Issues and guidelines for the emerging use of GPS and PDAs in agricultural statistics in developing countries

Naman Keita, *Senior Statistician, Food and Agriculture Organisation of the United Nations*
Viale delle Terme di Caracalla
00153 Rome, Italy
E-mail: Naman.Keita@fao.org;

Elisabetta Carfagna, *Professor, University of Bologna, Italy*
E-mail: elisabetta.carfagna@unibo.it

George Mu'Ammar, *Programme Officer, United Nations World Food Programme*
Email:: george.muammar@wfp.org

1. Introduction

The Global Strategy to Improve Agriculture and Rural Statistics adopted by the United Nations Statistical Commission recognises the importance of geo-referencing statistical units and use of geospatial information for the construction of master sample frames which is a key element for integration of agriculture into national statistical systems, one of the three pillars of the Strategy.

The use of handheld geo-referencing devices (global positioning system and personal data assistant equipped with GPS) is also emerging as an effective means of reducing the cost of crop area data collection and improving crop area data quality particularly in the African context.

The crop area (area planted and area harvested) is among the key variables included in the core data set defined in the Global Strategy which have to be produced by all countries on a regular basis. The availability of reliable estimates of crop area is an important requirement of any country agricultural statistics system, as this variable is an important parameter for estimating the production, together with crop yield. These three variables (area, yield and production) are among any minimum set of key indicators for planning, monitoring and evaluating agricultural and rural development programmes aiming at food security and raising the living standards of the population in a country.

However, in many developing countries, collecting reliable data on crop area still remains a challenge for agricultural statisticians. This was confirmed by a recent stakeholder survey conducted in the framework of the preparation of the implementation plan for Africa of the Global Strategy. The results of the survey clearly showed that improving the methodologies for estimation of crop area, yield and production remains the highest priority for research in Africa.

Different methods are currently used for collecting crop area data in developing countries, including field reporting system, eye estimation, interview of the farmers, objective measurement methods.

All these methods have their limitations in terms of reliability of crop area data, but the objective method of measuring areas is considered to give the most reliable data.

However, measuring areas is difficult, costly and time-consuming in developing countries, due to specific problematic situations. The shape of the fields, is not always polygonal but often a curvilinear closed figure, which has to be reduced to a polygon, with a small number of sides (e.g. less than 20) of an equivalent area. Measuring errors can be introduced by the surveyor or are inherent to the equipment used; thus the investigators have to be well trained on surveying techniques and on the proper use of the necessary equipment.

In hilly regions, when crops are grown on slopes which could be quite abrupt (more than 20%), the evaluation of the crop area is not simple. The crop area should not be the physical area measured on the slope (the inclined plan) but its projection on the horizontal plane. This is due to the fact that plants and trees grow vertically and not perpendicularly to the slope and thus require for their growth some kind of vertical cylinder of soil. If the crop area is measured on the slope and not projected horizontally, crop areas could be significantly over-estimated.

In addition to difficulties associated with measuring the area of a given plot, there are difficulties in determining the share of specific crops in presence of mixed or associated cropping which is a common practice in African countries. In some cases, two or more different temporary and/or permanent crops simultaneously grow on the same field or plot.

The problem of estimating crop areas in these situations gets more and more complicated as the number of crops in the mixture or in association increases and especially when the proportions of the different crops in the mixture vary from field to field. Moreover, the vegetative cycles of the crops may have different lengths (from less than three months to more than a year) and the crops may have different periods of sowing, planting and harvesting.

Thus, the number of crops in the same field may vary according to the period of the year and the time of the enumerator's visit. The allocation of areas to the component crops is a complex task.

Various methods are used in countries, which vary from simple methods to more complicated ones (often recommended but rarely used):

- when the number of crops in small and the number of combinations of crops is limited, affect total plot area to each distinctive combination
- attribute total area to principal or predominant crop. As a variation of this method, for each crop, report area as principal constituent of mixtures and secondary constituent of mixtures
- attribute total area to each crop in the mixture (large overestimation of area).
- divide total area between crops
- eye estimates of the proportion of area occupied by each crop

2. Review of various methods of crop area measurement

Two categories of area measurement methods are used: list frame with land surveying methods and methods based on mapping with geographic sampling

2.1. List frame and land surveying methods

Information on the crop area is often collected selecting a sample from a list of farmers and requesting the farmers an estimate in local units. This approach has many sources of inaccuracy. Crop area is often expressed in units other than metric units (ha, acres, square meters). Different units are used in different regions and some of them do not have a standard conversion factor which allows computing the corresponding measure in hectares. Serious problems exist when measuring units with different names are used locally in different regions within the same country and when no conversion factor is known. A still more serious problem is encountered in countries where measuring units with a specific name has different dimensions in different parts of the country and these dimensions are unknown.

Rapid approximation methods are also widespread (Triangulation, Rectangulation, Pacing, Perimeter squared):

- *Rectangulation*: when the shape of the field is not too complicated, the method of rectangulation can be useful and its reliability for estimating crop area may be acceptable. The simplest way is to measure the length of the parcel or field somewhere more or less across the middle and then determine through eye estimation the position of the average width. However, application of the method on the ground can be difficult and results in important errors
- *Triangulation*: the polygon is subdivided into triangles with common vertex (condition not indispensable but useful). The area of the field can be evaluated as the algebraic sum of the triangles. This method can be applied when the boundaries of the field are rectilinear and the field is a plane polygon with well identified vertices. When this approach is followed, the instruments used are: pantometer, compasses, clisimeter or clinometer in presence of slopes.
- *Perimeter square*: an order of magnitude of the area of a field can be quickly estimated by dividing the perimeter by a number between 4 and 5 and squaring the result. The choice of the divisor is subjective and is based on the degree of complexity of the boundary: number of sides, number of concavities, etc. For complicated fields, the divisor will be close to 5 and for a rectangle or square it will be close to 4.
- *Pacing: In devel*oping countries, most of the farmers do not know the magnitude of the areas under their crops and the fields has to be measured by the available field staff in collecting agricultural statistics. The staff, not being specialized in land surveying techniques, has to be trained on simple methods, namely, rectangulation using the length and width of the field and measuring distances through pacing, i.e. *walk*ing at a normal gait and counting the number of steps to cover the distance. The steps are then converted to standard units. When this approach is followed, the instruments used are: pacing, standardised cords, surveyor's chain, and others (Trumeter, Smith Wheel, Optical Range finder etc.).

For land surveying, the methods adopted can be classified into two groups:

Group 1: the primary data (distance and angle) is used to produce a sketch of the plot at a given scale, the area of which is measured or calculated and converted to actual crop area. Several techniques are used: Planimeter, Topochaix, others such as triangulation, weighting, etc.

Group 2: the primary data are used directly to produce plot area. This second group became a standard with the technological breakthrough of the apparition of handheld/pocket programmable calculators and their low cost. They allowed the calculation of crop area in the field by inputting directly the distances and angles.

For this method, FAO has developed several programmes for different models of programmable calculators which, in addition to the estimates of area, also provide an estimate of closing error (FAO, 1982). This standard method of crop area measurement (compass, tapes and calculator) is still widely used in most African countries and is the current standard method. It was a considerable advance in the mid 80's in terms of facilitation of the field work in agricultural surveys using land surveying methods.

2.2. Methods based on mapping with geographic sampling

Crop area measurement based on mapping is common in several countries. These methods are often called geographic sampling or area frame sampling. The units of an area frame can be points, transects (straight lines of a certain length) or pieces of territory, often named segments.

Area sampling: Strictly speaking, area sampling is performed when the selection probability of each segment is proportional to the area of the segment. The type of segments used for agricultural area sample surveys include:

(i) segments with permanent and identifiable physical boundaries (roads, rivers, canals, railroads, etc.) that are readily found and provide unambiguous identification of the segment;

(ii) square or rectangular segments that is segments defined by straight lines forming squares or rectangles whose end points are established by map coordinates. In this case, grid sampling procedures are often used;
(iii) segments that coincide with the land of agricultural holdings. In this case, point sampling procedures are used.

In most countries, land parcels and crop fields, have irregular boundary lines which are not straight lines.

FAO has produced in the mid-90's two complementary publications on the area frame sampling which can be used as reference for this technique: Multiple frame agricultural surveys Volume 1 (1996) and Volume 2 (1998).

An increasing number of countries are using or are interested in the methods based on mapping material, (particularly area frame and remote sensing) since the resolution of satellite images is becoming higher with substantial reduction in their cost and with also availability of powerful computer systems which can assist in the map based methods. Classified satellite images and land cover maps produced by photo-interpretation are useful tools for area frame construction as well as for improving the precision of crop area estimates through estimators which assign to classified satellite images the role of auxiliary variables (Gallego *et al.* 2010 and Gallego and Carfagna, 2005). Substantial advances are being made in this domain which needs a consolidated technical documentation on the state of the art and guidelines for practical application in countries.

However, for a variety of reasons, most African countries still use list frames with land surveying methods in their agricultural surveys for crop area measurement and the typical instrument is the compass, tape and calculator.

With the emergence of the new handheld geo-positioning and mapping devices such as the Global Positioning System (GPS) and the Personal Data Assistant (PDA), new perspectives exist as viable alternative to cumbersome distance and angle measurement.

In fact, the Global Strategy for Improving Agricultural Statistics advises to create a master sample frame for agriculture which will be the foundation for all data collections based on sample surveys or censuses and allows to use both households and farms as statistical units and provides a linkage between the census framework and land use. The master sample frame must provide the basis for the selection of probability based samples of farms and households with the capability to link the farm characteristics with the household and then connect both to the land cover and use dimensions. The area sample frame meets this requirement. The methodology using the population census recommended for the World Program for the Census of Agriculture 2010 will also meet this requirement if—households from the population census are geo referenced and used as the frame for the agricultural census—and linked to satellite images of land use. The use of GPS guaranties the geo-referencing of data collected.

Combination of digital thematic maps, digital administrative boundaries and geo-referenced statistical information opens to the possibility for spatial analyses of data. However, before such data can be used in Geographical Information Systems (GIS), a long process of data capture, geo-referencing/geo-coding, scanning and digitalization is required. Since geographical information will be found in different organizations it is crucial for common use to agree upon standards and formats.

By introducing geo-referencing (coordinates) and geo-coding (administrative division unit code) to statistical information of agricultural holdings sampled during survey and censuses and at the same time introduce similar coding for business and industry surveys/listings, new possibilities of spatial analyses of data occurs. The statistical information can also be combined with other sources of digitalized geographical data such as thematic maps.

Currently, with the new technological advances, survey statisticians need to take advantage of the tools available and revisit past methods. FAO is preparing a handbook which aims at providing practitioners with a new reference document based on the state of the art in terms of crop area measurement with GPS and PDA and their use for linkage with other layers of data in GIS.

3. Presentation of the GPS

The Global Positioning System was conceived by the US department of Defence to provide an accurate measurement of position and velocity of ground, sea, air and space objects. It became fully operational in the 80's, but with a limited precision for civilian uses (using a cryptographic technique named SA for Selective Availability, leading to accuracy levels in positioning measurement of nearly one hundred meters). In May 2000, the SA feature was turned off and since then, every standard GPS receiver gives precise positioning with a mean error of almost less than 10 meters. The Russian system named GLONASS covers only the Russian territory and Galileo, the European system should be available in 2014.

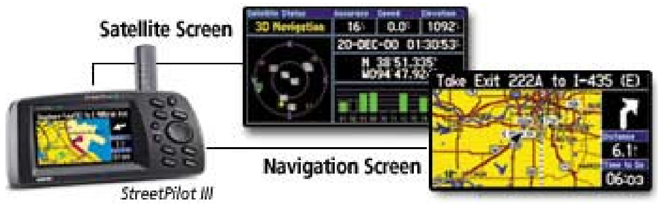
The GPS consists of the space segment and the control segment. The space segment is composed of satellites in medium earth orbit and also includes the boosters required to launch them into orbit. 24 satellites are generally in orbit, following 6 circular orbits (4 satellites per orbit) separated by 60° each inclined 55 degrees to the equator, placed at an altitude of 20,184 km. GPS satellites circle the earth twice a day in a very precise orbit, as shown in figure 3.1 and transmit signal information to earth. GPS receivers take this information and use triangulation to calculate the user's location.

Figure 3.1: The satellite orbits

The control segment is composed of a master control station, an alternate master control station, and a host of dedicated and shared ground antennas and monitor stations.

The GPS receiver compares the time a signal was transmitted by a satellite with the time it was received (see fig. 3.2). The time difference tells the GPS receiver how far away the satellite is. With distance measurements from a few more satellites, the receiver can determine the user's position and display it on the unit's electronic map.

Figure 3.2: The GPS system



Four GPS satellite signals are used to compute positions in three dimensions and the time offset in the receiver clock. Position in XYZ is converted within the receiver to geodetic latitude, longitude and height above the ellipsoid. Atmospheric factors and other sources of error can affect the accuracy of GPS receivers.

3.1. Position calculation with GPS and sources of GPS signal errors

Knowing the indicated time the message was received, the GPS receiver can compute the transit time of the message. Assuming the message travelled at the speed of light, the distance travelled or pseudorange can be computed.

In case the GPS receives the signal from only 3 satellites, computing a position in two dimensions fixing the altitude is still possible. Some handheld GPSs have a built-in altimeter based on pressure changes. This is in addition to the altitude calculation that is performed as part of the GPS solution and can provide more accuracy. These units always use the altimeter for altitude readings and will always show a 3D solution even if you only have a 2D GPS solution.

The geometry of the satellites affects the positioning error. When the GPS satellites are close together in the sky, the geometry is said to be weak and the Dilution of precision (DOP) or Geometric Dilution of Precision (GDOP) is high; when far apart, the geometry is strong and the DOP value is low. Other factors that can increase the effective DOP are obstructions such as nearby mountains or buildings.

Various factors can degrade the GPS signal and thus affect accuracy; the most common are the following:

- Number of satellites visible - The more satellites a GPS receiver can "see", the better the accuracy. Buildings, terrain, electronic interference, dense foliage or even the operators body, for example if the GPS is held close to the waist, can block signal reception, causing position errors or possibly no position reading at all. GPS units typically will not work indoors, underwater or underground.
- Satellite geometry/shading that is the relative position of the satellites at any given time. Ideal satellite geometry exists when the satellites are located at wide angles relative to each other. Poor geometry results when the satellites are located in a line or in a tight grouping.
- Signal multipath - This occurs when the GPS signal is reflected off objects such as tall buildings, large rock surfaces or even water bodies such as lakes or rice paddies, before it reaches the receiver. This increases the travel time of the signal, thereby causing errors. The result is a barrage of signals arriving at the receiver: first the direct one, then a bunch of delayed reflected ones. This creates a messy signal. If the bounced signals are strong enough they can confuse the receiver and cause erroneous measurements. Sophisticated receivers use a variety of signal processing tricks to make sure that they only consider the earliest arriving signals (which are the direct ones).
- Receiver clock errors - A receiver's built-in clock is not as accurate as the atomic clocks onboard the GPS satellites. This is responsible for inexpensive GPS inherent inaccuracies of 3-10 m
- Ionosphere and troposphere delays - The satellite signal slows as it passes through the atmosphere. The GPS system uses a built-in model that calculates an average amount of delay on a typical day to partially correct for this type of error, but atmospheric conditions are rarely exactly typical.
- Orbital errors - Also known as ephemeris errors, these are inaccuracies of the satellite's reported location. These errors are caused by gravitational pulls from the moon and sun and by the pressure of solar radiation on the satellites.

4. PDA with GPS

PDA (Personal Digital Assistant) is basically a handheld computer that can fit into the palm of your hand. Most PDA's do not have keyboards due to their size. They usually consist of:

- A screen which is touch sensitive;
- A stylus (a pen for PDA's) which functions like a computer mouse;
- A few buttons. Data entry occurs using the stylus on the touch sensitive screen.

The use of PDAs as a data collection instrument is rapidly reaching most areas of its potential application. The main advantages in using a PDA is that the software used for collecting the data can be customised to suit the necessities of the application. This means that issues such as user-friendliness, rapidity of data entry, data validation, data security and data availability can be supported to whatever level is necessary for the application.

Although the hardware characteristics of PDAs change and improve continuously, there are some features which are particularly useful for data collection, such having an incorporated GPS, being waterproof, having a flash memory card for holding data which can quickly be exchanged with a computer or with another PDA without connecting the PDA to the computer, having a screen and a wireless connectivity.

4.1 Steps to follow for using GPS and PDA for crop area measurement

The data acquisition process with GPS and PDA has to be made following a series of steps:

- Testing the device before use;
- Design of the acquisition strategy (individual plots or contiguous plots);
- Preparation of the material to check data in the field, such as maps with coordinates printed from user's GIS or from virtual globes and validation points (taking slope and landscape into account)
- Upload of existing data or maps into the receiver;
- Identification of the fields to measure;
- Marking the borders of the fields;
- Data storage and reporting. The data storage can be done on the paper (each measurement is annotated on the paper) or on the GPS. In the letter case, the coordinates of the perimeter of the plot are stored in the GPS and can be downloaded into a GIS and can be integrated with other geographic data;
- Looking at the shape of the plot on the screen of the devise in order to avoid systematic and gross errors, during the test as well as in the operational data acquisition phase. Whenever a gross error occurs, the operator should decide whether it is significant for the area measured. If it is the case, this fact should be noted in the field data form and the measurement should be repeated;
- Acquisition of redundant control points allowing post acquisition integrity checking (a standard form could be used for metadata capture);
- Post acquisition control, that is:
    - Checking the integrity of the data collection
    - Computing the perimeter/area ratio. The larger the perimeter/area ratio is, the more complex the shape of the plot is, and therefore the larger the effect of a potential GPS error would be on the area calculation of the shape.
    - Visualizing the plot borders
    - Overlay of the plot to existing digital maps or insert the plot in a virtual globe.

GPS hardware determines points in geographic coordinates and elevation. In the memory of the device, the points are tracked chronologically to make lines. The line is considered to define an area if the first point is connected to the last point. If the loop is not closed by the enumerator, then the program will compute the closed area obtained by connecting the last point in the log to the first point. This geometry is used by the GPS to calculate the area of the polygon. However, if the user walks in an overlapping path, this will define an impossible surface and this would lead to a grossly erroneous calculation. The calculation uses signed numbers so if the track is crossed over at some point creating a figure 8 like picture, then the program will compute the difference in area between two circles.

During the time needed for measuring the area of a plot logging along the perimeter, the GPS constellation can be considered as relatively stable, thus the positioning errors do not have effect on the measurement. Area measurement error is linked both to the operator speed and to the acquisition rate of the GPS device. Field area measurement errors can be limited if an appropriate combination of operator speed and GPS acquisition rate is selected, for parcels up to 4 ha, the ‘optimum’ range of speeds for operators on foot is between 0.5 m per second and 2 m per second (1.8–7.2 km per hour) (Bogaert *et al.* 2005).

4.2. Testing of the GPS devices before use in survey

Some GPS models give users control over the quality of the data to be collected, by giving a possibility of selecting threshold values like: the maximum DOP, horizon mask or the minimum signal to noise ratio. Availability of these settings should be carefully investigated prior to testing. However most of the recreational GPS devices will accept the entire signal without informing the user about its quality. Nevertheless it is not only the receiver with its configuration that influences the final quality of the data. The method of the data collection influences strongly the final result thus should also be considered and selected before testing, e.g., when measuring the area of irregular plots, logging along the perimeter gives more reliable result than when just logging the corner points of the object.

Area measurements with the GPS receivers are fast, thus time efficient, easy and digital, thus traceable and easy to incorporate into a database. On the other hand, the accuracy and precision of the results in terms of area measurement are unknown as not stated by the manufacturers (only the point positioning errors are usually stated). Therefore, in order to understand if the GPS receiver meets the precision level required, the measurement system (receiver, settings and method of the measurements) should be validated.

The validation of the measurement systems should be undertaken whenever one part of the measurement system is changed (e.g. an external antenna is added or the logging interval significantly changed).

In order to assess performance of a tool, the accuracy and precision of that device should be evaluated. Bias, as a systematic error, can be compensated by correction coefficients if needed. Random errors related to a certain level of precision of the instrument can be reduced for example by averaging of several measurements; however, measurement repetition strongly reduces the efficiency of the use of GPS devices compared with compass and rope (traditional instruments).

By testing, we can not only derive statements on the accuracy and precision of the measurements system but also find answers for the following questions:

- how comfortable, quick and easy to use is the receiver for the operators,
- how long will the batteries last in the operational mode,
- what are the advantages and limitations of the specific device model,
- could that device be a part of a GPS data collection system (metadata, data downloading, storing etc).

4.3. Conditions in which use of GPS and PDA for crop area measurement is recommended and not recommended

FAO and several other Institutions conducted empirical field experiences in various regions and under different conditions in order to provide a scientific basis for recommending the use of GPS and PDA for crop area measurement in specific conditions (see Keita and Carfagna, 2009).

These studies have collected information on 207 plots. A probability sampling technique has not been adopted for selecting the sample. On the same plots, the area is measured with the traditional method first and then with the various kinds of GPS receivers, often repeating the measurement until 3 times. The kinds of GPS devises considered in the study were: Garmin 12 xl (G12), Garmin 72 (G72), Garmin 60 (G60) Garmin Etrex Ventura (GE), Magellan Explorist 400 (M400).

If the area measured with compass and meter is considered as the true area we can say that most GPS measurements are very near the true area. According to some experiences, about 80 % of the measurements have an absolute value of the relative error less than or equal to 2.5 %.

A linear regression of the measures with compass and tape and the measures made by GPS shows that the linear model explains a very high percentage of the variability of the compass and tape measures (R-squared = 0.9633), with parameters significantly different from zero and the slope very near 1 (0.9600301).

The accuracy of the GPS measurements seems to be influenced by the slope of the plot and by the tree canopy cover. The accuracy is high with no tree canopy cover and lower with partial or dense tree canopy cover. Moreover, the accuracy of the GPS seems to be influenced by the weather conditions: the accuracy is higher with sun rather than with clouds.

The larger the size of the plots, the higher tends to be the accuracy of the GPS area measurements. The analyses made considering clusters of plots according to their size have suggested that for large and very large plots (from 10,200 square meters) the GPS measures are very similar to the compass and meter ones; for medium size plots (form 5,300 to 9,999), they are less similar. Finally, for small and very small plots (less than 0.5 ha) the two distributions are quite different. **This lower limit of half ha for measuring area with current GPS devices with acceptable accuracy was also confirmed by experiments conducted by a leading GPS maker**.

The time needed for measurement with compass/tape and GPS receivers shows a high variability. For some small and medium plots, the time needed by compass and tape is even 17 times longer than with GPS; however, the median of this ratio is 3.3 and the mean 3.8. Although the mean of the time needed for the measurement show a small decrease passing from the first to the third measurement, the three means cannot be considered significantly different, as well as the medians. Moreover, the distributions of the time need for the first, the second and the third measurements are not significantly different.

The accuracy of the measurements does not improve from the first to the third measurement. Since the repetition is a cost, these data do not suggest repeating the measurement.

We have to consider that, in the experiments carried out, the measurement with compass and tape has been performed before using a GPS, and thus the compass and tape measurement is a benchmark. In operational activities, if only one measurement with a GPS is performed, there is no possibility to have a benchmark which allows detecting mistakes. Moreover, the kind of GPS considered does not give any evaluation of the accuracy of the measurement, which is given for the compass and tape method by the closure error. The studies revealed also that some type of GPS perform better that other in different conditions. Therefore, the selection of GPS should be done with great care.

Regarding operational use, battery life, robustness under difficult field conditions, simplicity of use should be taken into account when selecting a GPS type.

5. GIS and geo-referenced data basics

A geographic information system (GIS) is any system that captures, stores, analyzes, manages, and presents data that are linked to location. In the simplest terms, GIS is the merging of cartography, remote sensing and database technology. GIS applications are tools that allow users to analyze spatial information, edit data, maps and data, create interactive queries based on spatial relationships and present the results of all these operations. In GIS, digital data can be used to represent real objects (features such as roads, land plots, towns, countries) or to represent measurements and classifications (for example land use, rainfall). Traditionally, there are two broad methods used to represent geographical data in a GIS for both abstractions: raster and vector. A raster data type may be an image or a grid whereas the vector format represents geographical features as geometrical shapes, which are based on points, lines and polygons. Lines and polygons are themselves based on sequences of points.

The third GIS data type are relational tables, as used in a relational database, which can be attributed to the raster or vector dataset, and are therefore called the “attributes” of the geographical dataset.

GPS data is essentially vector data, since it is based on location data which is interpreted as points. Elevation and time are also present as attributes of the GPS points.

In using inexpensive handheld GPS units with a clear sky with good satellite visibility, the accuracy of the data is approximately 5 meters. A rough calculation could be made to establish the scale of layer created by this data to be approximately 1:20,000.

When different kinds of information layers are overlapped in a GIS, much attention has to be devoted to the kind of projection adopted.

There are hundreds of locally-developed horizontal projections around the world, usually referenced to some convenient local reference point. The WGS 84 is a common standard datum and is the most common geodetic systems (or datum) used by GPS equipment. It is almost identical to the NAD83 datum used in North America and the ETRS89 datum used in Europe[1].

5.1. GPS and GIS exchange data formats and methods

The standard GPS receiver transmits a stream of data in a format called NMEA[2]. The simplest GPS receivers simply save the NMEA into a file, or transmit it to a computer. Modern handheld GPS units contain a small computer that manages NMEA data, extracting the coordinates, the time, the precision parameters, changing the projection, saving it in tracks, routes and waypoints, etc. These handhelds usually come with a computer software for downloading of data. The data can then be saved in NMEA files, GPX and/or other standard formats such as text or GIS formats.

GPX (www.topografix.com) is a modern XML format for GPS data exchange. GPX does not only contain strictly geographic and GPS related data but also metadata, such as author and keyword tags. Clearly GPX benefits from all the advantages of using an xml-based data format.

GPSBabel is a commonly used program to convert from one GPS format to another, and can convert from NMEA to GPX and to KML[3]. You can use GPSBabel either as a graphical user interface, via Gebabbel or a command-line program.

---

[1] WGS stands for World Geodetic System. NAD stands for North American Datum, ETRS stands for European Terrestrial Reference System

[2] NMEA is short for NMEA 0183, which is the name given to the data communication specifications made by the National Marine Electronics Association, used by GPS and other marine equipment

[3] KML stands for Keyhole Markup Language, and is an xml based language for representing simple geographic. It was developed by Keyhole technologies, the original author of Google Earth software

NMEA sentences normally store all the data output by the GPS including data which is poor. However this can easily be cleaned out by setting filtering options with GPSBabel. The most common need is to filter data which has poor horizontal accuracy due to a lack of satellite visibility or other signal performance issues.

To convert GPS data into GIS format for overlaying with other datasets, the easiest route to take using free software is to save the data in KML format for Google Earth. Then use Kml2Shp to convert into ESRI Shapefile[4] format, which is now readable by all GIS systems. Kml2Shp can be used online www.zonums.com/online/kml2shp.php and an offline version is also available at www.zonums.com/kml2shp.html (although currently not available due to bugs).

5.2. Using geo-referencing data collection for crop area measurement

A track data from a handheld GPS and converted into a GIS format appears as a sequence of points within the GIS dataset. These are not equivalent to waypoints, where each waypoint has an ID (or a label), that appears as an attribute in the feature dataset. The only attributes that will appear in a track are the date and time of each vertex point of the track. Note that some GPS download software may have a function to convert tracks into line features. However this will probably not give the desired result, in that each land plot will not be recognized as an individual line.

GIS editing tools are necessary to separate each individual land plot series of points and to convert them into a single polygon, attributing an ID of the land plot, in order to link the data collected in the field to the polygon. Once this is done, each land plot is delineated by a uniquely identified perimeter, whose area can be calculated, and whose surface can be identified on other GIS layers, such as remotely sensed imagery (see section 5.3. below).

GIS Tools

Although the conversion of a series of points into a polygon would be a very simple GIS task to be done, it is not a commonly found tool in GIS packages. In fact it is usually GPS data management tools that contain this kind of function. Often however, manual digitizing is performed using the GPS points as reference, in order to obtain a clean geometric shape. By activating the "snapping" tool in the GIS, it is possible to have the mouse pointer "snap" to each GPS point in the track simply by moving it close to that point. This enables the user to follow the track's points and to easily digitize a polygon. Each polygon will be given an attribute value that relates to the data for that land plot in the relational database.

Once a layer is created with a polygon for each land plot surveyed, the analysis can be performed on the land plot areas in combination with other GIS datasets. Some of the GIS tools that can be used are listed below, although by no means in an exhaustive list:

- **ZonalStatistics** calculates a statistical function for each zone. In this case the zones are the individual land plots. Useful functions include MEAN (useful on WRSI, NDVI or other quantitative layers – see section 5.3 below) or SUM (for example on rainfall)
- **Clip** for cutting out the assessed areas from another layer. This is useful for excluding data that is not present in the cropped areas assessed from remotely sensed images for example.
- **Distance functions** (names are not standardized amongst GIS applications) to determine accessibility of land plots and distances to other features such as roads and water bodies
- **Slope** and **aspect** to determine the slope and exposure to sunlight of the cropped areas using a digital elevation model
- **Watershed** functions to determine irrigation and erosion parameters

5.3. Linkages with spatial information and GIS

The mapping of the cropped areas in a GIS is an added value to the results of the assessment. This allows the dataset to be used in combination with remotely sensed images for modelling and ground-truthing,

[4] Shapefile is an open format for simple geographical vector data representation. It is used mostly as an exchange format or for simple mapping applications. It was originally developed by the Environmental Systems Research Institute (ESRI)

and for monitoring and early warning. Mapping of the cropped area perimeter will enable the data to be overlaid onto high resolution imagery.

5.3.1 Useful datasets for overlaying cropped area datasets

Following are some commonly used GIS layers that can be used for modelling and monitoring. It is by no means an exhaustive list, but can provide useful ideas for such applications.

Vegetation Indices

Normalised Difference Vegetation Index (NDVI) is an image product that can be derived from various satellite sensors. It relates to a number of vegetation properties (photosynthetic activity, biomass amount, vegetation cover). NDVI imagery is usually issued as 10, 16 or 30 day composites.

Precipitation

Precipitation data can be obtained from ground weather stations or from satellite imagery. Data from ground stations is more precise but is available only for specific locations – geographically extensive data records go back to the early 1900s. The most well known imagery is the RFE2 rainfall estimates developed by NOAA Climate Prediction Centre and distributed by the USGS/FEWNet

Water Resource Satisfaction Index

WRSI is an index developed by FAO that quantifies the degree to which a crop's water requirements have been met by available water (precipitation and/or irrigation) – it is a measure of the total water deficit over the whole crop development cycle.

Landcover and vegetation images

Landcover images and maps classify the territory into different coverage types whilst vegetation maps will classify different types of vegetation. Vegetation and landuse maps are derived from interpretation of satellite or aerial imagery.

Soils map

A soils map is an example of the usage of a thematic map for modelling purposes.

Digital Elevation Models (DEM)

Elevation is a dataset that can be used to derive several products such as slope and run-off for drainage and flood analysis, aspect (direction land is facing in) for solar and wind exposure, and watershed and river basins for irrigation analysis.

High resolution satellite imagery

Images such as those taken by Quickbird, GeoEye, Orbview 3 and Ikonos satellites are such a high resolution that it will be possible to clearly discern the cropped area mapped during assessment, although expensive to order.

Using publicly available GIS and mapping platforms for displaying results

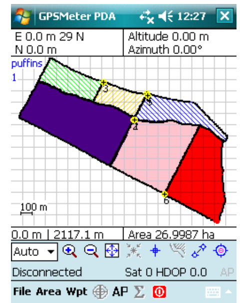
Web-based mapping platforms such as Google Maps, Bing Maps (Microsoft) or publicly available Web Map Services (WMS) provide a free and useful basemap to overlay crop assessment data on. Google and Bing Maps provide free high resolution imagery basemaps, although there is no control over what date the image is taken from.

Google Earth is a non-web-based application that presents the same (or equivalent) maps as seen on Google Maps, with the addition of historical images, as well as other information. The user can add crop assessment data by creating a file in KML format, which can be shared by email, for example.

6. Use of PDAs as a survey tool in agricultural statistics

6.1. Software applications for capturing GPS and PDA area measurements into a GIS system

- **Application software –** The number of applications on the market is growing at a very rapid pace. However software applications for capture of an area by GPS are actually very few and all run on Pocket PC. Some of these may be specific utilities, whilst others are actual GIS data capture packages, such as:



- **gvSig mobile (free)**
- **ESRI ArcPad**
- **Trimble PathFinder**
- **Encom**
- **FieldWorker**
- **SuperPad**
- **AgMapper**
- **GPSMeter PDA**

gvSig is a free software and AgMapper and GPSMeter are significantly cheaper than other applications.

6.2. World Food Programme (WFP) PDA Survey Software

WFP Vulnerability Assessment and mapping (VAM) has produced a PDA software for Food Security assessments. This software is made to handle very large and complicated questionnaires installed on many PDAs. The GPS component currently handles point data. The future release of the application will allow line and area data to be captured and will be released on an open license. Currently this functionality is under development.

6.3. Features that a user friendly software application must consider for proper area calculation.

GPS hardware determines points in geographic coordinates and elevation. In the memory of the device the points are tracked chronologically to make a line. The line is considered to define an area if the first point is connected to the last point. This geometry is used to calculate the area of the polygon.

Advantages of using PDA data collection compared to traditional data collection and handheld GPS usage:

- Questionnaire compilation: all data can be collected using a single instrument. This includes all metadata and data traditionally collected on paper regarding the survey and the crop. Photos and audio interviews can also be recorded for later reference.
- Accuracy data can be collected and reported to the user during the area track, if the software application is programmed to do so, including reporting of PDOP values for each point in the track, and the estimated error of the area calculation.
- Data security such as data backups using spare flash memory cards.
- Immediate validation on the collected data if the software application is appropriately programmed.
- User friendliness compared to handheld GPS units ultimately depends on the software application being used. Appropriate software would separate the procedures for configuration (system administration) from those for data collection (usage).
- PDAs can be used for a variety of data collection applications and therefore overhead costs can often be shared with other units in the organisation
- Handheld PDAs contain hardcoded software for trekking or boat navigation. This software smoothes the GPS track on the assumption that sharp corners will not be present in the track. This may introduce errors in the area estimation and may deform the shape of the polygon with reference to the mapping of the cropped plot.

Disadvantages:

- Overhead costs of procurement of PDAs compared to rope and compass
- Additional training in the use of PDAs is required
- Battery charging requires organised preparation and can cause limitations in usage
- Non-rugged PDAs are delicate and mishandling can cause damage
- Unnecessary risks may be caused by presence of PDAs in locations with high security issues

7. Concluding remarks

The use of GPS is essential when crop area is estimated through geographic sampling. It also allows geo referencing the households from the population census and linking them to satellite images of land use. This link allows using the households as a master sample frame for agricultural statistics in a list frame approach, as stated by the global strategy for improving agricultural and rural statistics.

Handheld GPS devices are also a viable alternative to time consuming and cumbersome distance and angle measurement when objective area measurement has to be performed by field staff in agricultural surveys. These devices are particularly relevant when the plots do not have a regular shape, since the accuracy of the GPS is acceptable and its use is faster and simpler. More evidence is needed for evaluating the accuracy of GPS measurements in the presence of abrupt slopes and under cloudy or raining weather conditions.

The GPS device also allows to integrate geo referenced data collected on the ground in GISs in order to overlay different kinds of spatial data and perform spatial analysis.

The use of PDAs has also many advantages, particularly GPS enabled PDAs. Several countries and institutions, including World Food Programme, are effectively using PDAs with electronic questionnaires directly in the field during surveys. This makes data collection faster and data quality better due to the reduction data entry procedures which may be error prone and due to immediate validation of the data collected.

On the other hand, the use of GPS and PDA has some disadvantages, e.g. overhead costs of procurement compared to rope and compass, specific training is required and battery charging can cause limitations in usage.

FAO, World Food Programme and the Joint Research Centre of European Union are preparing a Handbook on the use of GPS and PDAs in Agricultural Statistics.

References

Bogaert P., Delincé J, Kay S. (2005) Assessing the error of polygonal area measurements: a general formulation with applications to agriculture, Measurement Science and Technology, n. 16 pp. 1170–1178

FAO (1982), Estimation of crop areas and yields in agricultural statistics

FAO (1996) Multiple Frame Agricultural Surveys, vol. I Current surveys based on area and list sampling methods, "FAO statistical development series", n. 7, 119 pp., FAO, Rome, 1996

FAO (1998) Multiple Frame Agricultural Surveys, vol. II Agricultural survey programmes based on area frame or dual frame sample designs "FAO statistical development series", n. 10, 242 pp., FAO, Rome, 1998

Gallego F.J., Baruth B., Carfagna E. (2010) Accuracy, Objectivity and Efficiency of Remote Sensing for Agricultural Statistics in Benedetti, Bee, Espa, Piersimoni (Editors), Agricultural Survey Methods, Wiley, New York, April 2010, ISBN: 978-0-470-74371-3, 434 pages

Gallego F.J., Carfagna E. (2005) Using Remote Sensing for Agricultural Statistics" (with F. J. Gallego), International Statistical Review, volume 73, number 3, December 2005, pp. 389-404, ISSN 0306-7734

N. Keita, E. Carfagna "Use of modern geo-positioning devices in agricultural censuses and surveys", Bulletin of the International Statistical Institute, the 57th Session, 2009, Proceedings, Special Topics Contributed Paper Meetings (STCPM22) organised by Naman Keita (FAO) "Using advanced data collection methods and modern tools to improve agricultural statistics data quality", Durban, August 16-22, 2009