

Cross-Mission Validation of Cloud Masking for Optical Remote Sensing Missions Using Sentinel-5P Cloud Products

Efraín Padilla[1], Pertiwi A.[1], De Los Reyes R.[1], Alonso K.[2], Seo S.[1], Loyola D.[1], Storch T.[1]

1 - German Aerospace Center (DLR), Germany

2 - European Space Agency (ESA), Italy

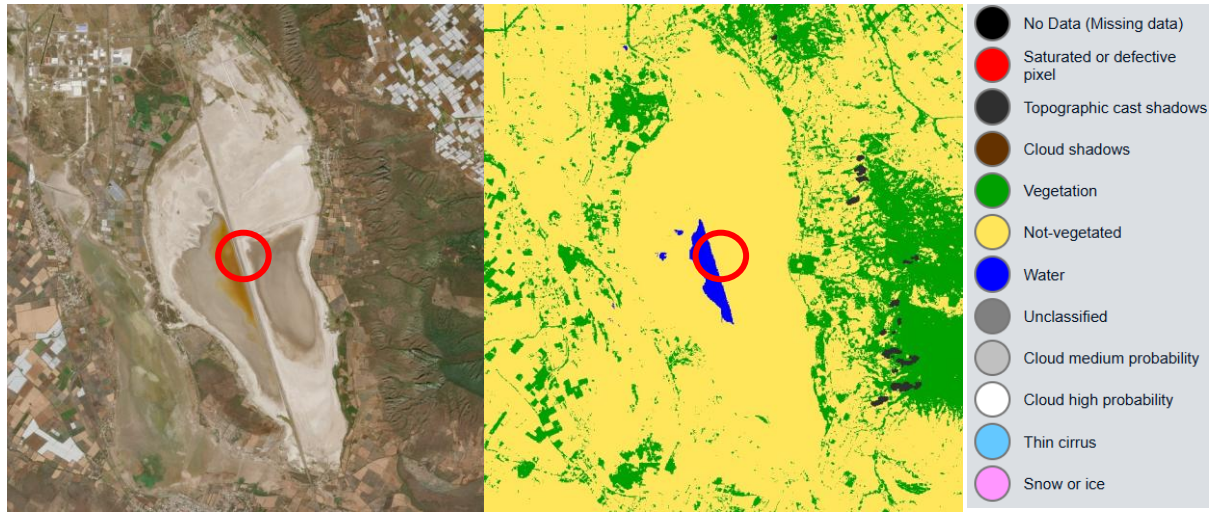
7th Sentinel-2 Validation Team Meeting

13 – 15 October 2025 | ESA – ESRIN | Frascati (RM), Italy

DLR

Introduction

- Validation of masking algorithms requires precise reference data ("ground truth").
- Currently, validations depend on manual annotations, which are labor-intensive and require expert knowledge for each new dataset.

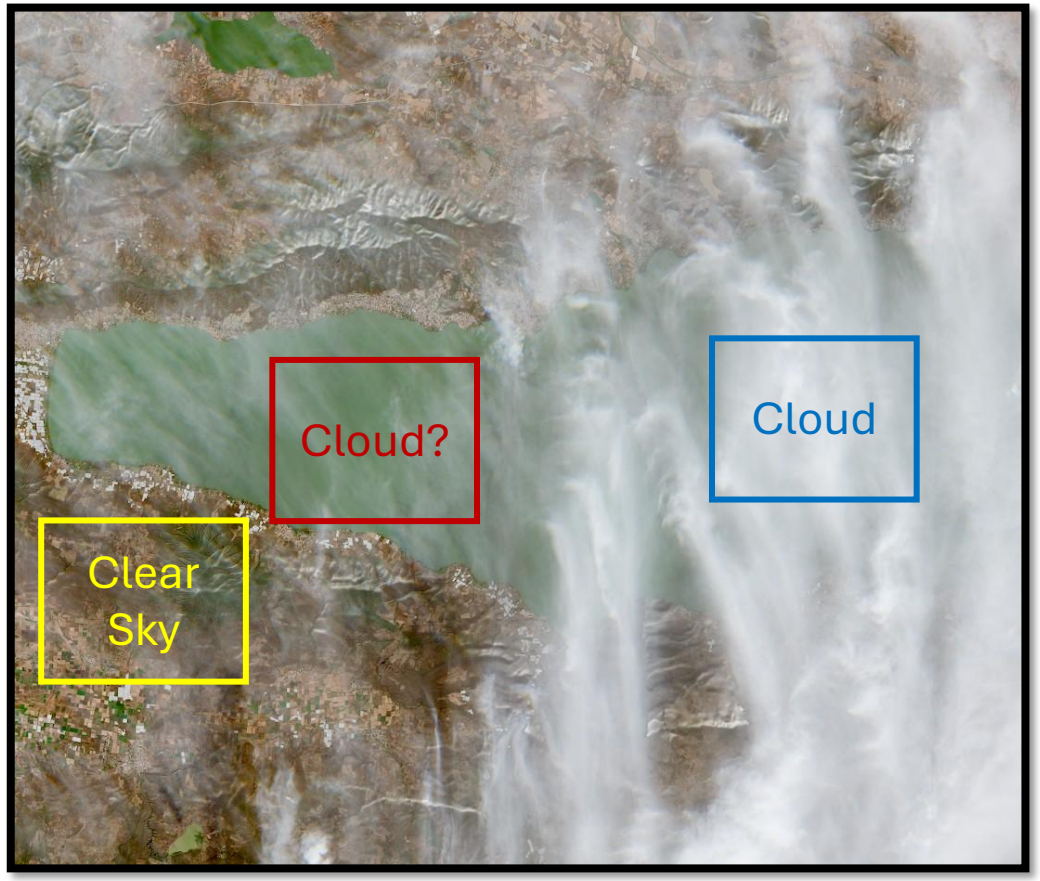


Sentinel-2 & SCL. San Marcos Lagoon, Mexico. 2022-12-13



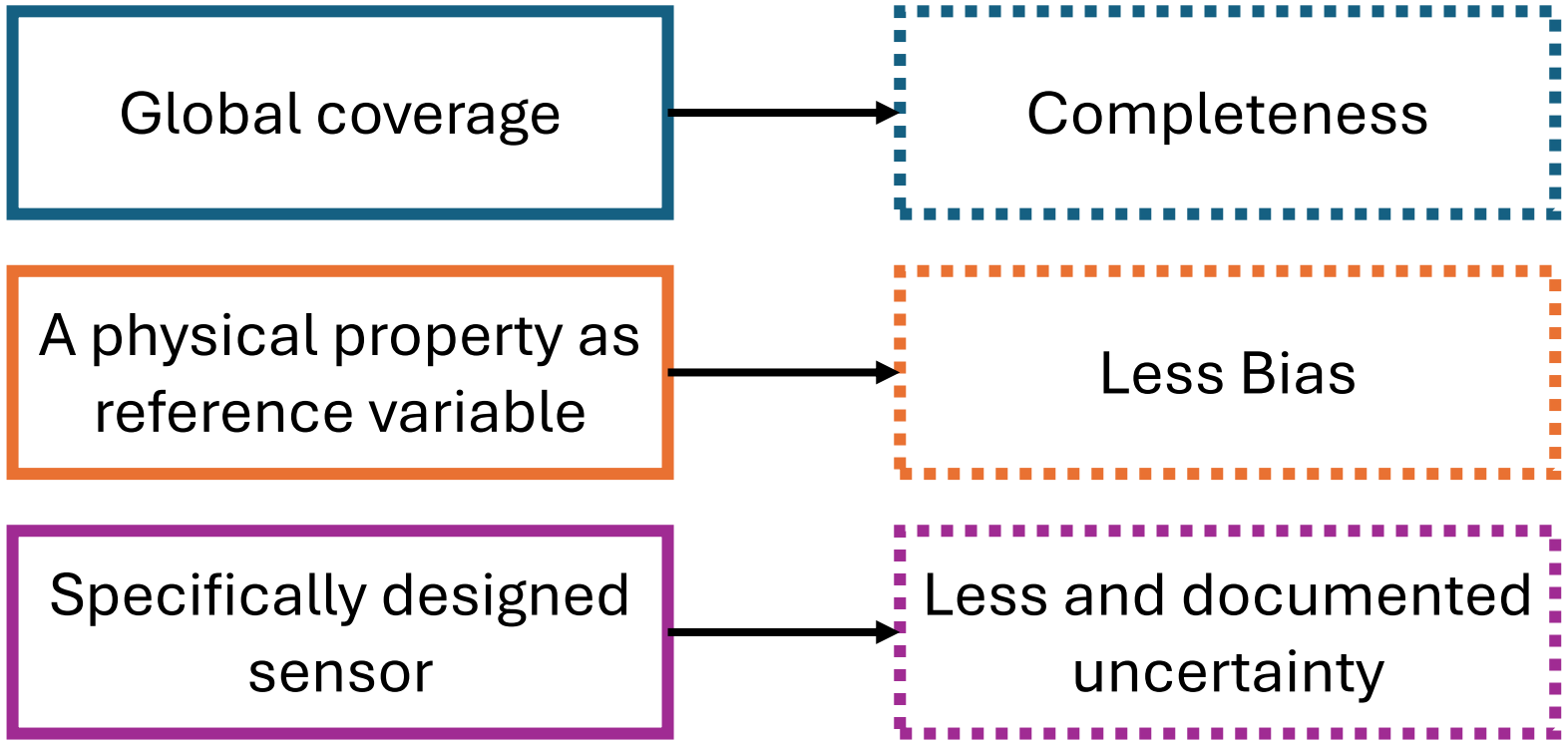
Introduction

- Additionally, class definitions often causes overlaps in annotations between classes.
- Q: How transparent a cloud should be to be flagged as a cloud?
- A: Could depend on the application.



Sentinel-2 RGB. Chapala Lake, Mexico. 2022-12-18.

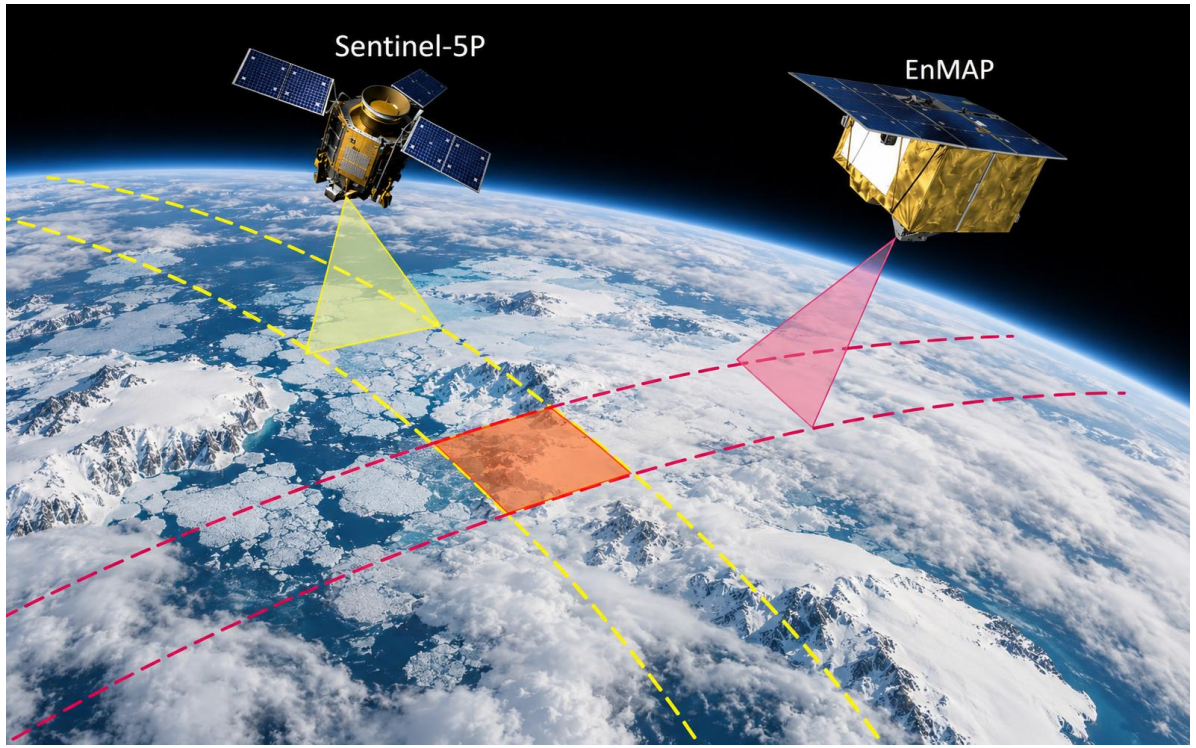
Characteristics of Data Source for Validation



Methodology

- A validation methodology leveraging physically based, globally available reference data to assess masking products, using matching overpasses between independent remote sensing missions.

PhD. Thesis →



Artistic illustration of a matching overpass between EnMAP and Sentinel-5P (not to scale). Generated with ChatGPT (OpenAI) for illustration.

Missions & Products

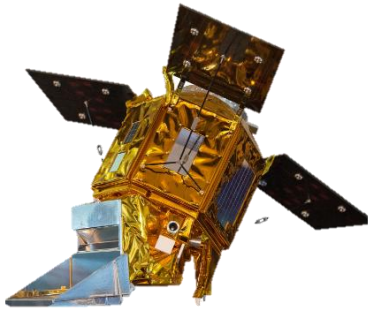
- Multispectral: Sentinel-2
  - **Scene Classification Map (SCL)**
  - 13 spectral bands (20m for SCL)
  - 290km swath
  - Revisit time of 5 days



- Hyperspectral: EnMAP
  - **Cloud Mask from PACO**
  - 224 spectral bands at 30m
  - 30km swath
  - Revisit time of 27 days

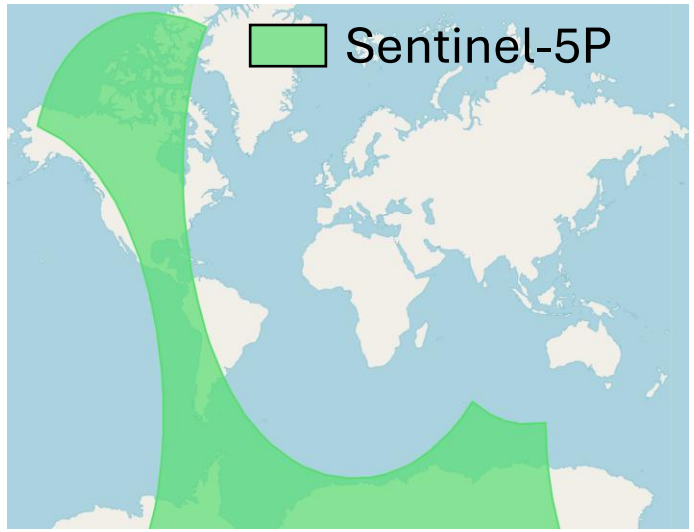


- **Validation**: TROPOMI sensor onboard Sentinel-5P
  - **Cloud Fraction Product** at 5.5 × 3.5 km
  - 2600 km swath
  - Full daily surface coverage

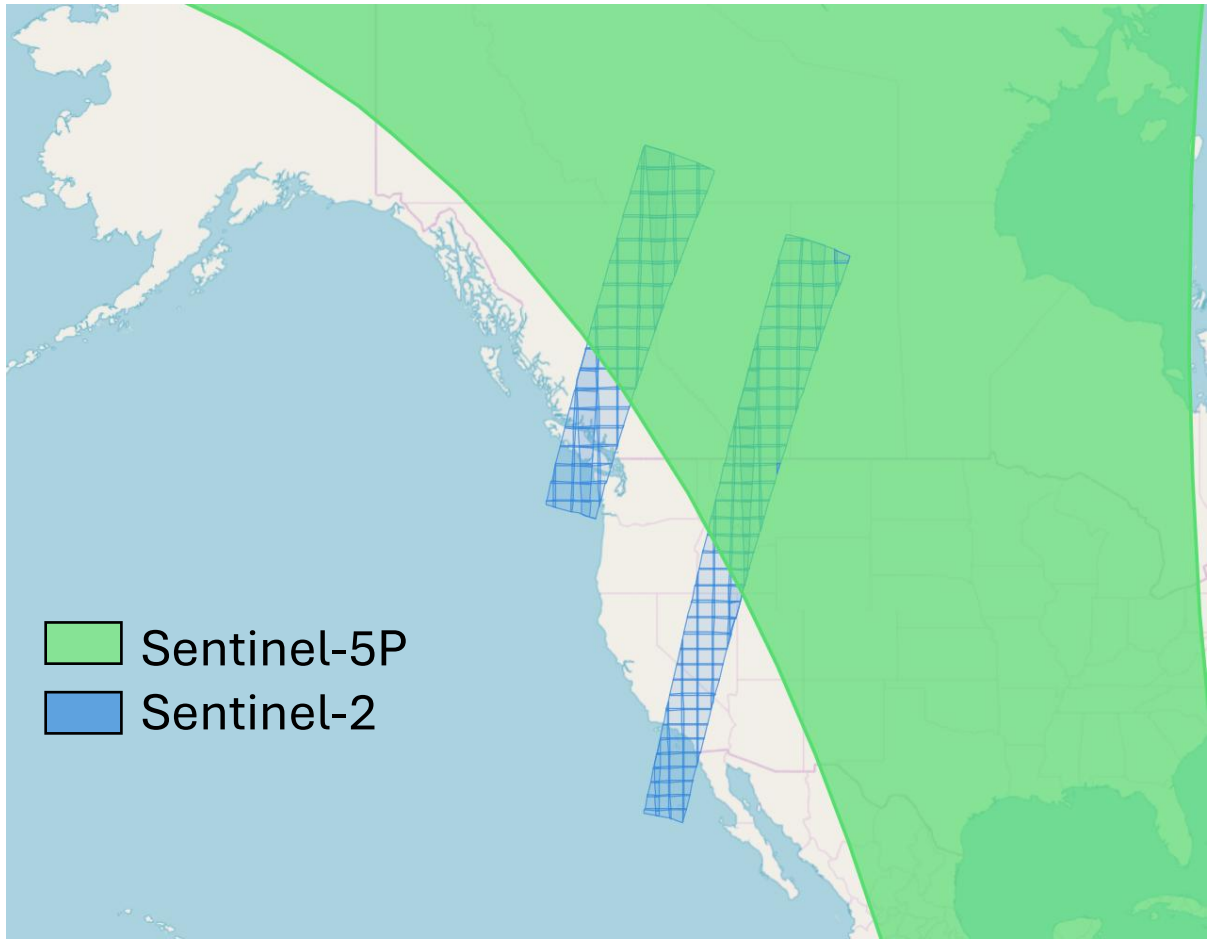


Matching Overpasses

- The search is conducted across the entire archive of both missions searching for matching in location and capturing time ($\Delta t \leq T_{max}$).

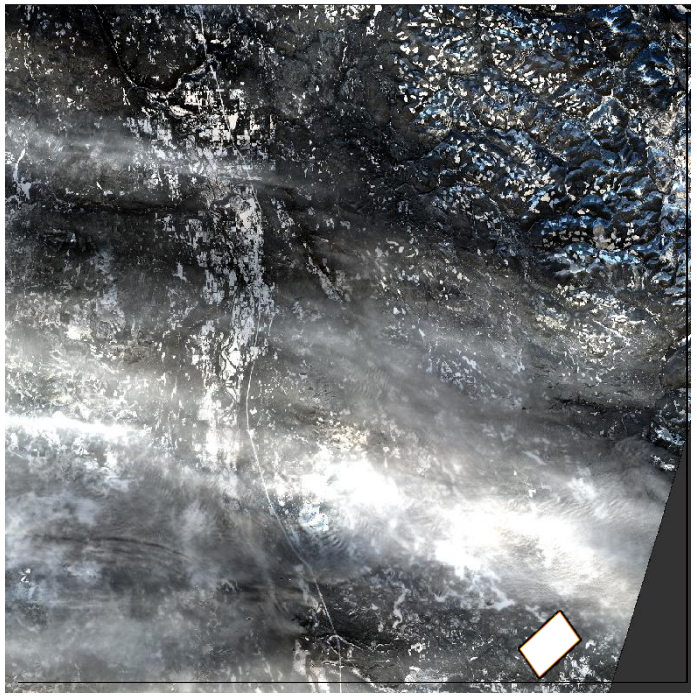


Sentinel-5P, 2025-01-02 18:13:30 - 19:55:00

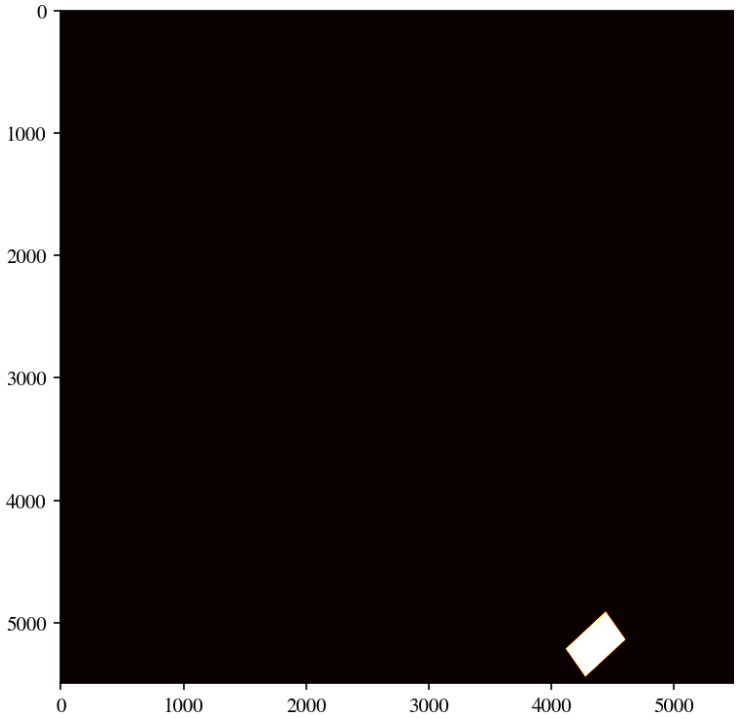


Matching Overpasses

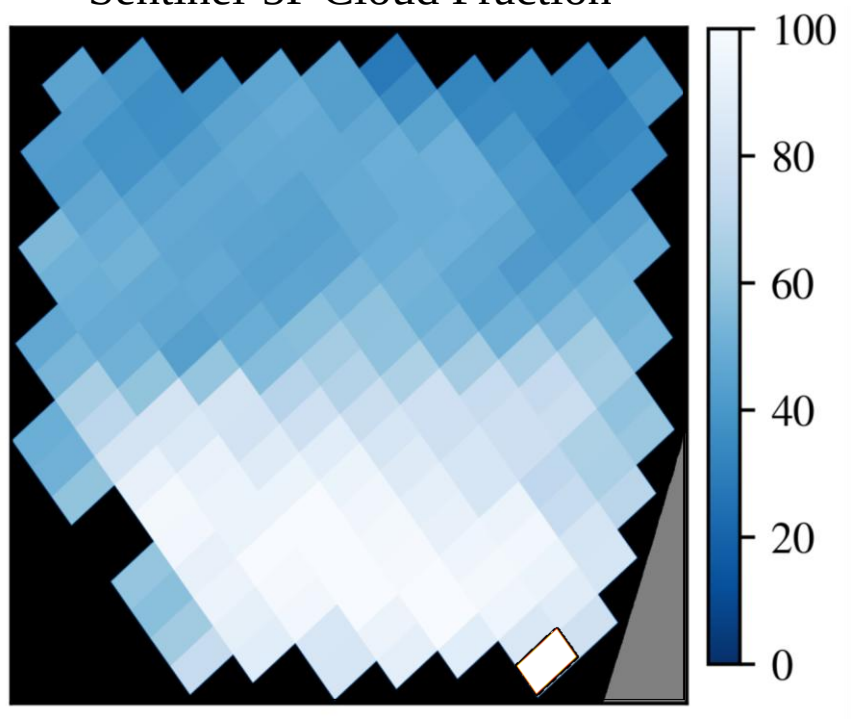
Sentinel-2 Quick-look



Sentinel-5P Pixels Mask



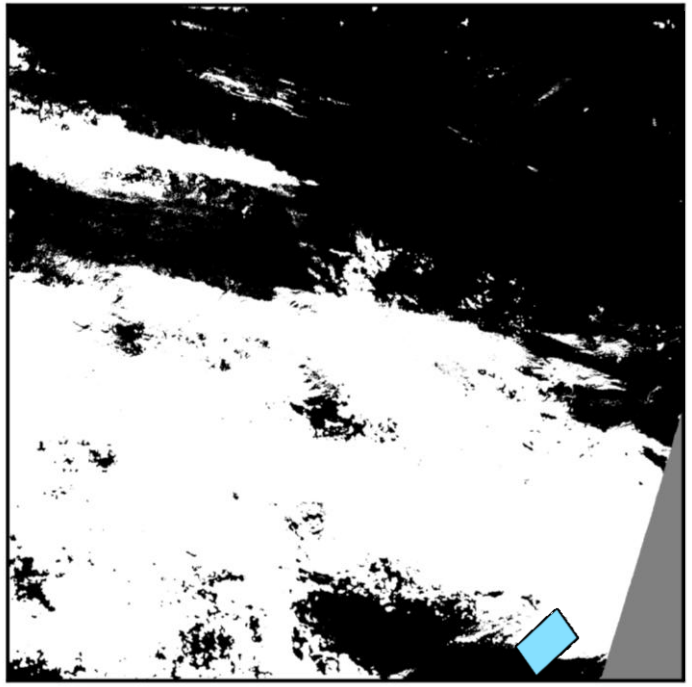
Sentinel-5P Cloud Fraction



Comparison of Sentinel-2 image and Reference Cloud Fraction

From Cloud Mask to Geometric Cloud Fraction

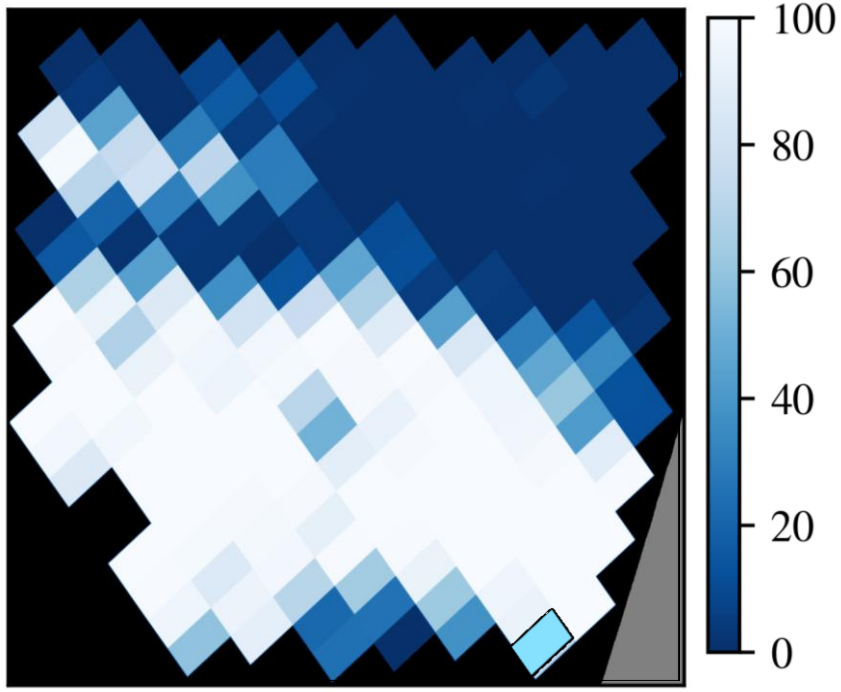
SCL Cloud Mask



CF_i: *Cloud Fraction of* i^{th}*pixel*

$$CF_i = \frac{Cloud\ pixels}{Total\ pixels}(\%)$$

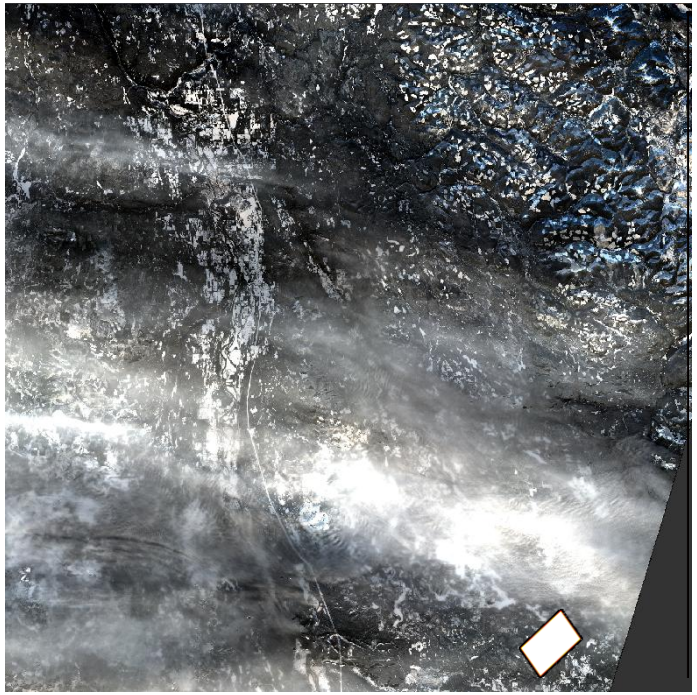
Sentinel-2 Cloud Fraction



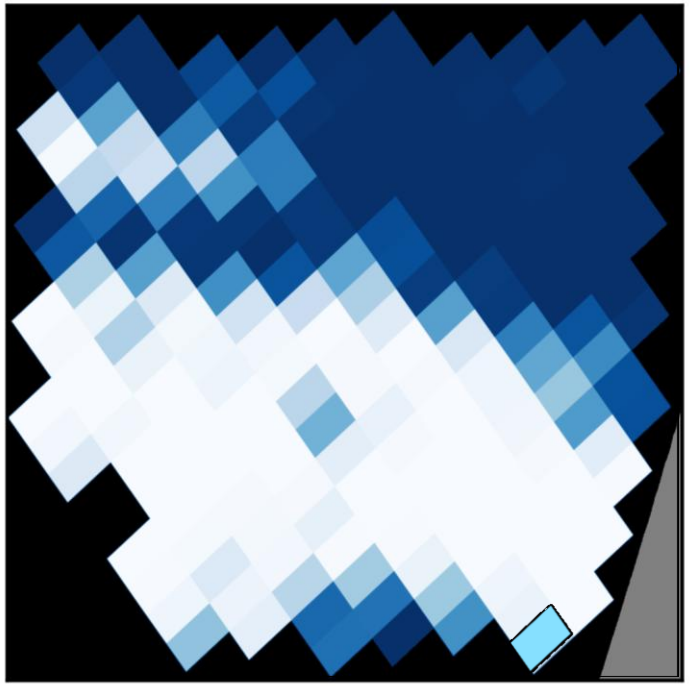
Comparison of Sentinel-2 image and Reference Cloud Fraction

Assessment

Sentinel-2 RGB

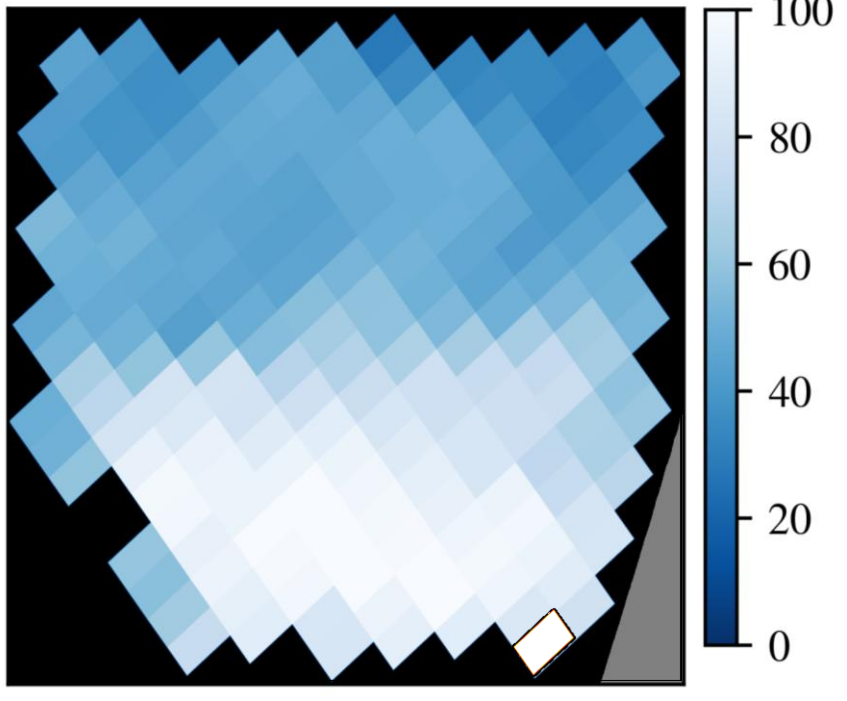


Sentinel-2 Cloud Fraction
$\hat{y}$: **Predicted CF**



Assessment

Sentinel-5P Cloud Fraction
y: **Reference CF**



Comparison of Sentinel-2 image and Reference Cloud Fraction

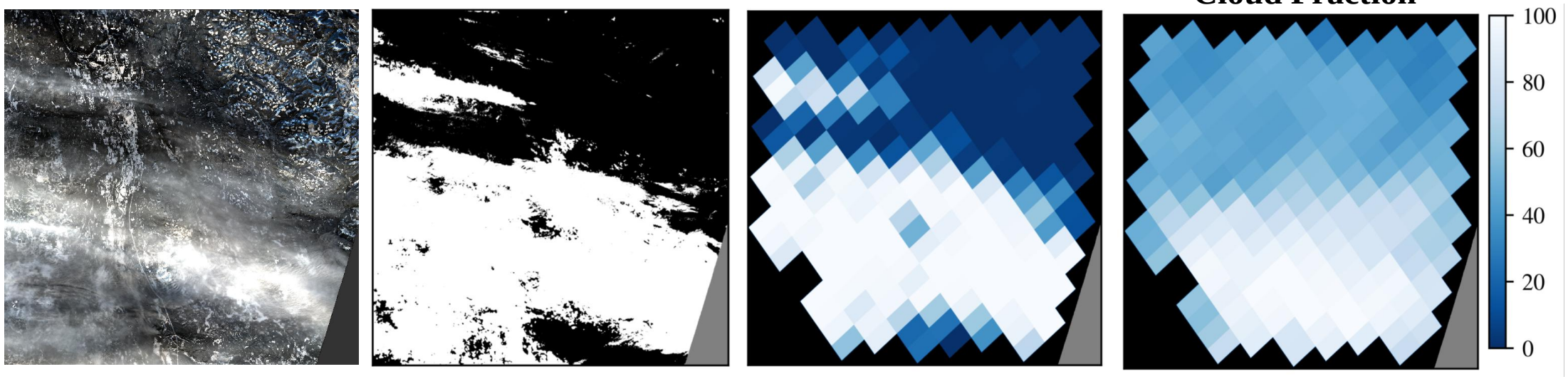
Geometric vs Radiometric Cloud Fraction

Sentinel-2 RGB

-High prob. clouds
-Medium prob. clouds
SCL Cloud Mask

Sentinel-2
Geometric
Cloud Fraction

Sentinel-5P
Radiometric
Cloud Fraction

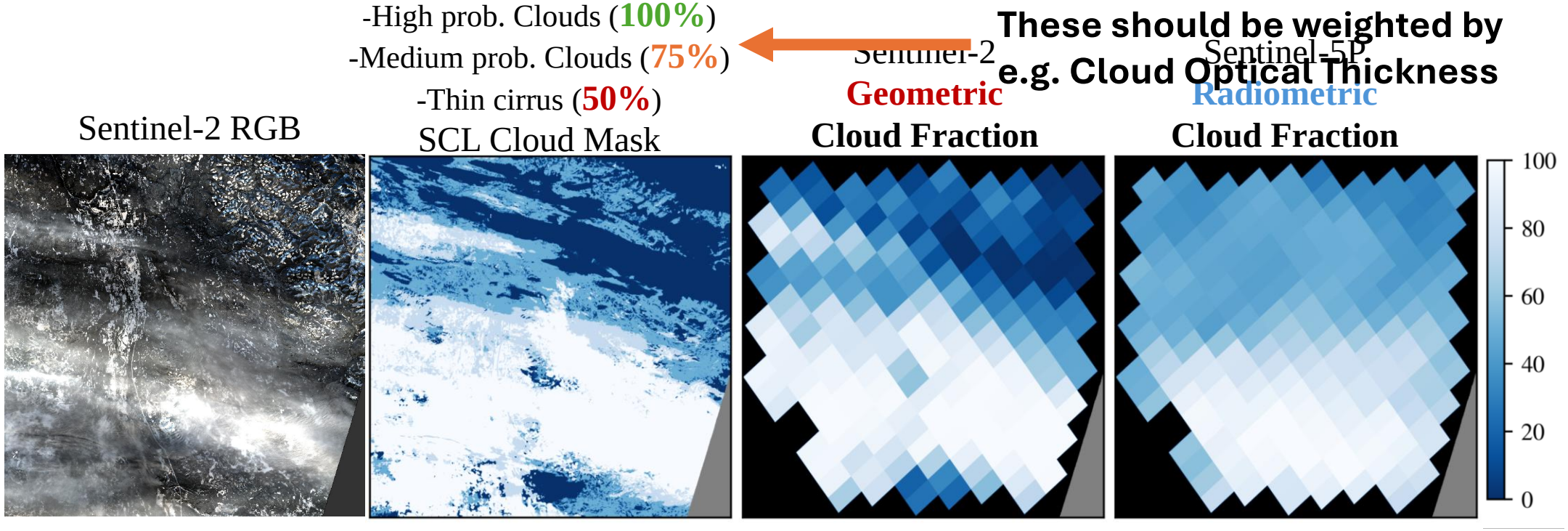


Comparison of Sentinel-2 image and Reference Cloud Fraction

Geometric vs Radiometric Cloud Fraction

-High prob. Clouds (**100%**)
-Medium prob. Clouds (**75%**)
-Thin cirrus (**50%**)

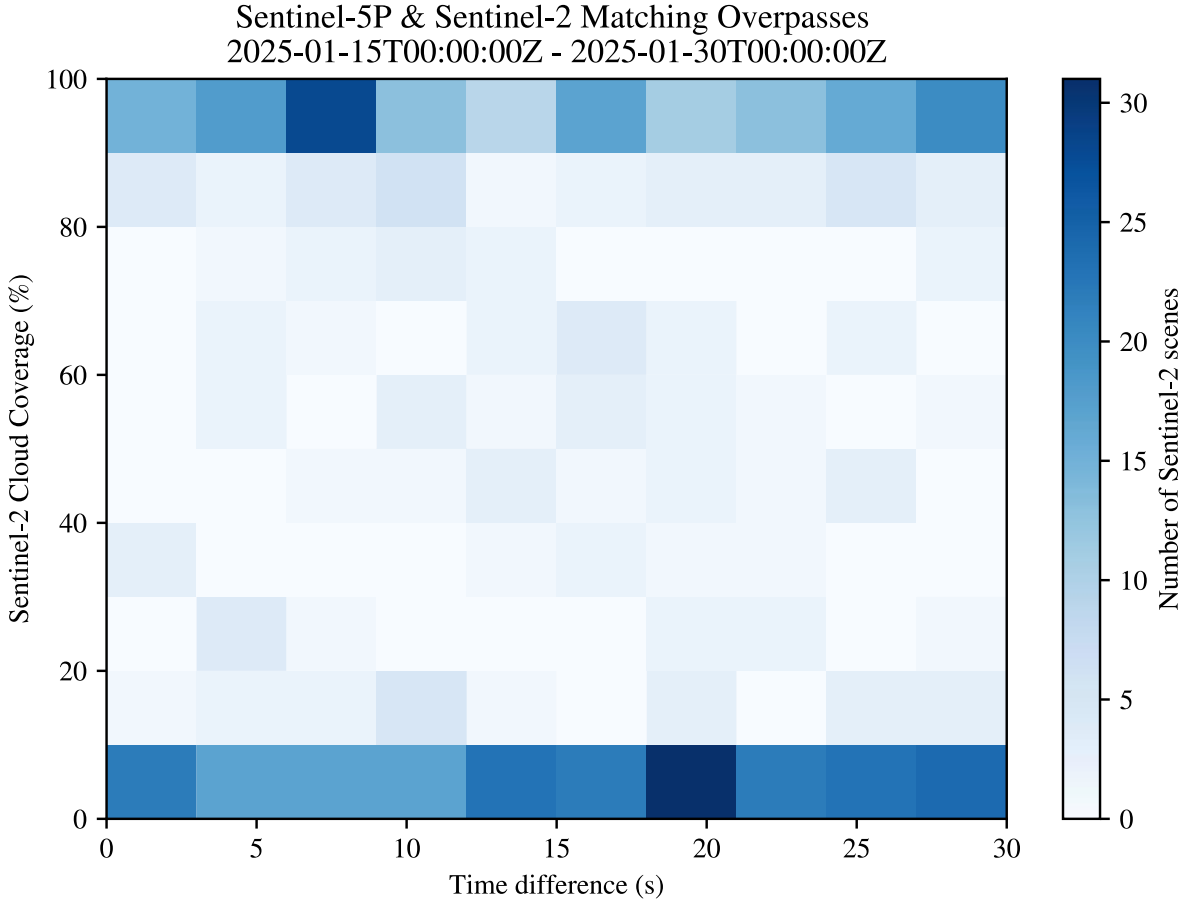
These should be weighted by e.g. Cloud Optical Thickness



Comparison of Sentinel-2 image and Reference Cloud Fraction

Cloud Displacement

- Clouds are highly **dynamic targets**: **minimizing the time difference** between acquisitions reduces validation uncertainty.
- Apparent cloud displacement in imagery also depends on **cloud altitude** and **sensor viewing geometry**.

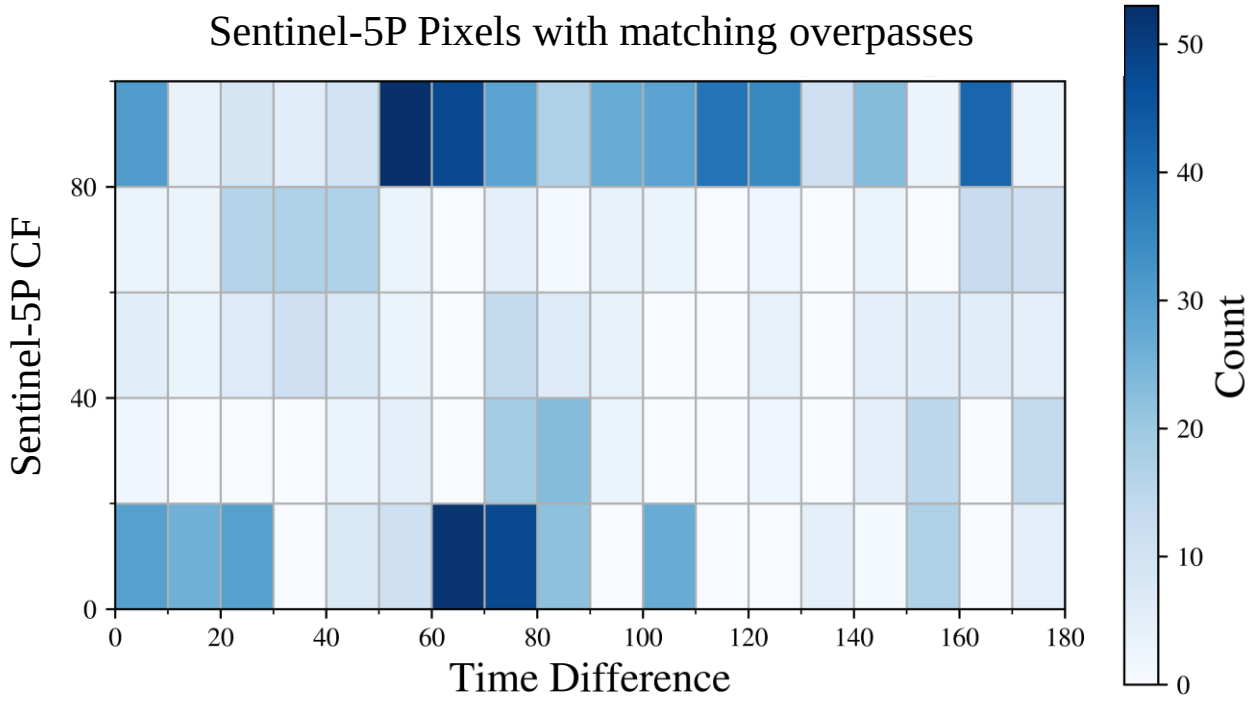


Distribution of matching overpasses by time difference and cloud coverage

EnMAP vs Sentinel-5P



- EnMAP (processed with PACO)
- **Time range:**
  - May 2023 to September 2024.
- **Matching overpasses:**
  - Sentinel-5P: 15 overpasses
  - EnMAP: 105 scenes.
- 982 Pixels

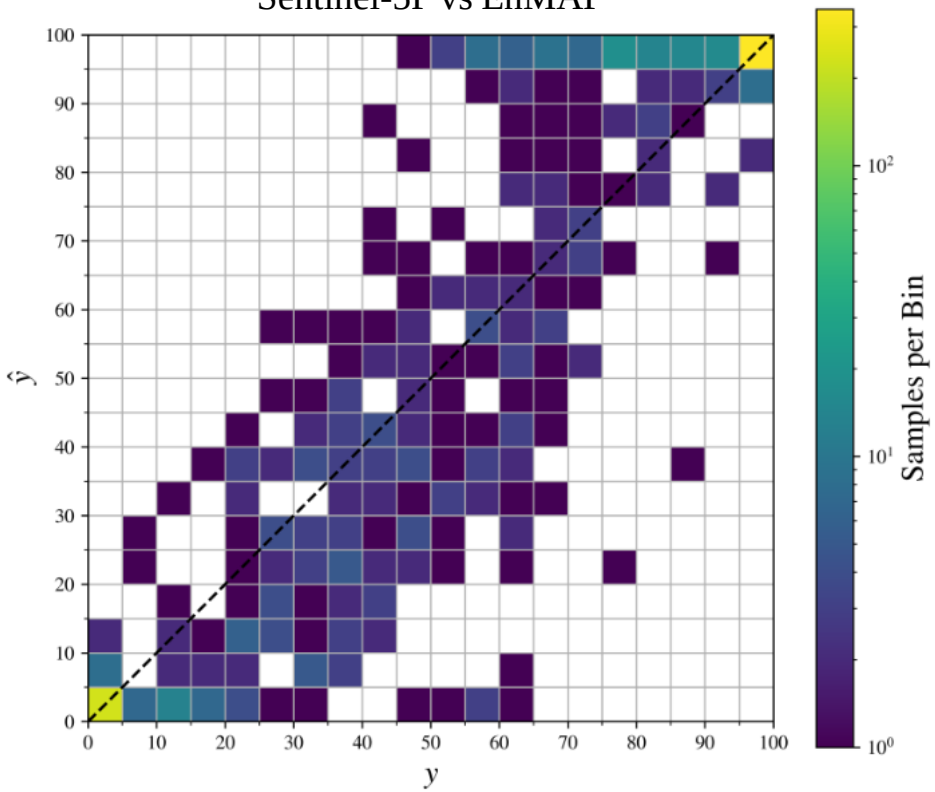


Distribution of matching overpasses by time difference and cloud fraction

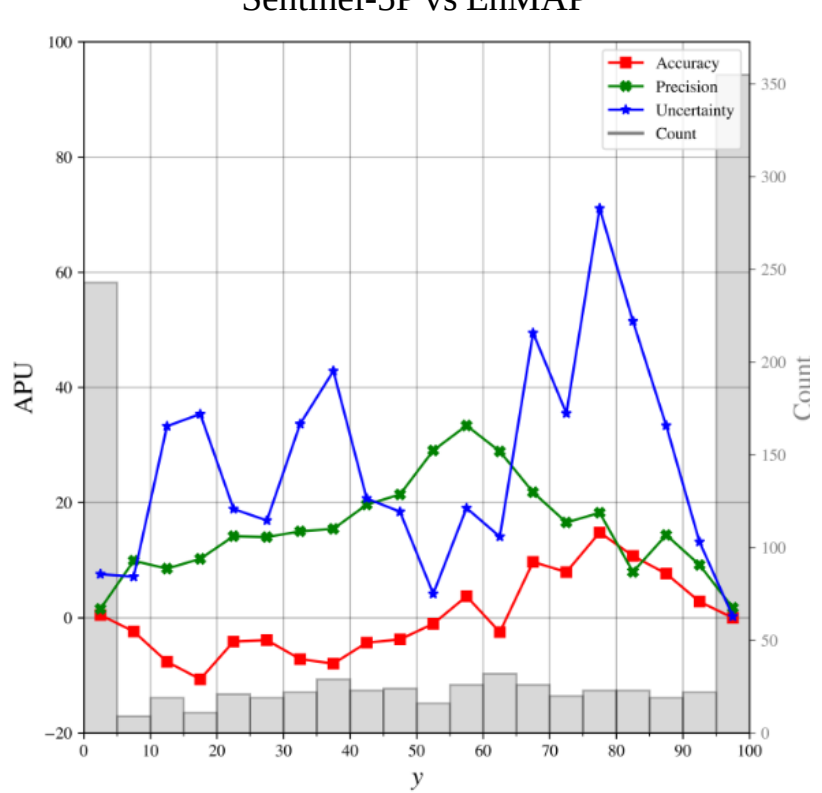
EnMAP vs Sentinel-5P



Not weighted



Regression plot with 2D Histogram. PACO vs Sentinel-5P. $R^2 = 0.9$.



APU Plot. EnMAP vs Sentinel-5P.

EnMAP vs Sentinel-5P



- EnMAP (processed with PACO)

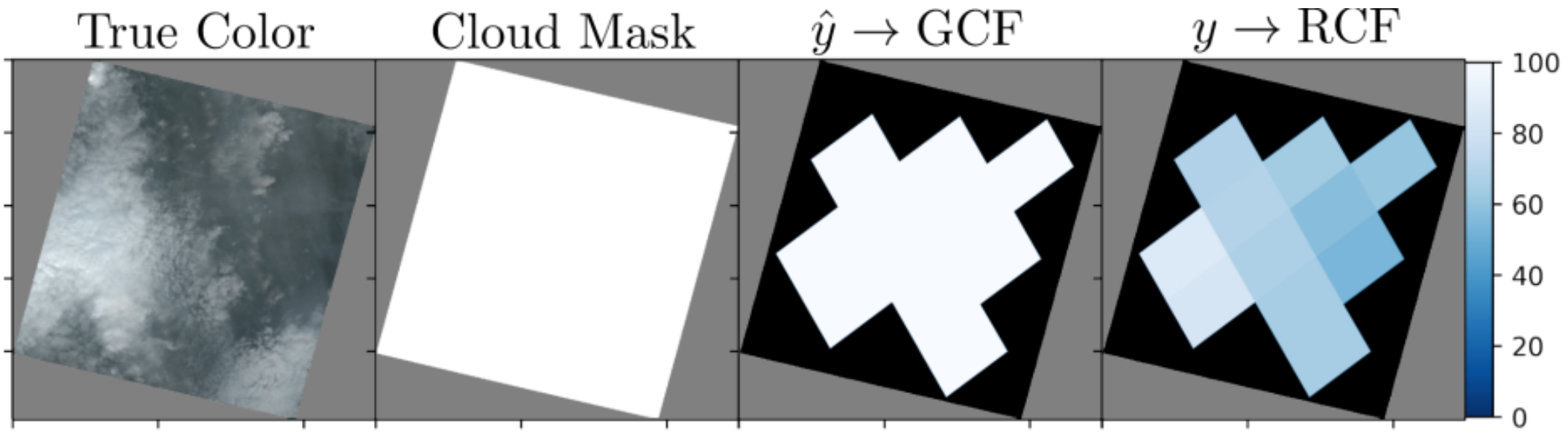
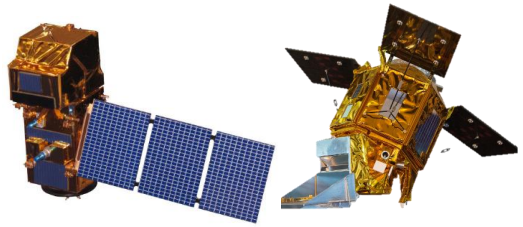
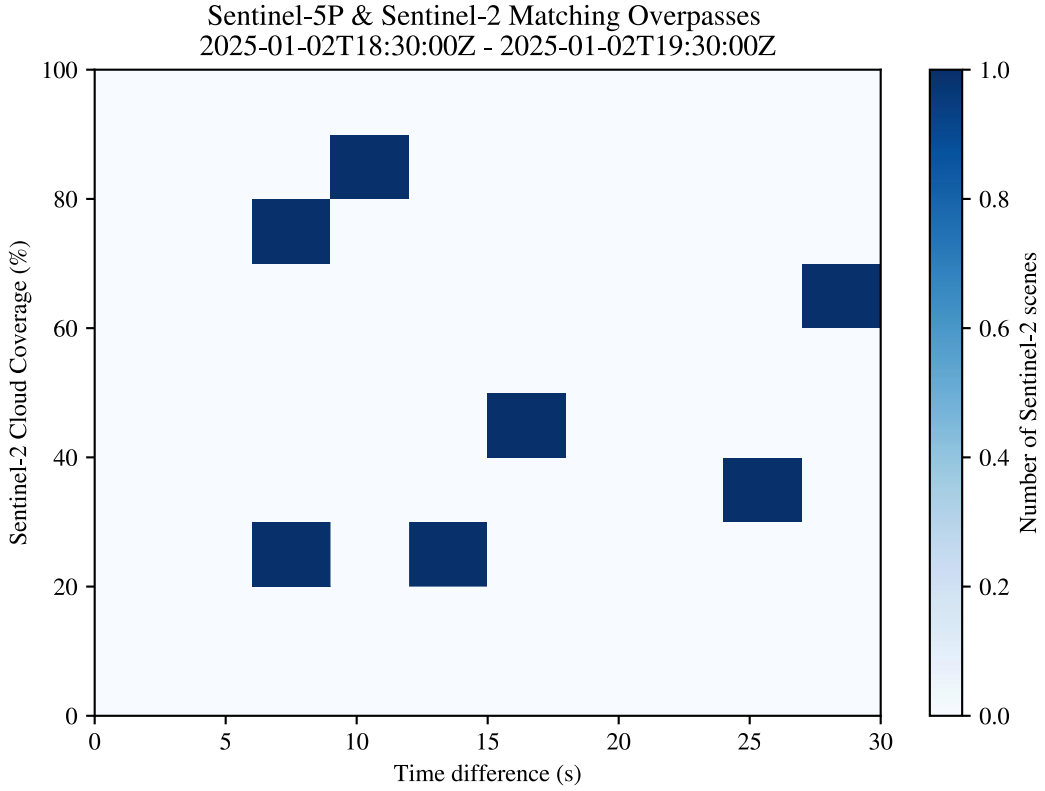


Image: EnMAP DT38193, tile-003. Captured at 2023/08/30. Time difference between captures of 25 s.

Sentinel-2 vs Sentinel-5P

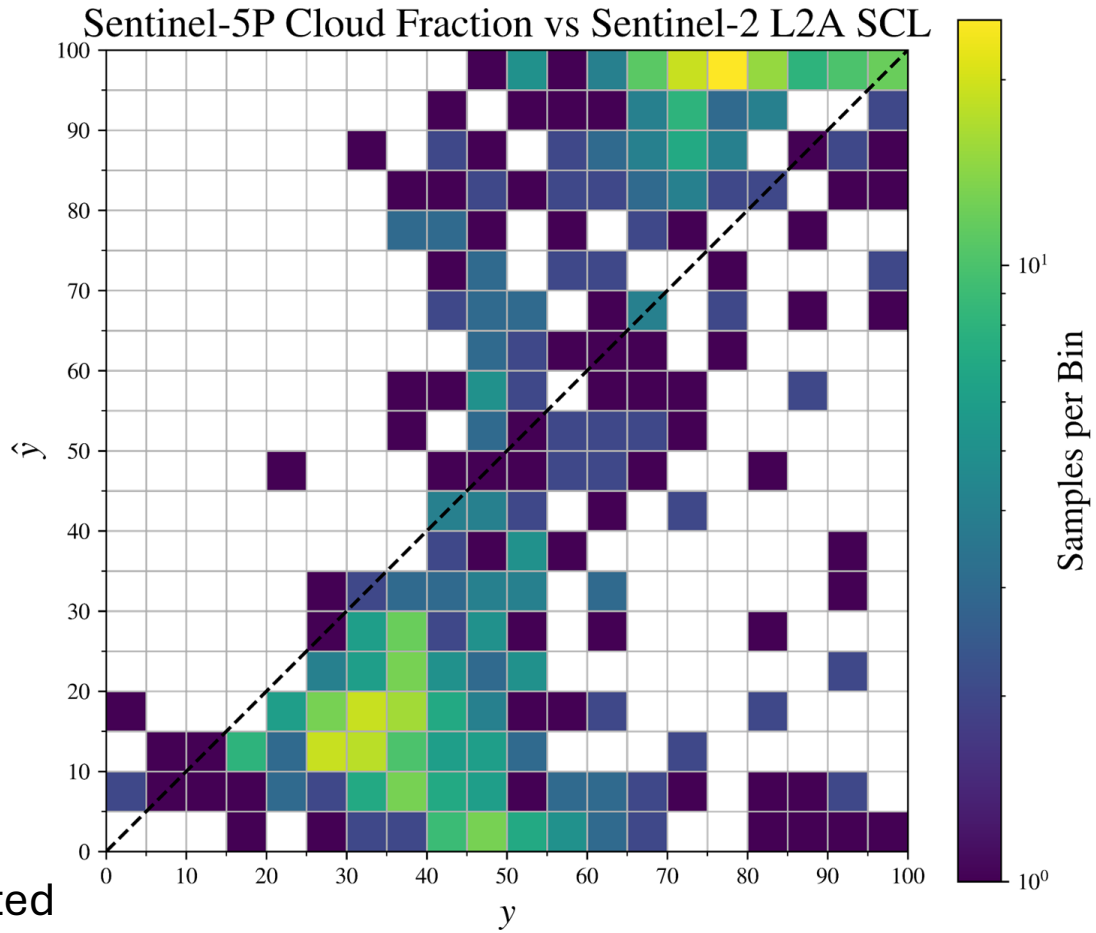


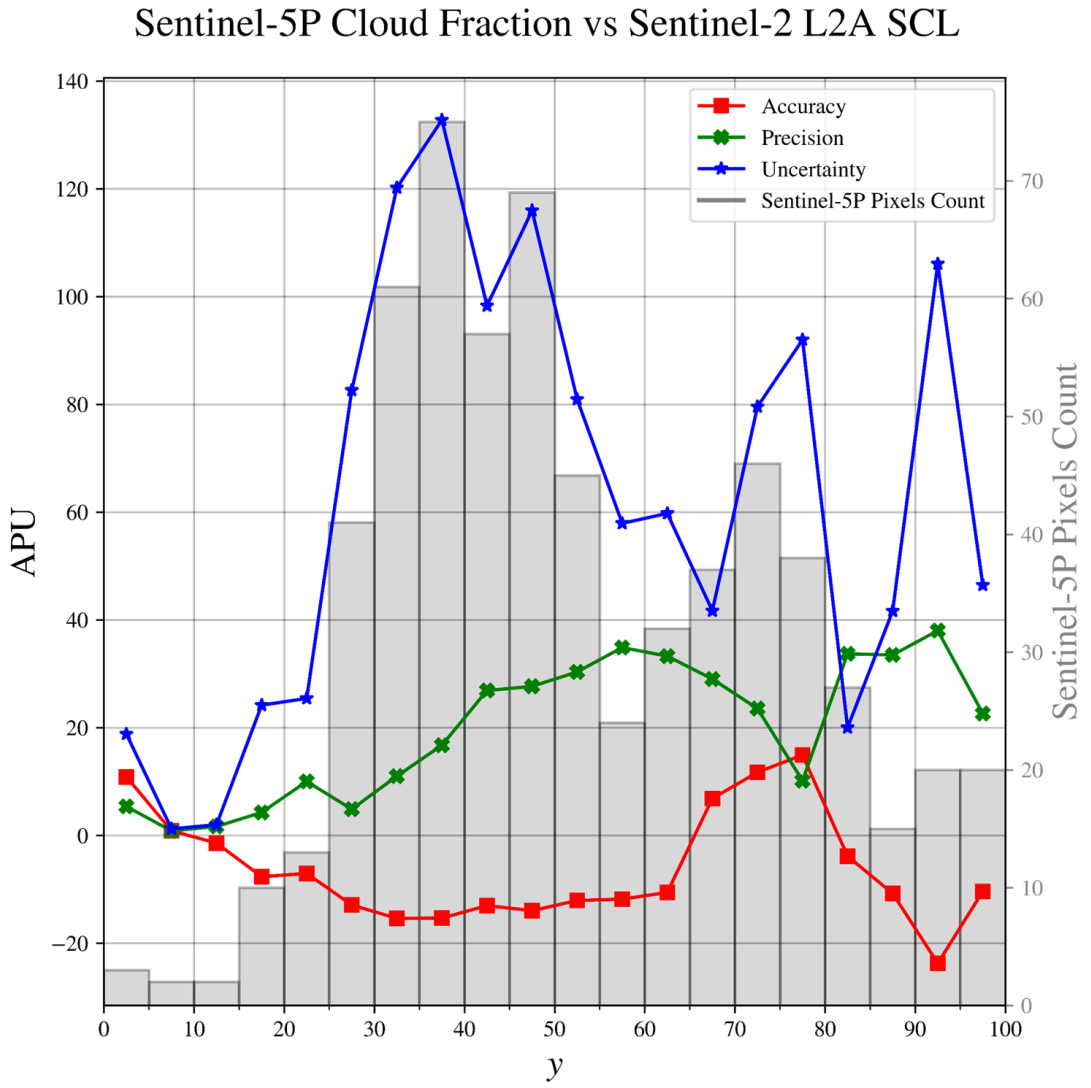
- **Time range:**
  - 2025-01-02 18:13 – 19:55
- **Matching overpasses:**
  - Sentinel-5P: 1 scene
  - Sentinel-2: 7 scenes
- 637 Pixels



Distribution of matching overpasses by time difference and cloud fraction

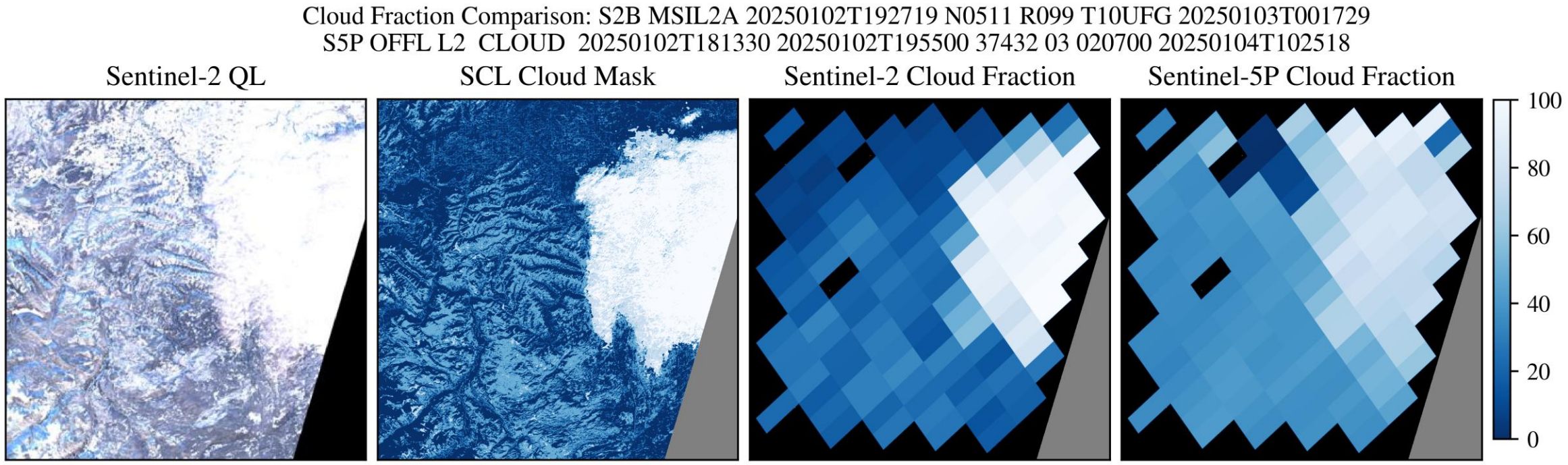
Sentinel-2 vs Sentinel-5P





Weighted

Sentinel-2 vs Sentinel-5P



Conclusions

- The methodology provides an **automatic way to retrieve reference data for validation globally.**
- It can be applied **to any pair of missions to validate any kind of mask** with matching overpasses available.
- The reference data source should be **more sensitive** than the mission being validated.
- The transformation to the reference physical variable must be properly **calibrated** (in this study: from SCL cloud mask → cloud fraction).
- Large data volumes (from missions with high acquisition frequency and wide swaths) further enhance the effectiveness of this methodology.

PROGRAMME OF THE EUROPEAN UNION

co-funded with esa



Efraín Alberto Padilla Zepeda

Postdoc at DLR | PhD in Electrical Engineering | Remote Sensing & Deep Learning Researcher

Thank you!

Efraín Padilla[1], Pertiwi A.[1], De Los Reyes R.[1], Alonso K.[2], Seo S.[1], Loyola D.[1], Storch T.[1]

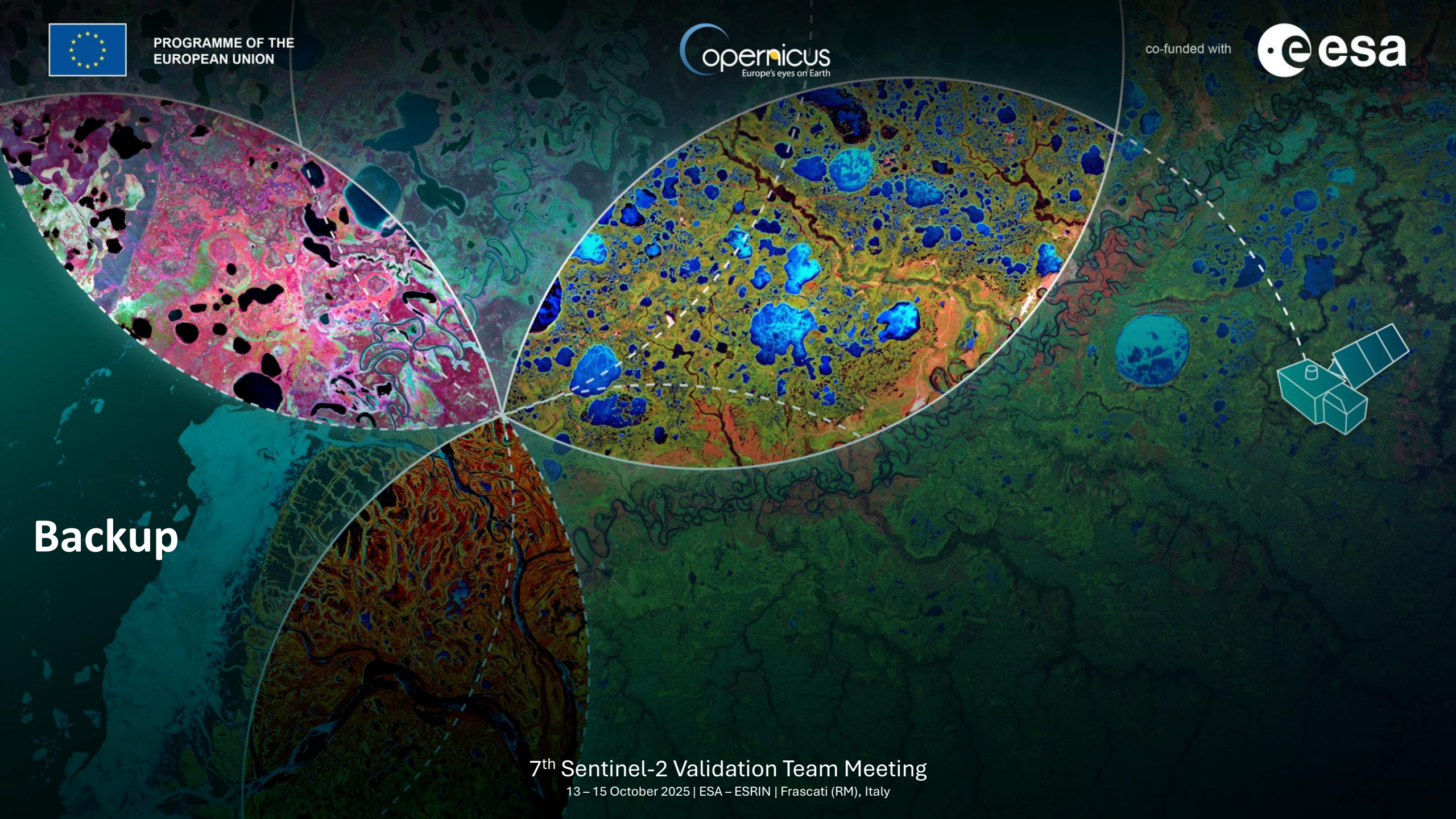
1 - German Aerospace Center (DLR), Germany

2 - European Space Agency (ESA), Italy

7th Sentinel-2 Validation Team Meeting

13 – 15 October 2025 | ESA – ESRIN | Frascati (RM), Italy

DLR

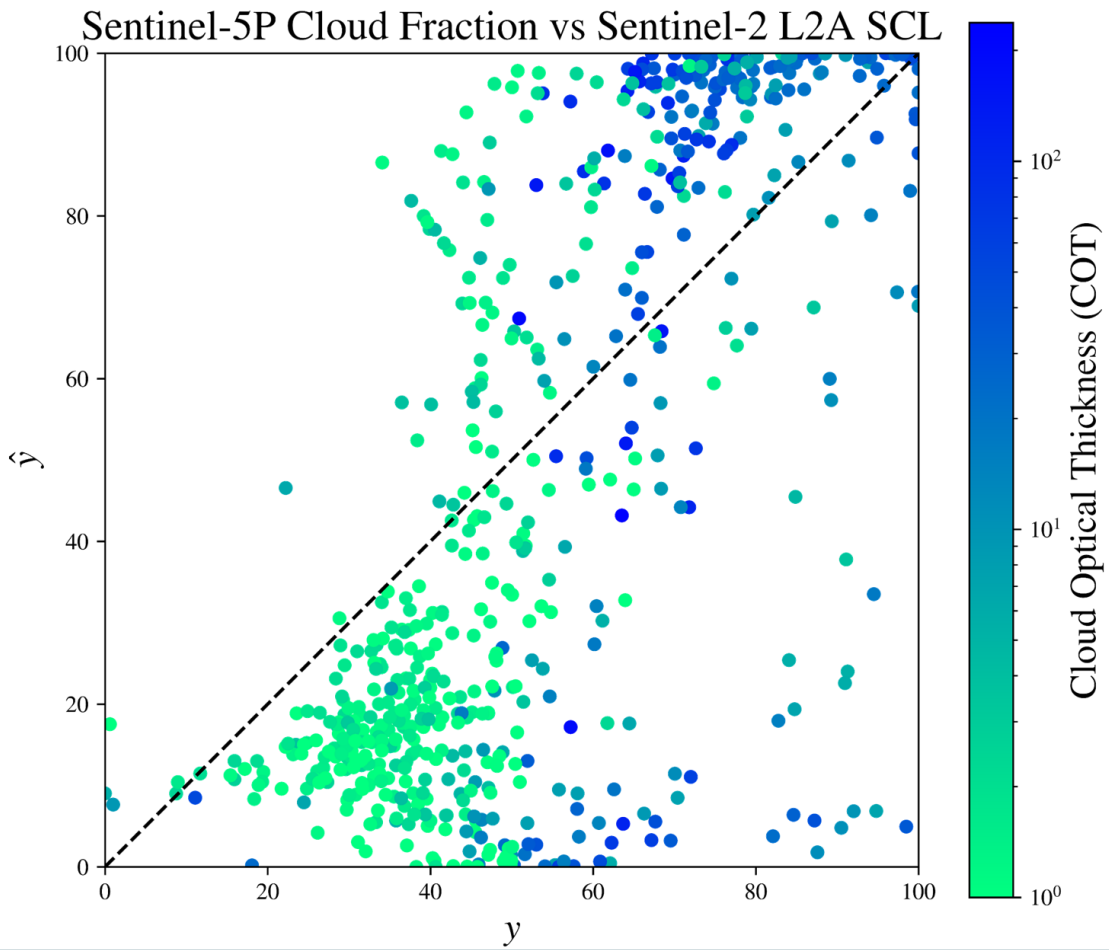


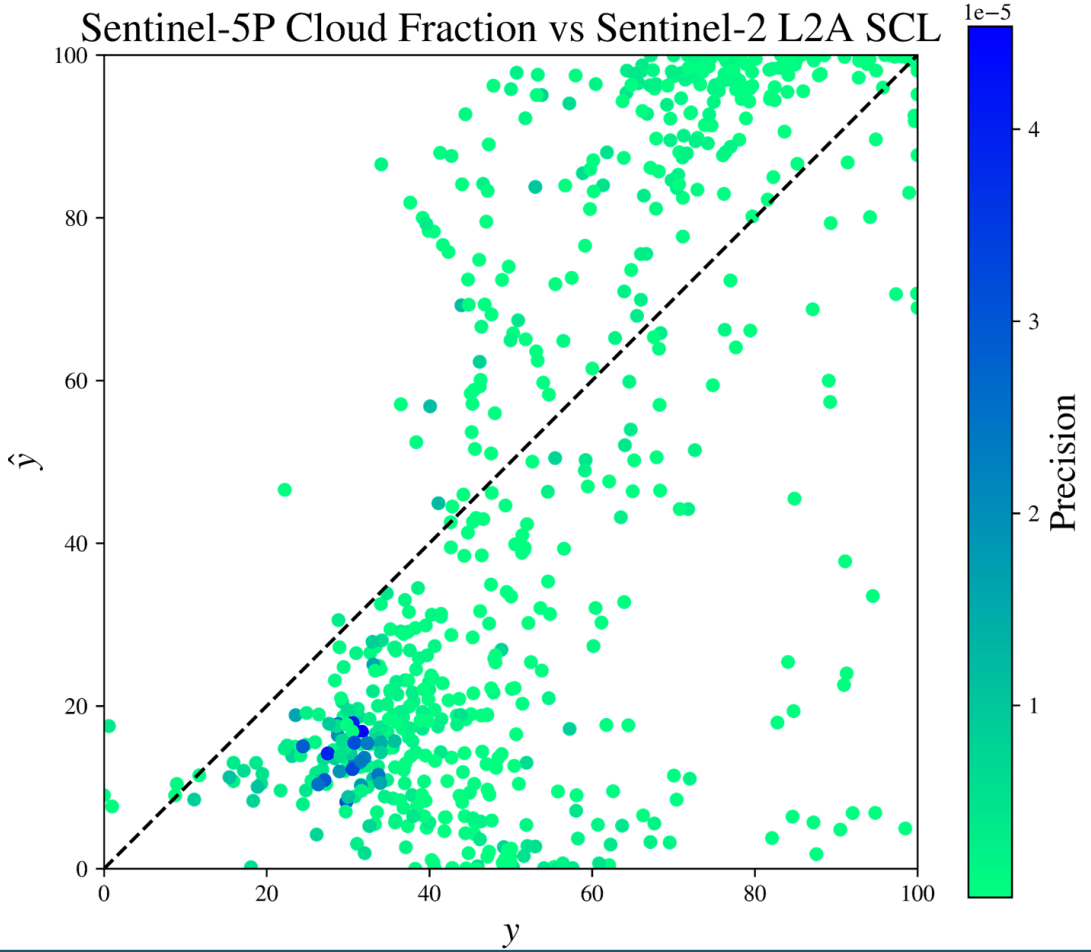
Backup

7th Sentinel-2 Validation Team Meeting

13 – 15 October 2025 | ESA – ESRIN | Frascati (RM), Italy

Sentinel-2 vs Sentinel-5P





APU Plot

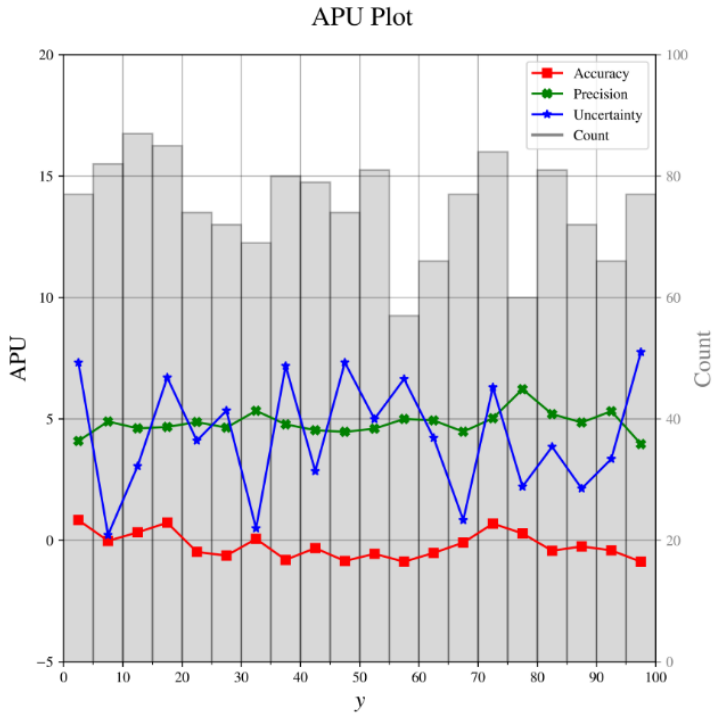
$$\Delta_{y_n} = \hat{y}_n - y_n \quad \hat{y}, y \in \mathbb{R}$$

y = Reference variable,
$\hat{y}$ = Predicted variable,
Δ_{y_n} = Difference between the predicted and reference variables,
n = As a subscript, is the n^{th} sample of N samples.

$$A = \frac{1}{N_b} \sum_{n=1}^{N_b} \Delta_{y_n}$$

$$P = \sqrt{\frac{1}{N_b - 1} \sum_{n=1}^{N_b} (\Delta_{y_n} - A)^2}$$

$$U = \sqrt{\frac{1}{N_b} \sum_{n=1}^{N_b} (\Delta_{y_n})^2}$$



Example of an APU plot showing Accuracy (A), Precision (P), and Uncertainty (U) metrics across B=20 bins using N=1500 samples.