Global Positioning System (GPS) data collection in the BASIS CRSP 'Dynamic Poverty Traps' Research Project[1]

Justine Wangila (ICRAF)

This memo is intended to provide some basic background information to members of the BASIS CRSP project research team with respect to the collection of spatially explicit data using a handheld global positioning system (GPS) receiver for subsequent analysis in geographic information systems (GIS). Those desiring more detailed or sophisticated information should consult any of several good texts in this area or GIS analysts in their institution.

1. What is GIS?

- GIS is a computer-based system normally used to capture, store, edit, display and plot geographically referenced data.
- A GIS database contains information on: 1) value of the individual observations, 2) their location, and 3) their spatial arrangement.
- GIS handles two types of information: 1) geographic information – location of features, and 2) attribute data … names, number of species, feed availability, etc. The attribute data are stored in "layers", or files that link the geographic coordinates of points to values for a specific variable at each of those coordinates.

2. What is GPS?

The Global Positioning System (GPS) is a satellite-based navigation system made up of a network of 24 satellites placed into orbit between 1978 and 1994 by the United States' Department of Defense about 12,000 miles above the surface of the planet. GPS was originally intended for military applications, but in the 1980s, the government made the system available for civilian use. GPS works in any weather conditions, anywhere in the world, 24 hours a day. There are no subscription fees or setup charges to use GPS.

GPS satellites circle the earth twice each day in a very precise orbit and transmit electronic signals back to earth. GPS receivers take this information and use triangulation to calculate the user's exact location by comparing the time a signal was transmitted by a satellite to the time it was received. The time difference tells the GPS receiver how far the signal traveled and, therefore, the distance from the satellite to the GPS receiver. Because the altitude and orbit of these satellites are known, the microprocessor in the GPS unit can compute the latitude and longitude of its location by triangulation using distinct signals from three different satellites (a fourth signal is necessary to get three dimensional coordinates, i.e., altitude in addition to latitude and longitude). Many GPS receivers cannot only determine the user's position; they can also

[1] This brief draws significantly from: Alex Awiti, 2000. Geographical Information Systems, Remote Sensing and Decision Support in AHI Research. African Highlands Initiative Report. ICRAF, Nairobi; from the Garmin Corporation's "What is GPS?" web site (http://www.garmin.com/aboutGPS/), Garmin Corporation's GPS Guide for Beginners (http://www.garmin.com/manuals/gps4beg.pdf), J. Hurn, GPS*: A Guide to the Next Utility*, Trimble Navigation, 1993; T. Jasumback, Luepke, D., McCullough, R. and Weigel, D., GPS Walk Method of Determining Area. Technology & Development Program United States Department of Agriculture (USDA, 1997) (http://www.fs.fed.us/database/gps/mtdc/97712321/), Dana, P. H, 1995. *Geodetic Datum Overview*. Department of Geography, University of Texas at Austin (http://www.colorado.edu/geography/gcraft/notes/datum/datum_f.html), and from personal communication and comments from Oscar Ochieng (DSS Lab, ICRAF), Chris Barrett (Cornell) and, especially, Steve DeGloria (Cornell).

display it on the unit's electronic map. GPS measurement accuracy ranges from a few millimeters for very high-end systems to tens of meters, depending on the hardware and software used and the length of observation times.

3. Types of GPS Systems

GPS receivers are small hand held devices, the size of first generation mobile phones or a pocket calculator. GPS systems can be separated into two accuracy ranges: 1) the differential GPS (DGPS) and 2) non-differential GPS.

The DGPS corrects for bias in measurement at one unknown location using the measured bias of the unit at a known position. A reference receiver, or base station, computes corrections for each satellite signal and broadcasts roaming signals to increase accuracy. DGPS has accuracy from 0.2-10 meters and is relatively expensive. Garmin, a leading manufacturer, reports current average accuracy of 3-5 meters for DGPS units.

Non-differential GPS are less accurate than DGPS units, with accuracy of 10-20 meters on average. However, newer Non-differential GPS receivers with WAAS (Wide Area Augmentation System) capability can improve accuracy to less than 3 meters on average. These are primarily handheld GPS receivers that can store feature attribute data. They are widely available and relatively inexpensive (US$100-200 for units without WAAS, $250 or so ones with WAAS). Non-differential GPS are sufficient to locate features at coarse level: village, household, homestead, community cattle dip, health center, etc., in a socio-economic study. This will suffice for the purposes of the BASIS CRSP project.

Keep in mind that measurement error with the GPS unit is a random variable, subject to lots of exogenous variation (e.g., day-to-day change in the ionosphere and troposphere, changes in the geometry of signal reception at different times of day or year, etc.). Some observations may yield errors near zero or very large ones, but neither case is of any particular significance (http://users.erols.com/dlwilson/gpsacc.htm).

The GPS instruments purchased for the Rockefeller students were Garmin 12 and cost US$149 per unit was using Ben Meadows on-line catalog. These lower cost Garmin 12 are preferred over the Garmin 12XL as the extra features of the 12XL are of little use outside the USA and one cannot estimate area of polygonal features using "tracks" with the 12XL.

4. Selecting a GPS Receiver

There are now so many different types of GPS receivers that it can be difficult for the inexperienced user to select the GPS receiver that best suits his or her needs. Here are four key considerations to help select the appropriate GPS: 1) suitability to your level of research, 2) price range, 3) preference in features and display, and 4) ease of operation. Popular handheld non-differential GPS receivers include Garmin GPS, currently the world’s best selling GPS model, and Magellan GPS.

GPS signals travel by line of sight, meaning they will pass through clouds, glass and plastic but will not go through most solid objects such as buildings and mountains. Measurement errors arise from any of several different sources. The ones that can be controlled by the user include: (I) multipath caused by signal reflection off objects such as tall buildings or large rock surfaces before it reaches the receiver and (ii) exposure to insufficient or minimal satellites due to line-of-sight blockages from buildings, rock, dense foliage or water. GPS units typically will not work, or at least not well, indoors.

There is sometimes significant difference in accuracy of measuring height (altitude), with units such as the Eagle Explorer and Garmin eMap exhibiting little or no bias in height measurement and other handheld GPS units, such as the Garmin 12XL and Garmin III+ exhibiting bias in altitude measurement of approximately 10 meters, on average.

5. Some Useful Terms

Coordinates - A unique description of a geographic position using numeric or alphanumeric characters.

Elevation - Distance above or below average sea level.

Initialization - The first time a GPS receiver orients itself to its current location. After initialization has occurred, the receiver remembers its location and acquires a position more quickly because it doesn't need a large amount of satellite information.

Latitude - A position's distance north or south of the equator measured by degrees from 0 to 90 …with the Equator at 0 degrees.

Longitude - A position's distance east or west of the prime meridian (usually measured in degrees) which runs from the North to South Pole through Greenwich, England.

Position - A geographic location on the earth, commonly measured in latitude and longitude.

Position Fix - The GPS receiver's computed x, y position.

Position Format – Commonly geographic (also called lat/long), and is the way in which the GPS receiver's position will be displayed on the screen. Commonly displayed as **degrees and minutes**, with options for **degrees, minutes, and seconds**; **degrees only**, i.e., degrees decimals (DD), degrees minutes decimals (DMD) and degree minutes seconds (DMS) or **one of several grid formats like Universal Transverse Mercator.**

Universal Transverse Mercator (UTM) - A worldwide coordinate projection system utilizing north and east distance measurements from reference point(s) in meters. UTM is the primary coordinate system used on United States Geological Survey topographic maps.

6. Relevance to BASIS

- In the BASIS CRSP project, our main aim is to use GIS to georeference socioeconomic information, farming systems information, information on climatic conditions, and information on soil, crop and forage conditions. In sum, we will use GIS as an interface between socioeconomic and biophysical data.
- GPS units can also be used for measurement of field area, as a check against respondent reporting of plot size.
- We will not concern ourselves here, in this memo, with the data integration procedures and processes but on basic data.

6. What and how to record?

- It is important to specify the datum for the coordinate system you plan to use. Different **geodetic datums** define the size and shape of the earth and the origin and orientation of the coordinate systems used to map the earth. Modern geodetic datums range from flat-earth models used for plane surveying to complex models used for international applications, which completely describe the size, shape, orientation, gravity field, and angular velocity of the earth. Referencing geodetic coordinates to the wrong datum can result in position errors of hundreds of meters. Different nations and agencies use different datums as the basis for coordinate systems used to identify positions in geographic information systems (GIS), precise positioning systems, and navigation systems. The diversity of datums in use today and the technological advancements that have made possible global positioning measurements with sub-meter accuracies requires careful datum selection and careful conversion between coordinates in different datums. The same X and Y coordinate values will plot at different locations on the Earth's surface when assuming a different datum. Those wishing to repeat your experiments or revisit your households will have difficulty doing so without knowing the datum used.

In the BASIS CRSP Project studies, the WGS84 datum or whatever datum is used on the topographic maps in Kenya and Madagascar are recommended. The WGS-84 Geoid defines heights for the entire earth. Any datum commonly used in Kenya or Madagascar, will be sufficient as long as the datum assumed by the surveyors, the datum setting on the GPS unit, the datum recorded by the surveyors, and the datum on the base maps are all the same.

- Oftentimes, and at the most basic and simple level, a geographic reference is simply a coordinate or address that specifies the location of a household, community or other feature.
- In geo-referencing data collection in socioeconomic or interdisciplinary surveys such as those conducted under the BASIS CRSP project, there are several key issues to keep in mind:
  - **Unit of reference** - Household.
  - **Point Format** - The way in which the GPS receiver's position will be displayed on the screen. Commonly displayed as **degrees and minutes**, with options for **degrees, minutes, and seconds**, **degrees only**, or **one of several grid formats**.
  - The **reference point** should be 5 meters from the doorstep (main door) of the main household house in a direction perpendicular to the wall. Put differently, record the reference point using 1) the distance of 5 meters from the doorstep of the main household, 2) the direction in azimuth degrees of arc, to the household from the reference point (in our case 90^{o}). In this way, those conducting a re-survey or re-visit will have an easier time deciding which household was actually surveyed, in that one can go in 360 different directions from the reference point (assuming one degree resolution). Also, with geodetic freeware ("Forward.exe"), the true coordinate location of the household can be calculated easily if one knows the distance and direction from the reference point.
  - **Acquisition Time** – This is the time it takes for your GPS unit to acquire a lock onto enough satellites (three for a 2D readings and four for 3D). This should be about two minutes to get a good fix.
- When measuring plot area, one must note *all* the turning points on the perimeter of the plot (i.e., the vertices of the polygon). With experience and a relatively high-end GPS receiver, one can record the positions of the turning points using the unit's *data logger,* which is set up to collect synchronous data at the 5-second logging interval. The simpler, more labor-intensive method is simply to collect data using the "walk and record on sheet of paper"

method by starting at one turning point and walking in a straight line to the next turning point, then to the next one until the perimeter of the polygon had been traversed (closing the polygon at the starting point). Remember to stay at each turning point long enough until a steady reading is found and record the latitude and longitude of the turning point appropriately before proceeding to the next.

Without Selective Availability (SA) active, Cornell specialists have consistently acquired point feature locations to within 2-5 meters RMSE positional accuracy using the Garmin 12 instrument. One does not need significant experience nor a high end GPS instrument to estimate area of polygonal features (e.g., fields, farms, building footprints). Using "tracks," no files need to be downloaded for post-processing, but GPS unit, track point management, and accurate field notes are important. Using "routes," files need to be downloaded using freeware (Waypoint +) and an ArcView script, written by Art Lembo at Cornell. Using routes is recommended as the area estimates are of higher accuracy and precision though a bit more data processing is required.

8. Unique Identification and Recordation of Coordinates

The purpose of the unique household identification code assigned to each respondent unit in the sample is to provide a link from the household name to all the survey data in a form that maintains respondent anonymity and minimizes the possibility of conflating responses across observations. Research ethics require that investigators and analysts safeguard the anonymity of human subjects. So the exact location of a respondent household should not be made publicly available in any circumstance when the respondents' name would not likewise be made public information. It is good practice; therefore to create a single master reference file containing the name, ID code and GPS coordinate readings for each respondent in the survey. This master reference file is for the purposes of replication and data verification and is typically held by the principal and co-principal investigators only. For the purposes of protecting respondent confidentiality, it is typically not made publicly available, even when all the rest of the data are made available.

For consistency and comparison across sites, household surveys under the BASIS CRSP Project should use the positional accuracy and precision shown in the two formats for the master reference file:

A) In degree, minute and seconds (DMS) to the nearest arc-second**, that is,**

Household Name	**Household ID Code**	**Latitude**	**Longitude**
		Degree Minutes Seconds	**Degree Minutes Seconds**
Internet Oloo	001	36 02 11.1	04 29 03.1
World cup Mwangi	103		
Senegal Shikanga	222		

OR

B) **In degree decimals** (DD) to at least five decimal places, that is,

Household Name	Household ID Code	Latitude	Longitude
		Degree Decimals	Degree Decimals
Senegal Shikanga	222	36.01304	04.48204

The above recommendation is based on the fact that one second of latitude and longitude in the vicinity of the household surveys represent 30.8 meters and 25.0 meters, respectively, in ellipsoidal distance. Recording, for example no decimal places for DMS coordinates **or** less decimal places for DD, three decimal places for the Internet Oloo site, could result in an error of approximately 1.5 decimal seconds, or approximately 50 meters simply through recording coordinates with inadequate precision.

Important: Use either of the above templates and record reading by hand.

Survey administrators should prepare a master reference list for each survey site, pre-filling the household ID codes. Enumerators then complete the remainder (name and geographic coordinates) at the time of the initial survey.