

## Dynamic Socio-economic Analysis: The Power of Persistence from Small Satellites

John M. Irvine, Richard P. Wood, Kimberly A. Slater

Charles Stark Draper Laboratory, Inc.  
555 Technology Square, MS 31  
Cambridge, MA 02139-3563

Phone: 617-258-4957, email: [jirvine@draper.com](mailto:jirvine@draper.com)

### ABSTRACT

Understanding of economic, social, and cultural characteristics of a society is critical to effective government policy and successful commercial undertakings. Obtaining this information, however, often requires direct observations and interactions with the local populace, entailing significant costs and potentially exposing data collectors to heightened risks. We address this challenge by combining automated processing of satellite imagery with advanced modeling techniques. We have developed methods for inferring measures of well-being, governance, and related socio-cultural attributes from satellite imagery. This research represents a fundamental innovation in the study of human geography by explicitly analyzing the relationship between the observable physical attributes and the societal characteristics and institutions of the region.

Through analysis of commercial satellite imagery and coincident survey data, we have developed and tested models for rural Afghanistan and selected countries in sub-Saharan Africa (Botswana, Kenya, Zimbabwe). The findings show the potential for predicting peoples' attitudes about the economy, security, leadership, social involvement, and related questions, based only on the imagery-derived information. When tested on sequestered data in Afghanistan, the image-based model predict with 79% accuracy whether villagers will volunteer their time to support a community project (an indicator of social capital), and with 78% accuracy whether the village will look to the government or local resources for protection. These models also predict the likelihood of a respondent supporting the village council, growing opium poppies, or going on Hajj. Similar models for the African region also provide useful indicators. Models for predicting economic attributes (presence of key infrastructure, attitudes about the economy, perceptions of crime, and outlook towards the future) all exhibit statistically significant performance.

These methods show significant promise for assessing key social indicators. However, the models only capture a snapshot in time. With the emergence of new small imaging satellites, the potential for temporal analysis offers a substantial improvement over previous work. By monitoring changes in physical structures and patterns of commercial activities, a far richer understanding of societies will be possible through image-based methods. Using a recent airborne imagery collection campaign, which provides a surrogate for the anticipated frequent coverage achievable with small satellites, we demonstrate a proof-of-concept analysis of traffic patterns and temporal analysis to understand local economic activities. The paper concludes with recommendations for future exploration.

### INTRODUCTION

Remote sensing offers a rich source of data that can overcome limitations associated with *in situ* observation and useful information can be extracted or estimated from the remote sensing data. The application of remote sensing to the social sciences is an emerging research area [1]. "Remote sensing can provide measures for a number of dependent variables associated with human activity—particularly regarding the environmental consequences of various social, economic, and demographic processes." [2] We recognize that people's behavior and values depend on the local environment. Similarly, people will modify the local

conditions to satisfy certain cultural needs. Researchers have explored these issues, including socio-economic attributes, ethnography, and land use [3, 4, 5]. A review of the literature indicates that investigators are exploring the connection between remote sensing and social science issues. Much of the research focuses on urban studies, demography, archaeology, land use and land cover. Our areas of investigation include deeper exploration of social, economic, and governmental attributes. The revolution spurred by small imaging satellites offers the potential to go beyond static analysis and quantify dynamic processes and activities.

Numerous indicators of economic status (housing, vehicles, crop land, livestock, infrastructure) are evident from overhead imagery and a few studies have explored the relationship between remote sensing data and the economy [6]. Indicators of changes in economic status can also be observed. Lambin, for example, found a relationship between deforestation and economic crisis in Cameroon [7]. Observables associated with economic well-being, including measures of wealth distribution, are evident from satellite imagery. Furthermore, these economic indicators are relevant to understanding governance and social capital. High social capital has been linked to good governance [8]. Similarly, evidence of economic growth can be associated with higher levels of social capital [9]. Furthermore, although direct measurement of social cohesion and connectedness is not possible, durable institutions (e.g., schools, places of worship) and infrastructure (e.g., roads) are indirect indicators of social connectedness.

## RATIONALE

Physical structures provide imagery observables that indicate local economic activity, i.e. indicators of overall well-being and the variation in wealth across the population. These economic indicators also have relationships to trends in governance and social capital. Some key indicators that have emerged from empirical studies in the social sciences suggest relations between the economic indicators and social/ political factors:

- Economic inequality negatively affects the establishment and maintenance of democratic practice [10, 11].
- Ethnic diversity reduces public goods provision [12].
- Lower GDP negatively affects democracy and public goods provision [13].
- Democratic practice increases overall wealth [14].
- Lower concentration of households leads to lower production of public goods [15].

## APPROACH AND METHODOLOGY

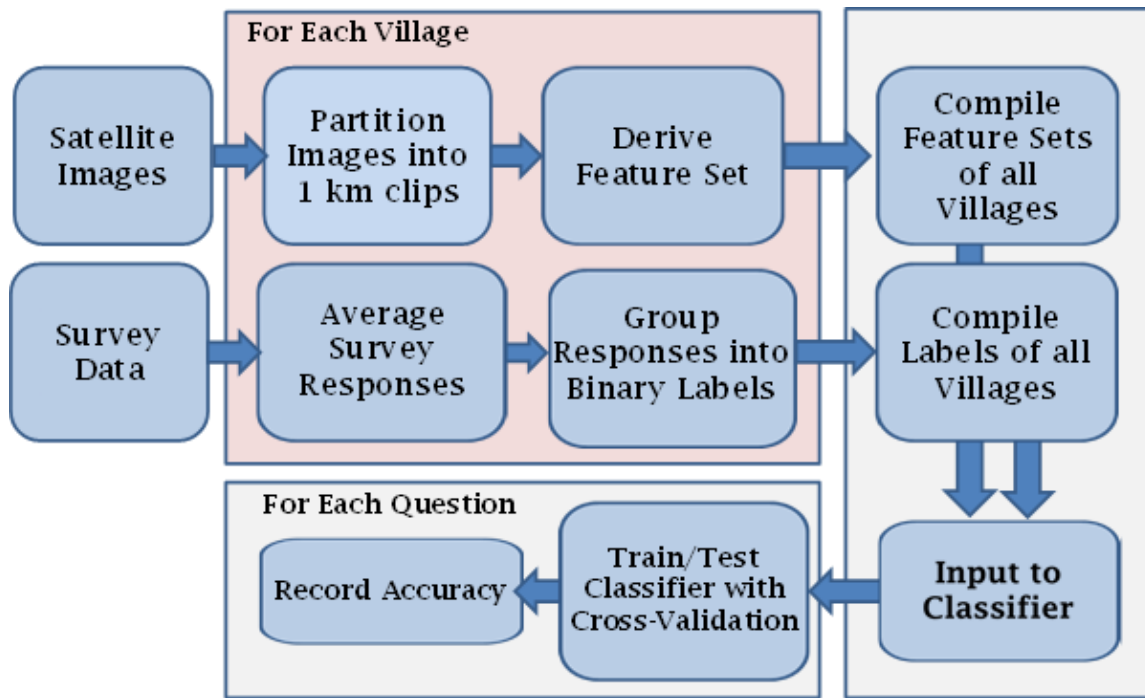
Combining political, socio-cultural and economic theory with rich sources of survey data and overhead imagery, we analyze the relationships between measurements acquired from direct surveys and features inferred from remotely-sensed data. The goal is to develop models that predict the values of specific indicators using the features extracted from the satellite imagery (Figure 1). The approach consists of five steps:

- *Analysis of survey data:* Analysis of the survey questions and responses produces indicators of local attitudes about economic well-being and governance. This establishes the “gold standard”

for what we want to predict from imagery-derived measurements.

- *Imagery analysis:* Using image processing techniques and visual inspection to extract relevant features from the imagery, we construct measures of the local conditions.
- *Model development:* Statistical analysis of the relationships between the survey-based indicators and imagery-derived features yields candidate models. The overhead imagery covers the same geographic region as the survey data and the geolocation of survey responses is the link between the two data sources.
- *Model validation and analysis:* Model validation requires a separate set of survey and imagery data that is held back (i.e., sequestered) during the model development process. Using this sequestered set of data, we compute survey-based indicators and imagery-derived features. Comparing the observed survey-based indicators to model predictions quantifies the model’s performance.
- *Applications of the model:* The result of this process is a model for predicting the survey-based indicators based on imagery-derived features. The model can now be applied on a broader scale to predict relevant social indicators.

In previous papers, we have reported our initial exploration of these issues for rural Afghanistan. [16, 17, 18, 19, 20]. This earlier work analyzed a large corpus of imagery data, collected by commercial imaging systems during 2007 – 2012. We focused on 500 villages across 10 Districts of northern Afghanistan. These villages were participants in the National Solidarity Programme (See web site: <http://www.nspafghanistan.org/>), which was the source of survey data. Applying the methodology described above, we developed models for a range of economic and social indicators. Analysis of commercial satellite imagery and coincident survey data for rural Afghanistan shows the potential for predicting peoples’ attitudes about the economy, security, leadership, social involvement, and related questions. The methodology for extracting relevant measures from the imagery combines human-guided and automated methods to identify salient features. Comparing the imagery-derived information to the survey responses for the 500 Afghan villages, we developed and validated models which predict survey responses. Across 76 survey-based indicators, classification accuracy was approximately 70-90% on the test data [18, 19, 20]. This paper extends these methods to a region in sub-Saharan Africa and compares performance based on two imaging systems: DigitalGlobe and PLANET.



**Figure 1. Overview of the Model Development and Testing Process**

## SATELLITE IMAGERY

Cultural, social, and economic factors that are critical to understanding societal attitudes are associated with specific phenomena that are observable from satellite imagery. Distinguishing among industrial, commercial, and residential areas, for example, is a standard use for imagery. Measures of socio-economic status (e.g., house size, crop area and crop vigor, presence of vehicles) can also be extracted from high-quality imagery. Other factors may also be inferred from indicators derived from the imagery such as population density, access to improved roads, distances to commercial and governmental centers, and attributes of communities. Patterns which emerge from the correlation of the geospatial analysis with the survey results can suggest phenomena observable in imagery that are reasonable surrogates for the direct measurements of public opinions. We explored several image processing techniques to extract and identify specific indicators that can be directly inferred from the imagery.

The imagery available for this study consists of approximately 40TB of commercial satellite images, both panchromatic and multispectral data, acquired over Kenya, Botswana, and Zimbabwe during the period of 2007-2012. A thorough discussion of the processing for the DigitalGlobe imagery appears in a

recent Technical Report [21]. By comparison, the imagery coverage was more extensive than for our earlier analysis of Afghanistan. Whenever possible given the available DigitalGlobe data, we selected imagery with matching panchromatic and multispectral data that were collected together.

The image processing began with geo-location of the survey data identified for this study (figure 2). The location could be a village or a neighborhood within a larger urban area. We chipped out a small region of fixed size covering the center point and surrounding area. Separate chips of panchromatic and multi-spectral imagery were extracted from the full scenes. The chips were then co-registered and fused to form a pan-sharpened product, using a fast pan-sharpening method [22, 23]. Although pan-sharpening was employed with the DigitalGlobe data, the PLANET imagery was already 4-band data at native resolution and no sharpening was performed.

Analysis of the imagery leveraged both spatial and spectral information to identify features of interest. The specific features extracted from the imagery include estimates of the number of buildings, building sizes, building density, extent and health of crops, and types of roads. Derived features also quantified the size, shape, and density of buildings. In addition to the automated processing, visual inspection of the imagery

produced ratings of relevant characteristics. Comparison of the computed features to observer ratings demonstrates the performance of the automated methods [17]. The image features used for model development were:

- Vegetation health as quantified by the Normalized Difference Vegetation Index (NDVI):
- Density of built up or urban areas: Man-made structures produce edge features and measures of edge density in the scene provide good indicators

- Spatial covariance metrics: Repeated patterns are indication of planned residential communities common in middle-class neighborhoods
- Spectral angle: The four-band multispectral data yields some information about material type. Spectral analysis shows that the material properties of the buildings differ from the surrounding region. Using the spectral angle, we identified regions as man-made (roads and buildings), vegetation, and bare soil.”

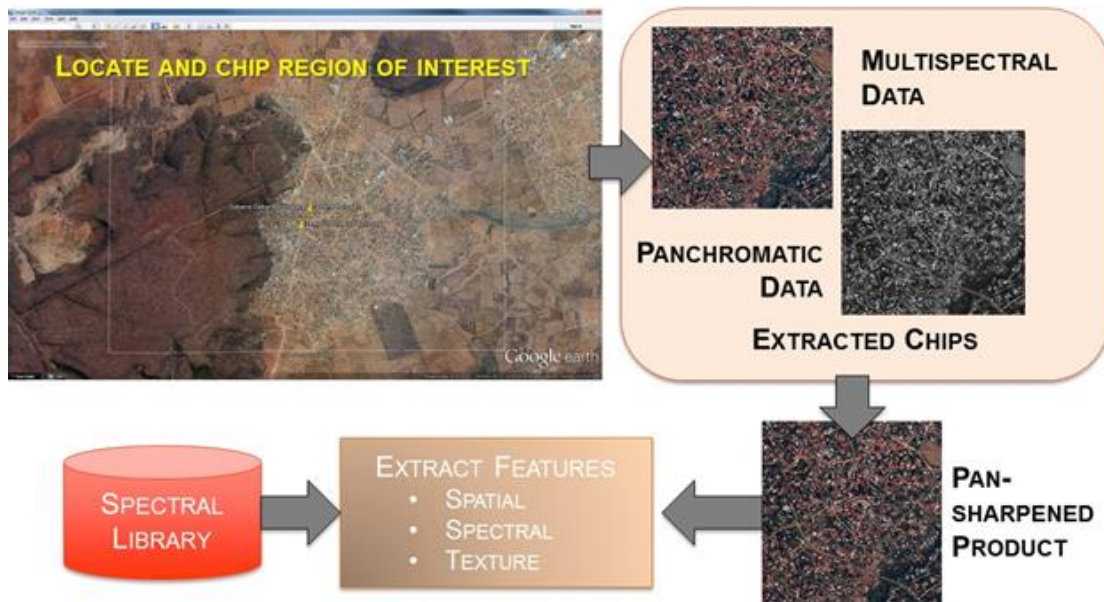


Figure 2. Image Processing Workflow

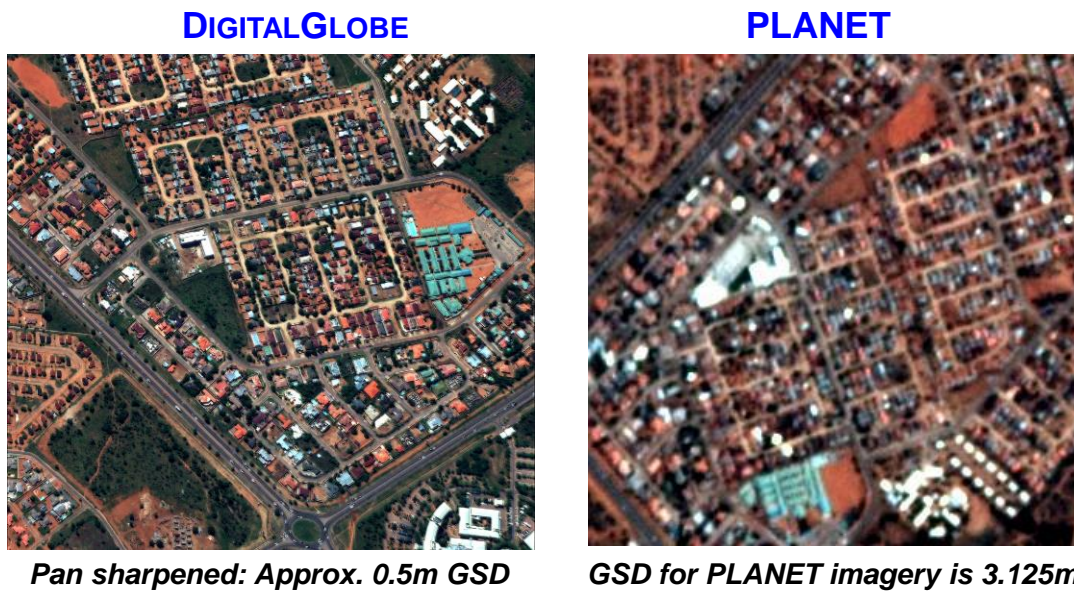


Figure 3. DigitalGlobe and PLANET imagery over the same village in Botswana  
 Note: All imagery was provided by NGA

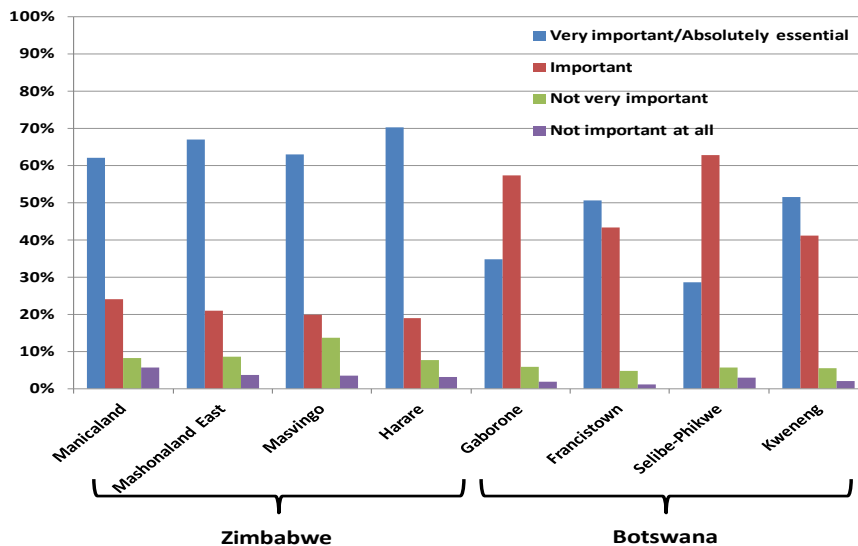
## SURVEY DATA

The survey data acquired by the Afrobarometer Program [24] provide a rich portrait of societal attitudes across several countries and multiple regions within each country (Table 1). For some countries, surveys have been repeated over multiple years, giving a temporal characterization of shifting

attitudes and opinions. For instance, one survey question explored attitudes concerning the importance of having a democratic society and the freedom to criticize the government. Responses show distinct patterns by country and, to a lesser extent, within countries (Figure 4). Deeper analysis reveals that within a country, attitudes vary with the level of urbanization and the economic status of respondents.

**Table 1. Issues addressed in Afrobarometer Surveys**

Democracy	Popular understanding of, support for, and satisfaction with democracy, as well as any desire to return to, or experiment with, authoritarian alternatives.
Governance	The demand for, and satisfaction with, effective, accountable and clean government; judgments of overall governance performance and social service delivery.
Livelihoods	How do families survive? What variety of formal and informal means do they use to gain access to food, shelter, water, health, employment and money?
Macro-economics and Markets	Citizen understandings of market principles and market reforms and their assessments of economic conditions and government performance at economic management.
Social Capital	Whom do people trust? To what extent do they rely on informal networks and associations? What are their evaluations of the trustworthiness of various institutions?
Conflict and Crime	How safe do people feel? What has been their experience with crime and violence?
Participation	The extent to which ordinary people join in development efforts, comply with the laws of the land, vote in elections, contact elected representatives, and engage in protest. The quality of electoral representation.
National Identity	How do people see themselves in relation to ethnic and class identities? Does a shared sense of national identity exist?



**Figure 4. “Democratic society and freedom to criticize the government.” Responses by country and region.**

## Model Validation

To develop and validate the models, we randomly partitioned the data into training and test sets. Training of the classifier, including feature selection and model estimation, relied only on the training data. We score the performance by comparing the image-based predictions to the values observed in the survey data, using only the testing data. The probability of correct classification computed from this analysis is our primary overall measure of the model's performance.

To demonstrate the process and illustrate performance, we present the results for selected survey questions using data from Botswana. The classifier was a k-nearest neighbor method with  $k = 3$ . The data were randomly partitioned such that 70% was for training and 30% was for testing. Our model classified the observations at the level of individual respondents. An alternative approach is to aggregate the data to the village level and perform classification for each location. This location-level analysis has been explored in related studies [21, 22].

The model validation demonstrates the level of performance currently achieved using only image-based features to predict survey responses. Feature selection indicated that certain features are critical to these predictions: spatial features indicative of urbanization and spectral features quantifying vegetation health and prevalence. Typical performance results using DigitalGlobe imagery appear in table 2, for individual response in Botswana. For comparison, table 3 shows performance for the same questions using models based on PLANET data.

## CONCLUSIONS AND FUTURE DIRECTIONS

In the analysis of selected sub-Saharan African regions, image-derived features provide useful information for predicting survey responses across a range of questions. Performance is generally strongest with questions about infrastructure, such as access to electricity, clean water, and sewage disposal. Some questions about social attitudes, however, are performing only slightly better than chance. When compared to the results from the earlier study in Afghanistan, the performance here is less compelling.

There are several important differences between Afghanistan and sub-Saharan Africa that could account for the differences in performance. First, the locations in Afghanistan are almost universally small villages, whereas the study areas in Africa include larger, more dispersed built-up areas. Some of the Africa data

includes major cities such as Gaborone. For these more extensive regions, the one-kilometer squares lack the spatial extent and context that could be very relevant to understanding local attitudes. A second factor is the difficulty in matching locations between imagery and survey data. For major urban areas, the initial data was insufficient to disambiguate locations within the city. We conducted further research into available gazetteers and were able to refine the location data, but this occurred late in the project and is not reflected in these results. Using the improved information, a more focused analysis of urban areas is possible.

We have identified several areas for future investigation that could improve the model performance. There is a wide range of possible features that can be extracted from the imagery data. The initial exploration of the feature space was guided by the previous research in Afghanistan and reasoning about the phenomena relate to socio-economic conditions. However, alternative features could provide better characterization of the phenomena and there is considerable room for improvement. In related work, McBee demonstrated improved performance using a bag of visual words approach, which quantified built up areas and distinguished between residential and industrial areas. [25, 26]

One limitation of the current work is that it captures a snapshot in time. The survey occurs over a fairly short time interval and the imagery was collected at dates close to the survey. For some locations, multiple images are available and it is possible to assess changes over a short time period. One change that is evident in the imagery arises from changes in vegetative health arising from access to water. The changes can be due to seasonal variation or differences in weather patterns. In 2010 and 2011, for example, Kenya experienced its worst drought in 60 years. To illustrate the effects on image-derived features, Figure 5 presents two images of Harare, Zimbabwe, showing the differences in vegetation health.

Temporal analysis with more frequent revisit offers the potential to quantify local activity indicative of commercial vibrancy (figure 6). Future research will explore the potential for temporal analysis to capture important indicators or activity. These indicators include periodic processes, such as seasonal variation, which is not captured by the single point in time. The non-stationary variation, such as upward or downward trends in vegetation cover and vegetation health could reveal economic shifts in the society. More subtle change detection can be employed to detect areas with high levels of vehicular traffic which is also indicative of commercial activity.

**Table 2. Classifier Performance Selected Questions in Botswana using DigitalGlobe Data**

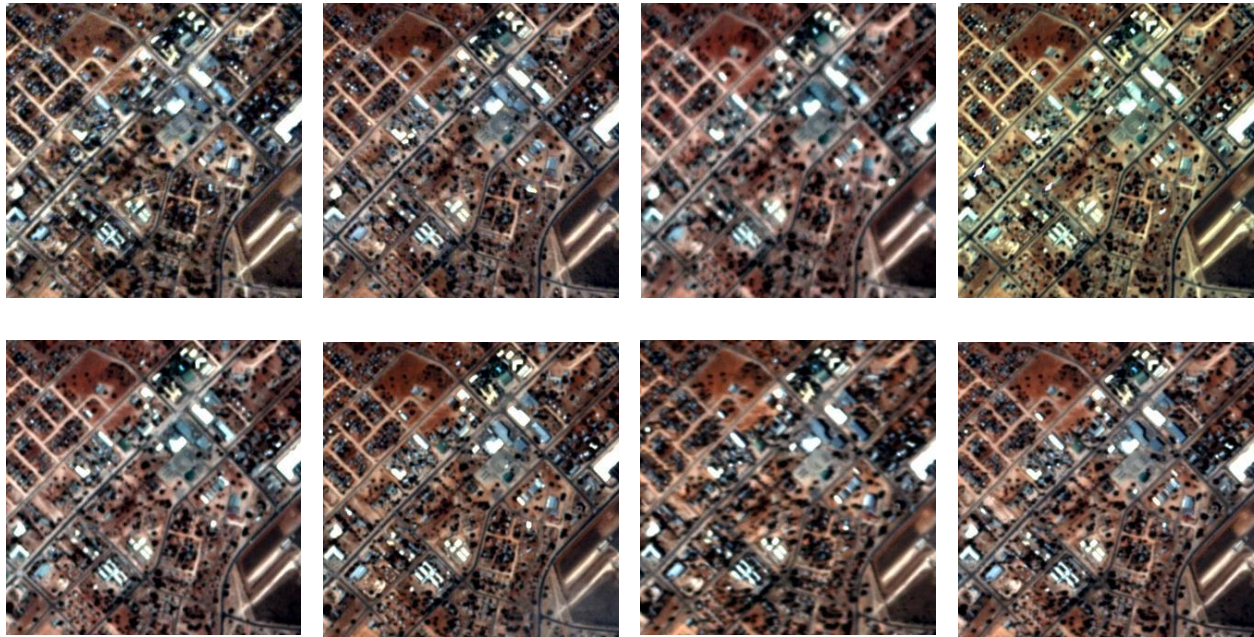
Survey Question	Valid Responses	Percent Correctly Classified
House has access to electricity	Yes or No	91.9
House has access to piped water	Yes or No	94.1
House has access to sewage system	Yes or No	85.2
Living condition in the country	Good/neutral/bad	41.7
Own living condition	Good/neutral/bad	45.6
Living conditions relative to other Batswana	5 categories	28.6
Access to enough food	Yes or No	57.1
Access to enough clean water	Yes or No	62.5
Fear of crime	Yes or No	58.5
Had something stolen from house	Yes or No	70.7

**Table 3. Classifier Performance Selected Questions in Botswana using PLANET Data**

Survey Question	Valid Responses	PCC: PLANET
House has access to electricity	Yes or No	98.7
House has access to piped water	Yes or No	98.0
House has access to sewage system	Yes or No	86.1
Living condition in the country	Good/neutral/bad	36.9
Own living condition	Good/neutral/bad	49.0
Living conditions relative to other Batswana	5 categories	34.5
Access to enough food	Yes or No	56.6
Access to enough clean water	Yes or No	72.6
Fear of crime	Yes or No	51.5
Had something stolen from house	Yes or No	60.2



**Figure 5. Temporal differences can include large changes in soil moisture and vegetation health due to seasonal patterns and drought conditions *Note: Imagery from DigitalGlobe provided by NGA***



**Figure 6. Temporal differences in PLANET show activity levels**

*Note: Imagery from PLANET provided by NGA*

## REFERENCES

1. Goodchild, M.F., L. Anselin, R. P. Appelbaum and B. H. Harthorn. 2000. "Toward Spatially Integrated Social Science," *International Regional Science Review* April 2000 vol. 23 no. 2 139-159
2. NRC. 1998. *People and Pixels: Linking Remote Sensing and Social Science*, Committee on the Human Dimensions of Global Change, National Research Council, 1998.
3. Jenson, J. R. and D. C. Cowen. 1999. Remote Sensing of Urban Infrastructure and Socio-economic attributes, *Photogrammetric Engineering and Remote Sensing*, vol.65, no. 5, pp.611-622, 1999.
4. Jiang, H. 2003. "Stories Remote Sensing Images Can Tell: Integrating Remote Sensing Analysis with Ethnographic Research in the Study of Cultural Landscapes," *Human Ecology*, Volume 31, Number 2, 215-232, June 2003.
5. Weeks, J.R. 2003. Using Remote Sensing and Geographic Information Systems to Identify the Underlying Properties of Urban Environments, In T. Champion and G. Hugo, Editors. *New Forms of Urbanization: Beyond the Urban-Rural Dichotomy*. Aldershot, UK: Ashgate Publishing Co., 2003
6. Elvidge C.D., Baugh KE, Kihn EA, Kroehl HW, Davis ER, Davis CW. 1997. "Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption." *International Journal of Remote Sensing* 1997; 18: 1373-9.
7. Eric F. Lambin, "Linking Socioeconomic and Remote Sensing Data at the Community or at the Household Level: Two Case Studies from Africa" in Fox, J., R. Rindfuss, S.J. Walsh, and V. Mishra (eds). 2003. *People and the Environment: Approaches for Linking Household and Community Surveys to Remote Sensing and GIS*. Boston: Kluwer.
8. Samuel Bowles and Herbert Gintis, "Social Capital and Community Governance" *Economic Journal*, Vol. 112, pp. 419-436, 2002
9. Knack, S. and P. Keefer. 1997. "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation," *Quarterly Journal of Economics*, 112(4), pp. 1251-1288
10. Carles Boix, *Democracy and Redistribution*, Cambridge, UK: Cambridge University Press.
11. Daron Acemoglu and James Robinson, *Economic Origins of Dictatorship and Democracy*, Cambridge, UK: Cambridge University Press.
12. James Habyarimana, Macartin Humphreys,



- Daniel Posner, and Jeremy Weinstein, "Why Does Ethnic Diversity Undermine Public Goods Provision?" *American Political Science Review* 101(4): 709-725.
13. Kennedy, Ryan. 2010. "the Contradiction of Modernization: A Conditional Model of Endogenous Democratization." *Journal of Politics* 72(3): 785-798.
  14. Dani Rodrik. (2000), "Institutions for High-Quality Growth," *Studies in Comparative International Development* 35(3): 3-31.
  15. James Rauk. 1993, "Productivity Gains from Geographic Concentration of Human Capital." *Journal of Urban Economics* 34(3): 380-400.
  16. John M. Irvine, Janet Lepanto, John Regan, Mon Young, "Deriving Economic and Social Indicators from Imagery" IEEE Applied Imagery and Pattern Recognition Workshop, 9-11 October 2012, Cosmos Club, Washington, DC
  17. John Irvine, Janet Lepanto, Natasha Markuzon, John Regan, Edward Vaisman, Mon Young, Fotini Christia, Roger Petersen, (2013) "Characterization of Villages in Rural Afghanistan" American Soc. of Photogrammetry and Remote Sensing (ASPRS) 2013 Annual Conference, March 25 – 29, 2013, Baltimore, Maryland.
  18. John M. Irvine, "Transforming Data into Information: Enabling Detection and Discovery for Socio-Cultural Analysis" in *Sociocultural Behavior Sensemaking: State of the Art in Understanding the Operational Environment*, Jill Egeth, Gary Klein, and Dylan Schmorrow (eds), MITRE Corporation
  19. John M. Irvine, Jennessa Kimball, John Regan, Janet Lepanto; "Application of Commercial Remote Sensing to Issues in Human Geography" *IEEE Applied Imagery Pattern Recognition Conference*, 23-25 October 2013, Washington, DC.
  20. John Irvine, Janet Lepanto, Remote Sensing and Indicators of Well-being and Governance, Draper Technical Report, September 2013.
  21. John M. Irvine, Richard J. Wood, Payden McBee, "Image-based Socio-Cultural Prediction Models", 2017 Technical Report (in press).
  22. Lee, Jonghwa and Lee, Chulhee, 2010. "Fast and Efficient Panchromatic Sharpening," *IEEE Transactions On Geoscience and Remote Sensing*, Vol. 48, No. 1, January 2010
  23. Kriegler, F.J., Malila, W.A., Nalepka, R.F., and Richardson, W., 1969. "Preprocessing transformations and their effects on multispectral recognition," *Proceedings of the Sixth International Symposium on Remote Sensing of Environment*, p. 97-131
  24. For a description of the Afrobarometer Program and survey data, see <http://www.afrobarometer.org/>
  25. J.M. Irvine, R. J. Wood, P. McBee, "Image-based Socio-Cultural Prediction Models" Technical Report, April 2017.
  26. P. McBee, "Exploring Feature Sets From Satellite Imagery In Socio-Economic Predictions In Sub-Saharan Africa", M.S. Thesis, Northeastern University, Mat 2017.