An integrative sampling approach for validating spatially non-contiguous remote sensing products

Chima Iheaturu, Vladimir Wingate, Felicia Akinyemi, and Chinwe Ifejika Speranza

University of Bern, Institute of Geography,
Land Systems & Sustainable Land Management (LS-SLM),
Bern, Switzerland.

European Research Council
Established by the European Commission

u^b

b
UNIVERSITÄT
BERN

Sharing not permitted

Background

❖ Remote sensing (RS) datasets are becoming more available and accessible.

❖ The **usability** and **reliability** of RS-derived products are dependent upon the product's accuracy and the **intended application and/or use**.
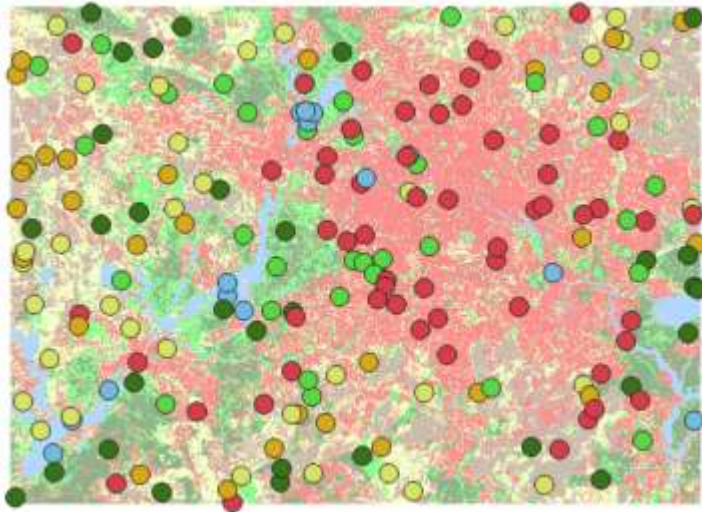
How accurate is the data that we use?

Global Forest Change 2000–2020 (Hansen et al., 2013)
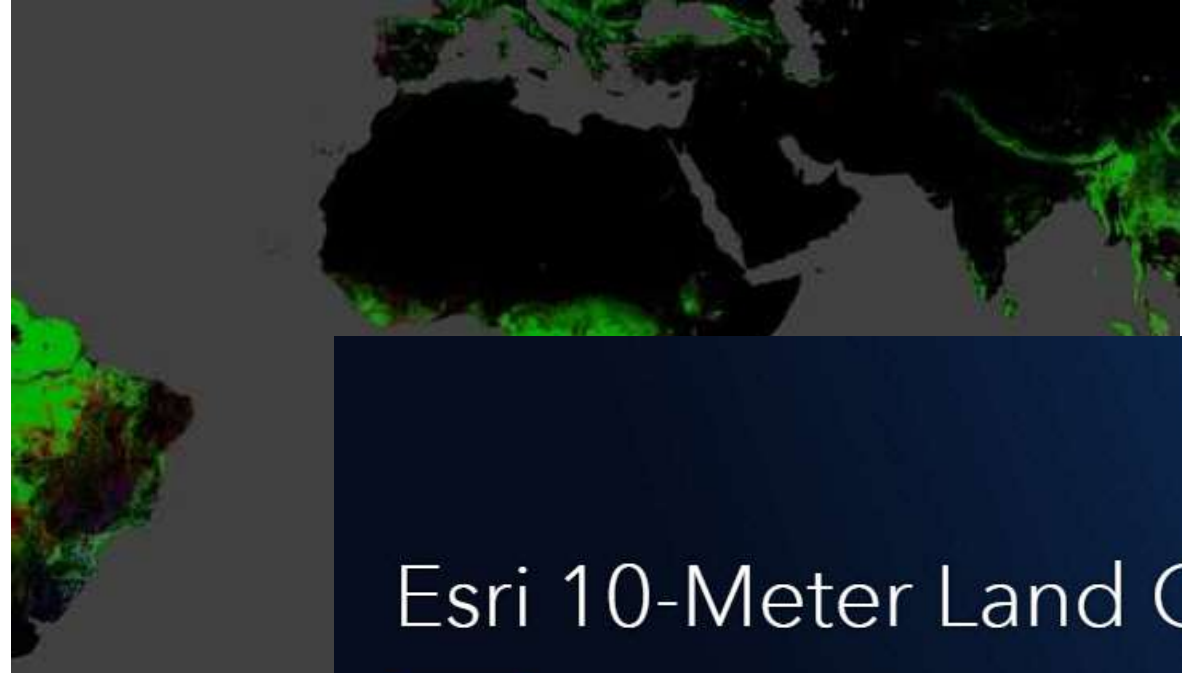
Esri 2020 land use/land cover (Karra et al., 2021)

Accuracy assessment for RS products

Sampling

Classified data

E.g, simple random sampling, systematic random sampling, stratified random sampling, proportional sampling, etc.

Reference data

Sample points with labels representing the 'reality' for the land use/cover class, e.g.,

- 🔵 Water
- 🟢 Forest
- 🔴 Urban

Cross validation

		Reference Data			
		Water	Forest	Urban	Total
Classified Data	Water	21	6	0	27
	Forest	5	31	1	37
	Urban	7	2	22	31
	Total	33	39	23	95

An integrative sampling approach for validating spatially non-contiguous remote sensing products

Spatially non-contiguous RS datasets

Spatially non-contiguous RS datasets are:

- Discrete
- Not spatially contiguous
- Not normally distributed

How do we validate **spatially non-contiguous** RS datasets **?**
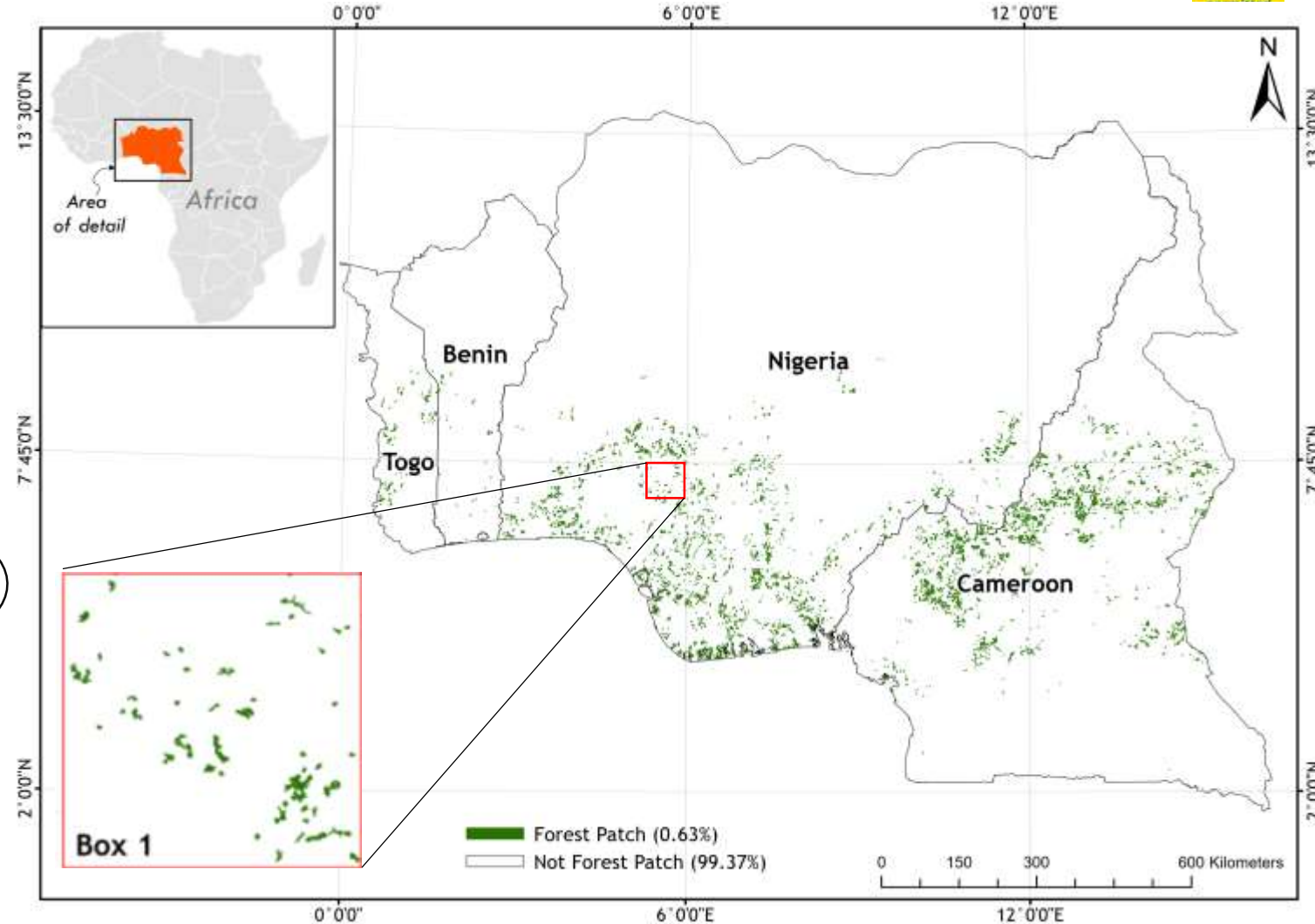
Forest patch dataset for Togo, Benin, Nigeria & Cameroon

An integrative sampling approach for validating spatially non-contiguous remote sensing products

The ISRS Approach

- ❖ ISRS optimally combines 3 commonly used sampling approaches

- ❖ It determines the sample size using Slovin's formula:

$$n = \frac{N}{1 + Ne^2}$$

- ❖ It uses K-means clustering algorithm to select the sample locations
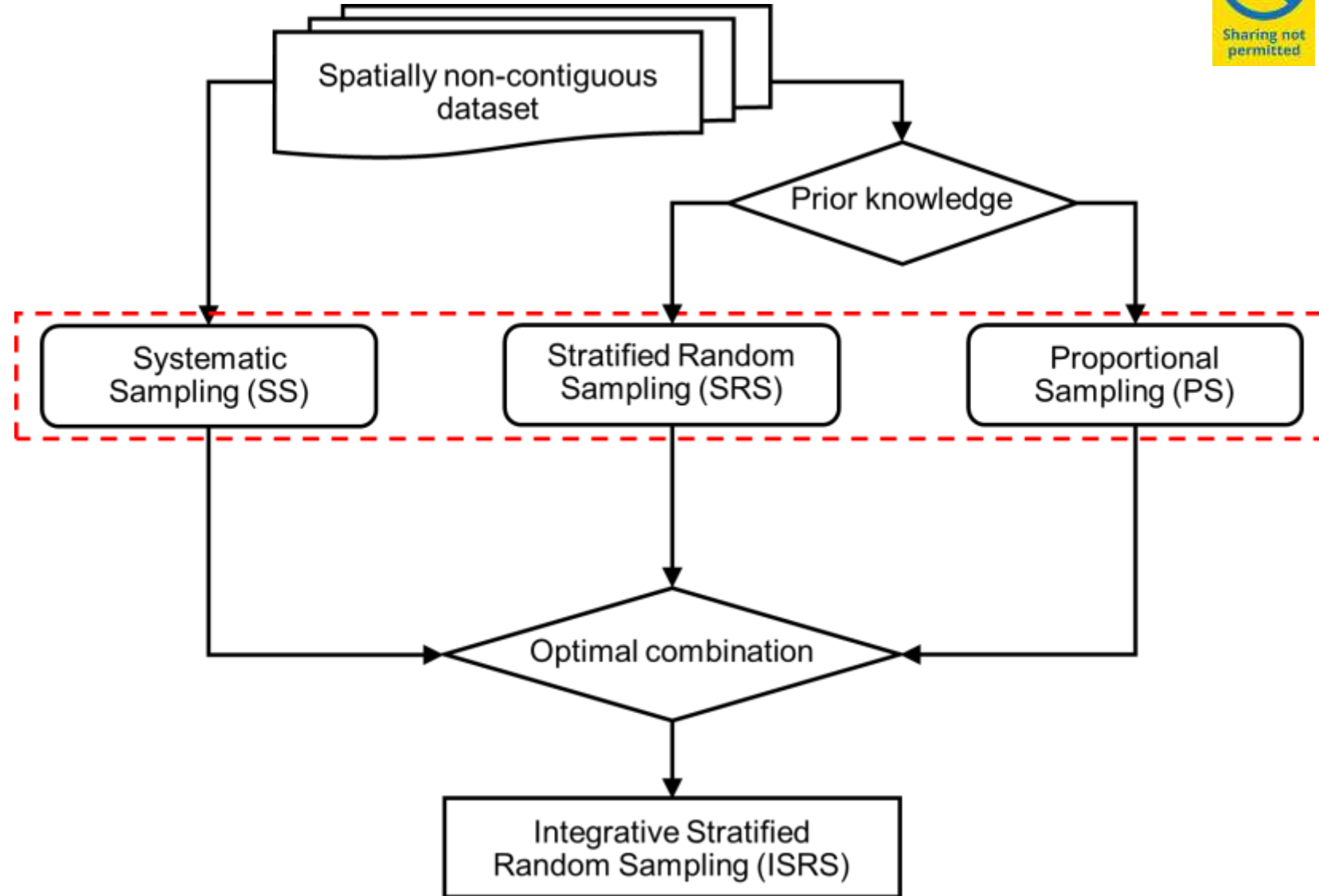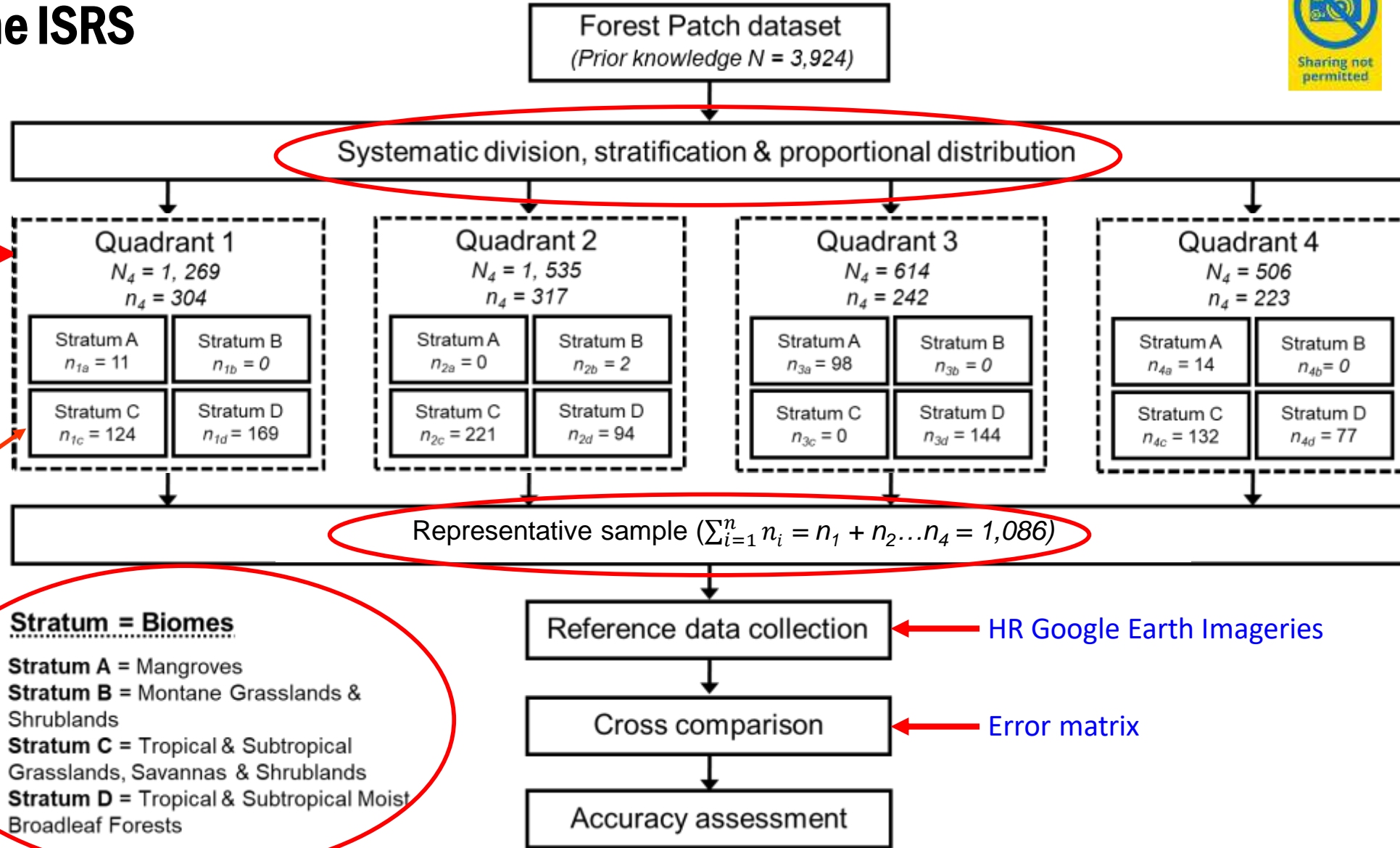
Operationalising the ISRS Approach

Forest Patch dataset
(Prior knowledge N = 3,924)

Systematic division, stratification & proportional distribution

Sample size

Slovin's formula $n = \dfrac{N}{1 + Ne^2}$

N = Population

n = Sample size

e = Margin of error

Quadrant 1
$N_4 = 1,269$
$n_4 = 304$

| Stratum A $n_{1a} = 11$ | Stratum B $n_{1b} = 0$ |
| Stratum C $n_{1c} = 124$ | Stratum D $n_{1d} = 169$ |

Quadrant 2
$N_4 = 1,535$
$n_4 = 317$

| Stratum A $n_{2a} = 0$ | Stratum B $n_{2b} = 2$ |
| Stratum C $n_{2c} = 221$ | Stratum D $n_{2d} = 94$ |

Quadrant 3
$N_4 = 614$
$n_4 = 242$

| Stratum A $n_{3a} = 98$ | Stratum B $n_{3b} = 0$ |
| Stratum C $n_{3c} = 0$ | Stratum D $n_{3d} = 144$ |

Quadrant 4
$N_4 = 506$
$n_4 = 223$

| Stratum A $n_{4a} = 14$ | Stratum B $n_{4b} = 0$ |
| Stratum C $n_{4c} = 132$ | Stratum D $n_{4d} = 77$ |

K-means clustering algorithm

Representative sample ($\sum_{i=1}^{n} n_i = n_1 + n_2 \ldots n_4 = 1,086$)

Stratum = Biomes

Stratum A = Mangroves
Stratum B = Montane Grasslands & Shrublands
Stratum C = Tropical & Subtropical Grasslands, Savannas & Shrublands
Stratum D = Tropical & Subtropical Moist Broadleaf Forests

Reference data collection ← **HR Google Earth Imageries**

Cross comparison ← **Error matrix**

Accuracy assessment

QGIS

ArcGIS

R

What makes the ISRS strategy reliable ?

❖ It minimises spatial autocorrelation by optimizing the distances between the sample locations

❖ It ensures that a representative sample is selected

Next steps:

❖ Present the results and complete the paper
❖ Publish in a suitable peer-review journal

An integrative sampling approach for validating spatially non-contiguous remote sensing products