

Use of Stratification: Design and Estimation

Steve Stehman

Lessons Learned from Accuracy
Assessments in the Context of REDD+

Oslo, 26-28 June

Brief Background

What determines sampling design for accuracy assessment and area estimation?

- **Project objectives**
- Response design
- Statistical inference framework
- **Use of auxiliary information**
 - **Design**
 - **Estimation**
- Prioritized design criteria

Prioritized design criteria influence choice of sampling design.

- Probability sample
- Yields small standard errors (precision)
- Practical to implement
- Simplicity of analysis
- Cost effective
- Spatially well distributed
- Easy to change sample size
- Unbiased estimator of variance

*** Prioritize – difficult to satisfy all**

Objectives determine need for strata and sample allocation to strata.

- Area estimates based on reference sample by
 - Geographic subregions
 - Forest type, emissions factor?
 - Change type
 - Deforestation primary and secondary forest
 - Degradation
 - Time
- Map accuracy (overall and per class)

Objectives: Time

- Single period
- Multiple change periods (**monitoring**)
- Other (use of estimates over time)
 - Establishing historical reference?
 - Compare change period to historical reference?
 - 2016-2018 to historical reference
 - Compare two change periods
 - 2016-2018 compared to 2018-2020
 - Trend in 2-year change periods

Use of auxiliary information is strongly influenced by objectives.

- Auxiliary information can be used in either
 - **Sampling design** (single or few objectives)
 - **Estimation** (multiple objectives)
- Sampling design
 - Fixed structure after sample selected
 - Optimizing for one objective may detract from other objectives
- Estimation is more flexible to use of auxiliary information – different information for different objectives

What can we learn from completed assessments?

Stratification – Comparing Standard Errors from Completed Studies

1) Did stratification help?

- Compare stratified to direct estimator (no map).

Use of strata

- Estimation: post-stratified
- Design: control allocation of sample to strata
- Proportional allocated design = Post-stratified

2) Proportional allocation versus actual allocation

3) Proportional allocation versus optimal allocation

4) Actual allocation versus optimal allocation

Variance of Stratified Estimator

$$V(\hat{p}) = \sum_h W_h^2 p_h (1 - p_h) / n_h$$

W_h = proportion of area in map stratum h

p_h = proportion of target class in stratum h

Costa Rica DEFOR 2001-2011

P = %area	<u>5.60</u>
SRS =	0.91
Opt =	0.71
Prop =	0.77
Actual=	0.74
Ratio SRS to Actual	1.23
Ratio Post to Actual	1.04
Ratio Post to Optimal	1.08
95% CI Actual	(4.12 to 7.08)
95% Post 5000	(5.05 to 6.15)

Republic of Congo

P = 0.70

SRS = 0.27

Opt = 0.17

Prop = 0.23

Actual= 0.19

Ratio SRS to Actual 1.42

Ratio Prop to Actual 1.21

Ratio Prop to Opt 1.35

95% CI Actual (0.32 to 1.08)

95% CI Post 5000 (0.50 to 0.90)

Deforestation, Actual 95% CI

1	Country	Strata	n	p(%)	Lower	Upper
2	Ethiopia	4	639	1.48	0.26	2.70
3	Costa Rica	4	639	5.60	4.12	7.08
4	Madagascar	6	594	5.21	3.89	6.53
5	Congo	5	931	0.70	0.32	1.08
6	Ghana	2	2929	14.81	13.97	16.13
7	Ivory Coast	8	2600	10.07	9.33	10.81
8	Vietnam	6	536	2.98	2.38	3.58

95% CI for Post-stratified, n=5000

1	Country	Lower	Upper
2	Ethiopia	1.14	1.82
3	Costa Rica	5.05	6.15
4	Madagascar	4.71	5.71
5	Congo	0.50	0.90
6	Ghana	14.17	15.45
7	Ivory Coast	9.45	10.69
8	Vietnam	2.75	3.25

Use of Strata: 1) Better than direct? 2) Stratified design or estimator?

1	Country	Dir/Actual	Post/Act	Post/Opt
2	Ethiopia	0.79	0.77	1.04
3	Costa Rica	1.23	1.04	1.08
4	Madagascar	1.38	1.09	1.26
5	Congo	1.42	1.21	1.35
6	Ghana	1.57	1.00	1.20
7	Ivory Coast	1.59	1.16	1.23
8	Vietnam	2.00	1.23	1.85

Summary of Use of Stratification

- Stratified design as implemented usually better than direct estimator (not Ethiopia)
- Post-stratified higher standard errors than actual (MGD) allocation
- Optimal allocation would improve relative to post-stratified
- Should look at optimal versus actual (MGD)

Summary of Use of Stratification

- Results only for single target – deforestation
- Evaluate other estimates (examples from Erik and Andres presentations)

Summary of Use of Stratification

- Optimizing allocation (design) for one target estimate may increase standard errors for other estimates
- Is post-stratified estimation best choice for multiple estimates?
- Optimize for (Activity x EF) ?

Other Topics

Impact of Omission Errors in Large Strata

- Look at sample allocation for specific examples where this problem occurred
- Proposed remedies:
 - Change allocation of design (future projects)
 - Post-stratify within “No change” strata
 - Example: buffer mapped change
 - Caution: Discourage searching for post-strata until the omission errors are removed

Example: ROC (Christophe)

- Change / No Change
- User=75%, Producer=53%
- P% = 0.93 change

Standard errors (n=1000)

	User 75%	User 50%	User 50%
	Prod 53%	Prod 86%	Prod 50%
Direct	0.30	0.30	0.30
Post	0.24	0.23	0.26
Opt	0.22	0.14	0.23

Relative Error versus p

- For SRS relative error is $(1/p-1)$

