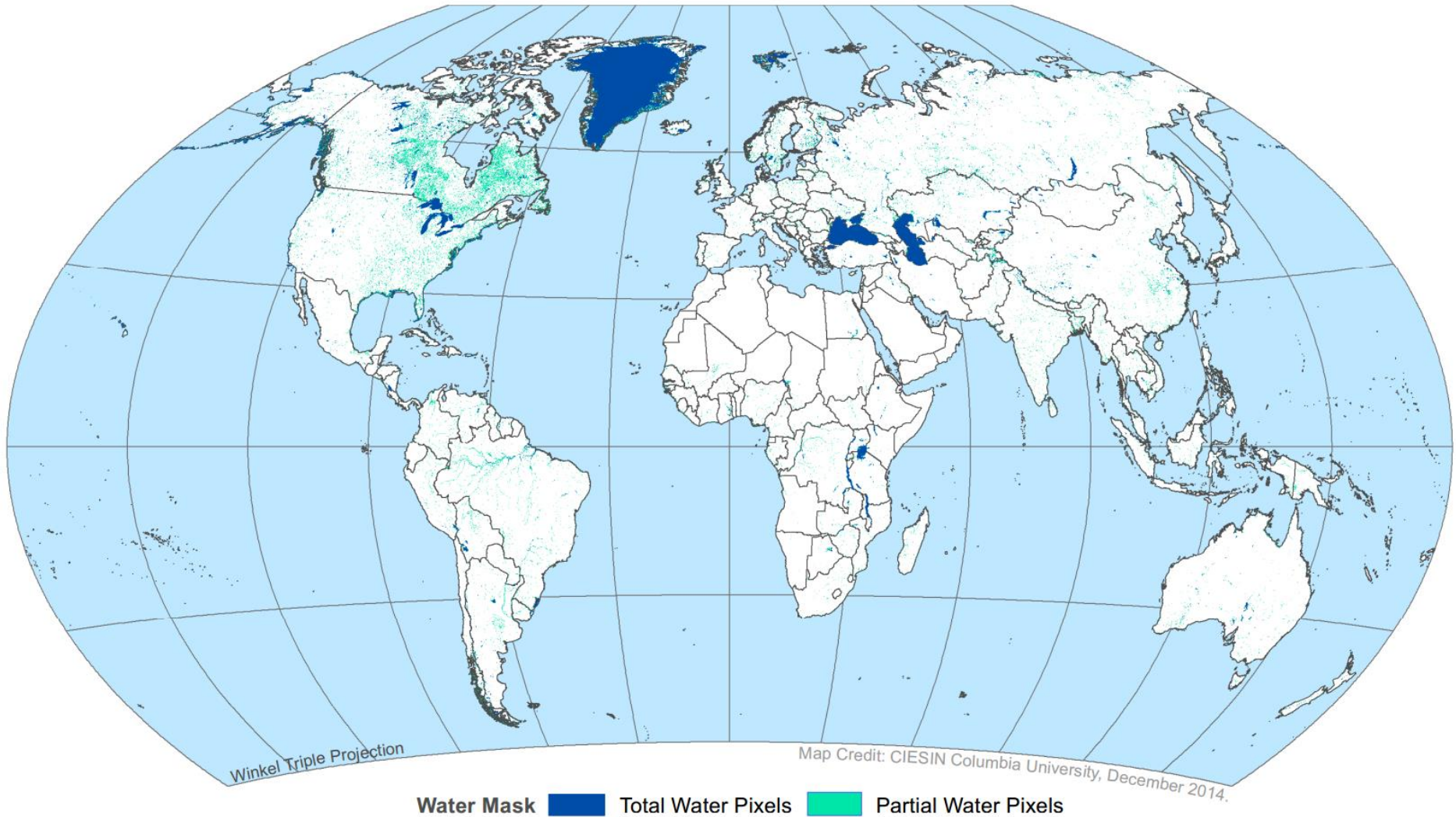
Distinguishes between pixels that are completely water and/or ice (Total Water Pixels) and pixels that also contain land (Partial Water Pixels).

Data Quality Indicators: Data Context

Uninhabited	Military District	Park or Protected Area	
No Households	Not Enumerated	Unknown	

Winkel Triple Projection

Map Credit: CIESIN Columbia University, December 2014.
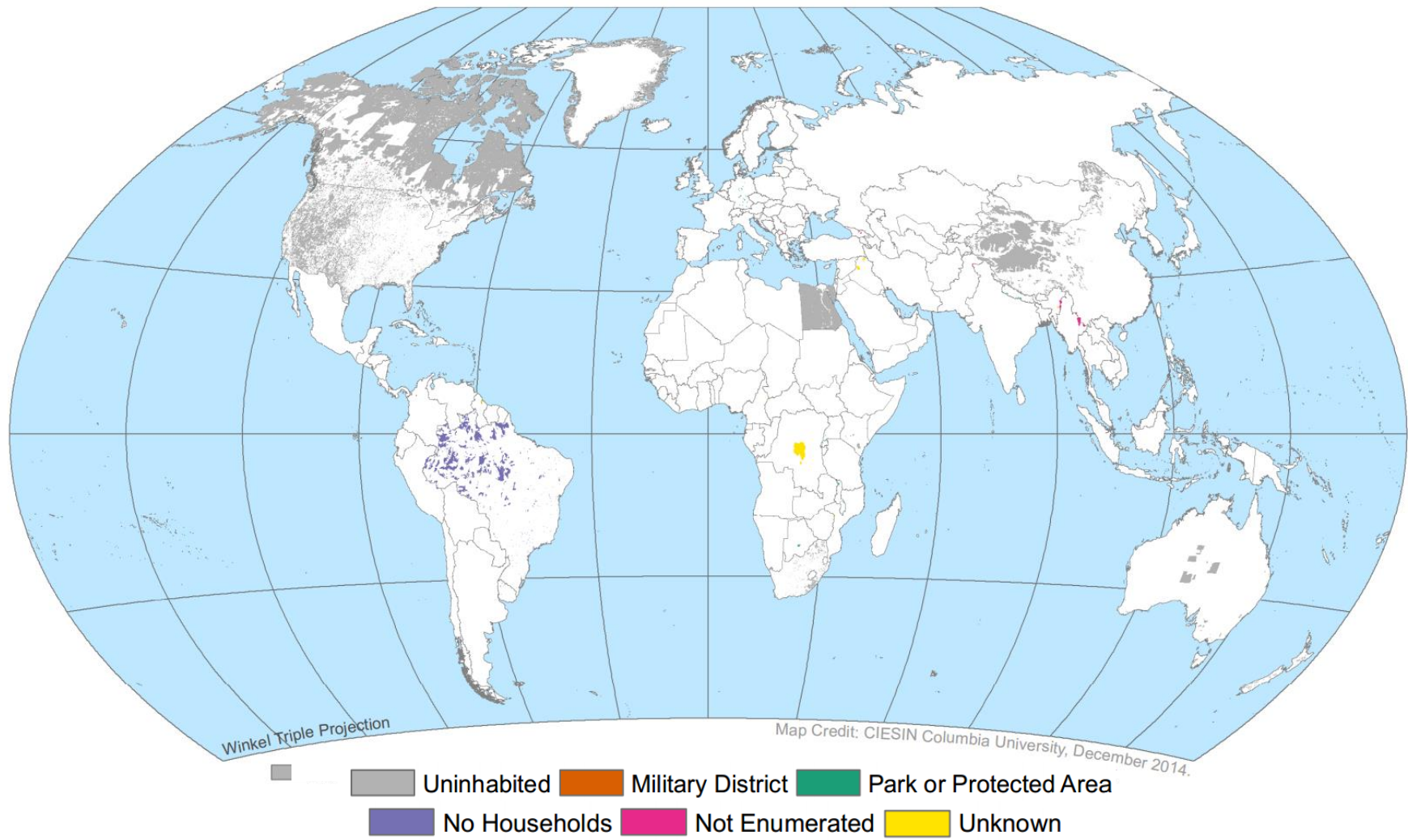
Categorizes pixels with a 0 population estimate in the population count and density grids, based on information included in the census documents.

Additional data sets
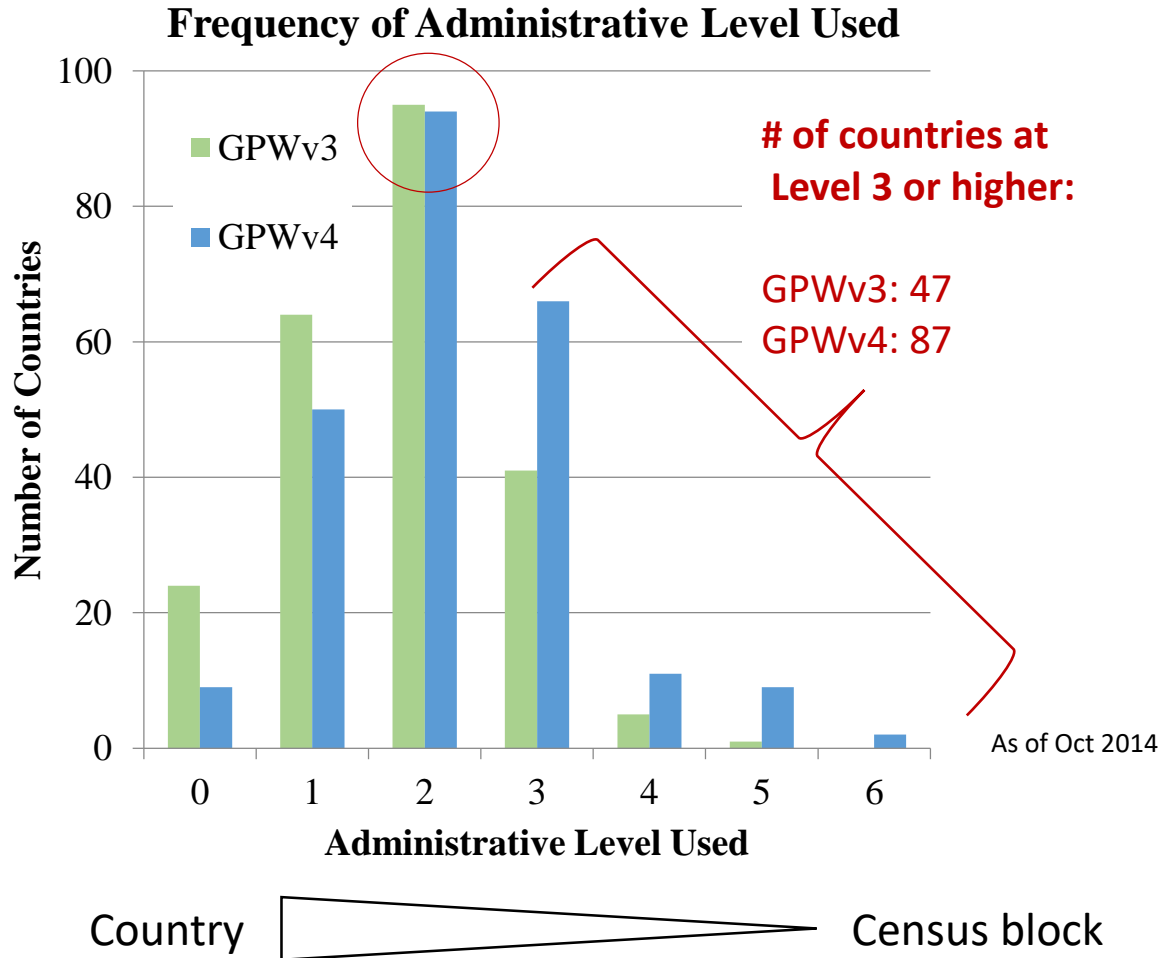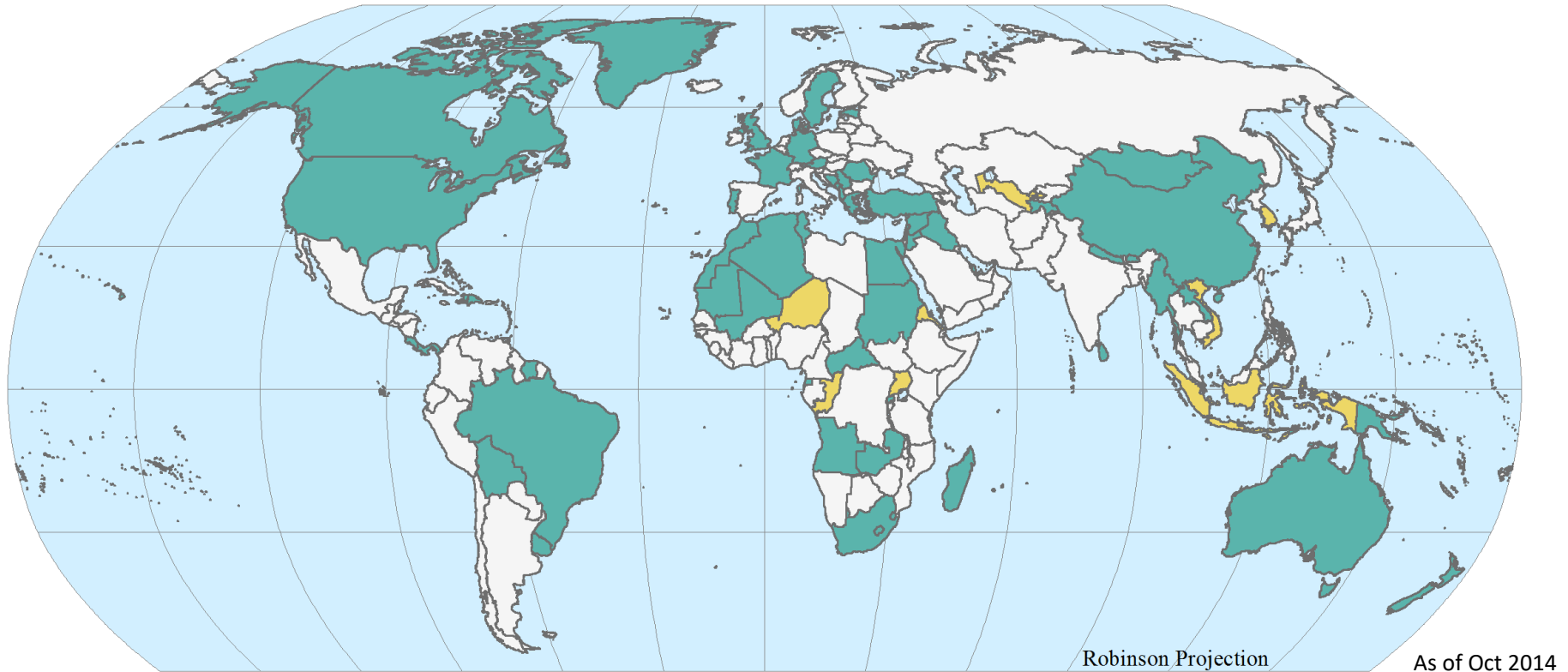
- Data Quality Indicators: Mean Admin Unit Area
  - The mean unit area data provides a quantitative surface that indicates the size of the input unit(s) from which population count and density grids are derived.

- State/Territory Identifier Grid
  - The identifier boundary data are derived from the pixels as polygons and thus have rectilinear boundaries at large scale.

- Centroids
  - Estimates of human population counts and densities for the years 2000, 2005, 2010, 2015, and 2020 by administrative unit centroid location.

Increased resolution of input census units

Frequency of Administrative Level Used

# of countries at
Level 3 or higher:

GPWv3: 47
GPWv4: 87

As of Oct 2014

Country — Census block

Improvements in Input Data Resolution

Robinson Projection

As of Oct 2014

| Lower resolution than GPWv3	Same as GPWv3	Higher resolution than GPWv3
11 countries	132 countries	98 countries

Addition of gridded census variables

Demand to include demographic information in global population grids

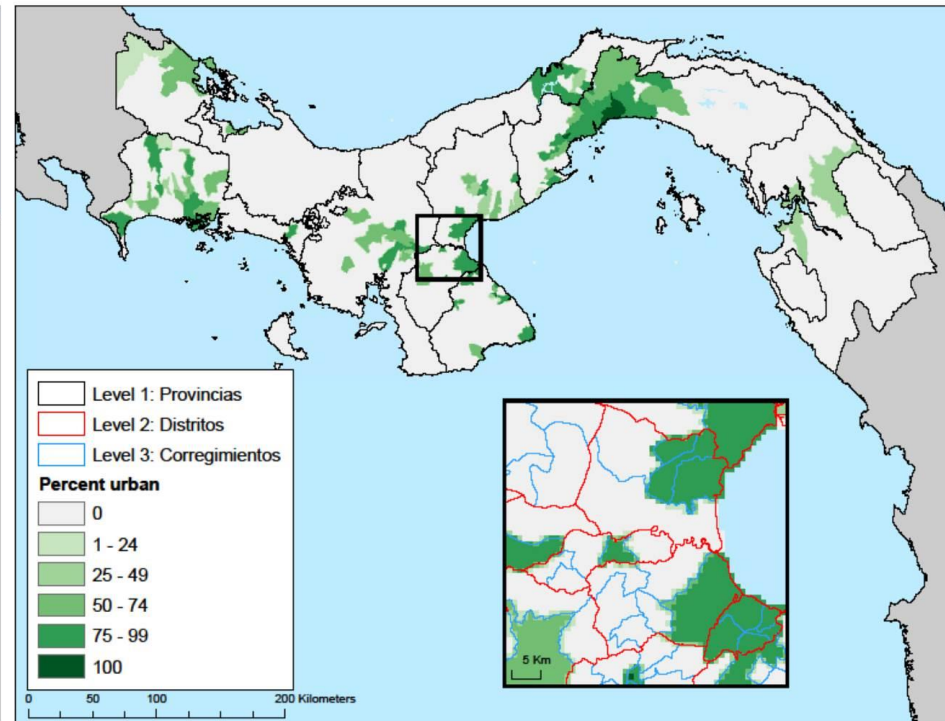**GPWv4 will include detailed global grids for:**
- Sex
- Age 5-year age groups
- Urban/Rural status

Addition of gridded census variables

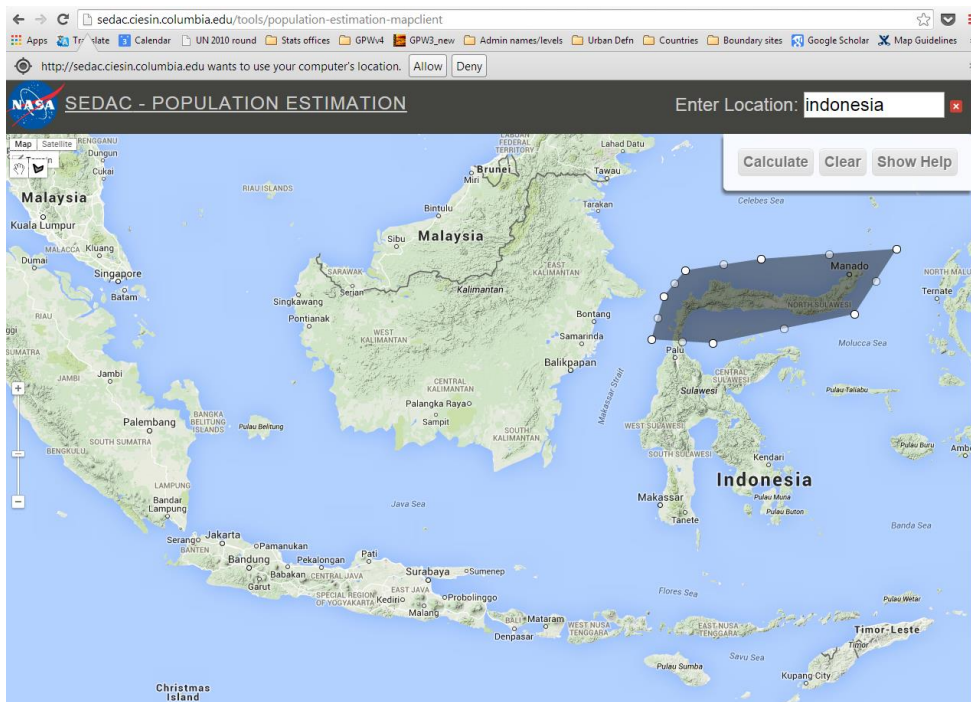
% Rural population, 2010, Panama
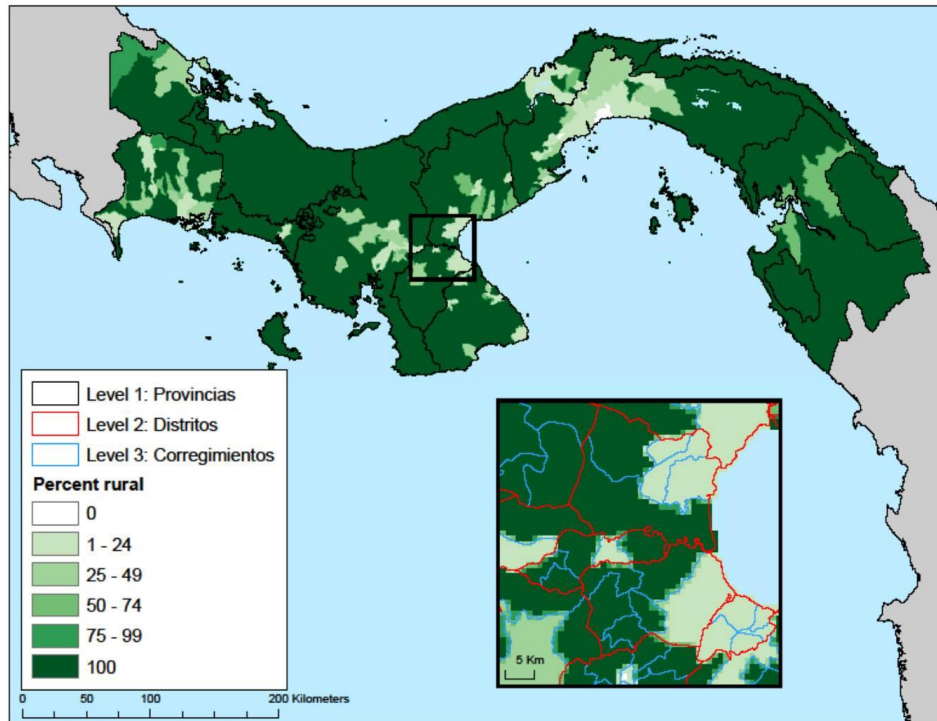
% Urban population, 2010, Panama
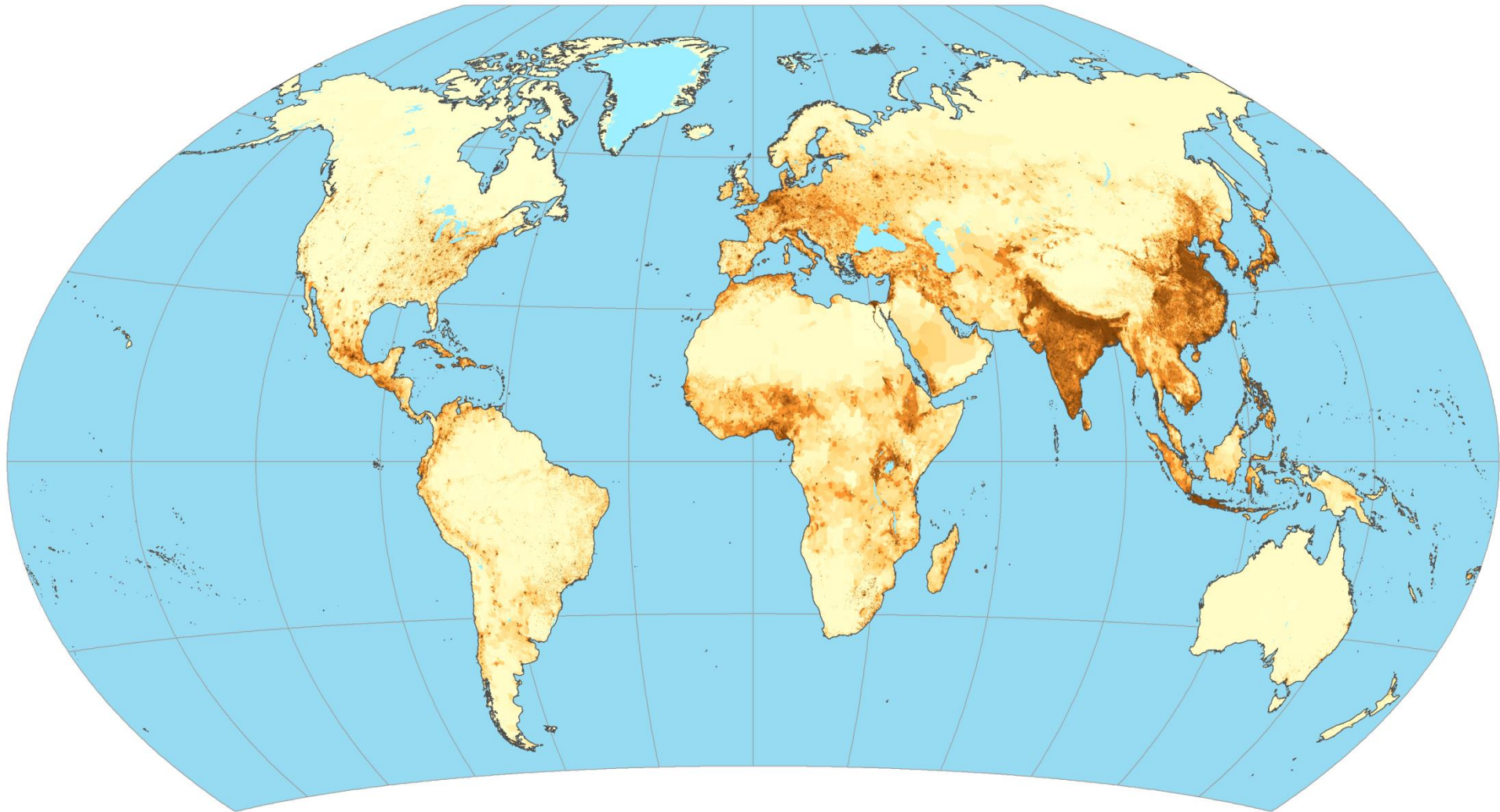
SEDAC Population Estimation Service

- Web-based tool for estimating population totals and related statistics from a user-defined region
- Uses GPWv3 dataset for the year 2005; will be updated with GPWv4

GPWv4: Population Density

Additional Modeling: GHSL

- Integration with GHSL via a joint project with JRC
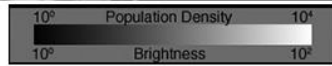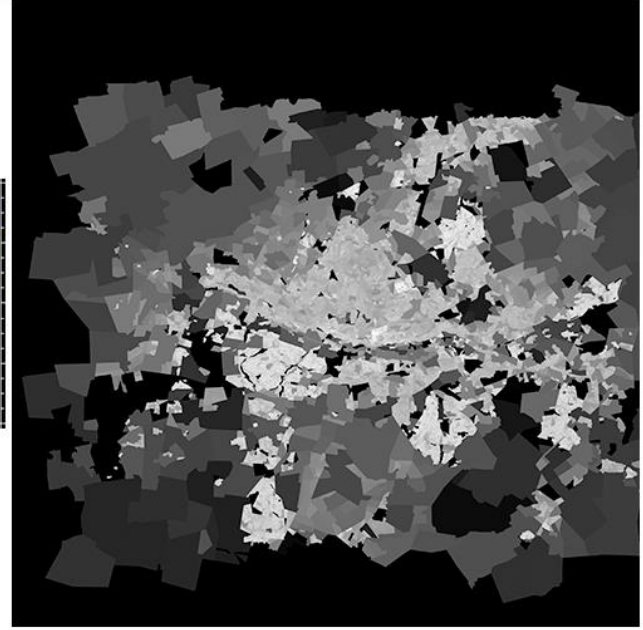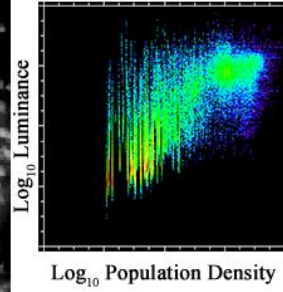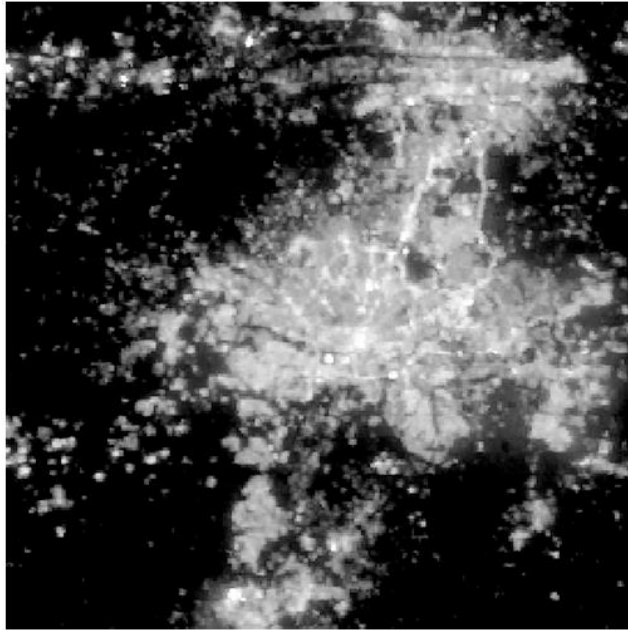   - Details to follow in 2016

Additional Modeling: VIIRS

- Day-Night band from the joint NASA-JAXA Suomi NPP satellite
- Detects lights with greater sensitivity than DMSP at a finer resolution (~300m)
- Working with beta lights composites
  - Final annual composites of stable lights scheduled for October 2015
- Relationship between high resolution census data (population density) and emitted lights
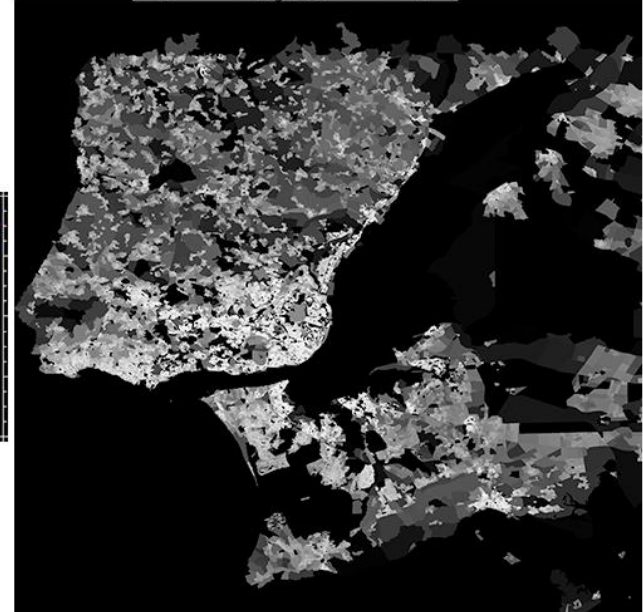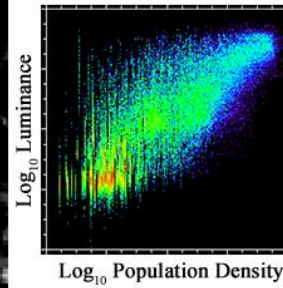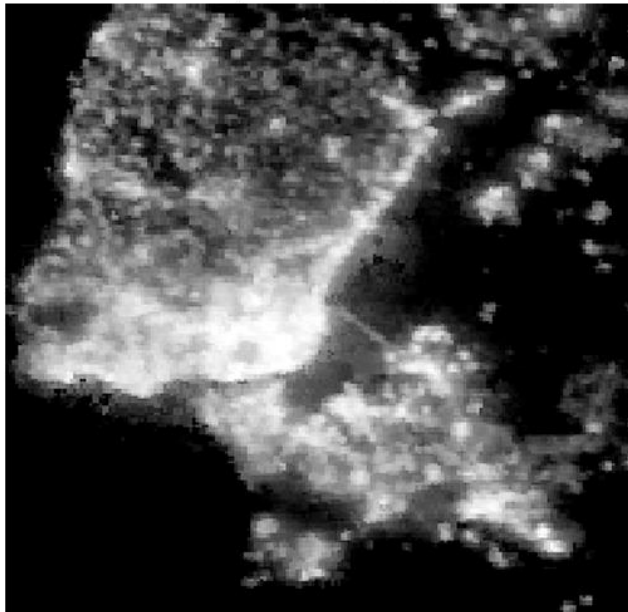
Northeast Corridor

Population Density 2010

Source: www.sedac.ciesin.columbia.edu
Analysis: www.LDEO.columbia.edu/~small/UrbanPopLight

Population Density
10^0 — 10^4
Brightness
10^0 — 10^6

VIIRS dnb 2012

Source: www.ngdc.noaa.gov/eog
Analysis: www.LDEO.columbia.edu/~small/UrbanPopLight

Population Density
10^0 — 10^5
Brightness
10^0 — 10^6

Johannesburg

Log$_{10}$ Luminance

Log$_{10}$ Population Density

Population Density 10^0 10^4
Brightness 10^0 10^2

Lisbon

Log$_{10}$ Luminance

Log$_{10}$ Population Density

Population Density 10^0 10^4
Brightness 10^0 10^2
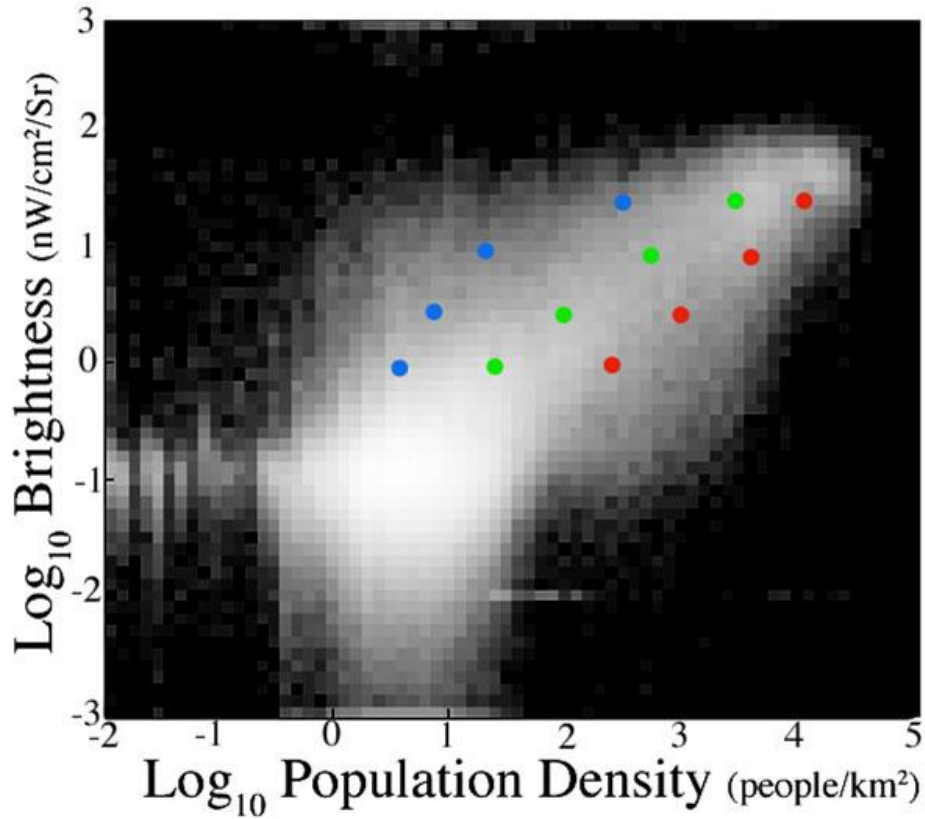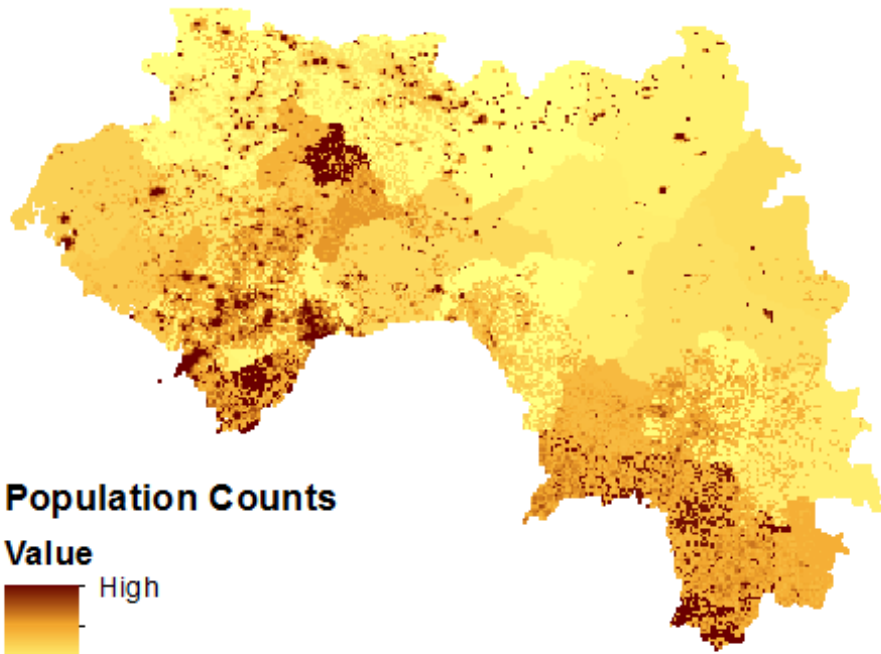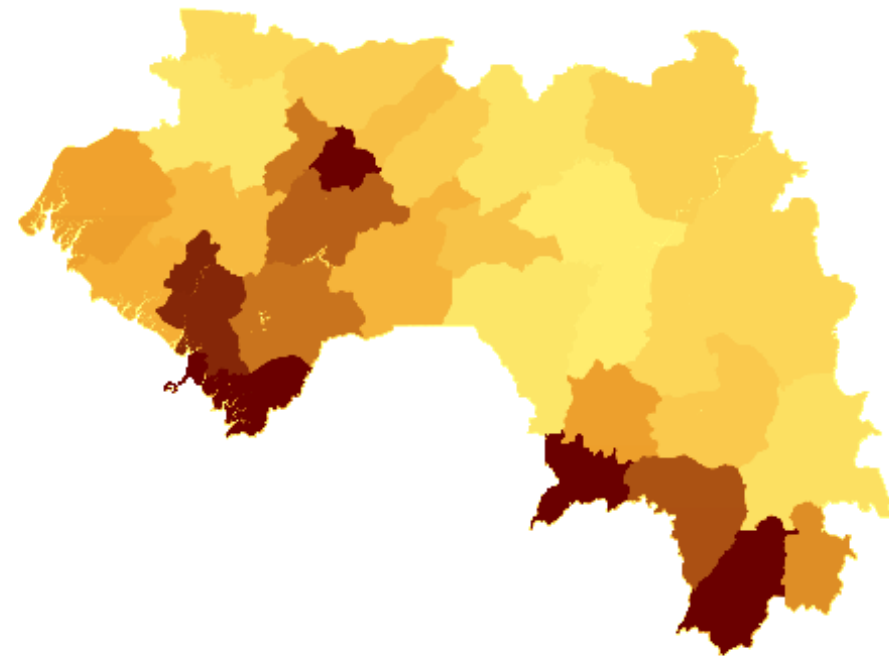
Population-Light Relationship
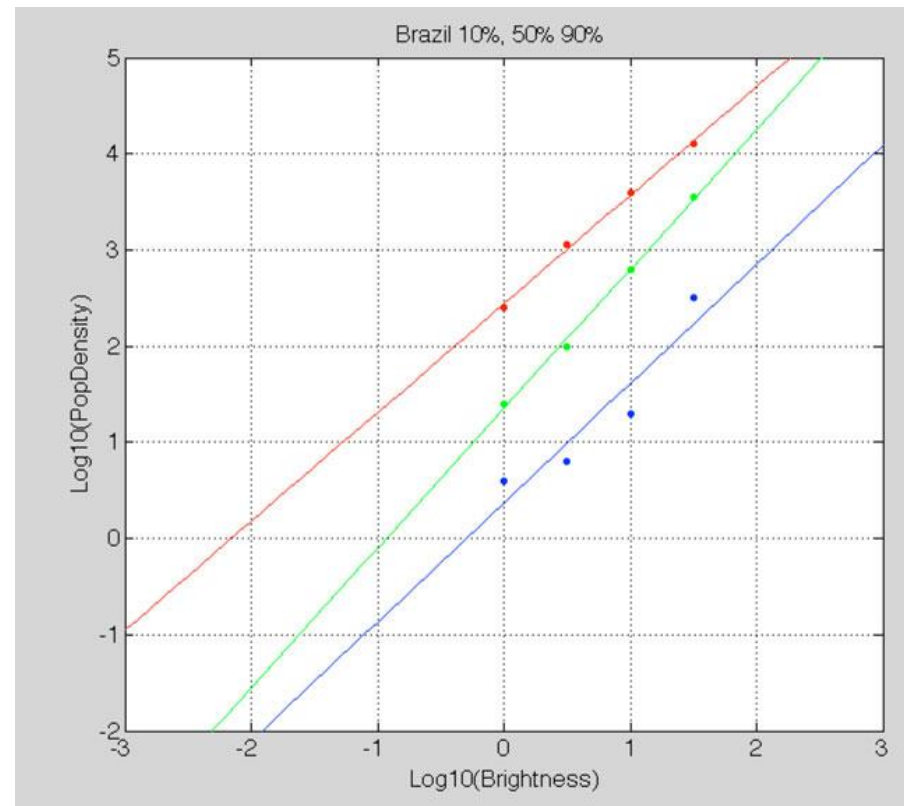
Test Case: Guinea
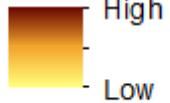

VIIRS Reallocation

Proportional Allocation

Population Counts

Value

High
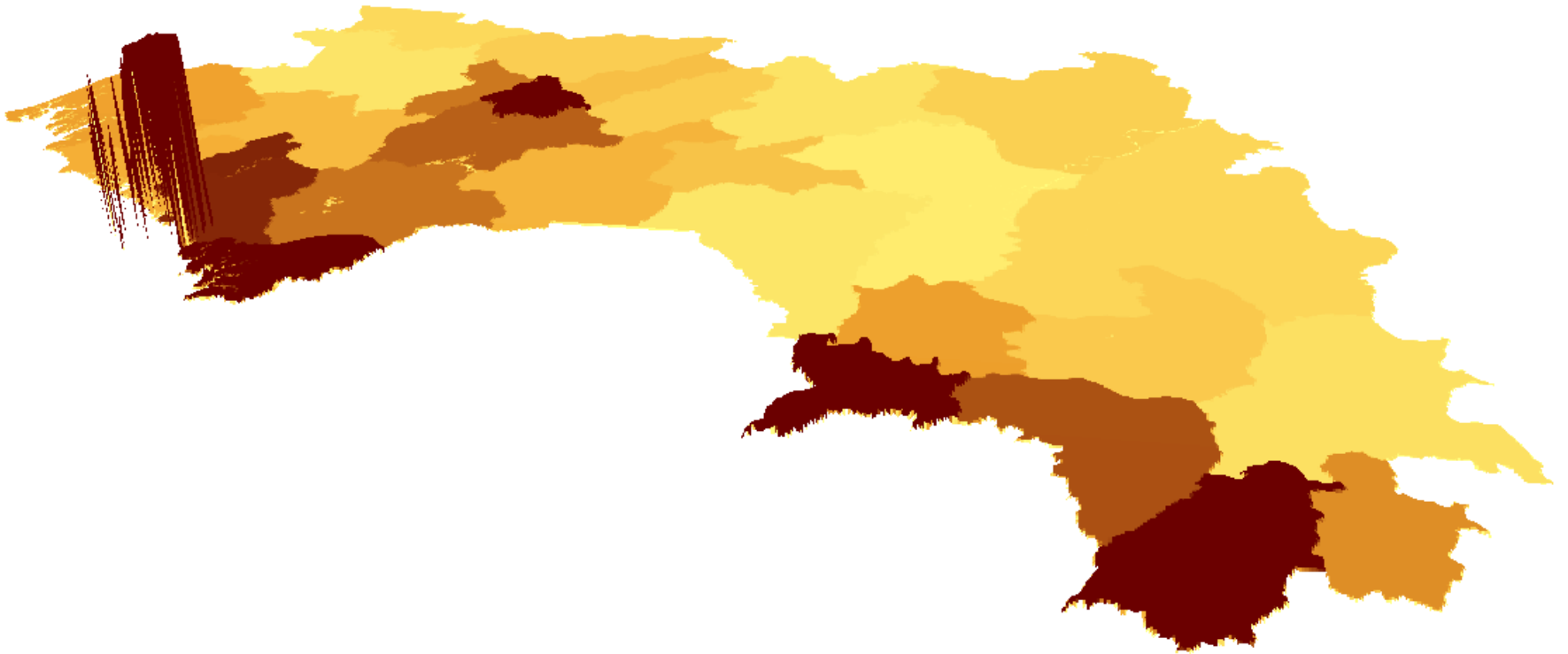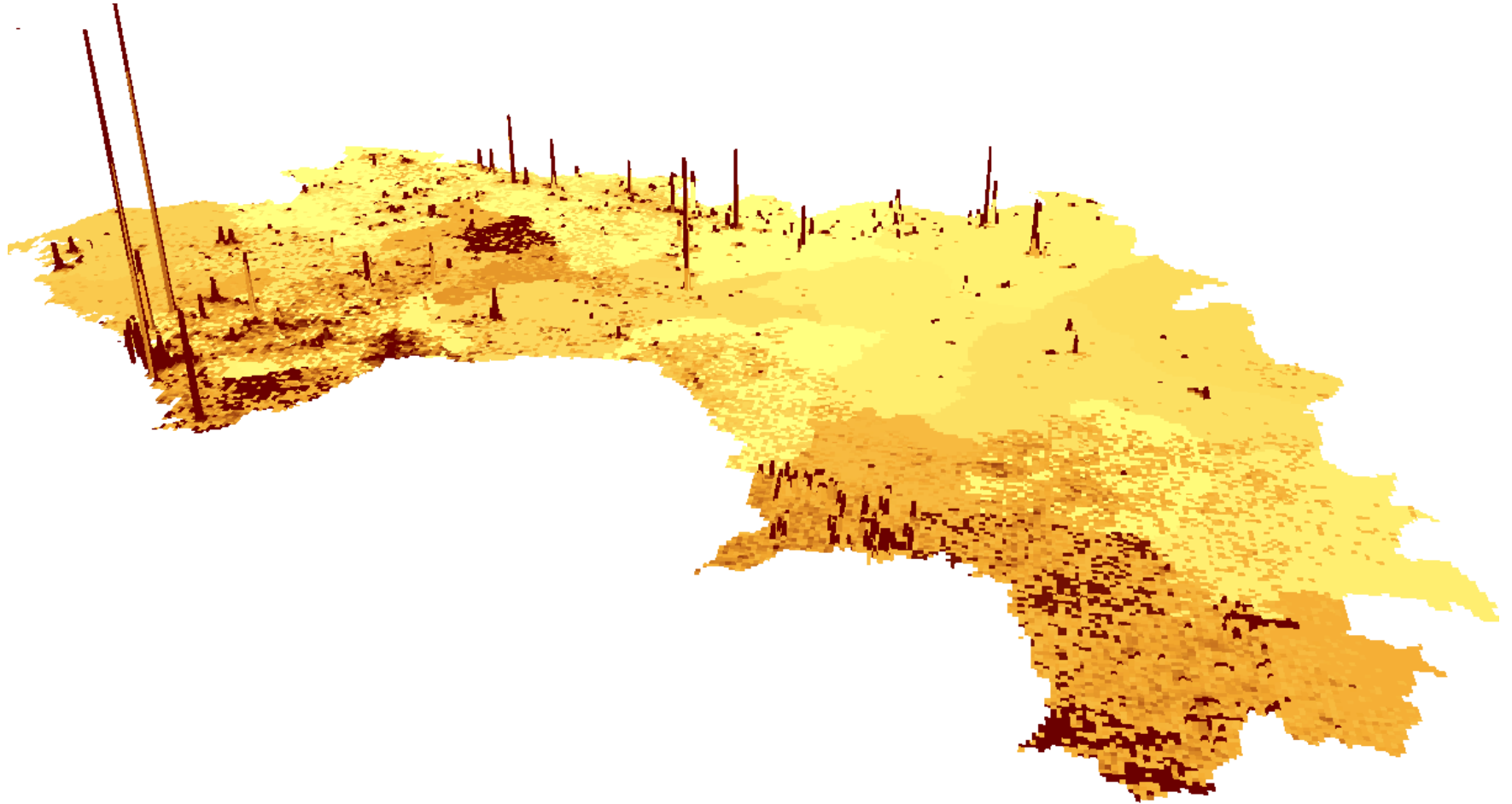
Low

Proportional Allocation
Height Proportional to Population Counts

VIIRS Reallocation
Height Proportional to Population Counts

GPW is a census-based global population grid that is:

- **Free and open source**
- **Well-documented**
- **Transparent**

- **Highly used**
- **Widely applicable**

How to access data:

- **GPWv3: http://sedac.ciesin.columbia.edu/data/collection/gpw-v3**
- **GPWv4: Email gpw@ciesin.columbia.edu to request alpha data**

Greg Yetman

gyetman@ciesin.columbia.edu