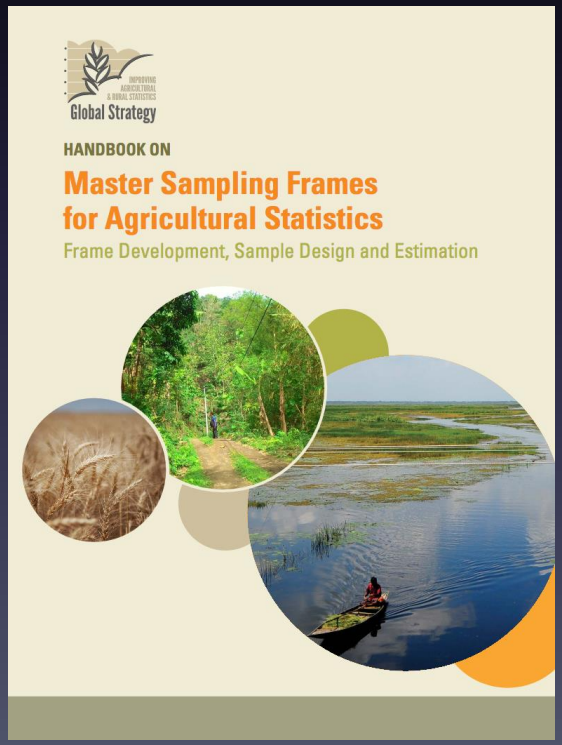
Testing Point Area Sampling Frame as candidate for Master Frame in Nepal

Jacques Delincé

Global Strategy, Statistics Division, FAO

Master Sampling Frame

- A priority of the Global Strategy
- See Handbook at: http://gsars.org/fr/category/publications/



What is a Master Frame

- A Master Frame is a frame that can be used for sampling in more than one survey
- It will be usually linked to Agriculture, Household or population census
- It will save time and money
- It will allow record matching among surveys
- It can be a list or an area frame

Field Tests Protocol

- Construction of the Frame
- Sampling Strategies
- Estimation Methods
- Costs & Benefits of the MF options

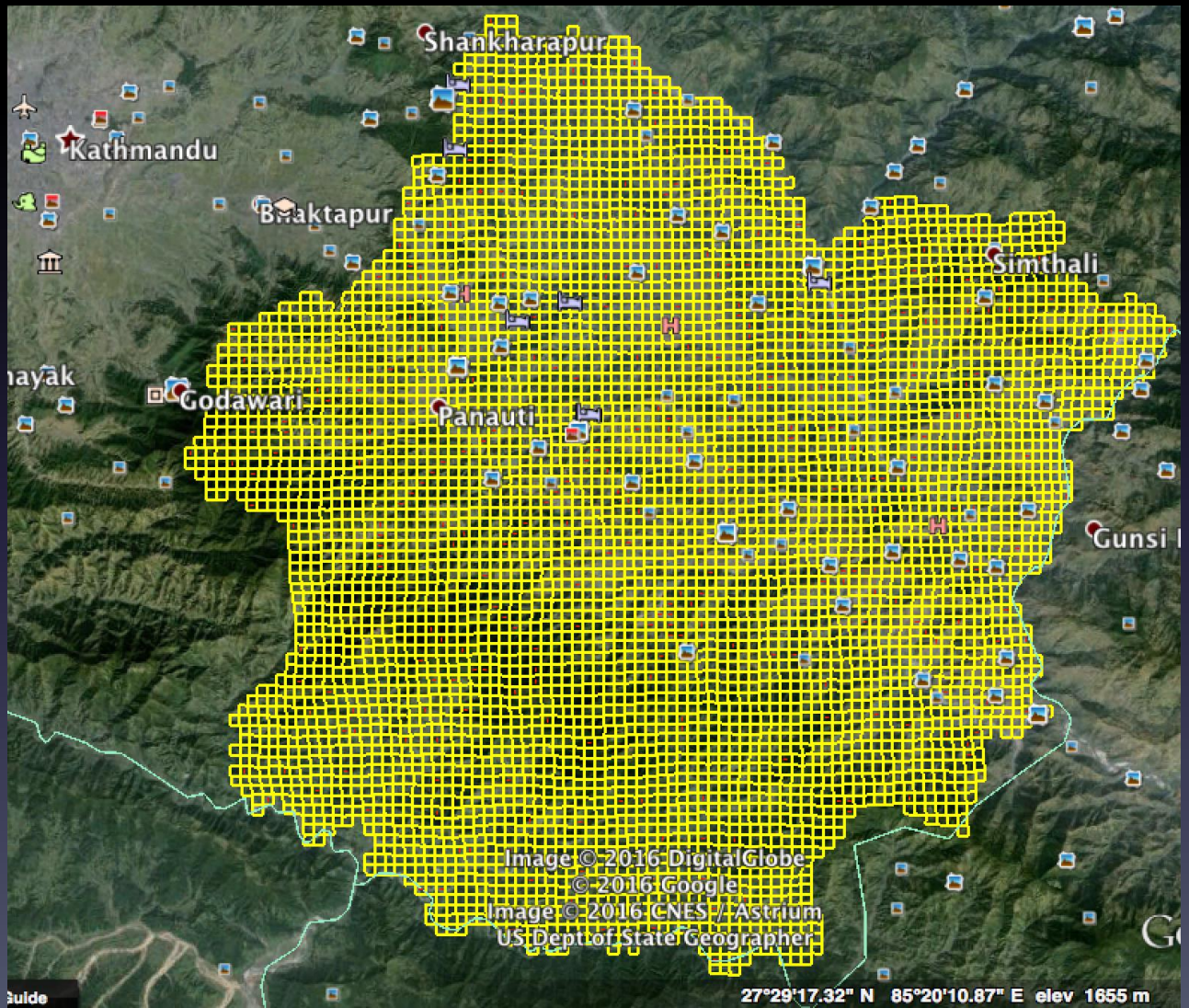
Field Tests Protocols

- Brazil (UPFE)
  - Pop 2010 census (2.6 M holdings) versus AG 2006 census (5.2 M holdings) -> List frames
  - SNPA (LF&R), PREVS (AF), Land Use Project (RS-MODIS)
  - Various registers exist
- Nepal (CBS)
  - No annual survey
  - 2011 Population census
  - 2012 census based on two stages sample (130.000 holdings)
- Rwanda (NSIR)
  - 2008 Agriculture census
  - 2013-2016 Multiple frame: Area Frame and Register
  - PSU (physical limits segments), SSU's of 10/50 ha, sample size 540
  - Register of 499 Special farms

Field Test in Nepal

- Point Area Frame in Kavre and Chitawan districts
- MSF: square segments of 25ha (5562+8968)
- Stratified sample of clusters (PSU's)
- 200 clusters of 4 points: land use & farmer interview
- Land Use/cover survey
- Farmers/Holdings survey

Frame Overview in Kavre



Software: Collect Earth

When needed: Bing imagery

Full detail for Photo-interpretation and Field Work

Land Use survey

- **Areas estimation** will be based on the formulas of stratified point sampling.
- Working at district (i) and stratum (k) level and A_{ik} being the area of stratum k in district i, the total area of a crop (c) in stratum (k) of district (i) will be:
-
- $Y_{cik} = A_{ik} (n_{cik}/n_{ik})$
-
- With A_{ik} = Nik *25ha (N_{ik} = number of population nodes in stratum k of district i).
- With n_{cik} being the number of points in stratum k of district i occupied by crop c.
- With n_{ik} being the number of points in stratum k of district i
-
- Its variance will be estimated considering that n_{cik} is Binomial (n_{ik}, p= n_{cik}/n_{ik}) by
- $V(Y_{cik}) = A_{ik}^2 p(1-p) / n_{ik}$
- $CV^2(Y_{cik}) = (1/ n_{ik}) (1-p)/p$

Land Use Survey

- Considering that paddy rice and maize occupy respectively 16.6% and 9.3% of the land in Chitwan and 7.2% and 14.1% in Kavre, in absence of stratification we can expect CV's respectively of 7.9, 11.1, 12.6 and 8.7 % with our 800 points per district.
- Considering a stratification efficiency of two, thus would result in an accuracy comparable to the one achieved in the Agriculture census with half the sample size and no need to measure/ask any area.

Holdings Survey

- A farm f, from district i and strata k, has a sampling probability equal to $\pi_{ikf} = n_{ik} (S_{ikf} / A_{ik})$
- So that the HT estimator of a total (based on the sampled farms of size n'_{ik}) is
- $$Y_{ik} = \sum_{f}^{n'_{ik}} Y_{ikf}/\pi_{ikf} = (A_{ik}/n_{ik}) \sum_{f}^{n'_{ik}} (Y_{ikf}/S_{ikf})$$
- In particular to estimate the number of farm we will take Y=1 and the estimator becomes:
- $Y_{ik} = (A_{ik} /H_{ik}).(n'_{ik}/n_{ik})$
- with H_{ik} being the harmonic mean of the farm agricultural area for the farms in the sample
- Average estimation for farm characteristics will thus be a ratio estimator.
- The variance of the estimation of a total will be given by:
- $$V(Y_{ik}) = (1/(n'_{ik}(n'_{ik} -1))) \sum_{f}^{n'_{ik}} (y_{ikf}/\pi_{ikf}) - Y_{ik})^2$$
- The variance of the estimation of an average will be based on the variance of a ratio:
- (Mood, Graybill and Boes 1974, p181)
- $V(X/Y) = (m_x/m_y)^2 (CV^2(X) + CV^2(Y) - 2COV(X,Y)/(m_x.m_y))$

Holdings Surveys

Other estimators exist, see:

Gallego F.J., Delincé J., Carfagna E., 1994,

Two-Stage Area Frame Sampling on Square

Segments for Farm Surveys.

Survey Methodology , vol 20, No. 2 , pp. 107-115

Work will consist in

- Estimation of the parameters of interest (total, average) and their variances for the variables of interest (+/- 20 on crops areas/production, nitrogen/water use, transportation costs, livestock –cattle, chicken, pigs, Sheeps/goats) either from the land use survey (crops areas), either from the holdings survey (all).
- Comparison of estimations and variances for farm data using both approaches (sample of points, sample of farms).
- Estimation of the stratification efficiency
- Quantification of times and costs associated to the various stages of the survey
- Comparison of the 2011 agriculture census and the current point area frame in term of their precision, the problems encountered and their costs.

Activity Planning

- June/July: Stratification
- Sept./October: Land Use and Holding surveys
- November/December: Simulations & Reporting

Thanks for your attention