WPS7329

Policy Research Working Paper 7329

Estimating Local Poverty Measures Using Satellite Images

A Pilot Application to Central America

Ben Klemens
Andrea Coppola
Max Shron

World Bank Group
Macroeconomics and Fiscal Management Global Practice Group
June 2015

Policy Research Working Paper 7329

Abstract

Several studies have used satellite measures of human activity to complement measures of economic production. This paper builds on those studies by considering satellite measures for improving poverty measures. The paper uses local-scale census and survey data from Guatemala to test at how fine a scale satellite measures are useful. Results show that supplementing survey data with satellite data leads to improvements in the estimates.

This paper is a product of the Macroeconomics and Fiscal Management Global Practice Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org. The authors may be contacted at acoppola@worldbank.org.

Estimating local poverty measures using satellite images: a pilot application to Central America

Ben Klemens[*], Andrea Coppola[†], and Max Shron[‡]

JEL classifications: I32, C55, C8.
Keywords: Geographic Information Systems, ICT Statistical Capacity Building, Poverty Measurement and Analysis

[*]Polynumeral. The authors are extremely grateful to Maria Eugenia Genoni and Leonardo Lucchetti who provided the poverty data used for the analysis and contributed with guidance, suggestions, and comments during the preparation of this paper. The authors would like also to thank Sandra Martinez and Mateo Salazar Rodriguez for valuable research assistance provided.

[†]World Bank
[‡]Polynumeral

1 Introduction

"The accuracy of global poverty numbers depends on the availability of household surveys," writes Chandy [2013, p 13], and "this remains one of the biggest constraints to poverty data today." Two-fifths of countries fail to conduct a household survey every five years [Chandy, 2013, p 14]. Even when household surveys are conducted, the poor tend to be hard-to-reach by survey or census takers [American Statistical Association, 2012], and data quality will be correspondingly lower. Sala-i-Martin and Pinkovskiy [2010] present evidence that household surveys are systematically biased and unreliable.

Conversely, satellites gather data at a constant rate throughout the year, regardless of physical or social hazards, and there are known means of cleaning the data to remove well-known inconsistencies. NASA's Moderate-Resolution Imaging Spectroradiometer (MODIS) project generates image data that is aggregated and cleaned for public use every eight days.

The bulk of the paper will concern itself with three exercises intended to explore the satellite data in the context of survey data. The first describes the quality of correlations between the survey and satellite data, the second is a panel of simple regressions including and excluding satellite measures, and the third a set of regressions that act as a stress test, asking whether there can exist survey-based models that obviate the need for satellite data. We chose to limit our modeling efforts to simple regressions, and reserve the problem of folding satellite data into a formal small-area model (such as that of Bauder et al. [2014], Elbers et al. [2003], or Molina and Rao [2010]) for future work.

For most of this paper, the scale of the data is roughly the level of the 338 municipalities in Guatemala, which is a smaller scale than the previous studies. Some municipalities are as small as a square kilometer, and they follow geographic and social boundaries rather than the neat grid that the satellite data follows. Further, the Guatemalan Instituto Nacional de Estadística (INE) provides data on urban and rural poverty separately, so we were able to do some exercises dividing even the municipalities into urban and rural.

So far, we have found that there are reasonable situations where satellite imaging data does add information at this scale. For ordinary least squares regressions with poverty measures as a dependent variable, the corrected Akaike information criterion (AIC_c) improves when luminosity measures are included. However, we also show that it is possible to construct regressions where satellite data does not add information beyond that provided by the survey data.

Adding satellite data to survey data is therefore not a guarantee of improved results. But, especially when compared with fielding a survey, it is virtually costless to try the experiment of augmenting a survey with satellite data.

In our regressions regarding rural poverty, luminosity data was significant; in our regressions regarding urban poverty, luminosity data was not.

We also explored satellite measures of leaf coverage and verdancy, but these showed less correlation to poverty measures than luminosity.

Section 1.1 describes some of the studies of luminosity and economic factors to date. Section 2 gives a basic overview of the data used in this paper, including correlograms displaying how well various measures correlate. Section 3 presents the basic contours of the data and correlations among the variables. Section 4 presents a panel of regressions with and without satellite measures. Section 5 presents an exercise using stepwise regressions to produce models using survey data, designed to seek a bound to how much information can be gained from survey data without luminosity data.

The use of luminosity data is not yet commonplace in the community of economists, so an appendix goes into some detail on the data-processing pipeline we used to merge map data with

survey data. We used only freely-available tools for the processing, some of which we wrote and made available at `http://github.com/polynumeral/satellite-pilot`.

1.1 Literature

Elvidge et al. [1997] and Elvidge et al. [1999] described technical methods for obtaining measures of cloud-free, stable nighttime illumination using satellite sensors. They did so “for the analysis of social, environmental, and energy issues” and “to detect the expansion of urban areas” [p 734].

Elvidge et al. [2012] used the imaging data to develop a Night Light Development Index, and found it to have good correlation with the Human Development Index from the United Nations Development Programme. Chen and Nordhaus [2011] and Henderson et al. [2009] wrote on using luminosity as a measure of GDP on a national level, and demonstrated a strong correlation between nighttime illumination and national GDP. Pinkovskiy and Sala-i Martin [2014] primarily consider the question of whether there is correlation between errors in GDP measures and errors in luminosity; under certain linearity assumptions, they find little evidence of any such correlation.

The studies above are on the national level. Sutton et al. [2007] develop a simple model of GDP at the subnational level for India, China, Turkey, and the United States.

Productivity measures are not necessarily the inverse of poverty measures, as people do not necessarily work and live in the same area, and the distribution of wealth may differ from area to area. Ebener et al. [2005] correlate subnational poverty data to luminosity, and find a good correlation between poverty and luminosity at lower-than-country levels.

Survey data is currently the most relied-upon method of estimating poverty, so we compare the efficacy of predicting poverty using satellite sensor data with prediction using survey and census responses. Bhattacharya and Innes [2006] used basic health survey data, and the other papers above did not include survey questions in their models; they also use satellite leaf coverage data, while all of the papers above used only luminosity data.

Our paper extends the existing literature by moving to a still smaller scale than previously done, and considering the luminosity data in the context of existing survey data. The Guatemalan Instituto Nacional de Estadística divides population, poverty counts, and locations between urban and rural, allowing us to more directly evaluate the efficacy of satellite data in urban and rural contexts. To the best of our knowledge, this is the first paper to look at the relationship between luminosity and poverty in the urban and rural contexts separately.

2 Methods and data

This section describes the raw data sets, their basic distributional characteristics, and some considerations in adapting them to be compatible.

We analyzed the relationship between existing small-scale poverty estimation (based on census and household survey information), nighttime illumination (from the U.S. National Oceanographic and Atmospheric Administration, NOAA), leaf coverage (from the U.S. National Aeronautics and Space Administration, NASA) and albedo (a measure of reflectivity, also from NASA).

2.1 Satellite data

We begin with the characteristics of luminosity and leaf coverage data in this region. We touch on the methods used to match satellite observations to geographies, but technical details on geoTIFFs

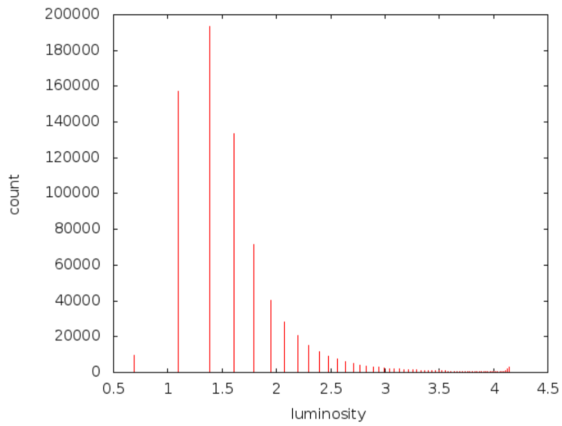

Figure 1: Histogram of nonzero log-luminosities for Guatemala.

and the toolchain we used to extract a list of data points amenable to use in typical statistics packages are presented in an appendix.

2.1.1 NOAA

For nighttime luminosity measurements, we used the stable lights geoTIFFs from NOAA's National Geophysical Data Center (NGDC). NOAA provides two "stable lights" composites for each year. The composites are an average for a given year after correcting for cloudy days, gas flares, forest fires, and fishing boats, via procedures derived from those described in Elvidge et al. [1997] and Elvidge et al. [1999].

Although NOAA corrects many common errors, it does not guarantee calibration of overall levels across years. Absolute year-to-year differences are unreliable. A variable measured with a constant additive or multiplicative error will bias downward the p-values on the coefficients in the linear regressions to follow [Greene, 1990].

Distribution of values The luminosity measure at any given pixel ranges from zero to 63. There are no observations in the data set with luminosity one.

In Guatemala, 25.90% of observed points are nonzero. Figure 1 shows the long-tailed distribution of log nonzero pixel values. There is a small uptick at the maximum value of 63, because the limits of the sensor group all areas with a brightness above a certain level at exactly 63. Top-coding affects relatively few pixels in the data set, so we did not make corrections for it.

Figure 2 shows the pattern of lights in 2001, 2008, and the change between them.

2.1.2 MODIS

NASA's Moderate-Resolution Imaging Spectroradiometer measures the reflections of light from the Sun in several wavelength bands. Albedo measures the reflectivity of the surface in a given band, and varies from 0% (full absorption) to 100% (full reflection). We use the .659μm (red) band. Areas with cloud cover are reported as N/A in the data.

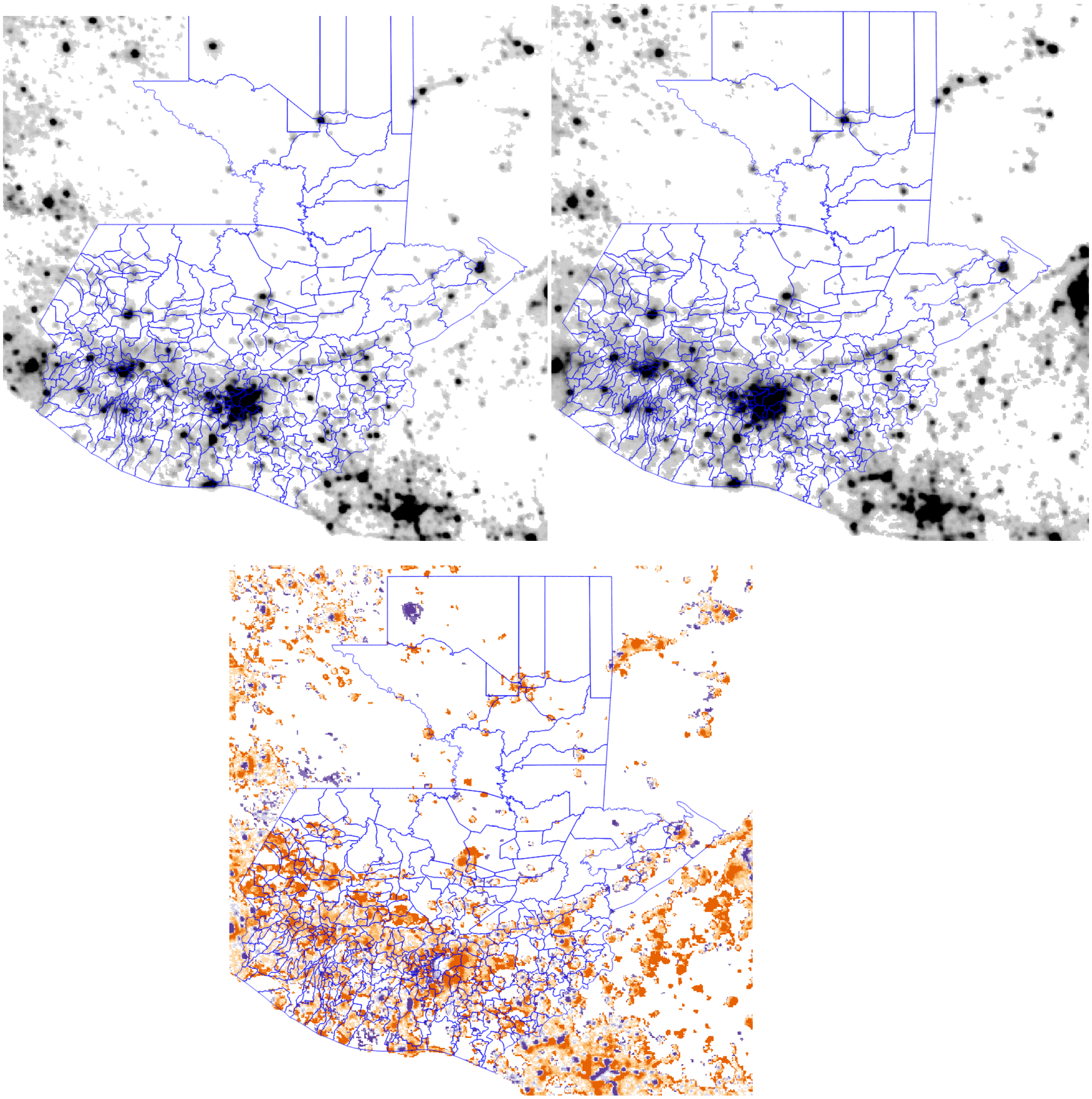

Figure 2: Left: 2001, black indicates nighttime lights. Right: 2008. Bottom: difference, where orange indicates new or more intense lights; purple indicates less intense light. The NOAA satellite is not calibrated across years, so the overall patterns are deemed accurate but the magnitude of changes may be consistently high or low. Overlaid are the municipalities of Guatemala.

Photosynthetic plants absorb a well-known set of wavelengths, so low intensity of light in those bands can be used as an indicator of plant density. We use leaf area index (LAI) and fraction of photosynthetically active radiation (FPAR). The MODIS data products handbook explains that "LAI defines canopy leaf area, while FPAR defines the amount of incoming solar radiation absorbed by the plant canopies."[1]

Summary files are available every eight days. We averaged three files per year (spaced four months apart). Figure 3 shows LAI in the region for 2001, 2011, and the change.

2.2 Census and poverty estimate data

This section describes the census and survey data used to measure poverty and population characteristics.

2.2.1 Guatemalan Census data

Guatemala's Instituto Geográfico Nacional (IGN) reports basic population information for the municipailites of Guatemala. Locations at a sub-municipality level are classified into urban and rural areas. This allows the unit of analysis for the regression to be the urban or rural portion of each municipality; we refer to these geographic areas as the *observed areas* for Guatemala.

The outcome variables of interest are the Foster Greer Thorbecke measures of poverty:

$$FGT_\alpha = \frac{1}{N}\sum_{i=1}^{H}(\frac{z-y_i}{z})^\alpha,$$

where α is a scaling factor (typically 0, 1, or 2), z is the poverty line, y_i is the individual income level, and H the count of people with incomes below the poverty line.

2.3 Processing

Our method in merging the satellite and terrestrial data was to use the NOAA data to divide the countries into squares, with one datum at the center of each square. Each datum from NASA was assigned to the single square it fell into.

The bounds of the municipalities are specified as polygons (sequences of line segments forming the area boundary) in shapefiles, freely available online.[2] We assigned every point in the grid defined by the NOAA data to a polygon. Thus, there was a unique mapping from each satellite datum to an observed area.

For Guatemala, we had the opportunity to go even below the municipality level. Guatemala's IGN provides a list, generated in 2002, of 27,352 named places, each with a point—not an area—marking its location, including 25,062 rural and 2,290 urban locations. For each square defined by the NOAA grid with Census-designated places inside, we counted how many were urban and rural, and specified the square as urban or rural based on a majority rule, with the 252 square kilometers that were a 50-50 split marked as urban. For each square defined by the NOAA grid that did not have a Census-designated place inside, we found the closest place and used its urban/rural status to classify the square.

[1] `http://modis.gsfc.nasa.gov/data/dataprod/pdf/MOD_15.pdf`
[2] `http://geocommons.com/overlays/35982.kml`

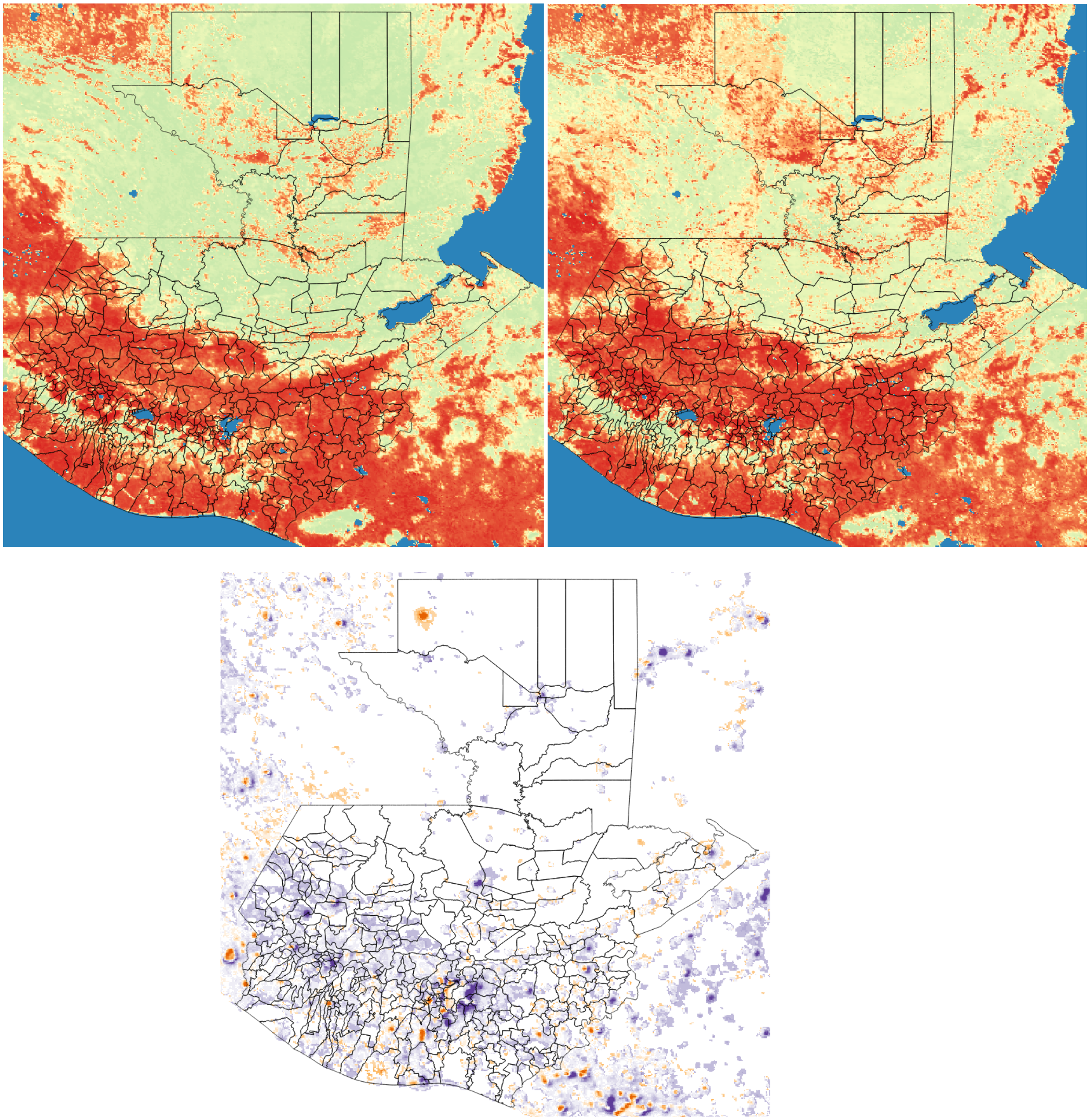

Figure 3: Left: Leaf Area Index, 2001, with municipality boundaries for Guatemala. Red=low LAI; blue=high LAI. Right: LAI, 2011. Bottom: difference, with more LAI=orange; less=purple.

This method resulted in 121,970 rural squares and 2,398 urban squares.

An appendix to this paper goes into greater detail on how the conversions were done.

3 Correlations

Figure 4 presents a correlogram relating a selection of variables for Guatemala. Bright red is a high positive correlation; bright blue is a high negative correlation; dim colors indicate weak correlation. Because each variable has perfect correlation with itself, the diagonal should be bright red, but we have censored the diagonal squares to zero to improve readability. With N in the hundreds, any correlation greater than about ± 0.15 is statistically significant.

All instances of FGT are well correlated. For rural and urban FGT in Guatemala, change in FGT is inversely correlated to FGT, meaning that there has been some regression to the mean over the course of the period studied.

Electrification, the percent of nonzero pixels in a region, and mean luminosity are all somewhat correlated. The leaf area index, FPAR, and albedo measures show a good correlation amongst themselves, but a weak correlation to the FGT and luminosity measures.

4 Regressions

This section presents a suite of regressions combining some luminosity measures and some traditional survey and census data.

For survey and census data, we used used the IGN data as above.

The dependent variables for the panels of regressions were percent change in FGT_2 for Guatemala. Regressions on other FGT measures were substantially similar.

The baseline is a prediction of the same change across all observed areas (the trivial regression with only a constant term). The other regressions include some common correlates to poverty, including percent indigenous, percent in agricultural work, and inmigration to the components of an observed area. The electrification regressions use the percent of households using electric lighting in a municipality. The luminosity regressions use the NOAA luminosity data, the Albedo/LAI/FPAR regressions use the NASA data, and the kitchen sink regression uses all of these things at once.

Tables 1 and 2 display the results for regressions on change in rural and urban poverty, respectively. Values in parens are p-values.

For the regressions regarding rural poverty, change in average log nonzero luminosity is statistically significant. This is true for both the set of luminosity pixels in a municipality marked as rural and as urban. This could be due to spillover effects such as commuters, the fact that many 1km^2 areas include a mix of urban and rural, or data quality/data handling issues. Percent indigenous was statistically significant, but other common predictors, including the percent using electric light and inmigration, were not.

For rural Guatemala, the corrected AIC shows less information loss relative to the base data in the regressions using luminosity [Burnham and Anderson, 1998]. For these regressions, the cross-validation $\sqrt{MSE}$ also shows better predictions using luminosity than without. The adjusted R^2 also improves with the inclusion of luminosity, indicating more variation is explained with luminosity than without.

For the regressions regarding extreme poverty in urban areas, the coefficients on change in average log nonzero luminosity had no statistical significance, while the survey measures had more

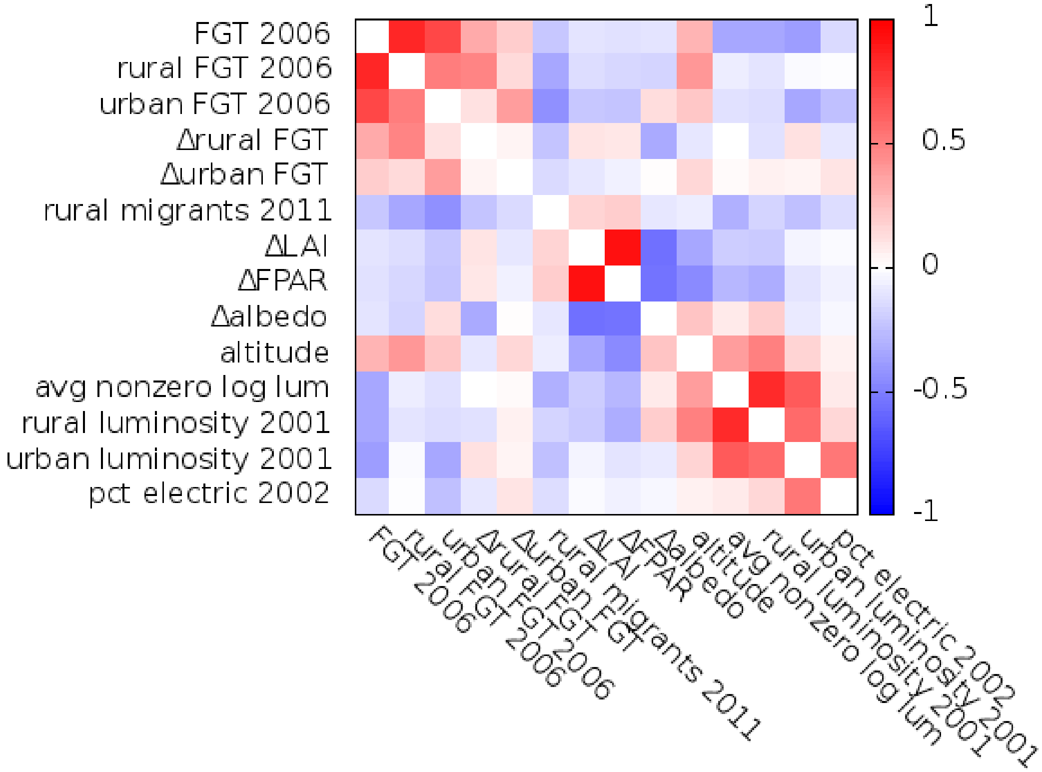

Figure 4: Correlations of variables over Guatemalan municipalities. Bright red is positive; bright blue is negative. Values on the diagonal (self-correlations) have been blanked out.

significant coefficients throughout. Model fit as reported by AIC$_c$ and adjusted R^2 showed a slight improvement relative to using survey data on electrification, while cross-validation $\sqrt{MSE}$ showed no improvement when using luminosity data.

In both panels, the results regarding verdancy were not significant, which is consistent with the correlograms above. Although a more sophisticated model could potentially generate better results, we dropped verdancy measures after this point.

5 Stress test

In this section, we run a set set of candidate regressions that explain as much of FGT_0 as is practicable using only survey data, selected via a stepwise regression. For the sake of simplicity, we restrict ourselves to linear models, and evaluate fit via comparisons of the Akaike Information Criterion for each regression. We ask whether the AIC of a regression built via stepwise procedure can improve with luminosity measures.

Our stepwise procedure begins with over forty survey variables for each country, and removes those with the least significance (exclusively according to p-value), until only very significant variables remain. Stepwise regressions are not known for producing especially useful models [Flom and Cassell, 2007]. They court problems with multiple testing, and the lists of variables our stepwise procedures selected are an implausible basis for models explaining poverty. But stepwise regressions do select models with a relatively high log likelihood and small residual errors. Adding another independent variable onto a regression built via stepwise procedure should produce very little change in model fit measures such as AIC, mean squared error, or R^2.

Thus, we use stepwise regression for a stress test: if we add luminosity measures to the model built to make the greatest use of the survey data, do the measures of model fit improve?

The stepwise regression for Guatemala includes these variables: rural population, rural literacy rate, rural indigenous population, urban indigenous population, health institutions, middle school, percent female head of household (HH), rural female literacy. Note that this set of variables was used for both rural and urban dependent variables. For regressions on poverty in 2000 and 2002, we used 2002 data; for 2011, we used 2011 data.

For Guatemala, we also include measures regarding electric light usage for some regressions, and added altitude to all regressions.

Figure 3 lists the AIC for regressions with and without log of nonzero luminosity included, and the change in AIC from adding the luminosity measure. A reduction in AIC is an improvement in the likelihood that the given model minimizes information loss. We do not report the regression coefficients and their significance levels because, as per the discussion above, the coefficients are not from a well-formed model and the significance levels are not reliable.

Luminosity did not do particularly well in our stress test. When added to a regression based on a constant and electrification, luminosity improved the AIC in most of our tests. But, with the exception of rural poverty in 2000, adding luminosity to the regressions built using the stepwise procedure did not significantly improve the AIC.[3]

[3] We downplay the importance of the one success of adding luminosity because, given seven trials adding luminosity to the stepwise regressions, this one success would not survive a correction for multiple testing.

6 Conclusion and future directions

This paper compared the data from surveys to the data from two types of satellite data, using a number of simple models. The comparison is on a very local scale, and distinguishes between urban and rural.

Consistent with the existing literature, we did find correlation between luminosity measures and poverty, and could use luminosity measures to improve (*ex post*) predictions of poverty rates. Luminosity seems unlikely to be sufficient to be a good predictor of changes in poverty rates by itself, but several measures showed that supplementing survey data with luminosity information was an improvement. In locations where survey data is not as reliable as that from Guatemala, one expects that the additional improvement from using sensor data to augment survey data would be greater. Especially given that satellite data is so freely available, there is promise in trying the experiment of augmenting survey data with satellite data in other contexts.

The measures of verdancy were not as effective. Their correlation with the survey measures was low, and the simple regressions correspondingly failed to show predictive power.

The Suomi National Polar-orbiting Partnership (between NASA and NOAA) is a satellite launched in 2011 with a set of sensors, including the Visible Infrared Imaging Radiometer Suite (VIIRS), intended as an improvement over MODIS.[4] One can expect that improved measurement would offer at least some reduction in the magnitude of errors in the regressions.

We did not precede this exploration of the data with a significant inquiry into how verdancy and poverty interact, and linear regressions may simply be the wrong model. Because this paper is a data exercise using simple models, one logical next step would be to incorporate all data using models more appropriate to small area estimation, such as those of Elbers et al. [2003] or Molina and Rao [2010]. Bauder et al. [2014] develop a Bayesian hierarchy that incorporates several data sources to describe small-area poverty rates in the United States; a similar exercise could be done using survey and satellite data.

Appendix: computing notes

To implement the analysis above, we used a sequence of freely-available tools in a pipeline that should be usable by any user comfortable with text files and basic scripting. This appendix describes the steps in the pipeline. All scripts that we wrote are available at `http://github.com/polynumeral/satellite-pilot`.

The set of line segments (the *polygon*) that forms the borders of a country or municipality are typically annotated using one of a few common file formats, including shapefiles and KML files (Keyhole markup language; Keyhole, Inc developed what is now Google Earth). Shapefiles and KML files for diverse geographies are readily available online. XML (extensible markup language) is a common format for expressing hierarchical information, and KML is a special-case use of XML. There exist programs and function libraries to read both shapefiles and XML (and by extension KML). We wrote a short script that takes in a KML file and a region name, and produces a plain-text list of latitude/longitude points within the boundaries of that point on a given grid. The in-or-out determination is made using the even-odd rule of Shimrat [1962].

The geoTIFF (TIFF=tagged image file format) is a relatively simple format for representing the value at a grid of points in a given space. The file includes a header describing the geography

[4] `http://npp.gsfc.nasa.gov/viirs.html`

covered, and a grid of values representing the value(s) at each pixel. At a resolution of a pixel per square km, this produces about 700MB of data for the full globe. geoTIFFs can be read by geographic information systems (GISes), such as ArcGIS, and some specialized libraries aimed at geoTIFF reading.

The *libtiff* C library, by Sam Leffler, provides a simple means of extracting a value from a given pixel in a TIFF. Like many C libraries, there are front-ends provided in many scripting languages, including Python, Ruby, et cetera. To minimize resource requirements, we used the plain C library. We wrote a short program to take in a list of latitude/longitude points (i.e., output from the KML-parsing program) and report the value in a geoTIFF file at that point. The output is a plain text list of latitude/longitude/value triplets, which can be read by any statistics package.

In both steps, we wrote a simple script to use an existing library that can read the file format (KML or geoTIFF) and output plain text usable anywhere. The tools needed to use the scripts are described in international standards (ISO/IEC 9899:2011 and ISO/IEC 9945:2008), and implementations are available on all common platforms.

There are point-and-click programs, some open and some with restrictive licensing requirements, that partially automate some of the above work. But one who hopes to replicate the original analysis will need to have the same program, which can create barriers to replication, the first of which may be the need to purchase licenses and compatible hardware. For example, the popular ArcGIS system is currently on version 10.1, which does not claim the ability to read ArcGIS databases from before version 9.2 (released November 2006).[5]

The MODIS data is released in HDF4 (Hierarchical Data Format). HDF4 had technical limitations and was not well-supported due to difficulties in use; it is largely replaced by HDF5. HDF is a container format, intended to embody a set of data sets; in the case of the MODIS data, the individual data sets are in geoTIFF format.

We used the HEG tool from NASA to convert from HDF4 format to geoTIFFs. It is also a filter, similar to the scripts we wrote and used for the geoTIFF $\to$ plain text and KML $\to$ plain text steps. HEG filters HDF4 $\to$ geoTIFF, which we can chain with the above tools to do the full HDF4 $\to$ geoTIFF $\to$ plain text extraction. Configuration notes for using HEG for the analysis here are also available at `http://github.com/polynumeral/satellite-pilot`.

In all cases, the plain text file we conclude with is a list of observations of the form (year, latitude, longitude, municipality, luminosity). A wide variety of familiar programs can read this list and calculate the mean luminiosity for a municipality in a given year, or other comparable statistics.

References

American Statistical Association. *H2R 2012: Survey Methods for Hard-to-reach Populations Conference Proceedings*. 2012. URL `http://www.eventscribe.com/2012/ASAH2R/`.

Mark Bauder, Donald Luery, and Sam Szelepka. Small area estimation of health insurance coverage in 2010–2012. Technical report, U.S. Census Bureau Social, Economic, and Housing Statistics Division; Small Area Methods Branch, February 2014. URL `http://www.census.gov/did/www/sahie/methods/files/sahie_tech_2012.pdf`.

Haimanti Bhattacharya and Robert Innes. Is there a nexus between poverty and environment in rural

[5] `http://resources.arcgis.com/en/help/main/10.1/index.html#//003n00000008000000`

india? In *American Agricultural Economics Association Annual Meeting*, Long Beach, CA, July 2006.

Kenneth P. Burnham and David Raymond Anderson. *Model Selection and Inference: A Practical Information-Theoretic Approach*. Springer Verlag, 1998.

Laurence Chandy. Counting the poor: Methods, problems and solutions behind the $1.25 a day global poverty estimates. Investments to end poverty working paper, Development Initiatives and Brookings Institution, May 2013. URL `http://www.brookings.edu/research/papers/2013/05/global-poverty-counting-the-poor-chandy`.

Xi Chen and William D Nordhaus. Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences*, 108(21):8589–8594, 2011. URL `doi://10.1073/pnas.1017031108`.

Steeve Ebener, Christopher Murray, Ajay Tandon, and Christopher C Elvidge. From wealth to health: Modelling the distribution of income per capita at the sub-national level using night-time light imagery. *International Journal of Health Geographics*, 4(5), 2005.

Chris Elbers, Jean O. Lanjouw, and Peter Lanjouw. Micro–level estimation of poverty and inequality. *Econometrica*, 71(1):355–364, 2003. URL `http://dx.doi.org/10.1111/1468-0262.00399`.

Christopher D Elvidge, Kimberly E Baugh, Eric A Kihn, Herbert W. Kroehl, and Ethan R. Davis. Mapping city lights with nighttime data from the DMSP operational linescan system. *Photogrammetric Engineering and Remote Sensing*, 63(6):727–734, 1997.

Christopher D Elvidge, Kimberly E Baugh, John B Dietz, Theodore Bland, Paul C Sutton, and Herbert W Kroehl. Radiance calibration of DMSP-OLS low-light imaging data of human settlements. *Remote Sensing of Environment*, 68(1):77–88, 1999.

Christopher D Elvidge, Kimberly E Baugh, Sharolyn J Anderson, Paul C Sutton, and Tilottama Ghosh. The night light development index (NLDI): A spatially explicit measure of human development from satellite data. *Social Geography*, 7:23–35, 2012.

P. L. Flom and D. L. Cassell. Stopping stepwise: Why stepwise and similar selection methods are bad, and what you should use. In *NESUG 2007 Proceedings*, 2007.

William H Greene. *Econometric Analysis*. Prentice Hall, 2nd edition, 1990.

J Vernon Henderson, Adam Storeygard, and David N Weil. Measuring economic growth from outer space. Technical Report 15199, 2009. URL `http://www.nber.org/papers/w15199`.

Isabel Molina and J. N. K. Rao. Small area estimation of poverty indicators. *Canadian Journal of Statistics*, 38(3):369–385, 2010. ISSN 1708-945X. doi: 10.1002/cjs.10051. URL `http://dx.doi.org/10.1002/cjs.10051`.

Maxim Pinkovskiy and Xavier Sala-i Martin. Lights, camera,... income!: Estimating poverty using national accounts, survey means, and lights. Working Paper 19831, National Bureau of Economic Research, January 2014. URL `http://www.nber.org/papers/w19831`.

Xavier Sala-i-Martin and Maxim Pinkovskiy. African poverty is falling...much faster than you think! Working Paper 15775, National Bureau of Economic Research, February 2010. URL http://www.nber.org/papers/w15775.

M Shimrat. Algorithm 112: Position of point relative to polygon. *Communications of the ACM*, 5, August 1962.

Paul C Sutton, Christopher D Elvidge, and Tilottama Ghosh. Sub-national scales using nighttime satellite imagery. *International Journal of Ecological Economics & Statistics*, 8(S07):5–21, Summer 2007.

Table 1: Regressions on $\%\Delta$rural extreme FGT, Guatemala from 2005 to 2009.

	Const	Electrification	Luminosity	Alb/LAI/FPAR	All
Constant	-0.06 ***	-0.04	-0.02	-0.00	-0.05 *
	(0.0)	(0.105)	(0.405)	(0.896)	(0.095)
Population density		2.07×10^{-7}	-0.08	-0.04	-0.05
		(0.363)	(0.112)	(0.479)	(0.344)
% indigenous, 2002		-0.04 *	-0.04 **	-0.05 **	-0.04
		(0.084)	(0.029)	(0.022)	(0.102)
% in migration, 2005		-0.37	-0.60	-0.52	-0.65
		(0.553)	(0.346)	(0.434)	(0.286)
% electric light, 2005		0.04			0.06
		(0.262)			(0.137)
Δlog lum, urban 2000–11			0.01***		0.01***
			(0.0)		(0.0)
Δlog lum, rural 2000–11			-0.01 ***		-0.02 ***
			(0.003)		(0.002)
Δ LAI 2001-11				0.01	0.01
				(0.366)	(0.2)
Δ FPAR 2001-11				-0.00	-0.01
				(0.574)	(0.344)
Δ Albedo 2001-11				-0.00	-5.53×10^{-5}
				(0.73)	(0.952)
degrees of freedom	139.00	134.00	133.00	132.00	129.00
R^2_{adj}	0.01	0.16	0.24	0.16	0.24
AIC$_c$	-320.49	-338.15	-351.75	-335.09	-346.91
Cross-validation $\sqrt{\mathrm{MSE}}$	0.08	0.07	0.07	0.07	0.06

Table 2: Regressions on %Δurban (extreme FGT), Guatemala from 2005 to 2009.

	Const	Electrification	Luminosity	Alb/LAI/FPAR	All
Constant	-0.00	-0.01	0.01	0.02	0.02
	(0.599)	(0.345)	(0.176)	(0.111)	(0.228)
Population density		1.20×10^{-7}	-0.09^{***}	-0.09^{***}	-0.09^{***}
		(0.365)	(0.002)	(0.001)	(0.002)
% indigenous, 2002		0.03^{**}	0.04^{***}	0.04^{***}	0.04^{***}
		(0.013)	(0.001)	(0.001)	(0.004)
% in migration, 2005		0.35	-0.03	-0.08	0.02
		(0.344)	(0.934)	(0.838)	(0.966)
% electric light, 2005		0.01			-0.01
		(0.56)			(0.556)
Δlog lum, urban 2000–11			-0.00		-0.00
			(0.912)		(0.906)
Δlog lum, rural 2000–11			0.00		0.00
			(0.711)		(0.539)
Δ LAI 2001-11				-0.00	-0.00
				(0.566)	(0.546)
Δ FPAR 2001-11				0.00	0.00
				(0.675)	(0.648)
Δ Albedo 2001-11				-0.00	-0.00
				(0.65)	(0.666)
degrees of freedom	139.00	134.00	133.00	132.00	129.00
R^2_{adj}	0.01	0.04	0.10	0.09	0.07
AIC$_c$	-489.94	-489.29	-496.22	-494.31	-487.92
Cross-validation $\sqrt{\text{MSE}}$	0.04	0.04	0.04	0.04	0.04

Rural GTM FGT 2000

regressors	AIC w/o luminosity	AIC w/luminosity	Δ AIC
stepwise	-1593.13	-1605.11	-11.98
electricity	-1461.77	-1492.99	-31.22
stepwise + electricity	-1588.6	-1612.82	-24.22

Rural GTM FGT 2006

regressors	AIC w/o luminosity	AIC w/luminosity	Δ AIC
stepwise	-1722.317	-1720.735	1.582
electricity	-1579.392	-1595.021	-15.629
stepwise + electricity	-1723.784	-1724.911	-1.127

Rural GTM FGT 2011

regressors	AIC w/o luminosity	AIC w/luminosity	Δ AIC
stepwise	-1655.635	-1653.821	1.576
electricity	-1632.216	-1630.52	1.696
stepwise + electricity	-1662.075	-1664.526	-2.451

Urban GTM FGT 2000

regressors	AIC w/o luminosity	AIC w/luminosity	Δ AIC
stepwise	-1825.75	-1824.552	1.198
electricity	-1751.622	-1767.863	-16.241
stepwise + electricity	-1867.577	-1870.399	-2.822

Urban GTM FGT 2006

regressors	AIC w/o luminosity	AIC w/luminosity	Δ AIC
stepwise	-1791.662	-1789.67	1.512
electricity	-1728.383	-1737.905	-9.522
stepwise + electricity	-1833.12	-1834.363	-1.243

Table 3: Comparisons of AICs for linear models with and without luminosity