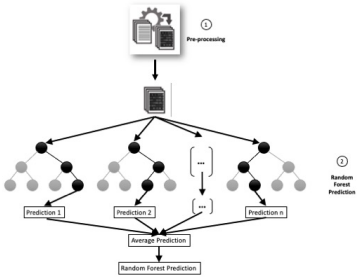
Enhancing Small Area Population and Poverty Estimates Using Geospatial Data and Satellite Imagery

Arturo Martinez Jr., Katharina Fenz, Martin Hofer, Thomas Mitterling, Tomas Sako, Joseph Bulan, Mildred Addawe, Ron Lester Durante, and Marymell Martillan

Availability of accurate, timely, and spatially disaggregated distributions of human population and poverty are important on many fronts. Through the use of advance computer vision algorithm, machine learning techniques, and combining small area estimates on population and poverty with innovative (geospatial) data, this study was able to present a computational framework for generating a more granular population and poverty map.

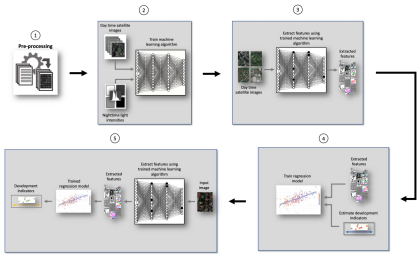
Methodology

Population distribution prediction using Random Forest Method



The population mapping applied a random forest of regression trees because the dependent variable (i.e., the number of people living in each 100 m by 100 m grid cell) is a continuous variable. The independent variables consist of satellite imagery data such as night lights, land cover classes, temperature, and precipitation. The vector of explanatory variables and the dependent variable are assumed to follow a joint distribution.

Poverty prediction using Convolutional Neural Networks and Ridge Regression

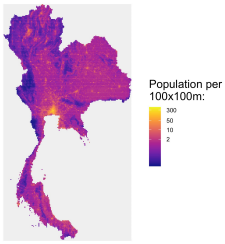


The methodology uses granular poverty data, daytime satellite images, and nighttime light intensities. After data cleaning and preprocessing, a pretrained Convolutional Neural Network (CNN) was retrained to categorize daytime images into various intensity levels of night light. The trained CNN is then used to extract visual features of the last layer. Regression is done to explain the relationship between the image features and the government-compiled poverty estimates.

Results

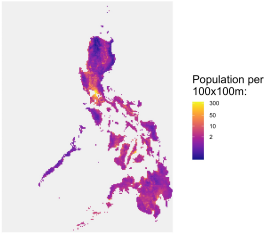
Population Mapping

Population Density in Thailand in 2015



R^2=91.29%

Population Density in the Philippines in 2015



R^2=91.82%

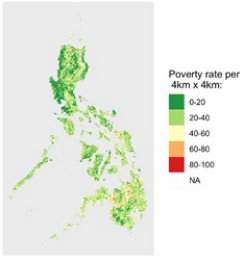
Poverty Mapping

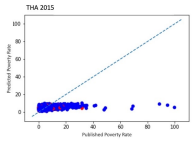
Root Mean Square Error by Country and Year

Country / Year	Validation Set	All
PHI, 2012	17%	17%
PHI, 2015	17%	15%
THA, 2013	12%	11%
THA, 2015	4%	5%

R^2 (Validation set): 42.30%
R^2 (Full data set): 52.48%

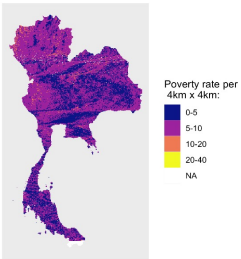
PHI 2015, Predicted





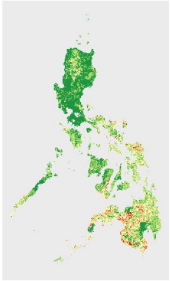
R^2 (Validation set): 4.62%
R^2 (Full data set): 10.66%

THA 2015, Predicted

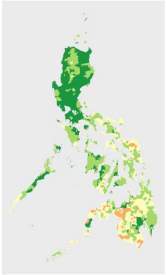


To address the issue of underestimation in the poorest areas, calibration methods can be employed. The machine learning predictions can be rescaled or calibrated such that they align more closely with the government-published numbers. The calibration should preserve the distributional structure of the grid-level poverty predictions, but still pay heed to the estimates published by the government at more aggregated levels.

PHI 2015, Calibrated

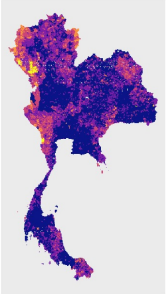


PHI 2015, Published

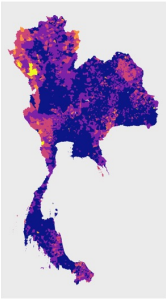


Poverty rate pe 4km x 4km:
0-20
20-40
40-60
60-80
80-100

THA 2015, Calibrated



THA 2015, Published



Poverty rate per 4km x 4km:
0-5
5-10
10-20
20-40
40-100

Summary

This study explores the feasibility of using geospatial information to provide granular estimates of population and poverty. The results are encouraging as they show capacity of machine learning methods to replicate the estimates from conventional data sources given information from satellite imagery.

Relevant Knowledge Products

- http://dx.doi.org/10.22617/FLS200215-3
- https://dx.doi.org/10.22617/TCS210076-2
- http://dx.doi.org/10.22617/TCS210112-2
- https://dx.doi.org/10.22617/SPR210131-2