Using GPS in Household Surveys

Siobhan Murray
Georeferencing: why

- survey management (visit verification, logistics)
- minimal additional cost and time, although need to invest some extra effort in training and data management
- enable integration with large and expanding pool of spatial datasets (LSMS-ISA geovariables)
- provide objective measurement of spatial attributes such as area, path/distance
Georeferencing: why

Objective vs. farmer self-reported measurement of plot area

Clear evidence of heaping/loss of precision in self-reported measurements!
Can have a significant impact on estimates of yield and productivity
Georeferencing: what

- Location variables (enumeration area, city, district, region..), which can be linked to administrative level boundaries, even EA in some cases.

- GPS position in specified coordinate system, defined by
  1. cs type (spherical vs projected)
  2. projection parameters
  3. units
  4. datum
Georeferencing: what

Datum refers to a combination of earth model (sphere or ellipsoid) and the position of the spheroid relative to the center of the earth.

Latitude and longitude are spherical coordinates measuring, respectively, angle above the equator or relative to prime meridian. Commonly based on WGS84 datum, but might also be local datum. The length of a degree of longitude varies based on latitude, but one degree at the equator = 111 km, so decimals are significant!

Standard map projections are made onto planar, conic or cylindrical surfaces. Each type preserves different properties (area, shape, distance) and requires different sets of parameters. Units are often meters or feet, so decimals are less important.
Georeferencing: what

Numbers ≠ Location if coordinate system is not well defined.

- DD MM.mmm
- DD.ddddd
- Meters UTM 37
- Meters UTM 36
- Meters UTM 38
- All Post QC
GPS: how does it work

With a minimum of 4 satellites in view, a GPS unit can calculate horizontal \((X, Y)\) and vertical \((Z)\) coordinates, using precise time measurement included in the signals to determine distance from each satellite, and overlapping spheres to identify position (trilateration).

- ideal arrangement of constellation
- dilution of precision indicators
- GPS + GLONASS = 50 satellites
## GPS: sources of error

<table>
<thead>
<tr>
<th>Source</th>
<th>Effect</th>
<th>Typical Error in Meters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satellite</td>
<td>Orbit errors</td>
<td>2.1</td>
</tr>
<tr>
<td></td>
<td>Clock bias</td>
<td>2.1</td>
</tr>
<tr>
<td>Signal Propagation</td>
<td>Ionospheric refraction</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>Tropospheric refraction</td>
<td>0.7</td>
</tr>
<tr>
<td>Receiver</td>
<td>Receiver Noise</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>Multipath</td>
<td>1.4</td>
</tr>
<tr>
<td>Data collection, entry, management...</td>
<td>Can be significant offset</td>
<td>Can be minimized</td>
</tr>
</tbody>
</table>
GPS Implementation

1. Equipment
2. Questionnaire
3. Training
4. Data Management
   • Data entry
   • Review & cleaning
5. Final Preparation & Dissemination
Equipment

Consider device features and “fitness for purpose”

- Battery life: plan for back-up batteries, vehicle chargers, power conservation
- Interface: USB, serial, wireless.
- Track log: storage capacity (expansion option) and downloading
- Camera
- Area Calculation
- Vertical (Z) coordinates
Equipment

Technical specs on accuracy (5-10 m for recreational, mapping grade LSMS-ISA surveys) refer to absolute measures.

Accuracy of relative measures (length, area) affected by

• feature (plot, track) size
• walking speed
• device frequency settings

*ongoing research…LSMS-ISA Land Area Measurement experiments
1. Questionnaire

Simple concept, but still need to put some thought into designing GPS section of questionnaire:

- What coordinate system will you use?
  - if UTM does study area cover multiple zones?
  - are there negative coordinates in your study area? (N/S, +/-)

- What units will be used?
  - if degrees, then DD.ddddd, DD MM.mmm, DD MM SS?
  - clearly indicate precision/number of decimals expected

- How do coordinates look on the unit chosen for survey?
  consistency with device in order of inputs

- if very high accuracy required, record DOP indicators, walking speed, potential interference (canopy, other structures)
## Questionnaire

### Coordinate precision & accuracy (DD):

<table>
<thead>
<tr>
<th>decimal places</th>
<th>degrees</th>
<th>N/S or E/W at equator</th>
<th>E/W at 23N/S</th>
<th>E/W at 45N/S</th>
<th>E/W at 67N/S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.0</td>
<td>111.32 km</td>
<td>102.47 km</td>
<td>78.71 km</td>
<td>43.496 km</td>
</tr>
<tr>
<td>1</td>
<td>0.1</td>
<td>11.132 km</td>
<td>10.247 km</td>
<td>7.871 km</td>
<td>4.3496 km</td>
</tr>
<tr>
<td>2</td>
<td>0.01</td>
<td>1.1132 km</td>
<td>1.0247 km</td>
<td>.7871 km</td>
<td>.43496 km</td>
</tr>
<tr>
<td>3</td>
<td>0.001</td>
<td>111.32 m</td>
<td>102.47 m</td>
<td>78.71 m</td>
<td>43.496 m</td>
</tr>
<tr>
<td>4</td>
<td>0.0001</td>
<td>11.132 m</td>
<td>10.247 m</td>
<td>7.871 m</td>
<td>4.3496 m</td>
</tr>
<tr>
<td>5</td>
<td>0.00001</td>
<td>1.1132 m</td>
<td>1.0247 m</td>
<td>.7871 m</td>
<td>.43496 m</td>
</tr>
<tr>
<td>6</td>
<td>0.000001</td>
<td>111.32 mm</td>
<td>102.47 mm</td>
<td>78.71 mm</td>
<td>43.496 mm</td>
</tr>
<tr>
<td>7</td>
<td>0.0000001</td>
<td>11.132 mm</td>
<td>10.247 mm</td>
<td>7.871 mm</td>
<td>4.3496 mm</td>
</tr>
<tr>
<td>8</td>
<td>0.00000001</td>
<td>1.1132 mm</td>
<td>1.0247 mm</td>
<td>.7871 mm</td>
<td>.43496 mm</td>
</tr>
</tbody>
</table>
Questionnaire Example 1

Check for consistency between modules in same survey

Household Qx:

8. WHAT ARE THE GPS COORDINATES OF THE DWELLING?

<table>
<thead>
<tr>
<th>LATITUDE (N)</th>
<th>LONGITUDE (E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>__ __ __ __ __</td>
<td>__ __ __ __ __</td>
</tr>
</tbody>
</table>

Agriculture Qx:

5. ENUMERATOR: RECORD THE COORDINATES FOR THE CENTER POINT OF THE [PLOT]. SAVE LABEL AS HOUSEHOLD ID AND PLOT ID

<table>
<thead>
<tr>
<th>LONGITUDE (E)</th>
<th>LATITUDE (N)</th>
<th>PLOT LABEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>__ __ __ __ __</td>
<td>__ __ __ __ __</td>
<td>__ __ __ __ __</td>
</tr>
</tbody>
</table>
Latitude and longitude values were reversed in approximately 16% of recorded plot locations.

Reversal of lat / lon values is not easily detected and/or fixed in areas where these values can be equal.
Check for consistency between Interviewer manual and questionnaire

### GPS Coordinates

<table>
<thead>
<tr>
<th>ParcelID</th>
</tr>
</thead>
<tbody>
<tr>
<td>36N</td>
</tr>
<tr>
<td>UTM</td>
</tr>
</tbody>
</table>

The manual states:

“GPS setup should be:
POSITION FRMT: hddd.ddddddo
MAP DATUM: WGS 84
-----------------------------------
Take note of the (N) North and (E) East coordinates. If the GPS is correctly setup, the co-ordinates should appear on the screen as decimal degrees with the following format:
N 00.00000°
E 000.00000°”

<table>
<thead>
<tr>
<th>PARCEL 01</th>
<th>Satellites:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N=1 S=2</td>
<td>D</td>
<td>M</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LAT</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LONG</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2005 GPS locations, interpreted as decimal degrees
no points below the equator

2009 GPS locations, interpreted as decimal degrees
Training

• Training and manuals should provide general guidance for operators on data collection and device maintenance
  1. GPS unit setup (CS, units, frequency)
  2. wait time for signal acquisition
  3. walking speed, perimeter definition for area measurement (start and closure, check sketch)
  4. monitor battery status
• Additional instructions for saving and downloading raw GPS data (supervisor task)
• Don’t underestimate time required to introduce GPS data collection. Plan for at least one day of training, including field practice (even in the parking lot!)
Training - use screen captures from device

How to capture a waypoint.

1. Check that you have enough satellites!!
2. Go to Main Menu
3. Go to Mark
4. Go to Top Field
5. Enter the **ID CODE** for the **HOUSEHOLD**
   - Use “Enter / Rocker Key”
6. Record the GPS Coordinate in the questionnaire.
7. Select “OK”
Training – be specific on process

IHS3:
To get the area, walk the plot and record the area.

IHPS:
Walk the plot, record the area and save the track.
Training – be specific on naming

PLOT NAME:

Rainy Season Plot:
HOUSEHOLD + LOWEST IHS3 PRESENT
HOUSEHOLD MEMBER ID FROM ROSTER

200901R1

CURRENT PLOT NUMBER
Training – stress important messages

Important!

- GPS units must all be set up to measure location and area in the same format
  Meters (not feet)
  Acres (not Meters or Hectares)

- Once you set up the GPS unit, make sure never to change the settings

- Periodically check that the GPS unit is set up correctly
Data Management: Data entry

Different ways to record data:

1. Manual data-entry (paper) in field or at HQ
2. GPS raw data: avoids confusion associated with GPS settings, as well as data-entry errors, but requires extra effort for
   - track, waypoint naming convention that ensures link to household or other survey records
   - downloading protocol to account for limits in device storage
   - conversion to standard format (open source tools GPS Babel, DNR Garmin & GPSU, or proprietary TerraSync, Waypoint Manager)
3. Automatic import to CAPI (future)
Data Management: Data entry

Range constraints imposed in data-entry application can eliminate large coordinate errors (wrong/missing digit in degrees, field reversal). If done before leaving EA, opportunity to recollect..

<table>
<thead>
<tr>
<th>DISTRICT</th>
<th>MIN_X</th>
<th>MIN_Y</th>
<th>MAX_X</th>
<th>MAX_Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mzuzu City</td>
<td>33.95292</td>
<td>-11.50042</td>
<td>34.07311</td>
<td>-11.36122</td>
</tr>
<tr>
<td>Ntchisi</td>
<td>33.69843</td>
<td>-13.53524</td>
<td>34.12824</td>
<td>-13.01336</td>
</tr>
<tr>
<td>Nkhata Bay</td>
<td>33.73531</td>
<td>-12.23483</td>
<td>34.32634</td>
<td>-10.99384</td>
</tr>
<tr>
<td>Nkhotakota</td>
<td>33.75265</td>
<td>-13.50490</td>
<td>34.34611</td>
<td>-12.16290</td>
</tr>
<tr>
<td>Nsanje</td>
<td>32.97952</td>
<td>-17.12500</td>
<td>35.31261</td>
<td>-9.36754</td>
</tr>
</tbody>
</table>
To implement distance calculation in data-entry application, with input in decimal degrees:

1. **conversion to planar coordinates**
   OGP Geomatics Guidance Note number 7: Coordinate Conversions and Transformations including Formulas (www.epsg.org)

2. **great-circle distance**
   \[ a = \sin^2(\Delta\varphi/2) + \cos(\varphi_1).\cos(\varphi_2).\sin^2(\Delta\lambda/2) \]
   \[ c = 2.\text{atan2}(\sqrt{a}, \sqrt{1-a}) \]
   \[ d = R.c \]
   *where \( \varphi \) is latitude, \( \lambda \) is longitude, \( R \) is earth’s radius (mean radius = 6,371 km)*

3. **degrees-to-meters lookup table by latitude**
   Simplest solution, relatively accurate for very small distances
Data Management: Data entry

Reliable coordinates carried forward enable panel verification across waves. Is this the same plot visited in previous wave?

<table>
<thead>
<tr>
<th>DIST METHOD</th>
<th>DIST_KM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coordinate conversion</td>
<td>0.02526</td>
</tr>
<tr>
<td>Great Circle</td>
<td>0.02539</td>
</tr>
<tr>
<td>Latitude lookup</td>
<td>0.02542</td>
</tr>
</tbody>
</table>
Data entry and validation could also include internal consistency checks. Is distance from household to plots within a reasonable range? Or households within expected range of village centerpoint?

<table>
<thead>
<tr>
<th>HHID</th>
<th>PLOTNUM</th>
<th>DIST_KM</th>
</tr>
</thead>
<tbody>
<tr>
<td>123456</td>
<td>M2</td>
<td>1.9</td>
</tr>
<tr>
<td>123456</td>
<td>M3</td>
<td>2.5</td>
</tr>
<tr>
<td>123456</td>
<td>M4</td>
<td>1.3</td>
</tr>
<tr>
<td>123456</td>
<td>M1</td>
<td>75.2</td>
</tr>
</tbody>
</table>
Rigorous data validation checks implemented in data-entry will hopefully minimize data cleaning required.

Simple filters can be used to generate a summary table of records for review:

<table>
<thead>
<tr>
<th>field name</th>
<th>value</th>
<th>description</th>
<th>records</th>
</tr>
</thead>
<tbody>
<tr>
<td>units</td>
<td>DD.ddddd</td>
<td>coordinate system = lat/lon, units = decimal degrees</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>DD MM.mmm</td>
<td>coordinate system = lat/lon, units = degrees and minutes</td>
<td>338</td>
</tr>
<tr>
<td></td>
<td>Meters utm36</td>
<td>coordinate system = UTM36N, units = meters</td>
<td>418</td>
</tr>
<tr>
<td></td>
<td>Meters utm37</td>
<td>coordinate system = UTM37N, units = meters</td>
<td>3413</td>
</tr>
<tr>
<td></td>
<td>Meters utm38</td>
<td>coordinate system = UTM86N, units = meters</td>
<td>260</td>
</tr>
<tr>
<td>ll_switch</td>
<td>1</td>
<td>fields reversed, hh_saq25_a approx equal to lon of woreda center &amp; hh_saq25_b approx equal to lat of woreda center</td>
<td>3857</td>
</tr>
<tr>
<td>ll_blank</td>
<td>1</td>
<td>missing or zero value in fields hh_saq25_a or hh_saq25_b</td>
<td>55</td>
</tr>
<tr>
<td>dups</td>
<td>1</td>
<td>other records with identical values in fields hh_saq25_a and hh_saq25_b</td>
<td>38</td>
</tr>
</tbody>
</table>
Data Management: Data review

Very obvious errors can be fixed in record-level review, but any changes should be tracked in separate variable (indirect measure of coordinate reliability)

<table>
<thead>
<tr>
<th>field name</th>
<th>value</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>qa_type</td>
<td>1</td>
<td>hh is less than 10km from EA central feature and the coordinates are unique</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>hh is less than 10km from EA central feature but the coordinates are duplicates</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>hh is less than 10km from EA central feature but the coordinates were modified during review (excluding lat-long switches)</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>hh is more than 10km from EA central feature but the hh moved between visits</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>hh is more than 10km from EA central feature</td>
</tr>
</tbody>
</table>
Raw data is a valuable output. Some post-processing steps possible only with access to raw GPS data.

In Afar Livestock Survey, sample sites were generated around randomly selected points, and sometimes fell in very rough and remote terrain. Enumerators were unable to access the whole site. As a result, weights had to be adjusted based on the extent of site covered.

Estimates of site coverage were generated in GIS using visibility analysis based on a terrain model and stored GPS tracks.
Dissemination

- Household GPS coordinates are confidential information, but location information can be valuable for end-users.

- For LSMS-ISA surveys we disseminate modified EA center-points, offset to prevent identification of communities AND a set of geovariables generated using the actual household locations.

- In order to preserve anonymity of households and individuals we adopted method used by MeasureDHS program, relies primarily on random offset of cluster coordinates within a specified range determined by population density.
Household locations (red squares) are averaged by EA (red dot), and random displacement applied to produce a public version of the location (blue dot).

Blue circle represents the known range of uncertainty (5 km radius of offset for 99% of dataset).
<table>
<thead>
<tr>
<th>theme</th>
<th>variable</th>
</tr>
</thead>
</table>
| distance                   | distance household to plot  
distance household to paved road  
distance household to major market (if available) |
| climatology                | annual mean temperature  
mean temperature of wettest quarter  
mean temperature of driest quarter  
annual precipitation  
precipitation of wettest quarter  
precipitation of driest quarter  
precipitation seasonality (CV) |
| landscape typology         | land cover class  
agro-ecological zone |
| soil & terrain             | elevation  
slope class  
topographic wetness index  
landscape-level soil characteristics |
| crop season (temporal)     | short-term average crop season rainfall total  
specific crop season rainfall total  
short-term average NDVI crop season aggregates  
specific crop season NDVI crop season aggregates |
GPS vs. Farmer Self-Reported:

– “Fact or Artefact: The Impact of Measurement Errors on the Farm Size-Productivity Relationship”
  • Gero Carletto, Sara Savastano, and Alberto Zezza
– “Missing(ness) in Action: Selectivity Bias in GPS-Based Land Area Measurements”
  • Talip Kilic, Alberto Zezza, Gero Carletto, and Sara Savastano
  • Forthcoming…

Upcoming Methodological Validation Experiments:

*GPS, Compass & Rope, and Farmer Estimates*

– Zanzibar: with validation of measurement of continuously harvested crops (cassava)
– Nigeria: on subsample of LSMS-ISA households
– Ethiopia: with validation of soil fertility and production measurements
References


• Incorporating Geographic Information Into Demographic and Health Surveys: A Field Guide to GPS Data Collection. MeasureDHS

• LSMS website – manuals and other survey documentation

• Forthcoming note on use of GPS in LSMS-ISA surveys
GPS Resources

- GPS Babel (http://www.gpsbabel.org/)
- GPSU (http://www.gpsu.co.uk/)
- DNR Garmin (http://www.dnr.state.mn.us/mis/gis/tools/arcview/extensions/DNRGarmin/DNRGarmin.html)
- proprietary software:
  - Garmin Waypoint Manager,
  - Trimble TerraSync, etc