WPS8374

Policy Research Working Paper 8374

Eyes in the Sky, Boots on the Ground

Assessing Satellite- and Ground-Based Approaches to Crop Yield Measurement and Analysis in Uganda

David B. Lobell
George Azzari
Marshall Burke
Sydney Gourlay
Zhenong Jin
Talip Kilic
Siobhan Murray

WORLD BANK GROUP
Development Economics
Development Data Group
March 2018

Policy Research Working Paper 8374

Abstract

Crop yields in smallholder systems are traditionally assessed using farmer-reported information in surveys, occasionally by crop cuts for a sub-section of a farmer's plot, and rarely using full-plot harvests. Accuracy and cost vary dramatically across methods. In parallel, satellite data is improving in terms of spatial, temporal, and spectral resolution needed to discern performance on smallholder plots. This study uses data from a survey experiment in Uganda, and evaluates the accuracy of Sentinel-2 imagery-based, remotely-sensed plot-level maize yields with respect to ground-based measures relying on farmer self-reporting, sub-plot crop cutting (CC), and full-plot crop cutting (FP). Remotely-sensed yields include two versions calibrated to FP and CC yields (calibrated), and an alternative based on crop model simulations, using no ground data (uncalibrated). On the ground, self-reported yields explained less than 1 percent of FP (and CC) yield variability, and while the average difference between CC and FP yields was not significant, CC yields captured one-quarter of FP yield variability. With satellite data, both calibrated and uncalibrated yields captured FP yield variability on pure stand plots similarly well, and both captured half of FP yield variability on pure stand plots above 0.10 hectare. The uncalibrated yields were consistently 1 ton per hectare higher than FP or CC yields, and the satellite-based yields were less well correlated with the ground-based measures on intercropped plots compared with pure stand ones. Importantly, regressions using CC, FP and remotely-sensed yields as dependent variables all produced very similar coefficients for yield response to production factors.

This paper is a product of the Development Data Group, Development Economics. 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 tkilic@worldbank.org.

Eyes in the Sky, Boots on the Ground

Assessing Satellite- and Ground-Based Approaches to Crop Yield Measurement and Analysis in Uganda[1]

David B. Lobell[a], George Azzari[b], Marshall Burke[c],
Sydney Gourlay[d], Zhenong Jin[e], Talip Kilic[f], and Siobhan Murray[g]

JEL Codes: C83, Q12.
Keywords: Maize, Crop Yield Estimation, Crop Cutting, Remote Sensing, Uganda.

[1] Following the lead author, the co-authors are listed alphabetically. [a] Corresponding author. Professor, Department of Earth System Science, and the Center on Food Security and the Environment (FSE), Stanford University, Stanford, CA. dlobell@stanford.edu. [b] Research Associate, Department of Earth System Science and the FSE, Stanford University, Stanford, CA. gazzari@stanford.edu. [c] Assistant Professor, Department of Earth System Science and the FSE, Stanford University, Stanford, CA. mburke@stanford.edu. [d] Survey Specialist, Living Standards Measurement Study (LSMS), Development Data Group (DECDG), Rome, Italy. sgourlay@worldbank.org. [e] Postdoctoral Scholar, Department of Earth System Science and the FSE, Stanford University, Stanford, CA. jinzn@stanford.edu. [f] Senior Economist, LSMS, DECDG, World Bank, Rome, Italy. tkilic@worldbank.org. [g] Technical Specialist, LSMS, DECDG, World Bank, Washington, DC. smurray@worldbank.org. The lead principal investigators were Lobell and Kilic on the Stanford University and the World Bank front, respectively.

1. Introduction

Improving the productivity of smallholder farmers is widely considered one of the most effective avenues for reducing poverty and food insecurity, and thus has been a longstanding goal in many African countries (Byerlee et al., 2007). The evidence concerning (i) agriculture contributing up to 69 percent of rural household income in Africa (Davis et al., 2017), and (ii) higher rates of expected poverty reduction associated with agricultural vis-à-vis nonagricultural growth (Dorosh and Thurlow, 2016) helps sustain the policy focus on achieving this goal at the national level. Similarly, at the international level, doubling productivity and incomes of smallholders have been identified as a key target within the Sustainable Development Goal (SDG) #2 of Ending Hunger.

Accurate measures of production and productivity are, therefore, essential to (i) tracking progress towards the relevant SDG targets; answering fundamental questions on the role of agriculture in household and individual welfare (Darko et al., 2018); and understanding which production factors have the most important role in determining productivity. Ongoing debates about the relationship between (land) productivity and (i) fertilizer use (Harou et al., 2017), (ii) plot/farm size (Bevis and Barrett, 2017; Desiere and Jolliffe, 2018; Gourlay et al., 2017), or (iii) soil quality (Berazneva et al., 2018) reflect the substantial knowledge gaps and the need to improve the accuracy and precision of recommendations for raising productivity in particular locations.

The most common way to assess outcomes related to economic productivity of smallholder farmers, including land productivity (e.g. crop yields), is by using information collected through in-person interviews for household and farm surveys. For example, the household surveys supported by the World Bank Living Standards Measurement Study – Integrated Surveys on Agriculture (LSMS-ISA) initiative measure plot areas with handheld GPS units and solicit farmer-reported information on crop production and input use, among other topics, at the plot level. This information, together with multi-topic household survey data, have informed a burgeoning field of development research on Africa over the last decade.

Compared to the body of methodological research that has shown severe systematic biases in farmer-reported plot area measures (Carletto et al., Forthcoming) and that have underlined the increasing use of GPS-based plot area measurement in national household surveys, there is a dearth of evidence on the accuracy of farmer-reported crop production. It is, however, known that the process of soliciting farmer-reported production information is mediated by complexities that include (i) the use of non-standard measurement units, (ii) various conditions and states of crop harvest; (iii) partial/early crop harvests; (iv) potential recall bias, and (v) tendency to round off numbers, among others (Carletto et al., 2015). Recent research highlighted the measurement errors in self-reported crop production estimates and their implications for the inverse scale-productivity relationship (Desiere and Jolliffe, 2018 in Ethiopia; Gourlay et al., 2017 in Uganda).

Less common but also well-established is to measure crop yields by physically harvesting a sub-section of a farmer's plot in a so-called crop cut (Fermont and Benson, 2011). Crop cuts provide a more objective way to measure grain production for a part of the plots, but heterogeneity within a plot can lead to sensitivities of crop cut yields to the precise location and size of the crop cut sub-plot vis-à-vis the entire plot (Fermont and Benson, 2011). An alternative is to harvest the entire plot, which avoids most of the problems of the prior methods, and therefore, is frequently considered the "gold standard" (Casley and Kumar, 1988; Fermont and Benson, 2011). However, full plot harvests require a substantial amount of labor and coordination with farmer harvest schedules, which makes them costly and difficult to scale.

Given the limitations of existing approaches, recent work has explored the ability of satellite data to track crop yields. Burke and Lobell (2017) (hereafter BL17) showed that 1m resolution data from Terra Bella's Skysat sensors were useful for mapping maize yields for farms in western Kenya. This usefulness was measured both by correlation of satellite-based yield estimates with traditional ground-based yield measures, as well as by the ability of satellite-based yields to detect positive yield responses to fertilizer and hybrid seed inputs. This latter aspect was considered especially important since (i) ground-based yield measures are inevitably imperfect themselves, and (ii) detecting response to inputs or some other aspect of farm management is a common motivation for collecting plot-level yield data in the first place.

The primary objectives of this paper are to assess the ability of satellite-based approaches to measure plot-level maize yields on African smallholder farms and to gauge the sensitivity of production function estimation to the choice of ground- versus satellite-based maize yield variant. The analysis uses data from the 2016 round of MAPS: Methodological Experiment on Measuring Maize Productivity, Soil Fertility and Variety, which was implemented during the first rainy season of 2016 (June-October) in 45 enumeration areas within a 400 square kilometer area spanning Iganga and Mayuge districts of Eastern Uganda; the leading maize-producing region of the country. The analysis extends the work presented in BL17 in at least three substantial ways.

First, the Ugandan maize systems are considerably more subsistence-focused and heterogeneous than the Kenyan counterparts in BL17, with generally smaller plot sizes, lower input use, and greater prevalence of under-canopy intercrops such as beans and groundnuts, and frequent occurrence of over-canopy intercrops such as cassava and bananas. Thus, Uganda represents a different and, in many ways, more challenging environment in which to test satellite-based crop yield measurement approaches.

Second, whereas BL17 relied on farmer self-reported data on maize production, this paper uses objective measures based on survey field team harvests of maize grain for 64 m^2 subplots within each plot ("crop cuts"), as well as whole plot harvests for approximately 1 random half of our sample ("full plot harvests"). Thus, we are able to compare different ground-based measures with each other, and with the satellite data.

Third, the study uses data from the Copernicus program's Sentinel-2A satellite, which has coarser spatial resolution but more spectral bands than the Skysat sensor used in BL17. Furthermore, whereas Skysat data are currently only available for selected locations, Sentinel-2A and its recently launched sister satellite Sentinel-2B each capture imagery every 10 days for the entire land surface of the Earth, with an effective 5-day repeat for the Sentinel-2 duo since June 2017. These data are quickly made available to the public at no cost. For these reasons, Sentinel-2 represents an attractive option for estimating yields over large regions.[2]

All plot-level measures of maize yield, including farmer-reported self-reported production per hectare (SR), sub-plot crop cut production per hectare (CC), full plot crop production per hectare (FP), and variants of remotely sensed production per hectare (RS), rely on GPS-based plot areas; are compared to each other using standard statistical approaches; and are used to study the sensitivity of the associations between maize yield and various production factors measured through a combination of a household survey and extensive soil sampling.

The paper is organized as follows. Section 2 describes the data. Section 3 presents the comparisons among ground-based yield measures; between ground- and satellite-based yield measures; and the results from the estimations of maize yield regressions for each yield variant of interest. Section 4 concludes.

2. Data

MAPS: Methodological Experiment on Measuring Maize Productivity, Soil Fertility and Variety is a two-round household panel survey that was conducted in Eastern Uganda to test the relative accuracy of subjective approaches to data collection vis-à-vis objective survey methods for maize yield measurement, soil fertility assessment, and maize variety identification. Both survey rounds were implemented by the Uganda Bureau of Statistics, with technical and financial assistance

[2] BL17 focused on field campaigns in 2014 and 2015, before Sentinel-2 was operational.

provided by an inter-agency partnership that was led by the World Bank Living Standards Measurement Study (LSMS).

2.1. Sampling Design

In Round I, the MAPS fieldwork was conducted during the first rainy season of 2015, from April to October 2015, in Eastern Uganda, the top maize-producing region of the country. A sample of 75 enumeration areas (EAs) were selected from the 2014 Population and Household Census (PHC) EA frame, with probability proportional to EA-level household counts. The sampled EAs were distributed across 3 strata, namely (1) Sironko district (15 EAs), (2) Serere district (15 EAs), and (3) a 400 square kilometer remote sensing tasking area spanning Iganga and Mayuge districts (45 EAs).

In each sampled EA, the original intention had been to select, at random, 6 households from each of the pure stand and intercropped universes of households of an EA, and ensure an even sample split by maize cultivation status. Within the remote sensing tasking area of interest, the MAPS I fieldwork started out with 540 households, of which 249 were pure stand (46 percent) and 291 (54 percent) were intercropped.[3] In each MAPS household, 1 maize plot, matching the household cultivation status, was selected at random by the Survey Solutions CAPI application for crop cutting, soil sampling, and variety identification components.

In MAPS II, the fieldwork was conducted during the first rainy season of 2016, from June to October 2016. The field teams attempted to track and re-interview 540 households that had been interviewed in 2015 within the 400 square kilometer remote sensing tasking area. Figure 1 provides an overview of the study region in MAPS II. Overall, 489 of the 540 households were successfully re-interviewed.[4] As in MAPS I, 1 maize plot was selected from each household for crop cutting and variety identification components. Whenever possible, the plot was selected among those that were matching the household cultivation status in MAPS I. Preference was given such that the plot would be selected from the same parcel that had contained the plot selected in Round I. If multiple plots matched the household cultivation status, the CAPI application selected one plot at random.[5]

[3] The uneven split by cultivation status was due to the low incidence of pure stand households, and the cases in which pure stand households would switch to intercropping status between the household listing and the first interview.

[4] 34 out of 51 households that we did not interview in MAPS II were due to the fact that they were not cultivating maize in the first season of 2016. The remaining 17 households can be broken down as follows: 5 households could not be tracked or were outside of the tracking area defined as the Iganga and Mayuge districts (5); 4 households had suffered total crop loss prior to post-planting interview; 7 households had already harvested their maize by the post-planting interview; and 1 household refused. Gourlay et al. (2017) report that attrition bias is not a concern.

[5] Refer to Gourlay et al. (2017) for MAPS II household tracking and plot selection protocols.

The MAPS I remote sensing findings were reported first by Gourlay et al. (2017). MAPS II implemented full-plot crop cutting for a random sub-sample of plots, and increased, on each plot, the area for sub-plot crop cutting (from 4x4m to 8x8m). These decisions were anchored in the concerns around intra-plot variability of maize yields. Given the enhancements in the scope of crop cutting data in MAPS II and the interest in the validation of satellite-based approaches to yield estimation, we rely solely on the MAPS II data on 463 households/plots for which sub-plot crop cutting data are available. The only exception, as explained below, is the plot-level data on soil fertility, which is sourced from MAPS I. Table 1 provides a breakdown of 463 plots in accordance with pure stand versus (type of) intercropped cultivation status.

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

2.2. Fieldwork

Three visits were made to each household during MAPS II. During the (first) post-planting visit, the questionnaire modules included those soliciting information on (1) demographic and socio-economic attributes of household members; (2) household dwelling characteristics and ownership of durable assets and agricultural implements; and (3) area, cultivation pattern, management, pre-harvest labor and seed inputs for all maize plots that were cultivated during the reference rainy season.[6] Following the completion of the household post-planting interview, each enumerator visited the maize plot that was selected in accordance with the protocol detailed in the previous section. At that time, he/she measured the plot area and saved its boundaries on a Garmin eTrex 30 handheld GPS device, and set up the crop cut sub-plot for later harvesting and weighing. The crop cut sub-plot location was chosen at random, in accordance with the protocol that is detailed by Gourlay et al. (2017) and in line with the international best practices.

During the (second) crop cutting visit, the enumerator harvested the crop cut sub-plots to obtain objectively measured harvest quantities, as detailed in the subsequent section. Finally, during

[6] A parcel is conceptualized as a continuous piece of land under a common tenure system, while a plot is defined as a continuous piece of land on which a unique crop or a mixture of crops is grown, under a uniform, consistent crop management system, not split by a path of more than one meter in width, and with boundaries defined in accordance with the crops grown and the operator. Therefore, a parcel can be made up of one or more plots. This distinction is key since for the purposes of within-farm analysis of agricultural productivity, the ideal is to capture within-parcel, plot area measurements linked with plot-level measurement of agricultural production

the (third) post-harvest visit, farmer-reported information on total plot-specific maize production, non-labor inputs and harvest labor inputs was solicited for all maize plots that were cultivated during the reference season. The post-harvest visit was scheduled within a 2-month period following the completion of each household's harvest.

2.3. Key Measurement Domains and Methods

2.3.1. Plot Area Measurement

After walking the perimeter of a given plot with the plot manager to identify the boundaries, the enumerators re-paced the perimeter and measured the area with a Garmin eTrex 30 handheld GPS device. The area was recorded on the questionnaire in square meters, and the raw GPS track outline was stored. The competing yield measures in our study are all anchored in GPS-based plot area measurement. In MAPS II, the median plot size was 0.11 hectare (ha) (roughly one-quarter of an acre), with 46 percent below 0.10 ha and 17 percent below 0.05 ha.

2.3.2. Soil Fertility Assessment

Gourlay et al. (2017) provides details on the collection of soil samples at each plot location in MAPS I. The soil sample collection was not repeated in MAPS II partly due to budget constraints and partly due to the MAPS II preference for the plots that were on the same parcels that had a plot selected in MAPS I, as explained by Gourlay et al. (2017). In MAPS I, four samples of the topsoil (0-20cm) were collected at random locations within each plot and were combined into one composite sample. A single deeper (sub-soil) sample (20-50cm) was collected from the plot center. All samples were shipped to the World Agroforestry Center (ICRAF) Nairobi office, and were subject to spectral soil analysis, with 10 percent of the top- and sub-soil samples also analyzed with conventional wet chemistry testing. The key soil attributes that were measured include pH, texture analysis (sand, % clay, % silt), cation exchange capacity, electrical conductivity (EC), and the concentration of organic carbon (OC), total nitrogen (TN), and potassium.

Following Mukherjee and Lal (2014), a composite soil quality index (SQI) was calculated for each MAPS I plot. Multiple approaches to index construction were employed, including simple additive and weighted additive approaches, as well as a principal component approach and each were computed using topsoil (0-20cm) and subsoil (20-50cm) depths. Bivariate analysis of each index and crop cutting yield estimates (not reported) suggested that the principal component method using top-soil properties was found to correlate more strongly with yield than other approaches, and thus, this index is used. Numerous versions of the principal component-based soil quality index were constructed, using different combinations of soil properties. In this approach, principal component analysis (PCA) was first conducted and components with eigenvalues greater than or equal to 1 were retained. Then, the most important variables in each component were identified, including all variables within 10% of the weight of the most important, if the correlation with the most important variable was less than or equal to 0.6. When two or more properties were retained from the same component (where they are weakly correlated and within 10% of the highest weighted property), each property received the same weight.

The index with the greatest predictive power with respect to crop cut yield was composed of organic carbon (%), soil electrical conductivity (an indicator of soil salinity), and pH. These variable values were transformed to a range from 0 to 1, where 1 represents the most optimal value in the sample (e.g., highest value for OC, intermediate values for pH), and 0 represents the lowest value in the sample. A composite index was then generated by weighting each variable by the fraction of total variance explained by its corresponding component. The relative weights for organic carbon, soil electrical conductivity, and pH are 68.3, 68.3, 31.7, respectively.[7] Given data limitations, the constructed index focuses on nutrient storage capacity but ignores the other two components of soil quality identified by Mukherjee and Lal (2014) related to root development and water storage.[8]

Although these soil samples were acquired in MAPS I, they still provide a useful measure of soil quality to compare with the various yield measures. Importantly, the selected maize plot for most households (n = 312) was part of the same parcel as in the previous year, so that the soil sample was from the same part of their farm. Concerning the remaining sample of households that had a MAPS II plot selected from a non-MAPS I parcel, the median distance between the MAPS II and the MAPS I plot locations was 0.56 kilometers, lending support to likely similarity in soil profiles of nearby plot locations. More importantly, the regression results using soil quality showed very little sensitivity to excluding those households where the parcel moved between years.

2.3.3. Ground-Based Maize Yield Measurement

2.3.3.1. Farmer Estimation

Plot managers were asked to report their estimate of maize harvest at the parcel-plot-level during the post-harvest visit, replicating the design of the Uganda National Panel Survey (UNPS)

[7] Organic carbon and soil electrical conductivity were both retained from the first component and, therefore, hold the same weight.

[8] The PCA-based soil quality index was constructed for the full MAPS 1 sample, and therefore, analyzes the correlation of soil properties and crop cutting yields on a larger sample than MAPS 2.

questionnaire modules.[9] Each plot manager was allowed to report production in non-standard measurement units, and was asked to report on both the condition (e.g. green harvested; dry after additional drying; etc.) and the state (e.g. with cob but without stalk or husk; grain; etc.) of up to three maize harvests that may have occurred on the plot over a period of time. The production measurement units, conditions, and states were borrowed directly from the UNPS, as also provided by Gourlay et al. (2017). The dry grain-equivalent harvest quantities in kilograms were calculated by using the conversion factor database that has been compiled by the UBOS during the 2007 Uganda Census of Agriculture (UCA) for each non-standard measurement unit-condition-state combination and that has been complemented by the data solicited during the UNPS 2009/10, 2010/11, and 2011/12 waves for the (rare) combinations that were not captured as part of the UCA exercise.[10]

2.3.3.2. Crop Cutting

Crop cutting has been recognized as the gold standard for yield measurement since the 1950s by the Food and Agriculture Organization of the United Nations (FAO). Gourlay et al. (2017) review the potential concerns regarding yield measurement concerning crop cutting and detail the way in which the MAPS approach to crop cutting and its hands-on supervision overcame them.

In MAPS II, one 8x8 meter sub-plot (divided into four 4x4m quadrants) was laid on each plot. Each subplot was cordoned off until harvest and was supervised by the EA-specific crop cut monitor between the post-planting and the crop cutting visits. Each plot manager was asked not to harvest any crop from the sub-plots until the crop cutting visit, and not to manage the sub-plot any differently than the rest of the plot. These messages, first communicated by the enumerator, were intended to be enforced by the local crop cut monitors. The shelled maize harvests tied to each of the four adjacent 4x4m quadrants were weighed and barcoded separately in the field and were reweighed at a central location in Kampala under strict supervision following additional drying. At the time of the final weighing, the moisture content of each sample was captured as to standardize all crop cut sample weights used for our analyses at 12 percent moisture. The

[9] It is important to note that the identification of parcels versus plots within parcels was anchored in the precise definitions that have been referenced above and that have been in effect since the UNPS 2009/10 wave. The operationalization of these definitions is such that each enumerator, prior to the administration of the post-planting questionnaire, has a detailed discussion with the holder regarding the organization of his/her farm. This conversation (1) ensures that the enumerator and the farmer are on the same page regarding what parcels versus plots within parcels mean, (2) often culminates in sketches of different parcels and plots within parcels that are being cultivated during that reference season, and (3) establishes how parcels and plots within parcels will be rostered in the questionnaire instrument. The established parcels and plots within parcels are then reviewed at each subsequent visit to the household.

[10] Refer to Gourlay et al. (2017) for more information regarding the conversion factors used in expressing farmer-reported production information in kilogram-equivalent terms.

MAPS II sub-plot crop cutting based plot-level maize production estimates are computed by multiplying the crop cut sub-plot production across the $64m^2$ area covered by the 8x8m subplot by the ratio of the entire GPS-based plot area in m^2 to $64m^2$.

Furthermore, half of the target household population within each of the pure stand and intercropped domain in each EA was selected at random prior to the start of the MAPS II fieldwork for a full-plot crop cut. This rare approach to crop production measurement entailed the harvesting of the entire plot area, shelling the resulting harvest, weighing it in the field, and capturing its moisture level. This operation was conducted by the enumerators with help from the EA-specific crop cut monitor and the crop cut assistant(s) recruited from within the households. On the MAPS II plots selected for full-plot harvest, the harvest of the designated 8x8m subplot was weighed separately from the full-plot harvest to allow for comparative yield analysis. The full-plot harvests were only weighed in the EAs as their transport to and additional drying and reweighing at a central location was deemed logistically infeasible. Moisture readings taken from the maize grain harvested from the full plot harvests were used to standardize the production quantity to 12 percent moisture. A total of 211 plots had full-plot harvests. Gourlay et al. (2017) detail the approach to full plot harvests. Although farmers were not told the final weight of their harvest, it is likely that the process of harvesting and bagging the maize improved their self-report production values compared to plots without full plot harvests. Therefore, the analyses that use self-reported maize production per hectare rely only on 252 plots without a full plot harvest.

2.3.4. Satellite-Based Maize Yield Measurement

Images from the Sentinel-2A Multispectral Instrument, processed to top-of-atmosphere reflectance (Level -1C) were accessed within the Google Earth Engine platform. Clouds and shadows were masked from the images using a random forest classifier trained on points visually selected from images throughout the region. Five vegetation indices (VIs) were then calculated for each pixel using the equations shown in Table 2. The average value of all bands and VIs within each plot polygon were then extracted for image date for further analysis. In addition, for comparison with the Sentinel-2A images, an image acquired by Terra Bella's Skysat sensor on May 29, 2016 was accessed. Skysat measures radiance in blue, green, red, and near-infrared channels at 1m resolution. As with the Sentinel-2 data, clouds and shadows were masked using a random forest classifier trained on several images in the region, including those used in BL17.

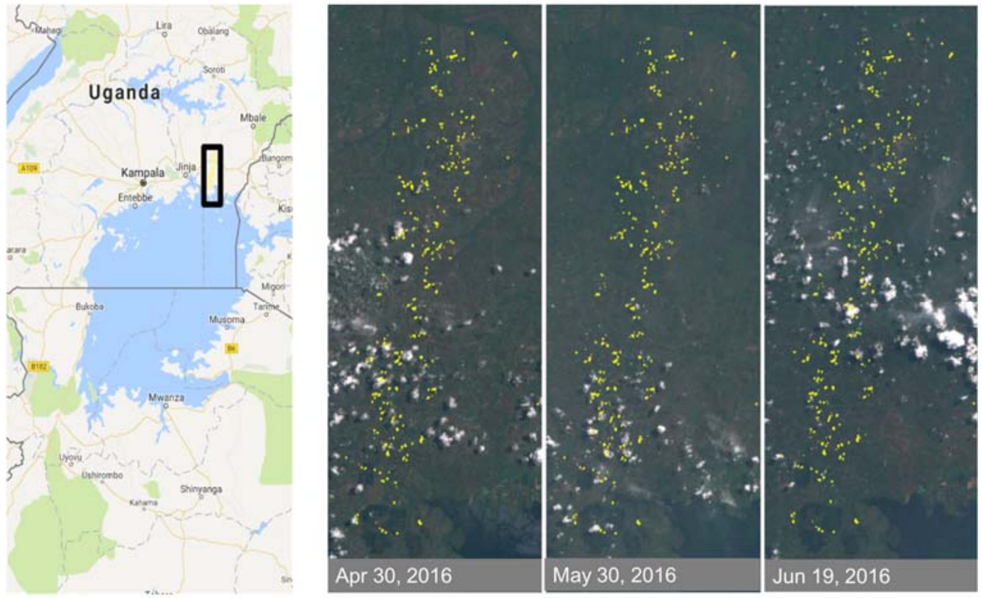

Figure 1. Study region in Eastern Uganda. Three images show Sentinel-2 images and dates used in the study. Yellow polygons indicate outlines of plots where surveys/crop cuts were performed.

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)

2.3.5. Methods

Ground-based SR and FP yields were derived by dividing the reported or measured mass of maize production by the area corresponding to the GPS-based plot area, or 64 m^2, in the case of the 8x8m crop cut sub-plot. Satellite-based yields were derived in two ways, following BL17.

First, "calibrated" remote sensing yields (**RS_cal**) were from a regression model of FP yields on MERIS Terrestrial Chlorophyll Index (MTCI) values on May 30 and June 19, 2016, using only pure stand maize plots that were at least 0.1 ha in size. The calibration focused on the pure stand plots since ground-based objective yield estimates were not available for non-maize crops on intercropped plots. The restriction in terms of plot area was driven by smaller plots having bigger problems with geolocation accuracies and mixed pixels in Sentinel-2. Since FP yields are expensive to obtain and cannot be considered as part of large-scale operations, an alternative version of the calibrated remote sensing yield was obtained (**RS_cal_cc**), which used CC, rather than FP, yields to calibrate the model.

The second satellite-based approach was to estimate "uncalibrated" yields (**RS_scym**) by using the scalable crop yield mapper approach (Lobell et al., 2015). In this approach, a crop model and local daily weather data were used to simulate crop growth and yield for various realistic combinations of on-farm management, such as sow date, seeding density, and fertilizer rate. The simulated values of total canopy nitrogen on the dates with available images were then translated into MTCI using published relationships (Schlemmer et al., 2013). As in the calibrated approach, the yields are then regressed on MTCI, except in the case of SCYM the regression uses simulated yield and MTCI rather than actual values. In this way, SCYM avoids reliance on any ground data for calibration, which is why it is referred to as an "uncalibrated" approach.

Both types of satellite-based yield estimates were tested in two complementary ways. First, the yields were compared directly with the ground-based estimates across both pure stand and intercropped plots. However, given that ground-based estimates are subject to (different types of) measurement error and neglect a potentially substantial amount of production from non-maize crops, the direct comparisons between the two yield measures is not a straightforward test of the satellite-based yields. That is, some of the discrepancy will also be due to errors in the ground-based estimates, or discrepancies in the types of outputs that are measured. As a second form of evaluation, we performed regressions of yield on different production factors for both ground-based and satellite-based yields and compared the resulting coefficients. Specifically, we regressed yields on key plot characteristics, including log of plot area, log of distance to household (km), presence of cover crops, log of seed planted (kg), use of inorganic fertilizer, log of household labor days and hired labor days, number of hired laborers, soil quality index (SQI), and household attributes, including wealth index, agricultural asset index, dependency ratio, household size, head of household age, gender, and years of education, and whether the manager was the survey respondent. For regressions including intercropped plots, two additional variables were included: a binary variable indicating the presence of an intercrop, and a variable indicating the log of the intercrop seed rate (i.e. the ratio of quantity of seed planted to quantity of seed that the farmer estimates would have been planted if plot was pure stand).

3. Results

3.1. Comparison of Ground-Based Yield Measures

The distributions of yields from the three ground-based approaches are displayed in Figure 2a and summarized in Table 3. Both objective, harvest-based approaches show very similar distributions, with a mean CC yield of 728 kilograms per hectare (KGs/Ha) and a mean FP yield of 676 KGs/Ha. These differences were not statistically significant ($p > 0.2$). In contrast, the farmer self-reported (SR) yields contained many more high yielding values, including 11 (out of 252 total) plots with SR yield greater than 5,000 KGs/Ha. The highest SR yields tended to occur on very small plots, with 8 of these 11 were on plots smaller than 0.05 ha. The average SR yield of 1,826 KGs/Ha was significantly higher, and indeed more than double, that for CC and FP yields.

Given that SR, CC, and FP yields are competing ground-based measures, a useful question is how well correlated they are across different plots. Correlation between CC and FP yields was significant ($p<0.01$) but only 0.51 overall (Fig. 2c). If one views full-plot crop cutting as the "gold standard" of ground-based measures, this indicates that 8x8m crop cuts capture only roughly one-quarter of the variability in actual plot yields. These discrepancies reflect the substantial intra-plot heterogeneity of yields in these systems. The 64 m^2 area of the crop cuts, despite requiring a costly and ambitious effort, are roughly just 6 percent of the median plot size (0.11 ha or 1100 m^2) or 4 percent of the average plot size. The effect of this heterogeneity appears to be greater in intercropped plots, as the correlation between CC and FP yields is higher on pure stand maize plots ($r = 0.70$).

The more subjective SR yields show almost no correspondence ($r = 0.04$) with the crop cutting-based measures (Fig. 2b). Because correlations may be heavily influenced by a few large values of SR yields, Figure 2b reports correlations that are based on the exclusion of plots with SR yields above 5,000 KGs/Ha. Despite the increase in the correlation coefficient to 0.28, there is still less than 10 percent of the variation in CC yields that is captured by SR yields.

Table 3. Summary Statistics for Ground-Based Maize Yield Measures

	All		**Purestand**		**Intercropped**	
Yields (Kg/HA)	Mean	Median	Mean	Median	Mean	Median
Self-Reported (SR)	1826	784	1878	1039	1805	685
Sub-Plot Crop Cutting (CC)	728	595	827	725	692	571
Full Plot Crop Cutting (FP)	676	511	842	740	623	472
	Different Means?	*Different Distributions?*	*Different Means?*	*Different Distributions?*	*Different Means?*	*Different Distributions?*
SR vs. CC	***	***	***	***	**	***
CC vs. FP	--	--	--	--	--	--

Notes: ***/**/* denote statistical significance at the 1/5/10 percent level, respectively, -- denotes significance at less than 10%. The mean differences are assessed based on the t-test, while distributional differences are assessed based on the Kolmogorov–Smirnov test.

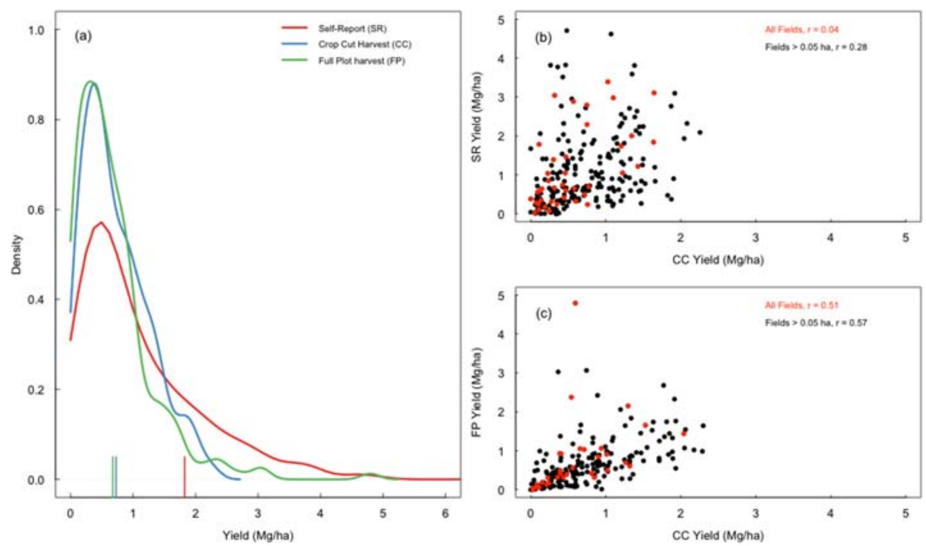

Figure 2. (a) Yield distributions for ground-based measures. Vertical bars at bottom indicate the mean yield for each measurement approach. (b) Scatter plot of SR and CC yields for all plots, and, separately, for plots above 0.05ha in size (black points). (c) Scatter plot of FP and CC yields.

3.2. Comparison of Ground- and Satellite-Based Yield Measures on Pure Stand Plots

We begin the evaluation of satellite VIs by presenting simple bivariate relationships between VIs on single dates and the objective ground-based yield measures (Fig. 3). For brevity, correlations with the subjective SR yields are not presented in this section, but they are generally lower than those for the objective yield measures.

Four important features are evident in Figure 3: (1) Correlations were generally higher between VIs and FP yields than between VIs and CC yields, which is consistent with the notion that full-plot crop cutting provides a better measure of plot-level productivity. (2) Correlations tended to improve when excluding the smallest plot sizes, consistent with the results in BL17. A likely explanation for this is the increased importance of georeferencing errors and mixed pixels on the smallest of plots. For example, a 0.05 ha plot covers an area of just five 10x10m Sentinel-2 pixels, and most of these pixels are likely to span the edge of the plot and contain some contribution from neighboring plots. (3) The MTCI consistently outperformed the other VIs on both image dates. The MTCI was designed to be sensitive to canopy chlorophyll concentration (Dash and Curran, 2004), which is likely a good proxy for yield in the low nutrient setting of Uganda. Perhaps more importantly, MTCI is much less sensitive to atmospheric conditions than other VIs such as NDVI or GCVI (Curran and Dash, 2005), because it uses the difference in reflectance between two nearby bands that will be similarly affected by atmospheric scattering. In both images, significant amounts of haze are evident above many of the plot sites in both the raw reflectance and NDVI or GCVI images. However, the MTCI images exhibit much lower sensitivity to haze (Fig. A1). (4) Finally, Figure 3 indicates that a substantial fraction of FP yield variability is captured by VIs on both dates, with MTCI capturing 37 percent of yield variability on plots at least 0.10 ha on May 30, and 49 percent on June 19. These values are similar to the amount of FP yield variability captured by CC yields on these plots (R^2 = 47 percent).

Satellite-based yields were estimated for all plots which did not contain clouds on either May 30 or June 19 (397 out of 463 total plots). The “calibrated” satellite yield estimates, obtained from a regression of FP yields vs. MTCI on May 30 and June 19, captured slightly more than half of yield variability for the pure stand plots above 0.10 ha (R^2 = 0.55, Fig. 4a). For comparison, calibration using CC rather than FP yields resulted in roughly half the amount of variability captured by satellite (R^2 = 0.26, Fig. 4b). Interestingly, though, the coefficients of the two regressions were very similar, with the model calibrated to CC yields having a slightly lower range of predicted yields. As a result, this model did nearly as well predicting FP yields (R^2 = 0.54) as the model calibrated to FP yields.

This important finding suggests that although CC yields are noisier measures of plot-level productivity compared to FP yields, this noise is mostly random and does not significantly bias the estimated coefficients in a model to predict yields from satellite data. Thus, one can expect models calibrated using CC yields (which are much more feasible and common than FP yields) to have lower R^2 but similar out of sample accuracy for predicting true plot productivity as models calibrated with FP yields.

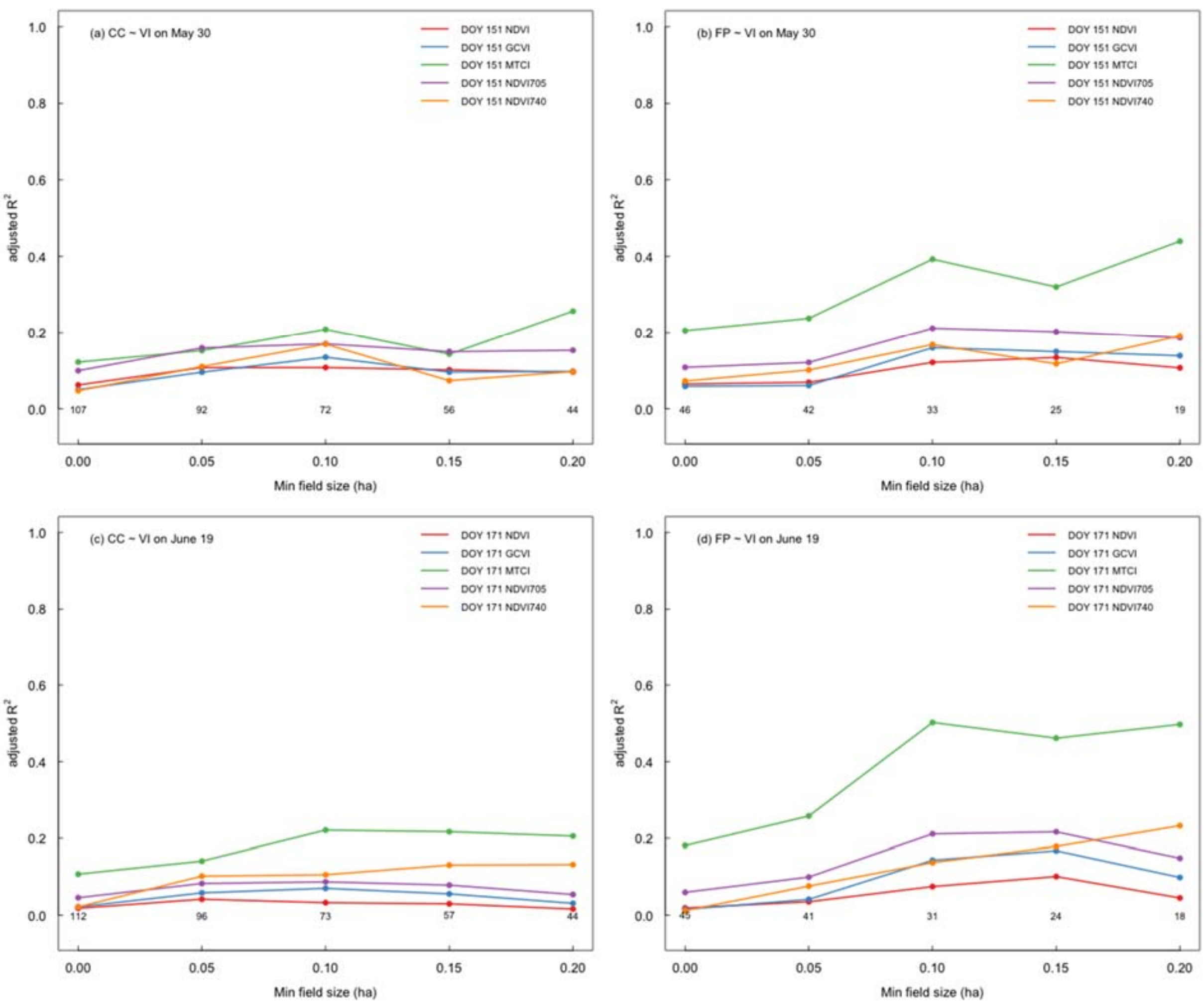

Figure 3. Adjusted R^2 of regressions of yields vs. VI, by VI type, date and type of ground-based yield measure. Top panels show results for May 30 image, bottom panels for June 19 image. Left panels show results for crop cuts, and right panels for full plot harvests. Models were run for successive subsets of data by excluding plots below indicated plot size. Numbers at bottom of plot indicate the sample size for each plot area threshold.

The "uncalibrated" estimates, obtained from a regression of simulated yields versus simulated MTCI on these same dates, resulted in a nearly identical R^2 to models calibrated with FP yields (R^2 = 0.54, Fig. 4c). The uncalibrated estimates did exhibit significant bias, with a tendency to overestimate yields by roughly 1 ton/ha, because none of the simulated yields were quite as low as the lowest of the observed FP yields. Nonetheless, the high correlation between uncalibrated estimates and true FP yields indicates that ground calibration is not a prerequisite for capturing a large fraction of spatial yield variability with satellite data.

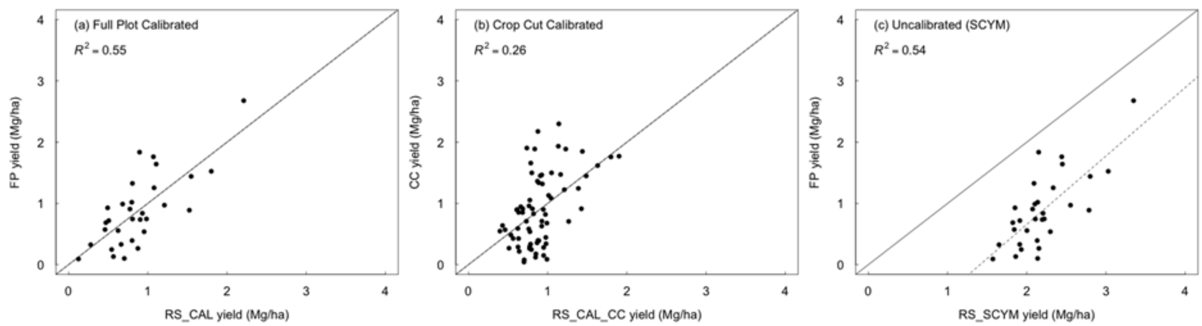

Figure 4. Comparison of (a) full plot yields vs. predictions from a remote sensing model calibrated to full plot yields, (b) crop cut yields vs. predictions from a remote sensing model calibrated to crop cut yields, and (c) full plot yields vs. "uncalibrated" remote sensing yield estimates, which are based on calibration to crop model simulations. All panels show results for pure stand maize plots at least 0.1 ha in size, which are the subset of plots used to calibrate the models in (a) and (b).

The superior performance of MTCI is noteworthy, especially given that several of the most recent satellite sensors, which possess higher spatial resolution than Sentinel-2, lack the red edge bands needed to calculate MTCI. In this study, we fortuitously had access to a relatively cloud-free image acquired by Terra Bella's Skysat sensor on May 29, one day before a Sentinel-2 image. Skysat was used in BL17, and in the context of smallholder mapping has the particularly attractive feature of 1m spatial resolution. Particularly for the small plot sizes in Uganda, we anticipated that the 1m resolution would offer substantial benefits compared to the 10m resolution of Sentinel-2's main bands, and the 20m resolution of Sentinel-2's red edge bands. Surprisingly, we found that Sentinel-2 and Skysat performed very similarly when using GCVI for both, even though many plots contained only a few Sentinel-2 pixels (Fig. A2). The large boost in performance when using MTCI with Sentinel-2 therefore more than outweighed any loss in accuracy from using coarser resolution. This result may be specific to the particular atmospheric conditions, time of growing season, and characteristics of the study site, and therefore we caution against overweighting the benefits of spectral versus spatial resolution. Nonetheless, it is an informative comparison made possible by having two images so close in time over a study site with large amounts of quality ground-based data.

3.3. Comparison of Ground- and Satellite-Based Yield Measures on All Maize Plots

Of interest in agricultural regions such as Uganda, where maize is typically intercropped with other species, is how well satellite measures can capture the performance of mixed-crop plots. Of course, ground-based yield measures are also beset by challenges from intercropping

(Carletto et al., 2015). Common practices include only measuring yields in pure stand plots, reporting yields separately for pure stand and intercropped plots, or correcting yields in intercropped plots based on either subjective or objective measures of the relative density of crops.

In our study, the ground-based measures of yield (SR, CC, and FP) in intercropped plots were obtained only for maize. We therefore compared the satellite-based yield measures to FP for different types of plots, grouped based on the presence and type of intercropping (Fig. 5). The performance on plots intercropped with legumes (beans or groundnuts) was significantly lower than on pure stand plots, with roughly 20 percent of yield variability captured for plots at least 0.10 ha in size (Fig. 5a). Maize yield estimates were even worse on plots intercropped with cassava (Fig. 5b) or both legumes and cassava (Fig. 5c), with less than 10 percent of the maize yield variability captured by the satellite estimates. The relatively better performance for legume intercrops presumably reflects the fact that both beans and groundnuts grow close to the ground, below the maize crop, whereas cassava intercrops often include very mature cassava plants that exceed the maize crop in height.

The worse performance for satellite- based maize yields on intercropped compared to pure stand plots makes sense, since non-maize crops can be a large contributor to the light reflected from the canopy and measured by satellite sensors, especially in the case of intercrops such as cassava that overhang maize plants. However, in these situations it is doubtful that the yield of maize is the best measure of land productivity. In the absence of other ground-based measures of productivity, we turn instead to assessing the sensitivity of the relationships between yield and factors of production to the choice of the ground- versus satellite-based yield variant.

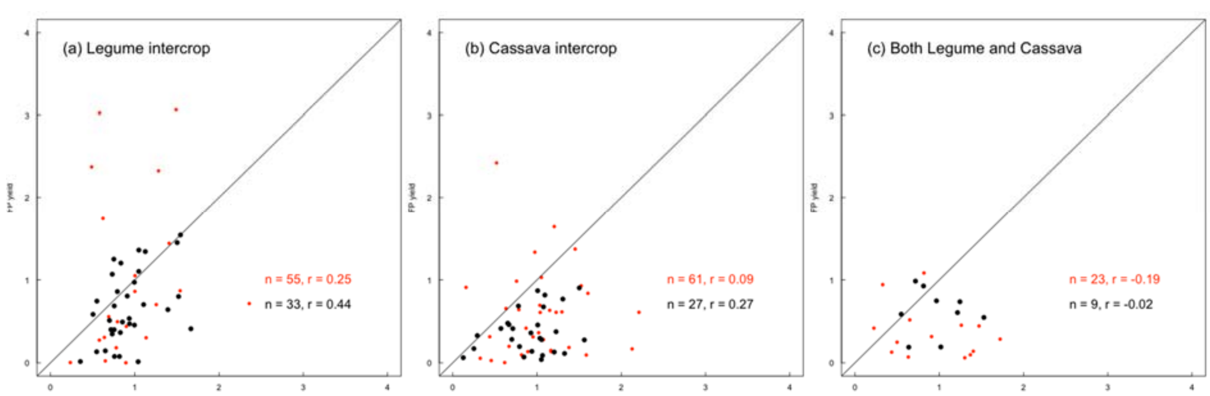

Figure 5. Comparison of calibrated remote sensing yields vs. full plot harvests for different types of intercropped plots: (a) maize intercropped with only legumes (beans, groundnuts), (b) maize intercropped with only cassava, (c) maize intercropped with both legumes and cassava. Red text shows sample size and correlation for all plots, while black points and text indicate values and corresponding sample size and correlation for only plots >.1 ha. All panels show remote sensing yields based on calibration to FP yields in purestand maize plots at least 0.1 ha in size (model shown in Fig. 4a).

3.4. Assessment of Inter-Relationships between Maize Yields and Factors of Production

Pure stand plot-level maize yield regressions resulted in similar coefficients for models using CC, FP and satellite-based yields (Table A1). The coefficients for three factors of production of interest – plot area, soil quality index, and incidence of inorganic fertilizer use – are visualized in Figure 6a. As also noted by Gourlay et al. (2017), the regression using SR yields resulted in a much stronger negative coefficient for plot area than the objective ground-based measures, indicating that the conventional wisdom of an inverse-relationship between farm size and productivity may be an artifact of measurement error. While the relationship between soil quality and any one of CC, FP and satellite-based yields was positive and statistically significant at least at the 5 percent level, the coefficient associated with soil quality failed to be statistically significant in the regression using SR yields. In line with the results of the CC and FP yield regressions, the relationship between fertilizer use and any one of the calibrated or uncalibrated satellite-based yields was positive and statistically significant at the 1 percent level.

The regressions for all plots, including both pure stand and intercropped plots, show qualitatively similar coefficients, as depicted Figure 6b and Table A2. The satellite-based regressions still find a significant positive effect of soil quality, whereas the coefficients on fertilizer remain positive but become statistically insignificant. A likely explanation for this result is that cassava biomass, which influences the satellite-based yield estimates on intercropped plots, is similar to maize in its responsiveness to soil quality, but less responsive to inorganic fertilizer. In comparison to regressions using FP yields, those using either CC or satellite-based yields generally had smaller confidence intervals for coefficient values, which reflects the fact that full plot harvests were only performed on 211 plots, whereas sub-plot crop cutting was done for all 463 and satellite estimates were available on 397.

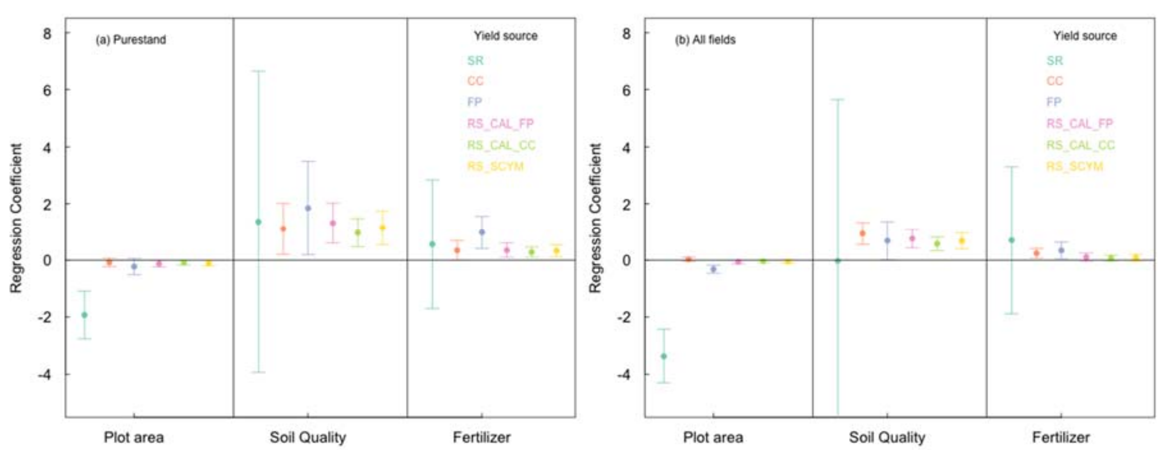

Figure 6. Summary of regression coefficients for three relevant factors using six different models corresponding to six yield measures. Error bars show +/- two standard deviations of the mean estimate.

4. Discussion and Conclusions

Despite the importance of agriculture for rural livelihoods, poverty alleviation, and food security across the developing world, household and farm surveys collecting micro data on agriculture exhibit substantial cross-country heterogeneity in terms of access policies, use of international best practice survey methods and dissemination standards, and data quality (Carletto et al., 2015). Given the rapid advances in the availability of 10-meter or sub-10-meter spatial resolution satellite imagery, the demand is increasing for understanding how these advances can be leveraged to measure and understand agricultural outcomes with greater accuracy and higher spatial resolution.

Although there is a concerted push to showcase the value of geospatial applications for monitoring and evaluation efforts in the agriculture sector, and for tracking the progress towards the Sustainable Development Goals, multi-disciplinary research efforts aimed at assessing the accuracy and feasibility of the proposed applications, particularly in smallholder production systems, are scant. If validated, satellite-based remote sensing, combined with georeferenced household and farm survey data that could serve as “ground truth”, could dramatically enhance not only our ability to fill the data gaps, but also our understanding of the linkages between development and human welfare.

Taking advantage of a unique range of ground-based plot-level maize yield measures based on farmer-reporting, sub-plot crop cutting and full-plot harvests that were collected as part of a methodological survey experiment that was conducted in Eastern Uganda, our study showcases the accuracy and empirical utility of satellite-based approaches to plot-level maize yield estimation in smallholder production systems with a median plot size of approximately one-tenth of a hectare.

The satellite-based yield estimates include those that are (a) anchored in a calibration model that relates maize yields from full-plot harvests to MTCI values on multiple dates on a subset of pure stand maize plots that were at least 0.1 ha in size; (b) based on the same calibration model that uses sub-plot crop cut, as opposed to full-plot, yield; and (c) based solely on crop model simulations, without reliance on any ground-based yield measure. While (a) and (b) are identified as “calibrated” variants of remotely-sensed maize yields, (c) is framed as the “uncalibrated” counterpart.

Overall, the accuracy of the satellite-based maize yield estimates is very encouraging. Having over 200 full plot harvests, which is very rare because of their cost, is a unique situation with which to test satellite estimates, and we find that both calibrated and uncalibrated approaches capture roughly half of the variance in full plot harvests when restricting the analysis to where both ground and satellite approaches are measuring the same output (pure stand plots) and where the satellite pixels corresponding to the plot are less likely to be contaminated by neighboring plots (plots > 0.10 hectare). The uncalibrated approach exhibits, however, a strong tendency to overestimate yields, but adequately captures spatial variation in yield. In fact, the satellite-based estimates explained roughly the same amount of variance in full plot harvests as sub-plot crop cuts performed within the plots.

Perhaps more convincingly, satellite-based estimates are able to faithfully reproduce the effects of key production factors such as soil quality and fertilizer use, even when including plots of all sizes and those that are intercropped. The significance levels of the coefficients informed by the satellite-based measures are often even higher than those underlined by the full plot harvests. This finding again emphasizes two important points. First is that any measure of yield is prone to errors, and thus an imperfect correlation with full plot harvests reflects errors in ground-based estimates as well as those in satellite-based estimates. Second, even if satellite-based measures are less accurate than full plot harvests, the greater sample size can compensate for any loss in accuracy.

Also noteworthy is the fact that satellite-based models calibrated to CC yields perform similarly to those calibrated to FP yields, in terms of both agreement with FP yields and estimation of yield response to soil quality and fertilizer. These results indicate that although CC yields are imperfect approximations of actual yields, the errors do not substantially bias remote sensing calibrations. Thus, sub-plot crop cutting appears to be a suitable replacement for full-plot harvests when the latter are not possible. Of course, crop model simulations can also be used as a replacement for any ground-based measures, if the potential bias in estimated yields is recognized and acceptable. The possibility of combining simulations with a small number of ground samples for providing improved accuracy at a minimal cost could be explored in the future.

Finally, even though our study placed emphasis on measuring plot-level yields, many applications, such as forecasting regional food supply or assessing local conditions for insurance payouts, will care more about accuracy at aggregate scales. What is expected to become increasingly more useful and insightful will be the ability to integrate georeferenced micro survey data on agriculture, such as the LSMS-ISA, with the expanding, publicly-available high-resolution satellite imagery. Such ability, combined with advances in remote sensing methods as well as mobile technology and handheld sensors for cost-effective, objective ground data capture, has the potential to create an unparalleled scope for research on entire landscapes of agricultural plots. Collectively, these measurement tools will allow more rapid feedback on the effectiveness of different efforts to raise productivity, which in turn can enable more effective food policy.

Acknowledgements

The study was conducted under the partnership between the World Bank Living Standards Measurement Study and the Stanford University Center on Food Security and the Environment (FSE) for developing and testing approaches for using satellite measurements, in combination with ground data, to estimate plot-level crop yields in smallholder production systems. The technical assistance to the 2015 and 2016 rounds of MAPS: Methodological Experiment on Measuring Maize Productivity, Soil Fertility and Variety was provided through the World Bank Living Standards Measurement Study (LSMS) "Minding the (Agricultural) Data Gap" Methodological Research Program, funded by UK Aid. The implementation of MAPS I (2015) was financed by the World Bank LSMS – Minding the (Agricultural) Data Gap Research Program, the Global Strategy to Improve Agricultural and Rural Statistics, led by the Food and Agriculture Organization of the United Nations (FAO), and the CGIAR Standing Panel on Impact Assessment. The implementation of MAPS II (2016) was financed by the World Bank Innovations in Big Data Analytics Program, the World Bank Trust Fund for Statistical Capacity Building – Innovations in Development Data Window, and the CGIAR Standing Panel on Impact Assessment. In MAPS I and MAPS II, Terra Bella provided free high-resolution satellite imagery for the MAPS remote sensing tasking area for research purposes. The technical partners included the World Agroforestry Center on soil fertility measurement, and the CGIAR Standing Panel on Impact Assessment on maize variety identification. MAPS I and MAPS II were both implemented using the World Bank *Survey Solutions* Computer-Assisted Personal Interviewing (CAPI) platform. The research team would like to thank the dedicated management and field staff of Uganda Bureau of Statistics regarding fieldwork implementation; Mr. Wilbert Drazi Vundru for Survey Solutions programming, fieldwork supervision and survey data quality control; and Ms. Madeline Lisaius for help with image processing.

References

Berazneva, J., McBride, L., Sheahan, M., Güereña, D., 2018. Empirical assessment of subjective and objective soil fertility metrics in east Africa: Implications for researchers and policy makers. World Dev. doi:10.1016/j.worlddev.2017.12.009

Bevis, L.E.M., Barrett, C.B., 2017. Close to the Edge : Do Behavioral Explanations Account for the Inverse Productivity Relationship?, pubdocs.worldbank.org.

Burke, M., Lobell, D.B., 2017. Satellite-based assessment of yield variation and its determinants in smallholder African systems. Proc. Natl. Acad. Sci. 114, 2189–2194. doi:10.1073/pnas.1616919114

Byerlee, D., De Janvry, A., Sadoulet, E., Townsend, R., Klytchnikova, I., 2007. World Development Report, 2008: agriculture for development. World Bank, Washington, DC, http//siteresources. worldbank. org/INTWDR2008/Resources/WDR_00_book. pdf, accessed 23.

Carletto, C., Jolliffe, D., Banerjee, R., 2015. From Tragedy to Renaissance: Improving Agricultural Data for Better Policies. J. Dev. Stud. 51, 133–148. doi:10.1080/00220388.2014.968140

Casley, D., Kumar, K., 1988. The collection, analysis and use of monitoring and evaluation data. The World Bank.

Curran, P.J., Dash, J., 2005. ALGORITHM THEORETICAL BASIS DOCUMENT ATBD 2.22 CHLOROPHYLL INDEX.

Darko, F., Palacios Lopez, A., Kilic, T., Ricker-Gilbert, J., 2018. Micro-level welfare impacts of agricultural productivity: evidence from rural Malawi Forthcomin.

Dash, J., Curran, P.J., 2004. The MERIS terrestrial chlorophyll index. Int. J. Remote Sens. 25, 5403–5413.

Davis, B., Di Giuseppe, S., Zezza, A., 2017. Are African households (not) leaving agriculture? Patterns of households' income sources in rural Sub-Saharan Africa. Food Policy 67, 153–174. doi:10.1016/j.foodpol.2016.09.018

Desiere, S., Jolliffe, D., 2018. Land productivity and plot size: Is measurement error driving the inverse relationship? J. Dev. Econ. 130, 84–98. doi:10.1016/J.JDEVECO.2017.10.002

Dorosh, P., Thurlow, J., 2016. Beyond Agriculture Versus Non-Agriculture: Decomposing Sectoral Growth–Poverty Linkages in Five African Countries. World Dev. doi:10.1016/J.WORLDDEV.2016.08.014

Fermont, A., Benson, T., 2011. Estimating yield of food crops grown by smallholder farmers: A Review in the Uganda Context. IFPRI Discuss. Pap. 01097 1–57.

Gourlay, S., Kilic, T., Lobell, D., 2017. Could the Debate Be Over? Errors in Farmer-Reported Production and Their Implications for Inverse Scale - Productivity Relationship in Uganda. doi:10.1596/1813-9450-8192

Harou, A.P., Liu, Y., Barrett, C.B., You, L., 2017. Variable Returns to Fertiliser Use and the

Geography of Poverty: Experimental and Simulation Evidence from Malawi. J. Afr. Econ. 26, 342–371. doi:10.1093/jae/ejx002

Lobell, D.B., Thau, D., Seifert, C., Engle, E., Little, B., 2015. A scalable satellite-based crop yield mapper. Remote Sens. Environ. 164, 324–333. doi:10.1016/j.rse.2015.04.021

Mukherjee, A., Lal, R., 2014. Comparison of soil quality index using three methods. PLoS One.

Schlemmer, M., Gitelson, A., Schepersa, J., Fergusona, R., Peng, Y., Shanahana, J., Rundquist, D., 2013. Remote estimation of nitrogen and chlorophyll contents in maize at leaf and canopy levels. Int. J. Appl. Earth Obs. Geoinf. 25, 47–54. doi:10.1016/j.jag.2013.04.003

APPENDIX

Table A1. Regression Coefficients for Pure Stand Plots Using Different Yield Measures

	Dependent Variable/Maize Yield Type					
	Self-report	Crop-cut	Full plot	RS_cal_fp	RS_cal,cc	RS_scym
	(1)	(2)	(3)	(4)	(5)	(6)
Log Plot Area (GPS, ha)	-1.94*** (0.42)	-0.08 (0.07)	-0.23 (0.14)	-0.13** (0.06)	-0.09** (0.04)	-0.11** (0.05)
Log Plot Distance from Dwelling (GPS, km)	0.10 (0.33)	-0.04 (0.06)	-0.19 (0.12)	-0.07 (0.05)	-0.04 (0.04)	-0.05 (0.04)
Cover Crops Present Prior to Planting †	-0.35 (0.99)	0.01 (0.20)	0.26 (0.48)	-0.04 (0.15)	-0.03 (0.11)	-0.04 (0.13)
Log Maize Seed Planting Rate (Kg/Ha)	1.19** (0.48)	0.09 (0.08)	0.18 (0.14)	0.12* (0.06)	0.09* (0.05)	0.10* (0.05)
Inorganic Fertilizer Application †	0.56 (1.14)	0.35** (0.17)	0.98*** (0.28)	0.35*** (0.13)	0.28*** (0.09)	0.33*** (0.11)
Log Household Labor Days	0.56* (0.30)	0.05 (0.06)	-0.01 (0.10)	0.04 (0.04)	0.05 (0.03)	0.05 (0.03)
Log Hired Labor Days	0.27 (0.42)	-0.01 (0.06)	-0.03 (0.10)	-0.11** (0.05)	-0.08** (0.03)	-0.09** (0.04)
No Hired Labor †	0.13 (0.96)	-0.24 (0.16)	0.09 (0.26)	-0.07 (0.12)	-0.04 (0.09)	-0.05 (0.10)
Soil Quality Index	1.36 (2.64)	1.11** (0.45)	1.84** (0.82)	1.31*** (0.35)	0.97*** (0.25)	1.14*** (0.30)
Wealth Index	0.46 (0.39)	0.09 (0.07)	-0.05 (0.12)	-0.08 (0.05)	-0.05 (0.04)	-0.06 (0.04)
Agricultural Asset Index	0.43 (0.32)	-0.01 (0.06)	0.09 (0.10)	0.07* (0.04)	0.05 (0.03)	0.06 (0.03)
Dependency Ratio	-0.16 (0.35)	0.01 (0.06)	0.01 (0.10)	-0.02 (0.05)	-0.02 (0.03)	-0.02 (0.04)
Household Size	-0.04 (0.11)	0.01 (0.02)	0.02 (0.04)	0.01 (0.02)	0.002 (0.01)	0.003 (0.01)
Manager = Respondent†	0.07 (0.83)	0.03 (0.16)	-0.05 (0.38)	0.04 (0.13)	0.06 (0.09)	0.07 (0.11)
Received Crop-Production Related Extension Services†	-0.08 (0.69)	-0.16 (0.12)	0.26 (0.19)	0.09 (0.09)	0.08 (0.06)	0.09 (0.08)
Female†	-0.20 (0.73)	-0.09 (0.13)	-0.04 (0.25)	-0.21** (0.10)	-0.15** (0.07)	-0.18** (0.09)
Age (Years)	-0.03 (0.02)	-0.004 (0.004)	0.003 (0.01)	-0.0003 (0.003)	-0.001 (0.002)	-0.001 (0.003)
Years of Education	-0.09 (0.07)	-0.01 (0.01)	0.03 (0.02)	-0.001 (0.01)	-0.003 (0.01)	-0.003 (0.01)
Constant	-4.35 (3.25)	-0.12 (0.60)	-2.01* (1.08)	-0.56 (0.46)	-0.13 (0.33)	0.95** (0.39)
Observations	73	124	51	105	105	105
R^2	0.4	0.19	0.47	0.36	0.37	0.37
Adjusted R^2	0.19	0.05	0.17	0.23	0.24	0.24
Residual Std. Error	2.25 (df = 54)	0.54 (df = 105)	0.55 (df = 32)	0.37 (df = 86)	0.26 (df = 86)	0.31 (df = 86)
F Statistic	1.96** (df = 18; 54)	1.33 (df = 18; 105)	1.55 (df = 18; 32)	2.69*** (df = 18; 86)	2.78*** (df = 18; 86)	2.78*** (df = 18; 86)

Notes: † denotes a dummy variable. ***/**/* denote statistical significance at the 1/5/10 percent level, respectively. Standard errors in parentheses.

Table A2. Regression Coefficients for All (Pure Stand + Intercropped) Plots Using Different Yield Measures

	Dependent Variable/Maize Yield Type					
	Self-report	Crop-cut	Full plot	RS_cal_fp	RS_cal,cc	RS_scym
	(1)	(2)	(3)	(4)	(5)	(6)
Log Plot Area (GPS, ha)	-3.37*** (0.47)	0.02 (0.04)	-0.32*** (0.07)	-0.06** (0.03)	-0.04* (0.02)	-0.05* (0.03)
Log Plot Distance from Dwelling (GPS, km)	0.21 (0.36)	-0.02 (0.03)	-0.06 (0.05)	-0.04* (0.02)	-0.03 (0.02)	-0.03* (0.02)
Cover Crops Present Prior to Planting †	0.18 (0.85)	0.05 (0.07)	0.01 (0.14)	0.03 (0.06)	0.01 (0.04)	0.01 (0.05)
Log Maize Seed Planting Rate (Kg/Ha)	1.74*** (0.46)	0.03 (0.03)	0.17*** (0.06)	0.04 (0.03)	0.03 (0.02)	0.03 (0.03)
Inorganic Fertilizer Application †	0.70 (1.30)	0.24*** (0.09)	0.34** (0.15)	0.10 (0.08)	0.06 (0.06)	0.07 (0.07)
Log Household Labor Days	0.97** (0.43)	0.01 (0.03)	0.10 (0.06)	-0.04 (0.03)	-0.02 (0.02)	-0.03 (0.02)
Log Hired Labor Days	-0.32 (0.52)	-0.001 (0.03)	0.03 (0.06)	-0.04 (0.03)	-0.03 (0.02)	-0.03 (0.03)
No Hired Labor †	-2.39* (1.22)	-0.09 (0.08)	-0.06 (0.13)	-0.05 (0.07)	-0.03 (0.05)	-0.04 (0.06)
Soil Quality Index	-0.03 (2.84)	0.94*** (0.19)	0.69** (0.34)	0.76*** (0.16)	0.58*** (0.12)	0.68*** (0.14)
Wealth Index	0.13 (0.37)	0.04 (0.03)	-0.06 (0.06)	-0.02 (0.02)	-0.01 (0.02)	-0.02 (0.02)
Agricultural Asset Index	-0.16 (0.37)	0.04 (0.03)	0.07 (0.05)	0.01 (0.02)	0.002 (0.02)	0.002 (0.02)
Dependency Ratio	-0.21 (0.37)	0.02 (0.03)	0.01 (0.04)	-0.002 (0.02)	-0.003 (0.02)	-0.004 (0.02)
Household Size	-0.10 (0.12)	-0.02* (0.01)	0.005 (0.02)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
Manager = Respondent†	0.48 (0.82)	-0.04 (0.07)	0.13 (0.14)	-0.13** (0.06)	-0.09** (0.04)	-0.11** (0.05)
Received Crop-Production Related Extension Services†	-0.01 (0.72)	-0.06 (0.05)	0.02 (0.10)	-0.04 (0.05)	-0.03 (0.03)	-0.03 (0.04)
Female†	0.43 (0.80)	-0.08 (0.06)	-0.04 (0.11)	-0.09* (0.05)	-0.07* (0.04)	-0.08* (0.04)
Age (Years)	0.01 (0.02)	0.0001 (0.002)	0.01** (0.003)	0.003** (0.001)	0.002** (0.001)	0.003** (0.001)
Years of Education	0.01 (0.08)	-0.002 (0.01)	0.02* (0.01)	0.003 (0.005)	0.003 (0.003)	0.003 (0.004)
Purestand †	-0.21 (0.78)	0.10* (0.06)	0.29*** (0.11)	0.03 (0.05)	0.01 (0.04)	0.01 (0.04)
Log Intercropping Seeding Rate (=100 for Pure stand Plots)	0.07 (0.69)	0.07 (0.05)	0.02 (0.08)	-0.05 (0.04)	-0.04 (0.03)	-0.05 (0.04)
Constant	-9.65** (4.57)	-0.05 (0.32)	-1.94*** (0.59)	0.48* (0.28)	0.66*** (0.21)	1.89*** (0.24)
Observations	252	463	211	397	397	397
R^2	0.21	0.14	0.21	0.13	0.13	0.13
Adjusted R^2	0.14	0.1	0.13	0.09	0.08	0.08
Residual Std. Error	4.96 (df = 231)	0.49 (df = 442)	0.59 (df = 190)	0.39 (df = 376)	0.29 (df = 376)	0.34 (df = 376)
F Statistic	3.07*** (df = 20; 231)	3.55*** (df = 20; 442)	2.57*** (df = 20; 190)	2.87*** (df = 20; 376)	2.73*** (df = 20; 376)	2.73*** (df = 20; 376)

Notes: † denotes a dummy variable. ***/**/* denote statistical significance at the 1/5/10 percent level, respectively. Standard errors in parentheses.

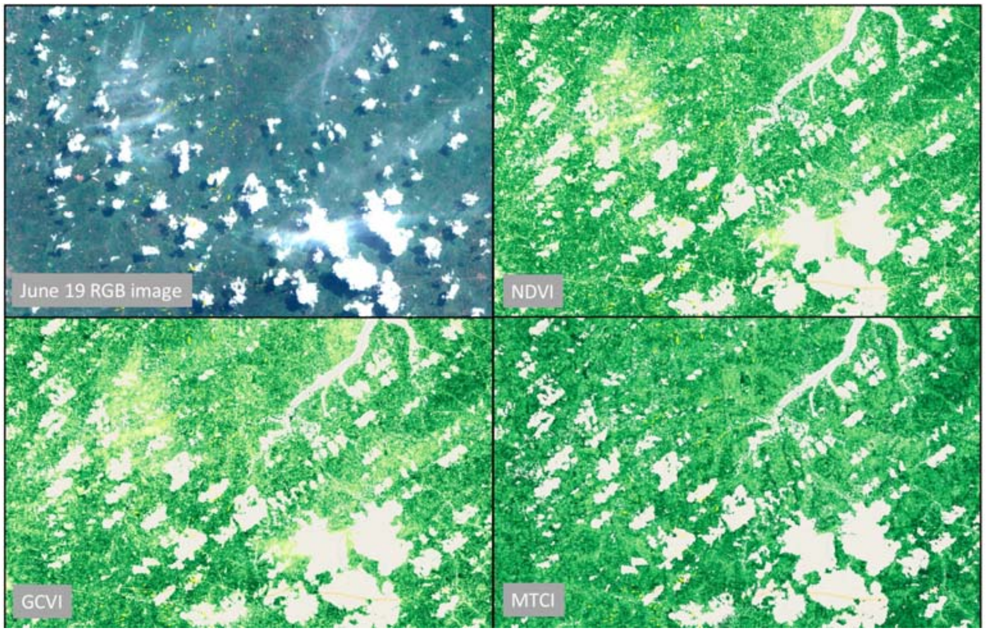

Figure A1. The effects of haze on a subsection of the (a) raw red-green-blue reflectance image from June 19, 2016, and the corresponding values of (b) NDVI (c) GCVI and (d) MTCI. For (b)-(d) darker green indicates higher values, and yellow indicates lower values (each VI has a different scale). Areas masked as cloud or cloud shadows are not shown. Both NDVI and GCVI show clear patterns associated with haze, whereas MTCI is less affected.

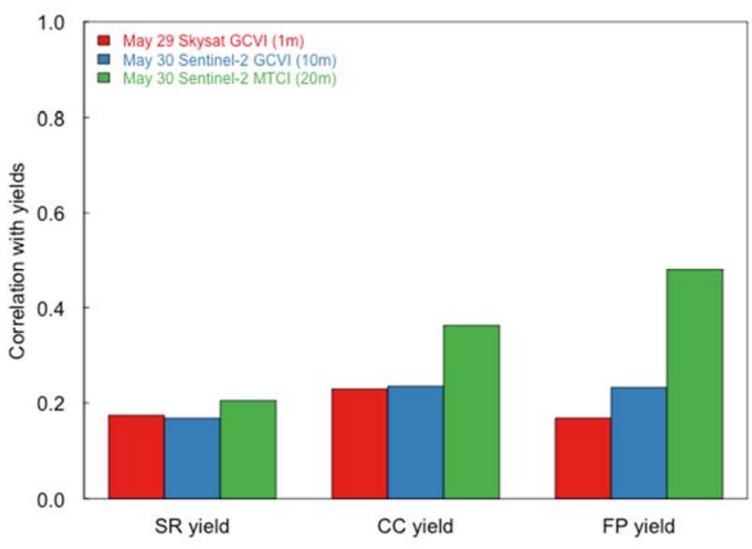

Figure A2. Correlation of different yield measures with VI from Skysat on May 29 or Sentinel-2 on May 30, 2016.