50x2030
DATA-SMART AGRICULTURE

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

TALIP KILIC
Senior Economist
World Bank
tkilic@worldbank.org

Joint work with George Azzari, Shruti Jain, Graham Jeffries and Siobhan Murray

UN World Data Forum | October 6, 2021

Background

- Role of agriculture in rural livelihoods

- 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

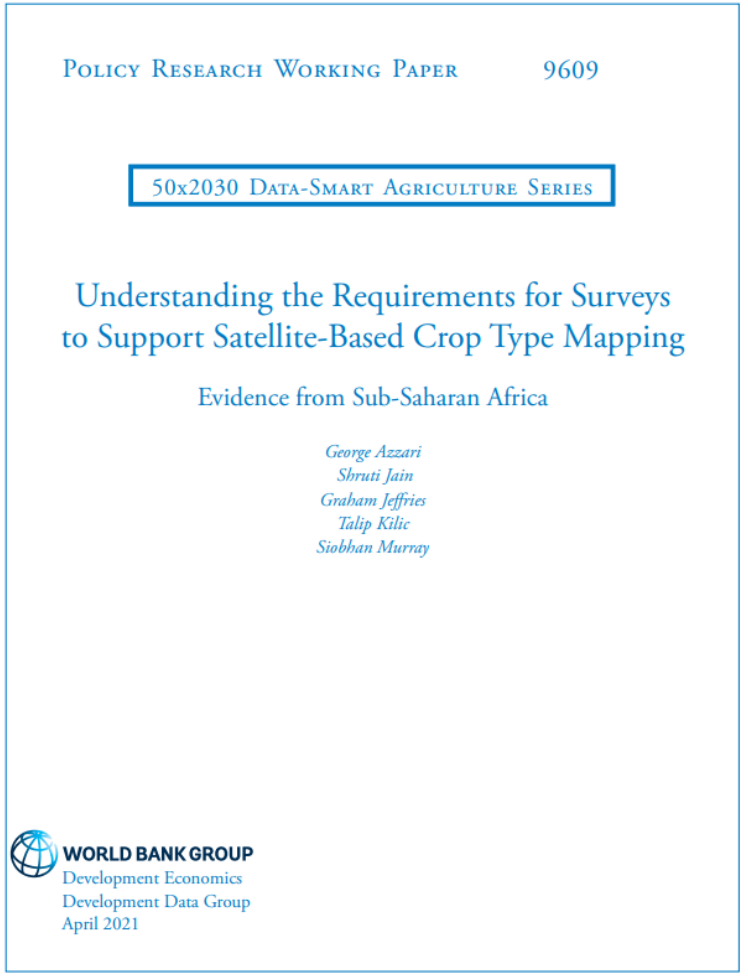
- Surge in high-resolution satellite imagery and research on remotely sensing outcomes in smallholder farming systems
    - Still need data to train and validate the underlying models

Key takeaways from the literature

- Training data have a bearing on the quality and spatial resolution of satellite-based estimates (Lobell et al. 2019, 2020)
- Research has largely been at sub-national levels, with heterogeneity in the type of and approach to training data collection
- Large-scale surveys *can* address training data needs of earth observation applications on crop area mapping and crop yield estimation in lower-income countries, but…
    - There are no clear recommendations on survey methods and fieldwork protocols to generate the right training data

What we do

- Address operational and inter-related research questions in the context of maize area mapping in Malawi and Ethiopia
  - How much training data do we need to reach an acceptable level of accuracy of a crop classification algorithm?
  - How does the approach to georeferencing plot locations in surveys impact algorithmic accuracy?
  - How do the type of satellite data and exclusion of plots under specific area thresholds affect algorithmic accuracy?
- Leverage ML, georeferenced national survey data collected by NSOs, and Sentinel-2 imagery + ancillary geovariables
- Inform the guidelines being developed by 50x2030 Initiative for georeferenced survey data collection for training and validating EO models for high-resolution crop type mapping and crop yield estimation in smallholder farming systems



https://bit.ly/wps9610

What we find

- 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).

- **Georeferencing 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.

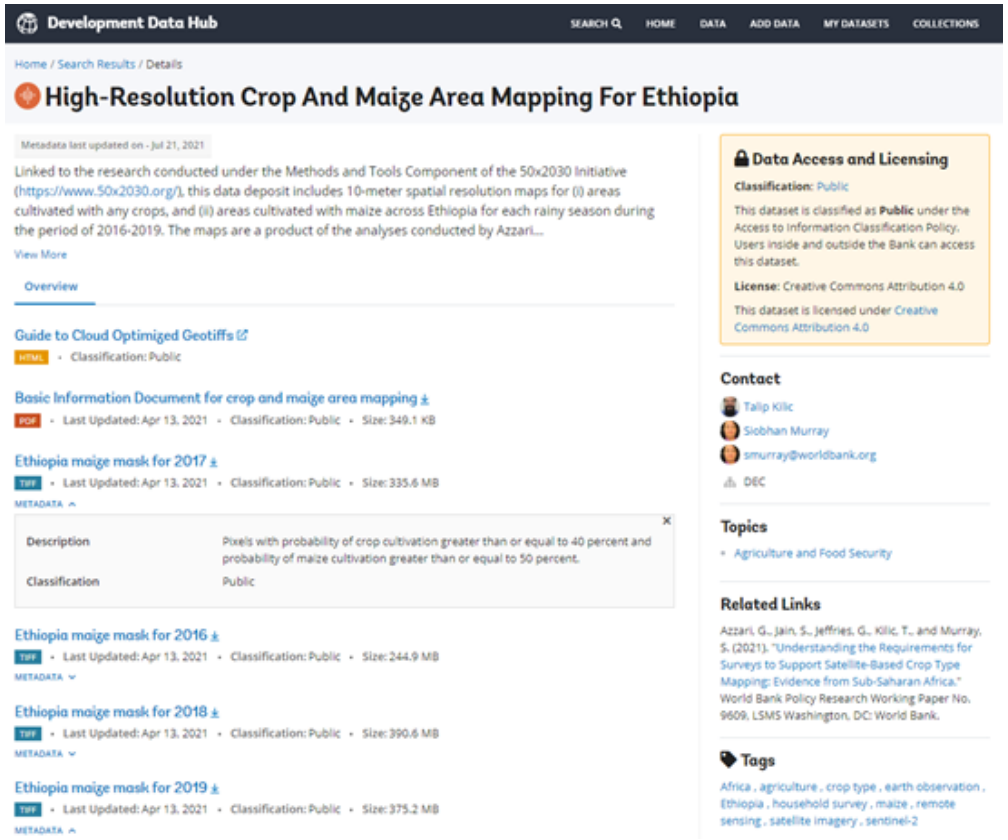
- **If only a single GPS point** can be collected, that location should be **near the plot centroid** rather than at the plot corner. With large datasets, the performance could be comparable to that of complete plot boundaries.

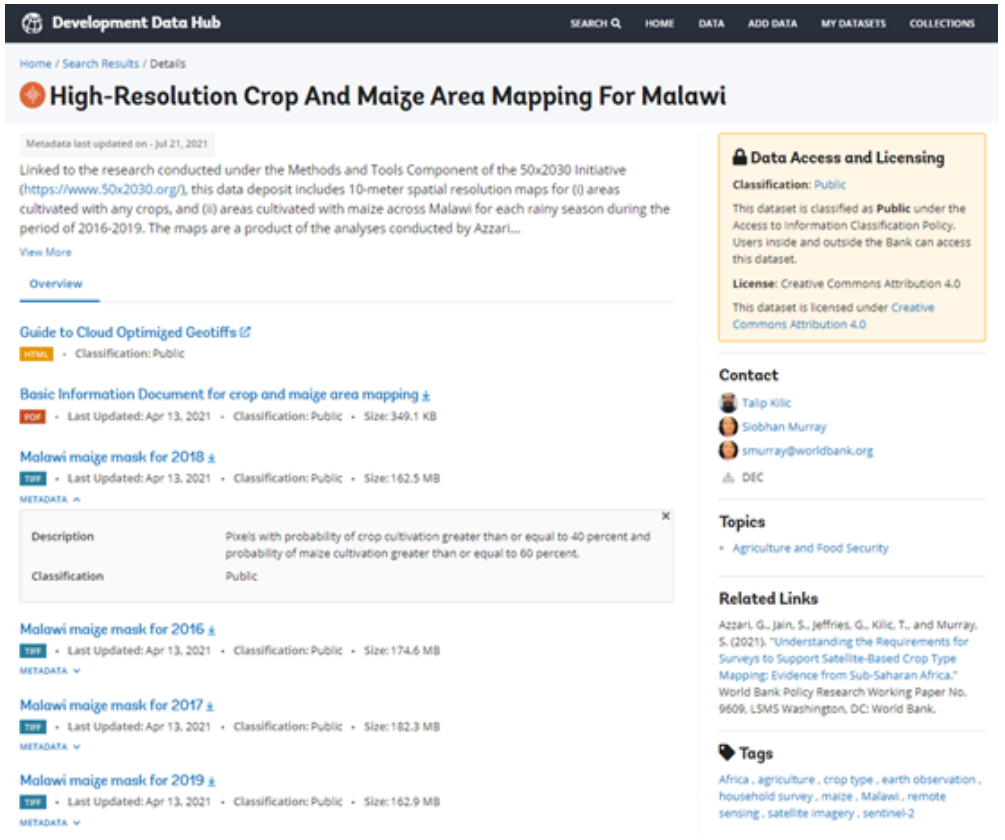
- **No plot observations should be excluded** from model training based on a minimum plot area threshold.

- **Optical features alone** can provide sufficient signal to maximize prediction quality.

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**



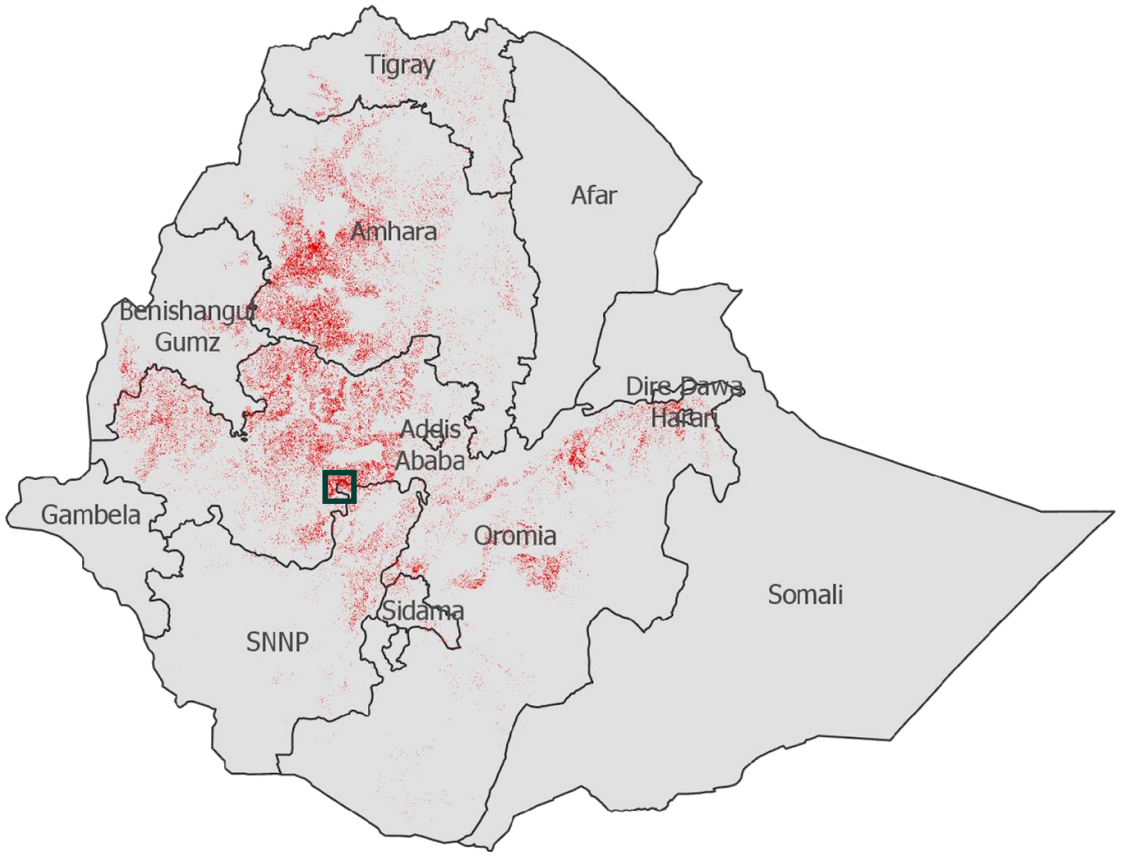


http://bit.ly/ethiopiamaps

http://bit.ly/malawimaps

Open access data assets

Ethiopia 2019: high-density production area



Looking forward

- Continuing 50x2030-supported research to formulate the guidelines for large-scale surveys to fulfill training data needs of satellite-based crop area mapping and yield estimation in smallholder farming systems
  - Leverage 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 & new cereals: barley, millet, rice, sorghum & wheat
    - Identify training data requirements for high-resolution yield estimation for maize & new cereals
  - Gauge sensitivity of recommendations to the choice of model (e.g., RF vs. CNN) & geospatial covariates
  - Document accuracy of out-of-season predictions & inter-temporal decay in model accuracy
  - Conduct research on object-based classification and automated detection of plot boundaries
  - If the COVID-19 pandemic allows, conduct additional survey experiments in 2022 in non-African settings

50x2030
DATA-SMART AGRICULTURE

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

TALIP KILIC
Senior Economist
World Bank
tkilic@worldbank.org

Joint work with George Azzari, Shruti Jain, Graham Jeffries and Siobhan Murray

UN World Data Forum | October 6, 2021