Estimation of area for REDD+ using stratified sampling estimator

Rémi D'Annunzio

16-19 May 2017, Phnom Penh, Cambodia

remi.dannunzio@fao.org

Uncertainty in UNFCCC context

- IPCC defines greenhouse gas (GHG) inventories consistent with "good practice" as those which "contain neither over-nor underestimates so far as can be judged, and in which uncertainties are reduced as far as practicable."

- In particular, information on uncertainty can be used to develop *conservative* REDD+ estimates, to ensure that reductions in emissions or increases in removals are not overestimated.

Source: GOFC-GOLD / WB FCPF training materials for REDD+ monitoring and reporting

Uncertainties in area changes

- In REDD+ context, a Tier 3 estimate of area and/or area change typically results from analysis of a RS product
- Maps only are subject to classification errors that induce bias into estimations and sampling only usually miss classes of interest
- A suitable approach is to **use the map as a stratification device** and collect reference data from points to produce area estimates.

FREL submissions

YEAR	FREL	Deforestation	Degradation	Method
2014	Brazil	Brazil		Map
2015	Colombia	Colombia		Map
	Ecuador	Ecuador		Map
	Guyana	Guyana	Guyana	Map
	Malaysia			
	Mexico	Mexico		
2016	Chile	Chile	Chile	Map + Strat Sampling
	Congo	Congo	Congo	Map + Strat Sampling
	Costa Rica	Costa Rica		
	Ethiopia	Ethiopia		Map + Strat Sampling
	Indonesia	Indonesia	Indonesia	Map
	Paraguay	Paraguay		
	Peru	Peru		Map + Strat Sampling
	Viet Nam	Viet Nam	Viet Nam	Map + Strat Sampling
	Zambia	Zambia		Map + Strat Sampling
2017	Cambodia	Cambodia		Map
	Côte d'Ivoire	Côte d'Ivoire		Map + Strat Sampling
	Ghana	Ghana	Ghana	Map + Strat Sampling
	Honduras	Honduras		
	Madagascar	Madagascar		Map + Strat Sampling
	Nepal	Nepal	Nepal	Map + Strat Sampling
	Papua New Guinea	Papua New Guinea	Papua New Guinea	Syst Sampling
	Sri Lanka	Sri Lanka		Map + Strat Sampling
	Tanzania	Tanzania		
	Uganda	Uganda	Uganda	Map + Strat Sampling

Objectives

- Understand how maps are used as stratification for area estimation
- Design a SAE system (sampling + reference collection)
- Create AA outputs (confusion matrix, adjusted estimates) and compare with other area estimations

Work plan

Day 1 Refresh on stratified sampling area estimation

Discussion FAO/WB/Donors on area estimation using maps
Review of national data, introduction to SEPAL
Creating a sampling design (hands on with SEPAL)

Day 2 R and GEE training

Obtaining data from GEE + Sentinel portals (hands on GEE)
Time series generation (hands on with R) / Collect Earth exercise

Day 3 Analysis of results

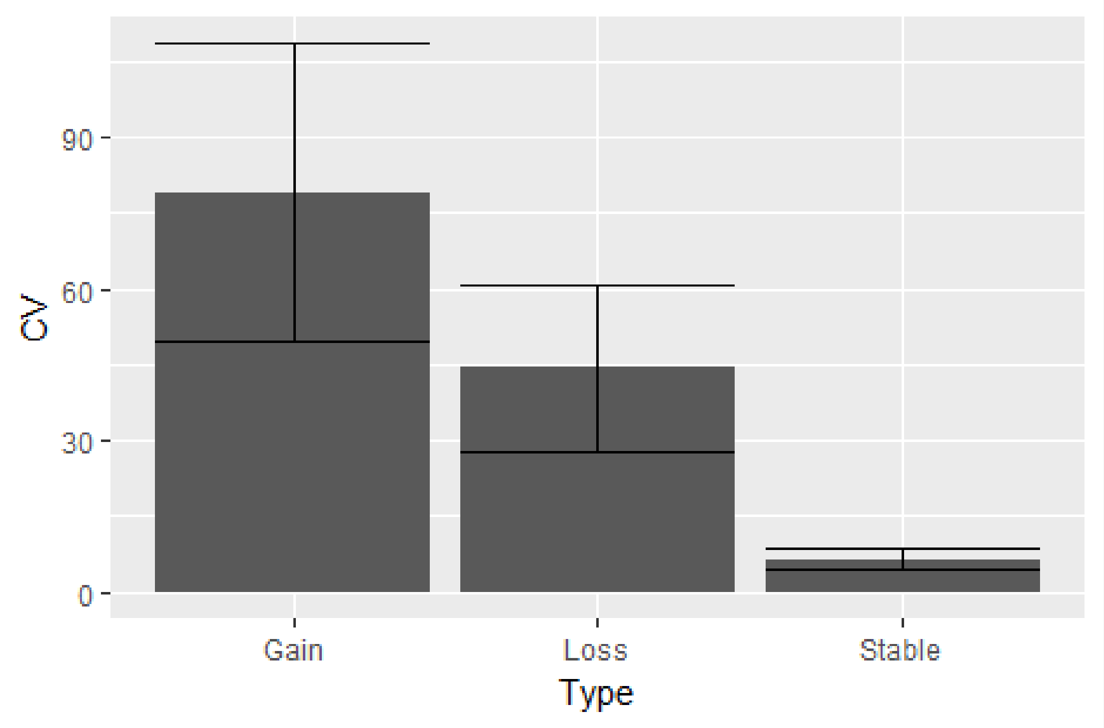
Confusion matrices, Stratified estimators (hands on with SEPAL)
Comparison with systematic sampling design results

Day 4 Evaluation + Discussion on further steps

Open session

Meeting April 2017 in FAO Rome

Name	Entity
Pontus Olofsson	Boston university
Stephen Stehman	University of New York
Ronald Mc Roberts	US Forest Services
Martin Herold	Wageningen University
Danny Donoghue	University of Durham
Frédéric Achard	Joint Research Center
Maarten Van der Eynden	Norway
Kay Kallweit	Germany
Andres Espejo	FCPF
Marco Van der Linden	FCPF
Rama Chandra Reddy	FCPF
Andy Gillespie	FCPF
Christophe Sannier	SIRS
Carly Green	GFOI
Naikoa Aguilar-Amuchastegui	WWF



Recommendations

In terms of area estimation and wording used

- The principle purpose for collecting reference data is for area estimation.
- We suggest not to include map estimates (pixel counts) in reports.
- Include map accuracy figures as a separate section from parameter estimates

In terms of assessing the quality of results and usefulness of the method

- Compute results from the sampling device as simple random
- Compare the parameters obtained with simple VS stratified random sampling

In terms of response design for stratified area estimates:

- No matter if the map is pixel-based or polygon-based : use point based interpretation

In terms of improvement of the method

- Assess the effects of adding samples to reduce the effects of errors
- Assess Effects of stratifying by error class

Issues on maps for area estimation

All maps have errors (bias)

change maps have cumulated errors

we must account for these

Reference data are required

of higher quality than data used to make map;

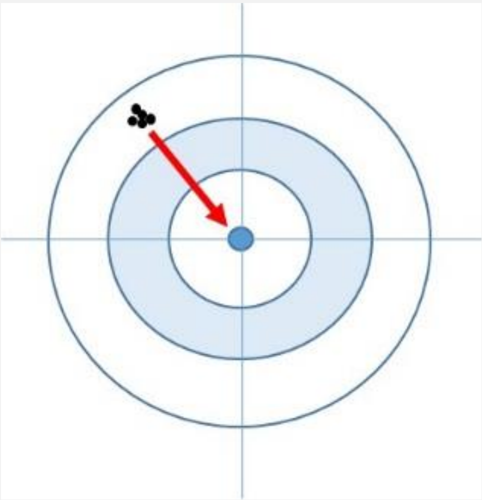
can be ‘better’ data, or ‘better’ interpretation

Purpose is to generate new area estimates

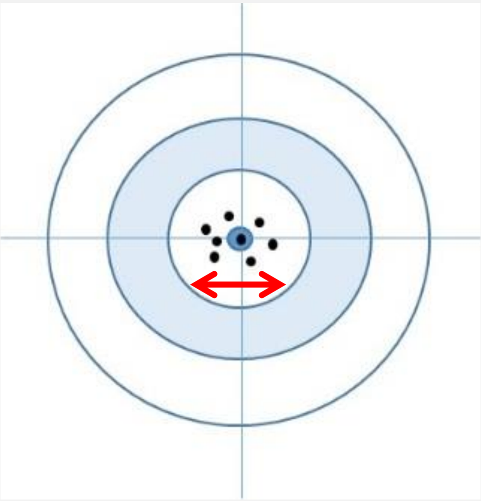
based on the stratified sampling

which generate with confidence intervals

What type of errors in a map ?



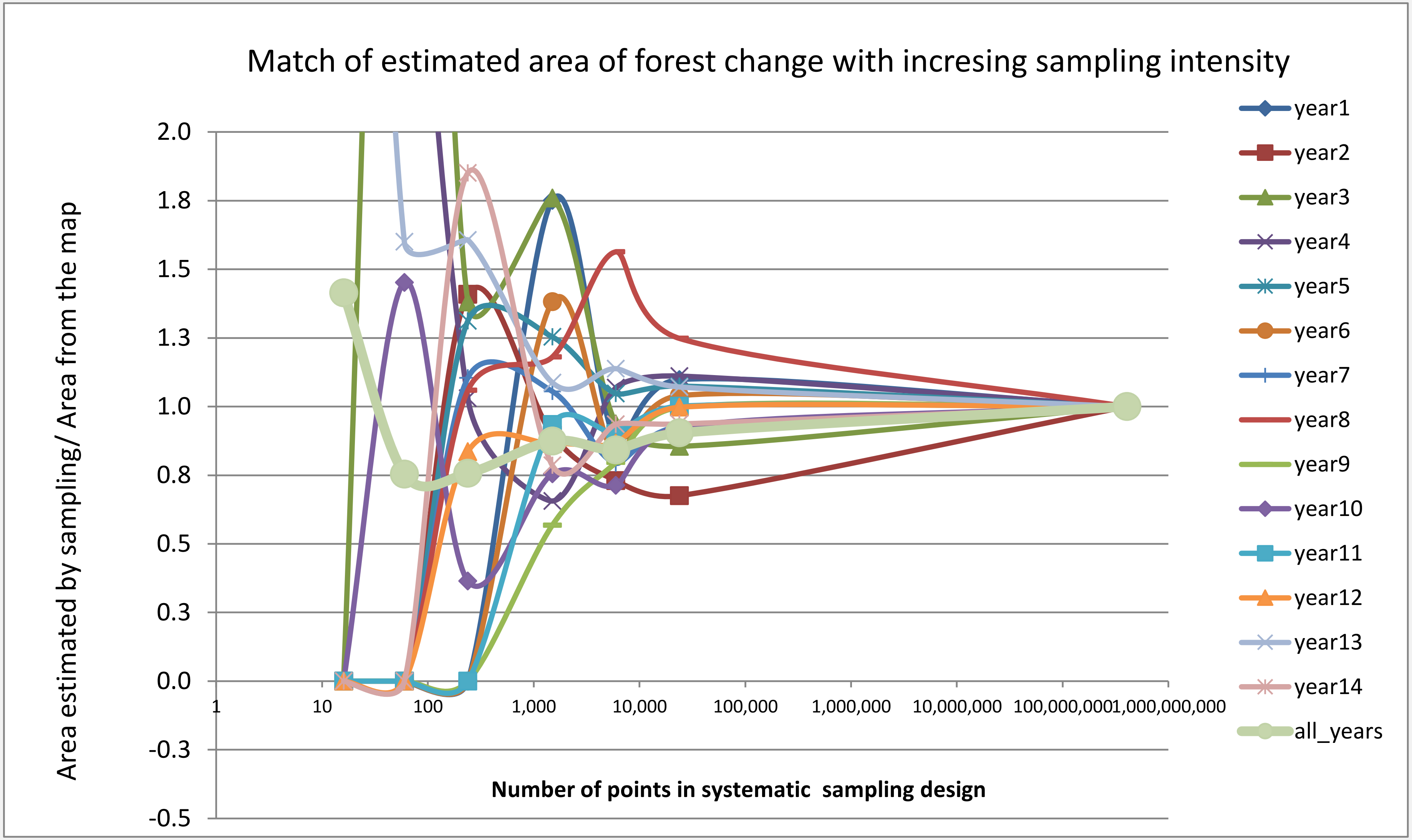
Systematic error: accuracy



Random error: precision

The accuracy assessment aims at providing a measure of both

Issues with pure sampling approaches



Good practices for assessing accuracy of land change

UN-REDD PROGRAMME

Remote Sensing of Environment 148 (2014) 42–57

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Review

Good practices for estimating area and assessing accuracy of land change

Pontus Olofsson [a,*], Giles M. Foody [b], Martin Herold [c], Stephen V. Stehman [d], Curtis E. Woodcock [a], Michael A. Wulder [e]

[a] Department of Earth and Environment, Boston University, 685 Commonwealth Avenue, Boston, MA 02215, USA
[b] School of Geography, University of Nottingham, University Park, Nottingham NG7 2RD, UK
[c] Laboratory of Geo-Information Science and Remote Sensing, Wageningen University, Droevendaalsesteeg 3, 6708 Wageningen, The Netherlands
[d] Department of Forest and Natural Resources Management, State University of New York, 1 Forestry Drive, Syracuse, NY 13210, USA
[e] Canadian Forest Service (Pacific Forestry Centre), Natural Resources Canada, Victoria, BC 12 V8Z 1M5, Canada

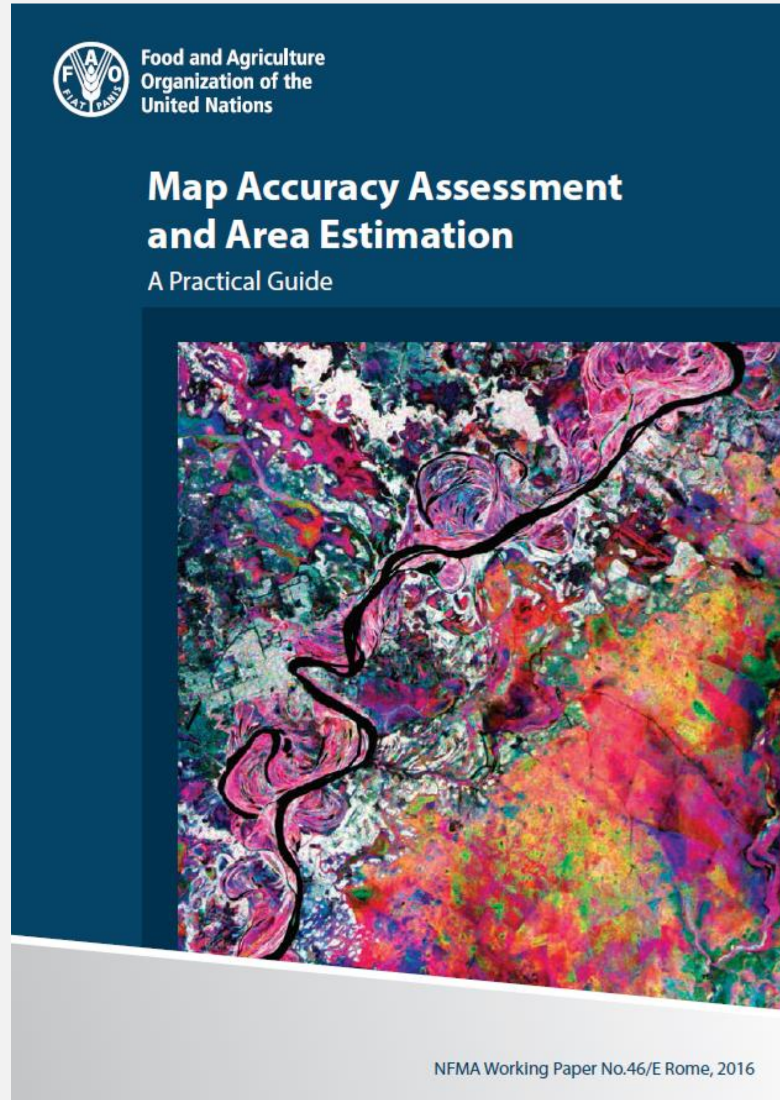
ARTICLE INFO

Article history:
Received 30 May 2013
Received in revised form 15 January 2014
Accepted 22 February 2014
Available online xxxx

Keywords:
Accuracy assessment
Sampling design
Response design
Area estimation
Land change
Remote sensing

ABSTRACT

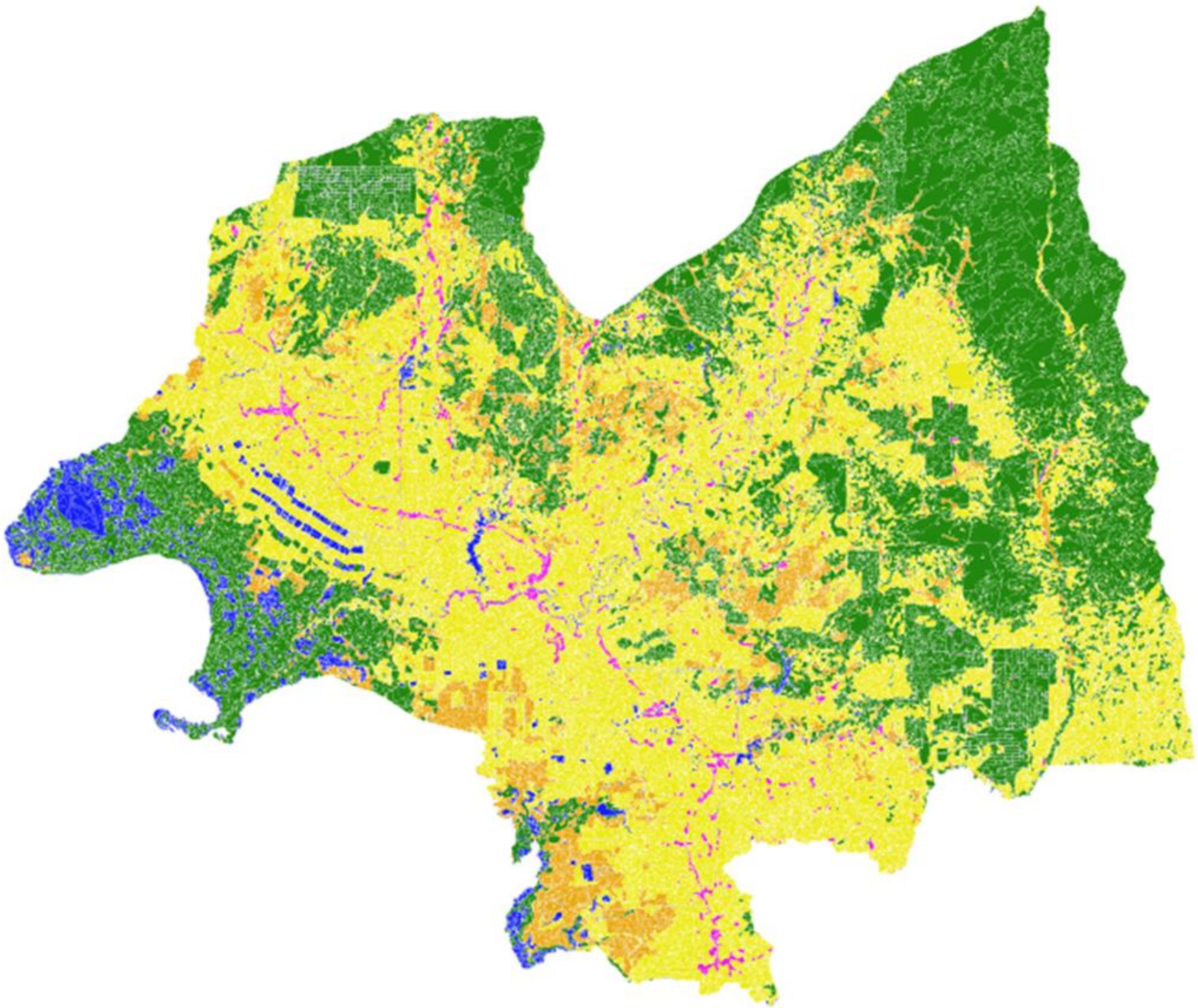
The remote sensing science and application communities have developed increasingly reliable, consistent, and robust approaches for capturing land dynamics to meet a range of information needs. Statistically robust and transparent approaches for assessing accuracy and estimating area of change are critical to ensure the integrity of land change information. We provide practitioners with a set of "good practice" recommendations for designing and implementing an accuracy assessment of a change map and estimating area based on the reference sample data. The good practice recommendations address the three major components: sampling design, response design and analysis. The primary good practice recommendations for assessing accuracy and estimating area are: (i) implement a probability sampling design that is chosen to achieve the priority objectives of accuracy and area estimation while also satisfying practical constraints such as cost and available sources of reference data; (ii) implement a response design protocol that is based on reference data sources that provide sufficient spatial and temporal representation to accurately label each unit in the sample (i.e., the "reference classification" will be considerably more accurate than the map classification being evaluated); (iii) implement an analysis that is consistent with the sampling design and response design protocols; (iv) summarize the accuracy assessment by reporting the estimated error matrix in terms of proportion of area and estimates of overall accuracy, user's accuracy (or commission error), and producer's accuracy (or omission error); (v) estimate area of classes (e.g., types of change such as wetland loss or types of persistence such as stable forest) based on the reference classification of the sample units; (vi) quantify uncertainty by reporting confidence intervals for accuracy and area parameters; (vii) evaluate variability and potential error in the reference classification; and (viii) document deviations from good practice that may substantially affect the results. An example application is provided to illustrate the recommended process.



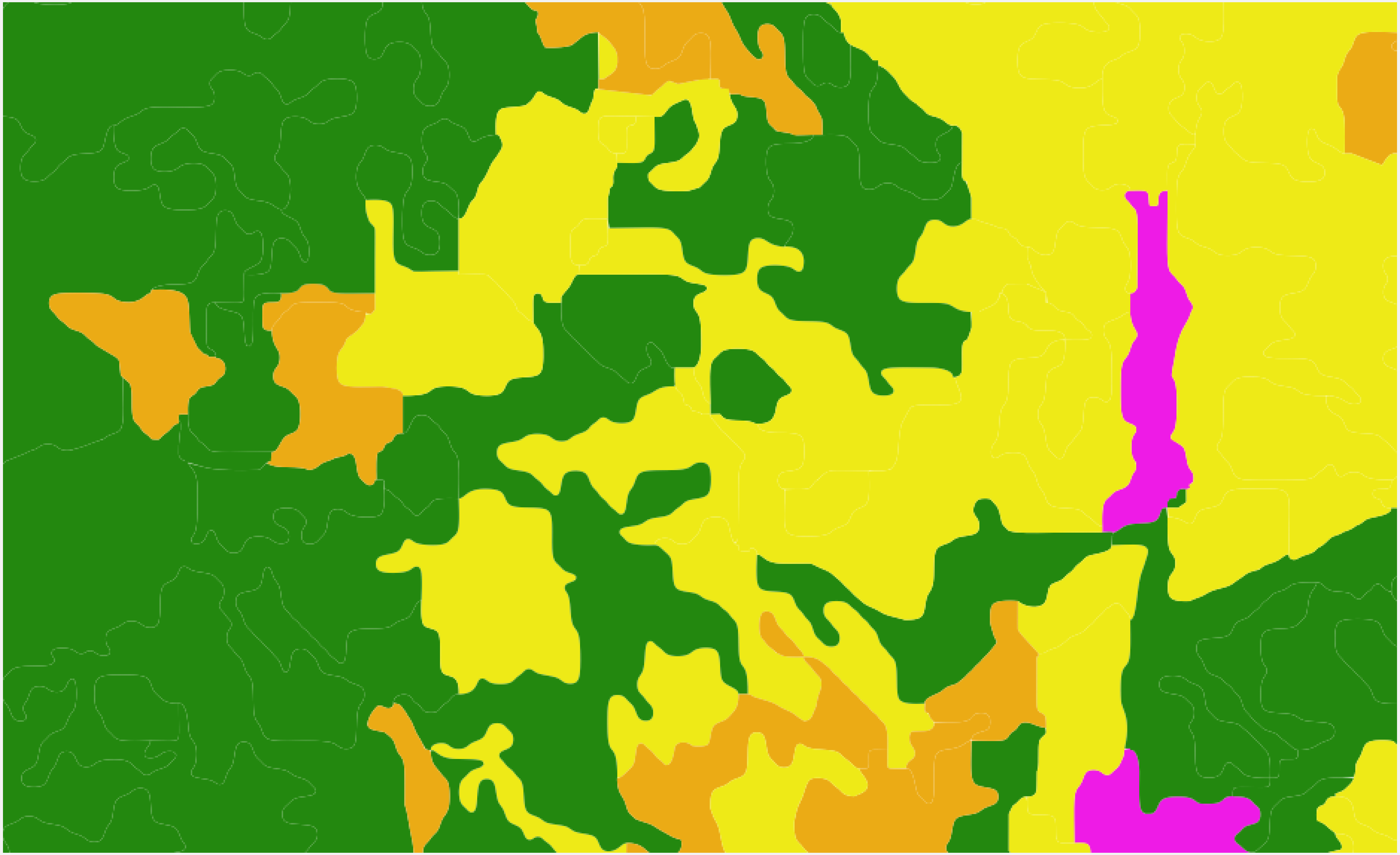
Steps of stratified area estimation

Finalized map data → Sampling design → Response design → Analysis

Land use change map to be assessed



Change in code AND shape



Map categories, transition matrix (1000 ha)

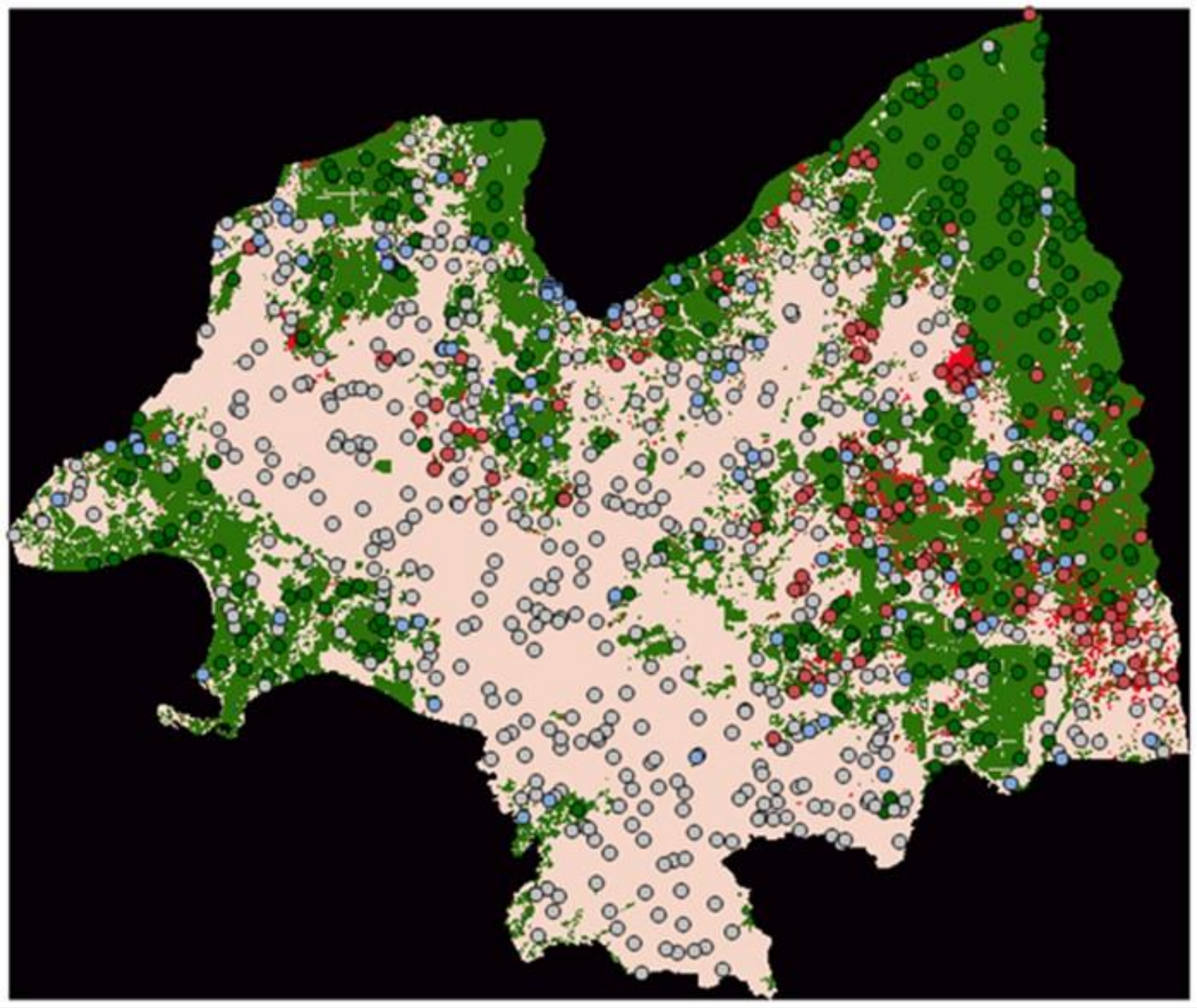
95% of the area is STABLE (1,179 Mha out of 1,245 Mha

		Ag		E		F										G		O			W	nodata	TOTAL
		Hc	Hr	Bt	Bu	B	D	E	Ff	Fr	Po	Pp	Rp	Se	Tp	G	Ws	R	S	UK			
Ag	Hc	215	0	1	0	0	0	0	0	1			1	0	0	0	0	0	0		0	0	219
	Hr	1	315	3	0	0	0	0	0	0			0	0	0	0	0	0	0		0	0	320
E	Bt	0	0	17	0	0	0	0		0			0	0	0	0	0		0		0		18
	Bu	0	0	0	0														0		0		1
F	B	0	0	0		2		0		0			0	0		0	0				0		2
	D	2	1	0	0		45	0	0	0			0	0	0	0	0	0	0	0	0	0	49
	E	20	0	0		0	0	195		0	0	0	3	0		0	1		0	0	0		219
	Ff	0	0						75							0	0		0		0		75
	Fr	8	0	0	0	0	0	0		63			1	0	0	0	0		0	0	0		73
	Po																						
	Pp																						
	Rp	0	0	0		0	0	0		0		0	70	0	0	0	0		0		0		70
	Se	1	0	0		0	0	0		0			0	15		0	0		0		0		16
	Tp	0	0	0			0		0	0			0		8	0	0		0				8
G	G	0	5	0		0	0	0	0	0		0	0	0	0	34	0		0		0		40
	Ws	2	4	0		0	0	0	0	0			0	0	0	0	79	0	0		0		87
O	R																						
	S	0	0	0			0	0	0	0	0		0	0	0	0	0		4		0		5
	UK																						
W		0	1	0	0	0	0	0	0	0			0	0		0	0		0		42	0	43
nodata		0														0							0
TOTAL		249	327	22	0	2	46	195	75	65	0	0	76	15	8	35	81	0	5	0	43	0	1,245

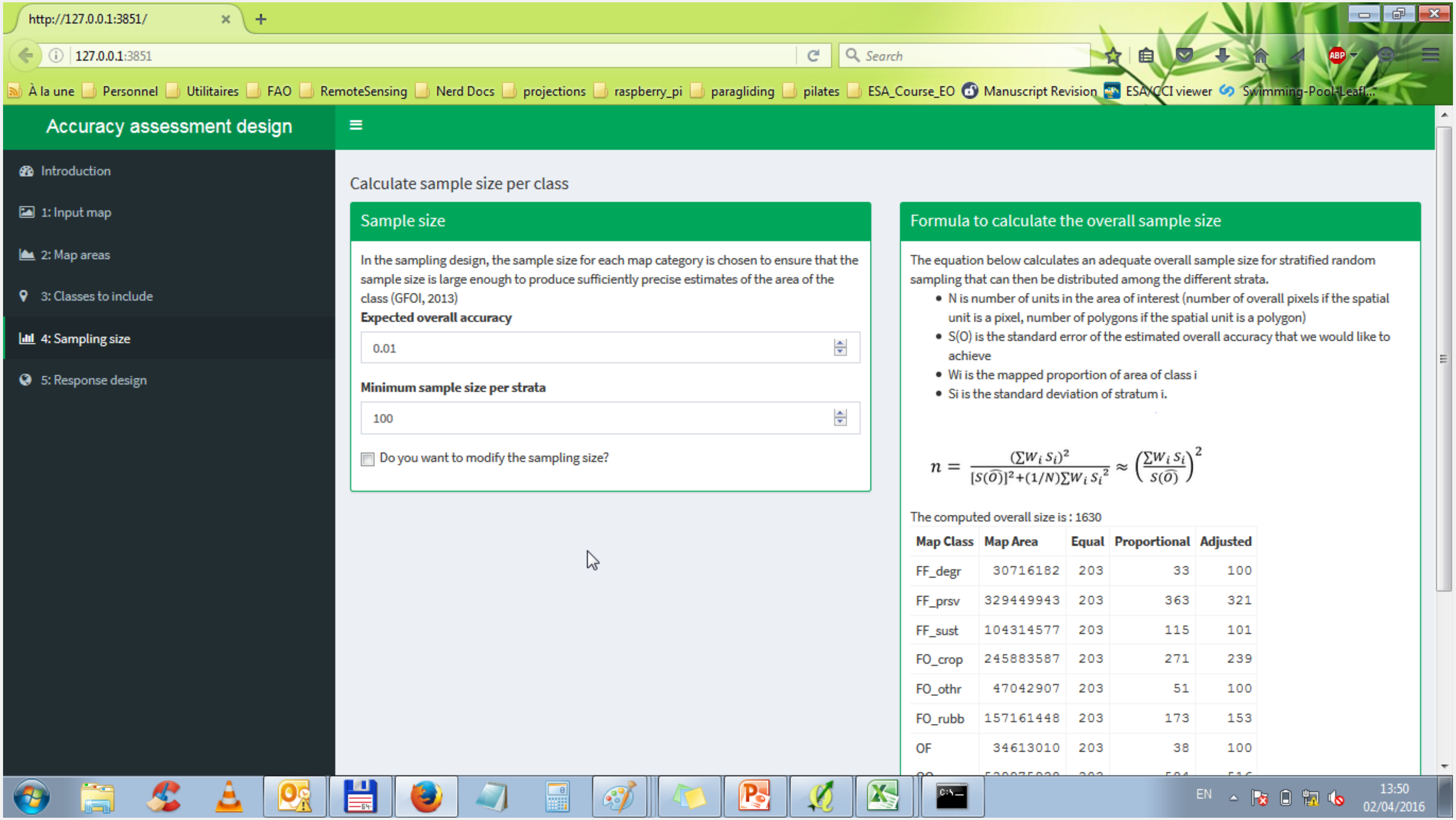
Steps of stratified area estimation

Finalized map data → **Sampling design** → Response design → Analysis

Sampling design



Sampling design: tool in https://sepal.io



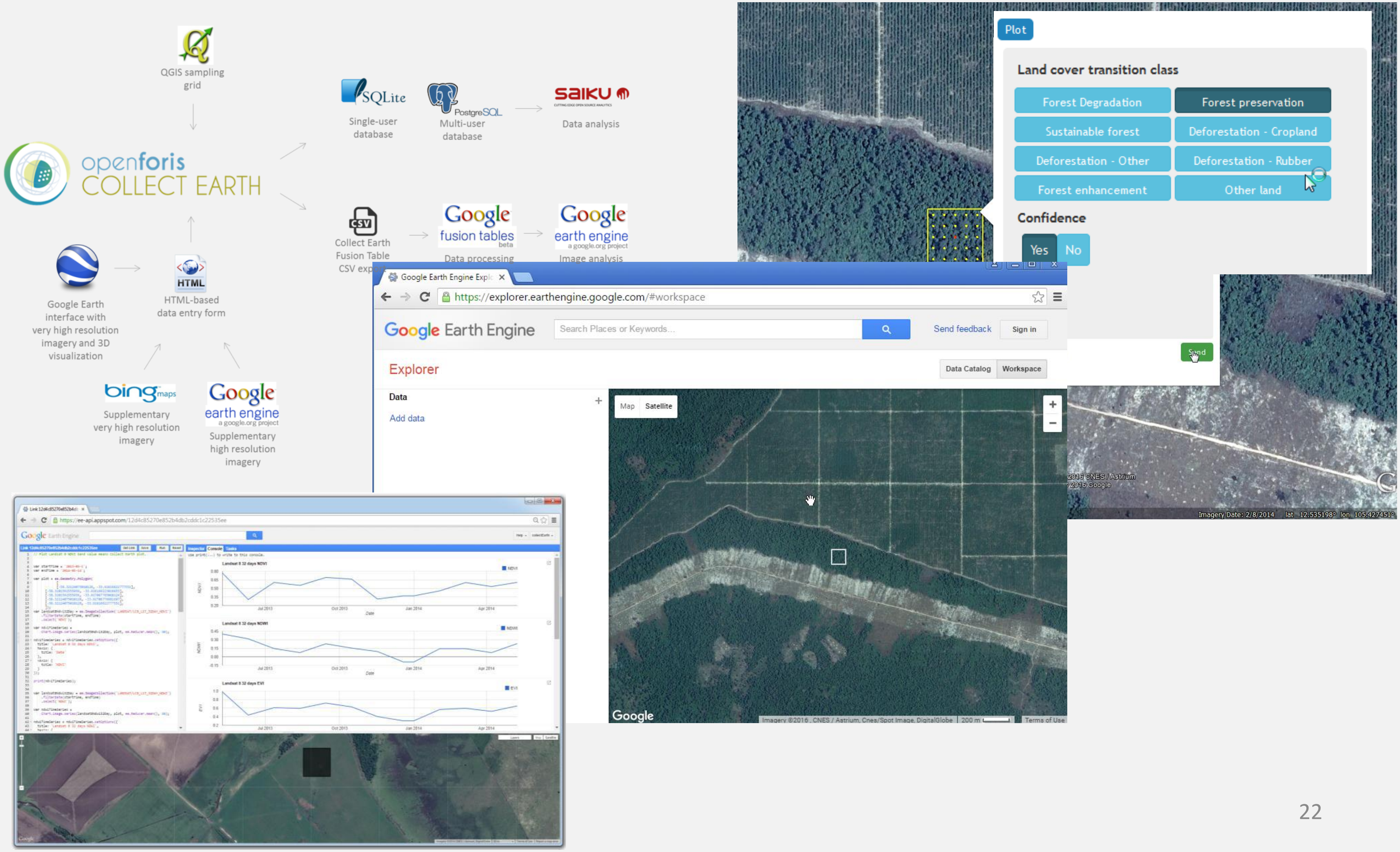
Steps of stratified area estimation

Finalized map data → Sampling design → **Response design** → Analysis

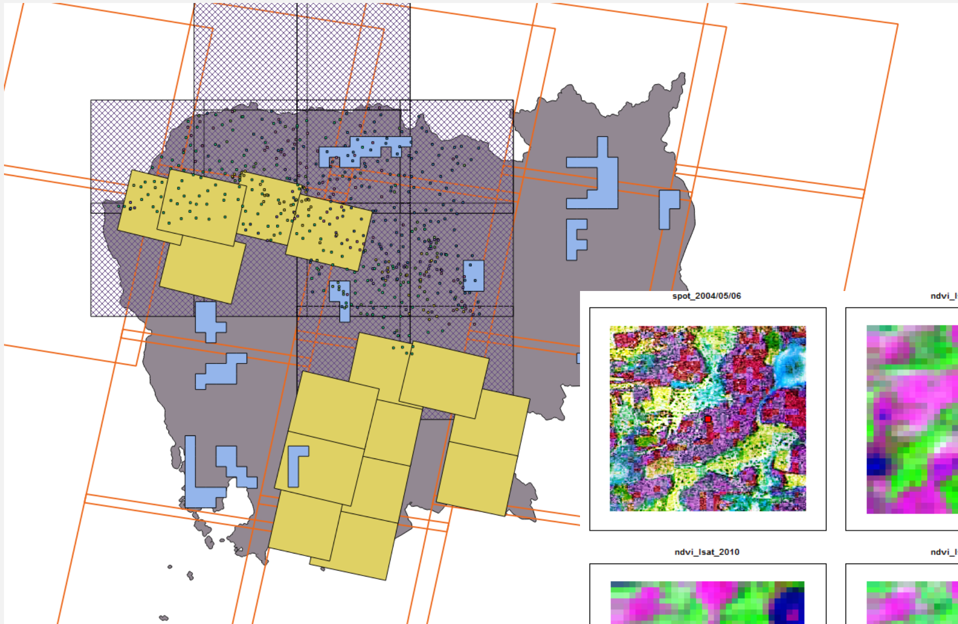
Response design

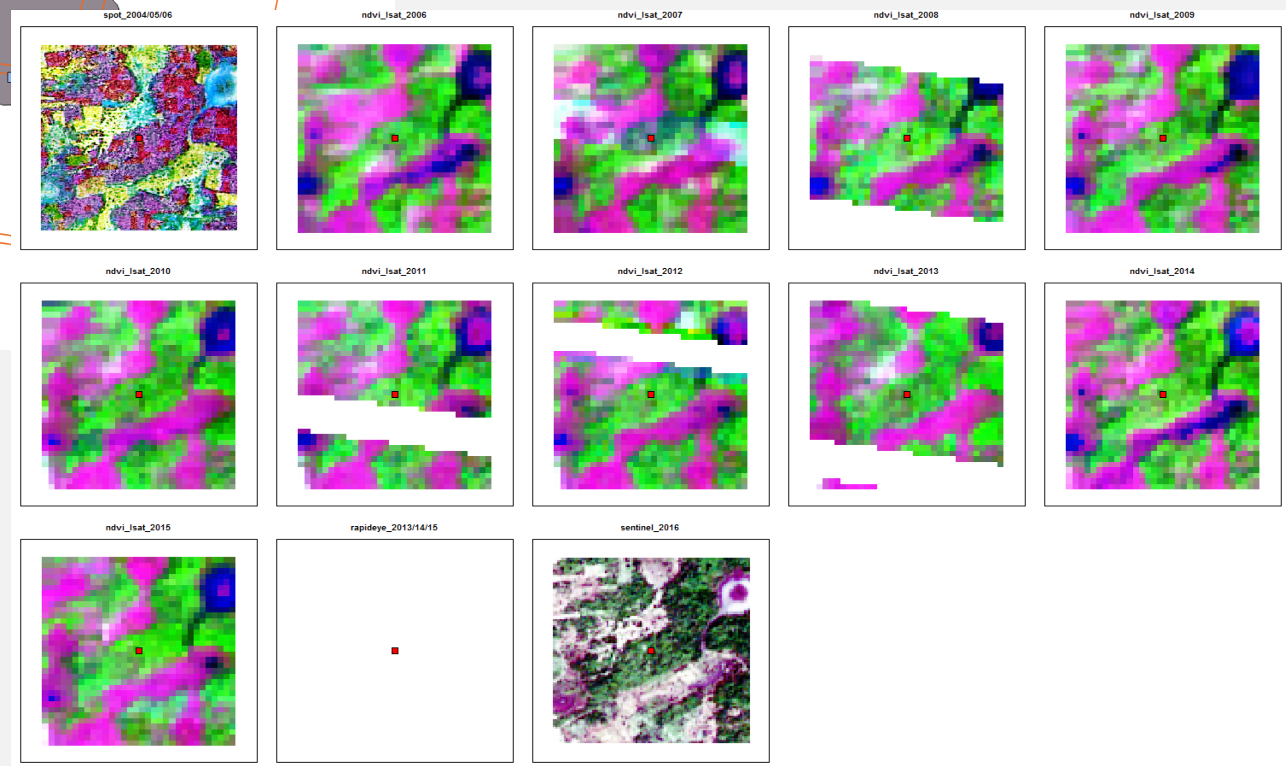
- The reference data MUST be of higher quality
  - Higher quality means higher spatial or radiometric resolution
  - Or the process to create the reference classification is more accurate, i.e. using a human to classify Landsat imagery rather than an algorithm
  - Should be from the same time period
- Spatial assessment unit:
  - **Points**
  - Pixel
  - Polygon
  - Pixel block (aggregated pixels)
- Sources of reference data
  - Google earth
  - National forest inventory

Response design: Collect Earth



Response design: Other data sources





All Landsat from 2006 to 2015
SPOT from 2006 (brown)
RapidEye from 2014 (blue)
Sentinel 2 from 2016 (hashed)

Response design wrap up

- Need to have clear understanding of the map classes
- Need higher quality data
- Reference labeling protocol
  - Establish rules for heterogeneous assessment units

Steps of Accuracy Assessment

Finalized map data → Sampling design → Response design → Analysis

Analysis – Calculate accuracy statistics

- Based on the proportional error matrix
- Entire map
  - Overall accuracy with confidence interval
- Per strata
  - User’s accuracy with confidence interval
  - Producer’s accuracy with confidence interval

Take home messages

- Need to assess accuracy of spatial data
  - Maps are not perfect
  - Extracting statistics from maps
  - Goal: Get adjusted area estimates of land cover classes
- Clear guidance available
  - "Easy" to implement
  - Cost and time efficient

The tools: free and open-source solutions

https://sepal.io

www.r-project.org

Design and Analysis

www.gdal.org

www.shiny.rstudio.com

Response

www.openforis.org

Thank you!

Questions?