Fifth GFOI Regional Technical Workshop: Accuracy Assessment

January 25-29 2016, Santa Rosa, Philippines

Statistical principles of design and analysis for accuracy assessment and area estimation

P. Olofsson and S.V. Stehman

BOSTON UNIVERSITY

"Why is this necessary?"

- A very common result of remote sensing is an estimate of an amount, or area (e.g. change map)
- The strength of remote sensing is that it often allows wall-to-wall coverage of the area under consideration
- The weakness of remote sensing is that results are never perfect (e.g., a classified map contains error)
- If the map has errors, then mapped areas of the map categories are incorrect and need adjustment for errors, and uncertainty in area quantified

BOSTON UNIVERSITY

Not new stuff! We're making use of Cochran (1977) and Särndal et al. (1992); Card (1982) used these techniques to improve map estimates

Sampling Techniques

third edition

WILLIAM G. COCHRAN

Professor of Statistics, Emeritus
Harvard University

JOHN WILEY & SONS
New York • Chichester • Brisbane • Toronto • Singapore

Copyrighted Material

Springer Series in Statistics

Carl-Erik Särndal
Bengt Swensson
Jan Wretman

Model Assisted Survey Sampling

Springer

Copyrighted Material

Don H. Card
NASA Ames Research Center
Moffett Field, CA 94035

Using Known Map Category Marginal Frequencies to Improve Estimates of Thematic Map Accuracy

Two statistical sampling schemes are discussed: simple random sampling of single points and random sampling stratified by map category.

Introduction

The introduction of satellite-acquired imagery in the early 1970s and the concurrent advances in techniques of computer classification of digital remotely sensed data have made possible the rapid generation of land-use and land-cover maps. However, user acceptance of this kind of computer-generated map has lagged behind the technology and prevented realization of its full potential, principally because of difficulties inher-

1979); Hord and Brooner, 1976; Hay, 1979; van Genderen and Lock, 1977; van Genderen *et al.*, 1978), a unified treatment of the subject has been lacking. This writer suggests that "contingency-table" analysis is the most natural framework for accuracy assessment, both for the convenient display of empirical results and for ease of statistical analysis. This idea is not new—other authors have displayed their data in the form of a cross-tabulation of map category versus true category; however, the statistical treatment of these tables has

Abstract: *It is generally recognized that estimating the accuracy of maps that are derived from remotely sensed data requires statistical sampling of photographs or ground plots to insure that the estimates are reliable and cost effective. The usual method is to cross-tabulate the categories identified for these plots with the categories associated with corresponding areas on the map in a table called a "contingency table." From this table, measures of map accuracy ("proportion-correct") are usually obtained by ratioing diagonal entries by marginal sums or by the total number of points in the table. Frequently, one has knowledge of the true map category marginal proportions, that is, the relative areas of each map category. These map category proportions can be used to improve estimates of "proportion-correct" for each map category. This paper derives these improved estimates with their asymptotic variances for two common sampling designs: simple random sampling of single points and sampling stratified by map category. A numerical example illustrates the computations.*

BOSTON UNIVERSITY

Some terminology

- We identify classification errors in a map by designing and implementing an ***accuracy assessment***
- ***Sample*** the map (i.e. the population) and collect ***reference observations*** for sample: best assessment of true class at a given location
- ***Reference data***: information used to obtain the reference class
- By comparing the map and reference labels – compute map ***accuracy***: the degree to which the map corresponds to the reference condition

BOSTON UNIVERSITY

More terminology

- ***Statistical inference***: collect information about population (map) by analyzing sample of population
- A conclusion of the inference is an ***estimate*** that best approximates a particular population parameter
- A ***confidence interval*** of the estimate is the interval that contain the true parameter value with a probability of the ***confidence level*** if the sampling is repeated many times
- ***Bias*** is the difference between an estimator's expected value and the true parameter value

Components of Accuracy Assessment

- ***Sampling design:*** Decide which elements of the population (map) to visit
- ***Response design:*** Determine the true class or "reference condition" at a location
- ***Analysis:*** Organize and summarize data to make inference (accuracy, area) about the population (map)

- *Where will we observe the true condition?*
- *What is the true (reference) condition?*
- *And how will we use the data?*

Sampling design

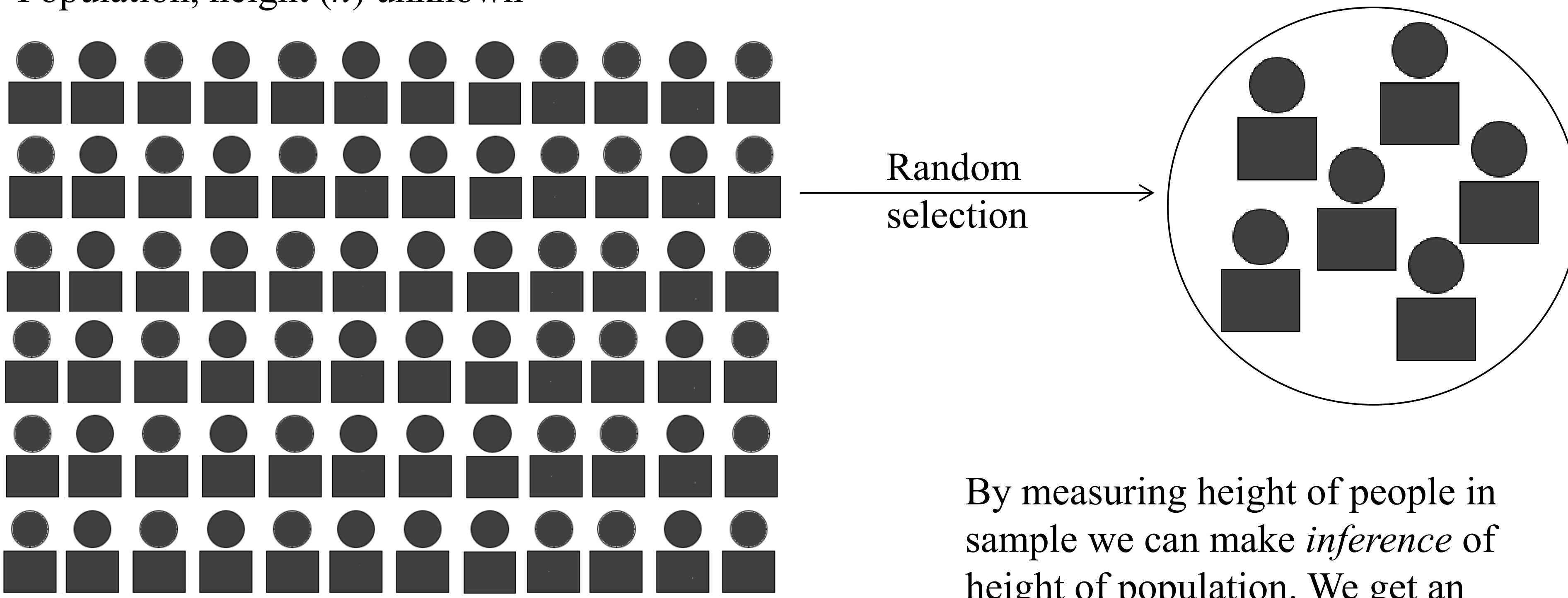
Many options but we want:

- A probability sample
- Practical to implement given response design
- Yields estimates with small standard errors
- Low cost
- Spatially well distributed across country
- Can allow for change in sample size
- Unbiased estimator of variance (not approximate)

Simple random sample

Population, height (h) unknown

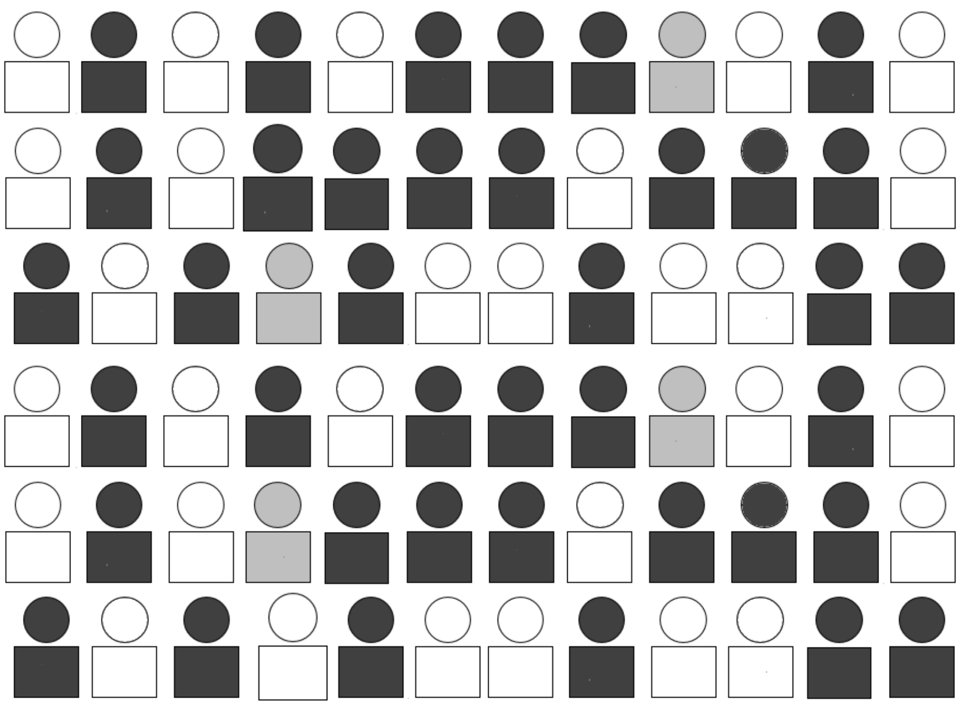
Random sample, height measured



By measuring height of people in sample we can make *inference* of height of population. We get an *estimate* of population height ($\hat{h}$).

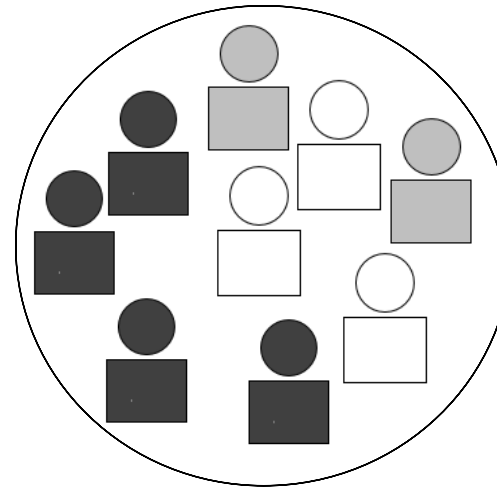
Stratified random sample

Population, height (h) unknown



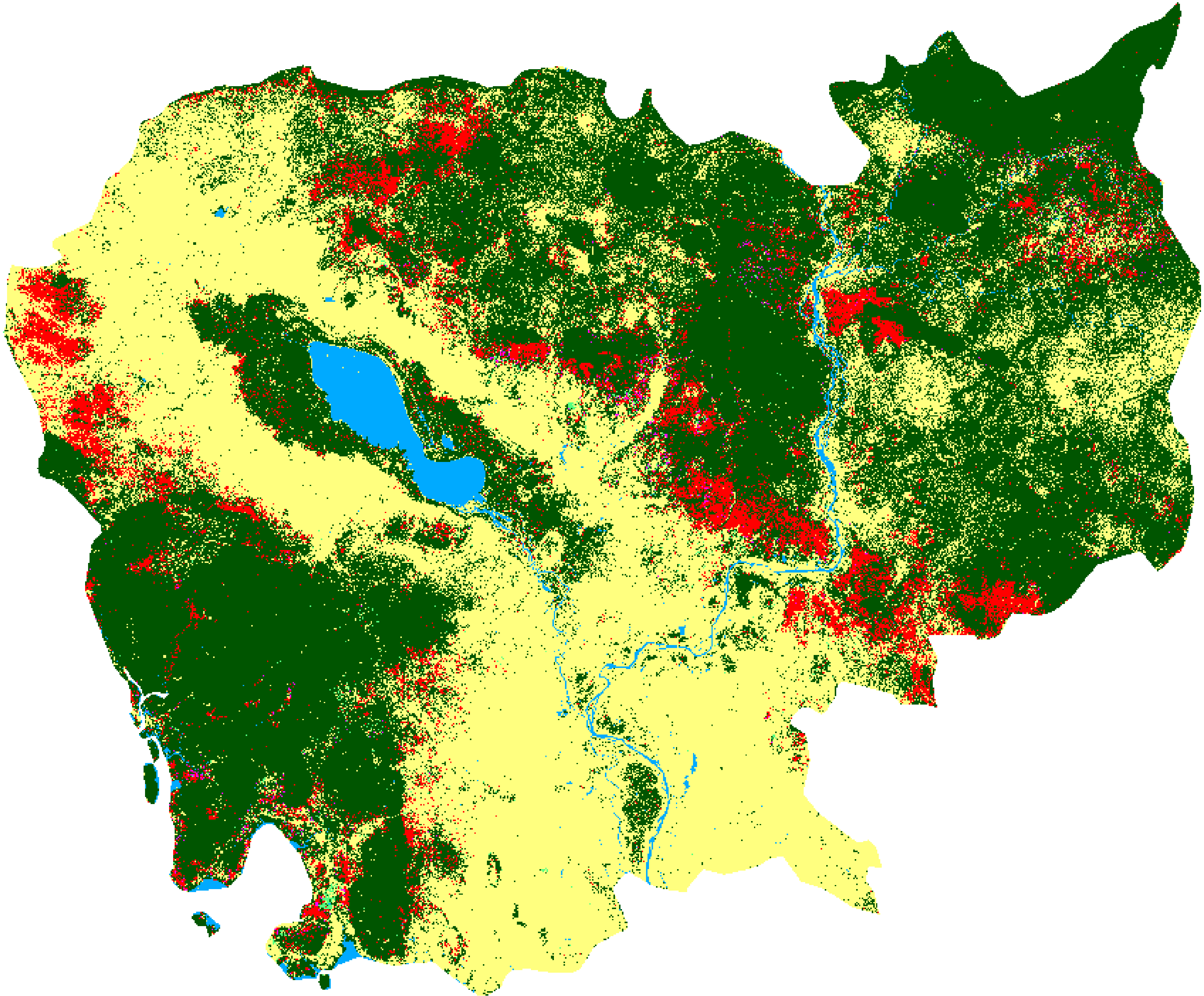
Stratified random selection; *3 strata*

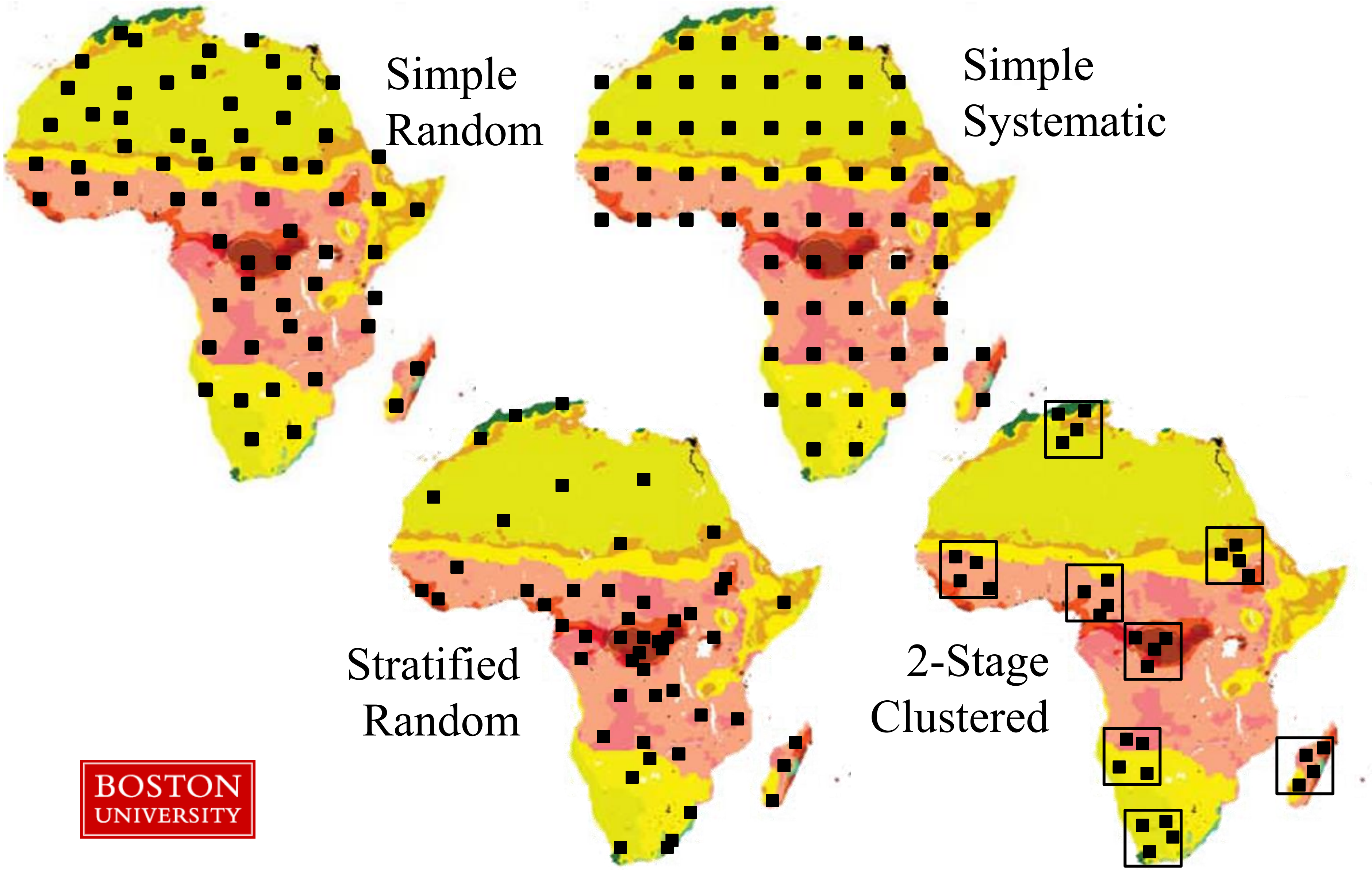
Stratified random sample, height measured



We get a more *precise* estimate of the of the height and we can get the height of each *stratum* in the population ($\hat{h}_g, \hat{h}_b, \hat{h}_w$).

BOSTON UNIVERSITY





BOSTON UNIVERSITY

Basic Sampling Design Questions

- *Strata* – group assessment units into strata?
- *Clusters* – sample assessment units individually or in groups?
- *Selection protocol* – select the sample of assessment units by a systematic or a simple random protocol?

Basic Sampling Design Questions

Decisions depend objective of analysis and design criteria

- *Clusters* – use for costs but complex variance estimation, larger variance and spatial correlation
- *Strata* – use for objectives and precision
- *Selection protocol* – SYS more difficult to combine with stratification by map class; SRS unappealing distribution

BOSTON UNIVERSITY

Response Design

- The method use to decide reference condition
- Reference condition decided without knowing map class of the location
- Four main features of response design
    - Information to decide reference condition
    - Spatial unit of the assessment
    - Assign reference condition
    - How do define if map class and reference condition agree

BOSTON UNIVERSITY

Spatial Unit of Assessment

- Partition region of interest into non-overlapping spatial units
- Common choices are
  - Pixel
  - Block of pixels (e.g. 3x3, 5x5)
  - Segment
- Sampling design depends on choice of spatial unit

BOSTON UNIVERSITY

Sources of Information to Determine Reference Condition

- Google Earth
- Landsat
- RapidEye
- Ground visit (e.g., National Forest Inventory)

- Decision will impact sampling design options (e.g., should we use clusters or not)

Analysis

- Accuracy
  - Define parameters that describe accuracy
  - Estimate parameters from sample
  - Estimate standard errors
- Area
  - Reference condition is basis of estimate
  - Increase precision by incorporating map information into area estimator
  - Estimate area and standard errors

BOSTON UNIVERSITY

Error matrix

		Reference		
		Forest loss	No loss	Map prop.
Map	Forest loss	p_{11}	p_{12}	p_{1+}
	No loss	p_{21}	p_{22}	p_{2+}
	Ref. prop.	p_{+1}	p_{+2}	1

BOSTON UNIVERSITY

Bias-adjusted estimator: add omission error and subtract commission error

Reference

Map		Forest loss	No loss	Map prop.
	Forest loss	p_{11}	p_{12}	p_{1+}
	No loss	p_{21}	p_{22}	p_{2+}
	Ref. prop.	p_{+1}	p_{+2}	1

$$\hat{p}_{+1} = p_{1+} + (\hat{p}_{21} - \hat{p}_{12})$$

Stratified/post-stratified estimator

Map		*Reference* Forest loss	No loss	Map prop.
	Forest loss	p_{11}	p_{12}	p_{1+}
	No loss	p_{21}	p_{22}	p_{2+}
	Ref. prop.	p_{+1}	p_{+2}	1

$$\hat{p}_{+1} = \sum_{i=1}^{q} \hat{p}_{ij} = \hat{p}_{11} + \hat{p}_{21}$$

BOSTON UNIVERSITY

Area estimators

Bias-adjusted estimator

- Unbiased for any sample size
- Known as a “difference” estimator in sampling texts
- More efficient if map class is continuous

Stratified/Post-stratified

- Unbiased (but problem if no units from a post-stratum)
- Allows use of all map classes as post-strata
- More efficient if map classes is categorical

Measures of accuracy

		Reference	
Map	Forest loss	No loss	Map prop.
Forest loss	p_{11}	p_{12}	p_{1+}
No loss	p_{21}	p_{22}	p_{2+}
Ref. prop.	p_{+1}	p_{+2}	1

$$O = \sum_{i=1}^{q} \hat{p}_{jj} = \hat{p}_{11} + \hat{p}_{22}$$

$$U_i = \hat{p}_{ii} \div \hat{p}_{i+} = \hat{p}_{11} \div \hat{p}_{1+}$$

$$P_j = \hat{p}_{jj} \div \hat{p}_{j+} = \hat{p}_{11} \div \hat{p}_{+1}$$

BOSTON UNIVERSITY

Conclusions

- All maps have errors – can't count pixels – *inference of area* from reference sample *necessary*
- Sample design depends on assessment objectives and practicalities
- We use the map to increase precision in area estimates – we are *not validating* the map
- *Area* estimation typically of first priority – *accuracy* secondary

BOSTON UNIVERSITY