Households from Space

Integrating Household Surveys with Geospatial Data Sources for Improved Monitoring of Development Outcomes

Talip Kilic
Senior Economist
Living Standards Measurement Study (LSMS)
Development Data Group, The World Bank

Applying Quantitative Analysis to Development Issues Conference
Bibliotheca Alexandrina | Alexandria, Egypt | February 18-19, 2018

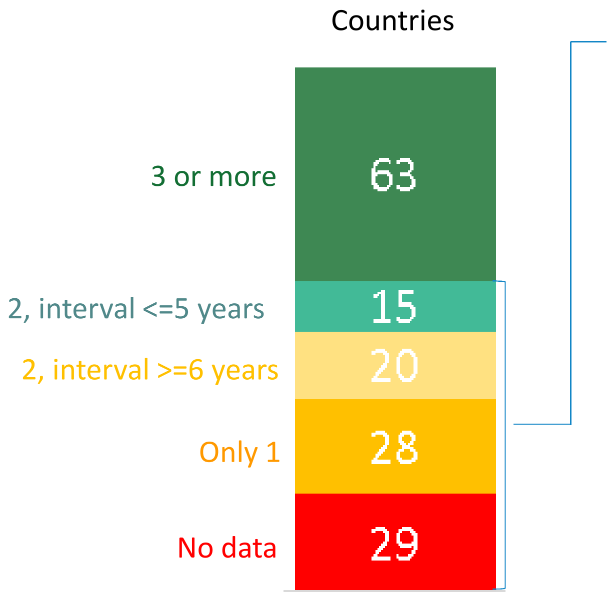
LSMS Living Standards Measurement Study

A propitious time for data

- Increased demand for data ...
  - Globally
    - World Bank: new data strategy under Development Data Council
  - At national and sub-national level
    - Increased accountability
    - More evidence-based policy decisions

- Household surveys at core of satisfying this demand

The sobering news: despite increasing demand ...



- 92 low/middle income countries are “Data Deprived”
  - Only 1 point: Mainly in Africa
  - > 5-year interval: 77 countries
  - Irregular survey implementation
- Beyond data deprivation, issues with:
  - Uncertainty of funding: many more (IDA) countries “at risk”
  - Data reliability, comparability and accessibility

Source: Serajuddin et al. (2015)

The SDG provide a unique opportunity, but …

… need to go beyond indicators!

=F(

; 

; 

; …)

For evidence-based policy making, need an integrated approach involving …

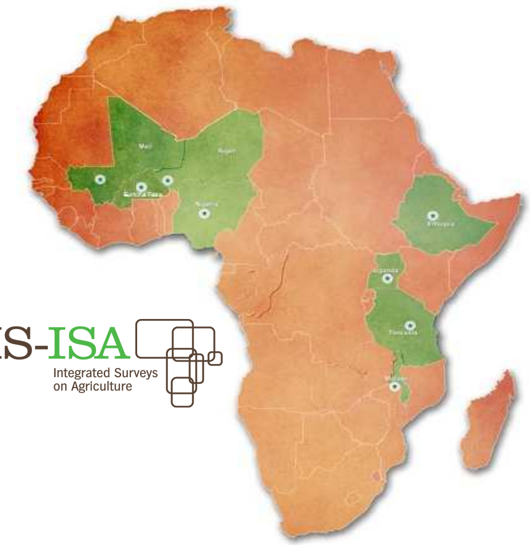
- Integration <u>within</u> same instrument
  - Cost saving
  - Analytical advantages … but also drawbacks!
- Integration <u>across</u> data sources
  - **Data from space**

LSMS–Integrated Surveys on Agriculture (LSMS-ISA)

Technical and financial assistance for the design and implementation of multi-topic panel household surveys, with a focus on agriculture.

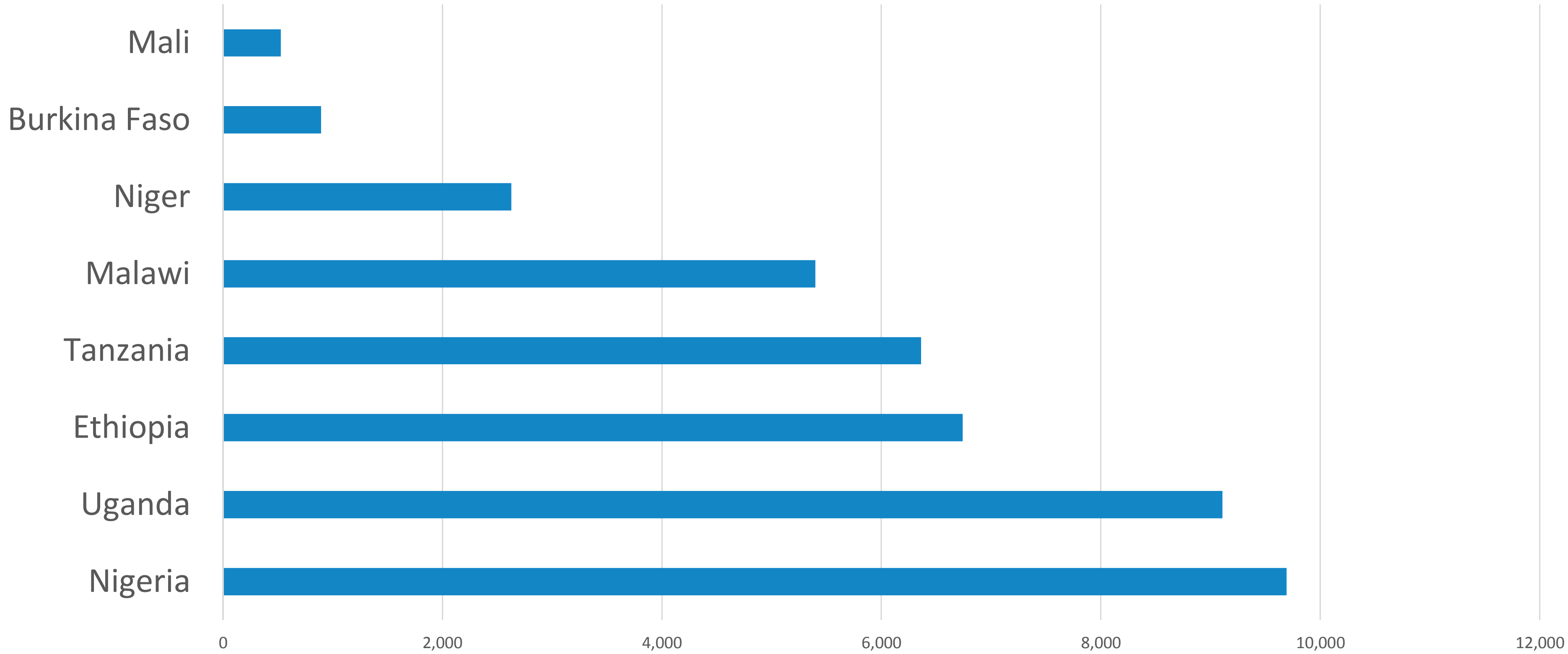
Since 2009, 20+ surveys, which :

- Are integrated into national statistical systems
- Are nationally & regionally representative
- Track households & individuals
- **Geo-reference household & plot locations**
- Collect individual-level data
- Use field-based data processing (CAPI)
- Are open access



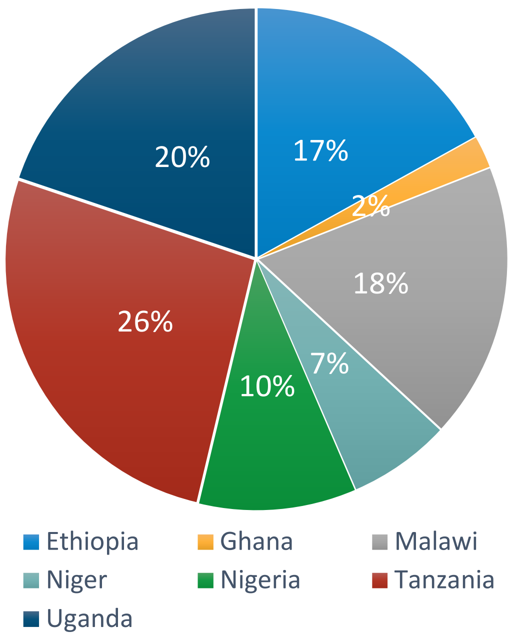
LSMS-ISA Integrated Surveys on Agriculture

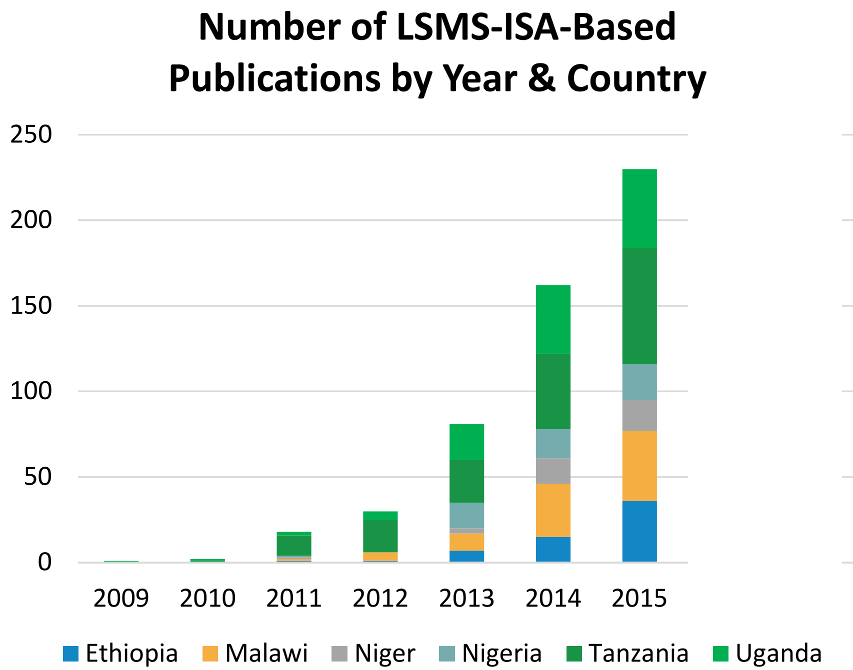
LSMS-ISA Downloads by Country



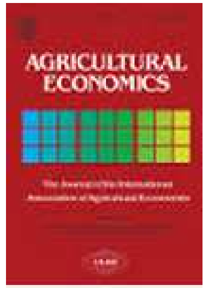
Total of 41,342 for these 8 countries (as of October 24, 2017)
** Lower bound: does not include direct downloads from NSO websites; more than ¾ are downloads of full datasets*

LSMS-ISA Research by Country



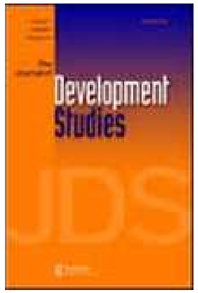


Examples of cross-country research



Gender & Agriculture

- Partners: IFAD, Africa Gender Innovation Lab, IFPRI, FAO
- World Bank Policy Research Working Papers
- World Bank-ONE Campaign Report – Leveling the Field
- *Agricultural Economics* Special Issue



Nutrition & Agriculture

- Partners: BMGF, IFPRI
- World Bank Policy Research Working Papers
- *Journal of Development Studies* Special Issue

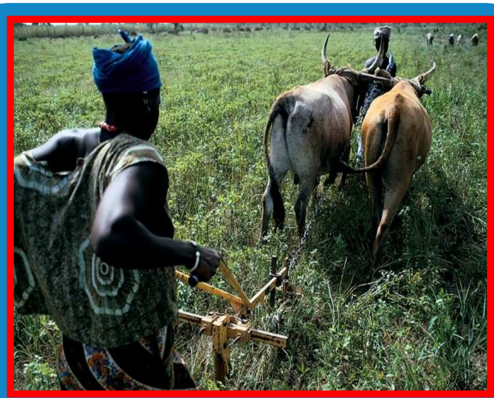
Agriculture in Africa: Telling Facts from Myths

- Partners: AfDB, World Bank Africa CE, Yale, Cornell, Maastricht
- World Bank Policy Research Working Papers
- *Food Policy* Special Issue

Scope of LSMS-ISA Data

Household

- **Dwelling GPS Coordinates**
- Demographics
- Education
- Health
- Housing
- Food & Non-Food Consumption
- Off-Farm Earnings
- Asset Ownership
- Anthropometry
- Food Security
- Safety Nets
- Shocks



Agriculture

- **Plot GPS Coordinates & GPS-Based Area Measurement**
- Parcels : Tenure, Ownership
- Plots: Physical Attributes, Labor & Non-Labor Input Use
- Crops: Cultivation, Production (Plot-Crop-Level), & Disposition (Crop-Level)
- Ag Asset Ownership & Use
- Extension Services
- Livestock Ownership & Production

Community

- Demographics
- Infrastructure
- Facilities
- Access to Services
- Facilities
- Collective Action
- Natural Resource Management
- Community Organizations
- Prices

LSMS-ISA Approach to Disseminating Geospatial Data

- Provide **Randomly Off-Set**, EA-Level Coordinates
  - Average household-level coordinates in a given EA
  - Apply a random offset of 0-2 km in urban, 2-5 km in rural areas
  - Similar to DHS Protocol

- Uses raw GPS coordinates to match household locations with publicly-available **geospatial variables**, disseminated alongside unit-record survey data
  - Depending on characteristics of source data, values may be rounded (distance) or ranged (population density) to maintain anonymity of place

LSMS-ISA Geospatial Variables

Theme	Variable
Distance	Plot distance to household
	Household to nearest main road
	Household to major agricultural market
	Household to headquarters of district of residence
	Household to nearest city or town with +20,000
	Household to nearest border post
Climatology	Annual mean temperature
	Mean temperature of wettest quarter
	Mean annual precipitation
	Precipitation of wettest quarter
	Precipitation of wettest month
Landscape	Land cover class
	Density of agriculture
	Population density
	Agro-ecological zone
Soil & Terrain	Elevation
	Slope

Theme	Variable
Soil & Terrain	Terrain roughness
	Topographic wetness index
	Landscape-level soil characteristics
Rainfall (TS)	Survey year annual rainfall
	Survey year wettest quarter rainfall
	Survey year timing of start of wettest quarter
Phenology (TS)	Average total change in greenness within primary ag season
	Average timing of onset of greenness increase
	Average timing of onset of greenness decrease
	Average EVI value at peak of greenness
	Total change in greenness in survey year
	Timing of onset of greenness increase in survey year
	Timing of onset of greenness decrease in survey year
	Maximum EVI value in survey year
	Specific crop season NDVI crop season aggregates

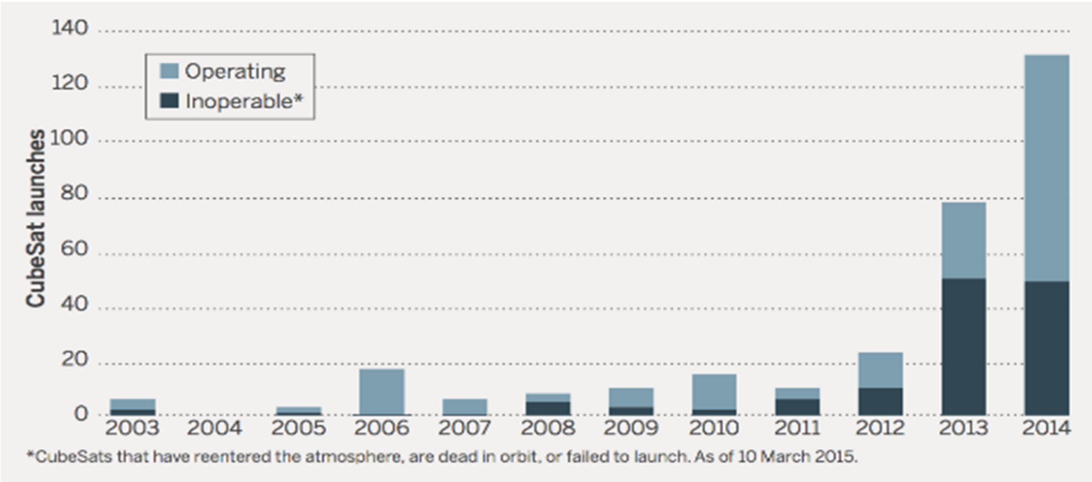
Why are we interested in integrating household survey data with geospatial data?

- At least two reasons…
  1. To study the relationships between farms/households/individuals and the environment

  1. Obtain higher-resolution/more frequent predictions of economic outcomes, at potentially lower costs
     - Today's highlighted applications will be on poverty and crop yields
     - Common thread: Use of household survey data as "ground truth"

The good news: We have more eyes in the sky than ever before!

Sensor	Wavelengths	Spatial Resolution	Revisit Frequency	Launch Year
Sentinel-1	C-band radar	20m	6 day	2014, 2016
Sentinel-2	Optical	10m	5 day	2015, 2017
Skysat	Optical	1m	~Weekly	2013-present
Planet	Optical	3-5m	~Daily	2014-present



Source: Hand, *Science News*, (2015).

Living Standards Measurement Study

POVERTY FROM SPACE

Engstrom, R., Hersh, J., and Newhouse, D. (2017). "**Poverty from space: using high-resolution satellite imagery for estimating economic well-being**." World Bank Policy Research Working Paper No. 8284

Feature-Based Approach

- Engstrom et al. (2017) predict poverty rates based on features derived from high-resolution satellite imagery

1. Generate features from satellite data
    - Convolutional Neural Networks
        - Identify cars, shadows, built-up area
    - Semi-automated classification
        - Identify road width, dirt vs. paved roads, roof type, roof area, simple land classification
    - Texture features from open-source Sp.Feas program
2. Use estimates of poverty and welfare from census-based poverty mapping exercise as "ground truth"
    - 10% and 40% relative poverty rates, and average expected log welfare
3. Regress satellite features on census-based welfare and poverty estimates

60 Percent of Variation in Welfare Explained by Satellite Features

Accuracy of Predictions	10% Poverty Rate	40% Poverty Rate	Average GN Log Expected Welfare
Out of sample R^2	0.59	0.60	0.60
Mean Absolute Error	3.2 pp	7.8 pp	0.139
Observations	1291	1291	1291

- Building density, roof type, and shadows are strongest predictors
  - In rural (urban) areas, poor areas have more (less) vegetation
- "Texture features" alone explain 40 to 50 percent of variation

Predictions Remain Accurate When Using Small Sample to Train Model

Out of sample R2	10% Poverty Rate	40% Poverty Rate	Average GN log expected welfare
Full sample	0.59	0.60	0.60
Small sample	0.53	0.59	0.58

- Use case is pairing imagery with a survey, not census
- Drew 1 percent synthetic sample from census
  - Comparable in size to HIES household survey
- Minor loss of performance when using 1 percent subsample

Living Standards Measurement Study

CROP YIELDS FROM SPACE

Preliminary Findings from: "**Eyes in the Sky, Boots on the Ground: Assessing Satellite- and Ground-based Approaches to Crop Yield Measurement and Analysis in Uganda**" (*Forthcoming)* – **DO NOT CITE**

Joint w/: David B. Lobell, George Azzari, Marshall Burke, Sydney Gourlay, Zhenong Jin, and Siobhan Murray

Objectives

- To test subjective approaches to measurement vis-à-vis objective methods for maize yield measurement, soil fertility assessment & maize variety identification
- To assess potential of using remote sensing for estimating crop yields

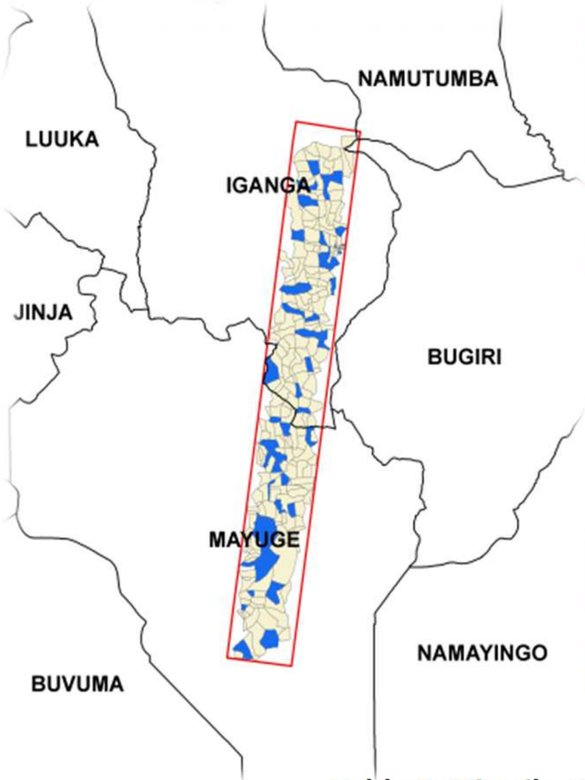
Methods

Methods Tested:	

Maize Production	• **Crop-cutting**
	  • 4m x 4m & a 2m x 2m subplot in Round I
	  • 8m x 8m sub-plot in Round II
	  • Full-plot crop cut in Round II (1/2 of sample)
	• **Remote sensing** based on high-res imagery
	  • First in testing the method in a smallholder production system against an objective measure
	• **Self-reported** harvest
	  • Conversion of quantities in non-standard unit-condition combos into KG-, dried grain terms ("official" methods)

Land Area	• **GPS measurement** (Garmin eTrex 30 handheld units) • **Self-reported** area
Soil Fertility (Round I)	• **Conventional** Soil Analysis (subsample) • **Spectral** Soil Analysis • **Self-reported** soil quality & attributes
Variety Identification	• **DNA fingerprinting** of grain sampled from the crop-cutting subplot harvest (4x4m in Round I, 8x8m in Round II) • **Self-reported** variety name, type & morphological attributes

Study Area & Sentinel-2 Imagery



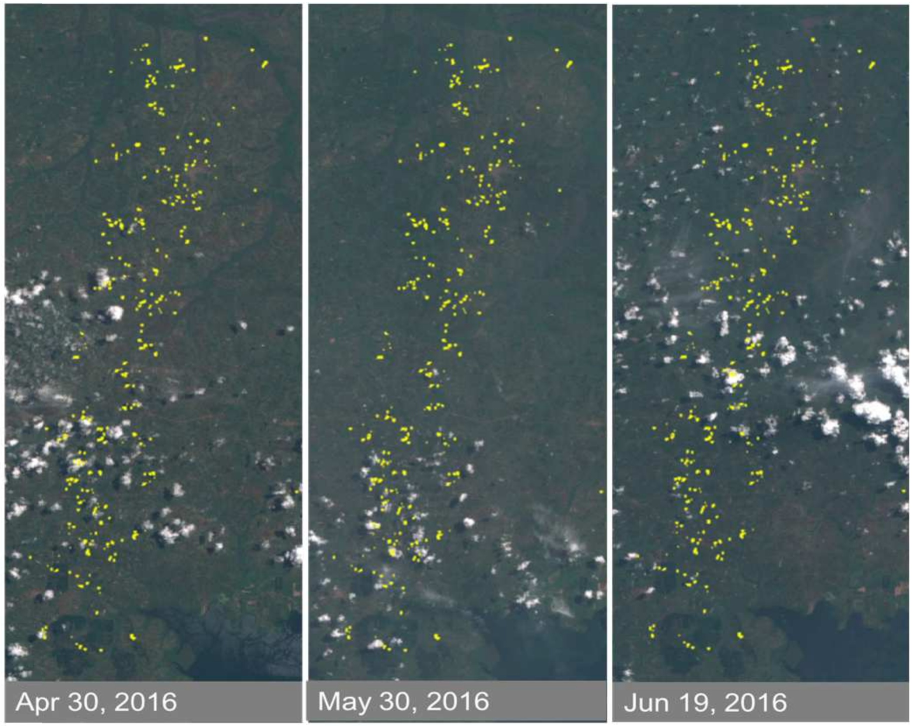


Table 1. Distribution of MAPS II Plots by Cultivation Status

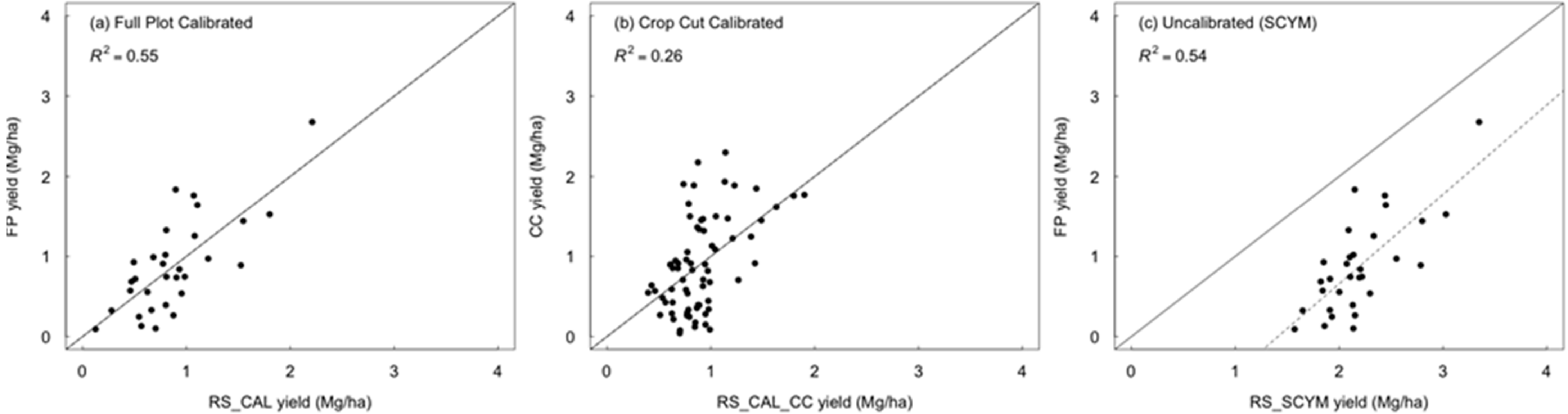
Purestand	Intercropped			
	Maize-Legume	Maize-Cassava	Maize-Legume-Cassava	Maize-Other
124	119	161	52	7

Vegetation Indices (Vis)

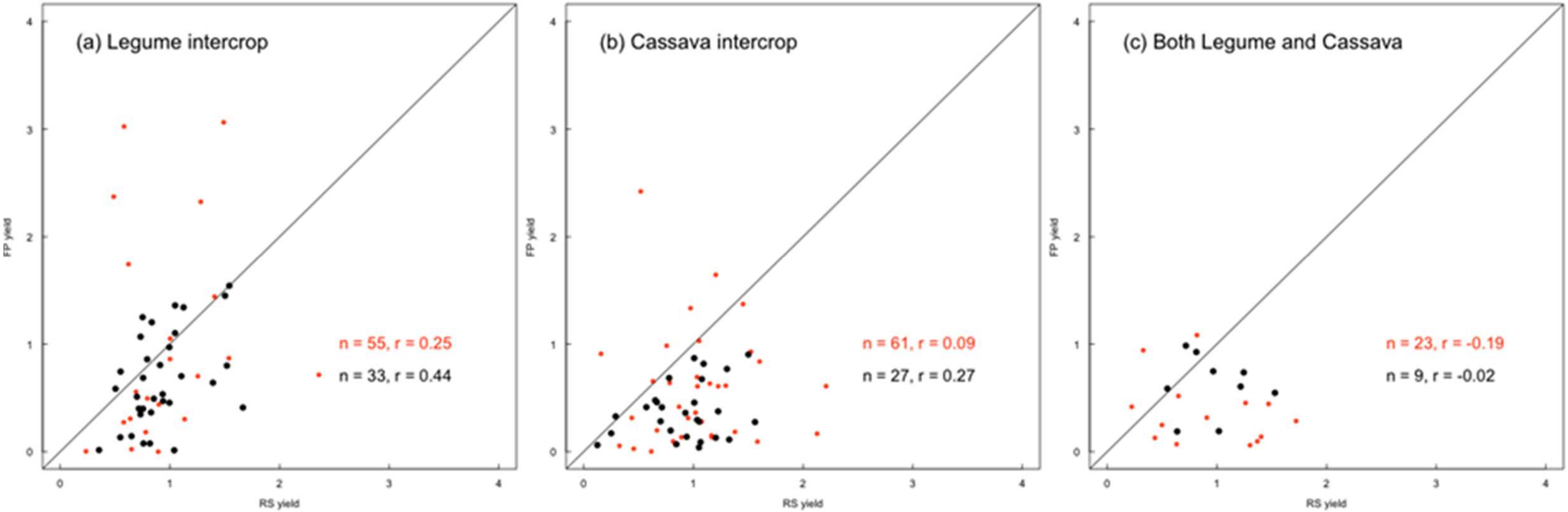
Table 2. Spectral Vegetation indices (VIs) Used

Name	Equation	Equation using Sentinel-2 bands	Reference
NDVI (Normalized Difference Vegetation Index)	$(R_{NIR} - R_{RED}) / (R_{NIR} + R_{RED})$	(B8 – B4) / (B8 + B4)	(Rouse et al., 1973)
GCVI (Green Chlorophyll Vegetation Index)	$(R_{NIR} / R_{GREEN}) - 1$	(B8/B3) - 1	(Gitelson et al., 2003)
MTCI (MERIS Terrestrial Chlorophyll Index)	$(R_{NIR} - R_{705}) / (R_{705} - R_{RED})$	(B8-B5) / (B5 – B4)	(Dash and Curran, 2004)
NDVI705 (Red-Edge $NDVI_{705}$)	$(R_{NIR} - R_{705}) / (R_{NIR} + R_{705})$	(B8 – B5) / (B8 + B5)	(Viña and Gitelson, 2005)
NDVI740 (Red-Edge $NDVI_{740}$)	$(R_{NIR} - R_{740}) / (R_{NIR} + R_{740})$	(B8 – B6) / (B8 + B6)	(Viña and Gitelson, 2005)

Remotely-Sensed vs. Ground-Based Yields on Purestand Plots



MAPS II Remote Sensing Performance on Intercropped Plots



Key Takeaways

On poverty…

- Tough to get R^2 above 0.5 when predicting welfare or poverty at large scale using integrated household survey and geospatial data applications
- Need to better understand strengths and weaknesses of different methods in different contexts
  - Feature-based approach by Engstrom et al. (2017) vs. Transfer learning approach by Jean et al. (2016) in Africa (Not reviewed here, uses LSMS-ISA data)
- Are satellite predictions better than census-based poverty maps? Not sure yet…

On agriculture…

- Promising performance of public-use, high-frequency 10m resolution imagery and remote sensing techniques in predicting maize yields at plot-level
- Importance of calibration using survey data

Some final thoughts

- Great potential for value addition of using data from space to inform policy research
- With increased availability of free/inexpensive spatial data, it will only get better!
- Household surveys remain indispensable source, including for validation
- Need to work on a global methodological research agenda to make value addition of integration more reliable and scalable
- Access to household-level geo-referenced data still problematic

Households from Space

Integrating Household Surveys with Geospatial Data Sources for Improved Monitoring of Development Outcomes

Talip Kilic
Senior Economist
Living Standards Measurement Study (LSMS)
Development Data Group The World Bank

Applying Quantitative Analysis to Development Issues Conference
Bibliotheca Alexandrina | Alexandria, Egypt | February 18-19, 2018

LSMS
Living Standards Measurement Study