Time to sow: validating adoption of cropping practices with spatiotemporal analysis

What is your idea?

During the past five years, the Government of Ethiopia (GOE) has invested heavily in the promotion of improved management practices to increase cereal production. A key element of its strategy has been the promotion of wheat and teff row planting - a practice that lowers seeding rates and increases yields. Efforts led by the Ministry of Agriculture and Natural Resources, the Agricultural Transformation Agency, and regional agricultural bureaus have heightened awareness and knowledge about the practice. Further, the GOE has engaged with partners such as Digital Green to promote row planting through a community-centric video-based approach to the provision of agricultural extension. However, data on both adoption rates and yield impacts remain unclear, with wide variability in reported achievements at different agroecological and administrative levels. **This study aims to validate row planting practices across different agroecological zones using a machine learning algorithm with high resolution spatiotemporal remotely sensed datasets of the growing season.**

The study builds off of a randomized controlled trial (RCT) being conducted by the International Food Policy Research Institute (IFPRI) of Digital Green's video-based extension approach. The trial covers approximately 350 villages in 30 districts - and collects data from a sample of 2,500 households and plots—to explore various outcomes associated with the Digital Green approach.

While the RCT can provide insight into both adoption patterns and yield impacts, it faces several challenges. First, reliance on self-reported data from household surveys opens the possibility for several types of measurement error. Second, because distances between sample sites are great and transportation costs are high, household surveys are very resource-intensive undertakings with low potential for scalability. Third, the external validity of the trial's findings and the robustness of out-of-sample estimation of effects is constrained by the high level of agroecological and socioeconomic heterogeneity in the country.

There is considerable potential in using remote sensing to address these constraints—to increase precision, reduce research costs, and extend coverage area in a credible and replicable manner—and unchain adoption studies from acute dependence on primary data collection through personal interviewing. The rapid development of remote sensing technology especially the improvements of spatial and temporal resolution make it more suitable to monitor agricultural activities in a precise and timely manner. The very high resolution imagery (~0.5 meter) from Digital Globe allows researchers to not only distinguish crop types but also to monitor land preparation and other management practices. Imagery from the European Space Agency's Sentinel II satellite provides earth observations at 10-day intervals, allowing one to observe crop growth over the complete growing season. As space borne remotely sensed data becomes more accessible and affordable, and as machine-learning algorithms are more capable to handle and analyze big datasets, new opportunities to complement survey-based data collection methods are emerging rapidly. The case of row planting in Ethiopia offers a unique, timely, and policy-relevant opportunity to leverage these advances.

Preliminary studies have demonstrated that row and broadcast planting can be distinguished if observed at the right time (e.g., shortly after crop emergence) using Google Earth 1-m base map. Other studies have shown the capability to retrieve crop phenology using time series satellite data (e.g., Sentinel II). When combined with artificial neural networks, one of the most powerful machine learning algorithms, it is possible to extract features across both spatial and temporal dimensions.

We propose to use a machine learning algorithm with limited or no human intervention to detect and estimate the adoption rate of teff and wheat row planting in Ethiopia by taking advantage of the high resolution satellite data available. The outcomes of this study will provide more precise, credible, and replicable estimates of adoption that serve the informational needs of the GOE and its development partners. **The novelty of this approach lies in the analysis of remotely sensed data across both time and space at a large scale with specific attention to spatial heterogeneity in agroecological zones.**

How will you pilot it?

The project will be piloted as a partnership between IFPRI, Digital Green, and DigitalGlobe, with direct linkages to the GOE through the first two partners. The project will first secure data being gathered as part of IFPRI's RCT to identify locations of plots within the geographical area of interest. A subset of these plots will be randomized to serve as training and validation plots. The 10-day composite of vegetation index (VI) derived from Sentinel II will be applied to retrieve critical phenology information at the plot level (e.g. emergence date, flowering date, and harvested date) using a time series data analysis algorithm with a necessary curve fitting method. This information will then be used to acquire DigitalGlobe's high resolution imagery for the appropriate time to ensure detection of row planting. The time series of VI together with potential texture features will be integrated with DigitalGlobe images through a data fusion process. These spatiotemporal data are linked with plot field data from the survey. A convolutional neural network (CNN) will be introduced with prepared training (model development) and testing (model validation) datasets because CNN accounts for spatial relationships which is more suitable for spatial pattern extraction than conventional neural networks. We recognize that the spatial patterns associated with row planting represent a complex relationship. Thus, relying only on CNN might not be sufficient. Hence, post processing and cleaning will be investigated using probability analysis (e.g., Gamma distribution along the row, Gaussian distribution across the row) and a continuity test (e.g., Snake). After validation, the model will be automatically applied broadly to the area of interest to identify row planting.

The budget will be allocated primarily to analytical activities at IFPRI and GMU (80%), and to data acquisition and related services (20%) from partners. Data collection and analysis activities will take place over a 12-month period in the following order: (1) household and geo-referenced data collection; (2) ordering and analyzing time series of Sentinel 2 data for crop phenology; (3) acquisition of high resolution DigitalGlobe data and spatiotemporal data development. (4) detection of row planting using a machine learning algorithm.

This study will generate estimates on adoption of row planting, create algorithms applicable to comparable geographies with different cropping cycles, and produce a dataset of vegetative indices for monitoring crop growth cycles. The next steps include developing tools to package and automate the comprehensive approach to identify plot boundaries, detect row-planting, and estimate yields that improve on survey-based self-reported estimates.

The project team is comprised of Zhe Guo, David Spielman, Tanguy Bernard, Simrin Makhija, and Gashaw Abate from IFPRI, Liping Di from George Mason University, Rikin Gandhi from Digital Green, and Rhiannan Price from DigitalGlobe.