Efficient Aggregate Land Cover Queries with Cloud-Optimized Raster Formats

4th September 2025 – MIGARS 2025, Bucharest

Luke McQuade, Dr. Martin Sudmanns, Assoc. Prof. Dirk Teide
@ EO Analytics, Dept. of Geoinformatics, PLUS

Introduction

- *“Find the images with less than 5% cloud cover between two dates”*
- *“Find the first image of the year where the vegetation content exceeds 20%”*

Sentinel-2’s Scene Classification Layer (SCL) can be used to help answer these.

But 30 megapixels per scene × 60 scenes per year => many pixels!

Q: Do we need to process every pixel?

Specifically:

- To what extent can downsampling be applied?
- How does this vary with size of the area of interest (AoI)?
- Assume an error tolerance of 1%.
- Follow up: What part (if any) does spatial distribution play?

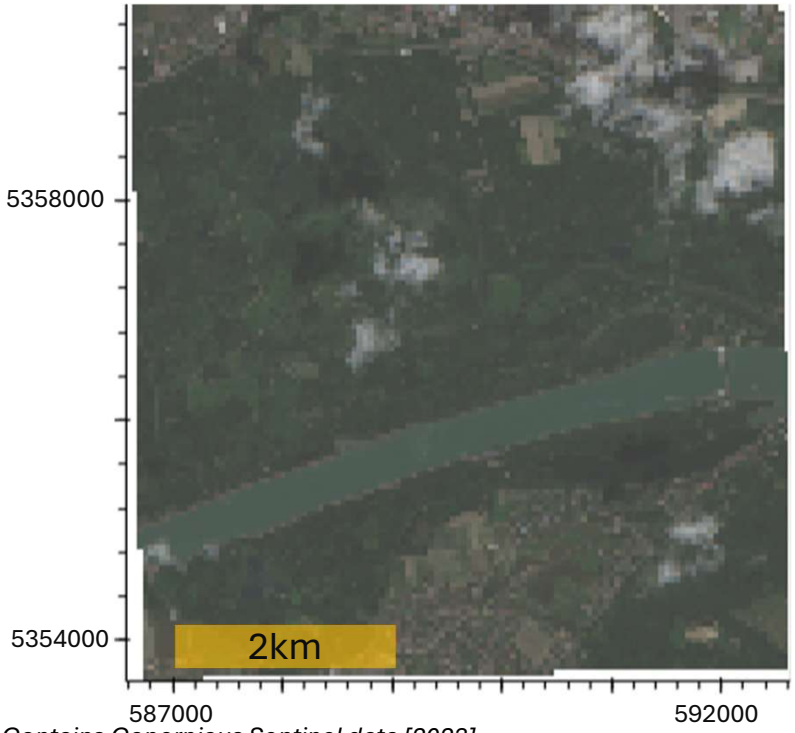
Background (I): Scene Classification Layer (SCL)

- Part of Sentinel-2 Level 2A product.
- Native resolution 20m.
- Aka. Scene Classification Map.

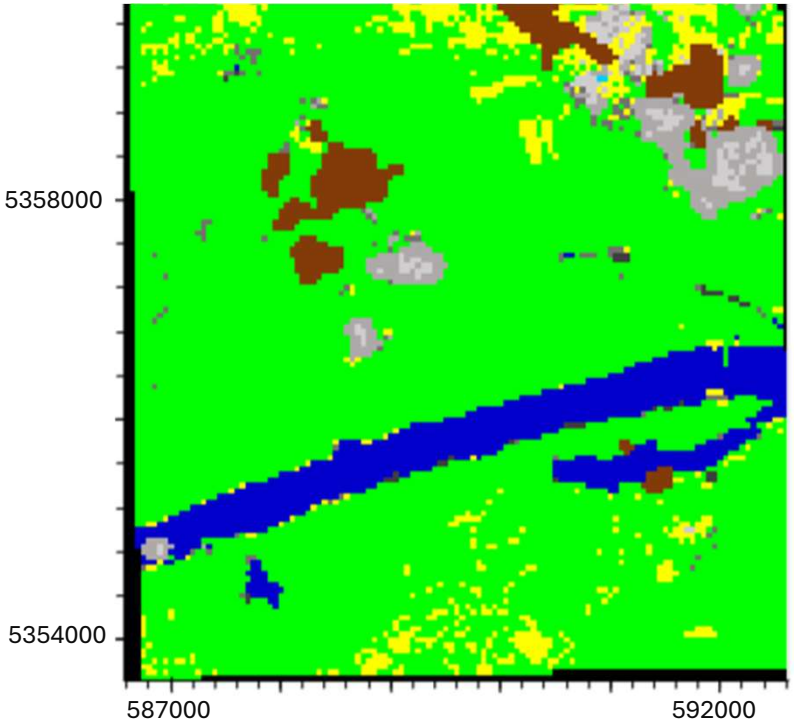
Source & how it works:
ESA Sentinel-2 Level-2A Algorithm Theoretical Basis Document, 2021 [1]

Image date 2023-06-29, granule 33UWP (NE Austria) (subset)

RGB



SCL



No data
Saturated
Dark
Shadow
Vegetation
Bare soil/built-up
Water
Unknown
Cloud (med. chance)*
Cloud (high chance)*
Thin cirrus*
Snow

*Combined in our analysis

Contains Copernicus Sentinel data [2023]

Background (II)

Downsampling:

- Uniform/nearest neighbor, random, various interpolations & adaptive algorithms, e.g., *Mîrț et al., 2022* [2].
- Random is less accurate than uniform for downstream tasks => spatial distribution is important [2].

Cloud Optimized GeoTIFF (COG):

- Tiling & pyramids/overviews => allows sub-setting and multiscale without reading the whole file [via HTTP range requests] (*https://cogeo.org* [3])
- Optimal usage with uniform downsampling.

Methodology: Study areas

Two base areas of interest (AoI):

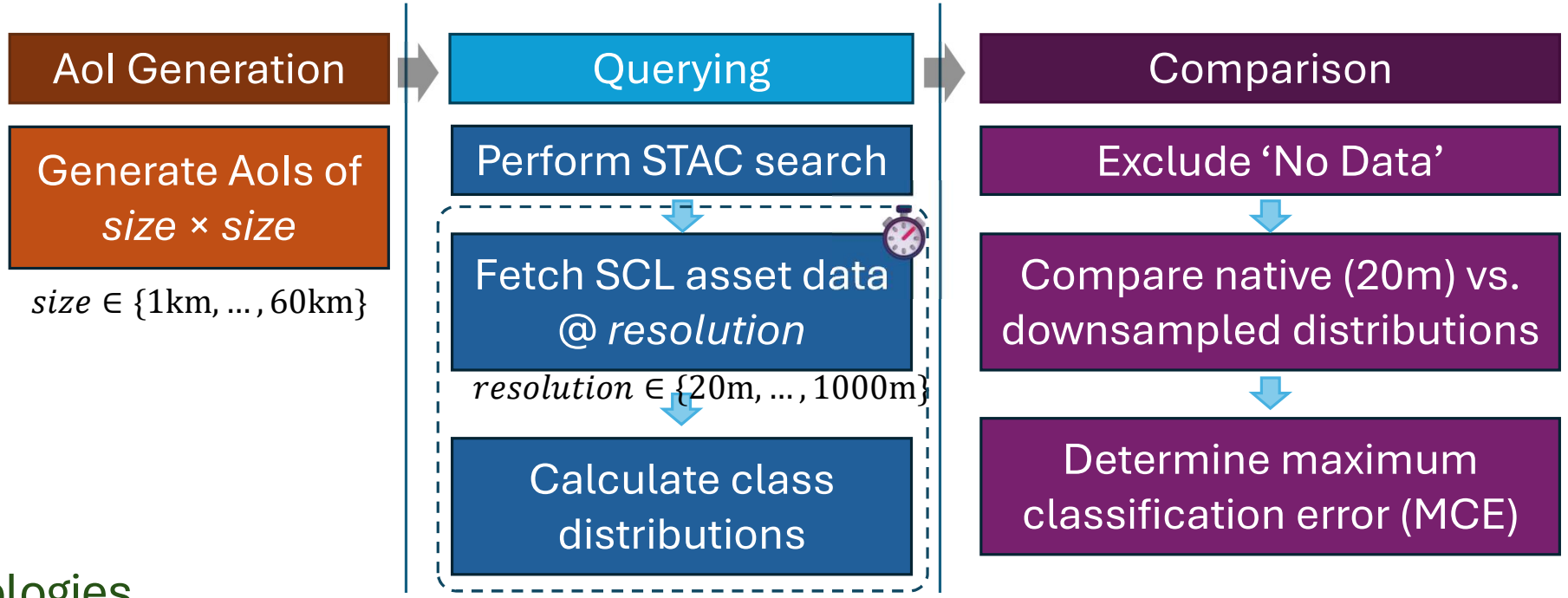
- West of Vienna, Austria
- Duoro Valley, Spain

→ Derived tiles 1km×1km to 60km×60km (synthetic AoIs)

Study Period:
1st June to 1st July 2023

=> Approx. 10 unique Sentinel-2 scenes

Methodology: Overview (accuracy & performance assessment)



Technologies

GeoPandas [4]

Microsoft Planetary Computer [5]

marimo [6]

stackstac [7]

HoloViews [8]

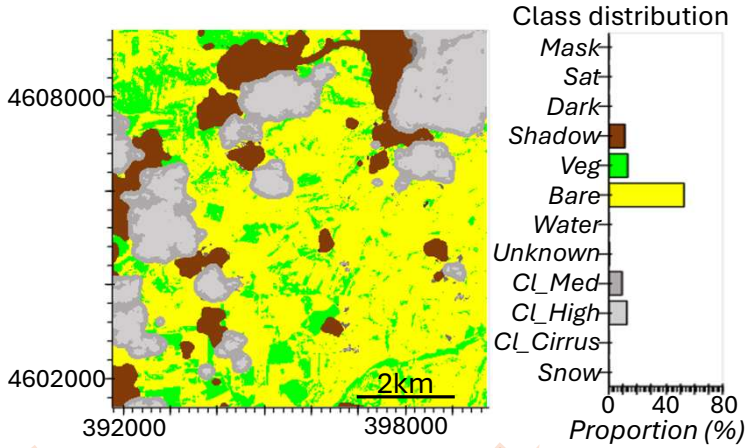
pandas [9]

xarray [10]

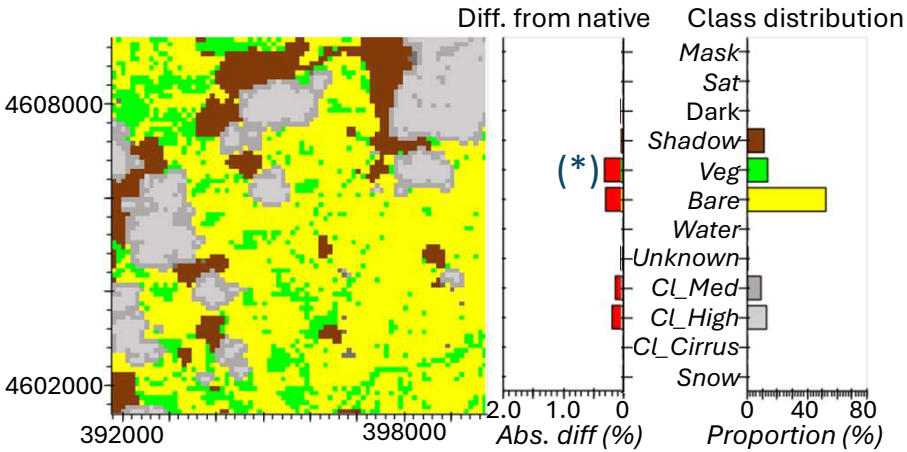
Methodology: Example

Spain 8km × 8km subset,
image date 2023-06-05,
granule 30TUM

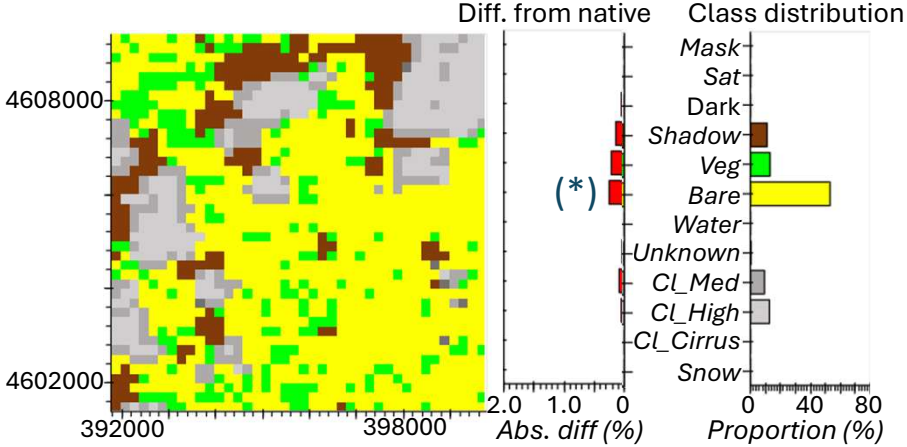
Native resolution (20m)



Downsampled to 100m



Downsampled to 200m



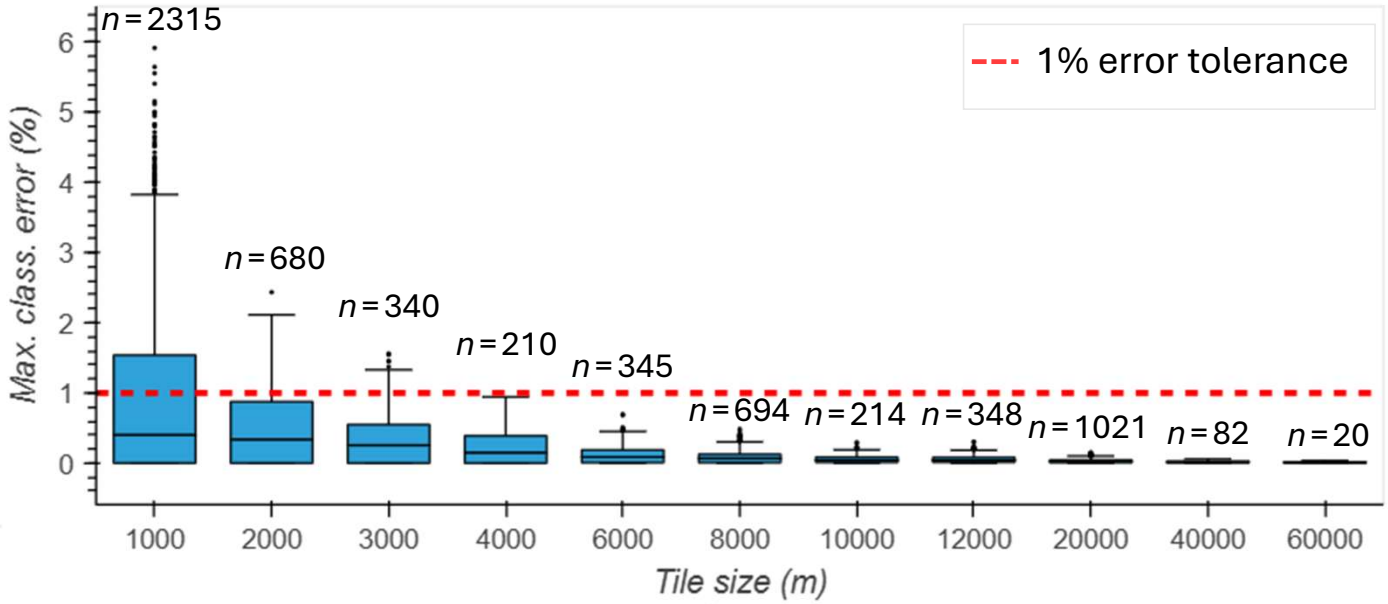
(*) Maximum classification error (MCE)

Contains modified Copernicus Sentinel data [2023]

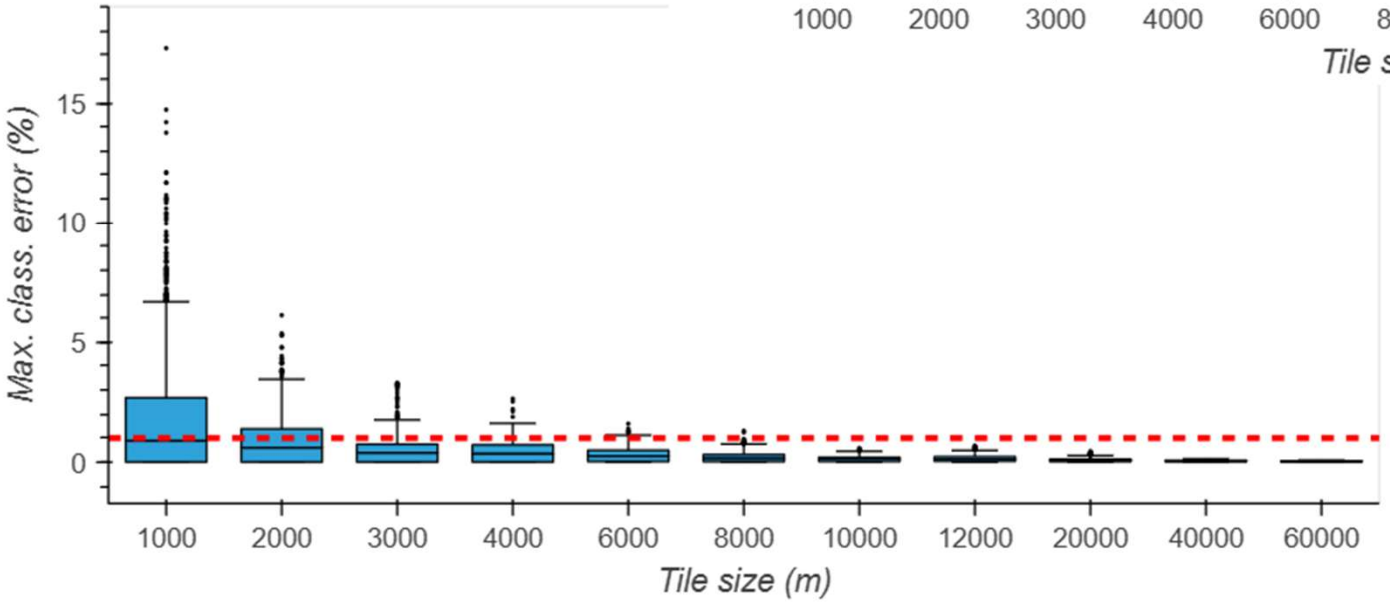
Results (I)

Distributions of MCE due to downsampling across AoIs

Downsampled to 50m

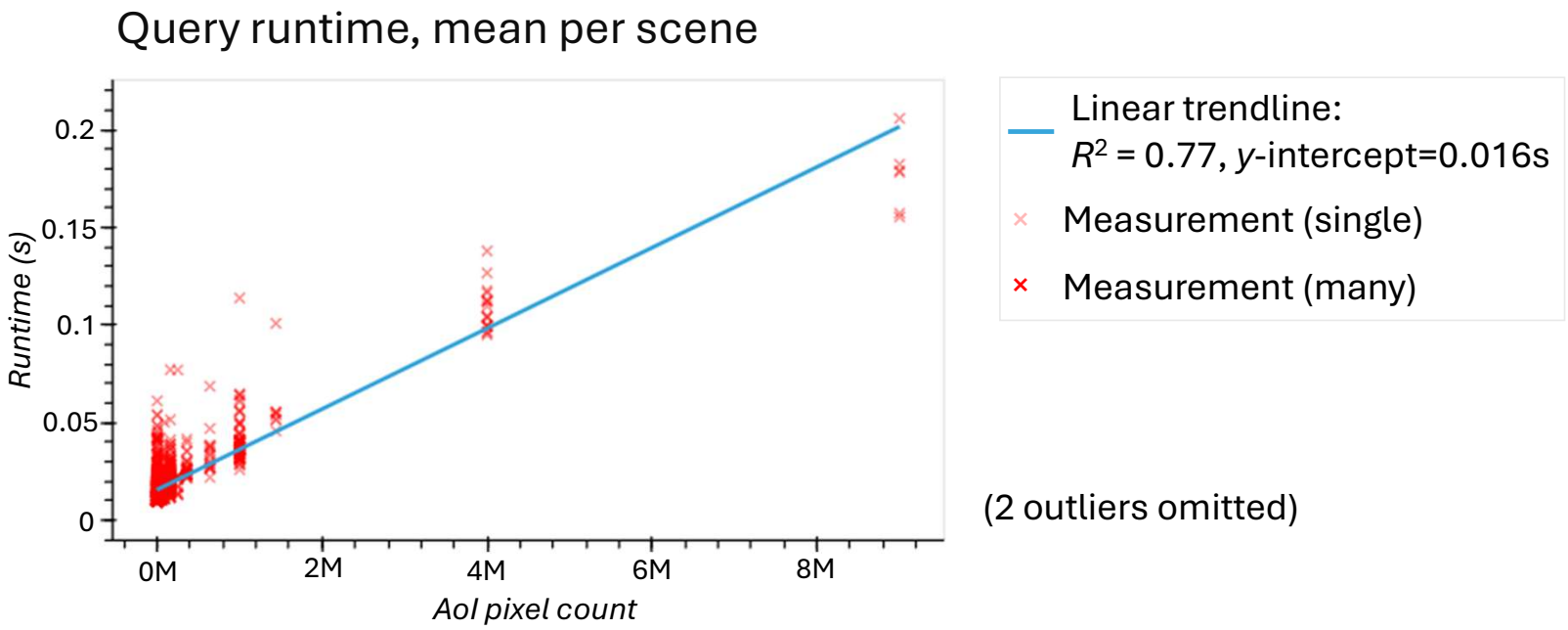


Downsampled to 100m



Note: *n* was later found to include duplicates due to multiple baselines; actual *n* is effectively half.

Results (II)



(2 outliers omitted)

Note: Performance measurements were taken before ‘No Data’ filtering step

Results (III)

Maximum downsampling resolution before MCE exceeds 1%, and associated data and estimated query runtime reductions

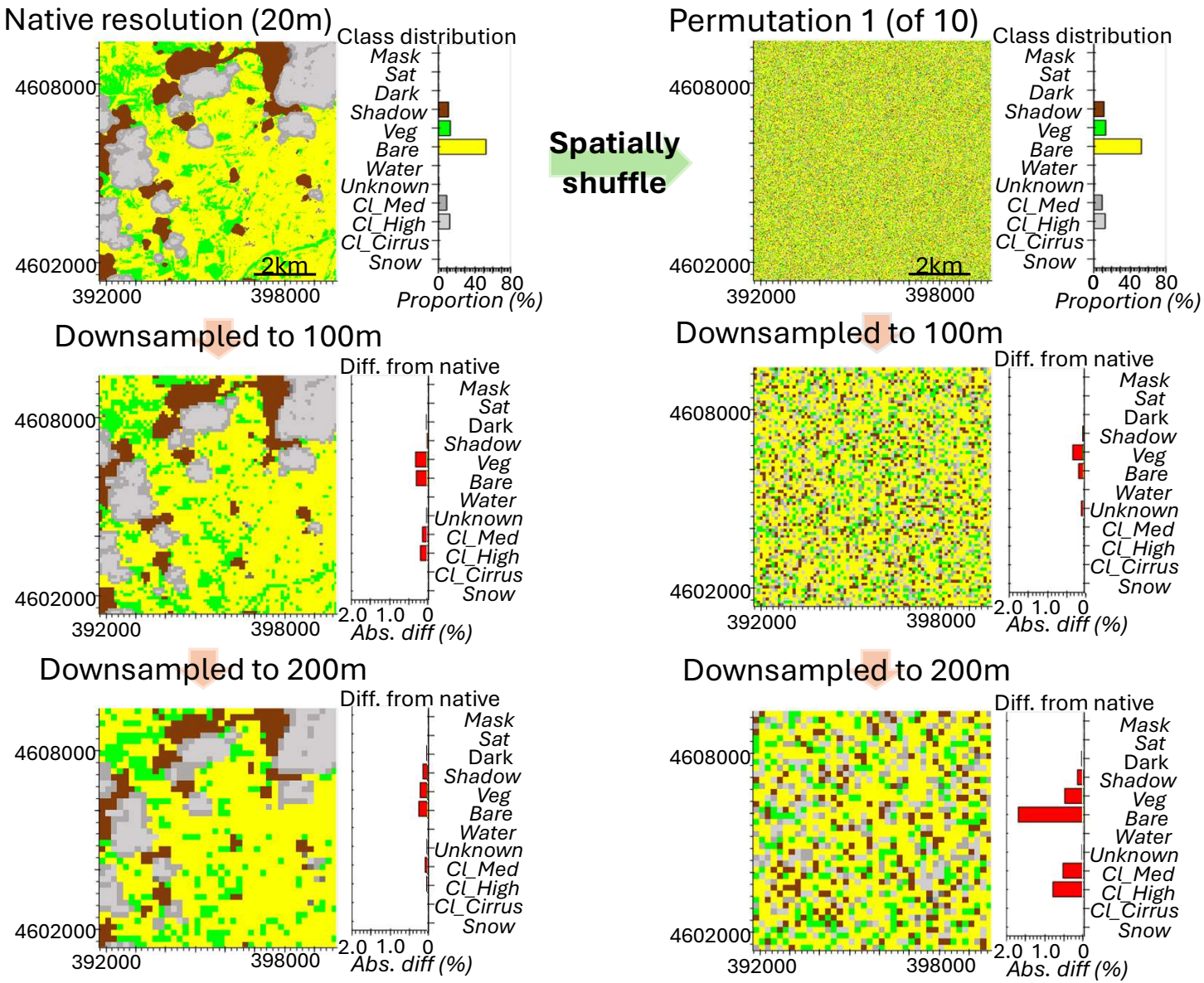
Tile size (m)	Max. downsampling resolution (m)	Pixels (native)	Pixels (downsampled)	Data reduction (multiple)	Est. runtime/scene (native) (s)	Est. runtime/scene (downsampled) (s)	Est. runtime reduction (multiple)
1000	N/A	2500	N/A	N/A	0.0157	N/A	N/A
2000	N/A	10000	N/A	N/A	0.0159	N/A	N/A
3000	N/A	22500	N/A	N/A	0.0161	N/A	N/A
4000	50	40000	6400	6.25	0.0165	0.01578	1.04
6000	50	90000	14400	6.25	0.0175	0.01595	1.10
8000	50	160000	25600	6.25	0.0190	0.01618	1.17
10000	100	250000	10000	25	0.0208	0.01586	1.31
12000	100	360000	14400	25	0.0231	0.01595	1.45
20000	200	1000000	10000	100	0.0363	0.01586	2.29
40000	500	4000000	6400	**625**	0.0983	0.01578	6.23
60000	500	9000000	14400	**625**	0.2016	0.01595	**12.64**

"N/A" = error tolerance was exceeded for all tested downsampling resolutions (50m, 100m, 200m, 500m, 1000m).

Follow-up:

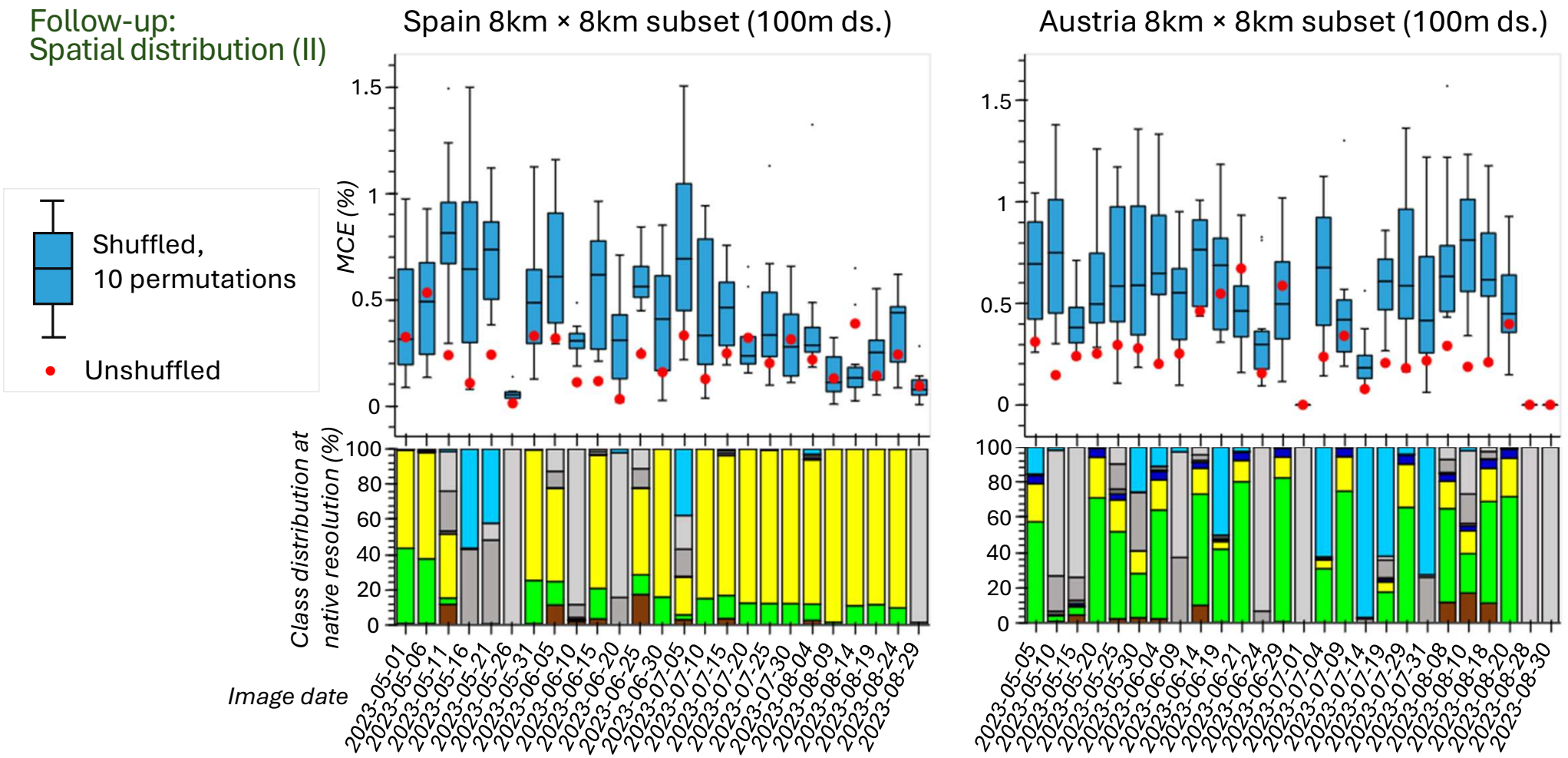
Spatial distribution

Spain 8km × 8km subset, image date 2023-06-05, granule 30TUM



Contains modified Copernicus Sentinel data [2023]

Follow-up: Spatial distribution (II)



Downsampling to 100m: of 51 images, 39 (**76.5%**) had lower MCE due to downsampling than the majority of spatially shuffled permutations. Of these, the mean reduction in MCE was **0.29%**.

Downsampling to 200m: of 51 images, 41 (**80.4%**) had lower MCE due to downsampling than the majority of spatially shuffled permutations. Of these, the mean reduction in MCE was **0.50%**.

Discussion

Why/when it works:

- Spatial autocorrelation: size of feature > spatial resolution

Future work:

- Test with more regions.
- Test with more resolutions, especially for small AoIs.
- Test with more satellites, e.g., Landsat 8/9 (work in progress).
- Finer-grained categories => less effective?

Tip: For AoIs with similar footprint to S2 granules, use STAC metadata, e.g., `s2:vegetation_percentage` [11]

Acknowledgements

- Many thanks to our colleague, Felix Kröber, whose earlier explorations and Jupyter notebooks on Sentinel-2 STACs and the SCL were a great help.

- Funded by the Horizon Europe programme of the European Union under the LEONSEGS project (project #101082493).

https://leonsegs.eu/

References

[1] European Space Agency, "Sentinel-2 Level-2A Algorithm Theoretical Basis Document," ESA – European Space Agency, Frascati, Italy, Technical Report S2-PDGS-MPC-ATBD-L2A, Nov. 2021.

[2] A. Mîrț, J. Reiche, J. Verbesselt, and M. Herold, "A Downsampling Method Addressing the Modifiable Areal Unit Problem in Remote Sensing," Remote Sensing, vol. 14, no. 21, Art. no. 21, Jan. 2022, doi: 10.3390/rs14215538.

[3] Cloud Optimized GeoTIFF Community, "Cloud Optimized GeoTIFF." Accessed: May 14, 2025. [Online]. Available: https://cogeo.org/

[4] Joris Van den Bossche et al., geopandas/geopandas: v1.0.1. (Jul. 02, 2024). Zenodo. doi: 10.5281/ZENODO.12625316.

[5] Microsoft Open Source, M. McFarland, R. Emanuele, D. Morris, and T. Augspurger, microsoft/PlanetaryComputer: October 2022. (Oct. 2022). Zenodo. doi: 10.5281/zenodo.7261897.

[6] A. Agrawal and M. Scolnick, marimo. (2025). GitHub. [Online]. Available: https://github.com/marimo-team/marimo

[7] G. Joseph, gjoseph92/stackstac. (May 01, 2025). GitHub. Accessed: May 18, 2025. [Online]. Available: https://github.com/gjoseph92/stackstac

[8] Philipp Rudiger et al., holoviz/holoviews: Version 1.20.2. (Mar. 13, 2025). Zenodo. doi: 10.5281/ZENODO.15019128.

[9] The pandas development team, pandas-dev/pandas: Pandas. (Sep. 20, 2024). Zenodo. doi: 10.5281/ZENODO.13819579.

[10] S. Hoyer et al., xarray. (Mar. 31, 2025). Zenodo. doi: 10.5281/ZENODO.15110615.

[11] M. Mohr and P. Varner, stac-extensions/sentinel-2. (Aug. 28, 2024). GitHub. Accessed: May 13, 2025. [Online]. Available: https://github.com/stac-extensions/sentinel-2

luke.mcquade@plus.ac.at