10TH ADVANCED TRAINING COURSE ON LAND REMOTE SENSING

esa

Accuracy in Forest Mapping

Dr Christophe Sannier

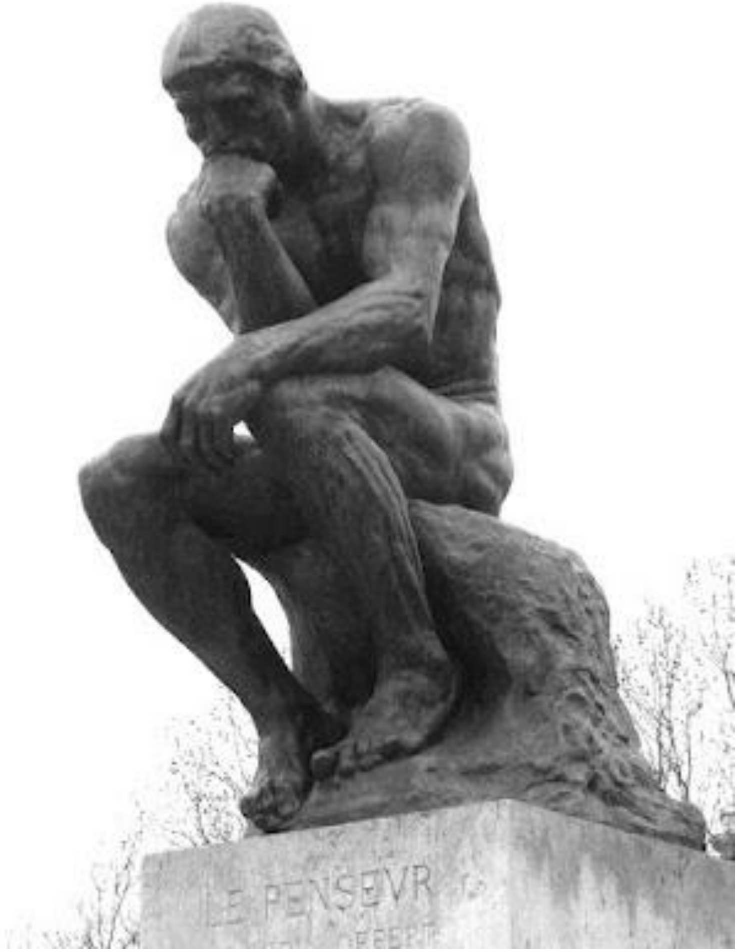
CLS
COLLECTE LOCALISATION SATELLITES

Overview

1. General Concepts
2. Stratification and sample design
3. Response Design
4. Analytical framework
5. Beyond Accuracy assessment
6. Summary

General concepts: Need for Accuracy Assessment

- Ensure that the map produced comply with the specifications.
- Map product finalised.
- Can the map be considered “true” and reliable statistics can be extracted from it.
- Anything else?



Validation / Verification / Accuracy

ISO Definition:

- **Verification**: confirmation of a claim, through the provision of objective evidence, that specified requirements have been fulfilled
- **Validation**: confirmation of a claim, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled
  - Was the product made in the right way? To which extent the products meet the technical specifications set part of validation / verification
  - Was the right product made? Fitness for purpose and limitations (if any) of the products with respect to their intended and potential uses part of validation
- Validation Plan:
  - Validation Framework: criteria for validation, analysis of product specifications and standards
  - Validation approach: methodology and guidance to implement validation

*may cover different concept

Validation Plan: standards and criteria

Examples of criteria:

- Geographic area to be reported: national, river basin, pan-European…
- Indicator threshold: e.g. less than 5 m RMSE
- Accepted uncertainty of accuracy measure: 90% overall accuracy, less than 10% omission and commission errors…

OGC	INSPIRE	FGDC
Positional Accuracy	Positional Accuracy	Positional Accuracy
Temporal Accuracy	Temporal Quality	
Thematic Accuracy	Thematic Accuracy	Attribute Accuracy
Completeness*	Completeness*	Completeness*
		Cloud Cover
Consistency and integrity	Logical Consistency	Logical Consistency
Definition (for semantic interoperability)	Usability	Lineage
Language		
Projection		
Scale		

Accuracy assessment: Product Specifications

Criteria for accuracy assessment, example of Pan-European Copernicus Land Monitoring Service Forest Type Product:

- The overall target thematic accuracy is 90% for both the broadleaf and the coniferous class.
- The 90% accuracy value must be understood as follows: 10% for commission errors and 10% for omission errors for the forest classes.

Forest Type 100m	Acronym FTY	Product category Aggregated status layer
Reference year *2018 (March to October)*		
Geometric resolution *Pixel resolution 100m x 100m, fully conform with the EEA reference grid*		
Coordinate Reference System *European ETRS89 LAEA projection*		
Geometric accuracy (positioning scale) *Less than half a pixel. According to ortho-rectified satellite image base delivered by ESA.*		
Thematic accuracy *Determined by the accuracy of the source Forest Type 2018 in 10m spatial resolution (FTY_2018_010m) and thus indirectly by the Forest Additional Support Layer (FADSL_2018_010m) at 10m spatial resolution and the Tree Cover Density 2018 at 10m spatial resolution.*		
Data type *8bit unsigned raster with LZW compression*		
Minimum Mapping Unit (MMU) *Pixel-based (no MMU)*		
Tree cover density threshold *10%*		
Necessary attributes *Raster value, count, class name, area (in km2), percentage (taking outside area not into account)*		
Raster coding (thematic pixel values) *0: all non-forest areas* *1: broadleaved forest* *2: coniferous forest* *3: mixed zones* *254: unclassifiable (no satellite image available, or clouds, shadows, or snow)* *255: outside area*		
Metadata *XML metadata files according to INSPIRE metadata standards*		
Delivery format *GeoTIFF*		

Validation Approach

- **Completeness**: is the whole area covered? Are there missing part? Percentage covered?
- **Positional Accuracy:** evaluation of the differences between the positions of the objects of the product with respect to reference data e.g. 5m RMSE
- **Thematic Accuracy:** accuracy assessment
- **Temporal quality**: are the source data compliant with the defined reference period
- **Usability**: does the product complies with the user needs?
- **Logical Consistency**:
  - **Conceptual**: Data model compliance, minimal mapping unit / Width
  - **Domain**: compliant attribute values
  - Format: file / attribute names and attribute coding and overall file format compliant
  - **Topological**: compliance with defined rules
  - **Symbology**: compliance with defined colour coding and or symbols
  - **Map projection**: presence and compliance of defined map projection

Accuracy Assessment

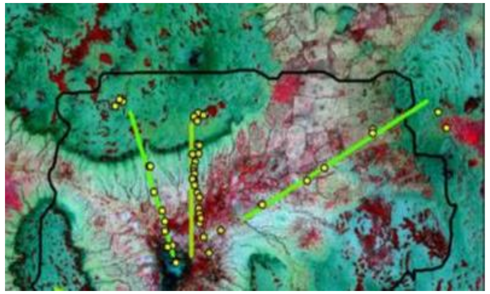
- **The implementation of an accuracy assessment plan consists of three stages:**
    1. **Stratification and sampling design**
    2. **Response design**
    3. **Confusion matrix and accuracy measures**
- **Thematic maps are normally assessed with a confusion or error matrix**
- Continuous data such as biomass or tree cover density maps requires a different analytical framework

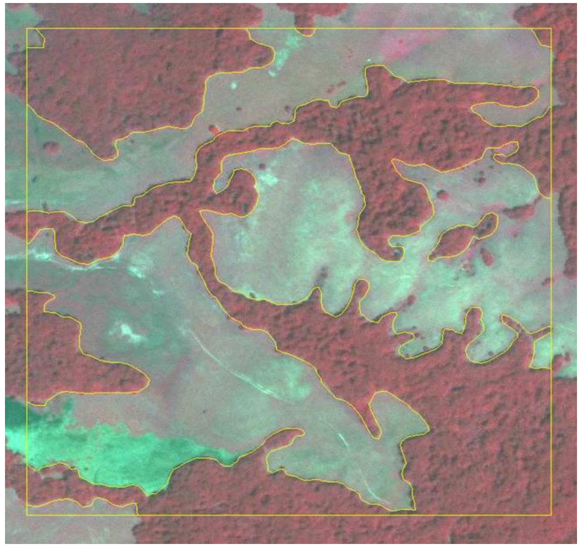
Stratification and Sample Design: Overview

- Probability sample: the probability of sample units to be drawn needs to be known
- A probability sample is obtained via a random or systematic approach
- Samples drawn from area frame (geography) rather than list frame
- Samples can be points, lines or polygons
- Examples of sampling approaches: Simple random sampling, stratified random sampling, stratified systematic random sampling…
- A stratification can be used:
    - To reduce the overall number of samples required thus reducing costs
    - To ensure that all thematic classes (rare classes) are included

Sample Units

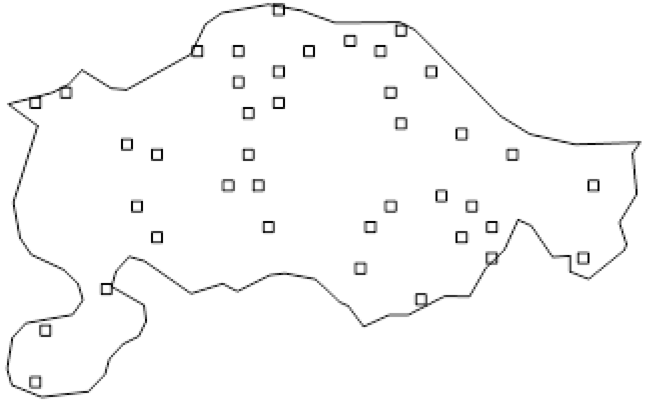
- **Point**: most often used, but need to consider the mapping rules such as minimum mapping area / width
- **Line or transect**: sometimes used in difficult terrain (accessibility)
- **Polygon or segment**: we apply the same input rules as in production but on restricted areas, allowing evaluation of aspects other than thematic precision such as geometry
- **Sampling** can be done in several **stages**: Primary Sample Unit (PSU), Secondary Sample Unit (SSU)



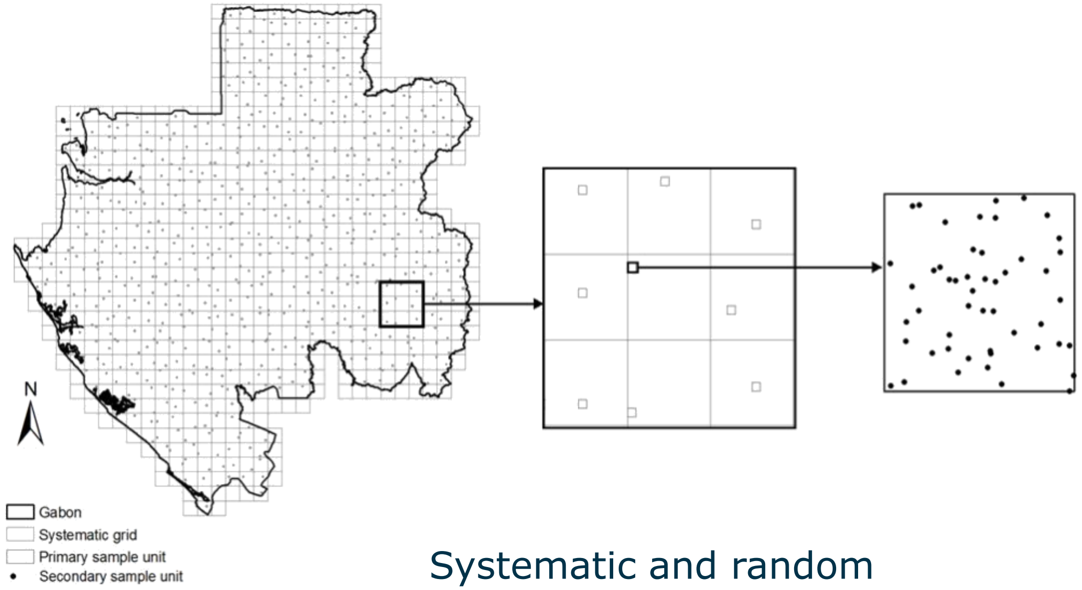


Sampling Approach

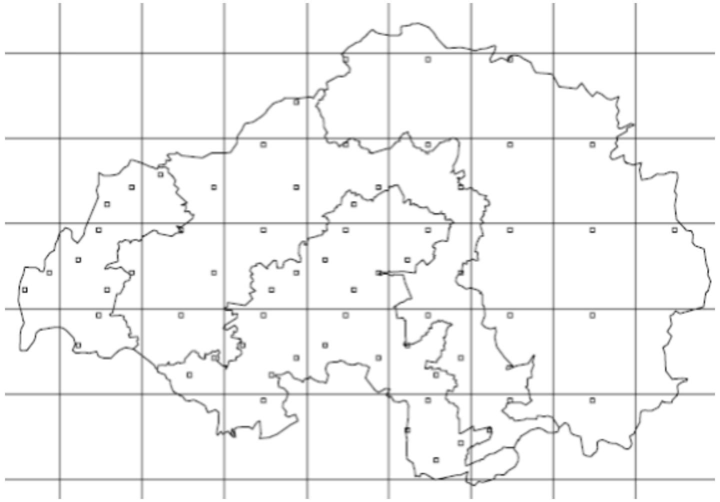
- **Systematic**: can be problematic if regular pattern in landscape
- **Random**: may be inefficient for rare classes
- **Stratified**: address rare classes



Random with minimum distance



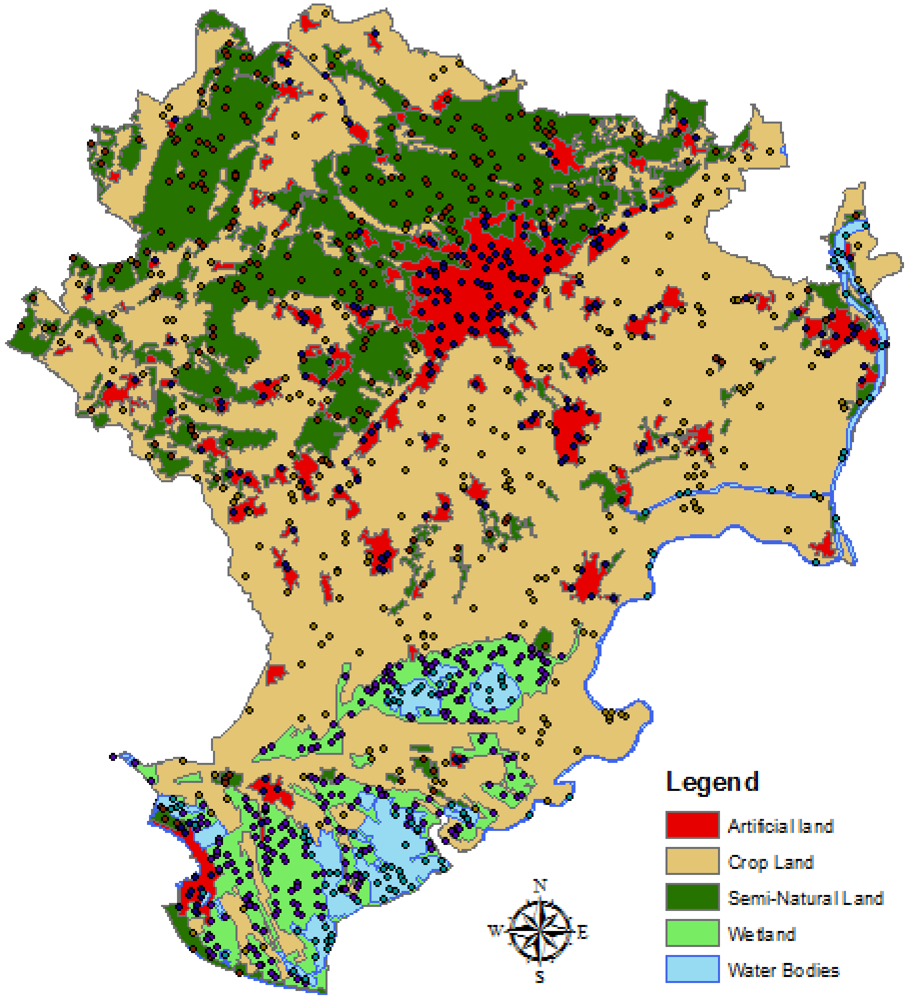
Systematic and random



Stratified Systematic

Stratification

- **Definition**: the partition of a study area into several distinct subsets whose goal is to improve the homogeneity of the data and therefore reduce their variance
- Allows a better representation of all the thematic classes
- Possibility of using the thematic map to be assessed as basis for stratification
- According to Cochran (1977) no need for more than 6-8 strata

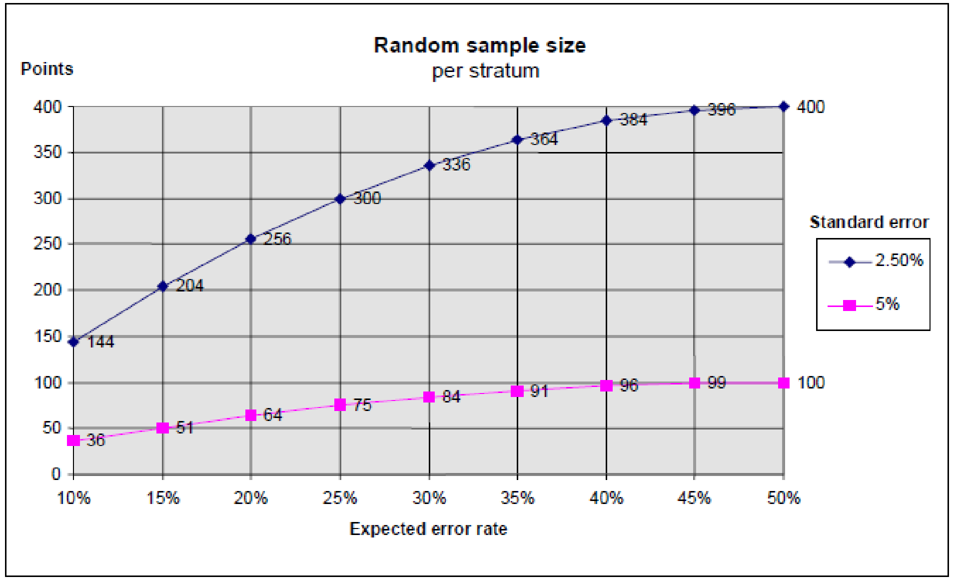


Number of Sample units per stratum

- Ensure sufficient level of precision at reporting level (e.g. administrative unit,…)
- Neyman allocation can be used for mono thematic maps: $n_h = \frac{p_h(1-p_h)}{\sigma_h^2}$

where n_h is the sample size for stratum h, p_h is the expected error rate and σ_h is the desired standard error

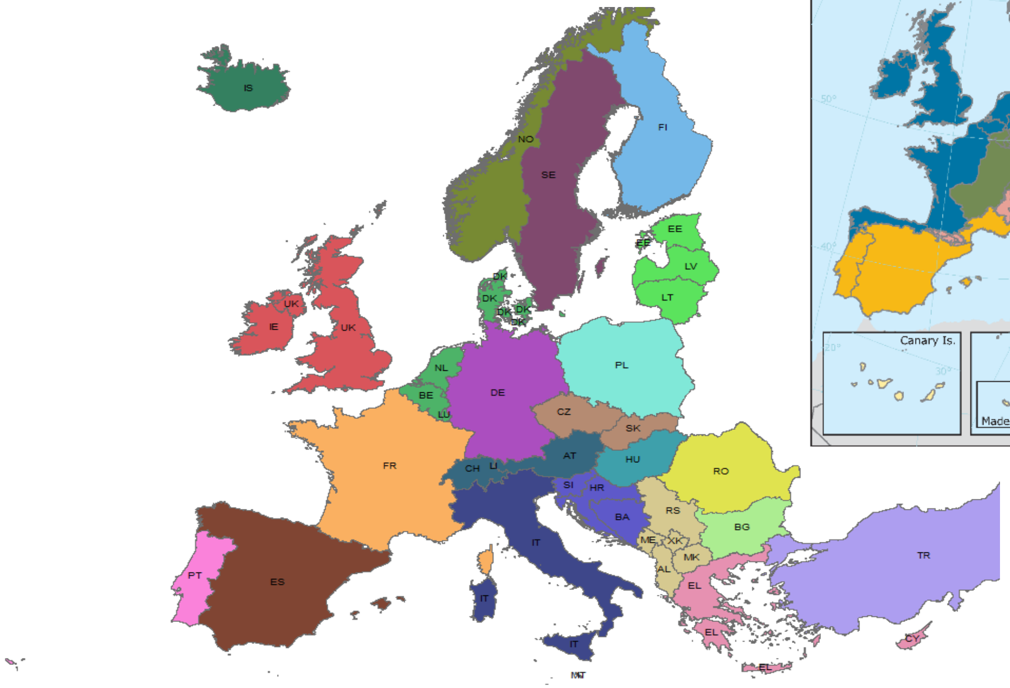
- For multi thematic maps some optimal allocation algorithms exist, but are not often easy to implement
- Equal allocation reduces uncertainty of commission errors
- Proportional allocation reduces uncertainty of omission errors and area estimates
- In practice a combination of proportional allocation and minimum number of samples for small strata (e.g. 50) is used
- Total number of sample units should not exceed 1000 per reporting unit, but this also depends on the number of thematic classes

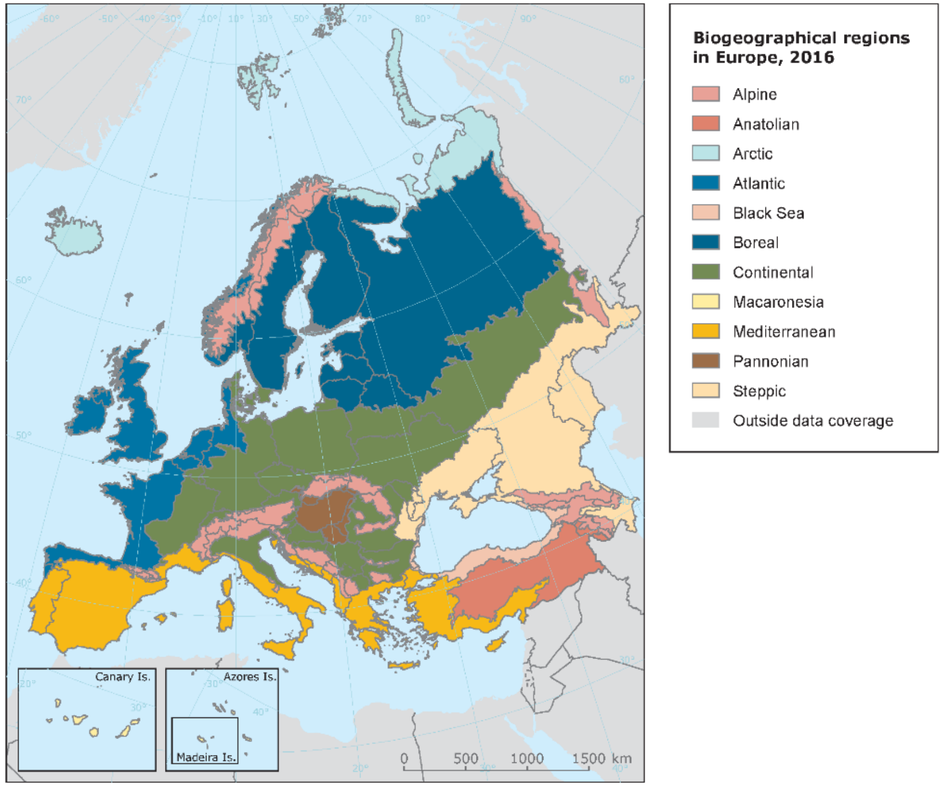


→ THE EUROPEAN SPACE AGENCY

Level of Reporting: Example of Pan-European Mapping

- Copernicus Land Monitoring Service: land.Copernicus.eu
- Need to find a compromise between number of sample units and representative results at sub-European level
- Grouping of countries < 90,000 km² area
- 23 main groups
  (including French DOMs)
- Biogeographical regions
- Production teams





Stratification & Sample design – Example of CLMS High Resolution Layer Tree Cover Density product (TCD)

- Different sampling intensity applied:
  → Focus on strata for which there is a higher probability that errors will be found.
- CLC tree covered classes were defined as follows based on CLC2006-12:
  - 2.2.2 = fruit trees, orchards
  - 2.2.3 = olives
  - 2.4.3 = agriculture with significant amount of natural vegetation
  - 2.4.4 = agroforestry
  - 3.1.1 = broadleaf forest
  - 3.1.2 = coniferous forest
  - 3.1.3 = mixed forest
  - 3.2.4 = transitional woodland, shrub

Stratification & Sample design – TCD

- First unstratified sampling at Pan-European level of 15,000 sample units
- The first level of stratification is defined according to countries or group of countries with an area greater than 90,000 km²
- The second level stratification was defined as follows:

Strata	Description	Number of sample units
Commission Low Probability	TC 1-100% & CLC impervious classes	minimum of 50 PSUs per country / group of countries
Commission High Probability	TC 1-100% & CLC non TC classes	minimum of 50 PSUs per country / group of countries
Omission High Probability	TC 0% & CLC TC classes	minimum of 150 sample units per country / group of countries
Omission Low Probability	Rest of the area	minimum of 50 PSUs per country / group of countries

Sample design

- Stratified systematic sample design based on the EU Land Use and Coverage Area frame Survey (LUCAS) 2 x 2 km sampling frame
- Selection of Primary Sampling Units (PSUs) based on LUCAS and densified LUCAS grid if required
- For HRLs TCD, PSU is the 100m pixel
- Selection of Secondary Sampling units (SSUs) for TCD based on a 5x5 20m grid

18km

9	23	66	44	68	10	48	16	42	76	23	66	44	68	10	48	16	42	76
8	71	12	69	25	51	29	60	37	7	71	12	69	25	51	29	60	37	7
7	20	79	18	72	78	3	31	63	70	20	79	18	72	78	3	31	63	70
6	33	**1**	80	11	59	32	38	9	64	33	1	80	11	59	32	38	9	64
5	28	40	26	49	55	17	53	50	77	28	40	26	49	55	17	53	50	77
4	45	27	41	67	6	65	15	73	5	45	27	41	67	6	65	15	73	5
3	35	39	13	36	62	21	57	24	47	35	39	13	36	62	21	57	24	47
2	8	58	74	46	14	75	2	56	34	8	58	74	46	14	75	2	56	34
1	43	30	4	54	61	19	81	22	52	43	30	4	54	61	19	81	22	52
9	23	66	44	68	10	48	16	42	76	23	66	44	68	10	48	16	42	76
8	71	12	69	25	51	29	60	37	7	71	12	69	25	51	29	60	37	7
7	20	79	18	72	78	3	31	63	70	20	79	18	72	78	3	31	63	70
6	33	**1**	80	11	59	32	38	9	64	33	**1**	80	11	59	32	38	9	64
5	28	40	26	49	55	17	53	50	77	28	40	26	49	55	17	53	50	77
4	45	27	41	67	6	65	15	73	5	45	27	41	67	6	65	15	73	5
3	35	39	13	36	62	21	57	24	47	35	39	13	36	62	21	57	24	47
2	8	58	74	46	14	75	2	56	34	8	58	74	46	14	75	2	56	34
1	43	30	4	54	61	19	81	22	52	43	30	4	54	61	19	81	22	52
Row in block	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*

Column in block →

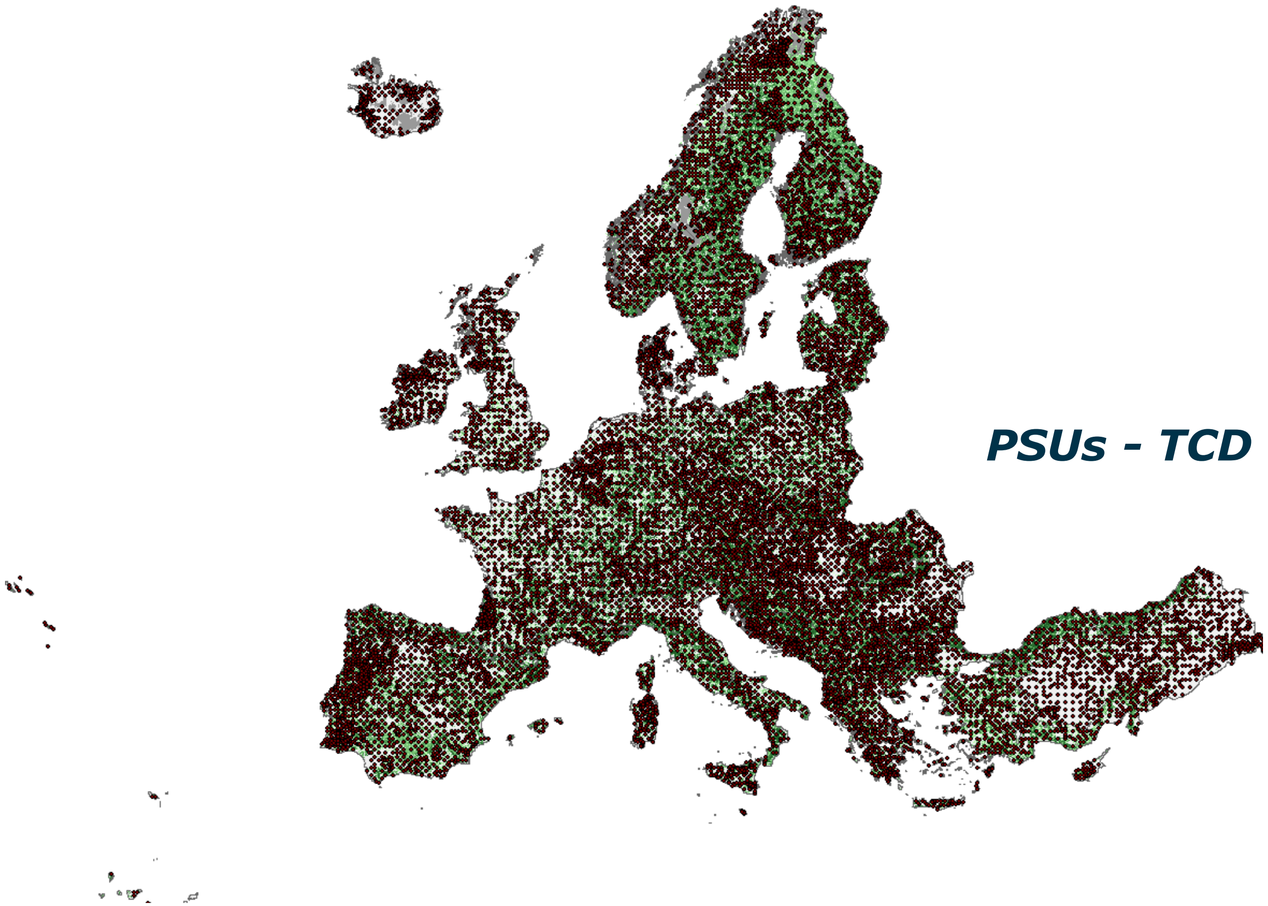
Sample design

- Stratified systematic sample design based on the EU Land Use and Coverage Area frame Survey (LUCAS) 2 x 2 km sampling frame
- Selection of Primary Sampling Units (PSUs) based on LUCAS and densified LUCAS grid if required
- For HRLs TCD, PSU is the 100m pixel
- **Selection of Secondary Sampling units (SSUs) for TCD based on a 5x5 20m grid**

Examples of SSUs organised in a 5x5 20m grid

Number of Sample units - TCD

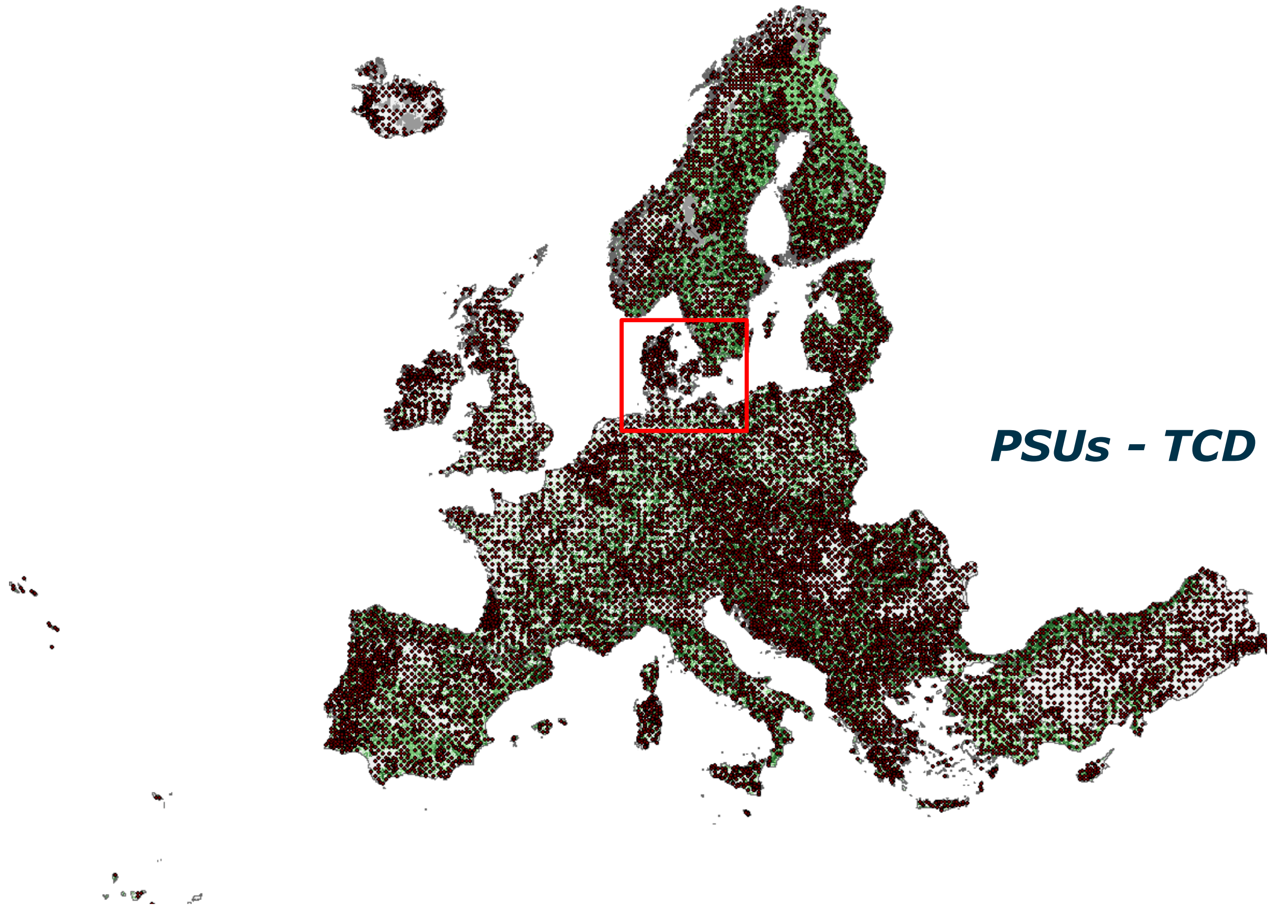
There was a total of 17,296 sample units selected



	Commission		Omission		
LABEL	High	Low	High	Low	Total
AL+ME+MK+RS+XK	82	111	273	111	**577**
AT + CH + LI	74	99	248	99	**520**
BA + HR + SI	74	100	250	100	**524**
BE + LU+ NL + DK	70	95	239	95	**499**
BG	72	94	237	94	**497**
CZ + SK	72	100	250	100	**522**
DE	113	189	426	189	**917**
EE + LT + LV	84	118	286	118	**606**
EL + CY	78	105	260	105	**548**
ES	247	149	543	247	**1186**
FI	132	179	411	166	**888**
FR	149	264	576	266	**1255**
FR DOMs	70	84	219	85	**458**
HU	68	87	223	87	**465**
IE + UK	111	173	395	207	**886**
IS	70	89	230	90	**479**
IT + MT	109	167	384	167	**827**
NO	109	182	402	176	**869**
PL	109	172	393	172	**846**
PT	68	86	222	86	**462**
RO	95	141	341	143	**720**
SE	154	225	498	204	**1081**
TR	202	353	756	353	**1664**
TOTAL	**2412**	**3362**	**8062**	**3460**	**17296**

Number of Sample units - TCD

There was a total of 17,296 sample units selected



	Commission		Omission		
LABEL	High	Low	High	Low	Total
AL+ME+MK+RS+XK	82	111	273	111	**577**
AT + CH + LI	74	99	248	99	**520**
BA + HR + SI	74	100	250	100	**524**
BE + LU+ NL + DK	70	95	239	95	**499**
BG	72	94	237	94	**497**
CZ + SK	72	100	250	100	**522**
DE	113	189	426	189	**917**
EE + LT + LV	84	118	286	118	**606**
EL + CY	78	105	260	105	**548**
ES	247	149	543	247	**1186**
FI	132	179	411	166	**888**
FR	149	264	576	266	**1255**
FR DOMs	70	84	219	85	**458**
HU	68	87	223	87	**465**
IE + UK	111	173	395	207	**886**
IS	70	89	230	90	**479**
IT + MT	109	167	384	167	**827**
NO	109	182	402	176	**869**
PL	109	172	393	172	**846**
PT	68	86	222	86	**462**
RO	95	141	341	143	**720**
SE	154	225	498	204	**1081**
TR	202	353	756	353	**1664**
TOTAL	**2412**	**3362**	**8062**	**3460**	**17296**

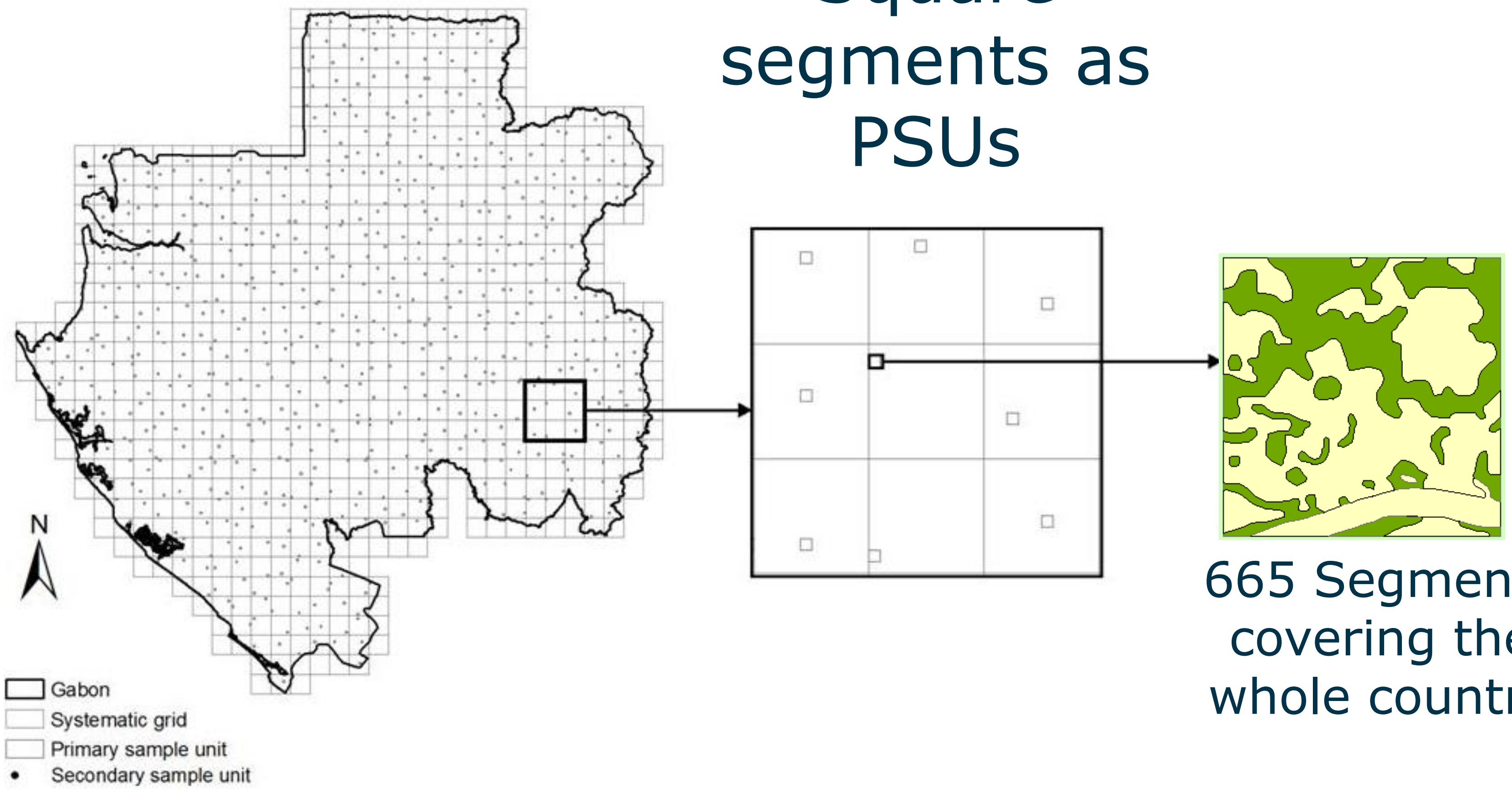
Number of Sample units - TCD

There was a total of 17,296 sample units selected

	Commission		Omission		
LABEL	High	Low	High	Low	Total
AL+ME+MK+RS+XK	82	111	273	111	**577**
AT + CH + LI	74	99	248	99	**520**
BA + HR + SI	74	100	250	100	**524**
BE + LU+ NL + DK	70	95	239	95	**499**
BG	72	94	237	94	**497**
CZ + SK	72	100	250	100	**522**
DE	113	189	426	189	**917**
EE + LT + LV	84	118	286	118	**606**
EL + CY	78	105	260	105	**548**
ES	247	149	543	247	**1186**
FI	132	179	411	166	**888**
FR	149	264	576	266	**1255**
FR DOMs	70	84	219	85	**458**
HU	68	87	223	87	**465**
IE + UK	111	173	395	207	**886**
IS	70	89	230	90	**479**
IT + MT	109	167	384	167	**827**
NO	109	182	402	176	**869**
PL	109	172	393	172	**846**
PT	68	86	222	86	**462**
RO	95	141	341	143	**720**
SE	154	225	498	204	**1081**
TR	202	353	756	353	**1664**
TOTAL	**2412**	**3362**	**8062**	**3460**	**17296**

Example in Gabon: Unaligned Systematic Random sampling

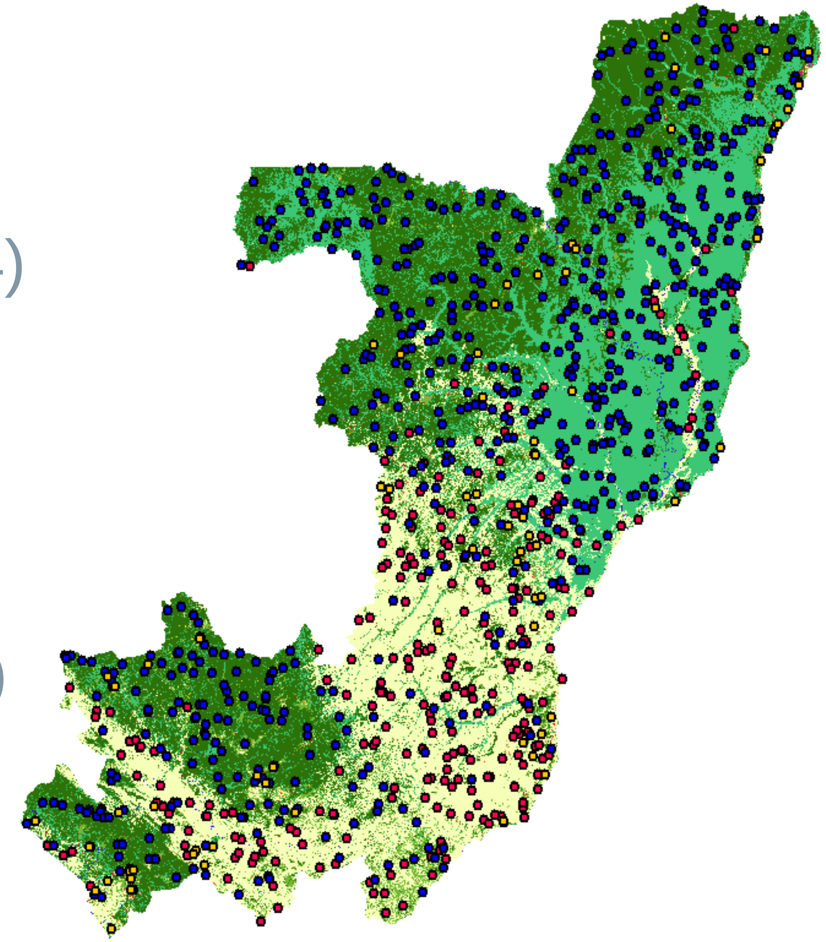
Square segments as PSUs



665 Segments covering the whole country

Stratified Random Sampling Design for the Republic of Congo

- 2000-2012 CNIAFF Forest loss map used for stratification
- Sampling design according to Olofsson *et al.* (2014)
- 1000 pixels sampled
- Minimum of 75 pixel/point samples for smallest stratum: F loss
- Rest of samples allocated proportionally for F (662) and NF (263) strata

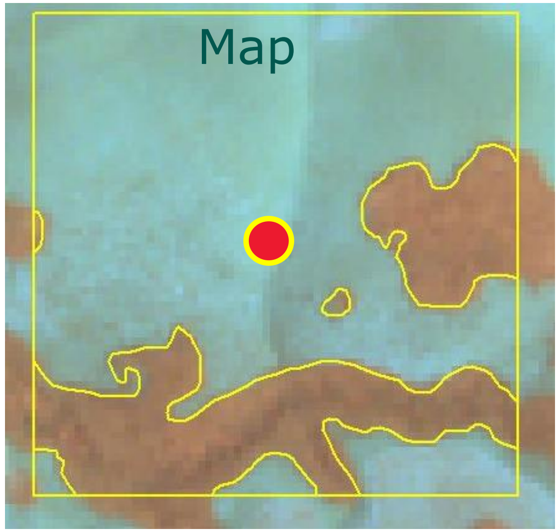


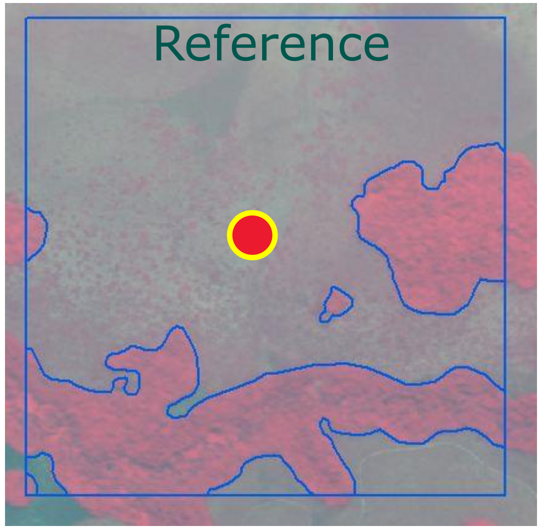
Response design

- How is the “ground truth” or better “reference” data collected
- Field visit or independent interpretation on THR data: the collection of reference data should be based on more accurate protocol and / or data sources
- Need to take into account the rules of interpretation used in production:
  - Definition of classes
  - Concept of minimum mapping area / width
- Double blind interpretation:
  - The production is not aware of the checkpoints
  - The validation team is not aware of the production data
- Plausibility analysis:
  - The production is not aware of the checkpoints
  - Validation based on whether the classification of samples is plausible

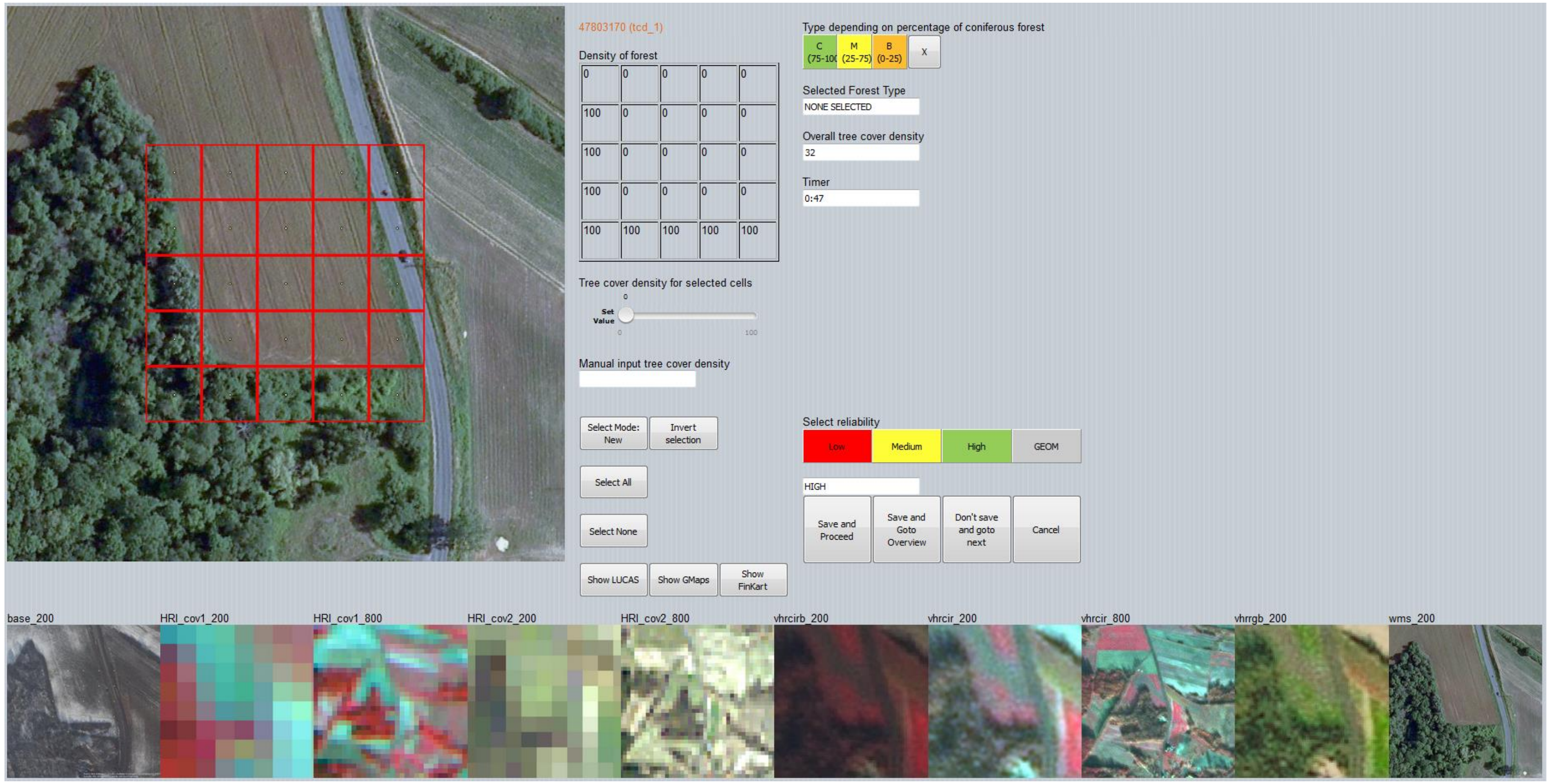
Response design F/NF for Republic of Congo

- Area Samples: F/NF polygons are digitized on screen independently or SSU / sampled pixels individually interpreted following selected forest definition and map product specifications
- Point samples: Point is labelled considering forest definition and mapping rules





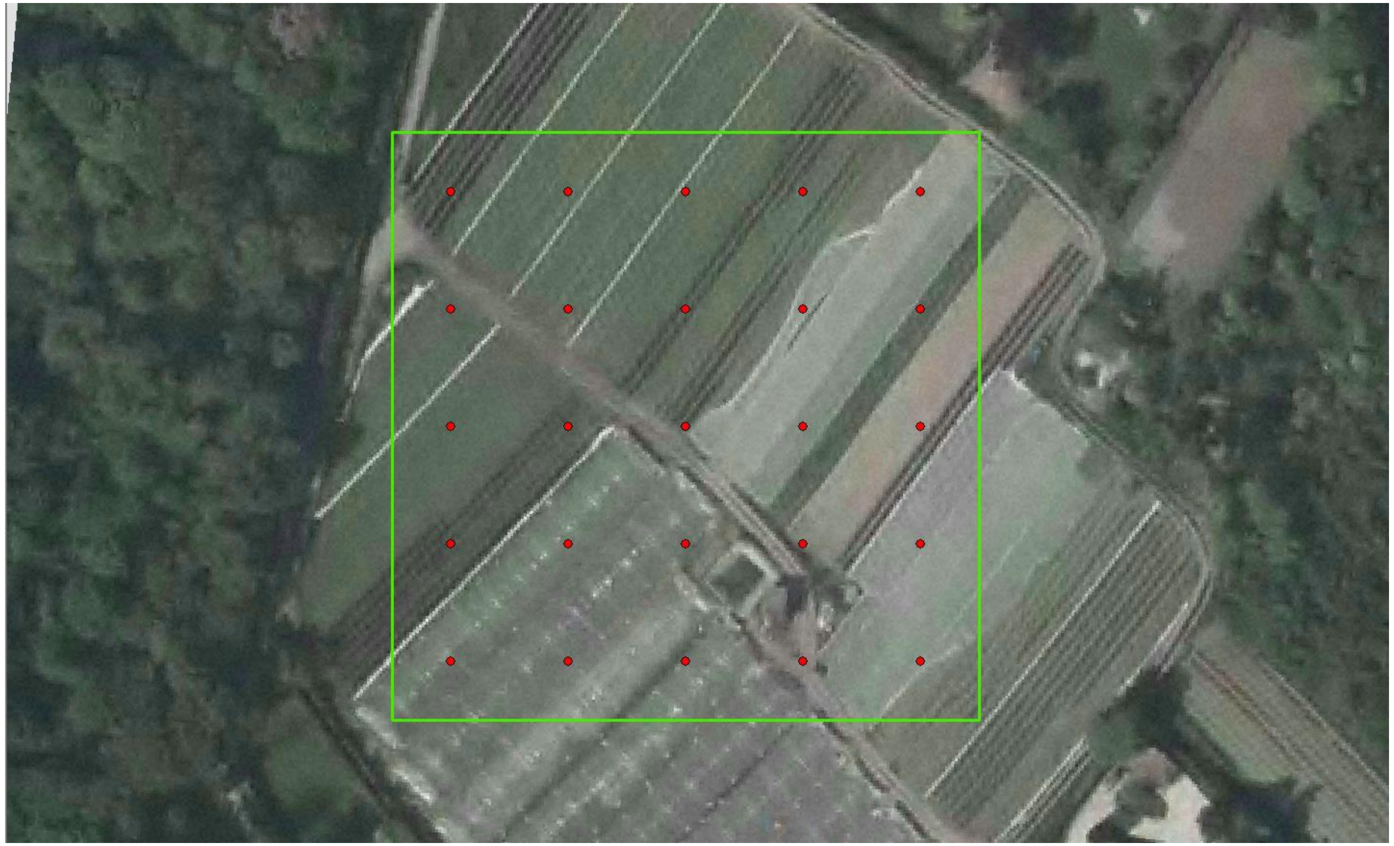
Thematic accuracy - Response design



TCD values collected for each 17,296 PSUs

Interpretation tool for reference data labelling

Response Design: Blind interpretation

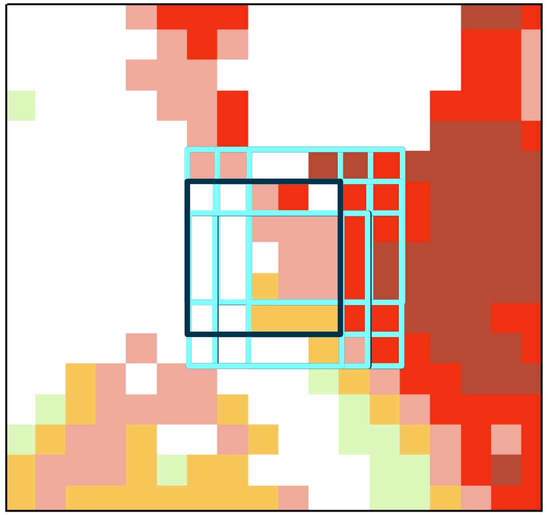


Example: response design & sample units for HRL IMD

Response Design: Blind interpretation

- Vector datasets: the geometric information is used to take into account potentials geometric shifts

- Raster datasets (HR Layers):
  - Apply a “8-Neighbors” shifts and re-interprete the sample units (SSUs) in each spatial position
  - Minimize the difference Prod. Vs Validation regarding the average sealing to take into account potentials geometric shifts



Response Design: Plausibility

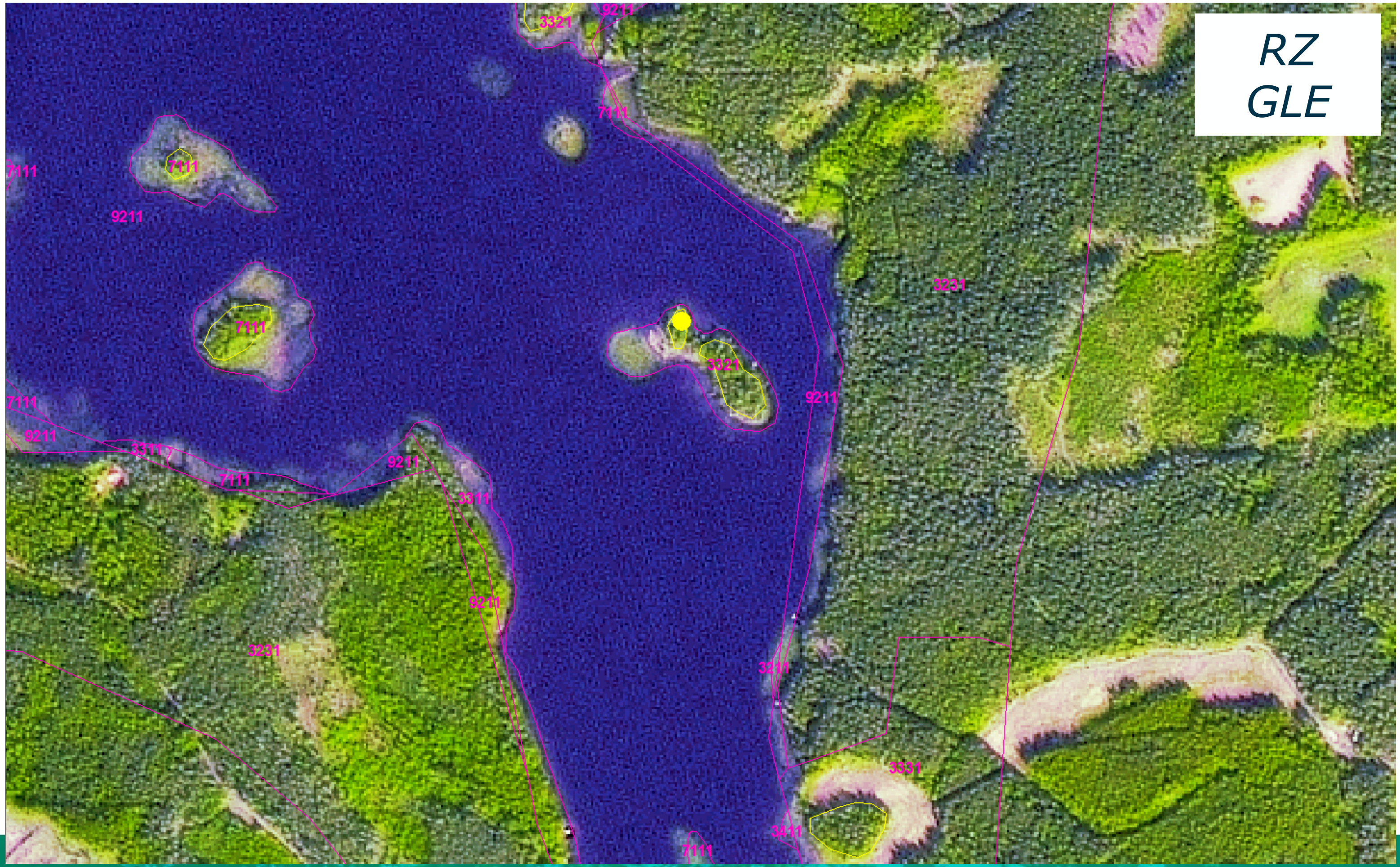
- Sample Units in disagreement are re-interpreted (Product vs. First blind interpretation)
- Interpretation carried out knowing the blind interpretation **AND the map products classification**.

- Code added during the Plausibility analysis:
  - 1 = both codes (product codification and bulk interpretation) are plausible
  - 2 = Product codification is correct
  - 3 = Bulk interpretation is correct
  - 4 = Both codes are wrong

Response Design: Plausibility

❑ Plausibility approach : 1 = both codes (product codification and bulk interpretation) are plausible

- Product: Patch of hedgerows
- Blind: Patch of trees
- Plausibility: Patch of shrub



Response Design: Plausibility

❑ Plausibility approach: 2 = Product codification is correct

- Product: Inland freshwater marshes (7111)
- Bulk/Blind: Mesic grasslands without trees (T.C.D. < 30%) (4222)
- CQ/Plausibility: Inland freshwater marshes (7111)

Response Design: Plausibility

- Plausibility approach: 3 = Blind interpretation is correct

- Product: Other features
- Blind: Patch of trees
- Plausibility: Patch of trees

→ THE EUROPEAN SPACE AGENCY

Response Design: Plausibility

❑ Plausibility approach: 4 = Both codes are wrong

- Product: patch of hedgerows
- Blind: Linear of trees
- Plausibility: Patch of trees

Analytical Framework: Formation and interpretation of confusion matrices

Thematic accuracy usually presented in the form of an error matrix. Cross-tabulate the frequency of occurrence of class combinations observed in a double sample of reference data and the classification:

Reference Data		API			
		1	2	3	4
Pine	1	**35**	4	12	2
Cedar	2	*14*	**11**	*9*	*5*
Oak	3	11	3	**38**	12
Cottonwood	4	1	0	4	**2**
Total		61	18	63	21

Because of the use of a stratified approach, there may be different sampling intensities for each stratum and a **weighting** correction factor for the **should be applied to each PSU within a given stratum**:

$$\hat{p}_{ij} = \left(\frac{1}{N}\right) \sum_{x \in (i,j)} \frac{1}{\pi^*_{uh}},$$

Where i and j are the columns and rows in the matrix, N is the total number of possible units (population) and π_{uh} is the sampling intensity for pixel u and stratum h.

Strata weight - TCD

	Commission		Omission		
LABEL	High	Low	High	Low	Total
AL+ME+MK+RS+XK	82	111	273	111	**577**
AT + CH + LI	74	99	248	99	**520**
BA + HR + SI	74	100	250	100	**524**
BE + LU+ NL + DK	70	95	239	95	**499**
BG	72	94	237	94	**497**
CZ + SK	72	100	250	100	**522**
DE	113	189	426	189	**917**
EE + LT + LV	84	118	286	118	**606**
EL + CY	78	105	260	105	**548**
ES	247	149	543	247	**1186**
FI	132	179	411	166	**888**
FR	149	264	576	266	**1255**
FR DOMs	70	84	219	85	**458**
HU	68	87	223	87	**465**
IE + UK	111	173	395	207	**886**
IS	70	89	230	90	**479**
IT + MT	109	167	384	167	**827**
NO	109	182	402	176	**869**
PL	109	172	393	172	**846**
PT	68	86	222	86	**462**
RO	95	141	341	143	**720**
SE	154	225	498	204	**1081**
TR	202	353	756	353	**1664**
TOTAL	2412	3362	8062	3460	17296

	Commission		Omission	
LABEL	High	Low	High	Low
AL+ME+MK+RS+XK	2.67971821	0.44241337	0.11482057	1.49857630
AT + CH + LI	2.04142316	0.59091710	0.01689773	1.61964669
BA + HR + SI	2.96075326	0.53185830	0.06208468	0.94230039
BE + LU+ NL + DK	0.74270409	0.51976168	0.05836892	2.36274750
BG	1.91698431	0.30546173	0.08449589	1.53966516
CZ + SK	2.30085347	0.42600981	0.05715658	1.59281736
DE	2.92235279	0.99965119	0.03625813	2.80125741
EE + LT + LV	3.09804540	0.72878076	0.04748719	1.36365352
EL + CY	1.95582270	0.64157431	0.10808149	1.65348760
ES	2.24904110	1.76061715	0.15712912	2.45634992
FI	5.42704464	0.59644387	0.03134921	0.74167927
FR	3.17816009	1.03453337	0.07047529	3.20369145
FR DOMs	2.61120243	0.04259741	0.01096788	0.08610773
HU	0.87779958	0.31664479	0.03525263	2.10267544
IE + UK	0.69171685	0.76967362	0.08052551	3.36372122
IS	0.00615944	0.00816788	0.00556458	2.38016521
IT + MT	3.21931227	1.02772151	0.07818544	2.06968553
NO	2.80792950	0.96698479	0.07602096	2.03529765
PL	2.88903204	0.70059505	0.05500502	2.75564626
PT	1.58613087	0.32302879	0.21713021	1.04878658
RO	2.43062670	0.39113601	0.11157068	2.70444111
SE	5.43348764	0.82180652	0.04890186	1.08740746
TR	2.81324789	0.66139620	0.34963660	3.56302852

Example confusion matrix

Reference Data				API	
		1	2	3	4
Pine	1	**35**	4	12	2
Cedar	2	*14*	**11**	*9*	*5*
Oak	3	11	3	**38**	12
Cottonwood	4	1	0	4	**2**
Total		61	18	63	21

- Diagonal elements represent agreement
- Errors of omission are off-diagonal elements which were identified in the reference data but omitted in the classification e.g. 4
- Commission errors are off-diagonal elements which were included in the classification but were other classes in the reference data e.g. 14

Accuracy measures

Reference Data		API				Total	Prod Accuracy %
		1	2	3	4		
Pine	1	**35**	4	12	2	53	66.0
Cedar	2	*14*	**11**	*9*	*5*	39	28.2
Oak	3	11	3	**38**	12	64	59.4
Cottonwood	4	1	0	4	**2**	7	28.6
Total		61	18	63	21	163	
User Accuracy %		57.4	61.1	60.3	9.5	Overall Accuracy % =	
MAI %		61.4	38.6	59.8	14.3		52.8

F-score is a measure of accuracy calculated from the Producer and User accuracies. F1-score is same as MAI: $F_1 = 2.\frac{(PA.UA)}{(PA+UA)}$

Overall Accuracy % = 100.(Sum of diagonal elements/Total observations)

Producer Accuracy % = 100.(Diagonal element/Row total)

User (mapping) Accuracy % = 100.(Diagonal element/Column Total)

Mean Accuracy Indicator (MAI) % = 100.(2.Diagonal /(Row + Column))

Accuracy Measures

- Overall accuracy and User and producer accuracy should be computed for all thematic classes and **95% confidence intervals** should be calculated for each accuracy.
- The standard error of the error rate can be calculated as follows: $\sigma_h = \sqrt{\frac{p_h(1-p_h)}{n_h}}$ where n_h is the sample size for stratum *h* and p_h is the expected error rate.
- The standard error is calculated for each stratum and an overall standard error is calculated based on the following formula: $\sigma = \sqrt{\sum w_h^2 . \sigma_h^2}$ In which w_h is the proportion of the total area covered by each stratum.
- The 95% confidence interval is +/- 1.96*σ.

Aggregated results example: CLMS HRL Forest Types 2015

	Plausibility Analysis												
	Overall	Prod B	CI 95%	Prod C	CI 95%	Prod M	CI 95%0	User B	CI 95%	User C	CI 95%	User M	CI 95%
EEA39	89.7%	85,8%	0,4%	82,6%	0,4%	65,1%	0,6%	75,1%	0,5%	88,4%	0,4%	71,9%	0,5%
Alpine	87.2%	85,5%	1,3%	83,7%	1,4%	65,6%	1,8%	73,9%	1,6%	87,0%	1,3%	70,5%	1,7%
Anatolian	94.9%	79,5%	1,1%	75,0%	1,2%	100,0%	0,0%	49,2%	1,3%	71,4%	1,2%	22,2%	1,1%
Arctic	99.8%	32,0%	0,1%	70,0%	0,1%	64,1%	1,3%	100,0%	0,0%	36,8%	0,1%		
Atlantic	92.3%	78,7%	1,2%	74,8%	1,2%	37,0%	4,7%	80,1%	1,1%	76,1%	1,1%	64,1%	1,3%
Black Sea	81.6%	86,5%	3,3%	90,0%	2,9%	70,3%	1,7%	67,4%	4,6%	78,3%	4,0%	37,0%	4,7%
Boreal	86.1%	89,5%	1,1%	84,9%	1,3%	56,8%	1,2%	63,5%	1,8%	92,8%	0,9%	70,3%	1,7%
Continental	91.6%	88,9%	0,8%	83,5%	0,9%			81,5%	0,9%	90,5%	0,7%	56,8%	1,2%
Macaronesia	100.0%	100,0%	0,0%	100,0%	0,0%	70,0%	1,3%	100,0%	0,0%	100,0%	0,0%		
Mediterranean	86.7%	83,9%	1,0%	78,2%	1,1%	80,0%	2,2%	73,5%	1,2%	84,5%	1,0%	70,0%	1,3%
Pannonian	97.1%	91,2%	1,6%	66,7%	2,6%			91,2%	1,6%	100,0%	0,0%	80,0%	2,2%
Steppic	97.9%	80,0%	2,7%			0,0%	0,0%	80,0%	2,7%	0,0%	0,0%		
AL+ME+MK+RS+XK	86.9%	81,2%	2,9%	84,6%	2,7%	62,5%	3,6%	85,2%	2,7%	78,6%	3,1%	55,6%	3,7%
AT + CH + LI	90.1%	83,9%	3,0%	93,1%	2,1%	75,0%	3,6%	66,7%	3,9%	93,1%	2,1%	75,0%	3,6%
BA + HR + SI	89.8%	92,9%	2,4%	87,5%	3,1%	71,4%	4,3%	86,8%	3,2%	80,8%	3,8%	80,0%	3,8%
BE + LU+ NL + DK	95.1%	93,1%	2,1%	88,9%	2,3%	33,3%	3,5%	81,8%	2,9%	80,0%	3,0%	80,0%	3,0%
BG	94.7%	94,1%	2,1%	95,5%	1,9%	66,7%	4,2%	90,9%	2,6%	95,5%	1,9%	100,0%	0,0%
CZ + SK	87.9%	91,2%	2,4%	69,4%	4,0%	50,0%	4,2%	76,5%	3,6%	94,4%	1,9%	45,8%	4,2%
DE	93.2%	83,7%	1,7%	86,8%	1,6%	69,4%	2,2%	83,1%	1,8%	92,6%	1,2%	81,1%	1,8%
EE + LT + LV	88.4%	89,0%	2,3%	72,6%	3,3%	69,2%	3,4%	80,2%	2,9%	89,8%	2,2%	60,0%	3,6%
EL	89.5%	83,8%	3,0%	77,4%	3,4%	100,0%	0,0%	86,5%	2,8%	82,8%	3,1%	55,6%	4,0%
ES	88.7%	82,1%	1,6%	73,9%	1,8%	83,0%	1,5%	78,7%	1,7%	82,7%	1,5%	58,2%	2,0%
FI	82.7%	83,8%	2,1%	83,4%	2,3%	63,0%	2,9%	60,8%	3,0%	91,5%	1,7%	75,9%	2,6%
FR	90.1%	88,7%	1,2%	73,9%	1,7%	47,7%	1,9%	77,7%	1,6%	78,5%	1,6%	66,7%	1,8%
HU	96.6%	92,3%	2,1%	100,0%	0,0%	80,0%	3,1%	87,8%	2,5%	100,0%	0,0%	100,0%	0,0%
IE + UK	95.9%	65,1%	1,9%	80,0%	1,4%	93,3%	0,9%	77,8%	1,5%	100,0%	0,0%	63,6%	1,7%
IS	99.8%	32,0%	0,1%	70,0%	0,1%			100,0%	0,0%	36,8%	0,1%	0,0%	0,0%
IT	86.7%	91,3%	1,6%	81,6%	2,2%	47,1%	2,8%	70,5%	2,5%	83,8%	2,1%	69,6%	2,6%
NO	86.4%	69,7%	2,4%	73,6%	2,2%	66,2%	2,3%	66,1%	2,3%	77,9%	2,0%	73,6%	2,2%
PL	90.5%	87,7%	1,6%	86,2%	1,7%	52,9%	2,5%	62,3%	2,4%	90,1%	1,5%	78,3%	2,0%
PT	84.1%	77,4%	4,2%	62,5%	5,1%	72,7%	4,5%	77,4%	4,2%	71,4%	4,6%	72,7%	4,5%
RO	94.8%	92,8%	1,4%	81,8%	2,2%	68,0%	2,6%	90,2%	1,6%	94,7%	1,2%	89,5%	1,7%
SE	87.9%	92,3%	1,3%	88,8%	1,6%	75,7%	2,2%	57,1%	2,5%	92,7%	1,3%	83,9%	1,9%
TR	89.2%	81,0%	1,1%	83,1%	1,1%	49,1%	1,4%	58,7%	1,4%	82,7%	1,0%	43,3%	1,4%

Effect of biased sample on accuracy estimates

- In unbiased confusion matrices row totals, expressed as proportions of the grand total, are also estimates of the proportion of the classes in the study region
- Column totals expressed as proportions of the grand total, are also estimates of the proportion of the classes in the map themes.
- In a biased confusion matrix this is not true and a confusion matrix becomes biased when it is not based on a random sample.
- This can occur when road transects are used to collect ground data thus making the assumption that the classes exist in the same proportions along roadsides as they do away from roads.
- Such assumptions are frequently untrue and lead to over-sampling of some classes with respect to others
- The problem is that the errors are invisible but can have profound consequences when interpreting the results of a classification

Effect of bias in confusion matrices

Reference Data		API				Total	Prod Accuracy %
		1	2	3	4		
Pine	1	**35**	4	12	2	53	66.0
Cedar	2	*28*	**22**	*18*	*10*	78	28.2
Oak	3	11	3	**38**	12	64	59.4
Cottonwood	4	1	0	4	**2**	7	28.6
Total		75	29	72	26	202	
User Accuracy %		46.7	75.9	52.8	7.7	Overall Accuracy % =	
MAI %		54.7	41.1	55.9	12.1		48.0

- The confusion matrix above is the same as the previous one except that bias has been introduced by oversampling the Cedar class by a factor of 2
- We expect that omission errors will occur in the same proportion and have simulated the effect by multiplying the row through by 2

Effect of bias in confusion matrices

Reference Data		API				Total	Prod Accuracy %
		1	2	3	4		
Pine	1	**35**	4	12	2	53	66.0
Cedar	2	*28*	**22**	*18*	*10*	78	28.2
Oak	3	11	3	**38**	12	64	59.4
Cottonwood	4	1	0	4	**2**	7	28.6
Total		75	29	72	26	202	
User Accuracy %		46.7	75.9	52.8	7.7	Overall Accuracy % =	
MAI %		54.7	41.1	55.9	12.1		48.0

Reference Data		API				Total	Prod Accuracy %
		1	2	3	4		
Pine	1	**35**	4	12	2	53	66.0
Cedar	2	*14*	**11**	*9*	*5*	39	28.2
Oak	3	11	3	**38**	12	64	59.4
Cottonwood	4	1	0	4	**2**	7	28.6
Total		61	18	63	21	163	
User Accuracy %		57.4	61.1	60.3	9.5	Overall Accuracy % =	
MAI %		61.4	38.6	59.8	14.3		52.8

- Producer accuracies are not changed but the amount of commission error introduced by Cedar occurrences within other classes is increased thus reducing their apparent User accuracies
- The apparent increase in the correctly identified Cedar leads to an increase in the estimate of its User accuracy
- Therefore, classes which are over sampled in a biased scheme will have their User accuracies overestimated and vice versa
- Since User accuracy by definition effects the client's use of the data, it is the client which is mis-led by the error

Accuracy Assessment

- The implementation of an accuracy assessment plan consists of three stages:
    1. Stratification and sampling design
    2. Response design
    3. Confusion matrix and accuracy measures
- Thematic maps are normally assessed with a confusion or error matrix
- **Continuous data such as biomass or tree cover density maps requires a different analytical framework**

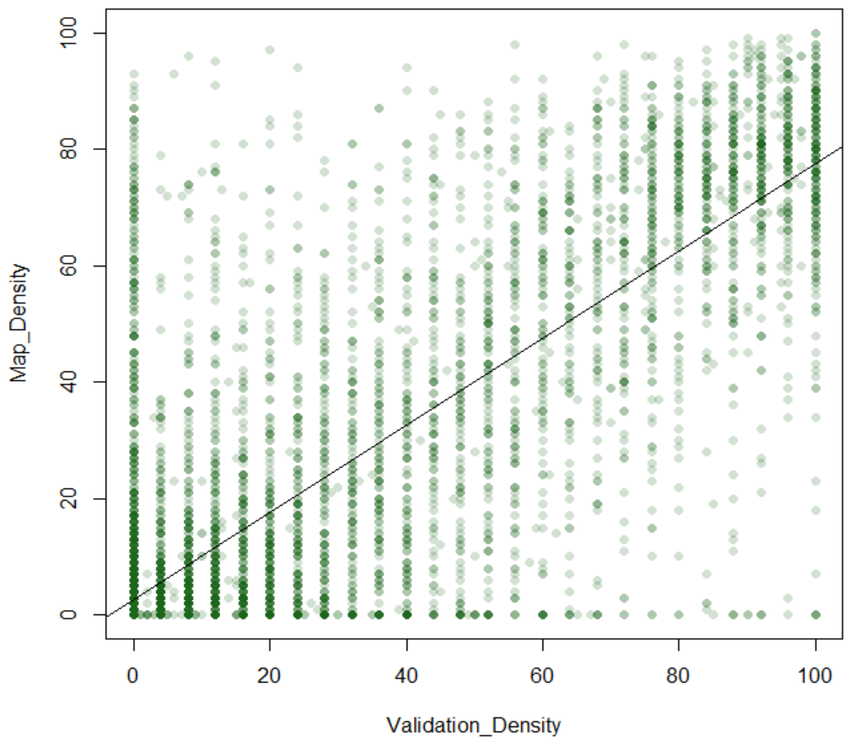
Continuous data – TCD

- Target thematic accuracy:
  - **Scatterplot**
  - 10-15-30% threshold to create binary mask with less than **10-15%** omission and commission error
- Scatterplots should not be constructed with all sample units (different sampling intensity)

  → A subsample was constructed based on equal sampling intensity: a total of 8,001 observations (out of 17,296)

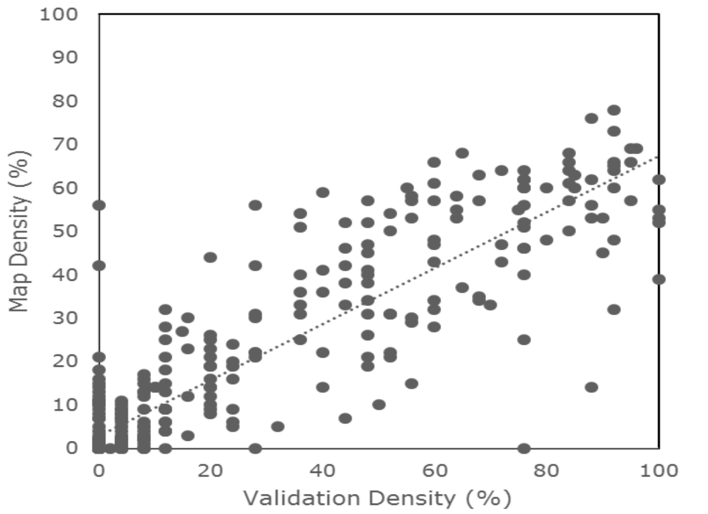
Examples of density scatterplots for TCD

Pan-European



Scatterplot of first replicate sample units for HRL TCD, darker green indicates a greater number of observations, n=8,001, R^2=0.67, Map_Density = 0.749*Validation_Density+2.542

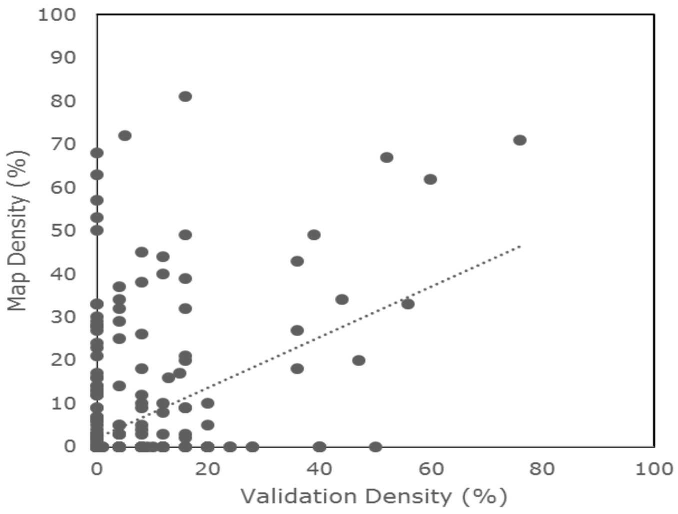
Finland



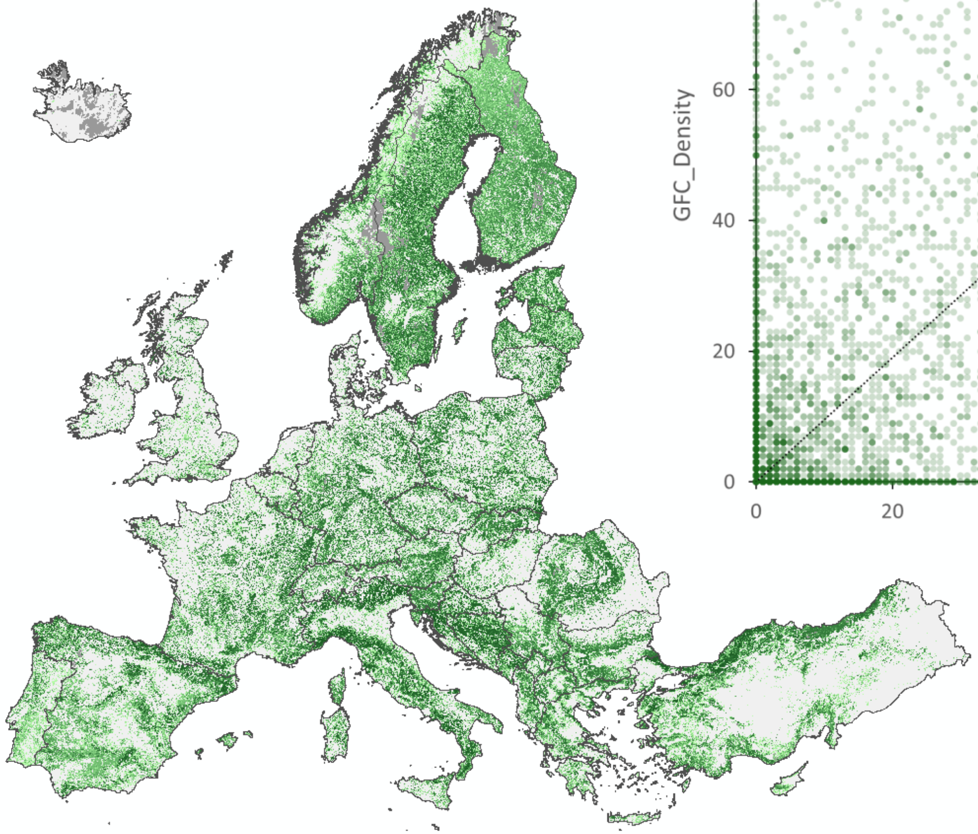
Example of high forest cover: Scatterplot of all sample units for FI HRL TCD, n=385, R^2=0.80, Map_Density = 0.645*Validation_Density+2.81

Anatolian Bio-Region

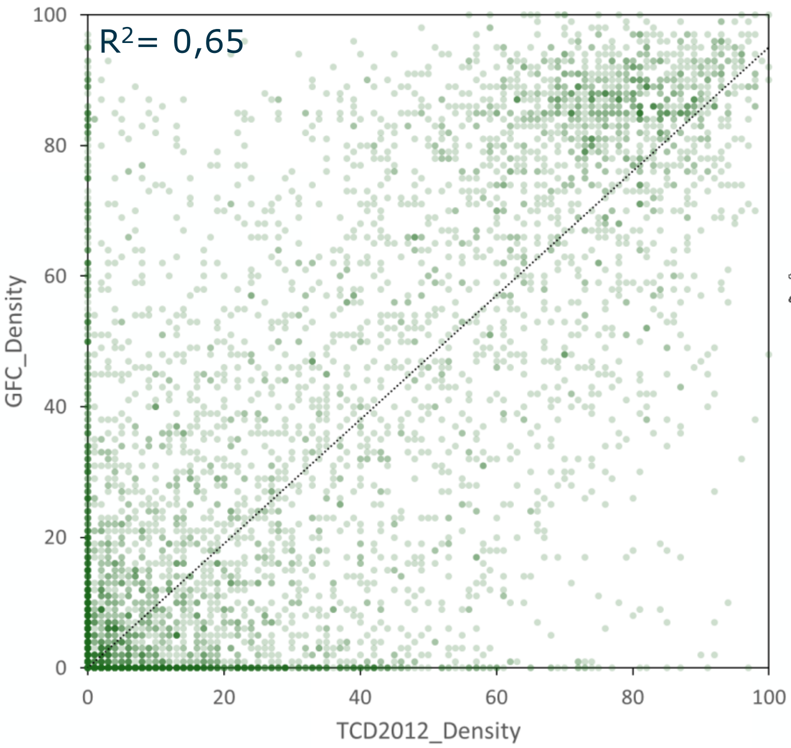
Example of low forest cover: Scatterplot of all sample units for the Anatolian bio-geographical region HRL TCD, n=554, R^2=0.19, Map_Density = 0.585*Validation_Density+1.960

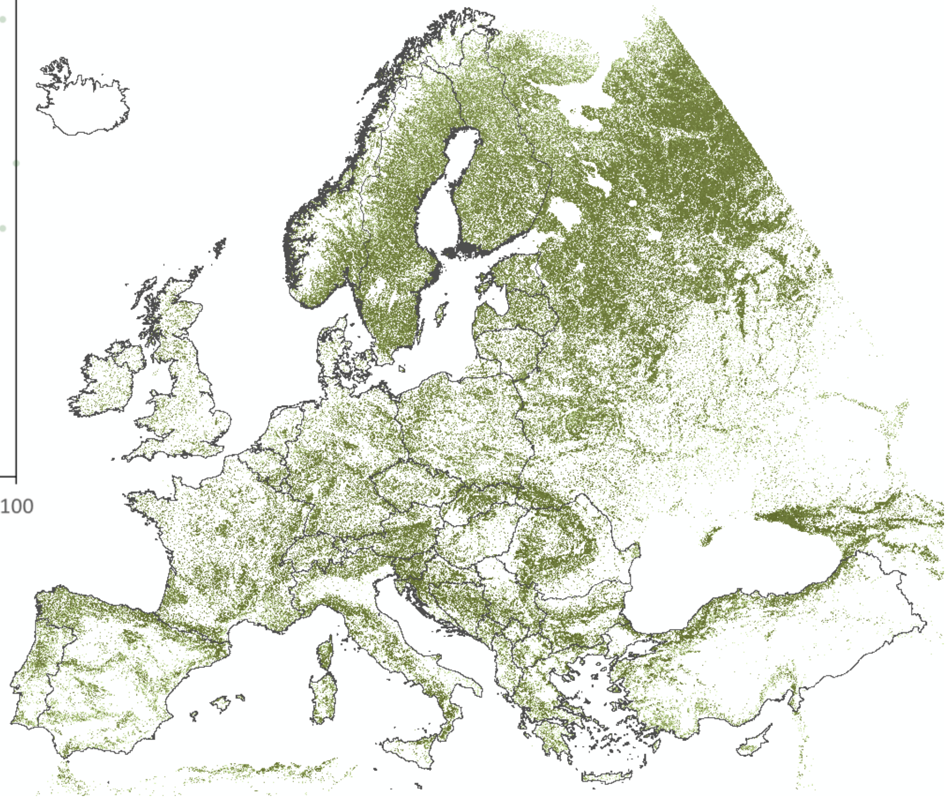


TCD/GFC Comparison



HRL TCD 2012



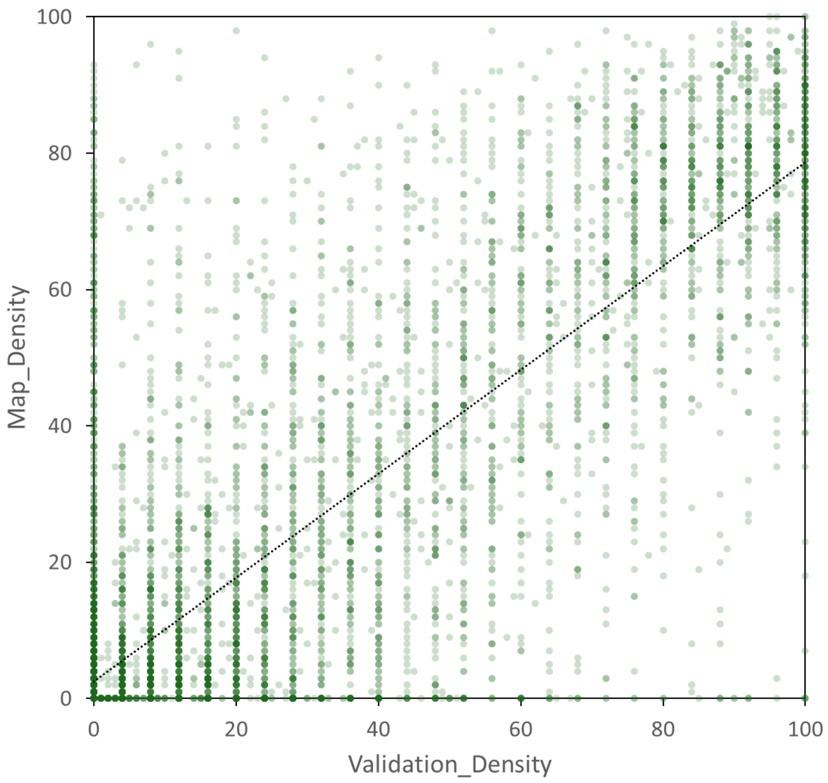


UMD GFC

TCD/GFC Comparison

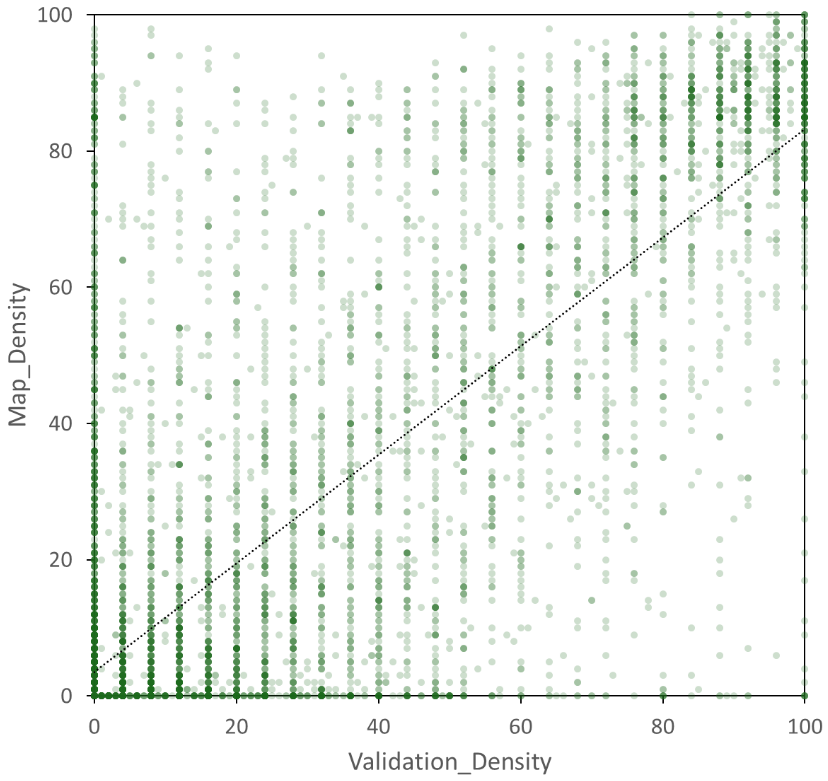
Map_Density = 0.798*Validation_Density+3.502

R^2=0.62



UMD GFC

Map_Density = 0.762*Validation_Density+2.464 R^2=0.68



HRL TCD

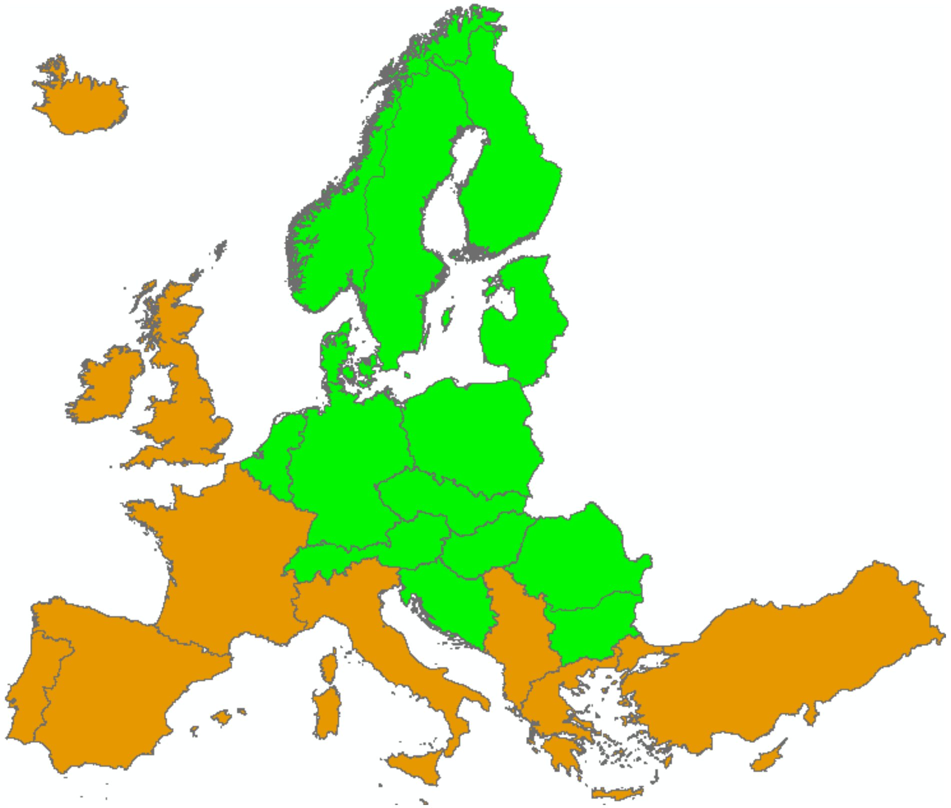
TCD/GFC Preliminary Validation results

with Gain & Loss for GFC		**Blind Validation TCD**				Blind Validation GFC			
		Prod TC	95% CI	User TC	95% CI	Prod TC	95% CI	User TC	95% CI
Pan-European	TC>10%	81.6%	1.9	89.0%	0.5	75.3%	1.85	87.0%	2.66
	TC>15%	82.0%	2.1	88.0%	0.6	76.5%	2.11	86.8%	1.42
	TC>30%	84.2%	2.2	85.9%	0.7	81.0%	2.22	85.1%	1.69

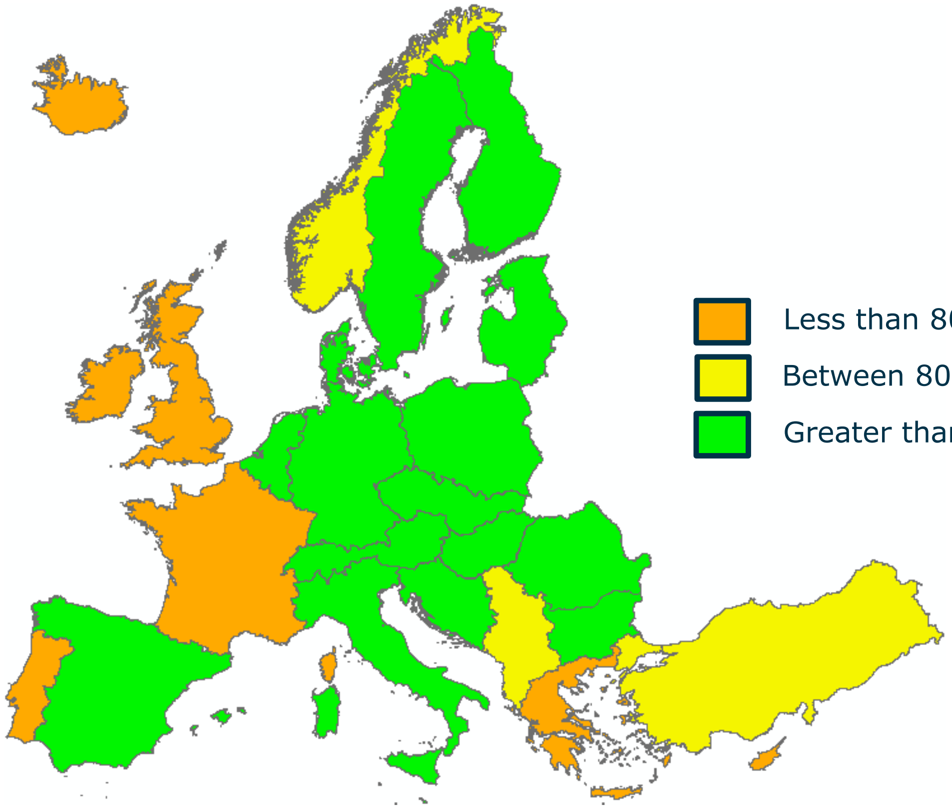
without Gain & Loss for GFC		**Blind Validation TCD**				Blind Validation GFC			
		Prod TC	95% CI	User TC	95% CI	Prod TC	95% CI	User TC	95% CI
Pan-European	TC>10%	81.6%	1.9	89.0%	0.5	77.9%	1.83	86.2%	1.33
	TC>15%	82.0%	2.1	88.0%	0.6	79.0%	2.09	85.6%	1.41
	TC>30%	84.2%	2.2	85.9%	0.7	83.0%	2.20	83.1%	1.67

Spatial distribution of Producer accuracies (omissions) - TC > 10%

UMD GFC



HRL TCD

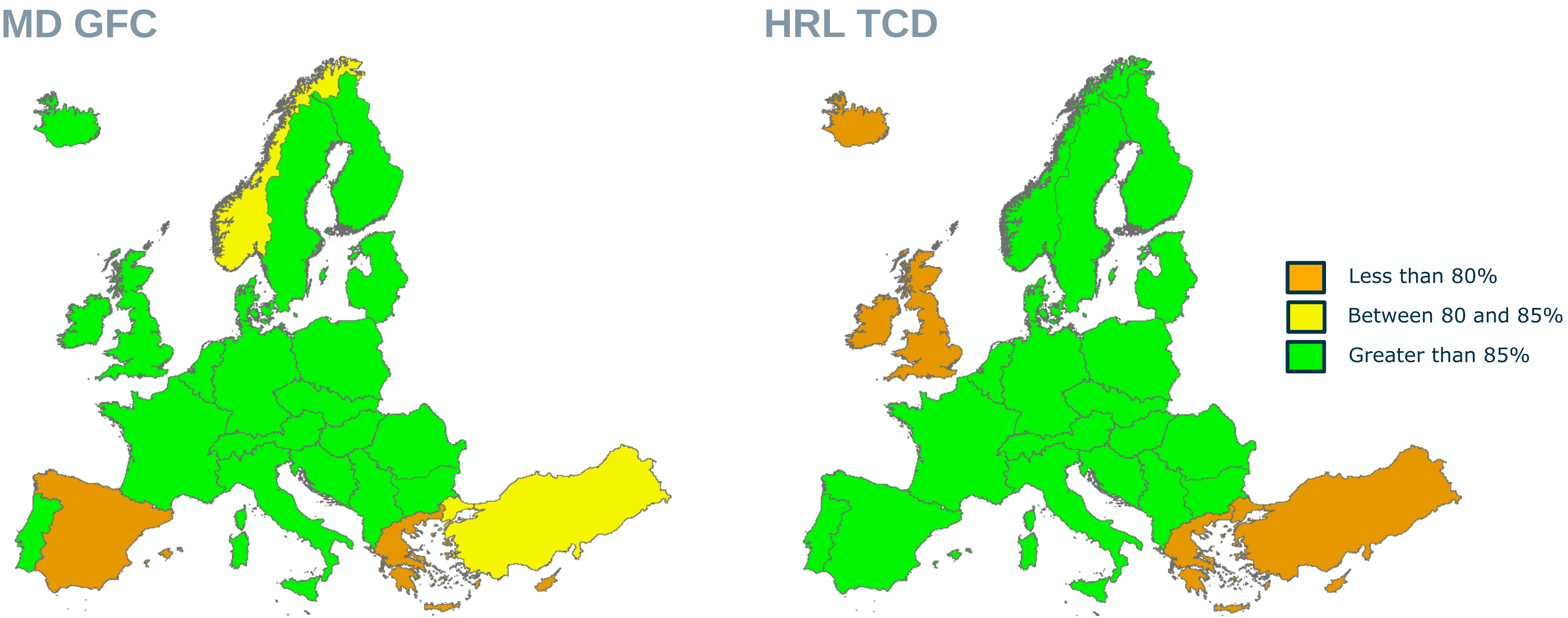


Less than 80%

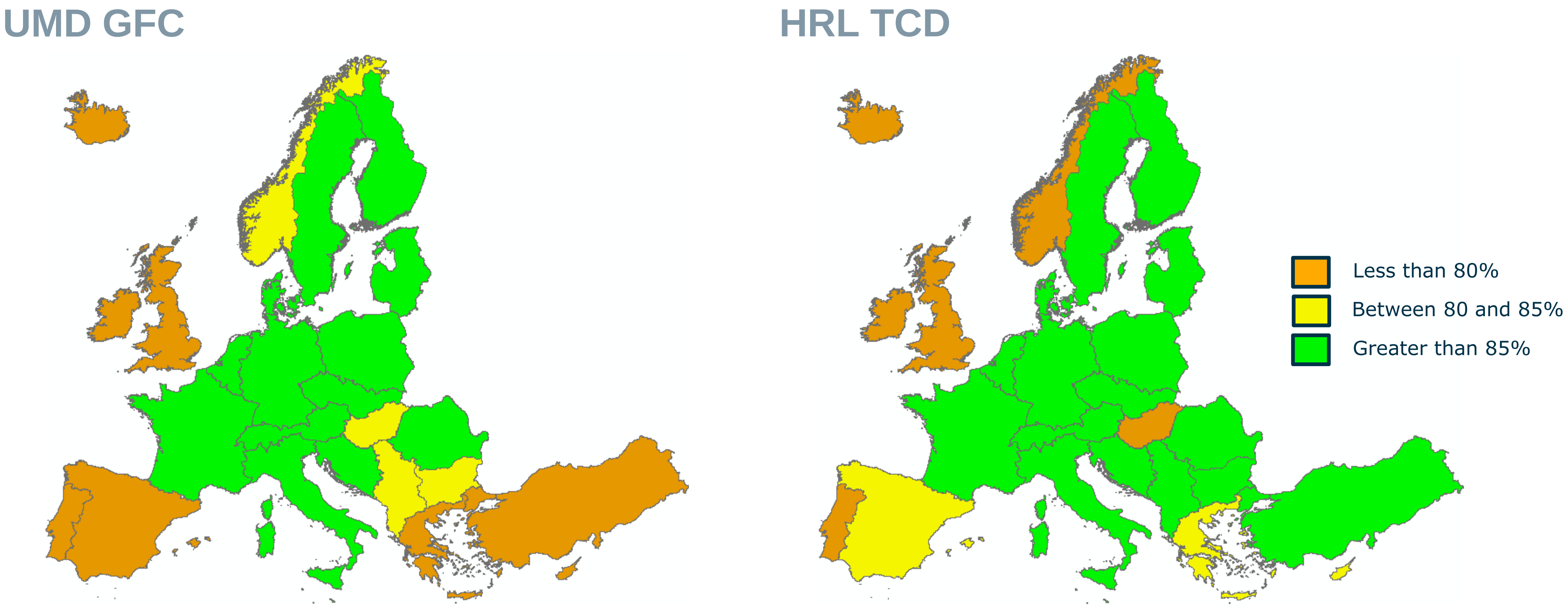
Between 80 and 85%

Greater than 85%

Spatial distribution of User accuracies (commissions) - TC > 10%

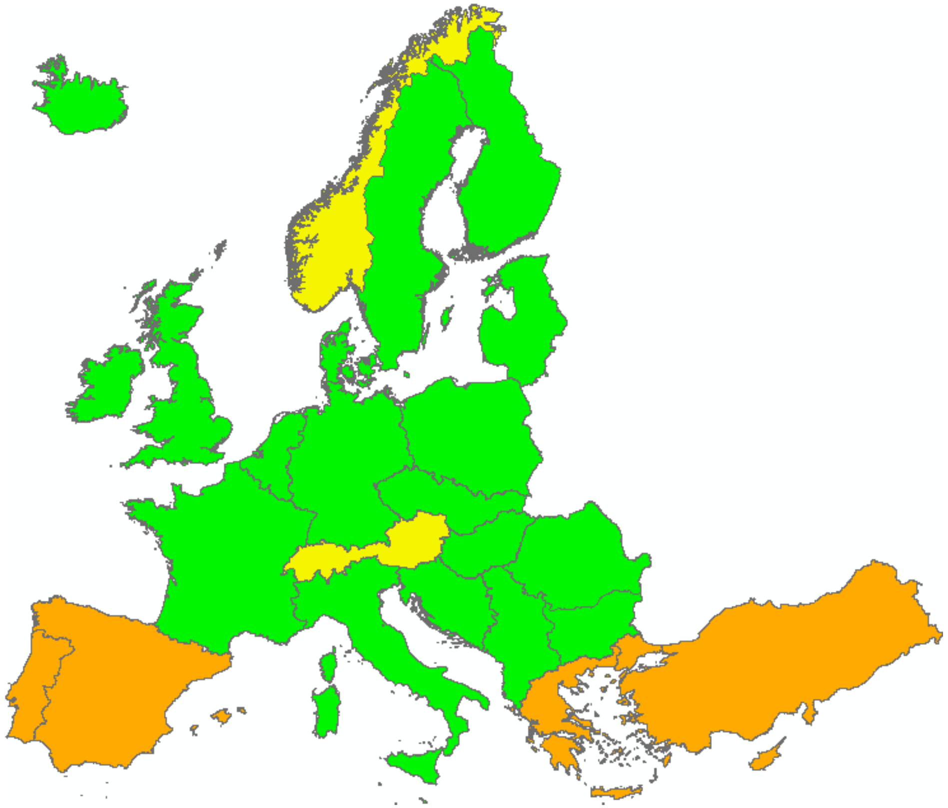


Spatial distribution of Producer accuracies (omissions) - TC > 30%

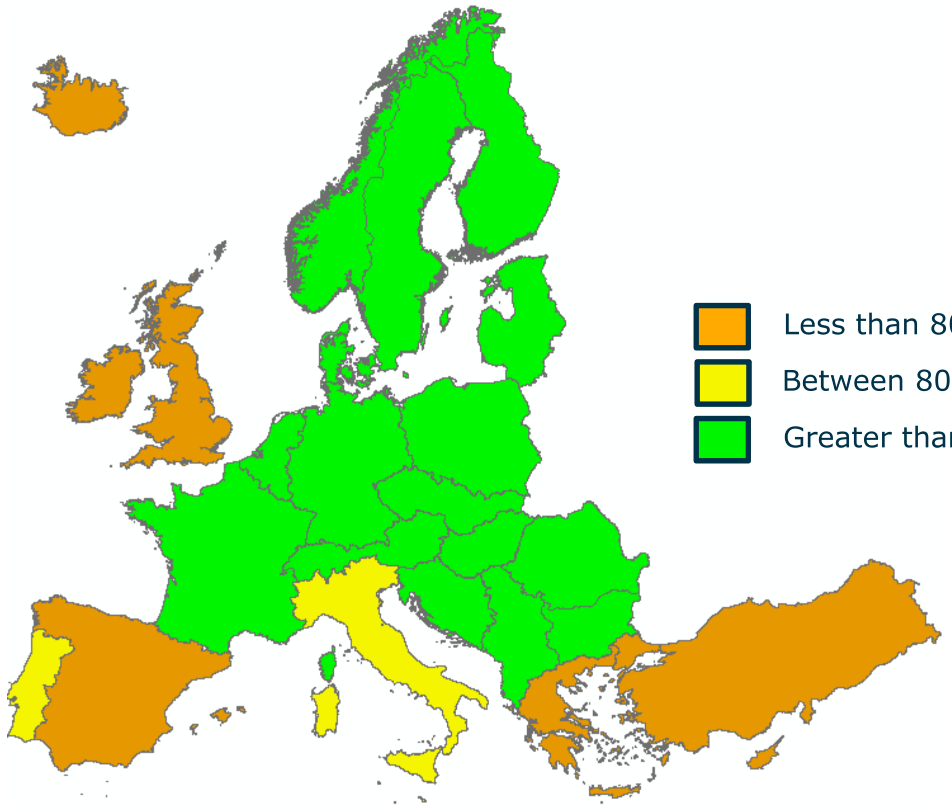


Spatial distribution of User accuracies (commissions) - TC > 30%

UMD GFC



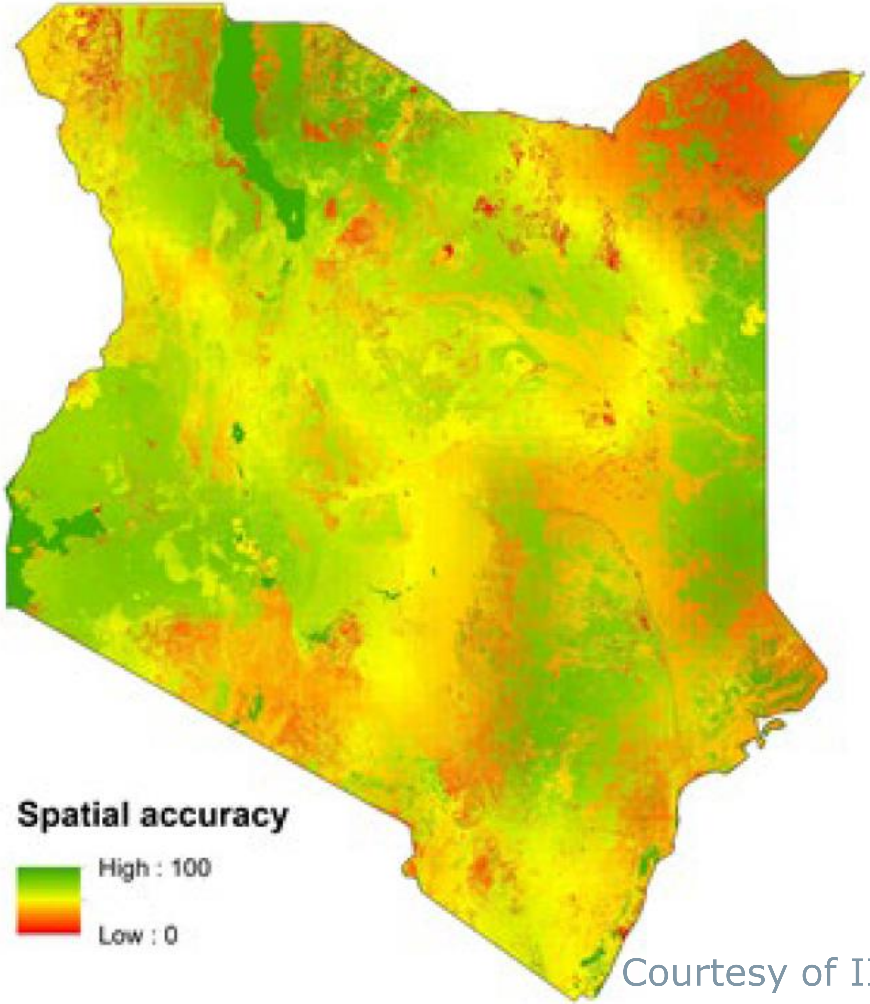
HRL TCD



Orange	Less than 80%
Yellow	Between 80 and 85%
Green	Greater than 85%

Spatial accuracy maps

- Information from accuracy assessment can be used to produce Spatial Accuracy maps
- Use of Kriging or other interpolation methods
- These maps ca provide insight on the spatial distribution of the map accuracy
- This can be used to combine several maps with data fusion
- Use in modelling approach



Courtesy of IIASA

Beyond Accuracy Assessment: statistical inference

		Reference				
		Stable NF	Stable F	Change	Total	User's accuracy
Map	Stable NF	276	6		**283**	0,977
	Stable F	20	691	1	**712**	0,970
	Change	1	2	2	**5**	0,427
Total		**297**	**699**	**3**	**1000**	
Producer's accuracy		0,930	0,988	0,651	Overall Accuracy	**0.969**

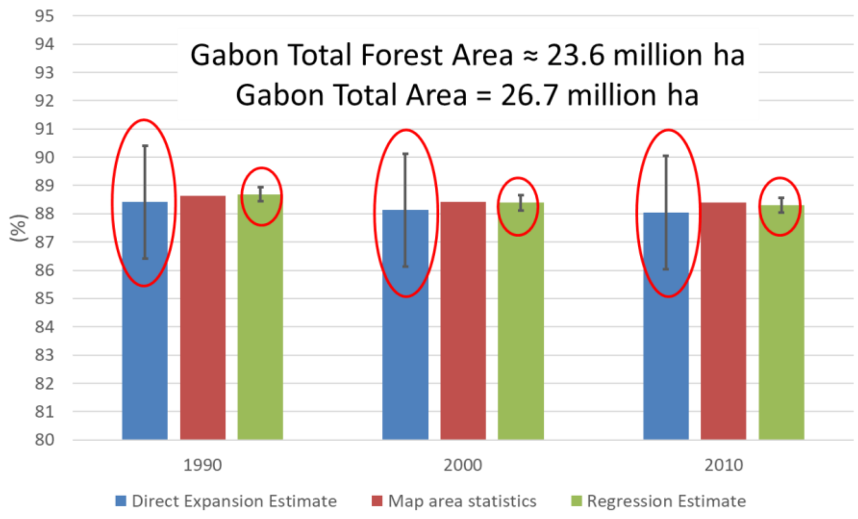
- High overall accuracy
- Accuracy of stable Forest class high
- Low accuracy of change class
- Cannot reliably extract change area statistics from the map
- Uncertainty is unknown

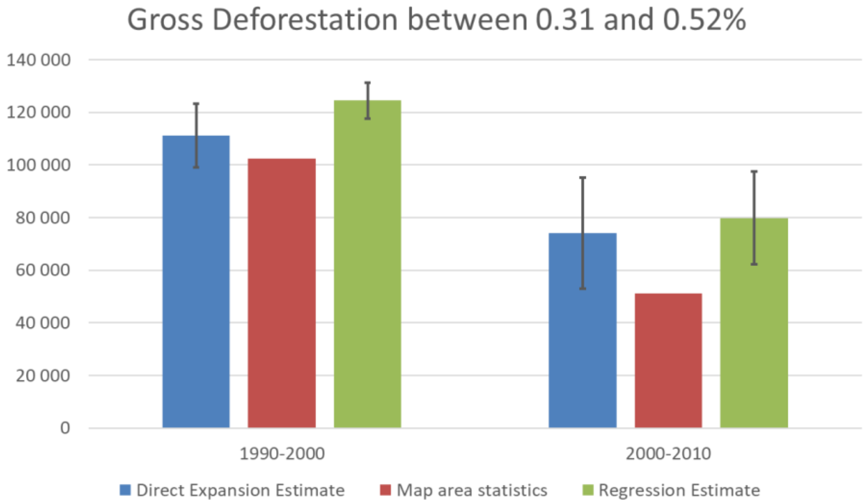
Area Estimates: Gabon Example

- Forest cover and forest cover change estimates can be produced based on samples alone (Direct Expansion)
- Observations from reference samples and the map can be combined to improve the precision of estimates (Model Assisted Regression):

$$\hat{\mu}^{MAR} = \hat{\mu}^{map} - \hat{Bias}\left(\hat{\mu}^{map}\right)$$

$$\hat{Var}\left(\hat{\mu}^{MAR}\right) = \frac{1}{m(m-1)} \sum_{i=1}^{m} \left(\Delta_i - \bar{\Delta}\right)^2$$





Summary

- Thematic Accuracy is only one yet crucial element of validating a map product, validation include additional considerations notably whether the product is fit for purpose
- The implementation of an accuracy assessment plan consists of three stages:
    1. Stratification and sampling design
    2. Response design
    3. Confusion matrix and accuracy measures
- A matrix must be unbiased in order to be able to draw reliable conclusions, it must therefore be based on a probability sample for which any difference between the sampling intensity for each stratu will have been corrected by applying a weighting factor per stratum
- Thematic accuracy is only useful for checking that the map complies with its specifications
- Statistical inference can be applied to produce Area Estimates from the analysis of the combination of reference data with the map product

References

Bellhouse D.R., 1988, Systematic sampling, Handbook of Statistics, vol. 6, ed. P.R. Krisnaiah, C.R. Rao, pp. 125-146, North-Holland, Amsterdam

Büttner G., Kosztra B., Maucha G. and Pataki R. (2012) Implementation and achievements of CLC2006, ETC-LUSI, EEA, 65p..

Cochran W., 1977, Sampling Techniques. New York: John Wiley and Sons

Czaplewski, R. L. (2003). Chapter 5: accuracy assessment of maps of forest condition: statistical design and methodological considerations, pp. 115–140. In Michael A.Wulder, & Steven E. Franklin (Eds.), Remote sensing of forest environments: concepts and case studies. Boston: Kluwer Academic Publishers (515 pp.).

Deville J.C., Särndal C.E., 1992, Calibration estimators in survey sampling, Journal of the American Statistical Association, Vol 87, pp. 376-382

Gallego, J. (1995). Sampling Frames of Square Segments. Official Publication of the European Communities, EUR 16317, 72 pp.

Gallego, J. (2004a). Area Frames for Land Cover Estimation: Improving the European LUCAS Survey. Third International Conference on Agricultural Statistics. November 2-4, 2004, Mexico.

Gallego, J. (2004b). Remote sensing and land cover area estimation. International Journal of Remote Sensing, 25, 3019-3047. DOI:10.1080/01431160310001619607

Gallego F.J., (2006) Review of the Main Remote Sensing Methods for Crop Area Estimates, Remote Sensing Support to Crop Yield Forecast and Area Estimates ISPRS archives, XXXVI, 8/W48, 65-70. http://www.isprs.org/publications/PDF/ISPRS_Archives_WorkshopStresa2006.pdf

Gallego J., Sannier Ch., Pennec A., (2016b), Validation of Copernicus Land Monitoring Services and Area Estimation. ICAS VII, Rome, October 26-28 2016. https://www.istat.it/storage/icas2016/g42-gallego.pdf

GFOI (2016), Integration of remote-sensing and ground-based observations for estimation of emissions and removals of greenhouse gases in forests: Methods and Guidance from the Global Forest Observations Initiative, Edition 2.0, Food and Agriculture Organization, Rome.

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau,S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend. (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. Science 342 (15 November): 850–53. Data available on-line from:http://earthenginepartners.appspot.com/science-2013-global-forest.

Laso Bayas, J. C., Lesiv, M., Waldner, F., Schucknecht, A., Duerauer, M., See, L., Fritz S., Fraisl D. MoorthyI., Mc. Callum, Perger Ch., Danylo O., Defourny P., Gallego J., . . . Wilson, J. W. (2017). A global reference database of crowdsourced cropland data collected using the geo-wiki platform. Scientific Data, 4 doi:10.1038/sdata.2017.136

References

Lesiv, M., Fritz, S., McCallum, I., Tsendbazar, N., Herold, M., Pekel, J. F., ... & Van De Kerchove, R. (2017). Evaluation of ESA CCI prototype land cover map at 20m.

Lowell K., Jaton A. (ed), 1999, Spatial Accuracy assessment: Land information uncertainty in natural resources, Ann Arbor Press, Chelsea, Michigan, USA. 443 pp.

McRoberts, R. E. (2011). Satellite image-based maps: Scientific inference or pretty pictures?. Remote Sensing of Environment, 115(2), 715-724.

McRoberts, R. E., & Walters, B. F. (2012). Statistical inference for remote sensing-based estimates of net deforestation. Remote Sensing of Environment, 124, 394–401. doi:10.1016/j.rse.2012.05.011

McRoberts, R. E., Vibrans, A. C., Sannier, C., Næsset, E., Hansen, M. C., Walters, B. F., & Lingner, D. V. (2016). Methods for evaluating the utilities of local and global maps for increasing the precision of estimates of subtropical forest area. Canadian Journal of Forest Research, 46(7), 924-932.

McRoberts, R. E., Stehman, S. V., Liknes, G. C., Næsset, E., Sannier, C., & Walters, B. F. (2018). The effects of imperfect reference data on remote sensing-assisted estimators of land cover class proportions. ISPRS Journal of Photogrammetry and Remote Sensing, 142, 292-300.

McRoberts, R. E., Næsset, E., Sannier, C. Stehman, S. V. and Tomppo, E. (2020) Remote sensing support for the gain-loss approach for greenhouse gas inventories. Remote Sensing, final review

Olofsson, P., Foody, G. M., Stehman, S. V., & Woodcock, C. E. (2013). Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. Remote Sensing of Environment, 129, 122–131. doi:10.1016/j.rse.2012.10.031

Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., Wulder, M.A., 2014, Good practices for estimating area and assessing accuracy of land change. Remote Sensing of Environment 148, 42 - 57

Rao, J.N.K. (2003). Small Area Estimation. Wiley, New York.

Sannier, C., McRoberts, R. E., Fichet, L. -., & Makaga, E. M. K. (2014). Using the regression estimator with landsat data to estimate proportion forest cover and net proportion deforestation in gabon. Remote Sensing of Environment, 151, 138-148.

Sannier, C.; Gallego, J.; Dahmer, J.; Smith, G.; Dufourmont, H.; Pennec, (2016a) A. Validation of Copernicus high resolution layer on imperviousness degree for 2006, 2009 and 2012. Spatial Accuracy Assessment, Montpellier, 5–8 July 2016. http://www.spatial-accuracy.org/system/files/Accuracy%202016%20Proceedings.pdf

Sannier, C., McRoberts, R. E., & Fichet, L. V. (2016b). Suitability of Global Forest Change data to report forest cover estimates at national level in Gabon. Remote Sensing of Environment, 173, 326-338.

Särndal, C.-E., Swensson, B., Wretman, J.H., 1992. Model assisted survey sampling. Springer, New York.

References

Selkowitz, D. J., & Stehman, S. V. (2011). Thematic accuracy of the National Land Cover Database (NLCD) 2001 land cover for Alaska. Remote Sensing of Environment, 115(6), 1401–1407. doi:10.1016/j.rse.2011.01.020.

Stehman, S. V. (2009). Sampling designs for accuracy assessment of land cover. International Journal of Remote Sensing, 30(20), 5243-5272.

Stehman, S. V., & Czaplewski, R. L. (1998). Design and analysis for thematic map accuracy assessment: fundamental principles. Remote Sensing of Environment, 64, 331–334.

Stehman, S. V., & Foody, G. M. (2008). Accuracy assessment. The SAGE handbook of remote sensing (pp. 297-314) doi:10.4135/9780857021052.n21

Taylor J., Sannier C., Delincé J, Gallego F.J. (1997), Regional Crop Inventories in Europe Assisted by Remote Sensing: 1988-1993. Synthesis Report . EUR 17319 EN, Office for Pubblications of the EC. Luxembourg. 71pp

Wack, Hirschmugl, Koboltschnig, Schmitt, 2012. Validation Methods for REDD, Technical Note for the ESA GSE Forest Monitoring REDD Extension services project.

Waldner F., Schucknecht A., Lesiv M., Gallego J., et al., 2018, Conflation of expert and crowd reference data to validate global binary thematic maps. Remote Sensing of Environment, vol 221, pp 235-246.

Westfall, J.A., Patterson, P.L., Coulston, J.W., 2011. Post-stratified estimation: within-strata and total sample size recommendations. Can. J. For. Res. 41, 1130–1139. https://doi.org/10.1139/x11-031

Thank you for your attention

Dr Christophe Sannier csannier@groupcls.com