When administrative data isn't enough: A gridding technique for geospatial matching

Dale Barnhart, Ioana Lungu, Belay Mohammed and Rik Linssen

Background

- As part of our evaluation of the Women's Empowerment through Gender Transformative Market Opportunities (WEGO) program, Laterite is evaluating whether CARE Ethiopia's Farmer Field and Business Schools (FFBS) improve women's empowerment and household income
- Our evaluation used a pre-post design with a concurrent control group
- CARE purposively selected intervention locations using diverse agricultural and infrastructural criteria, making it difficult to identify comparable control locations

Challenge

- Geospatial data could help identify potential control locations with similar characteristics across the two intervention regions.
- Because the FFBS program is implemented at the village or "gote" level, we wanted to match intervention gotes to control gotes.
- Administrative data is only available at the kebele level, which prevents geospatial match at the gote level.

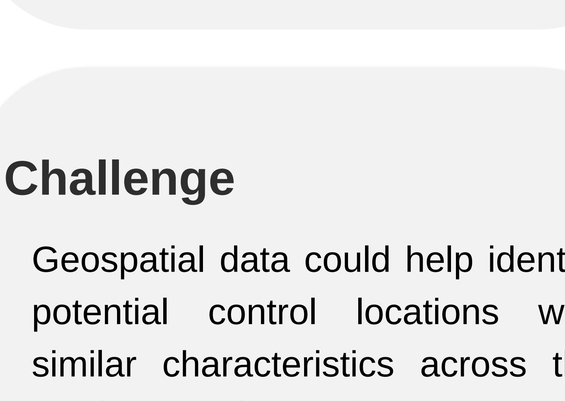

Fig 1. Sub-national administrative units in Ethiopia

Solution

1. We physically visited gotes where the FFBS program was being implemented and obtained a GIS coordinate for the center of each intervention gote.
2. We plotted the FFBS gotes on a map and drew a square 'cell' centered on that GIS coordinate (Figure 2, red squares).
3. We identified 26 potential control kebeles (Figure 3) and overlaid a grid over these kebeles to create a set of potential control 'cells' (green grid).
4. We extracted geospatial data for each of the FFBS intervention cells as well as each of the potential control cells from Laterite's geospatial database and Google Earth Engine.
5. We used nearest neighbor propensity score matching to select the *most comparable control cells for each treatment cell.*
6. We then visited the selected cells and conducted a household listing exercise in their surrounding gotes. Data from the household listing was used to select the *most comparable control and treatment households* for participation in the baseline survey

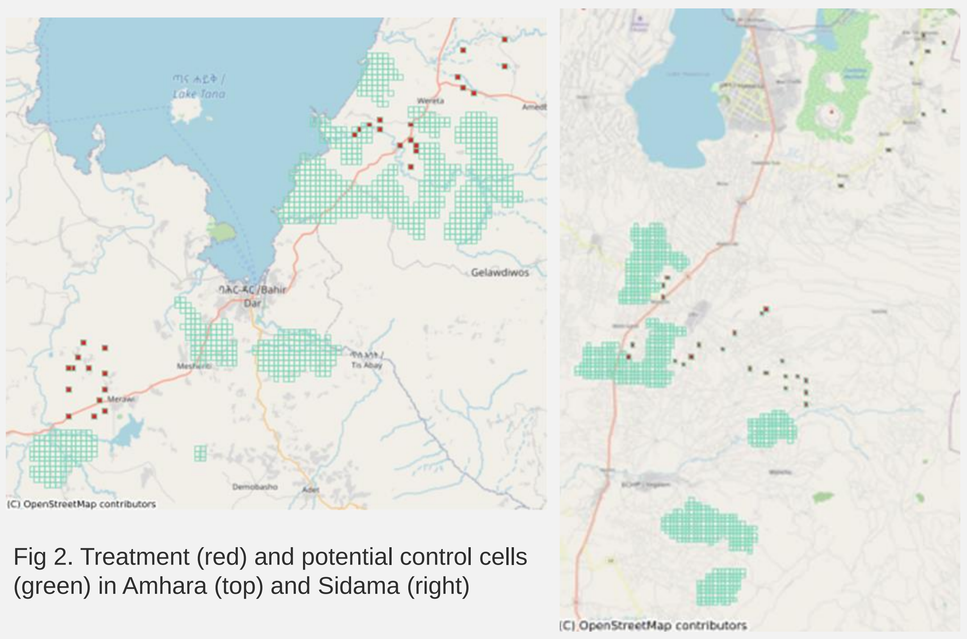

Fig 2. Treatment (red) and potential control cells (green) in Amhara (top) and Sidama (right)

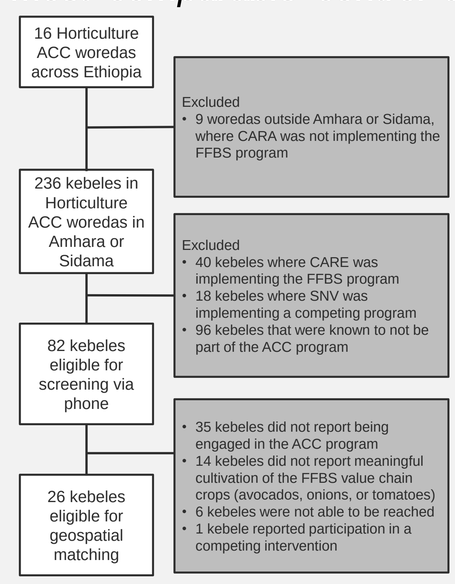

Fig 3. Inclusion and exclusion criteria for control kebeles

Results

- Our gridding process resulted in a pool of 59 treatment and 1,718 potential control cells from which we selected a final set of 40 FFBS and 40 matched control gotes
- In our final sample, we were successfully able to balance key agricultural, economic, and household composition variables.
- One key lesson learned is that selecting a match **set** of gridded locations rather than matched **pairs** (Step 5) would provide greater flexibility if it becomes necessary to replace locations.

Table 1. Characteristics of our final sample

	Amhara N = 444 Clusters = 32			Sidama N = 700 Clusters = 50		
	FFBS households N = 211 Clusters = 15	Control households N = 233 Clusters = 17	p-value[1]	FFBS households N = 350 Clusters = 25	Control households N = 350 Clusters = 25	p-value[1]
	%/Mean	%/Mean		%/Mean	%/Mean	
Crops grown last season[2]						
Avocado	23.7%	24.0%	0.974	99.1%	99.7%	0.407
Tomatoes	51.2%	39.9%	0.246	12.0%	4.0%	0.076
Onions	92.4%	82.4%	0.148	8.0%	4.3%	0.191
Share of female-headed households	13.7%	10.3%	0.440	17.4%	22.0%	0.231
Household has ≥5 members	74.4%	71.7%	0.541	73.4%	74.9%	0.716
ACC/Cooperative member	46.9%	46.8%	0.991	47.4%	39.7%	0.347
Head of household completed 5th grade or higher	15.2%	18.9%	0.404	56.6%	88.0%	0.264

[1]p-value for a pairwise t-test adjusting for clustering. [2]Households could farm multiple crops.

We would like to thank the team at the University of Washington's Evans School Policy Analysis and Research Group (EPAR) and CARE. This research is funded by the Bill and Melinda Gates foundation's Evaluating Inclusive Transformation in Agriculture (EVITA) program.

laterite
DATA | RESEARCH | ANALYTICS

More about our work:

Contact us:

dbarnhart@laterite.com
@Laterite
@LateriteAfrica
www.laterite.com