Reference Data Collection for Global Cropland Mapping

Introduction

Collection of reference data is an integral component of any thematic mapping project using remotely sensed imagery. The reference data are used for both the training of the classification algorithm(s) used to create the thematic map and also, independently, for assessing the accuracy of the final map. Therefore, no mapping project can succeed without the effective and efficient collection of reference data.

Goal

The goal of the reference data collection is to obtain the data necessary for completing the training and validation components of the project as efficiently and effectively as possible. All steps and considerations in the process must be thoroughly documented.

Important Terms

Sampling unit: The sampling unit is the term for the area covered by a single reference data sample. The sampling unit must account for positional error. Therefore, single pixels are not appropriate for a sampling unit. In this project and for Landsat Thematic Mapper imagery, a sampling unit of 3 x 3 pixels or approximately 1 ha. Is appropriate. The sampling unit must be homogeneous. That is the entire sampling unit must be the same land cover class.

Minimum mapping unit: The minimum mapping unit is defined as the smallest area that will be mapped in the project. While it is clear that no area smaller than a single pixel can or should be mapped, it is also true that the mmu is typically larger than a single pixel. The mmu should represent the required level of detail in the map while realistically considering the spatial resolution of the imagery.

Global Cropland Mapping Field Guide Procedures

1 Purpose

To collect reference data information to use for training or validation (must be independent of each other) of cropland distribution globally including crop type and crop intensity.

2 Objectives

- Crop types (8): Wheat, Maize, Rice, Barley, Soybeans, Pulses, Cotton, Potatoes
- Crop intensity (4): single, double, triple, and continuous cropping
- Irrigation (2): Irrigated or Rain-fed

3 Crop Definitions

- **Eight crop types**

1. **Wheat (spring and winter)**

Winter Wheat:

Planting: Winter wheat is planted from mid-August through October.

Harvest: Winter wheat is harvested from mid-May to mid-July.

Spring Wheat:

Planting: Spring wheat in northern mid-latitudes, like maize planting, generally occurs in April and May.

Harvest: Spring wheat is harvested from mid-August to mid-September.

2. **Maize:**

Maize is planted in the northern mid-latitudes generally occurs in April and May, with cooler regions planting later than warmer regions. Maize planting dates vary more widely in tropical and southern mid-latitude regions, where two neighboring countries can often have very different dates.

3. **Rice:**

The cycle of rice is 190 days from late March through April and the harvest season lasts for about 30 days in mid-September to October

4. **Barley:**

Winter barley is usually sown in the fall for exposure to low temperatures during the winter and then development is completed during the following spring and summer. Spring barley

does not require exposure to winter temperatures and can be sown in spring. Winter types usually mature somewhat earlier than spring types.

5. **Soybeans:**

Planting: Soybean crops are planted beginning in late April and last through June.

Harvest: Soybeans are mainly harvested in late September and is finished by the end of November

6. **Pulses/legumes**

Planting: March to October

Harvesting: May to December

7. **Cotton:**

Planting: April to June

Harvest: September to December

8. **Potato:**

Planting: Mid April to September

Harvesting: June to December

- **Crop intensity**: single, double, triple, and continuous cropping

Crop intensity is expressed as number of cropping cycles on the same area in a year.

- **Irrigation:**
  - **Irrigated Crops:** Irrigated Areas are those areas which are irrigated one or more time during crop growing season. Irrigation is the artificial application of any amount of water to overcome crop water stress.
  - **Rain-fed Crops**

Rain-fed areas have no irrigation whatsoever and are purely precipitation dependent

4 Materials and Tools

- Natural color, hard-copy high spatial resolution imagery of the area
- Compass
- GPS
- Camera (digital)
- Global cropland classification system and definitions (see above)

5 Time at Sample Unit

10-15 minutes

6 Frequency

Each crop type and intensity should be field-checked and must meet the classification scheme and field guide protocol requirements

7 Data preparation (pre-field activity)

Before going into the field to collect the actual data, it is imperative to obtain and review as much local information, imagery, and other geospatial data about the area as possible. The more knowledge of the area the field collection team has the better they can plan the actual collection of the reference data, especially in planning the optimal sampling strategies. Reviewing the following data would assist in planning and collecting the appropriate reference data.

1. Remote sensing images
   - MODIS
   - Landsat TM
   - Any other higher resolution images available of the area
2. Thematic maps about crops to determine crop types in the area
   - Cropland extent map
   - Agriculture Census data that might tell which crop types, intensities, or irrigation strategies are used in the area
3. Basic geographical data
   - Road network
4. Other geospatial data
   - Elevation data layer – slope and aspect
   - Ecological regions – important to find reasonable delineations of these areas long before going into the field

8 Field Data Collection Procedure and Requirements

8.1 Overview

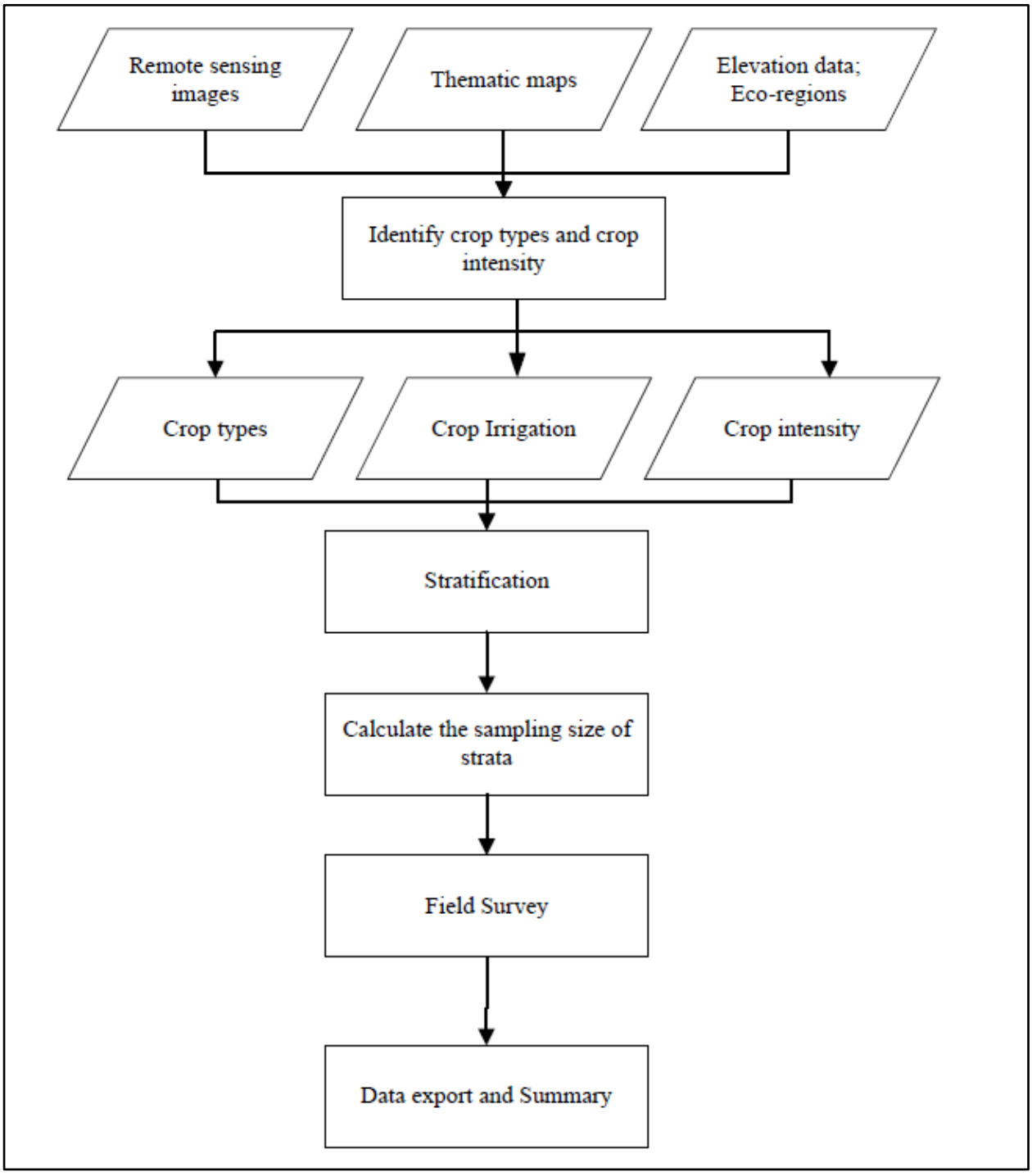


Figure 1: Flow chart of cropland investigation

8.2 Stratification

Stratification is a pre-field activity to be carried out based on local expert knowledge, ecological regions, climate variation and elevation to stratify the cropland extent map into several strata. In each strata, we implement an unsupervised clustering using the Landsat TM to determine how many spectrally unique clusters exist in each strata. The clusters will help us to decide the locations of sampling units in the field survey stage and also provide spectral uniqueness for each sample unit.

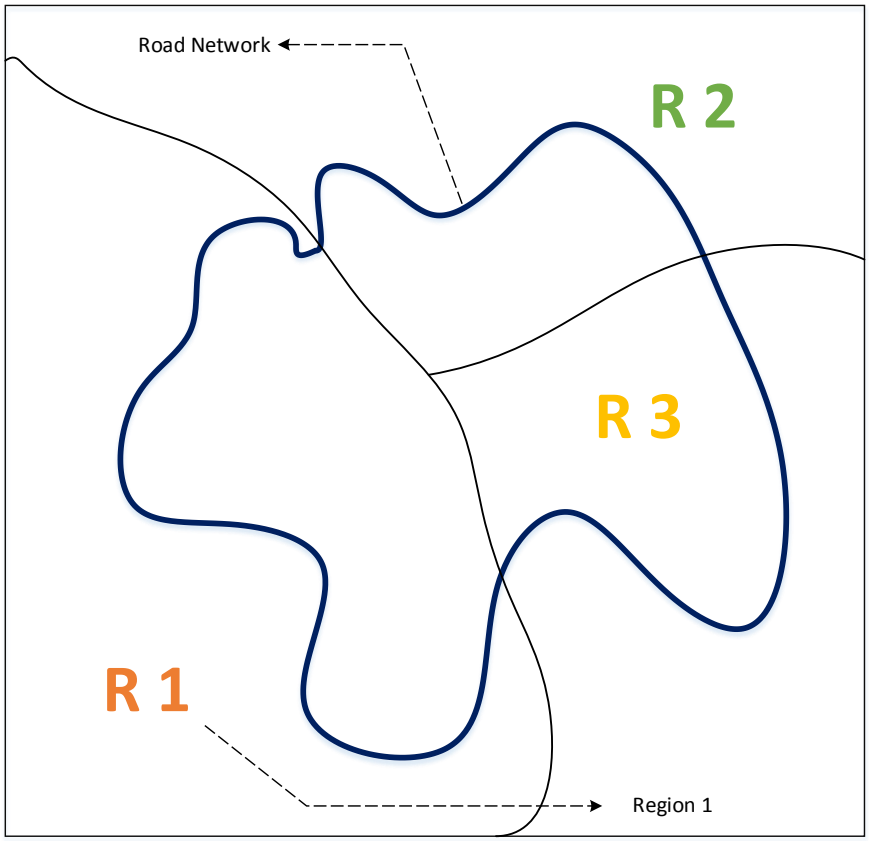


Figure 2: An example of stratification

In the example (Figure 2), the stratification divides the area into Region 1, Region 2 and Region 3. Each region consists of similar topographic, climatic and ecological regions which may have dominant crop type, intensity and irrigation. The sampling size of each region will be determined by its proportional area size and will be located within the maximum distance to the road.

8.3 Allocate the sampling size in each strata

Introduction: The stratified random sampling design is employed. The sample size of each strata is determined by the area proportion to the total area of cropland. Certain strata may have one dominant crop type with few rare crop types, so the minimum number of sampling units for the rare crop types are required.

Steps: Determine sampling size in each strata.

$$n_h = \frac{\text{Area of homogeneous strata h}}{\text{Total area of cropland}} \times n$$

n – The total sample size

n_h - The sample size of strata h

***Requirements*: *If some types are rare, make sure the minimum number of samples in these crop types*.**

8.4 Sample survey in the field

Introduction: In the field, equally distributing the sampling units along the road inside each strata is not very efficient. Unsupervised Clusters provide the roughly unique spectral clusters. Locating the samples in these clusters along the road will not only keep the precision but also make the sampling more efficient. Therefore, sample units will be taken where a number of different land cover types can be sampled at once, if possible.

***Step 1*:** Locate the samples in the actual field

- General principles

- The total number of reference data sampling units collected in certain strata should be equal or more than $n_{h.}$
- The goal is to stop to take samples at places where more than one crop type (unsupervised cluster) is located and to collect one or more sampling units for each of the crop types at that stop. Only one sample unit will be taken in a single unsupervised cluster.
- Sampling units should not be collected beyond the maximum distance from the road(D_{max}).
- The distance between the sampling units should be larger than the minimum distance (D_{min}) to minimize spatial autocorrelation.

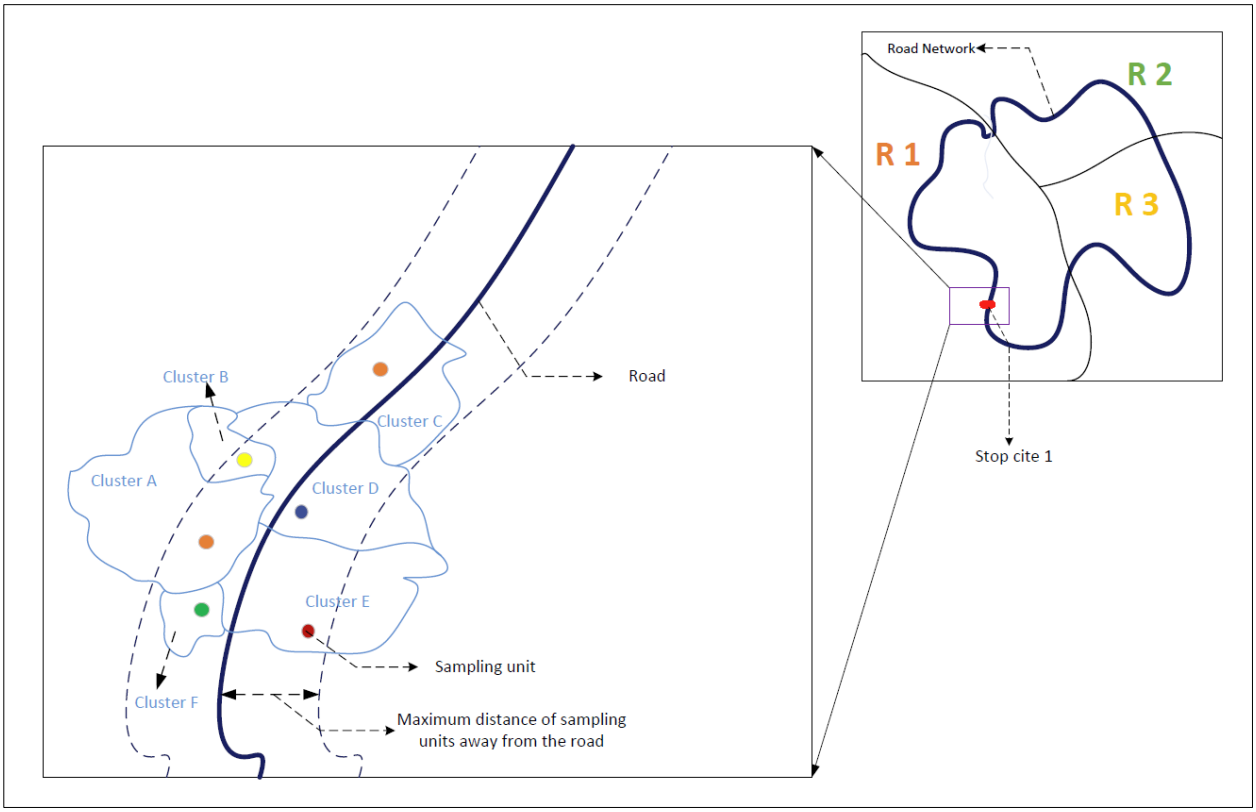


Figure 3 An example of stop site and data collection

Figure 3 is an example of sampling location in region 1 where only one or two crop types dominates. Pre knowledge or local knowledge shows that there are a number of possible clusters (crop types) on the map around stop site 1. We choose a suitable place to stop and see what happened in the field. In this example, we found that there are five types of crop in six spectral similar clusters or homogeneous field (Cluster A-F). So we located six sampling units around stop site 1: Two for crop A (shown in orange color), one for crop B (Yellow Color) and one for each crop C,D and E (Blue, Red and Green respectively), which accurately represent the proportion of each type of crop. In the field based on homogeneous condition, each cluster can split into two or neighbor clusters can be merged into one to represent one sampling unit in one homogeneous area or field. All sampling units are placed within a maximum distance from the road (D_{max}) and beyond the minimum distance from each other.

***Step 2*:** Travel to the sampling sites with GPS

- Travel to the center of sampling unit under the guidance of GPS and hard-copy of higher spatial resolution images of the sampling sites

Requirements: make sure the sampling units are within a homogeneous area of a single crop type

- Record 5 readings per minute at the center of the sampling site

Requirements*: *Make sure the center is clearly marked and the 5 records are valid.

Step 3: Take Photos with camera

- Take 5 photos at center of the site from the cardinal directions (N,S,E,W) and downwards
- Link the photo number with the sample site name and coordinates

***Step 4*:** Fill the survey form

Fill in the following form (See Figure 4) and make sure the information is correct.

8.5 Data export and summary

Export the data from the GPS and camera into the laptop every day. Check the link among the coordinates, forms and photos.

Appendix: Notations in this document

N-population

N_{h-} the population in the strata h

n_{h-} the number of samples in strata H

D_{min}- the minimum distance between samples

D_{max}- the maximum distance of sampling units from the road

Reference

William J. Sacks, Delphine Deryng, Jonathan A. Foley and Navin Ramankutty, Global Ecology and Biogeography, (Global Ecol. Biogeogr.) (2010) 19, 607–620

http://www.fao.org/agriculture/seed/cropcalendar/cropcalendar.do

http://www.pattyarnold.com/process.html

http://usda.mannlib.cornell.edu/usda/nass/planting/uph97.pdf

REFERENCE DATA COLLECTION FORM

Basic Required Information:

Date of Collection (month, day, year): __/__/____

Local Weather Conditions _____

Observer Name: _____ Expert: Yes or No

Village, Sub-Country, Country_____

GPS Coordinates of Center of Sampling Unit (WGS84?, DD) _____

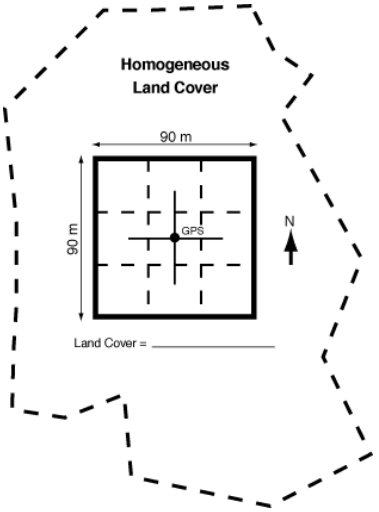
Datum_____

Latitude (X): _____ [Dec. Degrees] Longitude (Y): _____ [Dec. Degrees]

Estimated distance to nearest road: _____ [m]

Accessibility Class/Collection Method: ☐ Offset ☐ Vantage ☐ Center

Approximate Size of Sampling Unit (min. = 90mx90m): _____



Drawings of cropping pattern in the Training sample area (surrounding area including location of roads; must annotate with north arrow direction)

Crop Type (Wheat, Corn, Rice, Barley, Soybeans, Pulses, Cotton, Potatoes, or Other): _____

Were pictures acquired? Y or N

Associated Picture IDs

Photo ID	Direction	Photo ID	Direction
	North		East
	South		Up
	West		Down

Any issues or anomalies at site:_____

_____

_____

Additional Desired Information:

Irrigation: Irrigated or Rain fed

Cropping intensity: single, double, triple, continuous

Figure 4. Field Form