

BETTER DATA, HIGHER IMPACT:
Improving Agricultural Data Systems for Societal Change

Gero Carletto
Development Data Group
World Bank
gcarletto@worldbank.org

Setting the scene

- Based on European Review of Agricultural Economics paper and upcoming chapter in Handbook of Agricultural Economics
- Spotlight on:
 - low and middle-income countries
 - smallholder farming
 - household and farm surveys
- Background and motivation ... and a few opening tenets!
- Quick review of recent developments in agricultural data measurement
- “Manifesto” on how we can contribute to improving agricultural data systems

We hold these truths to be self-evident . . .

- Want higher impact? Start with **better data!**
 - Increasing the impact of agricultural economics research requires better agricultural data (and a **new paradigm!**)
- Stop focusing on *ex-post* solutions and start thinking more about ***ex-ante* design choices**
 - Better data requires shifting the balance from *ex-post* adjustments based on econometric models and data treatment to *ex-ante* data collection design choices
- **Incentives** make the world go round
 - Individual researchers have the power to improve agricultural data systems, but need to be incentivized accordingly
 - Development partners and professional associations can help create an enabling environment to support individual researchers and mitigate disincentives

The world of agricultural data (demand side)

- World experiencing massive, interconnected changes, requiring new and better data
 - Climate change, demographic transitions, increasing inequality, economic transformation
- Smallholders as producers, consumers, processors, and traders in ever more complex value chains and food systems
- Many gaps persist in terms of foundational data
- Yet, greater awareness of data limitations in both availability and quality
 - More demand for evidence
 - Burgeoning literature on measurement issues, more emphasis on *ex-ante* choices

The world of agricultural data (supply side)

- Technological innovation driving continuous improvements in data collection methods
- New data sources offer opportunities for more granular data at lower cost
- New modes of data collection increasingly used
- Greater financial commitment to supporting data production and innovation

Is demand outpacing supply?

What's the right balance between strengthening foundational data and pushing the frontier?

How can we contribute as a profession?

What's holding us back?

- Instruments are often **unfit-for-purpose** (e.g., COVID-19, extreme weather shocks, ...)
 - Recall methods can't handle the complexity of smallholder agriculture
- Data lack adequate temporal and spatial **resolution**
- Poor **integration and interoperability** between datasets
- Experiments are conducted at insufficient scale, lacking **external validity**
- Lack of investment in establishing **best practices** and methodological standards
- Scarcity of **analytical capacity**
- Inadequate **skill sets** on new areas like machine learning and Earth Observation
- Low **credibility** of data
- Limited **data access**, including lack of transparency and protocols
- Inability to **quantify the benefits** of data (including quality)
- **Institutional constraints** to interdisciplinarity

What's propelling us forward?

- **New data initiatives**, spurred by the 2008 World Development Report: *Agriculture for Development*
 - Global Strategy to Improve Agricultural and Rural Statistics
 - Living Standards Measurement Study – Integrated Surveys on Agriculture (LSMS-ISA)
 - FAO Agricultural Integrated Surveys (AGRIS)
 - 50x2030 Initiative to Close the Agricultural Data Gap
 - 2021 World Development Report: *Data for Better Lives*
- New, affordable **technologies**
- New **modes** of data collection
- Expansion in the availability and use of **new data sources**
- Innovative use of **imputation methods**

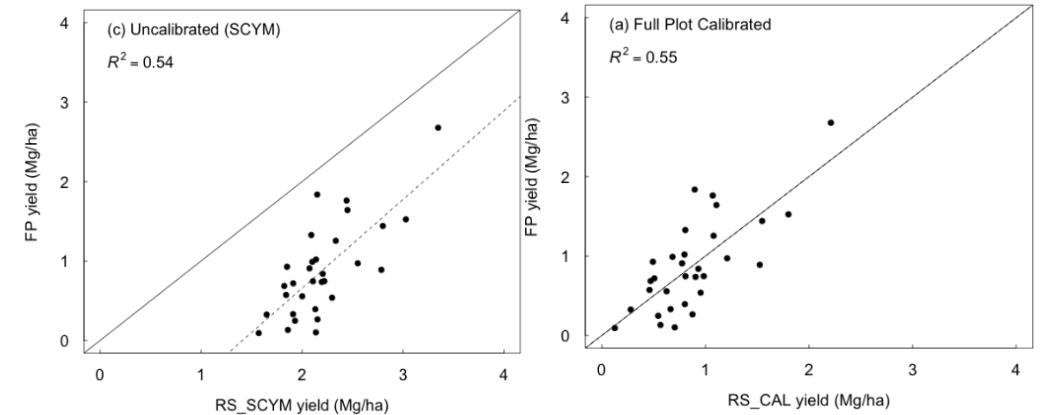
From Guesstimates to GPStimates to RemoteSenstimates!

- Growing literature on non-classical measurement error in **farmers' self-reported land area measurement**, proposing the use of GPS at scale (Carletto et al., 2013, 2015, 2017; Abay et al., 2019; Desiere and Jolliffe, 2018)
- **GPS as a viable, scalable solution**, although field visit costs and logistics may still pose challenges
- Alternatives include:
 - Using **maps** for plot boundary delineation (Masuda et al., 2010; Dillon and Rao, 2018)
 - **Imputation** methods (Kilic et al, 2017a, 2017b)
 - **Earth Observation** (Azzari et al., 2021; Gourlay et al., 2019)



Blue skies, muddy fields! Measuring crop yields from space

- Measuring agricultural productivity accurately in smallholder systems is key to designing policies to reduce poverty and food insecurity
- Stream of work to assess the accuracy of high-resolution remote sensing images for estimating plot-level crop yields (Lobell et al., 2019; Gourlay et al., 2019)
- **Calibration** of remote sensing models with georeferenced, objective plot-level ground survey data shown to reduce bias and **overestimation in maize yields in Uganda by 1 ton/ha**
- Defining minimum requirements for **scaling-up ground data collection** in large-scale surveys to inform satellite-based estimation of crop area under cultivation and yields (Azzari et al., 2021)



The sensors are coming!

- Ubiquity of **affordable sensors** radically changing measurement agenda
- Many applications: soil fertility, crop variety, time use, labor productivity
- Farmers' **self-reported** soil quality measures are **highly inaccurate** (Carletto et al., 2017; Gourlay et al., 2017)
- Use of portable spectrometers for calibration of Africa Soil Information Service data (Hengl et al., 2015)
- Wearable accelerometers to improve the accuracy of labor inputs in agriculture (Akogun et al., 2020)
- Next steps involve **scaling up**:
 - large samples or small sample with imputation
 - cooperation in production of gadgets



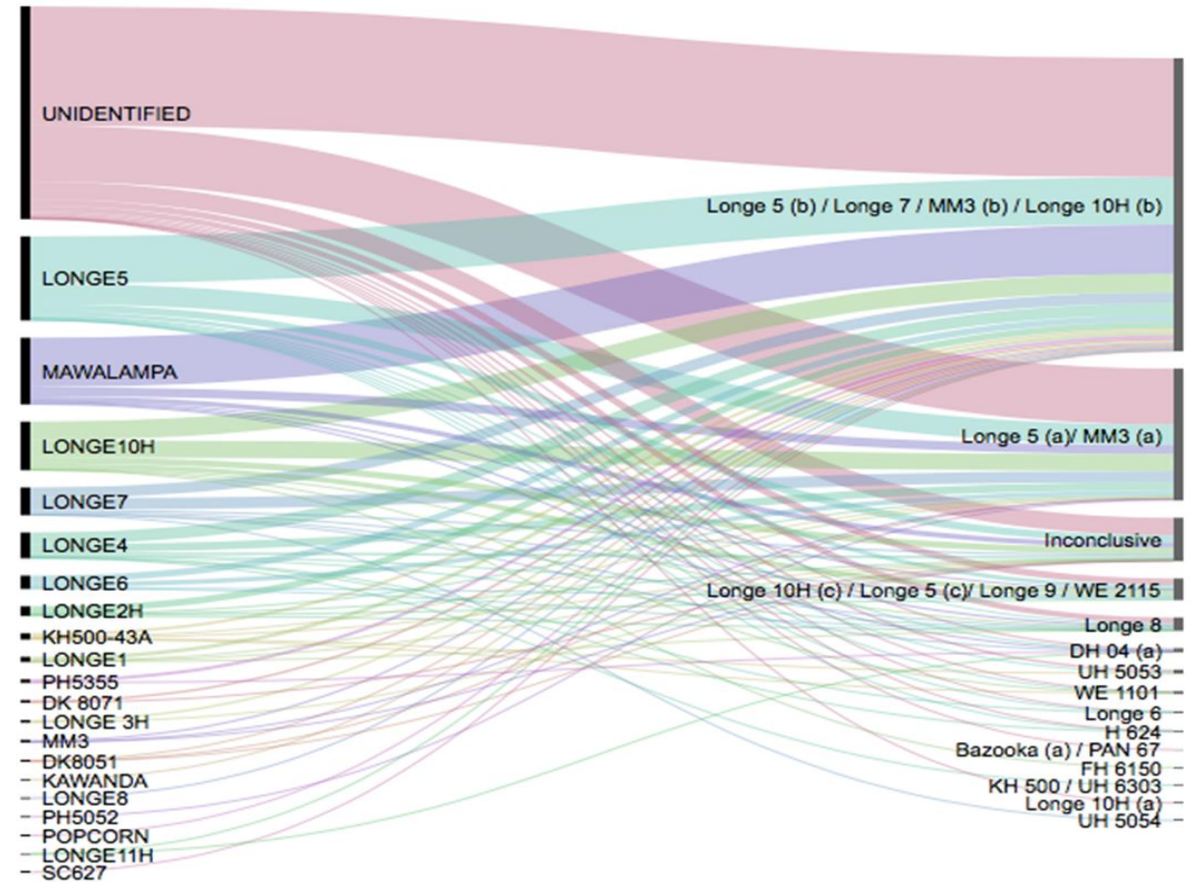
DNA don't lie!

- **Counterfeit seeds are common** – this has huge implications for input applications and yields
- Both farmers' self-reporting and expert opinions are **highly inaccurate**
- While infallible, the **cost and logistics** of DNA fingerprinting are still prohibitive, but feasible
- **Scale** remains a challenge - approach best used in combination with imputations on smaller samples



Farmer-Reporting

DNA Fingerprinting



Source: Ilukor, J., et al, ...)

Dial A for Agriculture, B for Bias!

- Phone surveys are proving to be a viable option to collect data on high-frequency, non-salient **events** that suffer from **large recall bias** (Arthi et al., 2017)
- COVID-19 exposed the possibilities and challenges of using phone surveys, including **non-response, attrition, and low coverage** both at the household and individual levels (Kilic et al., 2021; Brubaker et al., 2021)
- **Mode effects** may be pervasive
- Looking ahead, the use of **mixed-mode data collection** is being pursued at scale, but potential biases must be addressed



Data of the people, for the people, by the people!

- **Citizen-generated data** (CGD) and crowdsourcing are still underused in agricultural data, but contain great untapped potential (Fraisl et al., 2020)
- Some applications documented for **price data** (Zeug et al., 2017; Ochieng and Baulch, 2020), **climate data** (Minet et al., 2017; Van de Giesen, 2014) and **soil data** (Herrick et al., 2013)
- Many challenges remain in terms of **data quality, representativeness, and validation protocols** related to the recruitment, motivation, and retention of informants (Ebitu et al., 2021; Wiggins et al., 2011)
- **Interdisciplinary requirements** and **institutional constraints** add to the challenges
- Combining the strengths of CGD with the statistical rigor of probability sample surveys may have one of the most **consequential impacts** for the evolution of agricultural data systems



Gender-disaggregated data for gender-sensitive policy!

- Estimations of gender differentials in agricultural productivity have been shown to vary both across and within countries (World Bank and One Campaign, 2014; Kilic et al., 2015)
- Collecting agricultural data from “the most knowledgeable respondent” often results in asking the head of the household, usually male – **respondent bias** in gender-disaggregated data is widespread and well-documented
- **LSMS-Plus**: collecting individual-disaggregated data on asset ownership and labor
- Mainstreaming the collection of individual-level data will require a collective and coordinated effort by key stakeholders ... and funding!



Understanding transformation needs panel data!

- Addressing global challenges like **climate change** and informing policies for achieving **inclusive economic transformation** requires longitudinal data
- **Some progress**, with more than a dozen countries with regular panel data collection
- We need to both sustain and expand existing panel efforts – also through **inclusion** of more academic researchers, particularly those from low-income countries, in both the design and analysis stages
- The impacts of **extreme weather events** are best captured by establishing and maintaining long-term longitudinal data collection as an integral component of national data systems
- **High-frequency phone panel surveys**, built on existing pre-Covid19 panels, have been crucial for tracking the socioeconomic and health impacts of the pandemic to inform national responses



Achieving greater impact: the role of researchers

- Increase opportunities for publishing **methodological research** on data in top journals
- Reduce **publication bias**
- Make the right **design** choices when planning your data collection
- Think beyond your own, immediate objectives to the **public good**
- Systematically integrate **methodological experiments** into data collection
- **Document and disseminate** data for greater use and re-use
- Communicate findings, highlight **data quality improvements**

Achieving greater impact: the role of institutions

- Develop and expand use of **metrics for data quality**
- Create platform for **data sharing**
- Develop tools and protocols for **data anonymization**
- Incentivize the **use and re-use** of data
- **Recognize** and **reward** the production of high-quality data
- Encourage the inclusion of **methodological experiments** when funding data collection
- Facilitate **global collaboration** across academic institutions
- **Reduce transaction costs** to making data and research available

Concluding thoughts

- The “**credibility revolution**” in empirical (agricultural) economics of Angrist and Pischke starts with better data!
- **Incentives** to focus on agricultural **data quality** have been historically misaligned, but some progress has been made in recent years
 - More focus on *ex-ante* **design choices** rather than *ex-post* treatment
- Agricultural economists have a role to play in contributing to a “**social contract for data**” as called for in the 2021 World Development Report
 - **Fit-for-multi-purpose** data collection
 - Systematic integration of **methodological experiments**
- **Technology** and **new data sources/modes** are critical, but can only go so far without proper investment in infrastructure and capacity development
- Overcome institutional hurdles for **interdisciplinary research**
- Reduce cost of collaboration, **agricultural data as a public good**

BETTER DATA, HIGHER IMPACT:
Improving Agricultural Data Systems for Societal Change

Gero Carletto
Development Data Group
World Bank
gcarletto@worldbank.org