Crowdsourcing cassava yield data

What is our idea? We will pilot technology that enables cassava growers in Tanzania to collect and share crop yield data. Smart tools will assure data quality and incentivize farmers to contribute data by returning customized information on the yield and price evolution in their vicinity.

Insights in current crop yields at regional or national level, including variation in space and time is extremely valuable to the research community, policy makers, private sector enterprises engaged in processing and value addition, and the growers themselves. Yield data is invaluable for climate change research, crop modelling and the development of decision support tools for agricultural technologies. This research can in turn be used to anticipate and mitigate shortages in food supply, for example. Yield data also enables value chain actors to make informed decisions to improve efficiency and increase revenue of their production systems.

Satellite technology has gone a long way in forecasting yields for crops such as maize and soybean. For root crops, and particularly for cassava, much less progress has been realized. This is because the relationship between root yield and canopy features is highly dependent on environmental conditions during crop growth, and because of very large planting and harvesting windows, resulting in very diverse crop ages at any given time. For cassava, field observations are a must.

Researchers and national statistics services go to great length to obtain georeferenced datasets of crop yields. This is a tedious, expensive and time-consuming task, typically carried out by teams of trained and equipped enumerators. For cassava, this is an even more strenuous task due to its wide planting and harvesting window, often requiring almost year-round presence to conduct representative yield measurements. For these reasons, large cassava yield datasets remain an elusive target.

We aim to overcome this problem by applying citizen science approaches to acquire yield data at scale. Such approaches have shown that technology can make scientists of us all. Due to the breakthrough of mobile technology and the increasing penetration of smartphones and internet services, even in the smallest villages, an opportunity arises for volunteer farmers to directly collect and share yield data.

How will we pilot it? The technology developed will rely on a simple yield assessment protocol (already piloted), combined with advanced statistical approaches for yield estimation and mobile technology. Strategies will be put in place to overcome challenges related to data quality assurance, and how to reward the efforts of data contributors.

A recently piloted, simple yield assessment protocol has shown much promise to enable farmers to accurately measure yield. We will start by training extension workers to apply this protocol, and equip them with simple tools (hanging spring scales and tape measures) [month 1]. The method relies on a measurement of crop density, and recording the root weights of a few individual plants, supplemented with quick and simple plant canopy measurements. Extension workers will then carry out the assessment as part of their normal duties, as they visit and advise farmers throughout the year.

Incoming data will be fed into libraries [months 2-3], and the relationship between root yield and canopy features will be used as prior knowledge in advanced statistical models (Bayesian machine

learning) to improve the precision of yield estimates, as well as assess the reliability of submitted data [months 4-5]. Preliminary results and simulations have confirmed that this can be achieved with the proposed methodology, though distinctions need to be made for different varieties or growth types, which is feasible with sufficient data. Models will be improved as data accrues in the libraries [months 5-6]. The yield assessment protocol will be improved [months 5-6], and simplified to ensure the target end-users can obtain sufficiently accurate yield estimates in the easiest and quickest way possible. Yield data collection will then continue using this improved protocol [months 6-12].

Data entry will be done in the field using smartphones and free, open-source software (ODK) to capture yield parameters as well as basic crop cultivation parameters and GPS location. Data collection forms will be developed and demonstrated during training events [month 1]. The data will be submitted and aggregated in an open-access database, then automatically processed using the above-mentioned statistical procedures (implemented in R), and visualized (using R-Shiny) [months 2-6].

In a last step, we will evaluate how yield information is best packaged to be of practical use to data contributors, through a set of surveys and key stakeholder interviews [months 7-8]. This will then be provided to a software engineer to develop a prototype smartphone app with a user-friendly interface [months 9-12]. We will also identify supplementary output information to include in the app as an incentive for cassava growers to submit yield measurements. This could include a comparison with yields obtained in their vicinity, predictions of the evolution in yield given expected rainfall in the next 1-2 months, and the evolution of the storage root price at nearby markets, aiming to empower users and improve their position in negotiations with middle men and transporters to market their crop, and make decisions on the optimal timing of harvest.

The main outputs of this pilot will include (i) an optimized yield assessment protocol tested by laymen farmers, (ii) a suite of analysis tools and models to obtain yield estimates (with errors), (iii) a geo-referenced yield dataset, (iv) a prototype smartphone app for scaling beyond the target areas, and (v) insights in how to incentivize farmers to contribute yield data.

The African Cassava Agronomy Initiative (ACAI) led by IITA will provide expertise, exchange learnings and avail baseline datasets on cropping practices and storage root prices in the target areas. The technology will be tested within the network of cassava growers and processors of C:AVA-II, that will provide training and support through their network of skilled extension workers.

If the pilot is successful, the application will immediately be used within the wider network of C:AVA-II in Tanzania, and ACAI partners operating in similar agro-ecologies in Tanzania and Nigeria. Scaling experts within IITA will be contacted to evaluate options to improve the technological readiness and elaborate a strategy for further development of the app (coding, interface development, and applying efficient approaches for testing and validation beyond the pilot target areas), as well as scale the use of the tool by identifying and addressing gaps in the potential client network, and promotion events. The application can also be expanded to other crops.

Funds will be shared 50/50 between IITA and C:AVA-II, and serve to purchase tools and smartphones ($24,000), facilitation and travel of extension workers ($10,200), organize training events ($21,300), conduct field monitoring ($22,600), a software development consultancy for the app interface ($10,000) and institutional overhead ($11,900). Staff time will be covered by ongoing initiatives.