Mapping the Distribution of Human Population: The High Resolution Settlement Layer (HRSL)

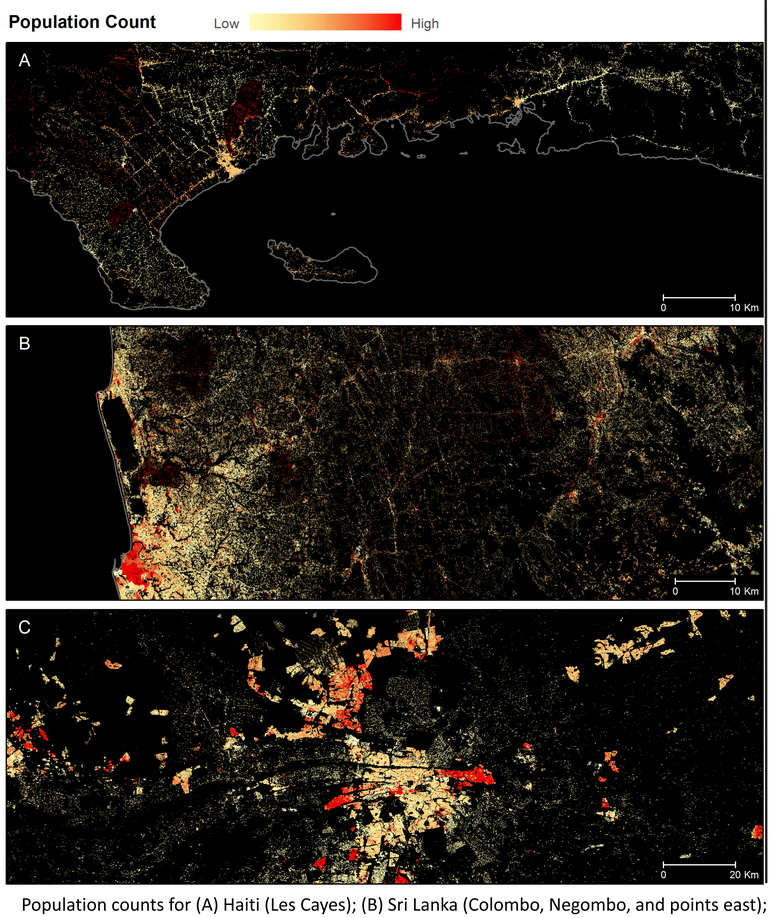
O. Borkovska, J. Gill, L. Pistolesi, J. Squires, G. Yetman

Center for International Earth Science Information Network (CIESIN), Columbia University

Overview

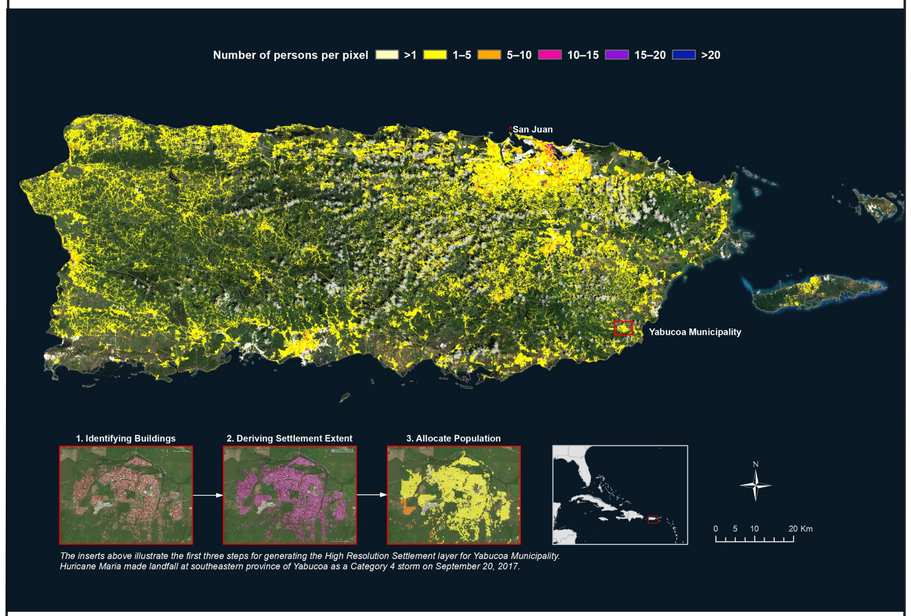
A new high-resolution population data developed in a joint effort by the **Facebook Connectivity Lab**, **CIESIN**, and **The World Bank** aims to help us better understand how people are distributed in many countries throughout the world. The High Resolution Settlement Layer data are part of a major effort by Facebook's Connectivity Lab to develop technologies to connect people in rural areas to the Internet. Better knowledge about patterns of rural settlement will help optimize the deployment of different communication platforms, including cell networks, high-altitude drones, and satellite-based connectivity.

State-of-the-art computer visioning techniques developed by the Connectivity Lab identified buildings from high-resolution commercial satellite images — the same type of imagery made available via publicly accessible mapping services. CIESIN then combined this information with census data to generate population estimates and worked with Facebook and the World bank to validate the results. We are making the data public now because learning how human settlements are distributed across the landscape—e.g., in clusters, along roads or waterways, or scattered widely—has valuable applications beyond the implementation of communication technologies.



Population counts for (A) Haiti (Les Cayes); (B) Sri Lanka (Colombo, Negombo, and points east); (C) South Africa (Pretoria)

High Resolution Population Allocation in Puerto Rico, 2015



The inserts above illustrate the first three steps for generating the High Resolution Settlement layer for Yabucoa Municipality. Hurricane Maria made landfall at southeastern province of Yabucoa as a Category 4 storm on September 20, 2017.

The High Resolution Settlement Layer (HRSL) models human population for the year 2015 at a resolution of approximately 30 meters at the equator. Following the tragic consequences of hurricane Maria, HRSL model was generated for Puerto Rico to identify population distribution of rural settlements. The Connectivity Lab at Facebook used computer vision techniques to classify blocks of high-resolution (0.5m) satellite imagery as settled (containing buildings) or not settled, and CIESIN used proportional allocation to distribute census data to the settlement extents.

Based on the results of the HRSL analysis (*Table 1*), the total population of Puerto Rico in 2015 was 3,672,370 covering a total settlement area of 1,173 square kilometers. Further analysis identified four settlements with populations larger than 100,000 people, constituting a total of 890,941 people or 24% of Puerto Rico's total population. As illustrated in *Table 2*, approximately 3,194,038 people or 87% of Puerto Rico's total population lives within 80 kilometers of the largest settlements.

See Table 2, for population distribution analysis in other countries.

Methods

1. Identifying Buildings

Buildings are identified in 0.5m optical satellite imagery (DigitalGlobe Basemap + Vivid) using labeled training data of buildings and a convolutional neural network (CNN) algorithm. The data are tiled and processed in parallel using computing infrastructure at the Connectivity Lab. The tiling scheme and the CNN result in some individual building being identified multiple times in the process. As a result, the classification only indicates the presence of buildings, not a random sample of the training data are excluded from the algorithm and used for validation. The overall accuracy achieved has been on the order of 90%. Specific statistics on precision, recall, and overall accuracy of the CNN algorithm are included in the Metadata for each published data set.

2. Derive Settlement Extent

The settlement extent locations were converted to a raster dataset at 1 arc-second resolution in geographic coordinates (approximately 30m per pixel). This layer is to facilitate analysis of settlement extents. It also indicates all pixels where buildings were identified; in some cases, buildings are located within administrative unit data with no population, such as national parks. In the population allocation step, these buildings are not allocated any population.

3. Allocate Population

Population from administrative polygons and census data are distributed to the pixels with buildings using a simple proportional allocation method. The total population of each polygon is divided by the number of pixels that have buildings (settlement extent above). The resulting value is assigned to each pixel in a new raster surface. This surface can be used with a zonal statistics function to calculate the sum of population over any overlapping geography.

4. Assessing Errors

The building location algorithm has been validated using comparison to other settlement extent datasets and random sampling of the building locations. False positive rates (errors of commission) from the random sampling were very low (1-2%). In most cases, false positives are related to linear features in the imagery being classified as buildings, such as terraced fields, jointed bedrock outcrops, and shallow-water waves. Coastline and water body datasets are being used to remove water areas. False negatives (errors of omission) are more common (10% or more, depending on the country); however, it is important to note that missing a single building counts as a false negative. The false negatives identified are commonly in areas with poor image quality (cloud, haze), dense canopy cover that obscures parts of buildings, and buildings with less distinct edges, such as thatched roofs.

Analysis

Settlement Extent Comparisons

The settlement extents were compared with two other built-up classifications: the Global Human Settlement Layer (GHSL) from the Joint Research Centre (JRC), and the Global Urban Footprint (GUF) from DLR. In both cases, the HRSL detected more settlements in rural areas. Given the differences in input data resolutions (0.5m DigitalGlobe data for HRLS, 30m Landsat data for GHSL, and 12m TerraSAR-X data for GUF) it is not surprising that the HRSL can find smaller settlements.

Population Distribution

The HRLS data allow characterization of rural population distribution in great detail. From the data, we can determine that most people live relatively close to large urban areas. Table 2 on the right shows the distribution of population near large urban settlements as defined by grouping contiguous pixels in the HRSL and summing the population totals.

Table 1: Settlement Area (square kilometers)

Country	Global Urban Footprint	Global Human Settlement Layer	High Resolution Settlement Layer
Puerto Rico	**866**	**1,103**	**1,173**

Table 2: Population Distribution

Country	Settlements with population > 100 k	Population of settlements >100k	Population within 80 km of settlements >100k
Puerto Rico	**4**	**890,941 (24%)**	**3,194,048 (87%)**
Haiti	3	3,134,633 (29%)	9,364,763 (87%)
Madagascar	4	5,575,438 (23%)	10,165,258 (42%)
Malawi	5	1,314,992 (8%)	10,953,484 (64%)
South Africa	42	21,981,011 (40%)	24,477,089 (56%)

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

For more information
Contact: Greg Yetman
gyetman@ciesin.columbia.edu

Data Download: https://ciesin.columbia.edu/data/hrsl

THE WORLD BANK

internet.org by facebook