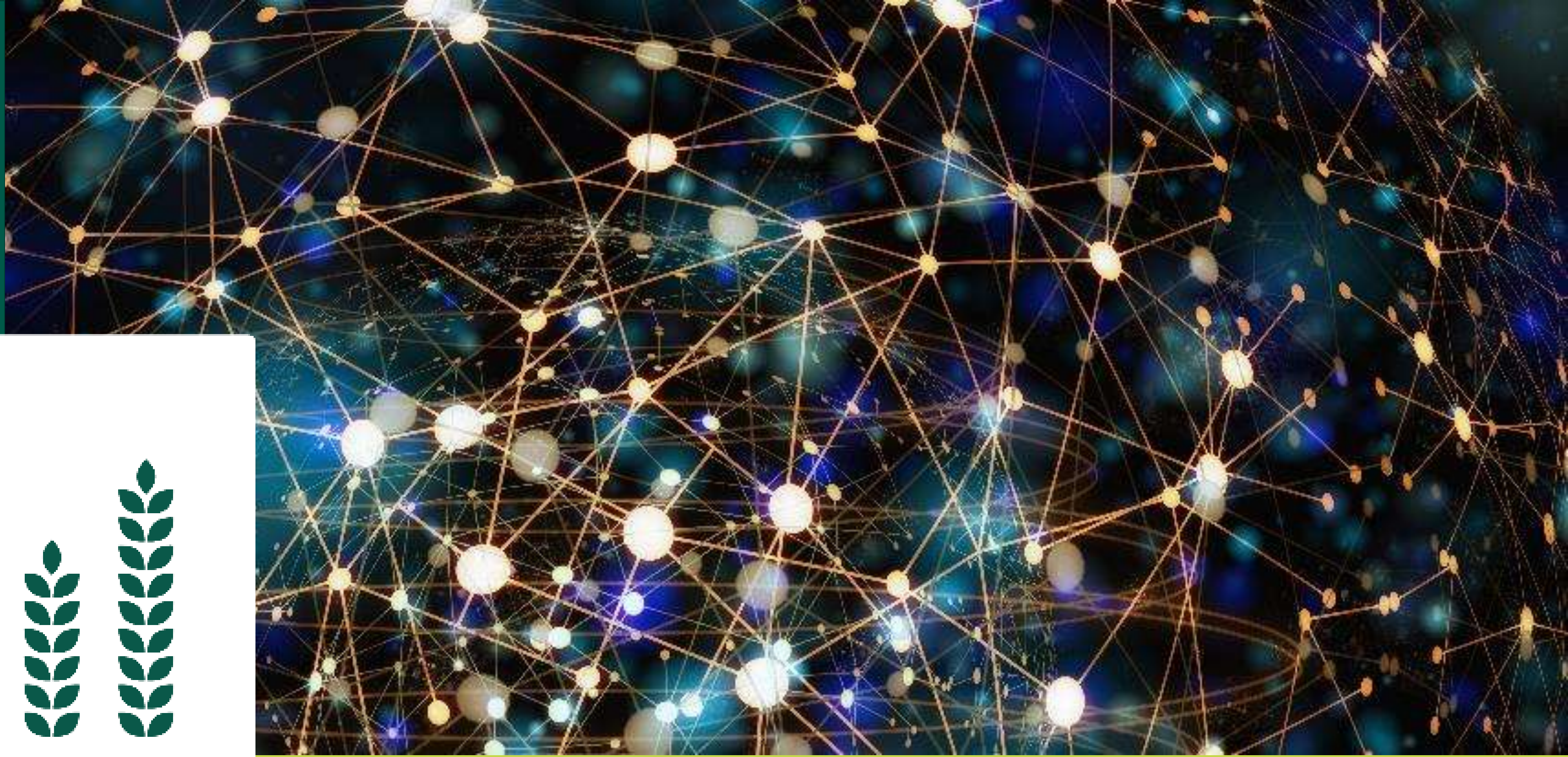


50x2030
DATA-SMART AGRICULTURE

SEMINAR SERIES

Integrating Surveys And Satellites for Agricultural Monitoring in Smallholder Farming Systems

CHAIR: ROBERT TOWNSEND, LEAD ECONOMIST, WORLD BANK

15 APRIL 2021

AGENDA

- **Presentation: "Understanding the Requirements for Surveys to Support Satellite-based Crop Type Mapping: Evidence from Sub-Saharan Africa"**

  Talip Kilic, Senior Economist, Development Data Group, World Bank

  George Azzari, Chief Technology Officer, Atlas AI

- **Discussants**

  Catherine Nakalembe, Africa Program Director, NASA Harvest

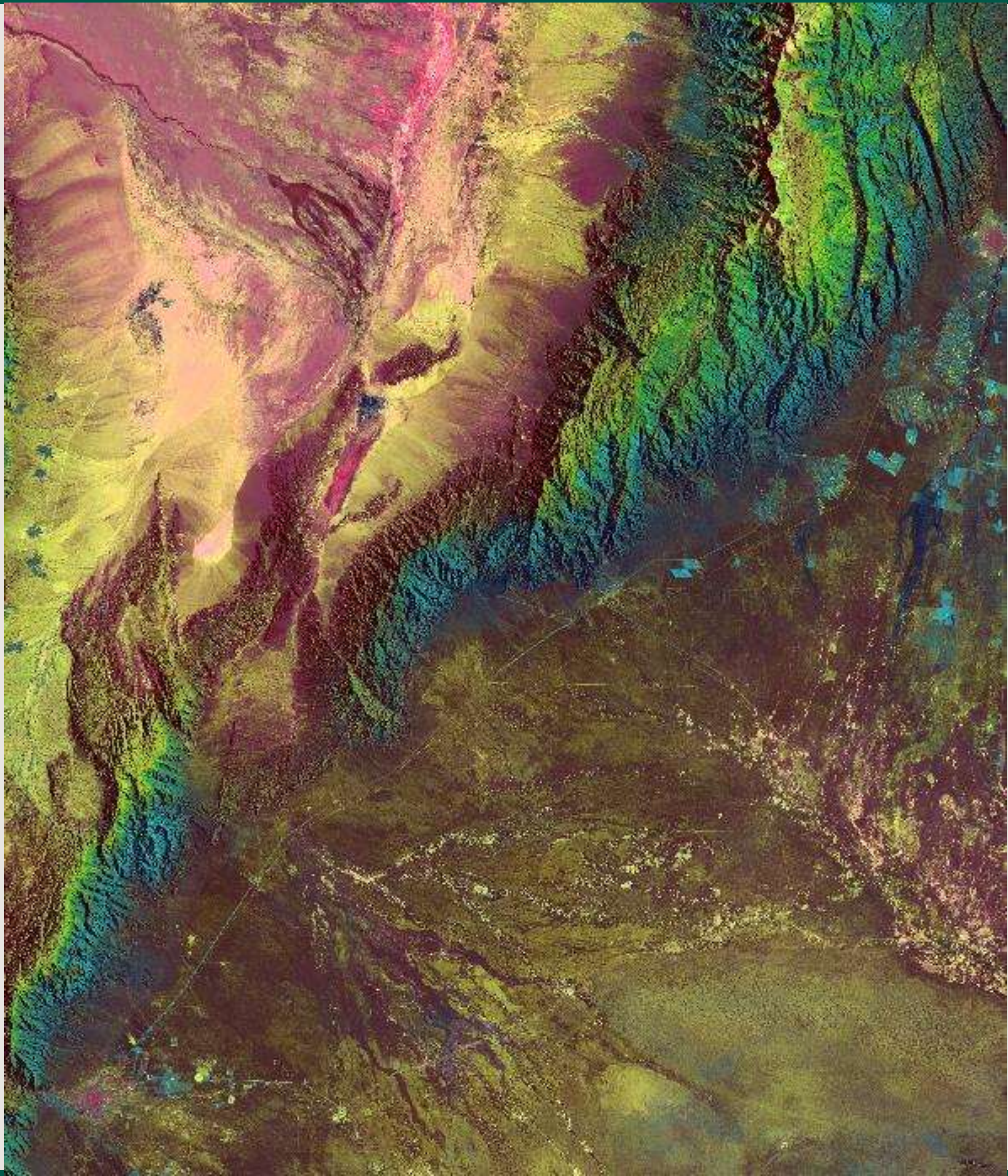
  Aberash Tariku, Deputy Director, Central Statistical Agency of Ethiopia

  Christophe Duhamel, 50x2030 Data Production Manager, Food and Agriculture Organization

- **Q&A**

WORKING PAPER PRESENTATION

"Understanding the Requirements for Surveys to Support Satellite-based Crop Type Mapping: Evidence from Sub-Saharan Africa"



Background

- Role of agriculture in rural livelihoods
  - Byerlee et al. 2007, Davis, et al. 2017
- Need for accurate, crop-specific measures of area under cultivation, production and yields – not only at the national-level but with enhanced within-country disaggregation
- Surge in availability of high-resolution satellite imagery and evidence on the feasibility of satellite-based monitoring of agricultural outcomes in smallholder farming systems
  - Need for data to train and validate the underlying models

Background

Training Data Source: Labelled Imagery

Remote Sensing of Environment 221 (2019) 551–568

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world

Pierre Defourny[a], Sophie Bontemps[a,*], Nicolas Bellemans[a], Cosmin Cara[b], Gérard Dedieu[c], Eric Guzzonato[d], Olivier Hagolle[c], Jordi Inglada[c], Laurentiu Nicola[b], Thierry Rabaute[d], Mickael Savinaud[d], Cosmin Udroiu[b], Silvia Valero[c], Agnès Bégué[e,f], Jean-François Dejoux[c], Abderrazak El Harti[g], Jamal Ezzahar[h,o], Nataliia Kussul[i], Kamal Labbassi[j], Valentine Lebourgeois[e,f], Zhang Miao[k], Terrence Newby[l], Adolph Nyamugama[l], Norakhan Salh[m], Andrii Shelestov[i], Vincent Simonneaux[c,n], Pierre Sibiry Traore[o], Souleymane S. Traore[p], Benjamin Koetz[q]

Nominal 30-m Cropland Extent Map of Continental Africa by Integrating Pixel-Based and Object-Based Algorithms Using Sentinel-2 and Landsat-8 Data on Google Earth Engine

by Jun Xiong [1,2,*], Prasad S. Thenkabail [1], James C. Tilton [3], Murali K. Gumma [4], Pardhasaradhi Teluguntla [1,2], Adam Oliphant [1], Russell G. Congalton [5], Kamini Yadav [5] and Noel Gorelick [6]

International Journal of Applied Earth Observation and Geoinformation

Volume 89, July 2020, 102087

Multiple factors influence the consistency of cropland datasets in Africa

Yanbing Wei [a], Miao Lu [a], Wenbin Wu [a], Yating Ru [b]

Training Data Source: Ground Data

remote sensing

MDPI

Article

Mapping Smallholder Yield Heterogeneity at Multiple Scales in Eastern Africa

Zhenong Jin [1,*], George Azzari [1], Marshall Burke [1,2], Stephen Aston [3] and David B. Lobell [1]

Remote Sensing of Environment 228 (2019) 115–128

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Smallholder maize area and yield mapping at national scales with Google Earth Engine

Zhenong Jin[a,1], George Azzari[a,1,2], Calum You[a], Stefania Di Tommaso[a], Stephen Aston[b], Marshall Burke[a], David B. Lobell[a,*]

PNAS

Satellite-based assessment of yield variation and its determinants in smallholder African systems

Marshall Burke[a,b,c,1,2] and David B. Lobell[a,b,1]

[a]Department of Earth System Science, Stanford University, Stanford, CA 94305; [b]Center on Food Security and the Environment, Stanford University, Stanford, CA 94305; and [c]National Bureau of Economic Research, Cambridge, MA 02138

Edited by B. L. Turner, Arizona State University, Tempe, AZ, and approved January 12, 2017 (received for review October 12, 2016)

Remote Crop Mapping at Scale: Using Satellite Imagery and UAV-Acquired Data as Ground Truth

by Meghan Hegarty-Craver [1,*], Jason Polly [1], Margaret O'Neil [1], Noel Ujeneza [2], James Rineer [1], Robert H. Beach [1], Daniel Lapidus [1] and Dorota S. Temple [1]

1 RTI International, Research Triangle Park, NC 27709, USA
2 Independent Agri-Consultant, Kigali 20093, Rwanda
* Author to whom correspondence should be addressed.

Remote Sens. **2020**, *12*(12), 1984; https://doi.org/10.3390/rs12121984

Remote Sensing of Environment 216 (2018) 647–657

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Estimating smallholder crops production at village level from Sentinel-2 time series in Mali's cotton belt

Marie-Julie Lambert[a,*], Pierre C. Sibiry Traoré[b,c], Xavier Blaes[a], Philippe Baret[a], Pierre Defourny[a]

Maize Cropping Systems Mapping Using RapidEye Observations in Agro-Ecological Landscapes in Kenya

by Kyalo Richard [1], Elfatih M. Abdel-Rahman [1,2], Sevgan Subramanian [1], Johnson O. Nyasani [1,3], Michael Thiel [4], Hosein Jozani [4], Christian Borgemeister [5] and Tobias Landmann [1,*]

American Journal of Agricultural Economics

Issues JEL Advance articles About

Article Contents
Abstract
Data
Results
Discussion and Conclusions

Eyes in the Sky, Boots on the Ground: Assessing Satellite- and Ground-Based Approaches to Crop Yield Measurement and Analysis

David B Lobell, George Azzari, Marshall Burke, Sydney Gourlay, Zhenong Jin, Talip Kilic, Siobhan Murray Author Notes

remote sensing

MDPI

Article

Sight for Sorghums: Comparisons of Satellite- and Ground-Based Sorghum Yield Estimates in Mali

David B. Lobell [1,*], Stefania Di Tommaso [1], Calum You [1], Ismael Yacoubou Djima [2], Marshall Burke [1] and Talip Kilic [2]

Background

- Role of agriculture in rural livelihoods
  - Byerlee et al. 2007, Davis, et al. 2017
- Need for accurate, crop-specific measures of area under cultivation, production and yields – not only at the national-level but with enhanced within-country disaggregation
- Surge in availability of high-resolution satellite imagery and evidence on the feasibility of satellite-based monitoring of agricultural outcomes in smallholder farming systems
  - Need for data to train and validate the underlying models
- Evidence on the impact of training data on the quality and spatial resolution of satellite-based estimates
  - Lobell et al. 2019, 2020
- Research largely sub-national in scope, with heterogeneity in the scope of and approach to ground data collection

Contributions

- Address several operational and inter-related research questions in the context of high-resolution maize area mapping across Malawi and Ethiopia
    - What is the minimum volume of training data to reach an acceptable level of accuracy of a crop classification algorithm?
    - How does the approach to georeferencing plot locations in household surveys impact the accuracy of the same algorithm?
    - How do the type of satellite data and exclusion of plots under specific area thresholds affect the algorithmic accuracy?

- Inform the guidelines being developed under the 50x2030 initiative for the collection of georeferenced data in large-scale surveys to train and validate earth observation models for high-resolution crop type mapping and crop yield estimation in smallholder farming systems

Survey Data

- Georeferenced plot-level survey data stem from nationally-representative, multi-topic surveys that were implemented by the Malawi NSO and the CSA of Ethiopia under the World Bank LSMS-ISA Initiative
  - Malawi Integrated Household Panel Survey (IHPS) 2019
    - Longitudinal sample, dating back to 2010
    - Reference season: 2018/19
    - Plot-level georeferenced information: Single plot corner point + plot boundaries
  - Malawi Fifth Integrated Household Survey (IHS5) 2019/20
    - Cross-sectional sample
    - Reference season: 2017/18 or 2018/19
    - Plot-level georeferenced information: Single plot corner point + plot boundaries
  - Ethiopia Socioeconomic Survey (ESS) 2018/19
    - Baseline for a new longitudinal sample
    - Reference season: 2018 *meher* season
    - Plot-level georeferenced information: Single plot corner point

Survey Data (2)

Table 1: IHPS 2019 and IHS5 2019/20 rainy season plots by georeferenced information availability

Plot category	IHPS 2019		IHS5 2019/20	
	Obs	%	Obs	%
Plots with no geolocation information	334	6.2	1,105	6.4
Plots with a corner point, but no polygon boundary	1,365	25.4	4,871	28.4
Plots with a corner point and a polygon boundary, but dropped from analysis	874	16.3	2,139	12.5
Plots with a corner point and a polygon boundary, used for analysis	2,792	52.0	9,059	52.7
Total # of Plots	**5,365**	**100.0**	**17,174**	**100.0**
Total # of Associated Households	**2,335**		**8,770**	

Table 2: IHPS 2019 and IHS5 2019/20 rainy season plots by maize cultivation status, conditional on being used for analysis

	IHPS 2019		IHS5 2019/20			
Season	2018/19		2017/18		2018/19	
Crop type	Obs	%	Obs	%	Obs	%
Maize	2,033	72.8	2,330	71.4	4,222	72.9
Non-maize	759	27.2	935	28.6	1,572	27.1
Total # of Plots	**2,792**	**100.0**	**3,265**	**100.0**	**5,794**	**100.0**
Total # of Associated Households	**1,470**		**1,926**		**3,506**	

Table 3: ESS 2018/19 *meher* season plots by georeferenced information availability

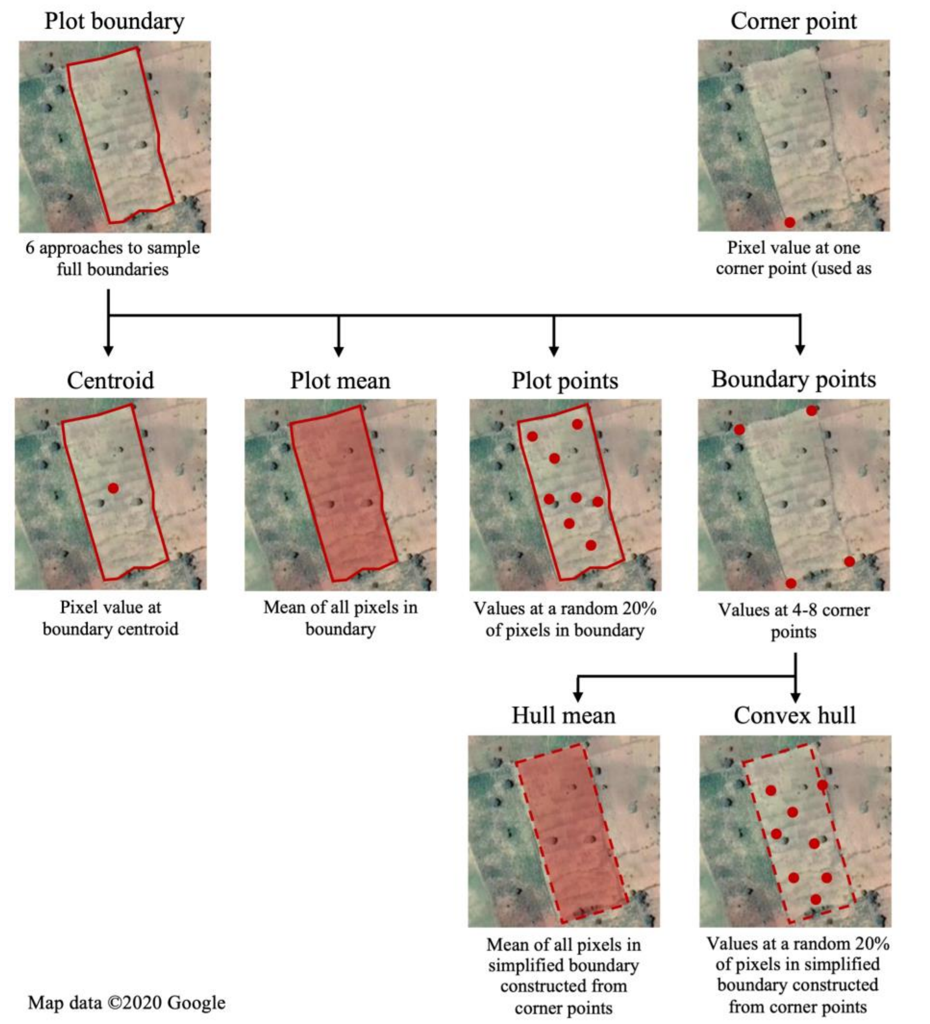
Plot category	ESS 2018/19	
	Obs	%
Plots with no geolocation information	1,168	8.7
Plots with a corner point, but dropped from analysis	299	2.2
Plots with a corner point, used for analysis	11,905	89.0
Total # of Plots	**13,372**	**100.0**
Total # of Associated Households	**2,199**	

Table 4: ESS 2018/19 *meher* season plots by maize cultivation status, conditional on being used for analysis

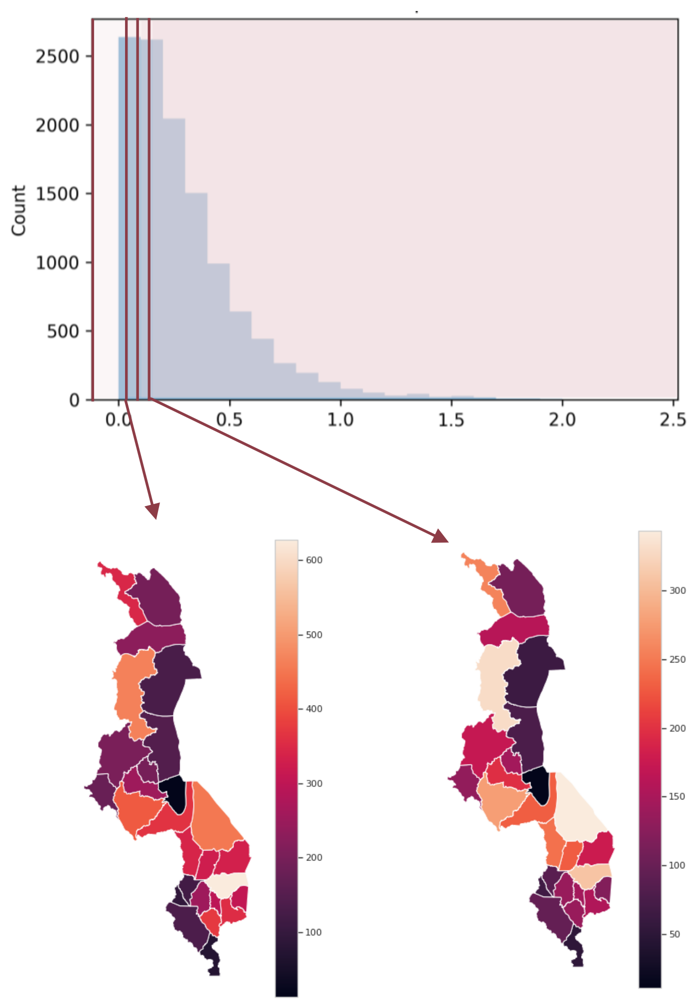
Crop type	ESS 2018/19	
	Obs	%
Maize	1,867	15.7
Non-maize	10,038	84.3
Total # of Plots	**11,905**	**100.0**
Total # of Associated Households	**2,090**	

Data Collection Scenarios

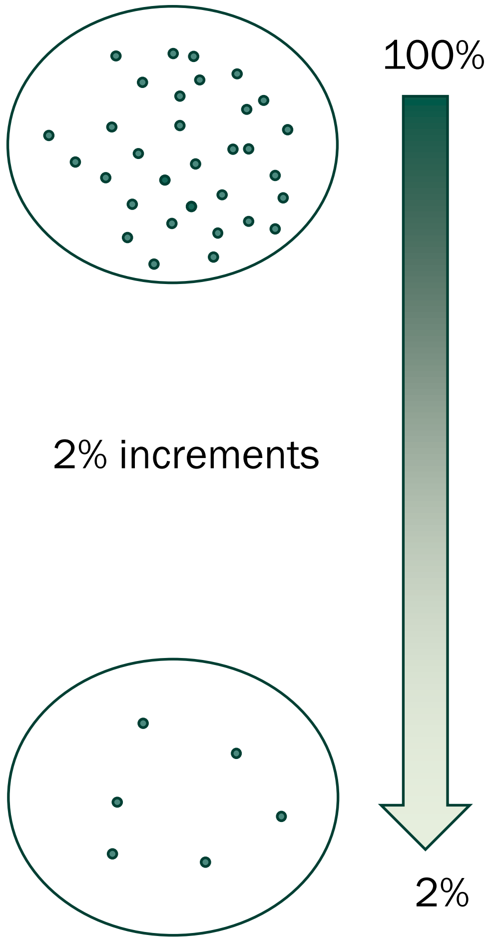
Geolocation



Field Size



Data Pool Size



Input Imagery and Phenological Metrics

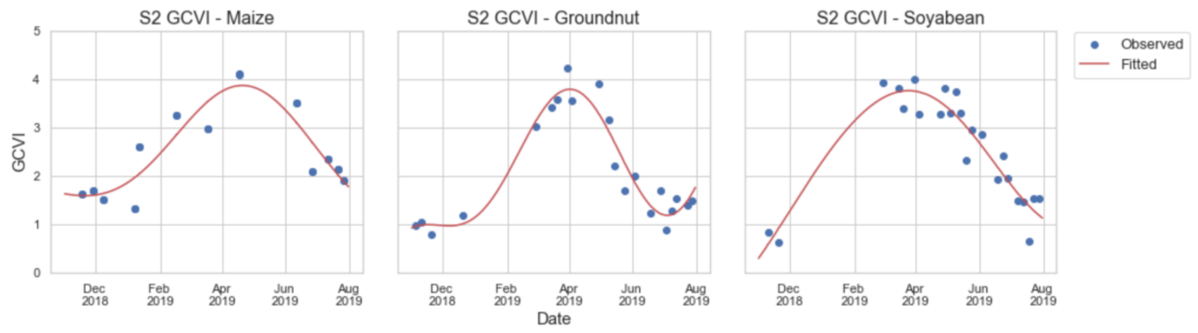
- Satellite observations from Sentinel 2 (optical features) and Sentinel 1 (SAR features), both at 10 m resolution.

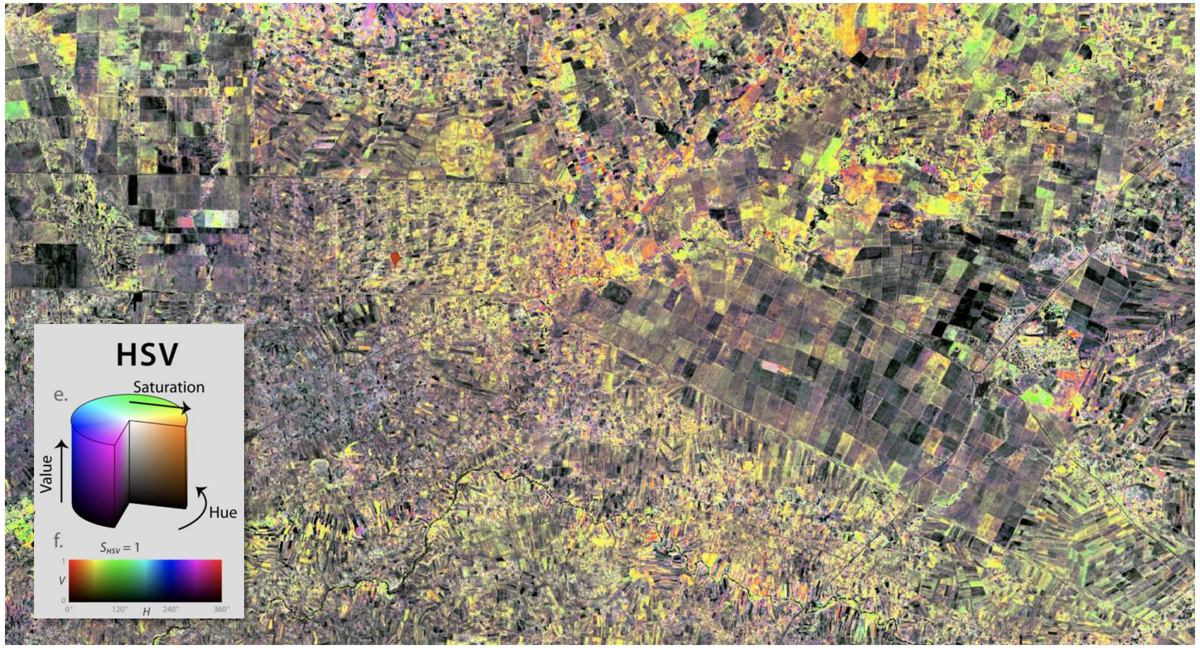
- We used harmonic regressions to process time series of satellite features.

- Harmonic regressions allow us to capture phenological metrics (phase and amplitude) by fitting a harmonic curve to the observations.

- Phase and amplitude are good differentiators of the seasonality of different crops.

- We also added topography and seasonal weather metrics.





Data Collection Scenarios and Modeling

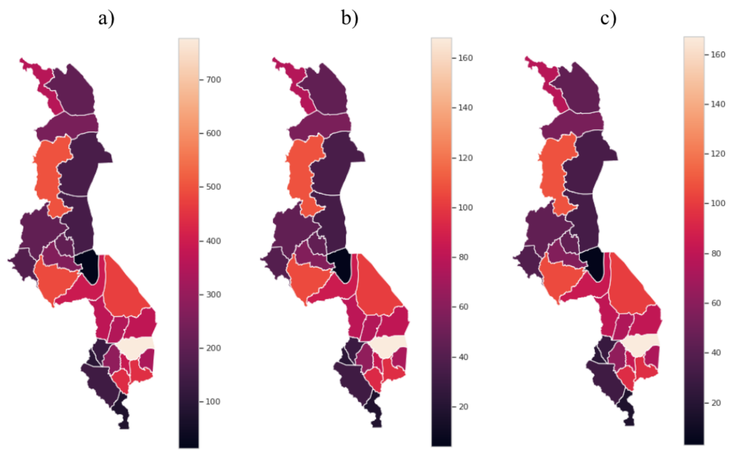
In Malawi, we tested a total of 26,250 scenarios:

- 7 geolocation methods - boundary points, centroid, convex hull, corner, hull mean, plot points, and plot mean.
- 50 data subsets - 2% to 100% subsets of training data, at an increment of 2% points.
- 5 area thresholds - 0, 0.05, 0.1, 0.15, and 0.2 ha.
- 3 feature types - optical only, radar only, both optical and radar.
- 5 replications to capture variability due to random sampling

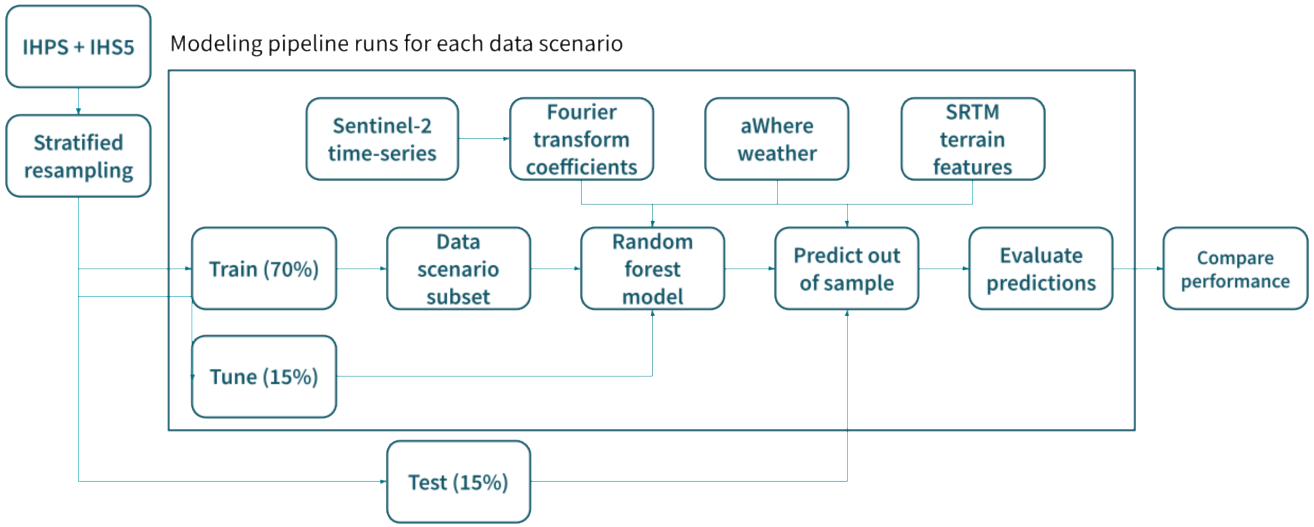
In Ethiopia, we tested a total of 250 scenarios – based on findings from Malawi and availability of only corner points:

- 1 geolocation method – corner point
- 50 data subsets - 2% to 100% subsets of training data, at an increment of 2% points.
- No area threshold, with optical data only
- 5 replications to capture variability due to random sampling

Figure 4: Distribution of plots across districts in a) train, b) validation, c) and test subsets

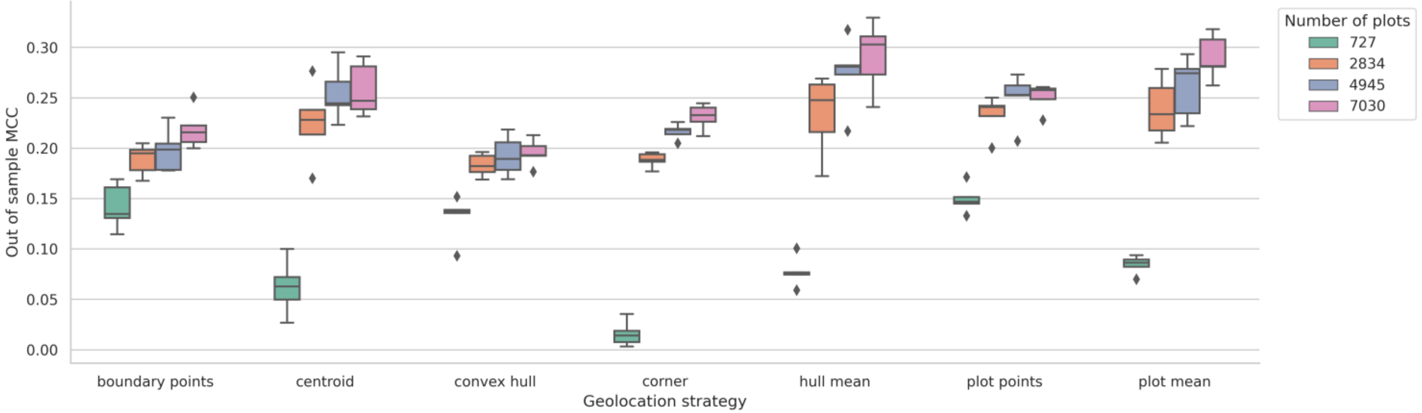


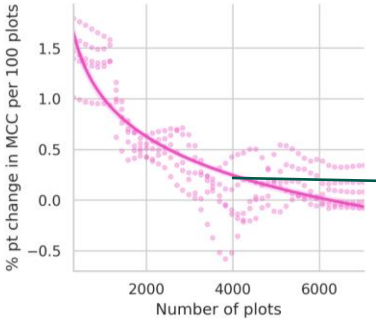
Notes: Color indicates number of plots.

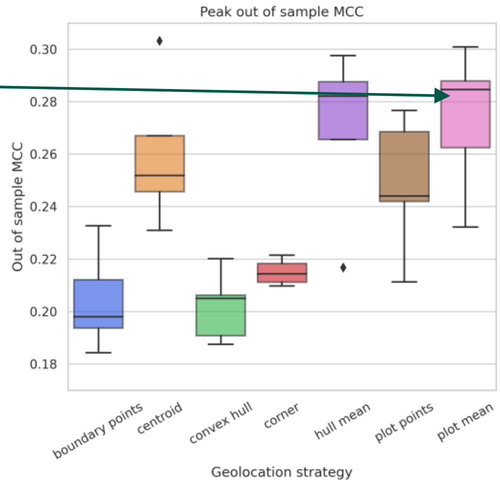


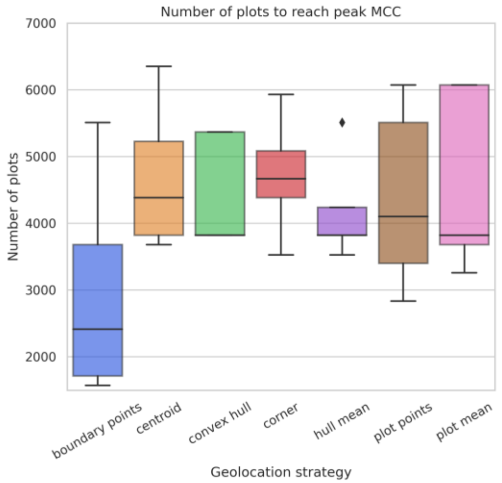
Results – Geolocation Methods and Sample Size

- *With less than 1,000 plots*: *multi-point approaches* perform better.
- Greater than 2,000 plots: *aggregation approaches - plot mean (based on plot boundary) and hull mean (based on all corner points)* - outperform all.
- Need about ~7,000 plots with a single corner point to reach performance with ~3,000 plots under *aggregation approaches.*
- Aggregation approaches had the fastest learning.
- Peak MCC can be achieved with ~ 60% of training data (~4,000 plots) under *plot mean* **(preferred)** and *hull mean* **(second best).**
- Corner point-based findings are comparable across Malawi and Ethiopia – though the peak MCC for corner point was reached at 3,000 plots as opposed to 4,500 in Malawi.
- Centroid method outperforms the single corner point method. If only a single georeferenced point can be collected by enumerators, that location should be near the center of the plot **(third-best).**



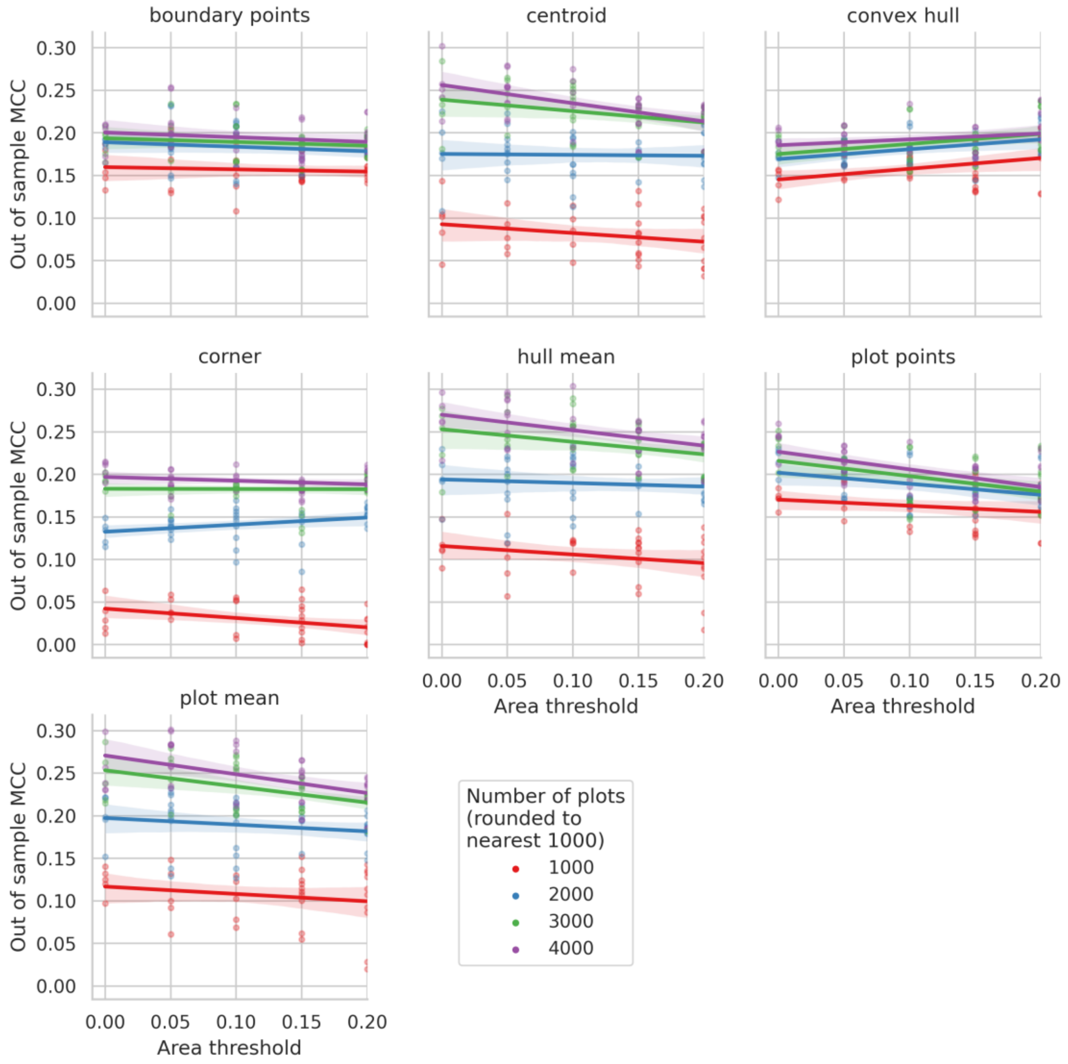






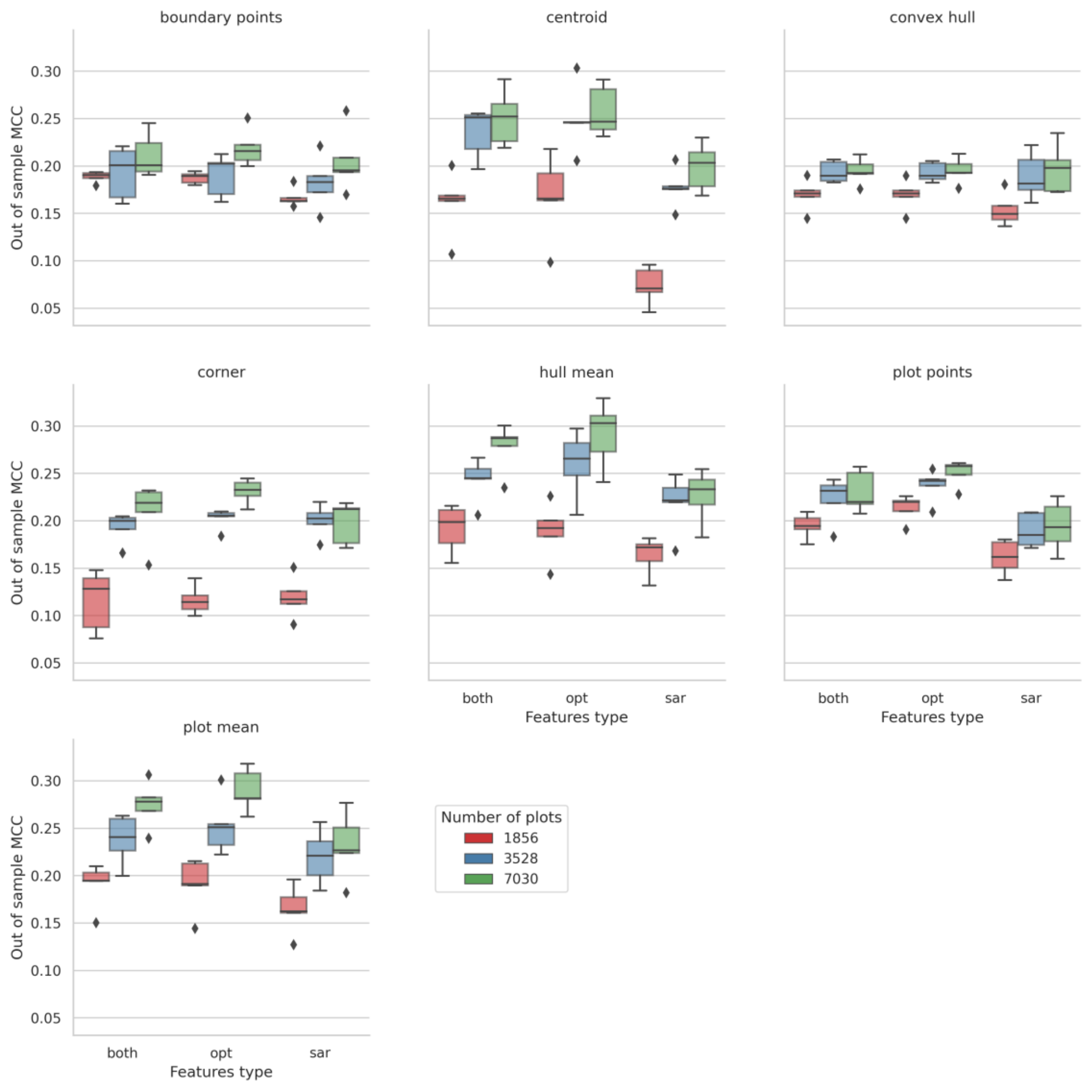
Results – Plot Size

- Limiting training data by excluding plots under specific area thresholds decreases performance.
- Exception: convex hull approach - likely due to geometrical approximation of plot boundaries.



Results – Satellite Type

- SAR alone generally lower performance.
- Optical alone highest.
- No gain in predicting power when combined.



Small Differences, Large Consequences

- Small differences in model performance may lead to large differences in estimated areas.
- Multi-points methods tend to overclassify.
- Aggregated methods are more conservative.
- There is value in achieving small performance gains anchored in better training data.

Table 8: Malawi maize area as obtained by seven different classification models

Classification model	Out of sample MCC	Total maize area - 2018/19 rainy season (million ha)
Boundary points	0.21	2.27
Centroid	0.24	2.17
Convex hull	0.21	2.46
Corner	0.23	2.15
Hull mean	0.25	1.94
Plot points	0.24	2.41
Plot mean	0.26	1.99
Mean across models	0.23	2.19

Classification model	Difference in out of sample MCC	Total area with disagreement (million ha)
Boundary points	-0.05	0.84
Centroid	-0.02	0.48
Convex hull	-0.05	0.69
Corner	-0.03	0.95
Hull mean	-0.01	0.22
Plot points	-0.02	0.55

Conclusions

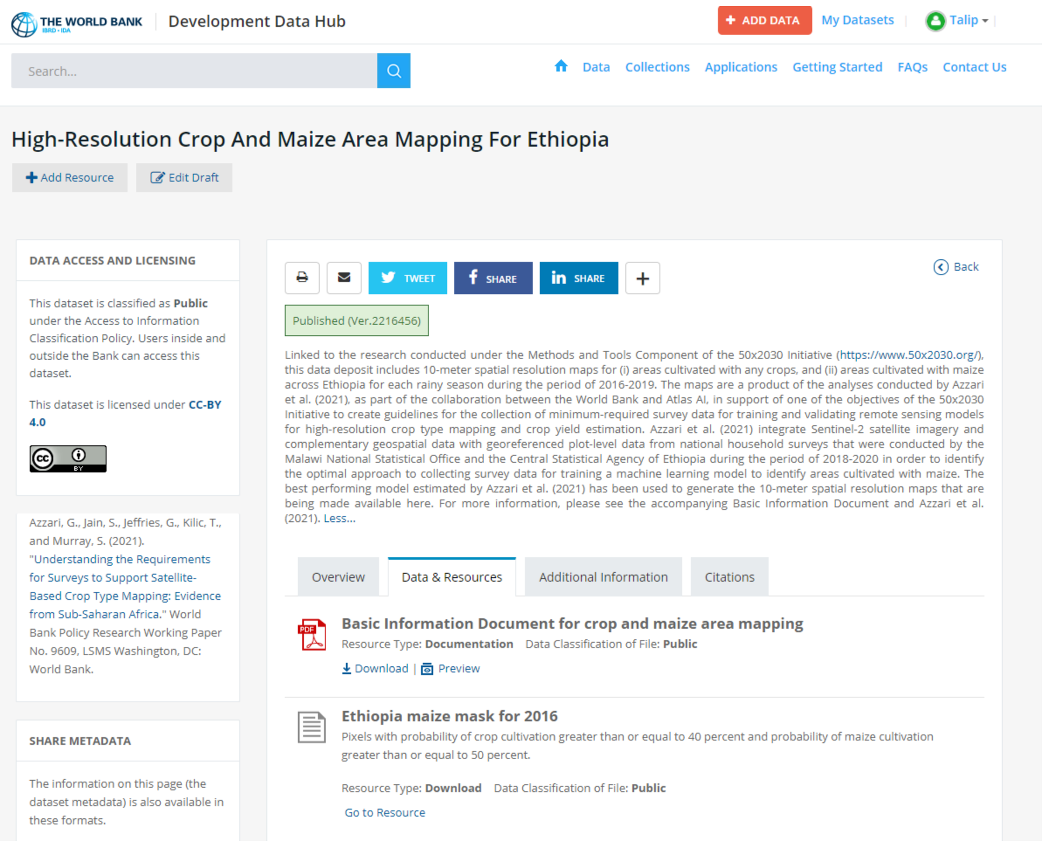
- Collecting a **complete plot boundary** is preferable to competing approaches to georeferencing plot locations in large-scale household surveys. This is particularly true if collection capacity is limited to fewer locations.
- Seemingly-small erosion in maize classification accuracy under less preferable approaches to georeferencing plot locations **consistently results in total area under maize cultivation to be overestimated** - in the range of **0.16 to 0.47 million hectares** (8 to 24 percent).
- Collecting GPS coordinates of the complete set of plot corners is a second-best strategy, can approximate full plot boundaries and can in turn train models with comparable performance.
- Classification performance peaks with ~60% of the training data under preferred and second-best approaches to georeferencing plot locations.
- If only a single GPS point can be collected, that location should be near the plot center rather than at the plot corner. With large datasets, the performance could be comparable to that of complete plot boundaries.
- No plot observations should be excluded from model training based on a minimum plot area threshold.
- Optical features alone can provide sufficient signal to maximize prediction quality.

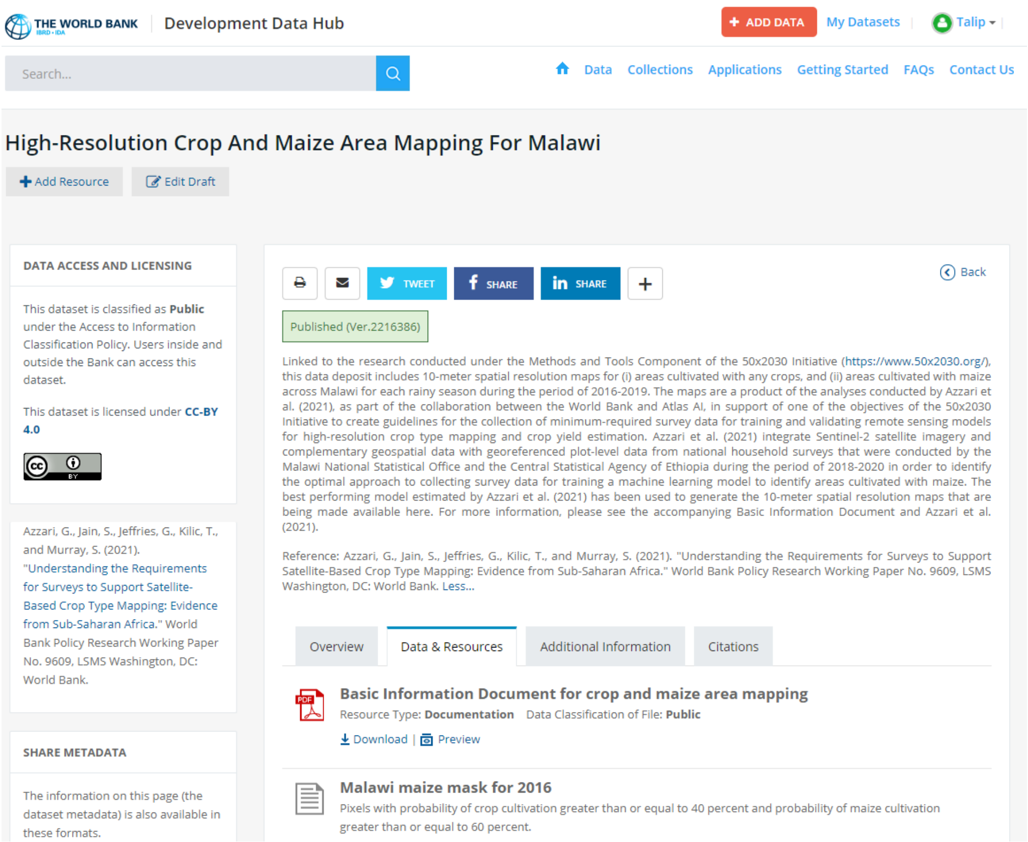
Looking Forward

- Continuing research under the 50x2030 Initiative to ultimately inform the guidelines for surveys to enable satellite-based crop type mapping and yield estimation in smallholder farming systems
  - Improving accuracy of maize classification, including through expanded suite of machine learning approaches and geospatial covariates
  - Leveraging additional existing large-scale survey data from **Mali**, **Malawi** and **Uganda** – with georeferenced plot outlines and objectives measures of yields based on *crop cutting* – to:
    - Expand crop classification to **new countries** and **new cereals**, including **sorghum**, **millet**, **wheat** and **rice,** to gauge the robustness of our recommendations
    - Conduct similar research to identify **training data requirements for high-resolution yield estimation** for maize and new cereals
  - Documenting (a) the accuracy of **out-of-season predictions** (e.g., using 2017/18 data from Mali to predict 2018/19 outcomes to be compared against actual 2018/19 data) and (b) the **decay in model accuracy** over time (i.e., over a 3-year period in Mali and Malawi)
  - Continuing research on object-based classification and **automated detection of plot boundaries** to potentially simplify ground data collection requirements
  - Depending on the COVID-19 pandemic, potentially conducting additional survey experiments in 2022 in non-African settings to increase the heterogeneity of farming systems subject to our research

Open Access Data Assets

10-m resolution crop area and maize area maps for Malawi and Ethiopia for each agricultural season from 2016 to 2019 on **World Bank Development Data Hub**



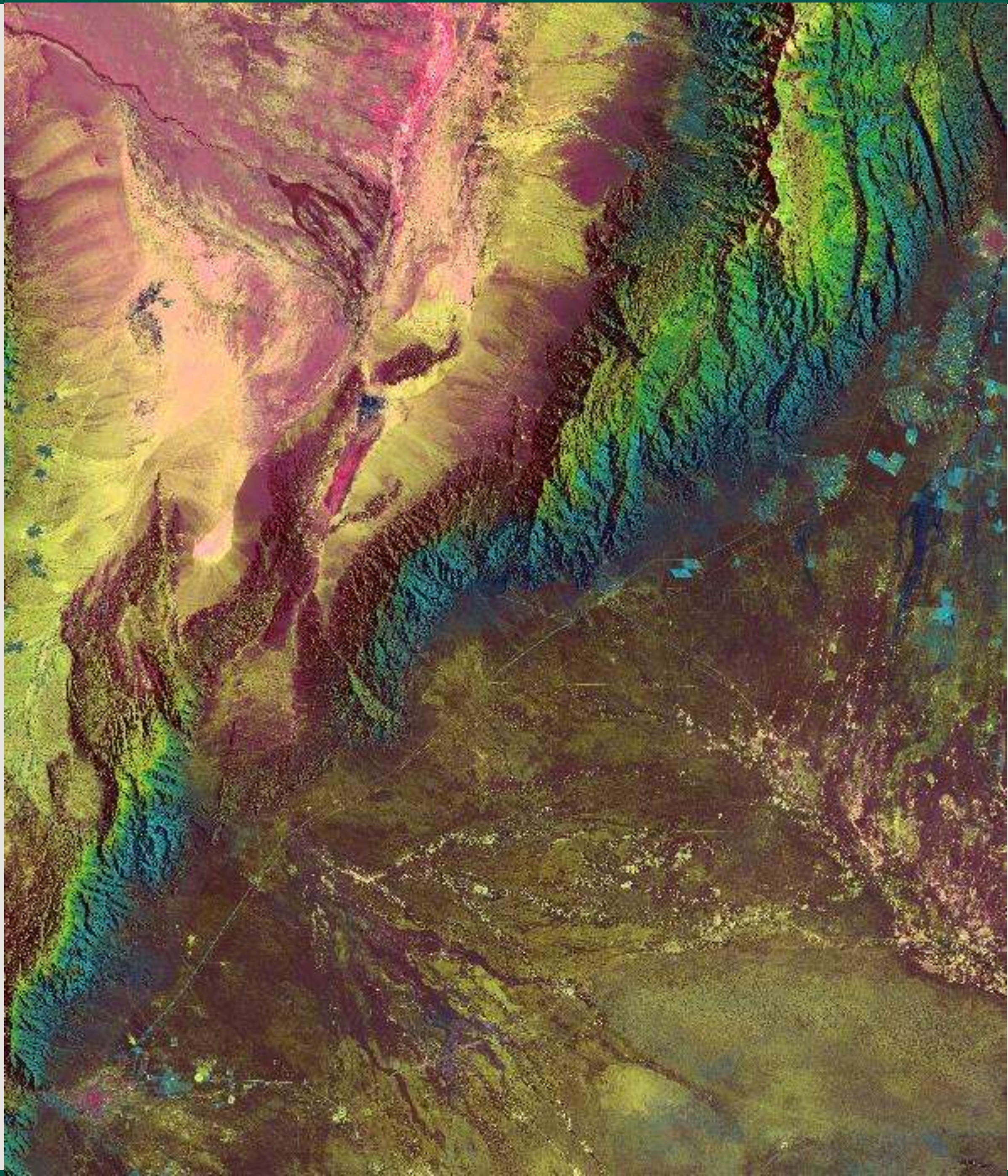


http://bit.ly/ethiopiamaps

http://bit.ly/malawimaps

DISCUSSANT

Catherine Nakalembe
Africa Program Director
NASA Harvest

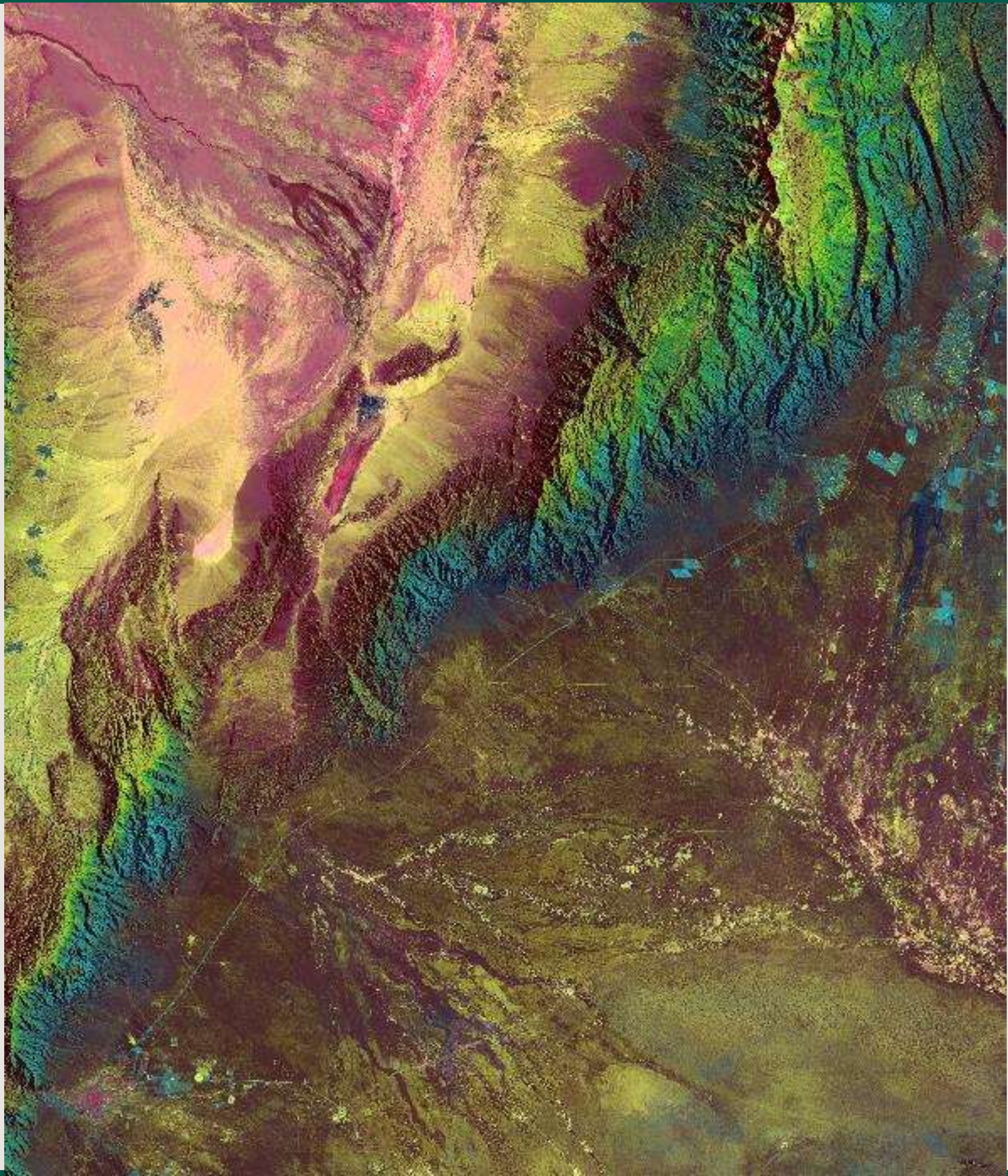


DISCUSSANT

Aberash Tariku

Deputy Director

Central Statistical Agency of Ethiopia



IMPLICATIONS FOR ETHIOPIA AND BEYOND

Highly relevant to **developing countries like Ethiopia** where **agriculture's contribution to economy is significant**.. Share of agriculture to GDP is 36.3%; rural population accounts for 78% of the total population

The research will:

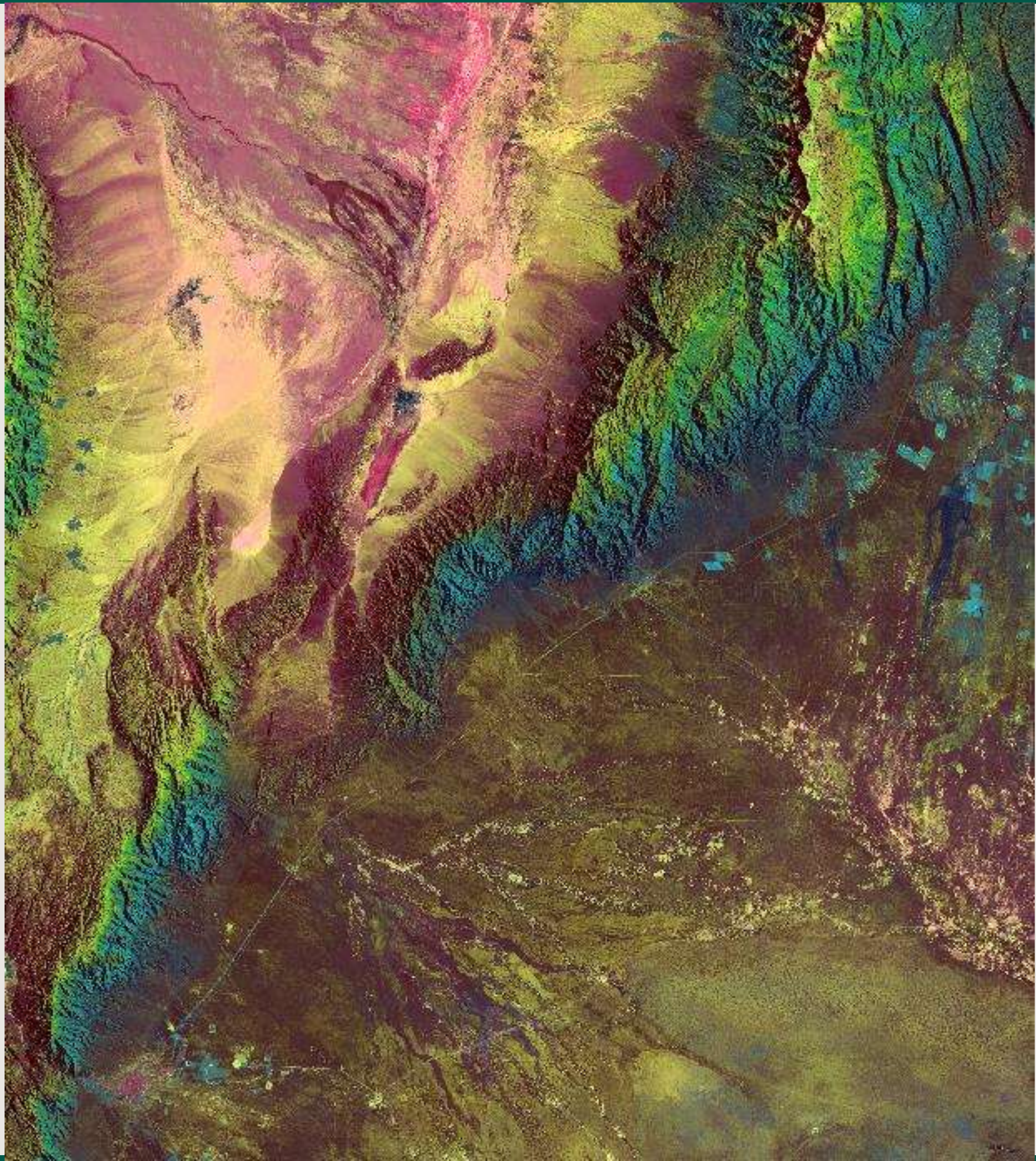
- Help provide more accurate agricultural data which will result in effective planning and monitoring in the sector
- Integrating remote sensing with survey data will provide geographically disaggregated data which is highly required by policy makers
- Help provide timely data specially for forecasting crop area and production and will also reduce workload in big farms

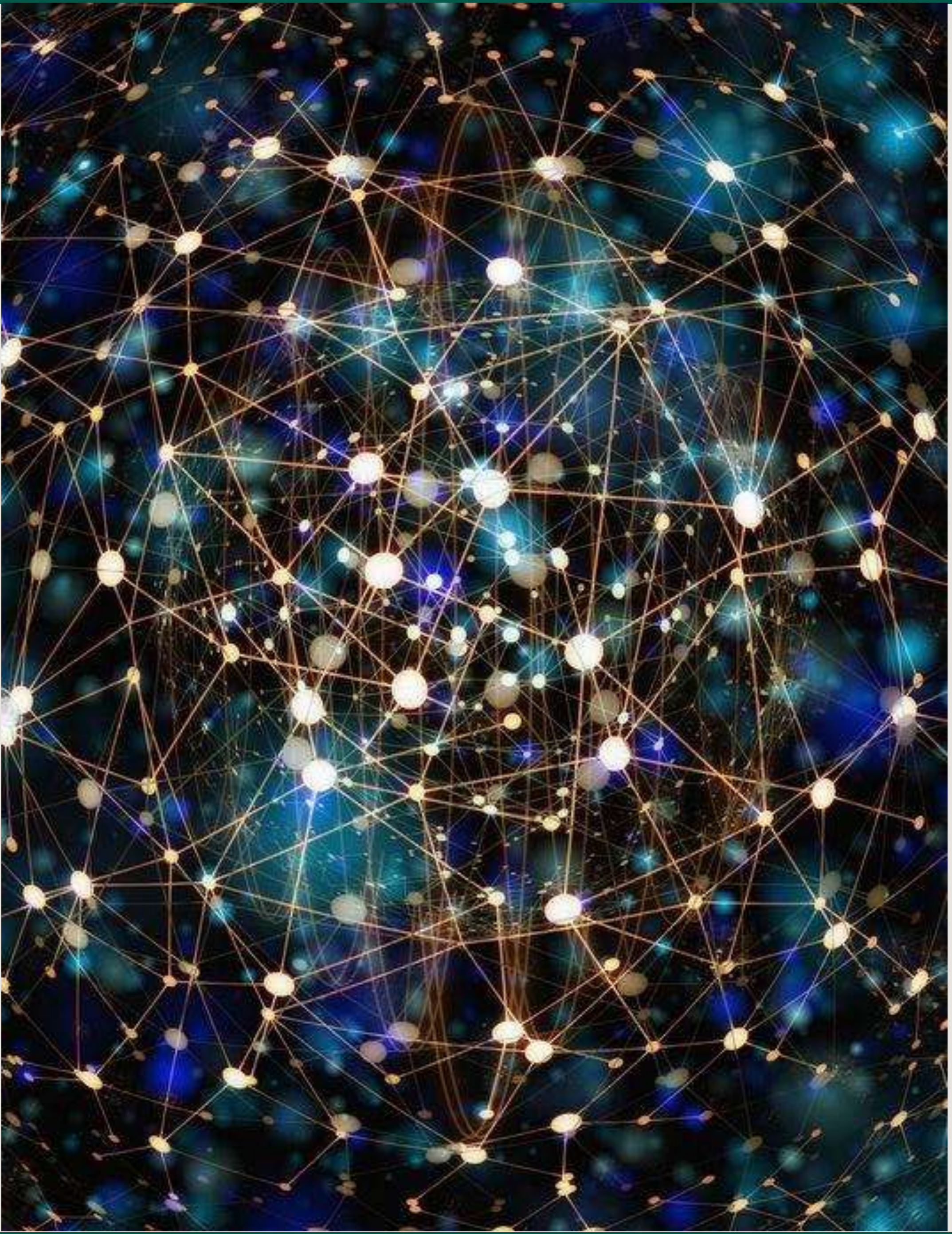
IMPLICATIONS FOR ETHIOPIA AND BEYOND

- Ethiopia will exercise remote sensing **to improve quality and timeliness of forecast survey and commercial farm survey** through 50x2030. These activities will be considered in next survey round
- In Ethiopia, official source of agricultural data is Central Statistical Agency. CSA and Ministry of Agriculture are working together to **facilitate implementation of the research planned for the use of remote sensing for forecasting and commercial farms survey.** Once the agreed methodology is designed, CSA will collect the data and provide timely and quality data for the Ministry to monitor progress
- **CSA and Ministry of Agriculture collaborating** within the 50x2030 Country Coordination Group. These activities are incorporated in the Program Implementation Plan of the project
- For effective utilization of the research, developing countries should be supported to produce **georeferenced survey data; satellite imagery should be availed; technical capacity required**

DISCUSSANT

Christophe Duhamel
50x2030 Data Production Manager
Food and Agriculture Organization





QUESTIONS
&
ANSWERS

JOIN THE CONVERSATION

#50x2030

50x2030
DATA-SMART AGRICULTURE

50x2030.ORG

Food and Agriculture Organization of the United Nations

SUPPORTED BY THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

IFAD
Investing in rural people

USAID
FROM THE AMERICAN PEOPLE

FEED THE FUTURE
The U.S. Government's Global Hunger & Food Security Initiative

BILL & MELINDA GATES foundation

Australian Government

Federal Ministry for Economic Cooperation and Development