Mapping Income Poverty in Belize

Using Satellite Features and Machine Learning

Jonathan Hersh
Chapman University

Ryan Engstrom
George Washington University

Michael Mann
George Washington University

Lucía Martín Rivero
Inter-American Development Bank

Alejandra Mejía
Inter-American Development Bank

Cataloging-in-Publication data provided by the
Inter-American Development Bank- Felipe Herrera Library

Mapping income poverty in Belize using satellite features and machine learning / Jonathan Hersh, Ryan Engstrom, Michael Mann, Lucía Martín, Alejandra Mejia.

p. cm. — (IDB Monograph ; 806)
Includes bibliographic references.

1. Poverty-Belize-Data processing. 2. Machine learning-Belize. 3. Geographic information systems-Belize. 4. Belize-Economic conditions- Data processing. I. Hersh, Jonathan. II. Engstrom, Ryan. III. Mann, Michael. IV. Martín, Lucía. V. Mejía, Alejandra. VI. Inter-American Development Bank. Country Office in Belize. VII. Series.

IDB-MG-806

JEL Classification: I3, I32, I38, O31, O35

Keywords: Belize, Central America, big data, satellite imagery, poverty maps, measurement and analysis of poverty, economic development, social innovation, machine learning, geographic information systems

Contents

Prologue

The last poverty assessment of Belize was undertaken in 2009, which hinders the country's capacity to make an accurate diagnosis of the social situation. Measuring poverty not only allows governments to design better social programs, target their beneficiaries and monitor their progress - it also allows a country to assess how close it is to eradicating poverty.

For decades, poverty has been measured using costly on-the-ground surveys that can only hope to sample a fraction of the poor. In part because of the high cost, the poorest countries often have the weakest poverty statistics. Despite international efforts, most governments measure poverty too infrequently to adequately inform policy, using samples that are representative of larger geographic areas than is necessary to combat increasingly isolated pockets of poverty.

With the advent of big data, such as satellite-based remotely sensed data, we are now able to measure correlates of poverty that span the entire universe of the world's poor. Until recently, however, these methods were too expensive; for countries that did not have the budget to produce traditional surveys, such costs were unaffordable. Fortunately, the widespread availability of free imagery with global coverage and frequent revisit rates now provide an important opportunity to produce poverty mapping at a lower cost.

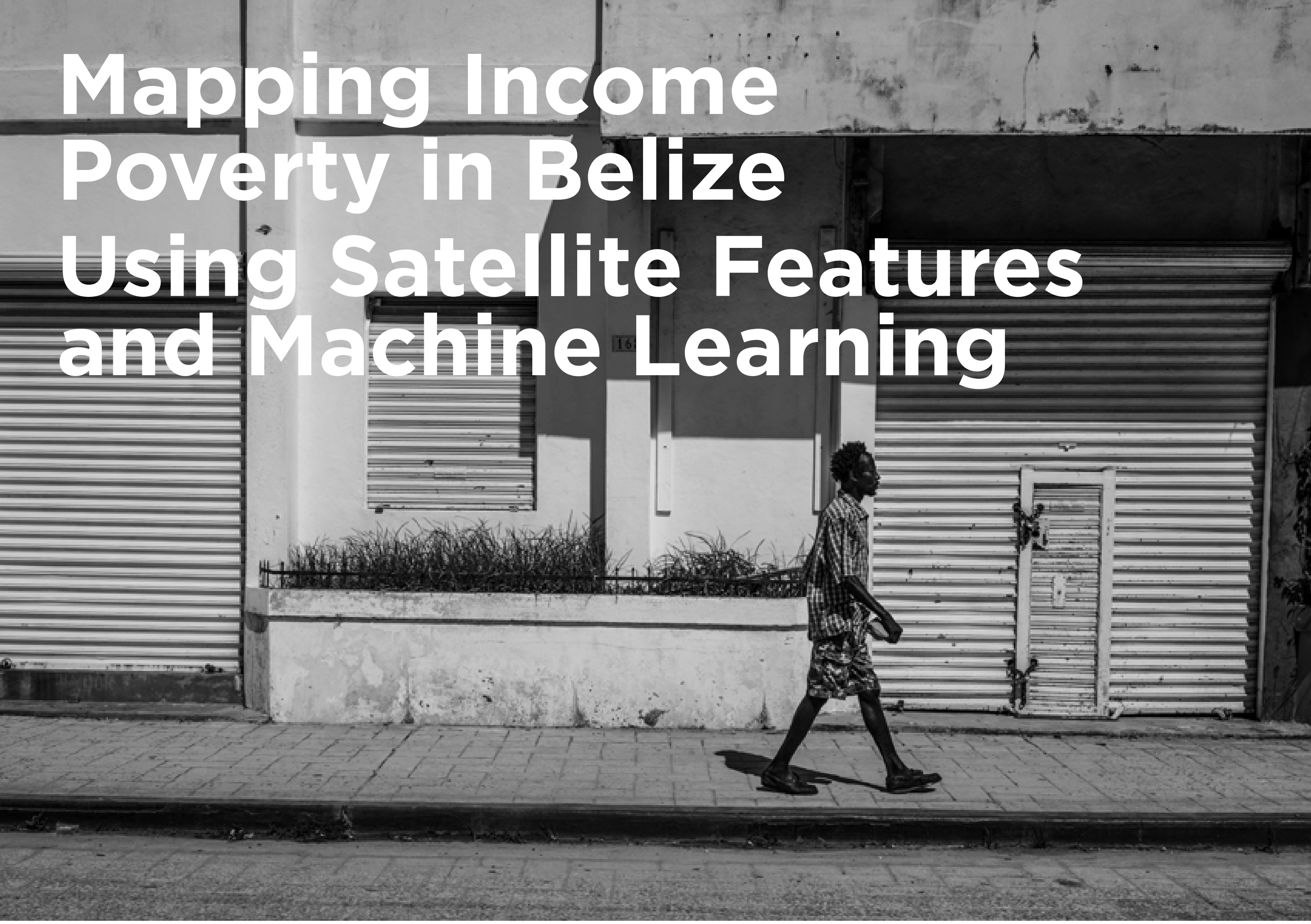
Belize is an exemplary country where big data and machine learning hold promise for generating poverty maps at a reduced cost. More than ten years have passed since Belize produced its last poverty assessment, and the sample sizes of the surveys that the country produces are too small to generate poverty statistics. Moreover, agriculture is a major source of income for Belizean households, especially the poorest, and satellite imagery is particularly adept at capturing intangibles in rural areas that are difficult to measure in surveys.

For all the reasons cited above, the Inter-American Development Bank (IDB) initiated this project with the assistance of researchers at Chapman and George Washington Universities over a year ago. Today, this monograph makes available for the first time a poverty map at the enumeration district level, leveraging the latest techniques in big data and the use of satellite imagery. It is important to note that the project included a one-week training session for the Statistical Institute of Belize (SIB) in an effort to provide the necessary skills for SIB staff to produce poverty statistics with greater frequency and at a lower cost.

At the IDB, we trust that this monograph will be a tool to assist policymakers in Belize to make better-informed decisions as well as contribute to the research in this field, so that more countries could benefit from new technologies to fight poverty.

Cassandra Rogers
IDB Country Representative in Belize

Mapping Income Poverty in Belize Using Satellite Features and Machine Learning



This study creates poverty maps for Belize leveraging free and open source methodologies that link satellite imagery and existing survey data with machine learning. Belize is an exemplar of country for which Big Data and machine learning hold promise for generating poverty maps at a reduced cost as the last poverty assessment was conducted in 2009 and no consumption survey with the intention of producing sub-national estimates of poverty has been conducted since then.

Even though poverty mapping provides clear information about the location and extent of poverty within a country, mapping the spatial distribution of poverty or incomes within a country remain a challenge for countries. Budget constraints restrict the use of poverty maps for policy targeting, leading to waste and inefficiency.

Recently, several methods have been proposed that incorporate features from satellite imagery to either improve model performance or supplant existing small area estimation methods. However, all the currently proposed methods require expensive high-spatial resolution imagery which, given their high cost and infrequent acquisition, may render these advances impractical for most applications.

This paper investigates how small area estimates of average income may improve when incorporating features derived from Sentinel-2 and MODIS imagery. Both satellites provide free imagery, have global coverage, and frequent revisit rate. It estimates a poverty map for Belize at the Enumeration District level which incorporates 423 contextual spatial features, and 203 time series features derived from these sensors, with and without 37 survey derived variables. The paper compares four machine learning methods, Ridge, Elastic Net, Random Forests, and Extreme Gradient Boosted Trees. The paper documents a nine percent improvement in model performance when including these satellite features into machine learning models estimating average incomes. Critically, these improvements are most significant for the poorest households. In particular, models without satellite features are biased upward for the poorest households. This suggest that surveys alone do not contain sufficient information to recover poverty rates and incomes for the poorest households.

This paper finds that the poorest districts in Belize are Corozal, in the north, and Toledo, in the south, and suggests significant reduction in poverty for the districts of Orange Walk and Stann Creek in the last decade.

Executive summary

Problem	Approach	Results
Collecting routine poverty statistics is challenging and expensive. Current small area estimate methods overstate model performance or require expensive proprietary inputs. Budget constraints restrict the use of poverty maps for policy targeting, leading to waste and inefficiency.	We leverage machine learning and publicly available data to boost model performance. This captures critical information about urban and rural poverty from satellite imagery. Combined this creates a low cost and easily reproducible methodology that can be implemented anywhere.	We explain over 91% of variability in enumeration district average income, and 55% out-of-sample. In particular we see significant improvements in accuracy for lowest quintile and rural households.

Poverty mapping provides clear information about the location and extent of poverty within a country. These maps are generally used for planning and resource allocation. In many countries, budget constraints and access to clear and open source methodologies limit the routine creation of spatially explicit maps of poverty. Combining surveys with Big Data, such as information in cell phone call detail records or satellite imagery, may help reduce the costs of generating routine poverty maps. Further, Big Data might provide useful information on subpopulations of interest – the very poor, or individuals in rural areas, for example – that might otherwise be absent using traditional surveys.

In this study, we create poverty maps for Belize leveraging free and open source methodologies that link satellite imagery and existing survey data with machine learning. We extract, from freely available satellite images, meaningful characteristics that are possibly correlated with local area income, including variables tracking the physical characteristics of a neighborhood, and time series information on rainfall and vegetation patterns in the year prior to the survey. This method is a significant jump forward in poverty mapping methodologies as it largely overcomes issues of cost and data availability.

Belize is an exemplar of country for which Big Data and machine learning hold particular promise for generating poverty maps at a reduced cost. The last poverty assessment for Belize was conducted in 2009, during which the national poverty rate was assessed at 41.3%. During this assessment, sub-national estimates of poverty were provided only at the district-level. While labor force surveys in Belize are routinely conducted, no consumption survey with the intention of producing sub-national estimates of poverty has been conducted since then. For this application, we leverage a small, routine Labor Force Survey (LFS) and its linkages with a recent household census to infer household income.

Raw satellite data does not capture information that is readily useful in poverty mapping. Therefore spatial and time series features are used to identify correlations between spatial and temporal patterns and poverty. Spatial patterns such as differences in complexity urban areas can help identify informal settlements. Time series patterns can provide information critical to drivers of rural income such as the health of agricultural plants. These satellite data, when combined with the census, create a large number of independent variables for estimating spatial patterns of poverty. Satellite variables also prove crucial in accurately estimating poverty of rural households.

In generating poverty maps for Belize we follow the Small Area Estimation method that combines information in a survey, which does not sample all areas for which we want to have poverty estimates, with a Census, which does sample all areas. We deviate from the standard linear models in Small Area Estimation and leverage the latest advances in machine learning to find linkages between satellite imagery, and census data to predict household income recorded in the smaller LFS survey. Here machine learning is particular useful in fitting models with large number of co-variates, as is the case of our models with satellite-derived variables. We utilize a multi-model approach, taking advantage of the benefits of a variety of modeling approaches. Model accuracy is assessed on both the data used to estimate the model (i.e., the training data) and the data used to validate the model (i.e. the testing data). Assessing model accuracy on data not used to estimate the model (out-of-sample) is necessary to ensure we have built a model that is broadly representative of the country as a whole.

Results from this study indicate that the models explain 90% of the variation in enumeration district (ED) average income in the training data, and 55% of the variation in ED income in the testing data.

Maps are produced showing relative poverty rates at the ED level, indicating that the poorest districts are Corozal, in the north, and Toledo, in the south. Viewing the last poverty map that was completed in 2009, four districts were classified as having high poverty rates – Corozal, Orange walk, Stann Creek and Toledo. Viewing our analysis in light of the previous map, it appears there has been significant reduction in poverty for the districts of Orange Walk and Stann Creek. Given the higher resolution of the new poverty map, we see substantial within-district heterogeneity.

One important caveat with these maps is that while they provide an excellent view on the distribution of poverty within Belize, as currently designed they cannot provide an update to the national poverty rate. In order to do so, Belize would need to conduct a consumption survey that would provide the input to the machine learning models in substitution for the Labor Force Surveys used. However, using Big Data and machine learning paired with a consumption survey for this purpose will likely reduce the total costs of the consumption survey, given that Big Data has shown to reduce the number of households needed to survey for a given level of survey accuracy.

A second important caveat is that, because we model household income, prediction variance should improve with the population of the underlying enumeration district. This is a basic feature of statistics, akin to the fact that a survey which samples more households will have a smaller margin of error in comparison to a survey that samples fewer households. One important point to remember, however, is that the increased variance for enumeration districts with smaller population will affect their income estimates both positively and negatively. Outside of systematic model bias, it will not be the case that ceteris paribus higher population districts will have higher estimated income.

Overall, these model results indicate that this open source method provides an in-expensive approach that performs as well or better than more complicated and costly methodologies. Importantly, this methodology can easily be applied over time as new survey data arrives, or can be expanded to new countries at low cost.

Introduction

Eradicating poverty is the first of the UN sustainable development goals (SDGs). However, "data gaps" in the coverage of poverty statistic persist, despite international efforts to increase their development. As many as 57 countries produced one or fewer data points on poverty in the decade between 2002 to 2011 (Serajuddin et al., 2015). Unless better, cheaper methods are developed for identifying poverty it seems unlikely we will meet the goal of eradicating global poverty by 2030.

Producing sub-national poverty estimates at regular frequencies is burdened by the expense of conducting reliable consumption surveys. Kilic, Serajuddin, Uematsu, and Yoshida (2017), estimate an average direct survey cost in Latin America of $105 per household surveyed, and technical assistance costs on average are $613,000. Taken together, these costs amount to $2M per survey. Despite the innovations of rapid poverty assessment approaches such as Pape and Mistiaen (2018) and SWIFT (Yoshida et al., 2015), the expense of conducting reliable poverty surveys with sufficient frequency remains prohibitive in most countries.

Given finite budgets and limited technical capacity to mount these surveys, several researchers have proposed using Big Data to assist in the generation of sub-national estimates of poverty, such as metadata from cellular phones (Blumenstock, Cadamuro, and On, 2015) or satellite imagery (Jean et al., 2016, Engstrom, Hersh, Newhouse, 2017). Most of these methods, however, rely on the use of expensive and proprietary cellular metadata, or high spatial resolution satellite imagery. For the hypothetical statistical agency that cannot commit $2M to mount a survey, the proposition of purchasing expensive Big Data such as high-resolution satellite imagery amounts to a sick patient considering two medicines, the expense of either of which is too dear to bear.

This paper investigates the extent to which "open-source" Big Data can be meshed with existing survey data, to alleviate the lack of frequent sub-national poverty estimates. Using Belize as a test case, we utilize freely available, open-source satellite imagery to build sub-national estimates of income poverty, and determine the extent to which features from satellite imagery act as substitutes and complements to survey-based estimates of poverty.

1. *Belize Context and Satellite Feature Motivation*

Belize is an exemplar of a country for whom open-source Big Data methods may greatly reduce the cost of generating poverty statistics. Firstly, the country last produced a poverty assessment in 2009 (Belize National Human Development Advisory Committee, 2010). Since then, no sub-national poverty statistics have been produced to our knowledge. The expense of producing a specific poverty survey may be prohibitive. Secondly, the Statistical Institute of Belize (SIB) conducts bi-annual Labor Force Surveys with sample sizes that are insufficient for generating poverty statistics directly, but with the addition of variables derived from satellite imagery, they may be able to cheaply produce small area estimates of poverty. Thirdly, agriculture is a major source of income for much of the population, in particularly poorer households. Satellite imagery is particularly adept at capturing intangibles in rural areas – such as periods of drought or excessive rain, whether fields are green or fallow – that are difficult to impossible to measure in surveys. While well designed for its intended purpose of tracking employment patterns, survey questions in the Labor Force Survey are designed to capture labor income rather than the intransient incomes inherent with agricultural labor. Fourthly, the distribution of poverty in Belize is such that poverty is higher in rural areas than urban areas. Thus to sufficiently measure poverty in Belize we have to accurately measure rural poverty, something that is greatly improved with the addition of satellite features.

Using two waves of surveys from the 2017 Labor Force Survey in Belize, we estimate machine learning models to predict household labor income as a function of survey & satellite variables. Features generated from satellite imagery are derived to capture both cross-sectional information as well as time-series properties. This information may be able to capture, for example, drought conditions in remote agricultural areas otherwise unobserved in surveys. We utilize four advanced machine learning models – Ridge regression, Elastic Net Regression, Random Forest, and Extreme Gradient Boosted Trees – to predict household income from satellite and survey characteristics. Despite the models' relative agreement as to the income of each area, we use a technique called model ensembling to reduce the prediction variance. In model ensembling, we take each model's prediction and average it across the models. This is done to weigh the strengths and weaknesses of any particular model, and has been shown to outperform using individual models alone. Additionally, we create an ensemble estimate of poverty rates using information from all four estimated models, which may be more robust to model uncertainty than a single poverty model.

We find that household-level income models, used to generate Enumeration District (ED) poverty rates, improve when incorporating satellite variables. Satellite and survey models explain 55.3% of the variation in average incomes between predicted and true average ED income in the validation sample, compared to 50.7% of the variation using survey data alone, an improvement of 9 percentage points. Satellite models alone explain 30% of the variation between predicted average ED income and true average ED income. Altogether these are not stunning arguments for the use of open-source Big Data. However, we find that models improve precipitously for the poorest households. Average residuals for the lowest income decile households decline by nearly a third in magnitude. Given that much of the poverty in Belize occurs in rural areas, we believe satellite variables capture important features of income that are not observed by surveys.

Poverty Map Motivation and Background

There are several motivations that justify the production of frequently updated poverty maps. One key function of a poverty map is to increase the targeting efficiency of anti-poverty programs. These include not only direct transfer programs, but any fiscal policy that has the reduction of poverty as an outcome of interest. To the extent that efforts can be directed towards poor areas, the efficiency of these programs crucially depends on accurate information on the location of the poor. Regarding direct income transfers, the standard method of assigning transfers involves a proxy means test, where individual household characteristics are used to assign poverty status in absence of a full consumption survey (Grosh and Baker, 1995). Firstly, poverty maps can help design the sample frame over which these proxy means tests are enacted, ensuring poor areas are adequately covered by income transfers (Bah et al., 2018). Secondly, the efficiency of these transfers depends on whether they are timed according to aggregate shocks (Bazzi et al., 2015). Given the likely spatially heterogeneous income shocks accurtate, up-to-date poverty maps increases targeting efficiency.

Another motivation to produce frequently updated poverty maps is to ensure information about the location of the poor is transferred to higher levels of government. This is particularly crucial if fiscal policy has an anti-poverty aim as one of its goals, such as centralized transfers to support school construction in poor areas. Local representatives may be aware of which areas are poor, and can design anti-poverty efforts efficiently given their local resources. However, the information on sub-national poverty may not transfer to higher levels of government. Worse still, while qualitatively local representatives may know the rank ordering of poverty for areas under their representation, aggregating rank order information from multiple representatives cannot ensure a national rank ordering is representative of true poverty rate. If maps are then frequently produced they may then be useful as an outcome to measure the relative effectiveness of anti-poverty efforts. This could improve the efficiency of anti-poverty policy. While the costs of such policies are well known, learning the benefits of such programs requires accurately measuring sub-national poverty rates, as aggregate statistics may be confounded by numerous factor outside the scope of the anti-poverty policy.

Poverty maps may also be used by democratic societies in holding their elected leaders responsible for a key welfare measure of interest: the fraction of their constituents that are in poverty. Poverty maps are easily understood by individuals with varying educational backgrounds. By informing constituents on the changing or static poverty rates of their local areas, voters are enabled to make informed decisions about which elected leaders to keep in or oust from office. Elected leaders often make grand promises regarding various outcomes of interest, and yet poverty is a measure agreed on by most as an important metric to track. Defining this metric locally, and updating it frequently, aids in the political process to hold elected leaders responsible.

Specifically for the Belize, the challenge is whether to implement a transfer based anti-poverty program or if fiscal policy should be designed with an anti-poverty goal, and if so, what type of policy should be enacted. Specifications on policy are beyond the scope of this document, however these estimates of sub-national poverty will be important inputs to a Hausman, Rodrik, and Velasco style growth diagnostics analysis. (2004).

2. *Motivation for Open Source Date and Methods*

Open-source and freely available satellite images may hold many potential benefits for resource constrained agencies. For one, statistical agencies can commit to the price of 0$ for open-source imagery in perpetuity. In comparison, a statistical agency that incorporates proprietary data into their statistical pipeline opens themselves to price gouging as proprietary data providers have pricing power due to "lock-in" effects (Arthur, 1989). Data lock-in effects could occur if there are considerably costs moving from one data provider to another, for example because of costs of staff adapting to new software or methods. Thus an initially low cost for proprietary data could balloon into larger costs if firms are profit maximizing and choose to exert pricing power.

At the extreme end of firm profit maximizing behavior, it's possible that even with competition among data providers, any surplus from using Big Data at statistical agencies may eventually be captured by proprietary data providers because of these lock in effects. Thus, it's crucial to consider open-source alternatives to proprietary providers. This paper fills a necessary gap in the literature whereby we explore whether these open-source alternatives may be of use to the prototypical statistical agency.

Data

Survey Data

In order to generate estimates of household income we utilize two datasets: first, a labor force survey that asks critical questions about income but has a relatively small sample size, and the second, the national census, which covers all households but does not directly ask about household income.

Labor Force Survey and Census

We derived household income statistics from the April and September 2017 waves of the Belize Labor Force Survey, and use the 2010 Belize Census. The April wave surveyed 2,331 unique households and the September wave surveyed 2,320 households, which were repeated cross-sections and not necessarily the same sampled households. This resulted in 4,651 unique households across the two waves. After removing missing values this resulted in a data set of 3,658 household level observations. We split the observations into 75% training data, which will be used to estimate our household level model, and 25% testing data, which will be used to validate our model at the household and enumeration district level. This testing and training split of the data is a standard method in machine learning (Gareth, et al., 2013), which is a necessary step to prevent overestimation of model performance. However, this testing/training splitting of the data is seldom if at all used in generating poverty maps. In comparing our model performance to other models, the directly comparable statistics will be the "in-sample" performance metrics, which is almost always more optimistic regarding model fit than the "out-of-sample" criteria.

We derived 37 co-variates, or household-level variables likely correlated with income, from both the Census and Labor Force Surveys (LFS) which were derived from identical survey questions. While the set of co-variates may seem small, consider that a popular software for building poverty maps provided by the World Bank PovMap[1] can estimate at most 25 variables.

Remotely Sensed Data

In order to summarize satellite images at a given geographic level, we processed images to create summary statistics that capture important spatial and temporal aspects of the imagery. These statistics then provide information that correlates with conditions on the ground, such as building and vegetation patterns that are correlated with poverty. Satellite data provides several clues about the condition of households on the ground. Images might be able to provide information about the size of homes, the spatial layout of the buildings, the types and intensity of land-use, or even indicators of successful or poor agricultural seasons. The challenge then is to extract a set of metrics from imagery that can describe some of these attributes.

In this study we explore two approaches, one Contextual Features that looks at spatial and spectral patterns within neighborhoods within a single time period. Contextual Features can help us understand texture (spatial patterns observable in a single image), orientation, complexity, and continuity of neighborhoods or groups pixels. The second approach looks at Time Series features, examining the change of each pixel over time. Here we can extract features like maximums, means, trends, sudden shifts, for a variety of metrics including rainfall and greenness.

1. *Sentinel-2 Imagery: Contextual Features*

In the creation of Contextual Features we used imagery from the Sentinel 2 sensor, which is onboard two satellites owned and operated by the European Space Agency (ESA)[2]. The European Space Agency provides images from these sensors for free, and because it is on two satellites, all of the earth's land mass is observed every 5 days. Sentinel 2 imagery measures reflected energy from the sun in 12 wavelengths from the visible bands (Blue, Green and Red) into the Near Infrared (NIR) and Short-Wave Infrared (SWIR). Recall that a visible color image is composed of different visible "bands", each capture different wavelengths, these are typically Blue, Green and Red bands. We focused on the visible (Blue, Green, Red) and NIR bands of Sentinel 2 because they have the highest spatial resolution with a pixel size of 10m.

Because the sensor measures reflected sunlight, one of the difficulties in working with these data in a country such as Belize is cloud cover. In order to overcome these issues we use Google Earth Engine to create a cloud free mosaic of the entire country by selecting a cloud free pixel for each location over a period of time. This was done by selecting the median pixel in each band from January 1, 2017-March 31, 2018 for each time an image was collected by the Sentinel 2 sensors. These data provide the spatial detail required to observe spatial patterns across the landscape.

2. *MODIS Imagery: Time Series Data*

We additionally use imagery from the Moderate Resolution Imaging Spectroradiometer (MODIS), which is owned and operated by the National Aeronautics and Space Administration[3] (NASA). MODIS sensors are on two satellites and acquire images for entire world twice daily. As such, while MODIS has a low spatial resolution (250m) it is compensated by its high temporal resolution – that is high revisit rate. Thereby these data provide rich information for time series statistics. Moreover, because the MODIS sensors have been in orbit for many years, a longer time series is available that allow us to summarize the properties of the five years leading up to the study date, from Jan 1 2013 – Dec 31 2017.

The time series we construct from MODIS, is the normalized difference vegetation index (NDVI). NDVI is commonly used to monitor the status of crops, forests, and ecosystems. NDVI is sensitive to the amount of chlorophyll in any location and used to observe approximate levels of plant productivity. Given the relatively small scale of agriculture in Belize, we derive the NDVI using the 250m vegetation products from the MODIS sensors.

3. *CHIRPS Rainfall: Time Series Data*

We also examine the time series properties of rainfall as measured by the Climate Hazards Group Infrared Precipitation with Station data (CHIRPS). CHIRPS is a 30+ year quasi-global rainfall dataset. CHIRPS incorporates 0.05° resolution satellite imagery with in-situ station data to create gridded rainfall time

1 *https://www.worldbank.org/en/research/brief/software-for-poverty-mapping*
2 *https://sentinel.esa.int*
3 *https://modis.gsfc.nasa.gov*

series for trend analysis and seasonal drought monitoring. In this case we resample the rain data to 75m spatial resolution to ensure that each enumeration area has an observation associated with it. We collect precipitation by *dekad*[5] (Funk et al. 2014). There are three dekads in a month, the first two being 10 days long, and the third being the remaining days in the month. Because CHIRPS data has a similar high frequency and availability as MODIS data above, we provide the denser set of summary statistics outlined in Table 2 for low-spatial resolution data.

4. Synthetic Aperture Radar

We utilize the Japanese sensor PALSAR/PALSAR-2 mosaic data to provide 25m resolution synthetic aperture radar (SAR) data. SAR can be used to create three-dimensional reconstruction of objects, such as mountains and landscapes (Kirscht 1998; Kirscht and Rinke 1998). SAR has been successfully used to create global forest/non-forest maps (Shimada et al. 2014), for assisting in remote crop classification (McNairn et al. 2009), to mapping flooding events (Shan et al. 2010).

Methods

Machine Learning Small Area Estimation

Most surveys which measure income or consumption do not sample all areas where policy makers would like to understand poverty or welfare. Many low-income areas are largely inaccessible or sparsely populated. Even when they do sample all areas, there may not be sufficient observations to generate welfare statistics at the spatial resolution required to inform policy decisions. As a result, most estimates of welfare at the local level are generated through small area estimation, techniques which typically match a target survey, which measures the variable of interest (poverty, consumption or income), and a census, which contains sufficient observations from which one can accurately calculate welfare. These procedures first estimate household level models of income using the consumption survey, then use these models to predict household level income or consumption in the census.

We follow this methodology but deviate in that we use machine learning methods rather than parametric approaches, as used in Elbers et al., (2003), to estimate household level models of income. We use machine learning methods because they have favorable properties for building poverty maps when the number of variables used is large, as is the case with our models where we have access to a rich set of satellite-derived variables (Afzal et al., 2015).

We utilize four machine learning models to estimate household level income and eventually local area poverty: 1) Ridge Regression (Hoerl and Kennard, 1970), 2) Elastic Net regression (Zou and Hastie, 2005), 3) Random Forest (Breiman, 2001), and 4) Extreme Gradient Boosted Trees (Friedman, 2001). We lastly estimate model 5) which creates a simple average of the four estimated models.

A detailed discussion of the methods is beyond the scope of this article[4], however, one important note is that the first two methods are linear models that use machine learning for variable selection. The second two are tree-based ensemble methods of multiple regression trees. This is important in that linear models may perform worse in extrapolation if the space into which they are predicting are sufficiently different from the areas in which they have been estimated. The tree models (models 3-4) may be more robust in their predictions.
The Appendix contains additional information on the machine

Satellite Imagery Methods Overview

learning modeling process, and we refer the curious reader there. Two primary methods are used to extract useful information from raw remotely sensed imagery, contextual features and time-series features. The original, raw image of a location, while useful for human interpretation, provides little useful information for poverty modeling. Instead we must devise ways to extract specific

4 *For more detail on the methods we refer the reader to the book Gareth et al., (2013) available at: http://faculty.marshall.usc.edu/gareth-james/ISL/*

5 *Dekad refers to a period of ten days*

components of an image that might be useful. For instance, does the image have many long lines pointing in the same direction, or are they short and complex? In this case a residential area with lots of short complex lines might indicate that it is an informal settlement. This type of information will be captured by "Contextual Features" in the following section. Or alternatively we can look a series of images over time and see if there were any sudden shifts in rainfall or greenness over time that may impact agricultural productivity. This type of information will be described in "Time-Series Features" in a later section.

Contextual Features

One powerful set of methods to summarize satellite images is known as contextual features. Contextual features are information that represent the spatial and spectral values derived from satellite imagery based on neighborhoods or groups of pixels. In the past we have shown that these features are strongly correlated with population and poverty variation within Sri Lanka and Ghana (Engstrom et al., n.d.). For the most part, past research of this nature has used very high spatial resolution imagery (2m spatial resolution and lower). While these data provide a tremendous amount of detail, there are major drawbacks including high cost, and difficulty covering large areas. Recent research has used data from the freely available, Sentinel-2 sensors, which have extensive spatial coverage, thus allowing us to easily and freely, collect imagery over large areas (i.e., entire countries).

A cloud free image, Sentinel-2 mosaic was used as the input to calculate contextual features using the Python package SpFeas. SpFeas is an open-source Python library for processing contextual image features from satellite imagery. The 11 contextual features calculated are as follows:

Name	Description	Interpretation	Source
Gabor Filter	A linear Gaussian filter used for edge detection	Finds edges of buildings and determines if they are in similar directions.	(Mehrotra, Namuduri, and Ranganathan 1992)
Histogram of Oriented Gradients (HOG)	Captures the orientation and magnitude of the shades of the image	Finds the orientation of edges of buildings and groups them together.	(Dalal and Triggs 2005)
Lacunarity (LAC)	Describes the extent of gaps and holes in a texture.	Finds gaps within areas. Can determine if buildings are close together or have space between.	(Myint, Mesev, and Lam 2006)
Local Binary Patterns Moments (LBPM)	Define contiguous regions of pixel groups and sorts them into a histogram	Finds buildings and neighborhoods of different sizes.	(Ojala, Pietikäinen, and Mäenpää 2002)
Line Support Regions (LSR)	Characterize line attributes	Characterizes the lengths of lines, typically roads and building edges.	(Ünsalan and Boyer 2005)
Normalized Difference Vegetation Index (NDVI)	The most widely used vegetation index that provides information about the health and amount of vegetation	Determines the presence or absence of vegetation.	(C J Tucker 1979)
Oriented FAST and Rotated Brief (ORB)	Selects key points for image matching and object recognition.) It is similar Speeded Up Robust Features (SURF).	Finds bright things such as buildings in imagery.	(Rublee et al. 2011)
PanTex	Is a built-up presence index derived from the grey-level co-occurrence matrix	Used to determine if areas have buildings or not. If buildings present, can help understand size.	(Pesaresi, Gerhardinger, and Kayitakire 2008)
Structural Feature Sets (SFS)	Statistical measures to extract the structural features of direction lines	Finds road and building edges and characterizes the size and length.	(Huang, Zhang, and Li 2007)
Fourier Transform	Detects high or low frequency of lines	Can be used to determine if neighbourhoods are on a grid pattern.	
Mean	The average brightness in the Blue, Green, and NIR bands	Finds bright and dark areas. Can help find vegetation.	

Table 1: *Description of contextual features used in the analysis*

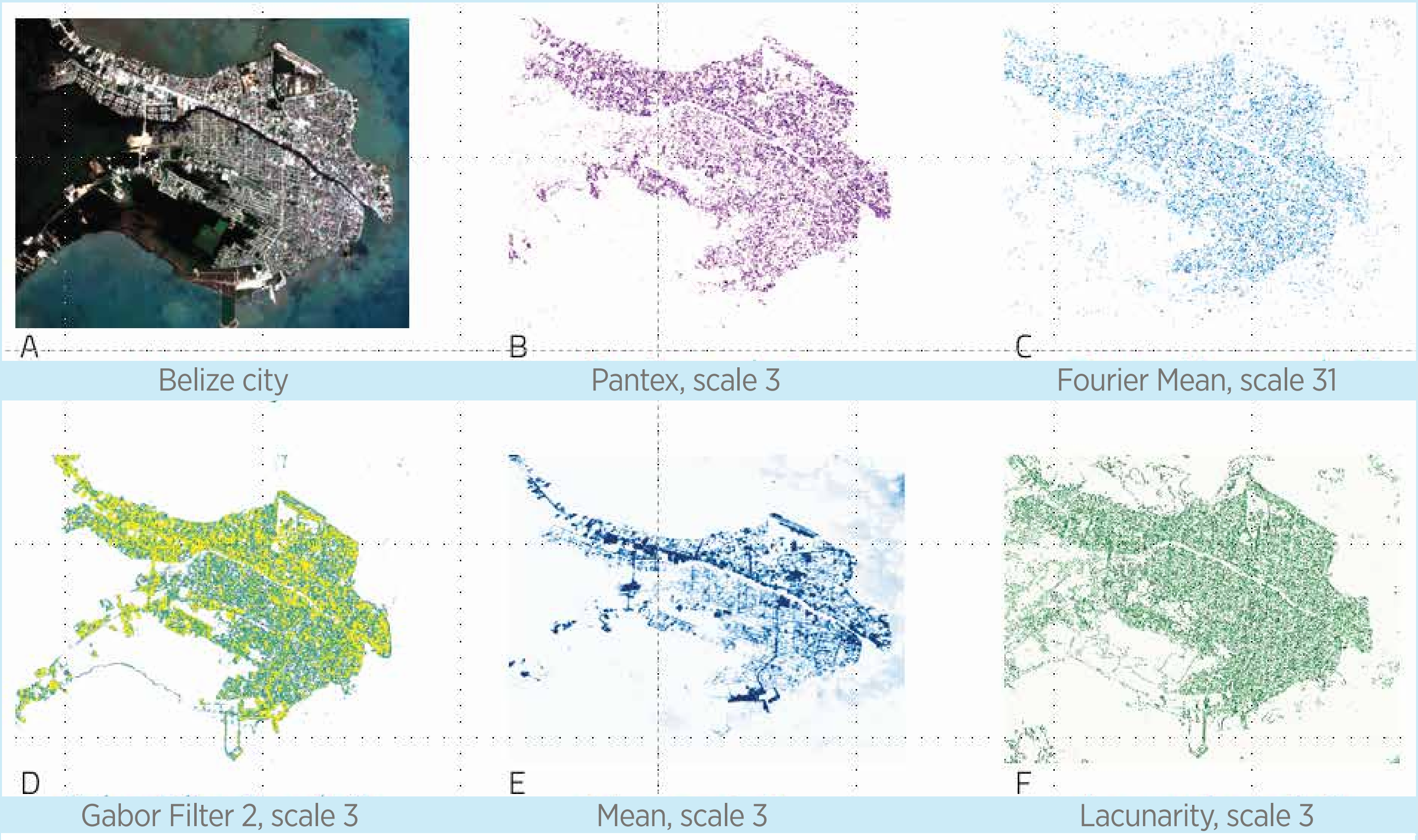


***Figure 1** Belize City contextual features*
*Displayed in **A** is the true color (Blue, Green, and Red) Sentinel 2 image for Belize City, **B** is the Pantex measure derived for this area at scale 3 (30m), **C** is the Mean Fourier transform at scale 31 (310m), D) is the second Gabor Filter at scale 3 (30m), **E** is the Mean brightness at scale 3 (30m), and **F** is the Lacunarity measure at scale 3 (30m).*

Contextual features are created by comparing central pixels with their neighbors and then reporting this value back to the central pixel (in this case 10m). Thus, contextual features measure the "context" in which an individual pixel is situated, using information from surrounding pixels. The number of neighboring pixels considered in the comparison is the "scale", which varies by the feature being calculated.

When applied to satellite imagery, the features capture "texture" or spatial variability and spectral values of neighborhoods. As an example, the Pantex feature captures the minimum contrast between a pixel and its neighbors. Highly built-up neighborhoods tend to have greater contrast in all directions, which will create high values of this feature. In contrast, in rural areas the pixel's brightness will likely be similar to a neighbor in at least one direction, which will create a low minimum contrast. To help visualize what contextual features capture we present a number of features for Belize City in Figure 1. For more detail on Contextual Features please refer to the Appendix.

While the contextual features we developed may not predict as well as features derived from convolutional neural networks, the latter require many times more data than are currently available in most poverty surveys. In contrast, our method does not require a massive amount of training data, are simpler and cheaper to calculate, and have proven robust in other settings.

Time-Series Features

To complement the contextual features described above, we also calculate several time series properties. Time series properties can play an important role in predicting household well-being. For instance, if an agricultural community has experienced below average rainfall for the last five years, this can be determined by looking at the time series for precipitation. Moreover, a variety of statistics can provide invaluable information, for instance, the maximum greenness of an agricultural area is correlated with agricultural yields and plant productivity (Mann, Warner, and Malik 2019; Mann and Warner 2017). Time series can also pick up on the effects of drought, flooding, or even the slow sustained loss in productivity.

1. Dense temporal feature extraction

TS-Raster (TS) is a python package for analyzing time-series characteristics from raster data. It allows feature extraction, dimension reduction and applications of machine learning techniques for geospatial data.

TS's primary significance is the ability to provide an extensive set of time-series properties, including simple metrics like minimums or maximums, but also more complex ones like the number of peaks observed within a year, or the number of observations above or below the mean. TS should be able to meaningfully characterize the time series of high frequency data products like those from MODIS or CHIRPS. For a visual example of what kinds of properties TS extracts see Figure 2.

A summary of time series attributes are provided below (Tables 2). The feature name indicates the naming convention used for data storage, the description provides a simplified description of that statistic, and use descriptions provide some context for how that attribute might be useful in our modeling. Table 2 provides a partial list of statistics collected from data with very high temporal resolution (MODIS, CHIRPS). More summary statistics can be provided for this data because the time series has more observations (see Table 2A in the appendix), and therefore is more complex. Table 6 in the appendix, provides a full list of statistics gathered from sensors that are collected less often (ranging from once every 5 to 30 days depending on cloud cover). This is temporal data from sensors such as Sentinel-2 and PALSAR/PALSAR-2.

2. LandTrendr

LandTrendr (LT) is a broadly used algorithm that detects sudden shifts in an index. For this study we examined NDVI (a satellite measure of vegetation health), on a pixel-by-pixel basis. Due to it's lack of importance in the final models, we have moved the description of this data product to the appendix.

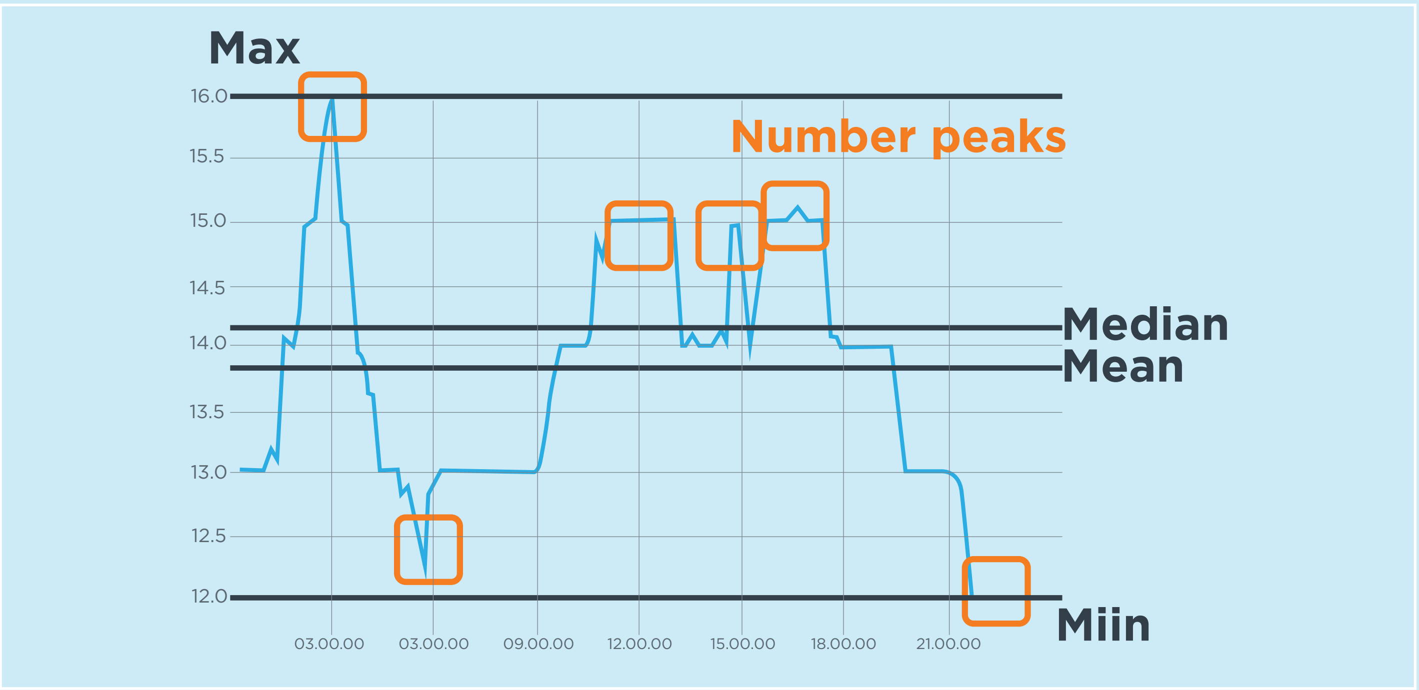
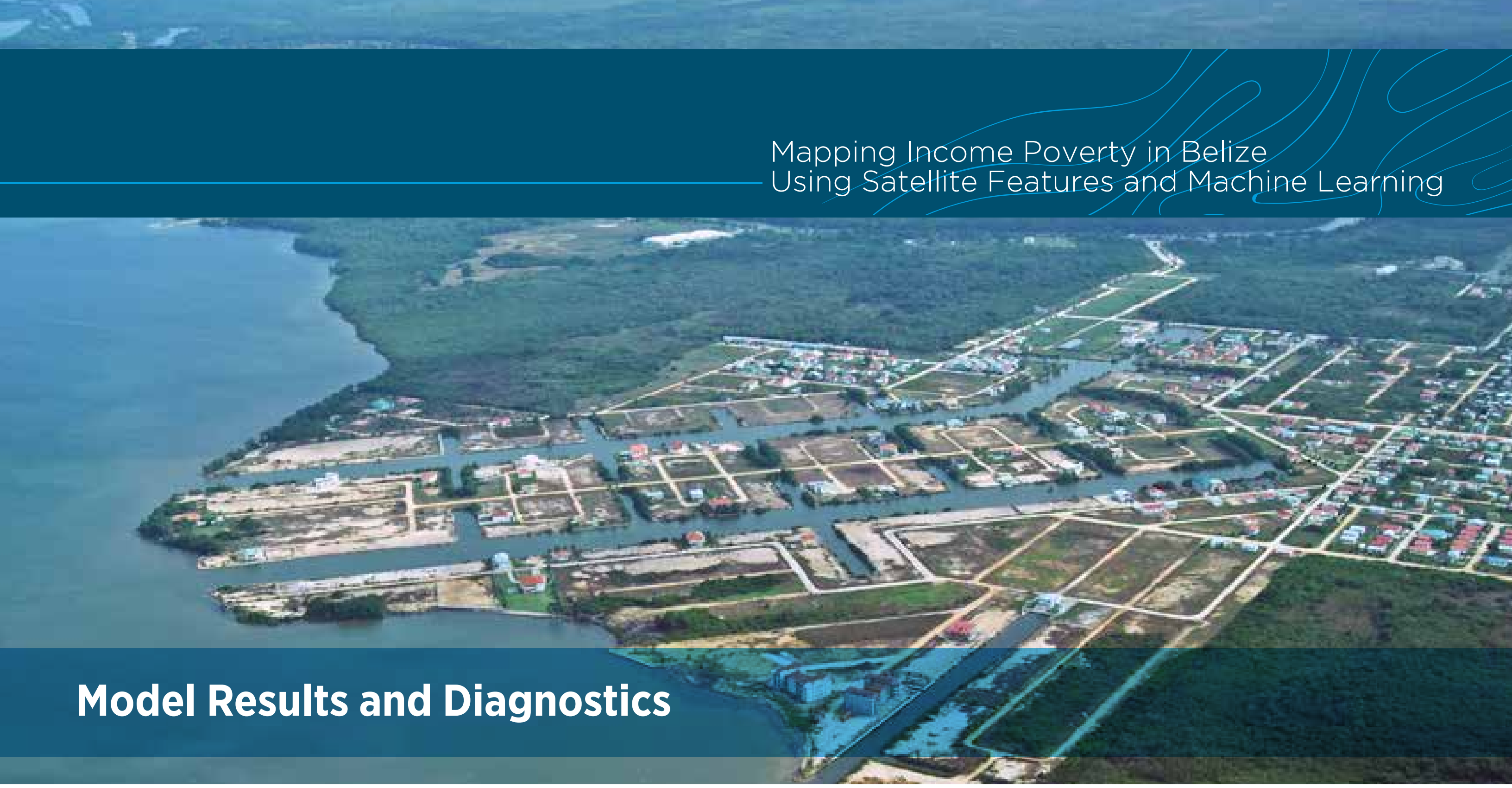


Figure 2 *Examples of TS-Raster time series properties*
In this example the light blue line is plot of a time series for a single pixel in an image. Red boxes are used to highlight a series of time series attributes that can be extracted with TS-Raster. '

Opportunities	Challenges	Interpretation
maximum	Global maximum value	Highest observed greenness / rainfall
mean	Global mean value	Average observed greenness / rainfall
mean_change	Average change between any two periods in series	Instability in time series (irregular rain)
median	Global median value	Average observed greenness / rainfall
minimum	Global minimum value	Minimum observed greenness / rainfall
sum_values	Sum of all values across the time period	Excess or insufficient rain, or crop failure
agg_linear_trend_f_agg“max-”__ chunk_len_6_attr“slope”	Maximum observed trend during any 6 periods	Sudden positive shocks (flood)
agg_linear_trend_f_agg“min”__ chunk_len_6_attr“slope”	Minimum observed trend during any 6 periods	Sudden negative shocks (drought, land use change)
longest_strike_above_mean	Maximum number of observations sustained above the mean value	Measure of sustained excess rain
longest_strike_below_mean	Maximum number of observations sustained above the mean value	Measure of sustained drought
count_above_mean	Number of observations above the global mean	Persistent shifts up (increase in rainfall)
count_below_mean	Number of observations below the global mean	Persistent shifts down (decrease in rainfall)

Table 2 Description of high temporal resolution time-series feature



Model Results and Diagnostics

To validate the household-level income models that we used to produce sub-national estimates of poverty, we first examined diagnostics from those household models in Table 3. We show results for the separate machine learning models used, as well as for the set of variables the model can access: 1) satellite variables and survey variables, 2) variables in the LFS and census surveys only, and 3) the satellite derived variables only. We consider two performance metrics: root mean squared error (RMSE), which gives a measure of average error interpreted in units of log household income, and R^2, which measures the coefficient of determination between predicted household income and true household income. Model performance metrics at the household level are shown in Table 3. Note that lower RMSE values indicate a better model performance. An R^2 close to 1 indicates a perfect relationship between the predicted and true household income, and a value of 0 indicates no relationship between predicted and true.

To better characterize our true out-of-sample performance, remember that we separated our dataset into two groups. The first 75% of observations, called the "training" set, is used to fit all models. The remaining 25% of observations are held out as an independent "testing" set and provides a much more realistic measure of model performance.

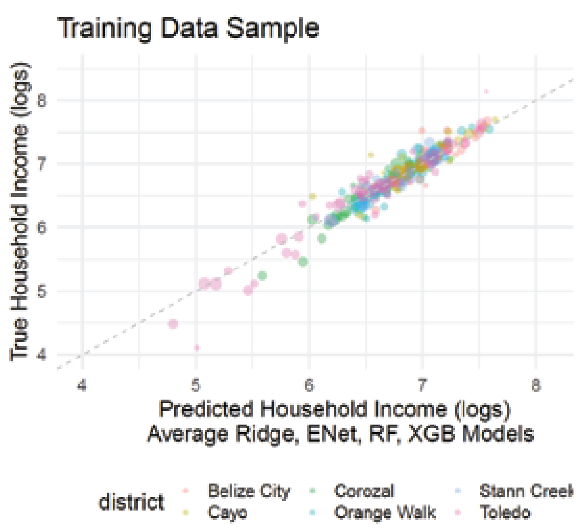
1. *Model Performance With and Without Satellite Variables*

Comparing the performance in the training and test sets, we note that in-sample performance on the training set significantly overstates model performance. R^2 values indicate we explain between 42%-86% of the variation in household income when just looking at performance in the training set (using the survey & satellite variables). Out-of-sample performance in the testing set drop as expected. R^2 values indicate that we can explain 30-36% of variation at the household level using survey & satellite variable models. In-sample model performance can exceed 80%. Clearly, any study reporting solely results from the training set would dramatically overstate the true explanatory power of their model. Comparing across the set of variables employed, we see, unsurprisingly, that models with the most variables – survey & satellite -- tend to perform best. This is followed by models that use information only available in the survey, which have R^2 values that vary between 0.30-0.34, indicating we can reliably explain between 30-34% of the variation in household income using survey variables alone. In comparison, models that use only contextual and time series satellite derived variables can explain 12.4%-14.4% variation in household level income.

2. *Machine Learning Model Comparison*

Across machine learning models we see high levels of variability across training set model performance. Meanwhile, actual performance in the test set is more consistent between machine learning models. In the test set R^2 values vary between 0.30-0.36, for the preferred models using satellite and survey variables. The best performing individual model is Extreme Gradient Boosted Trees, with an R^2 of 0.349, followed by the Elastic Net model with an R^2 of 0.34. Ridge performs only slightly worse with a test-set R^2 of 0.338, and finally random forest performs the worst with an R^2 of 0.298. The highest performing model overall is the combined model which averages all the model predictions, which has combined R^2 score of 0.36, and an RMSE value of 0.65.

Model	Test (LFS, 25%) RWSE	Test (LFS, 25%) R2	Train (LFS, 75%) RWSE	Train (LFS, 75%) R2	Variable Set
Elastic Net	0.656	0.340	0.652	0.418	Elastic Net
Ridge	0.660	0.338	0.641	0.437	Ridge
Extreme Gradient Boosted Trees	0.654	0.349	0.538	0.613	Extreme Gradient Boosted Trees
Random Forest	0.676	0.298	0.369	0.861	Random Forest
Combined (Average of Models)	0.645	0.360	0.536	0.630	Combined (Average of Models)
Elastic Net	0.675	0.302	0.695	0.337	Elastic Net
Ridge	0.676	0.302	0.695	0.338	Ridge
Extreme Gradient Boosted Trees	0.656	0.342	0.621	0.475	Extreme Gradient Boosted Trees
Random Forest	0.650	0.351	0.357	0.876	Random Forest
Combined (Average of Models)	0.653	0.344	0.582	0.560	Combined (Average of Models)
Elastic Net	0.759	0.126	0.741	0.247	Elastic Net
Ridge	0.760	0.126	0.738	0.253	Ridge
Extreme Gradient Boosted Trees	0.749	0.144	0.730	0.274	Extreme Gradient Boosted Trees
Random Forest	0.775	0.125	0.709	0.310	Random Forest
Combined (Average of Models)	0.753	0.138	0.723	0.286	Combined (Average of Models)



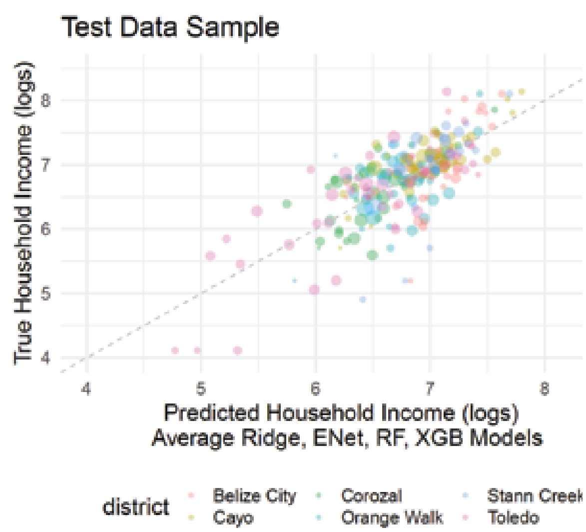


Figure 3 *Predicted versus true plots for ED level household income predictions*
These plots show estimated average income versus true average income using the "combined" model averaging estimates between the Ridge, Elastic Net, Random Forest, and Extreme Gradient Boosted Trees models. Note the training sample is the data on which our model has been estimated, and the test data sample is the validation sample, which was not used to directly estimate our model.

3. *Enumeration District Level Prediction Comparisons*

Satisfied with a household level model of income that performs adequately well, we then generate predictions for every household in the Census. These household level predictions are averaged at the ED level to generate average ED income. Results in the previous analysis lead us to believe the best performing model is the combined model, which averages household income predictions across the four machine learning models.

Figure 3 above shows predicted versus true plots at the enumeration district level in the testing and training samples of the data. Each point shows, therefore, the predicted average income for an ED (on the x-axis) against the true average income (on the y-axis). We scale the size of each point by the total population in each ED, and color code them by district. We see a better fit to the 45 degree line (indicating perfect prediction performance) for the training dataset (data which was used to estimate the model) versus the test dataset (data used to validate the model). However, we see consistently good performance in both of these datasets.

To complement the information in Figure 3, we present numerical summaries of the same information in Table 4 . We see that the R^2, or the coefficient of variation between predicted average ED incomes is highest when using both the satellite and survey variables. In the test set, the most reliable measure of out-of-sample performance, we see an R^2 value of 0.55 using both satellite and survey variables. This falls to 0.50 when we use survey variables only. When we use only satellite variables, we find an R^2 performance of about 0.30, indicating satellite variables alone explain 30% of the variation in ED level average incomes. Again, we see the large decline in performance from the training versus test sets, indicating a clear need to use testing sets as proper validation of performance.

R2	RWSE	DATA	Variable Set
0.553	0.438	Test	Satellite & Survey
0.298	0.554	Test	Satellite Only
0.507	0.462	Test	Survey Only
0.91	0.165	Train	Satellite & Survey
0.883	0.195	Train	Satellite Only
0.707	0.296	Train	Survey Only

Table 4 *Enumeration District Level Performance Metrics. Metrics compare true ED average income with predicted average ED income using the "Combined" model, with variable sets.*

Model Validation and Residuals Diagnostics

1. *Variable Importance*

Exploring which variables – derived from satellites or surveys -- have the most impact on the predictive power of the models is complicated by the fact that there are so many variables, and the variables interact non-linearly in our tree based models (random forests and extreme gradient boosted trees). With simple linear models applied to a more reduced set of variables, we can view estimated coefficients and the resulting t-statistics to determine statistical significance. This is not feasible here given the large set of variables used. One way we can examine which variables improve model performance is by calculating variable importance. This at least provides the relative importance for each variable within a given machine learning model. Variable importance for linear models – Ridge and Elastic Net -- is calculated as the absolute value of the t-statistics, which is the coefficient divided by the standard error. For the tree-based models, the variable importance is calculated by averaging all the trees that do not contain a particular variable[6], and comparing mean decrease in final classification purity (or accuracy) against models that do contain these variables. Again, we cannot compare how much each variable improves the models across model types, but we can compare within models, thus we scale the variables importance scores within a model such that 100 is the most important variable, and 50 is half as important as the most important variable.

In the appendix, Figures 6 and 7 present the variable importance metrics across the four machine learning models, plotting the top 30 most important variables for each. The Ridge, Random Forests, and Extreme Gradient Boosted Trees select head of household years of education as the most important variable. Elastic Net models also considers this variable important, selecting it third behind number of dependents and number of children. Next most important variables are a set of asset variables – whether a household has high quality cooking fuel, toilet, number of computers, cable access, TV, refrigerator, number of vehicles, and electric washers.

Following these variables, we see several satellite derived variables appear, both generated from the time series as well as contextual cross-sectional information. For the time series variables, the longest period below or above mean precipitation tend to be important variables, indicating that areas with long periods of drought or excess rain correlates with changes in income, likely in agricultural areas. Similarly, streaks of NDVI above or below the average level for these areas tend to be predictive of income. Contextual satellite information also appear to be strongly predictive of average incomes. Elastic Net models pick many of these variables – NDVI at 3 pixel scale, SFS at 31 and 71 pixel scale, oriented fast and rotated brief (ORB) at scale 71 pixels, and local binary pattern moments at 3 pixel scale.

6 Recall that in these models variables are randomly selected at each node and therefore some trees will not contain particular variables.

Overall, while variables derived from surveys tend to be the strongest predictors of household income, almost all machine learning models improve with the addition of satellite derived variables, and within models outside of the top 5 or 10 most important variables, satellite features tend to be strongly predictive of household income.

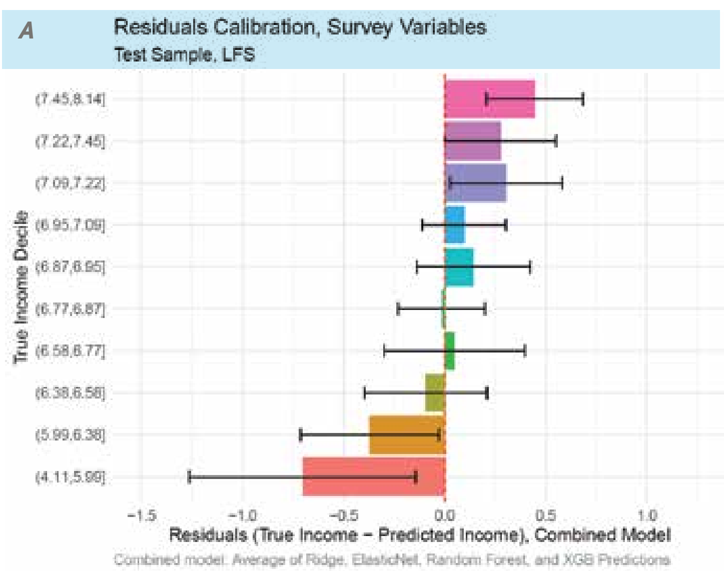
2. *Residuals Diagnostics*

One concern with small area estimates is that our model may be biased, not on average, but for particular sub-populations whose outcomes we would like to measure with high precision. For instance we need to ensure we are not producing biased estimates of incomes for the poorest populations. Therefore, residuals diagnostics by subgroups is a crucial component of any small estimation model.

Figure 4 A presents, for the survey only models, the average residuals (true household income minus predicted income) and standard errors by household income decile in the test sample. We use the predictions from the combined model, which calculates the predicted household income as the mean of the four machine learning models. For each income decile, we calculate the average (bar graph length) and the estimated standard error (black bar). Note that the lowest errors are seen for households in the middle of the income distribution, from log incomes of 6.38 to 7.22. For households within this range, errors are roughly symmetric around zero, and small in magnitude. This indicates we can assume, for households within this income range, our models of income are accurate and unbiased.

As we move to the two highest and lowest income deciles, we see the average residuals grow. For the richest households, residuals are positive, indicating we under predict income for these households. For the poorest households, residuals are negative, indicating we over predict incomes for these households. In general, our models of poverty tend to understate the true variance of household incomes.

However, in comparing the top panel A (survey variables only) with the bottom panel B (satellite and survey variables) we see that the residuals for the poorest and richest households are smaller for the satellite and survey models. In particular, residuals for the poorest decile decline from around -0.75 to -0.5. The black standard error bars show that the two poorest deciles for the survey only models (panel A) are biased, whereas the standard errors bars for the satellite and survey models (panel B) cross the zero threshold. We see similar improvement in the richest deciles when comparing the satellite and survey to the survey models alone, indicating satellite features help predict rich households as well. Taken together, the satellite features help recover critical characteristics of the most important income deciles.



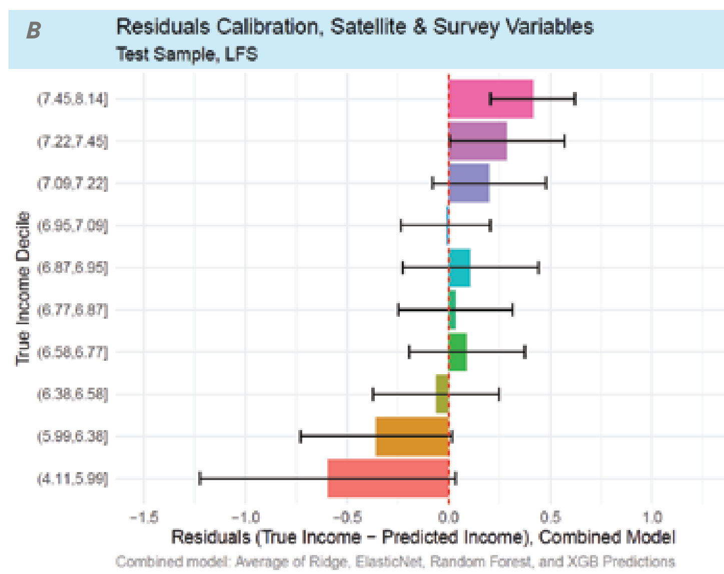


Figure 4 A-B *Residuals calibration, survey only models and satellite & survey models*
These plots show average error by actual income quintile. Black bars show estimated standard errors specific to each quintile.

Map Results and Implications for Policy Design

To calculate poverty scores, we need to choose a poverty line, and calculate headcount poverty rates, or FGT0 counts (Foster, Greer, Thorbecke, 1984). Because of data release restrictions, we calculate poverty lines at the 5/10/15 and 20th percentiles of national income. One important caveat with these maps is that while they provide an excellent view on the distribution of poverty within Belize, as currently designed they cannot provide an update to the national poverty rate. In order to do so, Belize would need to conduct a consumption survey that would provide the input to the machine learning models in substitution for the Labor Force Surveys used. However, using Big Data and machine learning paired with a consumption survey for this purpose will likely reduce the total costs of the consumption survey, given that Big Data has shown to reduce the number of households needed to survey for a given level of survey accuracy.

The modeling calibration section showed the best model performance in the test set was the combined model, which averaged all household predictions together. We further define the poverty status of each household using an algorithm where each machine learning model "votes" whether a household is poor or not, details of which are provided in the appendix.

Discussion of Poverty Maps

The maps showing poverty rate using these four relative poverty lines (5/10/15 and 20th percentiles of national income) are presented in figure form in the appendix. We see from these figures that the poorest districts are Corozal, in the north, and Toledo, in the south. If we use a threshold of the 5th percentile of national income as our poverty line, it appears that almost all of the poverty is confined to these two areas. The last poverty map for Belize was produced by the Statistical Institute of Belize in 2009 using information in the 2009 Census. For the 2009 map, four districts were classified as having high poverty rates – Corozal, Orange walk, Stann Creek and Toledo. Viewing our analysis in light of the previous map, it appears there has been significant reduction in poverty for the districts of Orange Walk and Stann Creek, which is immediately apparent when viewing the 5th percentile of national income relative poverty map. Our new poverty maps have enumeration district as their resolution, and the previous maps only provide district level poverty disaggregation, thus it's difficult to make a direct comparison.

Given this higher resolution of a poverty map – at the enumeration district rather than district level – other interesting patterns emerge as well. Even within a poorer region such as Stann Creek we see there is substantial heterogeneity in the relative poverty rate. The city of Dangrige appears significantly less poor than the surrounding areas. Whether this indicates improvement from 2009 we cannot say as the previous poverty map did not produce enumeration district poverty levels. The city of Punta Gorda in Toledo, itself a relatively poorer district, appears to be lower in poverty and surrounded by higher poverty EDs. Interestingly the city of Corozul in Corozal still appears as poor as surrounding areas, a pattern that is different from the previous districts discussed.

Given the fact that satellite features have been shown in the previous section to improve the modeling of rural (poorer) households, it is likely that the inclusion of satellite variables allows for the increased accuracy of rural households. This allows

us to see details in the poverty rates of households surrounding cities and not restricting to the cities themselves.

Cost Discussion and Considerations For Implementation

The prototypical government agency considering incorporating open source satellite features into their statistical agency should be aware of likely costs for building the satellite features pipeline, as shown in the table below. One important note is that these are almost all fixed costs. Improving poverty modeling for the poorest households in other ways – such as redesigning surveys or increasing sample sizes – add to variable costs.

Countries which are likely to benefit the most from open source satellite features are those with substantial rural poverty, whose traditional surveys are not capturing sufficient variation in rural incomes. To some degree satellite features and surveys are substitutes, thus countries who are already performing frequent poverty maps are unlikely to benefit as much from these efforts as those who produce infrequent poverty maps.

Implications for Policy Design

The substantial within-district heterogeneity in poverty rates discussed above suggests certain policy mechanisms may be more effective at eradicating poverty than others. First, blanket untargeted transfers are not likely to be cost-effective given the starkly contrasting poverty rates across EDs. It's clear that poverty is a phenomenon that is not restricted to one district and thus means tested transfers and subsidies are likely to be more effective than blanket transfer at direct poverty reduction. Other fiscal policies under consideration such as educational or labor programs should additionally consider firstly the greater poverty in Corozal and Toledo, but secondly the differing poverty within districts.

Frequently updated poverty maps can aid in the design of fiscal and means-tested programs. By viewing changes in ED level poverty rates the government can build more cost-effective programs that accomplish the goal of poverty reduction at a lower cost.

One important added benefit of producing poverty maps is that now the government has at its disposal predicted incomes for every household in the country. The government could implement a policy simulation sandbox whereby each policy under consideration is modeled in a general equilibrium framework, and new simulated household incomes are derived. The government could in this fashion forecast the cost-benefit tradeoffs of policy before it is implemented, which is all the more important given the distributed nature of poverty in the country. Costs for such an enterprise would range from $250,000 to $750,000 USD and require a couple staff to design and score each policy. The cost savings in terms of policy implementation would likely be many times that cost.

R2	Time	Cost
Workshop with		
Statistical agency	3 days	$30,000
Project Investigation	2-6 months	$30,000
Project Research and Implementation	6-9 months	$100,000-$250,000
Capacity Building	2-5 months	$100,000
Data Warehousing and API Building	3-6 months	$100,000

Table 5 *Cost Discussion of Open Source Satellite Features Poverty Map*

Conclusion

What can the proto-typical department of Statistics learn from this example in Belize? We have shown that open-source satellite features have the potential to improve poverty estimation, and that they can be incorporated into a machine learning prediction framework with relative ease. We document a 9% improvement in average coefficient of variation when these models incorporate open-source satellite features. However, these averages do not tell the whole story. The model substantially improves the accuracy of income estimates for the poorest households when we add these satellite variables, indicating they hold great potential for identifying the poorest of the poor. This ability to better differentiate the poorest communities, at a high degree of spatial resolution, is critical to the meaningful targeting of poverty interventions.

In light of these results, what is the value of "cheap" Big Data? Given the typical poverty modeling approach of merging a Census to a Survey, this necessarily restricts models to intersections of questions asked in both questionnaires. We have seen that using only survey information, our models perform poorly for the poorest households. Big Data, in this case from satellite imagery, provides information not captured by surveys, suggesting its value could be very high.

Appendix

Machine Learning Poverty Modeling Methods

The following section outlines our use of machine learning for small-area estimation.

Most surveys which measure income or consumption do not sample all areas where policy makers would like to know estimates of poverty or welfare. Many low-income areas are largely inaccessible or sparsely populated. Even when they do sample all areas, there may not be sufficient observations to generate welfare statistics at a spatial resolution. As a result, most estimates of welfare at the local level are generated through small area estimation, techniques which typically match a target survey, which measures the variable of interest (poverty, consumption or income), and a census, which contains sufficient observations from which one can accurately calculate welfare. This approach requires a model of household-level welfare, $y_{h,c}$, which is the income or consumption level of household h measured in local area c. One approach is to assume a (log) linear relationship between household characteristics $\chi_{h,c,k}$ and income/consumption which takes the form:

$$\ln\ y_{h,c} = \sum_{k=1}^{K} \chi_{h,c,k} \cdot \beta_k + \epsilon_{h,c}$$

where $\epsilon_{h,c} \sim N(0,\sigma)$ is the unexplainable error term and we have k linear coefficients, β_k, to use to model income or consumption. Note that if we build a sufficiently accurate model of income or consumption, and the true income process is linear, $\widehat{y_{h,c}} = y_{h,c} = \sum_{k=1}^{K} \chi_{h,c,k}\, \beta_k$. If we restrict our household characteristics $\chi_{h,c,k}$ to those that are also available in the census, and we have recovered a model of household income or consumption that remains sufficiently accurate in the survey and census, we can apply the parameter estimates obtained from the census $\widehat{\beta_k}$ to derive $\widehat{y_{h,c}}$ for every household in the census. This provides a method to compute welfare for each household in the census. The resulting estimate of welfare is:

$$\ln \overline{\widehat{y}}_c = \frac{1}{n_{,c}} \sum_{h \epsilon c} \ln \widehat{y}_{h,c} = \sum_{k=1}^{K} \chi_{h,c,k} \cdot \widehat{\beta}_k$$

where $n_{,c}$ is the number of households or population in cluster c, and $\ln \overline{\widehat{y}}_c$ is the average welfare statistic of interest in the cluster.

Several problems may arise. For one, the true relationship between welfare $y_{h,c}$ and household characteristics $\chi_{h,c}$ may be non-linear and difficult to model via ordinary linear regression (OLS). A common refinement of this method is to use simulation methods to improve the asymptotic properties of $\overline{\widehat{y}}_c$. We take a different approach, and instead use machine learning to model $\widehat{y}_{h,c}$ under the logic that machine learning will recover a better household-level model of welfare. In particular we aim to better capture the likely non-linear relationship between satellite and census features and income. We estimate models of the form:

$$\ln y_{h,c} = f\left(X_{h,c}\right) + \epsilon_{h,c}$$

where $f(.)$ is estimated using four separate machine learning models. We estimate the following four models: 1) Ridge Regression (Hoerl and Kennard, 1970), 2) Elastic Net regression (Zou and Hastie, 2005), 3) Random Forest (Breiman, 2001), and 4) Extreme Gradient Boosted Trees (Friedman, 2001). We lastly estimate model 5) Combination, which creates a simple average of the four estimated models. This last model is a simple ensemble of several models that may retain the separate strengths of each model.

Ensemble Poverty Metric

We calculate what we define as the "ensemble" poverty metric, as shown in the equation below. Suppose we want to know the poverty status of a household h in cluster c. For each household, we have calculated predicted income or consumption $\widehat{y^{m}_{h,c}}$ from each model m, from a total of M models. The poverty status of each household is defined as:

$$poverty_{h,c} = \frac{1}{M} \sum_{m=1}^{M} 1\left(\widehat{y^{m}_{h,c}} < ProvLine\right)$$

Where $poverty_{h,c}$ is the poverty status as defined by the ensemble poverty metric. Each model "votes" according to whether it defines a given household as being poor or not. In using the ensemble poverty metric we can remain agnostic as to the best model, as some models may perform better for one subgroup or another. In this fashion, the ensemble poverty metric better captures model uncertainty than relying on one model alone.

Model Estimation

Across all four machine learning prediction methods we use repeated five-fold cross-validation to tune necessary parameters in the model. The shrinkage parameters for models 1 and 2 are selected via repeated five-fold cross-validation. For model 3 we utilize 1,000 regression trees and cross-validate the parameter 'mtry' using a grid from 1 to 40, in steps of 3. For model 4, we cross validate the following parameters: number of trees grown from 50 to 400, maximum tree depth from 1 to 5 in steps of 5, learning rate from 0.3 to 0.4, and variables sampled from 60% to 80%. All models are estimated in R using the package caret[7].

Contextual Features

In this analysis, we use scales of 3, 5, 7, which are squares of 3 pixels by 3 pixels, 5 pixels by 5 pixels, to 7 pixels by 7 pixels for the majority of features. This constitutes looking at an area of 30 meters, 50 meters, and 70 meters for the "neighborhoods" which will constitute the windows of analysis for our contextual features. For the features ORB, SFS, Fourier and LSR the scale was increased by a factor of 10 because these features need more area to properly capture the variation in the landscape.

Each of the contextual features may have several different outputs depending upon the statistical properties of the features as those features are calculated. For Pantex and Lacunarity, the actual values themselves are outputted. For NDVI, Mean, and Fourier, just the mean and variance are outputted. For HOG, LPBPM, and ORB, we output the Mean, Maximum, variance, Kurtosis, and Skewness for these measures. LSR outputs the line contrast, line length, and line mean. SFS maximum line length, minimum line length, mean, weight mean, standard deviation, and maximum ration of orthogonal angles. Finally, Gabor outputs mean and variance for each of the filters, which in this study we used 14.

In total this produces 46 total outputs for all of the features and, because each feature is run at 3 scales, in sum our method produces a total of 144 outputs from the contextual features.

The eventual geographic area to which we link these satellite features is the Enumeration District (ED), thus for each ED area we summarize the features using the mean, standard deviation, and the sum for each. Together this produces 432 contextual feature values, which summarize various contextual aspects of satellite imagery for each ED.

As is displayed in Figure 5, the spatial and spectral patterns of the urban area visible within the imagery is well captured by contextual features.

These features are primitive versions of the features constructed using machine learning techniques such as Convolutional Neural Networks (Jean et al. 2016). Both approaches summarize images by comparing pixels with their neighbors. The main difference is that the Convolutional Neural Networks require survey data on welfare to determine which features to calculate[8]. In other words, the computer selects parameters for layers of filters, which when applied to the imagery construct textures that are optimized to distinguish between low and higher welfare areas. In order for the computer to select the best parameters for these layers, the general method is to use millions of data points when training the algorithms. Because of the limited training data for poverty surveys, a method known as "transfer learning" is used to hot-start the intermediate layers of the convolutional neural network which defines the filters (Jean et al. 2016; Babenko et al. 2017), often using intermediate filters that have been trained against large corpuses of images such as ImageNet (Deng et al. 2009). In practice this assigns filters intended for the purpose of recognizing features in traditional photography to satellite images.

In contrast, the contextual features used in this analysis are constructed using pre-determined algorithms. Therefore, they are independent of the survey data. To be clear, both methods must use external information to inform the choice of filters which summarize the imagery. Both methods use pre-defined filters given the paucity of survey data, only ours are designed for summarizing satellite images and not photographs of dogs and cats.

LandTrendr

LandTrendr (LT) is a broadly used algorithm that detects sudden shifts in an index, in this case NDVI, on a pixel-by-pixel basis. Effectively LT fits local regressions, the uses a series of metrics to detect sudden shifts in slope, or intercept on a year-by-year basis. As such, LR is effective at identifying land cover change, for

7 http://caret.r-forge.r-project.org/
8 Convolutional neural networks are feed-forward neural networks that have one or more convolutional layer which as a filter that heightens or depresses image characteristics depending on

instance conversion of forest to agriculture, or agriculture to urban settlement. Because LT only needs one clear observation per year it is effective for use with high resolution data like Sentinel or Landsat another higher resolution satellite. LT has a series of underlying assumptions and parameters, which we will not cover here for the sake of clarity. Detailed information on the algorithm is available (Cohen et al. 2018; Kennedy, Yang, and Cohen 2010). For a visual example of LandTrendr's output see Figure 5 below:

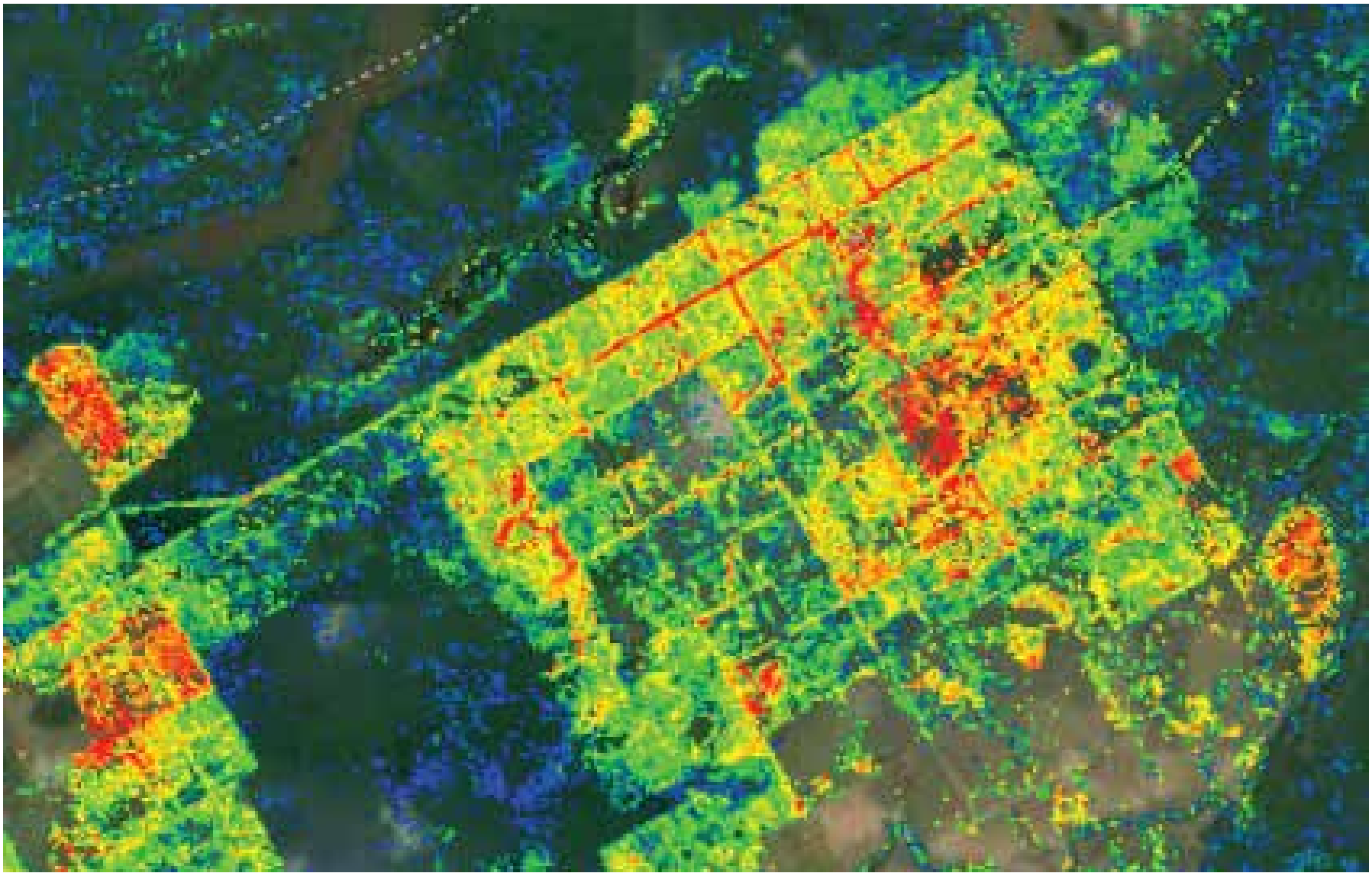


Figure 5 *LandTrendr magnitude of disturbance*
In the image above we can see that LandTrendr's algorithm can detect one-time shocks, such as paving or resurfacing roads, conversion of crop type (or perhaps crop loss). Areas in red indicate a strong negative shock, such as the creation of a road, or reduction in greenness. Areas in light green indicate mild shocks perhaps indicating typical planting, harvest cycle

Time Series Statistics

Name	Description	Interpretation
agg_linear_trend_f_agg"max"__chunk_len_6_attr"slope"	Maximum observed trend during any 6 periods	Sudden positive shocks (flood)
agg_linear_trend_f_agg"min"__chunk_len_6_attr"slope"	Minimum observed trend during any 6 periods	Sudden negative shocks (drought, land use change)
count_above_mean	Number of observations above the global mean	Persistent shifts up (increase in rainfall)
count_below_mean	Number of observations below the global mean	Persistent shifts down (decrease in rainfall)
last_location_of_maximum	Location of the periods maximum value	Time since maximum observed value (declining productivity)
last_location_of_minimum	Location of the periods minimum value	Time since minimum observed value (increasing productivity)
longest_strike_above_mean	Longest period of values observed above the global mean	Duration of persistent shifts up (flooding)
longest_strike_below_mean	Longest period of values observed below the global mean	Duration of persistent shifts down (drought)
maximum	Global maximum value	Highest observed greenness / rainfall
mean	Global mean value	Average observed greenness / rainfall
mean_change	Average change between any two periods in series	Instability in time series (irregular rain)
median	Global median value	Average observed greenness / rainfall
minimum	Global minimum value	Minimum observed greenness / rainfall
number_cwt_peaks__n_12	The highest number of peaks that occur in 12 periods	Number of crop rotations, unstable rainfall
number_cwt_peaks__n_6	The highest number of peaks that occur in 6 periods	Number of crop rotations, unstable rainfall
quantile__q_0.05	Value of the 5th percentile	Minima correcting for outliers
quantile__q_0.15	Value of the 15th percentile	Minima correcting for outliers
quantile__q_0.85	Value of the 85th percentile	Minima correcting for outliers
quantile__q_0.95	Value of the 95th percentile	Maxima correcting for outliers
ratio_beyond_r_sigma__r_2	Ratio of values that are more than 2*std(x) away from the mean	Frequency of extreme values, flooding, shocks
skewness	Ratio of values that are more than 3*std(x) away from the mean	Frequency of very extreme values, flooding, shocks
sum_values	Sample skewness of x (calculated with the adjusted G1 coefficient)	Changes in distribution over time

***Table 2A** Description of high temporal resolution time-series feature*

Name	Description	Interpretation
Med/Mn	Global median/mean NDVI	Average greenness, vegetation productivity
Min/Max	Global minimum/maximum NDVI	Max/Min vegetation productivity
P5/P25/P75/P95	5th, 25th, 75th, 95th percentile value NDVI	Min(Max)imal measures of greenness robust to outliers such as clouds
Sum	Sum of all NDVI values	Persistence of vegetation and productivity
Std	Standard deviation of NDVI values	Stability of greenness and productivity
LS_distr_mag_2012_2017	Magnitude of LandTrendr observed shock to NDVI	Sudden positive or negative shock (drought, land use change)
LS_distr_dur_2012_2017	Duration of observed LandTrendr shock to NDVI	Indicator of severity/duration of shock

Table 6 *Description of low temporal resolution time-series features*

Variable Importance

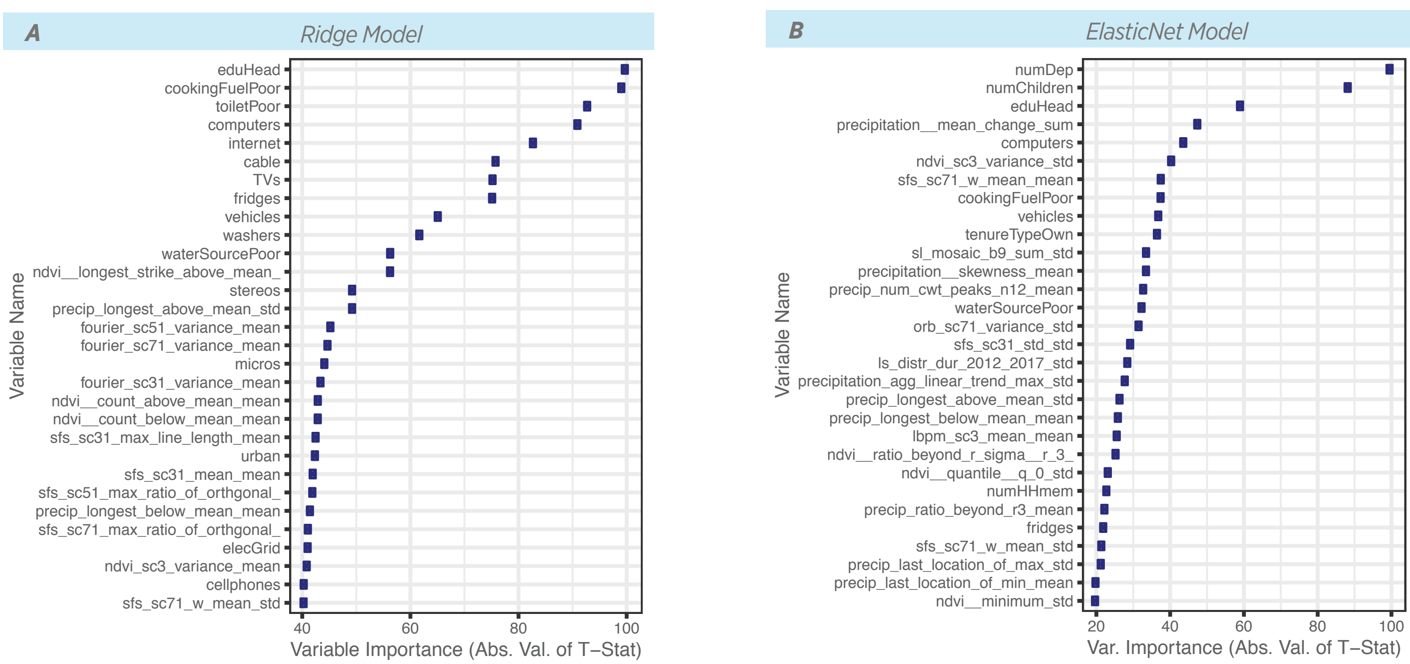


Figure 6 A-B *Variable importance machine learning models of household income, Ridge and Elastic Net models These plots show the top 30 most important variables for predicting household income. Each variable's importance metric is scaled by the top variable, which is given an importance metric of 100, and other variables scores are relative to that variable*

IDB
Inter-American Development Bank

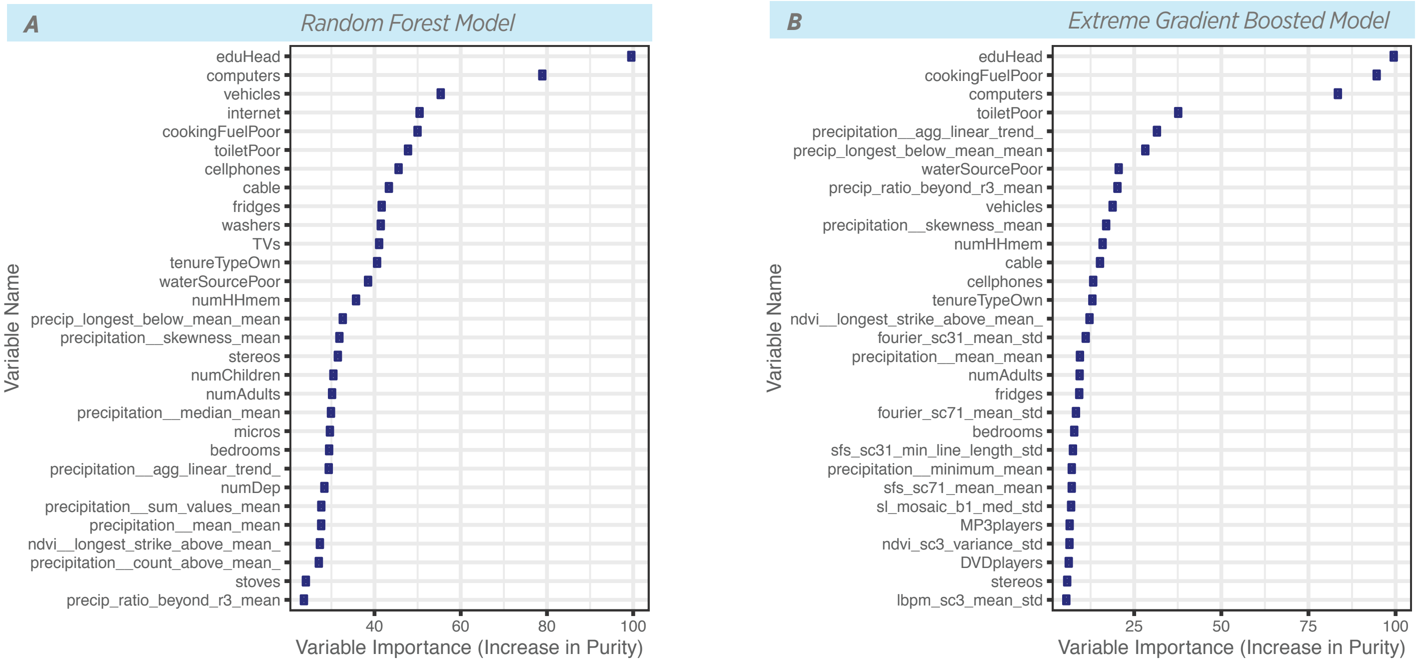
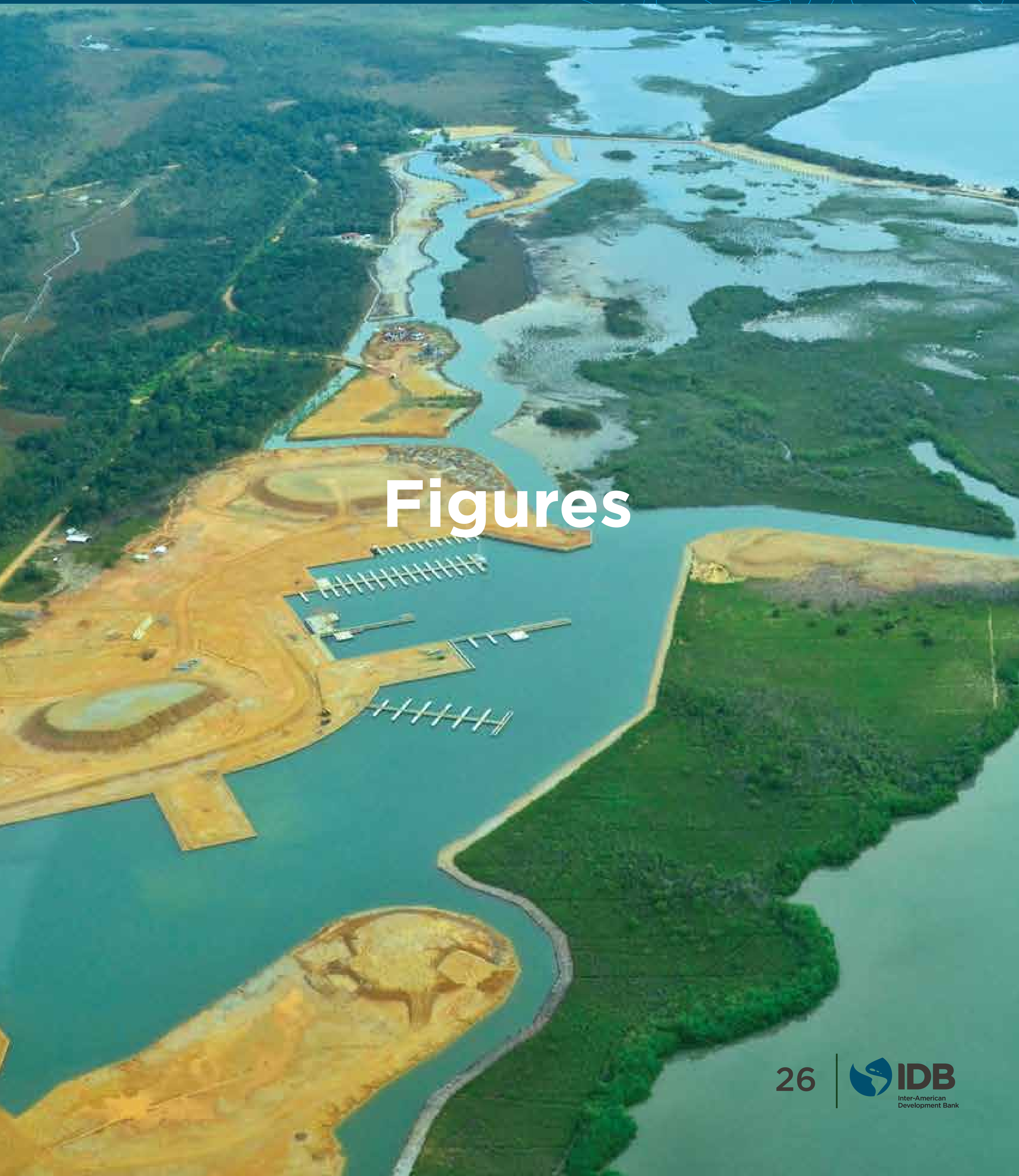
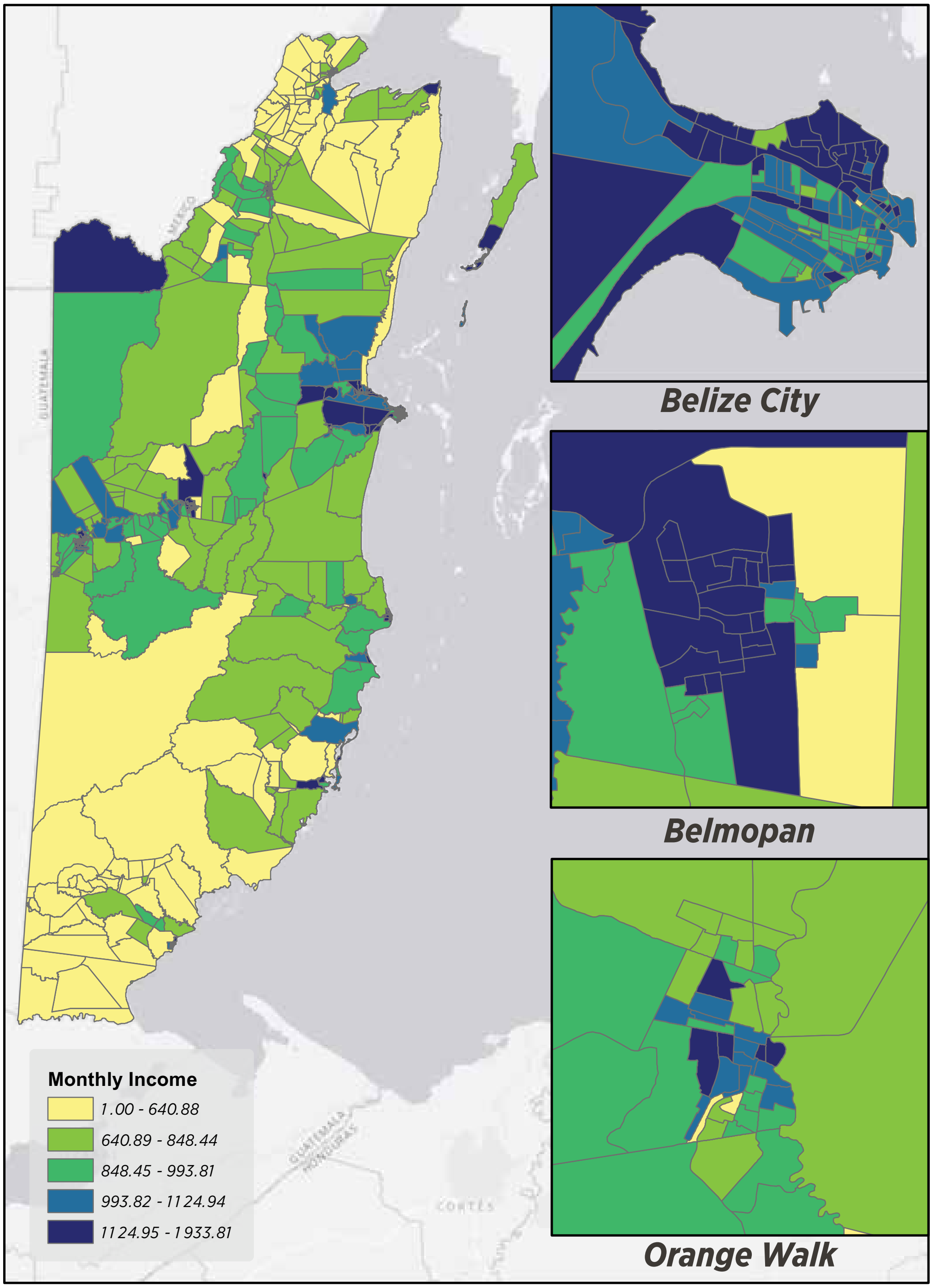


Figure 7 A-B *Variable importance machine learning models of household income, Random Forest and Extreme Gradient Boosted models These plots show the top 30 most important variables for predicting household income for Random Forest and XGBoost. Each variable's importance metric is scaled by the top variable, which is given an importance metric of 100, and other variables scores are relative to that variable*

Figures

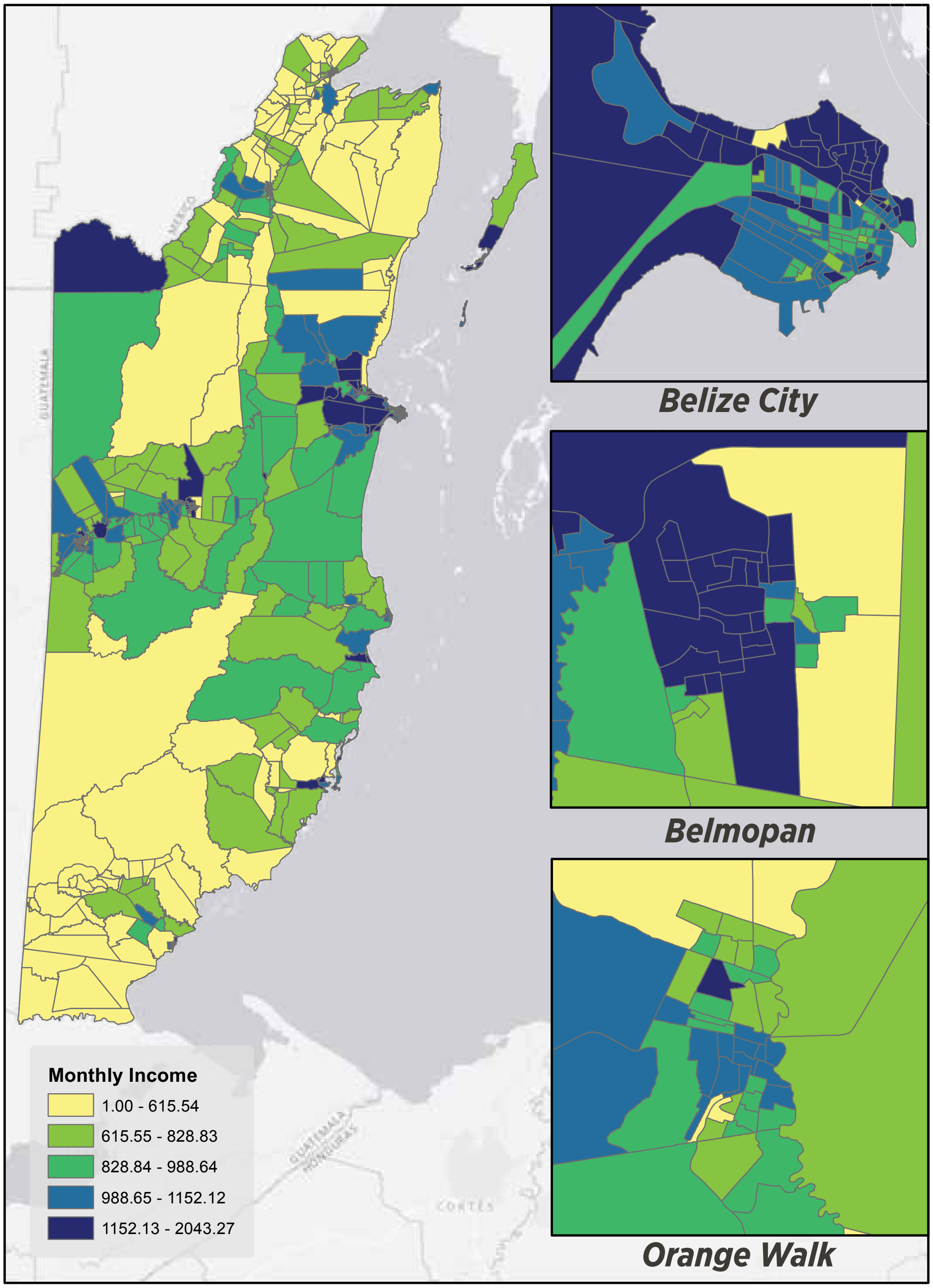


Combined Average Monthly Income Estimates for Belize



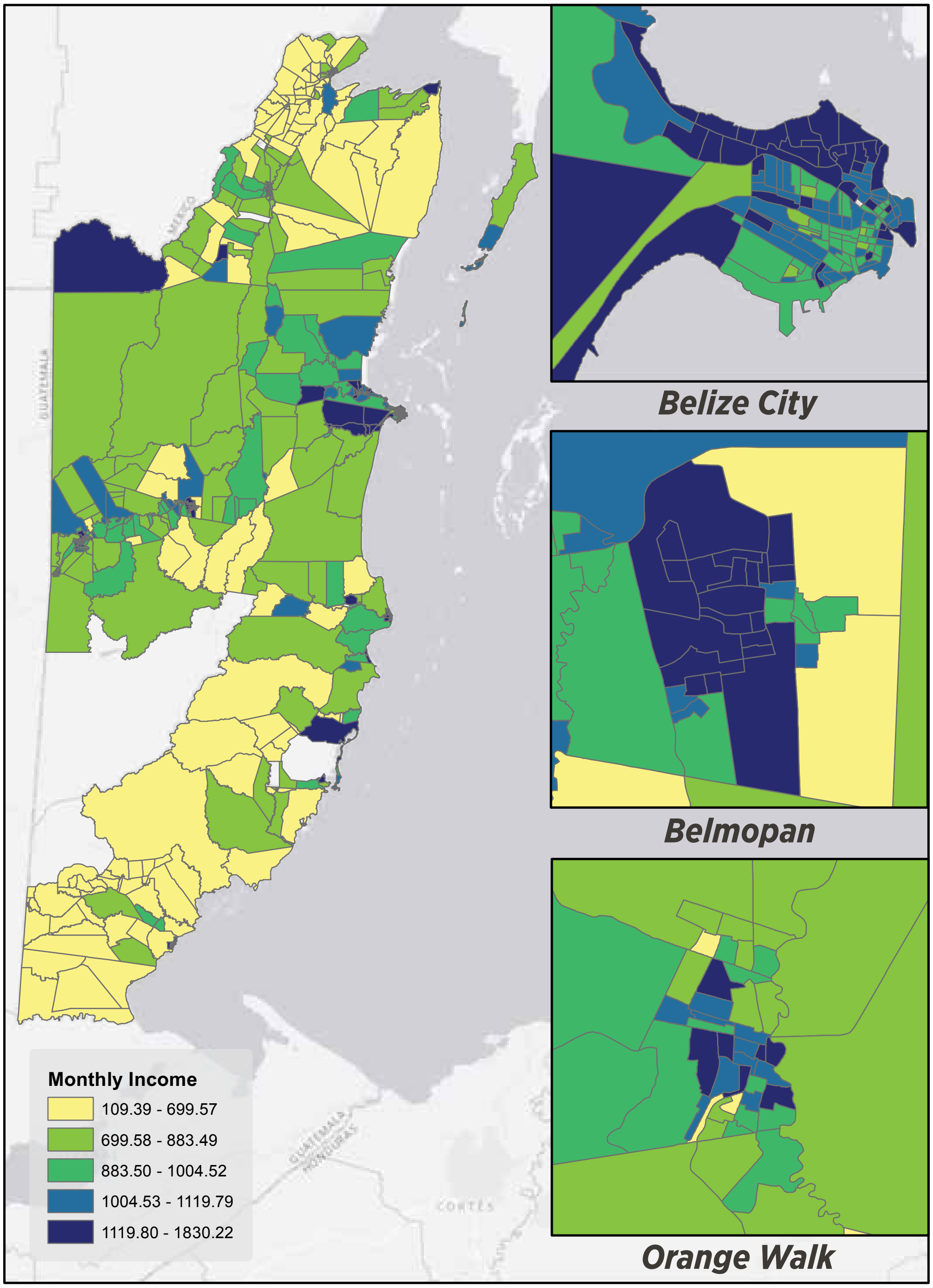
Source: "Mapping Poverty in Belize Using Satellite Features and Machine Learning"
Jonathan Hersh, Ryan Engstrom, Michael Mann, Alejandra Mejia, and Lucia Martin Rivero.
Inter-American Development Bank, 2020

XG Boost Average Monthly Income Estimates for Belize



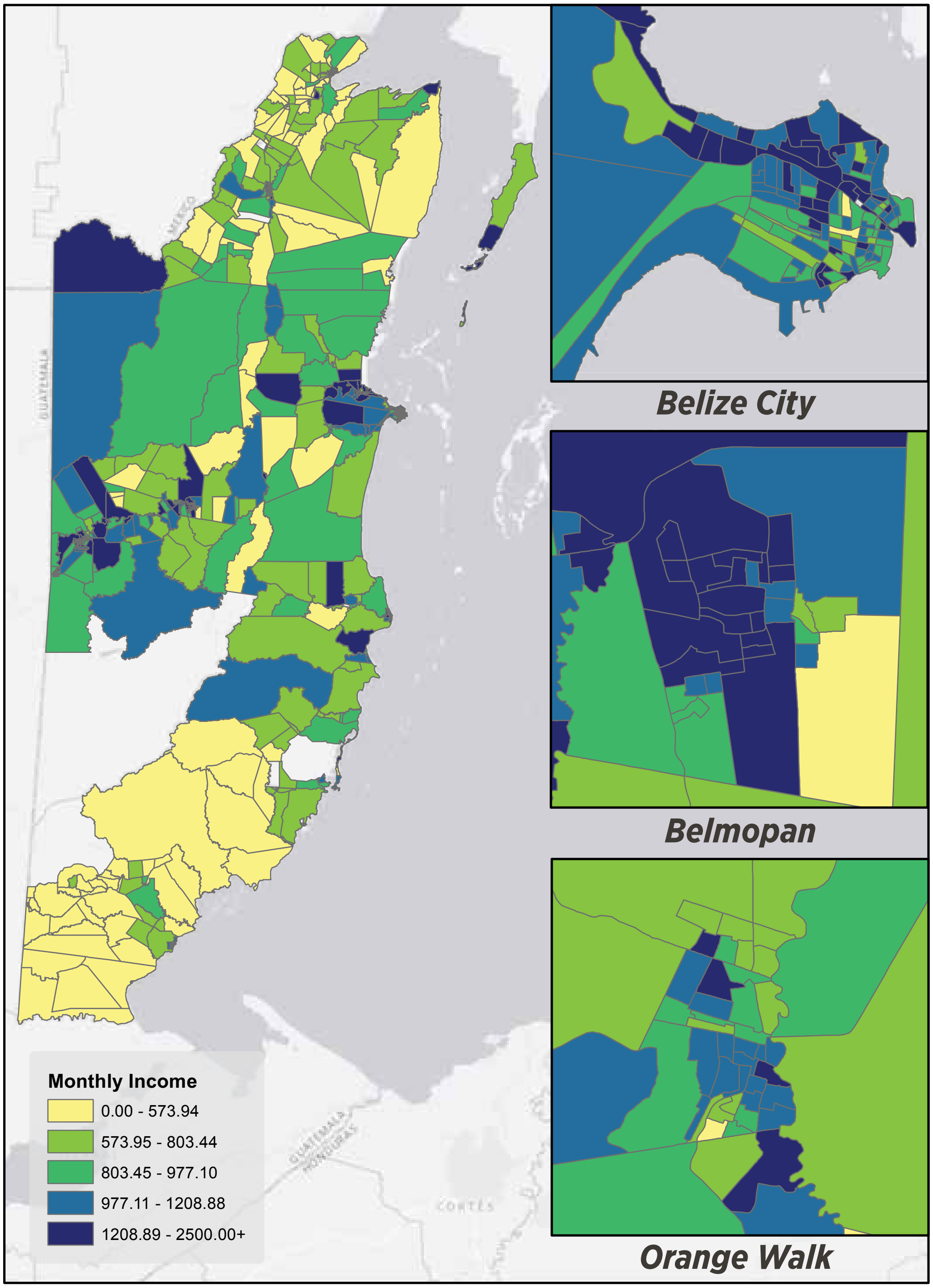
Source: "Mapping Poverty in Belize Using Satellite Features and Machine Learning"
Jonathan Hersh, Ryan Engstrom, Michael Mann, Alejandra Mejia, and Lucia Martin Rivero.
Inter-American Development Bank, 2020

Random Forest Average Monthly Income Estimates for Belize



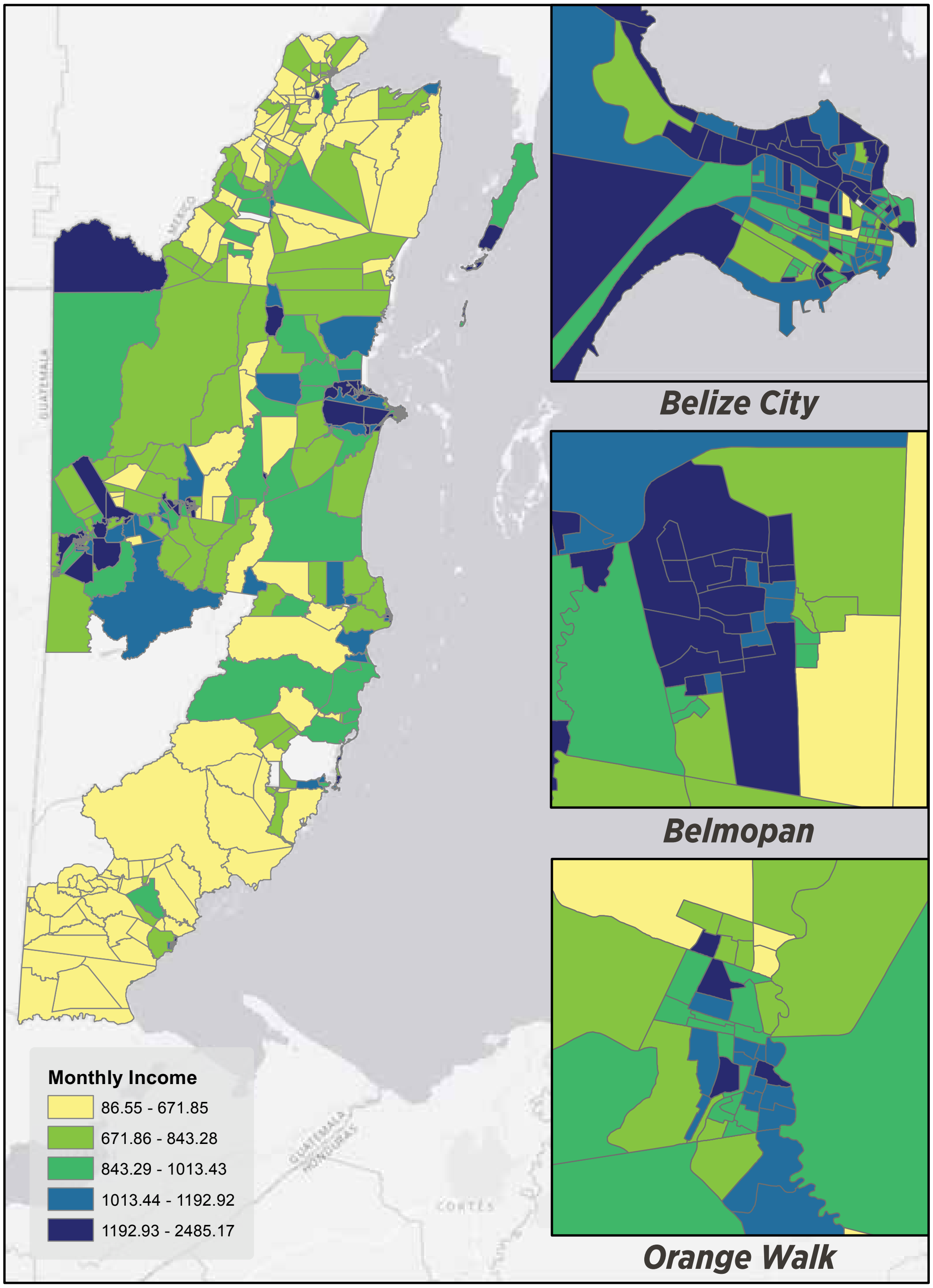
Source: "Mapping Poverty in Belize Using Satellite Features and Machine Learning"
Jonathan Hersh, Ryan Engstrom, Michael Mann, Alejandra Mejia, and Lucia Martin Rivero.
Inter-American Development Bank, 2020

Ridge Average Monthly Income Estimates for Belize



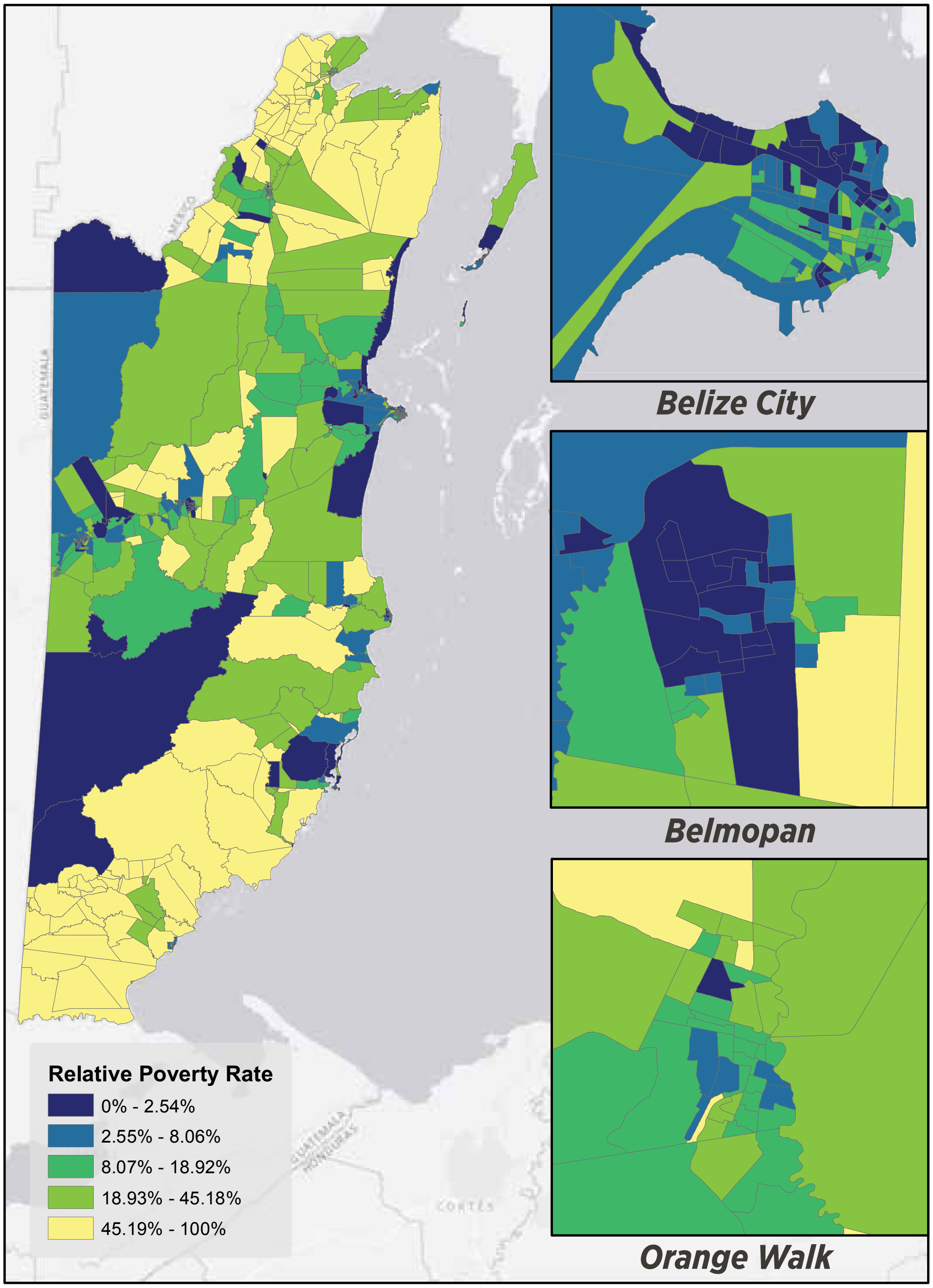
Source: "Mapping Poverty in Belize Using Satellite Features and Machine Learning"
Jonathan Hersh, Ryan Engstrom, Michael Mann, Alejandra Mejia, and Lucia Martin Rivero.
Inter-American Development Bank, 2020

Ridge Average Monthly Income Estimates for Belize



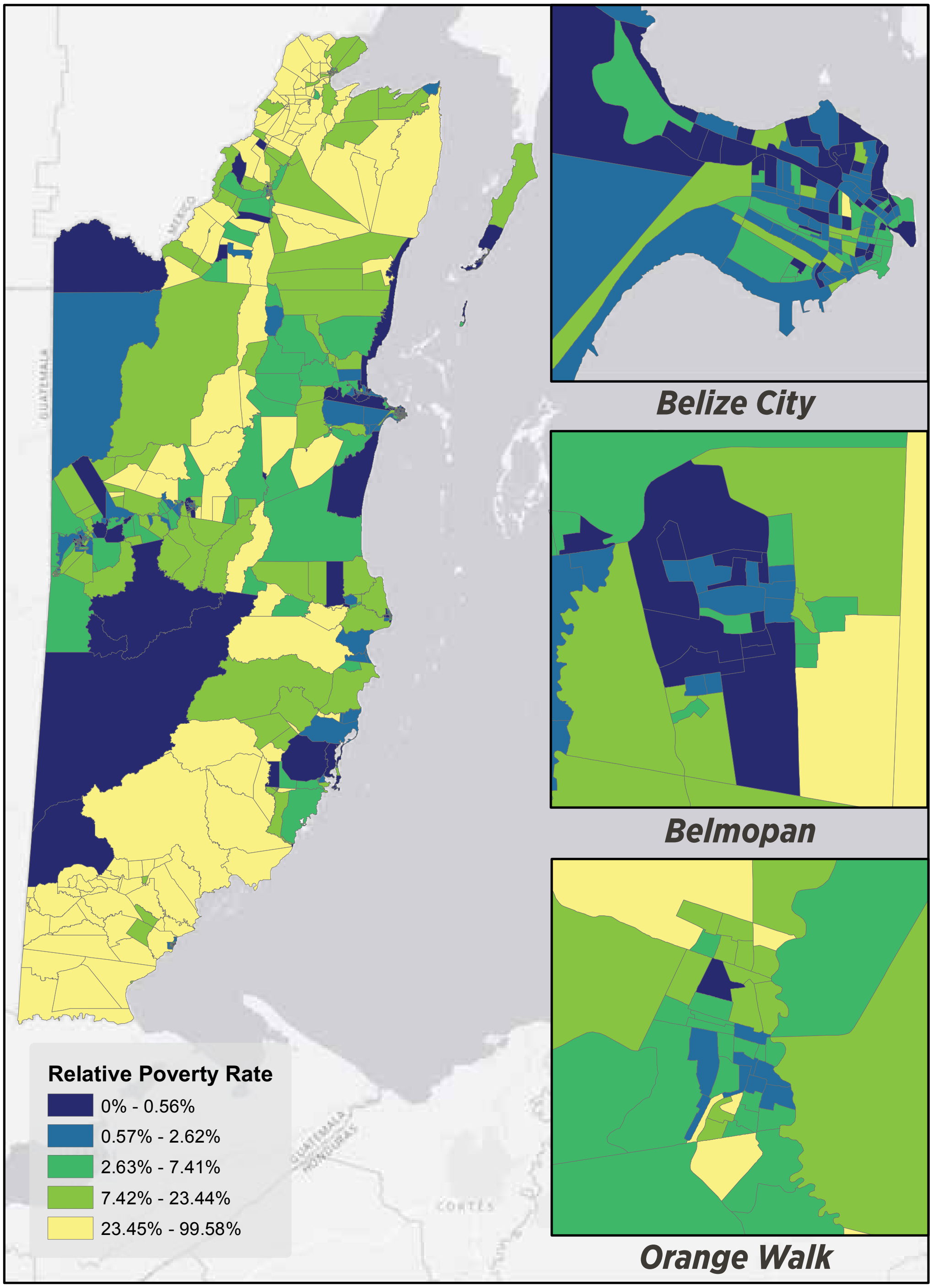
Source: "Mapping Poverty in Belize Using Satellite Features and Machine Learning"
Jonathan Hersh, Ryan Engstrom, Michael Mann, Alejandra Mejia, and Lucia Martin Rivero.
Inter-American Development Bank, 2020

Household with less than 20th Percentile of National Income



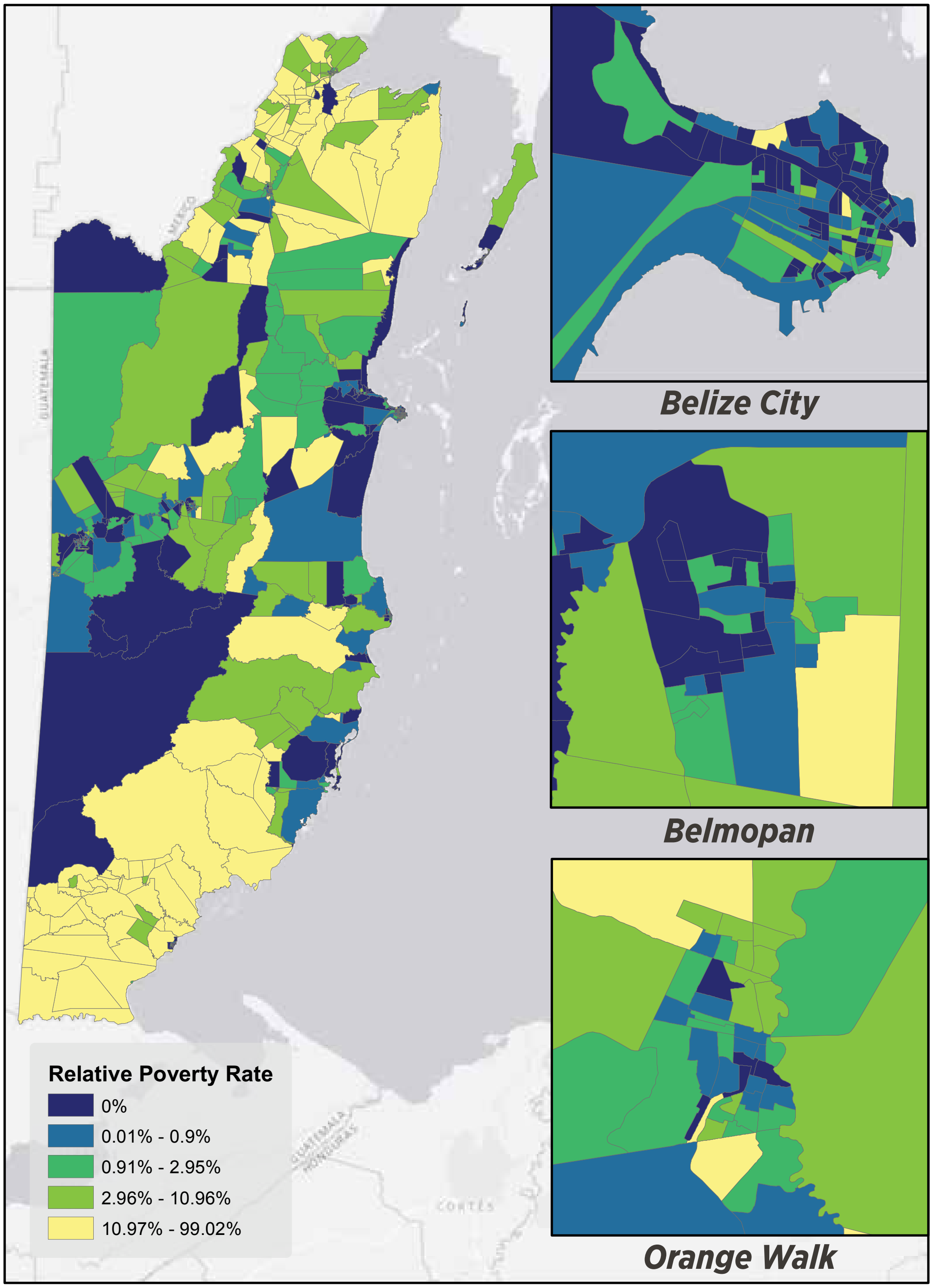
Source: "Mapping Poverty in Belize Using Satellite Features and Machine Learning"
Jonathan Hersh, Ryan Engstrom, Michael Mann, Alejandra Mejia, and Lucia Martin Rivero.
Inter-American Development Bank, 2020

Household with less than 15th Percentile of National Income



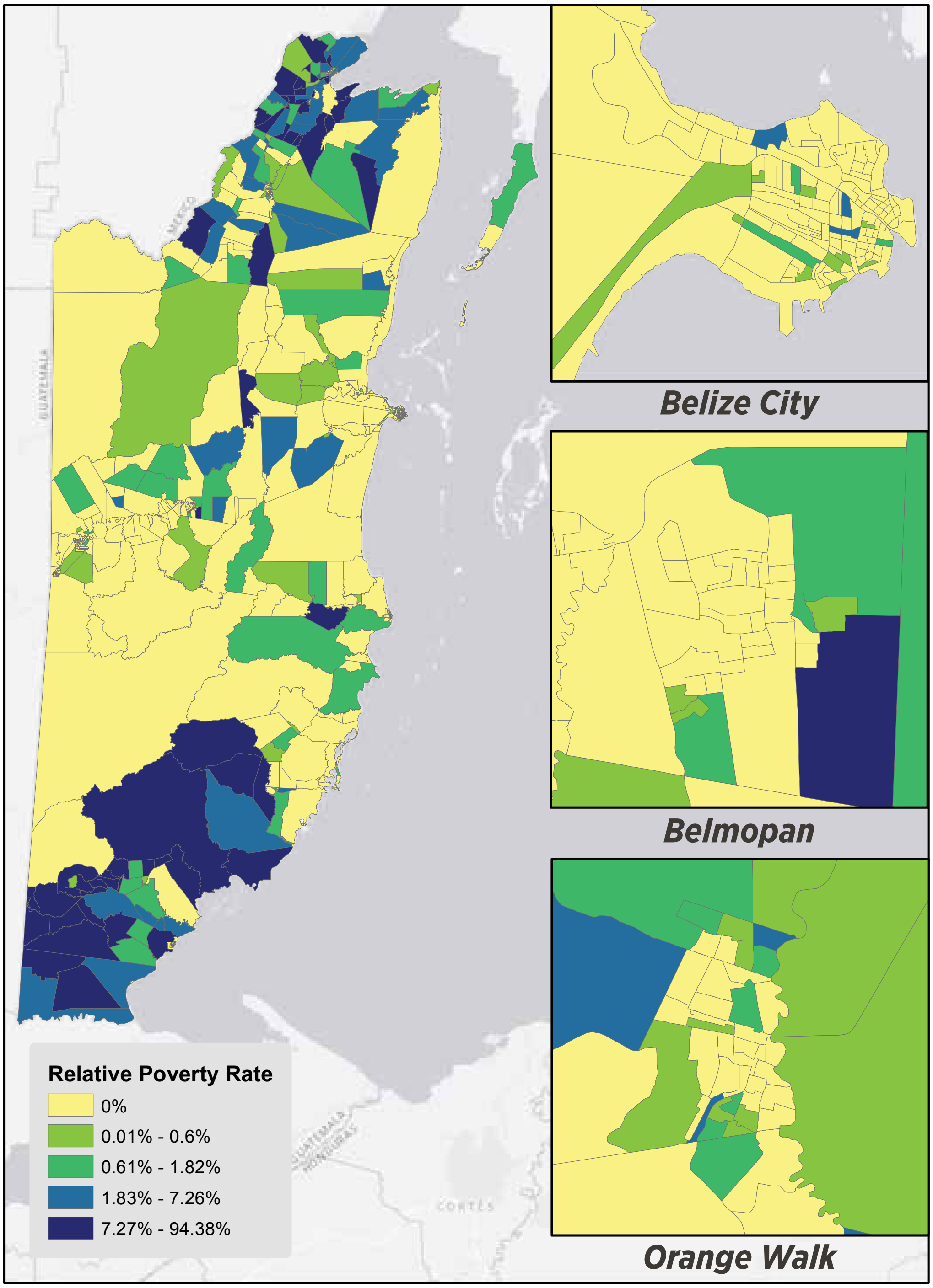
Source: "Mapping Poverty in Belize Using Satellite Features and Machine Learning"
Jonathan Hersh, Ryan Engstrom, Michael Mann, Alejandra Mejia, and Lucia Martin Rivero.
Inter-American Development Bank, 2020

Household with less than 10th Percentile of National Income



Source: "Mapping Poverty in Belize Using Satellite Features and Machine Learning"
Jonathan Hersh, Ryan Engstrom, Michael Mann, Alejandra Mejia, and Lucia Martin Rivero.
Inter-American Development Bank, 2020

Household with less than 5th Percentile of National Income



Source: "Mapping Poverty in Belize Using Satellite Features and Machine Learning"
Jonathan Hersh, Ryan Engstrom, Michael Mann, Alejandra Mejia, and Lucia Martin Rivero.
Inter-American Development Bank, 2020

References

Afzal, M., Hersh, J. and Newhouse, D., 2015. Building a Better Model: Variable Selection to Predict Poverty in Pakistan and Sri Lanka.

Arthur, W.B., 1989. Competing technologies, increasing returns, and lock-in by historical events. The economic journal, 99(394), pp.116-131.

Babenko, Boris, Jonathan Hersh, David Newhouse, Anusha Ramakrishnan, and Tom Swartz.

2017. "Poverty Mapping Using Convolutional Neural Networks Trained on High and Medium Resolution Satellite Images, with an Application in Mexico." ArXiv Preprint ArXiv:1711.06323.

Bah, A., Bazzi, S., Sumarto, S., & Tobias, J. (2018). Finding the poor vs. measuring their poverty: exploring the drivers of targeting effectiveness in Indonesia. The World Bank.

Bazzi, S., Sumarto, S. and Suryahadi, A., 2015. It's all in the timing: Cash transfers and consumption smoothing in a developing country. Journal of Economic Behavior & Organization, 119, pp.267-288.

Blumenstock, J., Cadamuro, G. and On, R., 2015. Predicting poverty and wealth from mobile phone metadata. Science, 350(6264), pp.1073-1076.

Breiman, L., 2001. Random forests. Machine learning, 45(1), pp.5-32.

Chatfield, C., 1995. Model uncertainty, data mining and statistical inference. Journal of the Royal Statistical Society: Series A (Statistics in Society), 158(3), pp.419-444.

Cohen, Warren B, Zhiqiang Yang, Sean P Healey, Robert E Kennedy, and Noel Gorelick. 2018. "A LandTrendr Multispectral Ensemble for Forest Disturbance Detection." Remote Sensing of Environment 205: 131–40.

Dalal, Navneet, and Bill Triggs. 2005. "Histograms of Oriented Gradients for Human Detection." In .

Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. "Imagenet: A Large-Scale Hierarchical Image Database." In 2009 IEEE Conference on Computer Vision and Pattern Recognition, 248–55.

Elbers, C., Lanjouw, J.O. and Lanjouw, P., 2003. Micro–level estimation of poverty and inequality. Econometrica, 71(1), pp.355-364.

Engstrom, R., Hersh, J. and Newhouse, D., 2017. Poverty from space: Using high-resolution satellite imagery for estimating economic well-being.

Engstrom, Ryan, Dan Pavelsku, Tanaka Tomomi, and Ayago Wambile. n.d. "Mapping Poverty and Slums Using Multiple Methodologies in Accra , Ghana." In Joint Urban Remote Sensing Conference 2019, 1–4.

Foster, J., Greer, J. and Thorbecke, E., 1984. A class of decomposable poverty measures. Econometrica: journal of the econometric society, pp.761-766.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Annals of statistics, pp.1189-1232.

Funk, Chris C, Pete J Peterson, Martin F Landsfeld, Diego H Pedreros, James P Verdin, James D Rowland, Bo E Romero, Gregory J Husak, Joel C Michaelsen, and Andrew P Verdin. 2014. "A Quasi-Global Precipitation Time Series for Drought Monitoring." US Geological Survey Data Series 832 (4): 1–12.

Grosh, M.E. and Baker, J.L., 1995. Proxy means tests for targeting social programs: simulations and speculation. The World Bank.

James, G., Witten, D., Hastie, T. and Tibshirani, R., 2013. An introduction to statistical learning (Vol. 112, p. 18). New York: springer.

Hoerl, A.E. and Kennard, R.W., 1970. "Ridge regression: Biased estimation for nonorthogonal problems." Technometrics, 12(1), pp.55-67.

Huang, Xin, Liangpei Zhang, and Pingxiang Li. 2007. "Classification and Extraction of Spatial Features in Urban Areas Using High-Resolution Multispectral Imagery." IEEE Geoscience and Remote Sensing Letters 4 (2): 260–64.

Jean, Neal, Marshall Burke, Michael Xie, W Matthew Davis, David B Lobell, and Stefano Ermon. 2016. "Machine Learning To Predict Poverty." Science 353 (6301): 790–94.

Kennedy, Robert E, Zhiqiang Yang, and Warren B Cohen. 2010. "Detecting Trends in Forest Disturbance and Recovery Using Yearly Landsat Time Series: 1. LandTrendr—Temporal Segmentation Algorithms." Remote Sensing of Environment 114 (12): 2897–2910.

Kilic, T., Serajuddin, U., Uematsu, H. and Yoshida, N., 2017. Costing household surveys for monitoring progress toward ending extreme poverty and boosting shared prosperity. The World Bank.

Kirscht, Martin. 1998. "Detection, Velocity Estimation and Imaging of Moving Targets with Single-Channel SAR." In EUSAR'98- European Conference on Synthetic Aperture Radar, Friedrichshafen, Germany, 587–90.

Kirscht, Martin, and Carsten Rinke. 1998. "3D Reconstruction of Buildings and Vegetation from Synthetic Aperture Radar (SAR) Images." In MVA, 228–31.

LeCun, Y. and Bengio, Y., 1995. Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks, 3361(10), p.1995.

Levy, S. 2007. Progress against poverty: sustaining Mexico's Progresa-Oportunidades program. Brookings Institution Press.

Mann, Michael L, and James M Warner. 2017. "Ethiopian Wheat Yield and Yield Gap Estimation: A Spatially Explicit Small Area Integrated Data Approach." Field Crops Research 201: 60–74.

Mann, Michael L, James M Warner, and Arun S Malik. 2019. "Predicting High-Magnitude, Low-Frequency Crop Losses Using Machine Learning: An Application to Cereal Crops in Ethiopia." Climatic Change 154 (1–2): 211–27.

McNairn, H, J Shang, X Jiao, and C Champagne. 2009. "The Contribution of ALOS PALSAR Multipolarization and Polarimetric Data to Crop Classification." IEEE Transactions on Geoscience and Remote Sensing 47 (12): 3981–92. https://doi.org/10.1109/TGRS.2009.2026052.

Mehrotra, Rajiv, Kameswara Rao Namuduri, and Nagarajan Ranganathan. 1992. "Gabor Filter-Based Edge Detection." Pattern Recognition 25 (12): 1479–94.

Myint, Soe W., Victor Mesev, and Nina Lam. 2006. "Urban Textural Analysis from Remote Sensor Data: Lacunarity Measurements Based on the Differential Box Counting Method." Geographical Analysis 38 (4): 371–90. https://doi.org/10.1111/j.1538-4632.2006.00691.x.

National Human Development Advisory Committee, Ministry of Economic Development, Commerce and Industry, and Consumer Protection. 2010. "2009 Country Poverty Assessment". Available at *http://www.ncabz.org/wp-content/uploads/2015/05/Country-Povert-Assessment.pdf*

Ojala, Timo, Matti Pietikäinen, and Topi Mäenpää. 2002. “Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns.” IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (7): 971–87. *https://doi.org/10.1109/T-PAMI.2002.1017623.*

Pape, U.J. and Mistiaen, J.A., 2018. Household expenditure and poverty measures in 60 minutes: a new approach with results from Mogadishu. World Bank Policy Research Working Paper, (8430).

Pesaresi, Martino, Andrea Gerhardinger, and François Kayitakire. 2008. “A Robust Built-up Area Presence Index by Anisotropic Rotation-Invariant Textural Measure.” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 1 (3): 180–92.

Rodrik, D., Hausmann, R., & Velasco, A. (2004). Growth diagnostics. Part I The Washington Consensus: From Its Origins to Its Critics, 356.

Rublee, Ethan, Vincent Rabaud, Kurt Konolige, and Gary R Bradski. 2011. “ORB: An Efficient Alternative to SIFT or SURF.” In ICCV, 11:2.

Serajuddin, U., Uematsu, H., Wieser, C., Yoshida, N. and Dabalen, A., 2015. Data deprivation: another deprivation to end. The World Bank.

Shan, Jie, Ejaz Hussain, KyoHyouk Kim, and Larry Biehl. 2010. “Flood Mapping with Satellite Images and Its Web Service.” Photogrammetric Engineering & Remote Sensing 76 (2): 102–4.

Shimada, Masanobu, Takuya Itoh, Takeshi Motooka, Manabu Watanabe, Tomohiro Shiraishi, Rajesh Thapa, and Richard Lucas. 2014. “New Global Forest/Non-Forest Maps from ALOS PALSAR Data (2007–2010).” Remote Sensing of Environment 155: 13–31. https://doi.org/https://doi.org/10.1016/j.rse.2014.04.014.

Tucker, C J. 1979. “The NDVI Has Been Proven to Be Well Correlated with Various Vegetation Parameters, Such as Green Biomass.” Rouse et Al, 605–790.

Tucker, Compton J., Ryan Engstrom, Jonathan Hersh, David Newhouse, Soe W. Myint, Victor Mesev, Nina Lam, et al. 2017. “Spatial Panel Data Models.” Edited by Alluvial Fan Task Force. Remote Sensing of Environment, Policy Research Working Paper, 10 (3): 3373–79. https://doi.org/10.3390/rs10050691.

Ünsalan, Cem, and Kim L Boyer. 2005. “A System to Detect Houses and Residential Street Networks in Multispectral Satellite Images.” Computer Vision and Image Understanding 98 (3): 423–61.

Yoshida, Nobuo; Munoz, Ricardo; Skinner, Alexander; Lee, Catherine Kyung-Eun; Brataj, Mario; Durbin, Spencer William; Sharma, D..2015. Survey of Well-Being via Instant and Frequent Tracking (SWIFT) Data Collection Guidelines (English). Washington, D.C. : World Bank Group. *http://documents.worldbank.org/curated/en/591711545170814297* Survey-of-Well-Being-via-Instant-and-Frequent-Tracking-SWIFT-Data-Collection-Guidelines

Zou, H. and Hastie, T., 2005. "Regularization and variable selection via the elastic net." Journal of the royal statistical society: series B (statistical methodology), 67(2), pp.301-320.


IDB
Inter-American
Development Bank