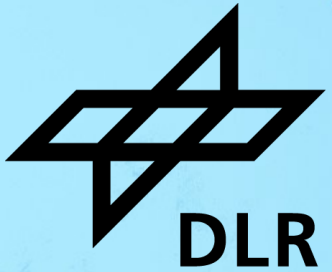
# Total column water vapor for Sentinel-4 and Sentinel-5 algorithm and initial results

Luca Lelli
Temenuzhka Avramova
Diego G. Loyola

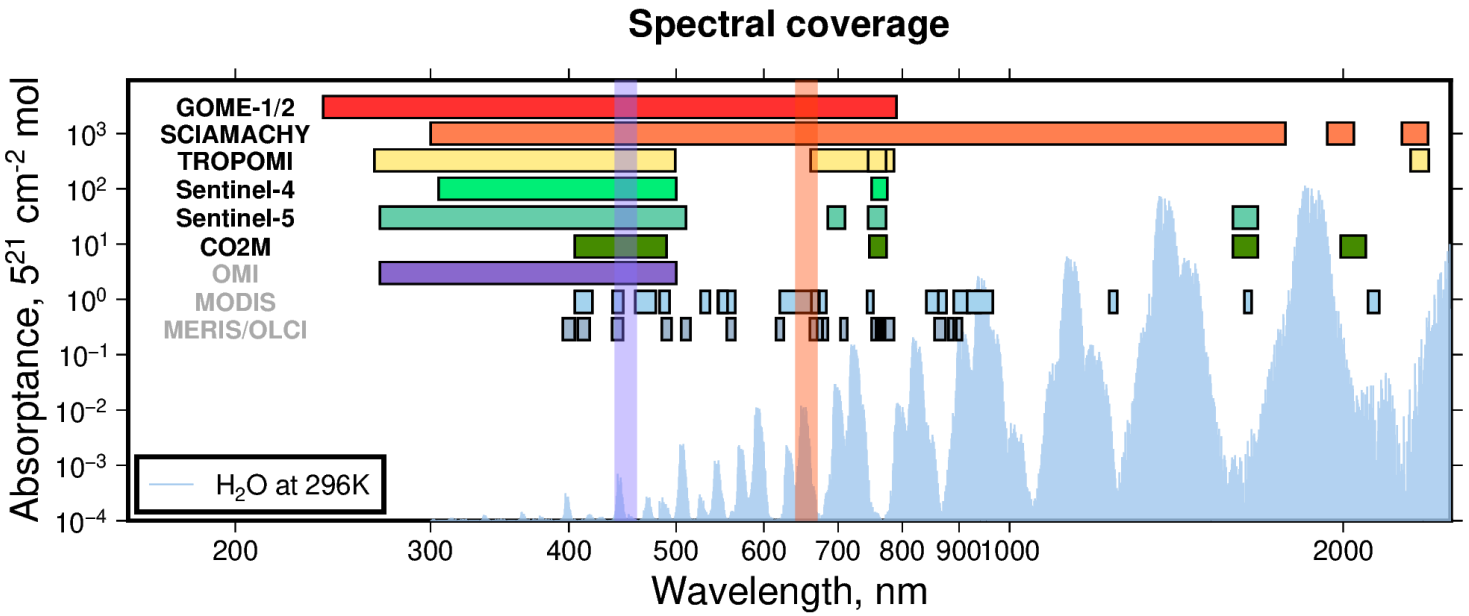


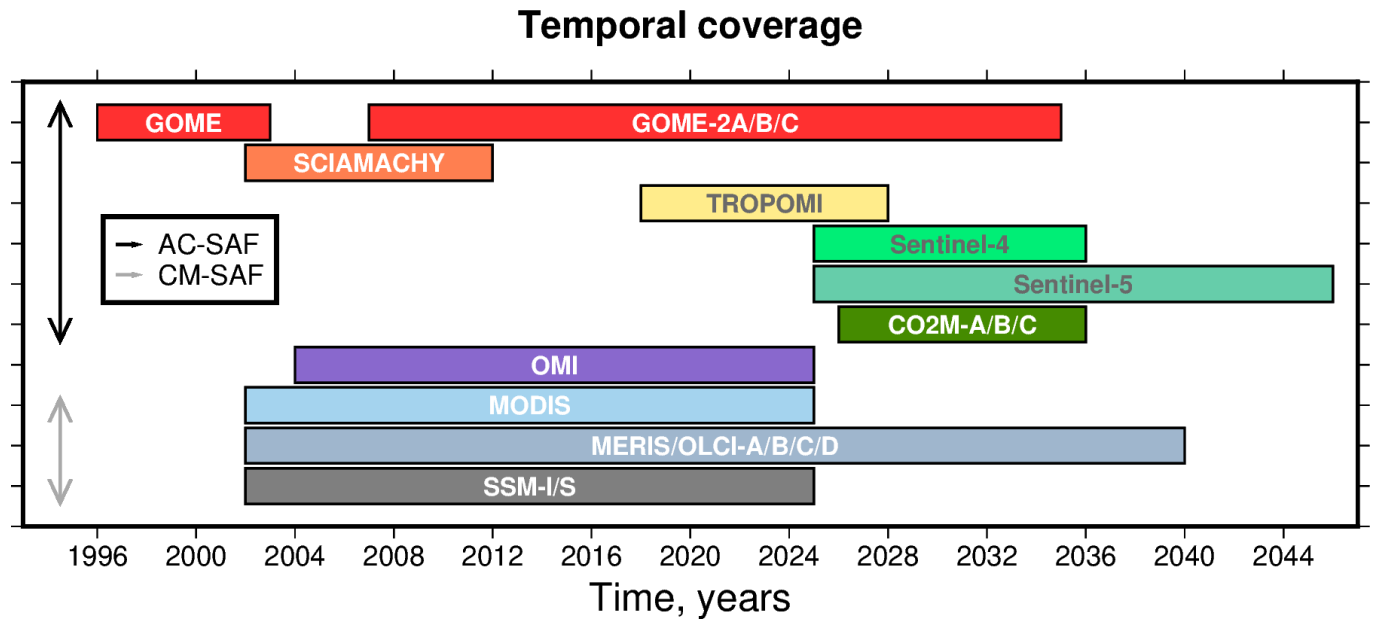



# Outline

1) Motivation
2) Algorithmic updates : DOAS + AMF
3) Full error budget and propagation
4) Sentinel-4 : initial results with CalVal datasets
(March 05-11, July 20-30)
5) Sentinel-5 : initial results (one day: 25-11-2025)
6) Sentinel-4 : possible applications – hygroscopic growth

# Motivation: platforms




The **blue band of water vapor** is the weakest but it is sensed by more sensors and the surface is less reflective in the UV

→ best trade off between sensitivity and coverage over both land and oceans

**Fitting windows** are different from other works

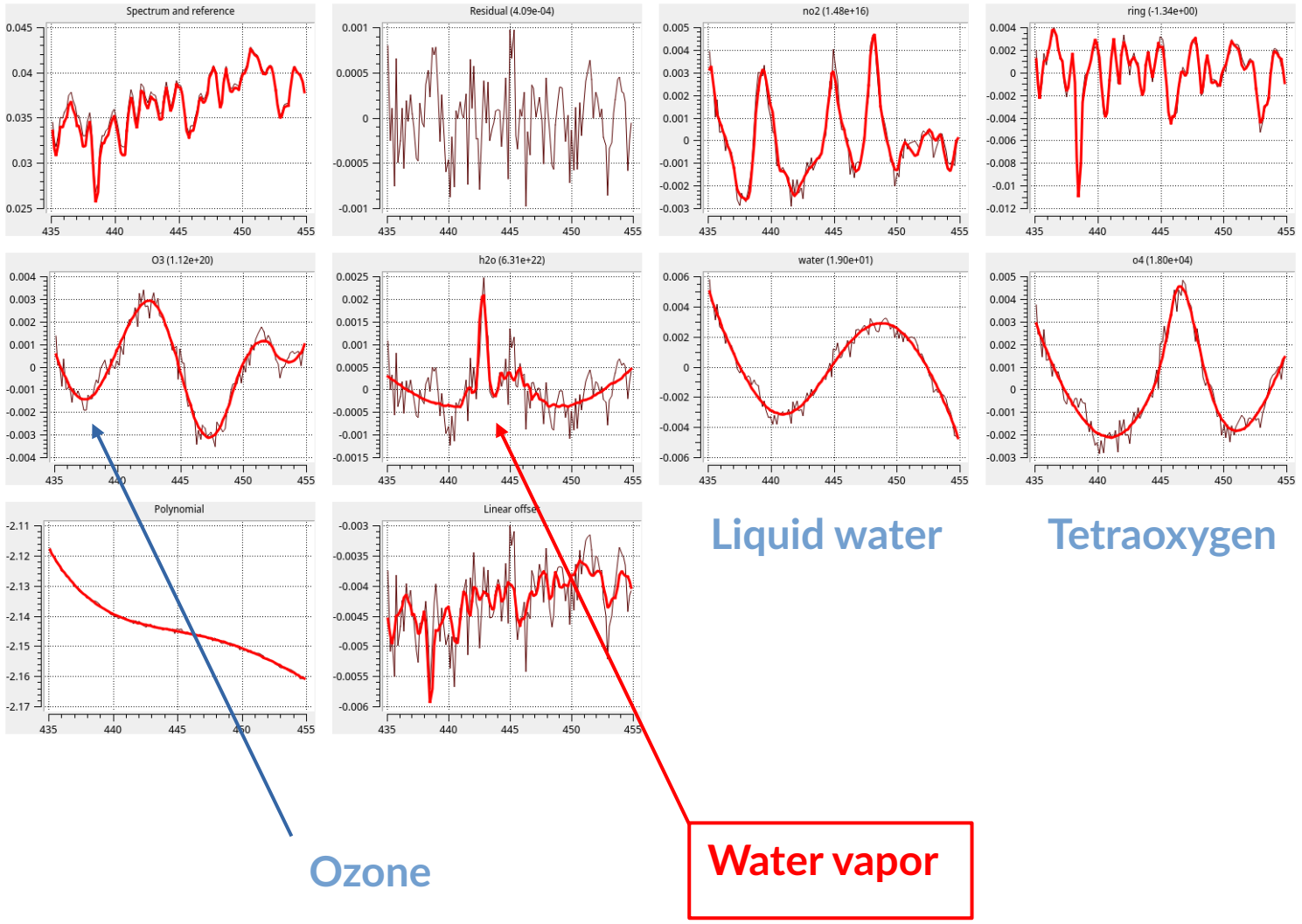
TROPOMI (S5P) 435 – 455 nm
GOME-2 (MetOp) 427.7 – 455 nm

S4 and S5: yet to be fixed

For now:

- S4 : 435 – 455 nm
- S5 : 425 – 455 nm

# Algorithm: DOAS




The **blue band of water vapor** is the weakest but it is sensed by more sensors and the surface is less reflective in the UV

→ best trade off between sensitivity and coverage over both land and oceans

**Fitting windows** are different from other works

TROPOMI (S5P) 435 – 455 nm
GOME-2 (MetOp) 427.7 – 455 nm

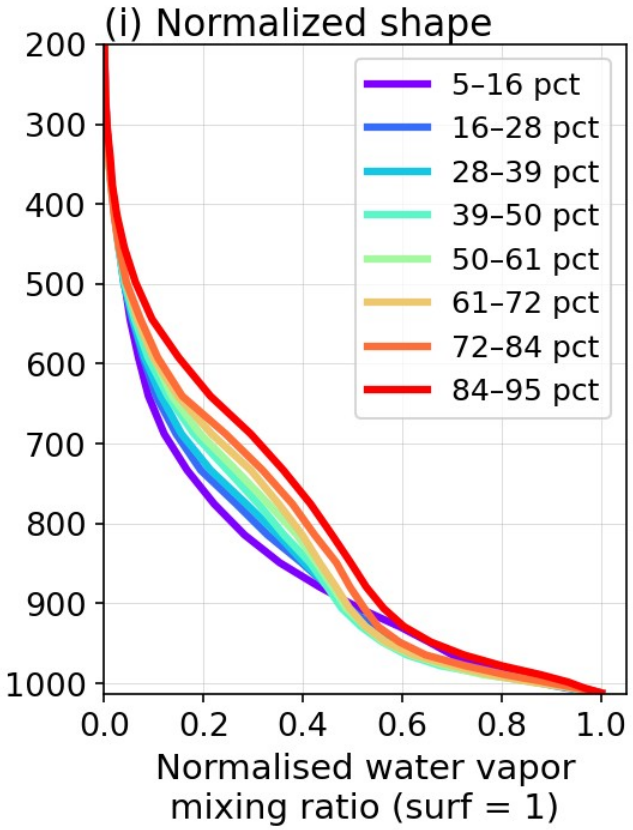
S4 and S5: yet to be fixed

For now:

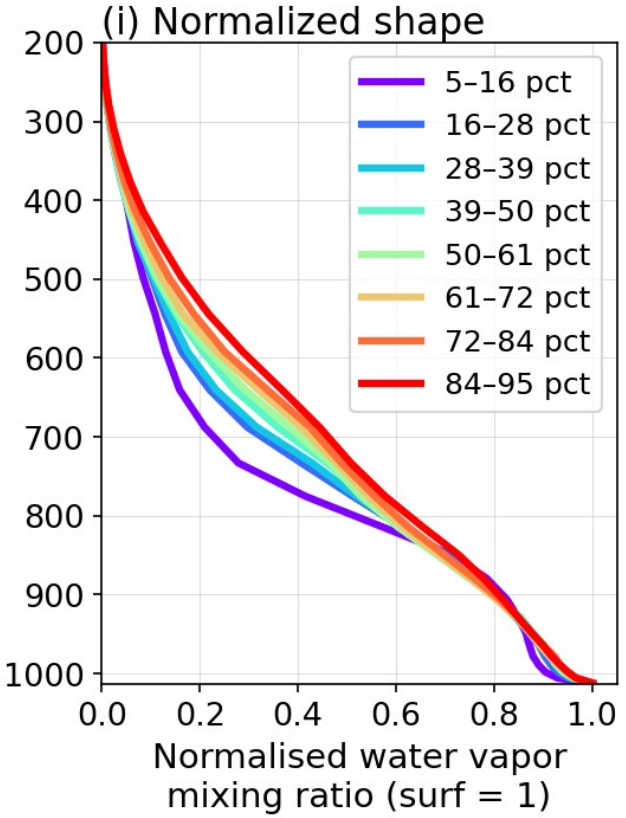
- S4 : 435 – 455 nm
- S5 : 425 – 455 nm

# Algorithm: a-priori profiles for overpass time and daily cycle

Mediterranean Sea
35-42°N, 10-25°E
July 8:00 UTC

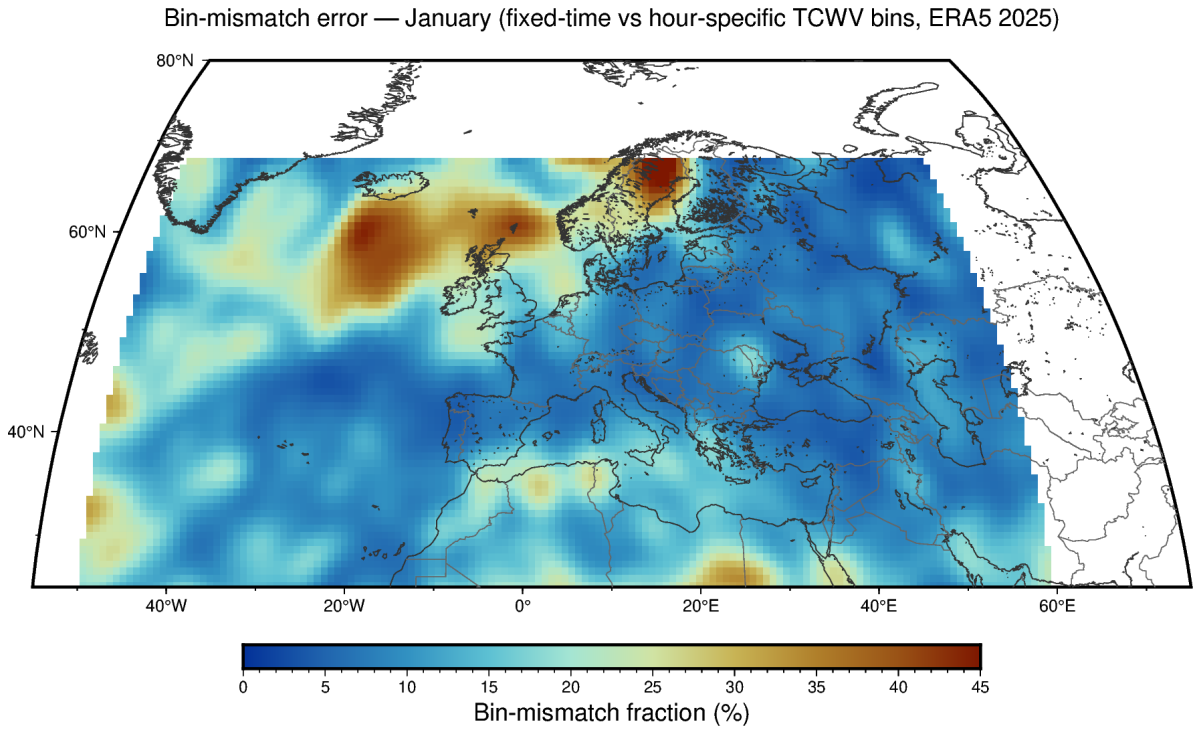


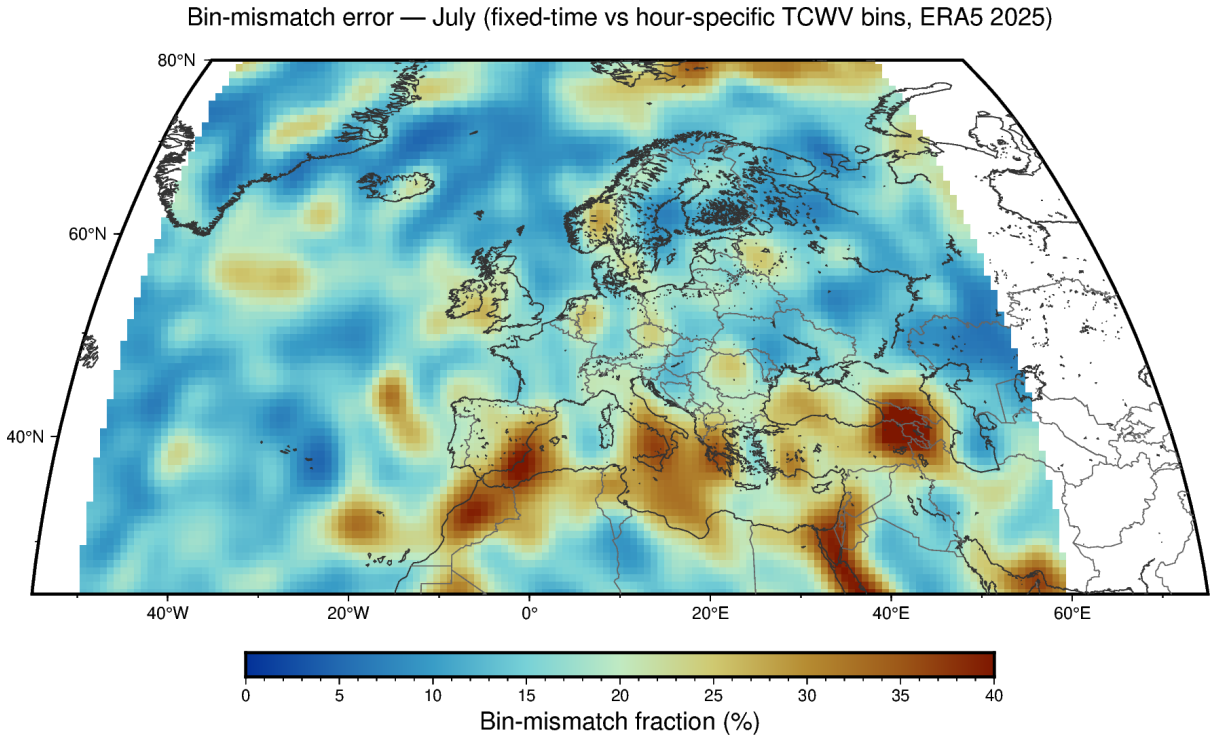

Continental Europe
47°-54°N, 5°-20°E
July 8:00 UTC




- S4_TARGET_HOURS: HH + 0.8793 h for HH = 0..23 (scan midpoint ≈ HH:52:45 UTC, derived from March 2026)
- ERA-5 for 2023-2025
- ERA5 resolution
  0.25° 1-hourly Sentinel-4
  0.75° 6-hourly Sentinel-5
- Iteratively compute VCD with box air mass factors to match the proper TCWV range for a given profile

# Algorithm: bin-mismatch fraction in fix-time prior profiles




On average across all daylight pixels, about 1 in 6 of the 8 TCWV bins would be misclassified if daily-mean bin boundaries are used instead of hour-specific ones.

| | mean | median | p95 | max |
|---|---|---|---|---|
| January | 14.5% | 11.1% | 39.1% | 70% |
| July | 17.0% | 14.2% | 40.2% | 80% |

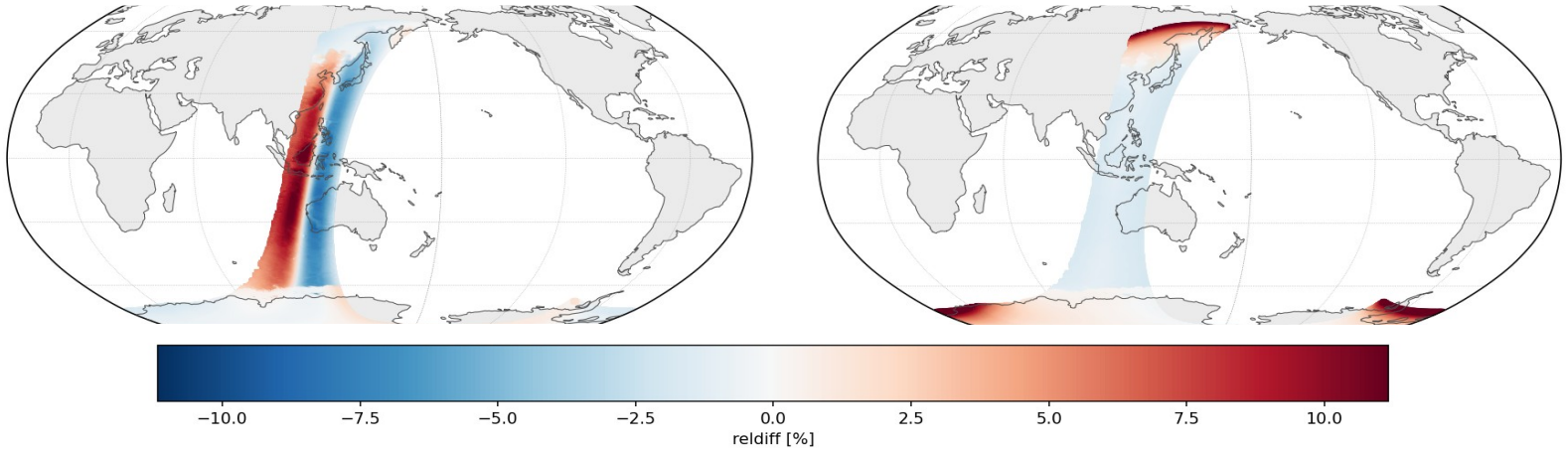
# Algorithm: full spherical box air mass factors

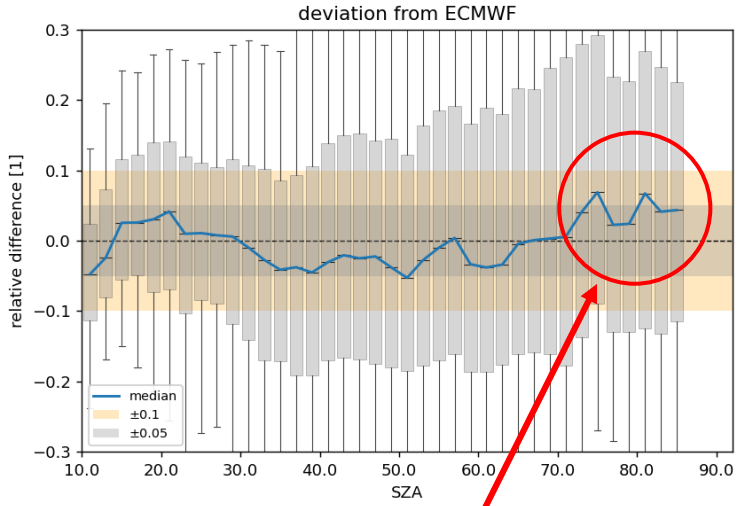
pseudo-spherical = plane-parallel (diffuse beam, Earthshine) + spherical (direct beam, Sun)

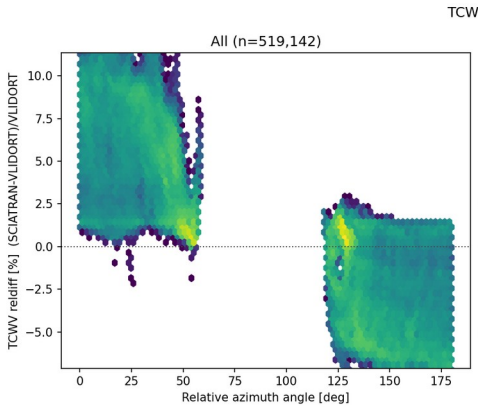
full-spherical = spherical in both diffuse (Earthshine) and direct (Sun) beam

VLIDORT pseudo-spherical
SCIATRAN full-spherical

VLIDORT pseudo-spherical
SCIATRAN plane-parallel




*The AMF difference is not viewing zenith but azimuthally dependent*

Full-spherical AMF reduces degeneracy for long light paths. (SZA > 75 degrees)

# Algorithm: full error budget and propagation – from S5P to S4 and S5

| Error type | Before | Now | Technique |
|---|---|---|---|
| SCD noise (QDOAS) | spikes uncorrected | spike_tolerance=4 | QDOAS spike clipping |
| Cloud covariances As,Ps,Ac,Pc,CF | diagonal $S_b$, fixed $\delta$'s | per-pixel cross-terms | Rodgers OE, $\hat{S} = (K^\top S_y^{-1} K + S_a^{-1})^{-1}$ |
| Smoothing error | degenerate LUT lookup | analytic $\sigma_{sm}$ per pixel | $(A-I)^\top S_a (A-I)$, ERA5-calibrated<br>$S_{a,zz'} = f^2 b_z b_{z'} e^{-\lvert\Delta \ln p\rvert / L}$<br>$L$=decorr. length in $\ln p$ |
| Quality flag | binary $crf$>0.5 cut | continuous $r_{sm}$ ladder | $qa = qa_{sys} \times f(\sigma_V)$<br>$f(\sigma_V)$: random-error factor<br>$\chi^2$ vs in-situ validation |
| Structural/Systematic RTM error | none | RTM ensemble<br>VLIDORT → SCIATRAN | baseline + physics-variant deltas<br>aerosol (OT, type) + BRDF |

**Before:**

$$\sigma^2_{VCD} = \left(\frac{\sigma_{\mathrm{SCD}}}{AMF}\right)^2 + \left(\frac{VCD}{AMF}\right)^2 \underbrace{\left[\sum_i \left(\frac{\partial AMF}{\partial b_i}\right)^2 \delta^2_{b_i} + \varepsilon^2_{\mathrm{pro}}\right]}_{\text{diagonal } S_b,\ \text{degenerate smoothing}}$$

**After:**

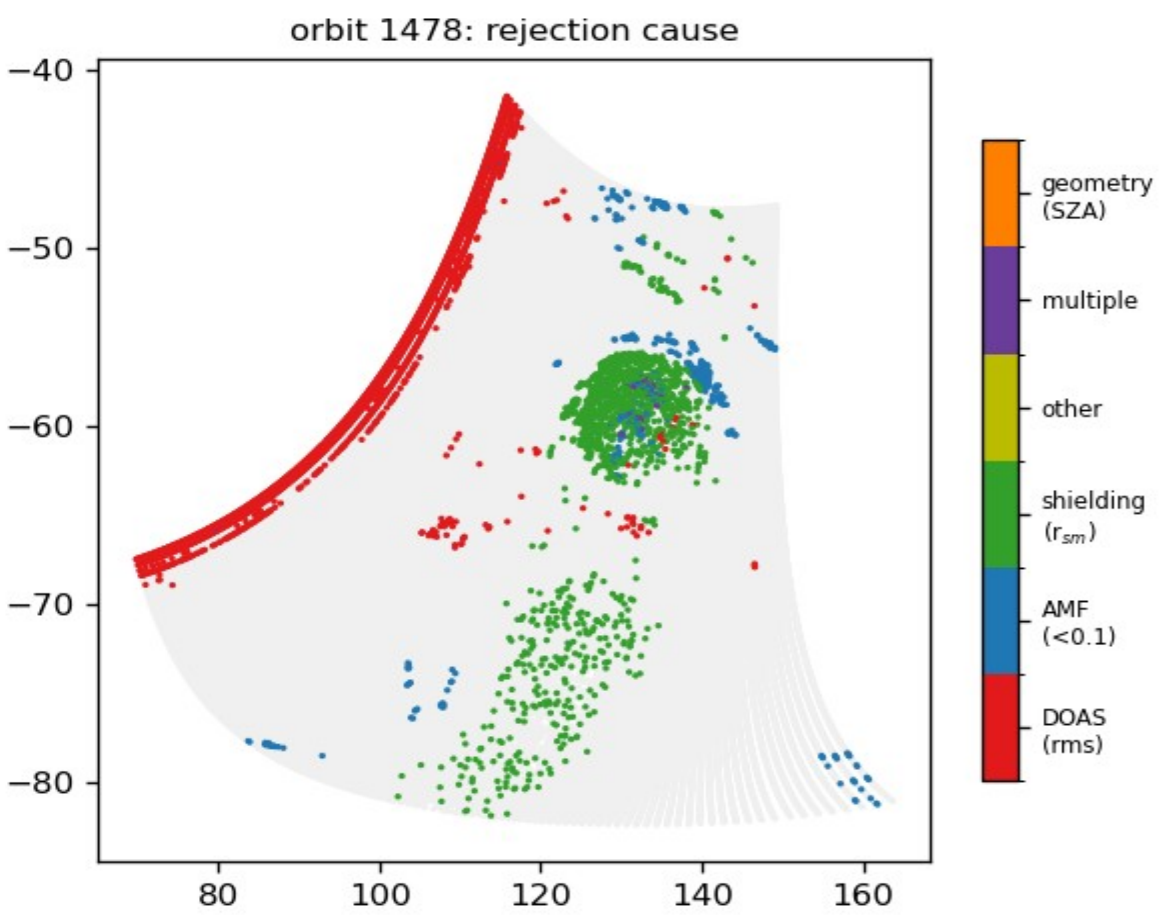
$$\sigma^2_{VCD} = \left(\frac{\sigma_{\mathrm{SCD}}}{AMF}\right)^2 + \left(\frac{VCD}{AMF}\right)^2 \underbrace{\left[\mathbf{J}^\top S_b \mathbf{J} + \sigma^2_{sm} + \sigma^2_{struct}\right]}_{\text{full } S_b \text{ (cross-terms)},\ \text{real } \sigma_{sm},\ \text{RTM ensemble}}$$

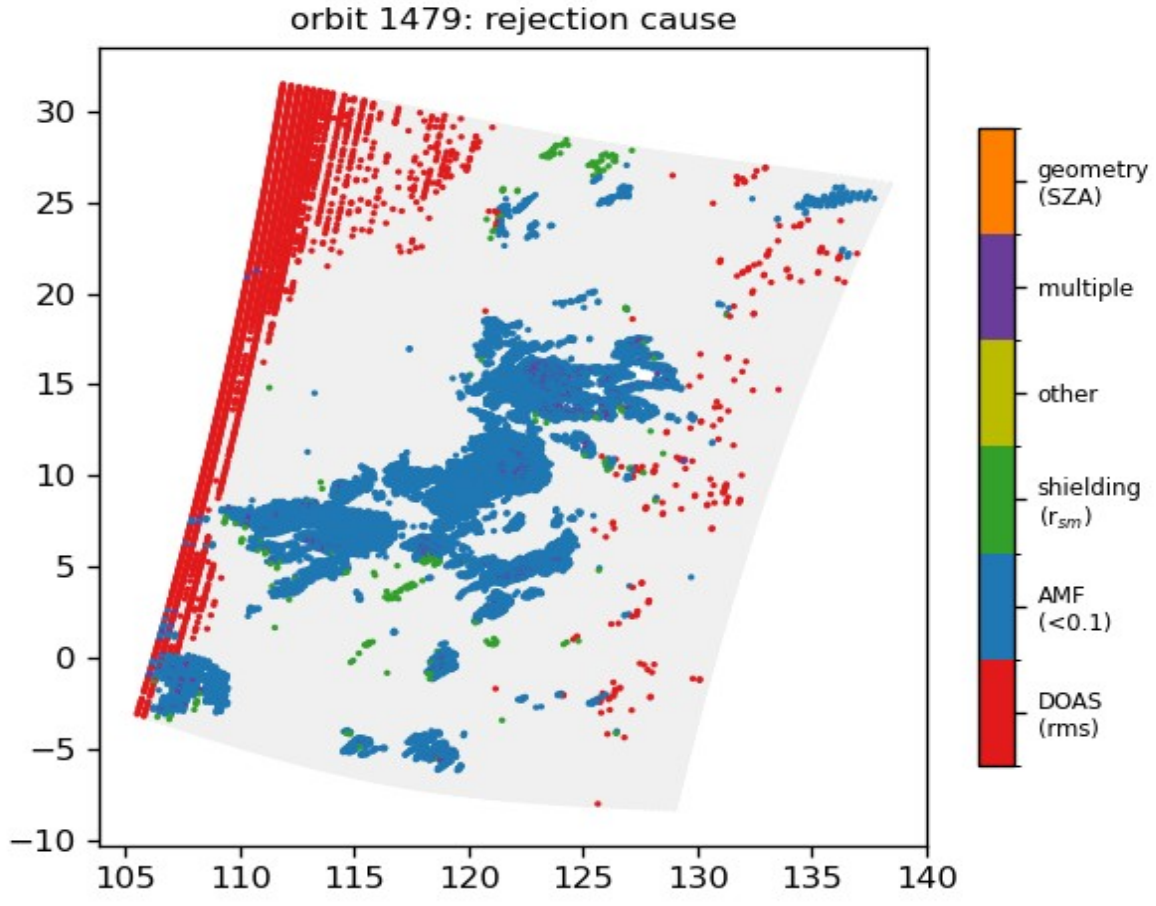
$\mathbf{J} = \partial AMF / \partial \mathbf{b}$, $\mathbf{b} = [A_s, P_s, A_c, P_c, CF]$.

- Cloud and surface off-diagonal covariances from the NN-model of the cloud retrieval algorithm
- Propagated smoothing error from prior profile (Enkes and Boersma)
- Air Mass Factors LUT with MODIS-based aerosol climatologies
- Split errors for AMF_above and _below clouds

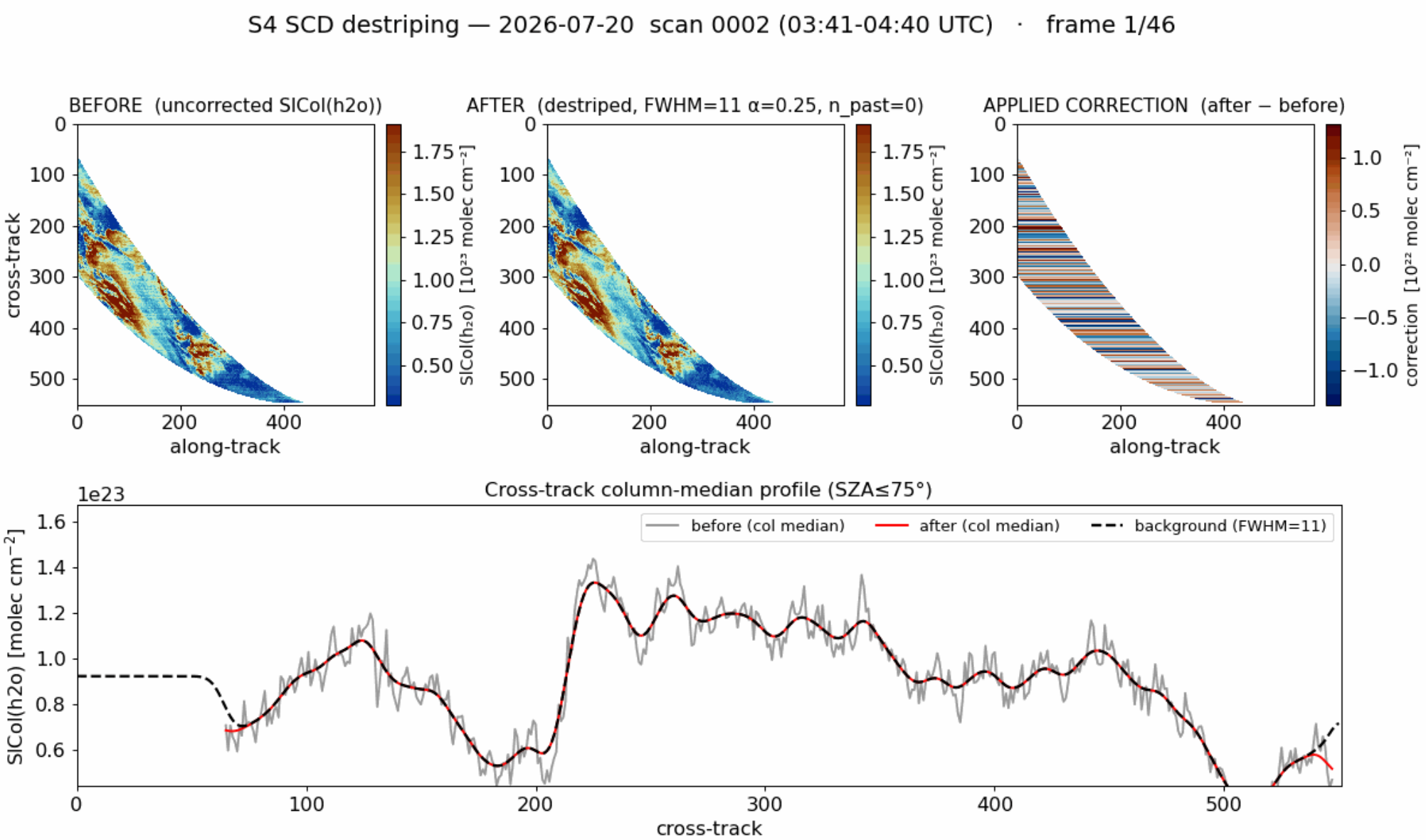
# Algorithm: graded QA for granularity and error attribution

Example of two Sentinel-5 granules



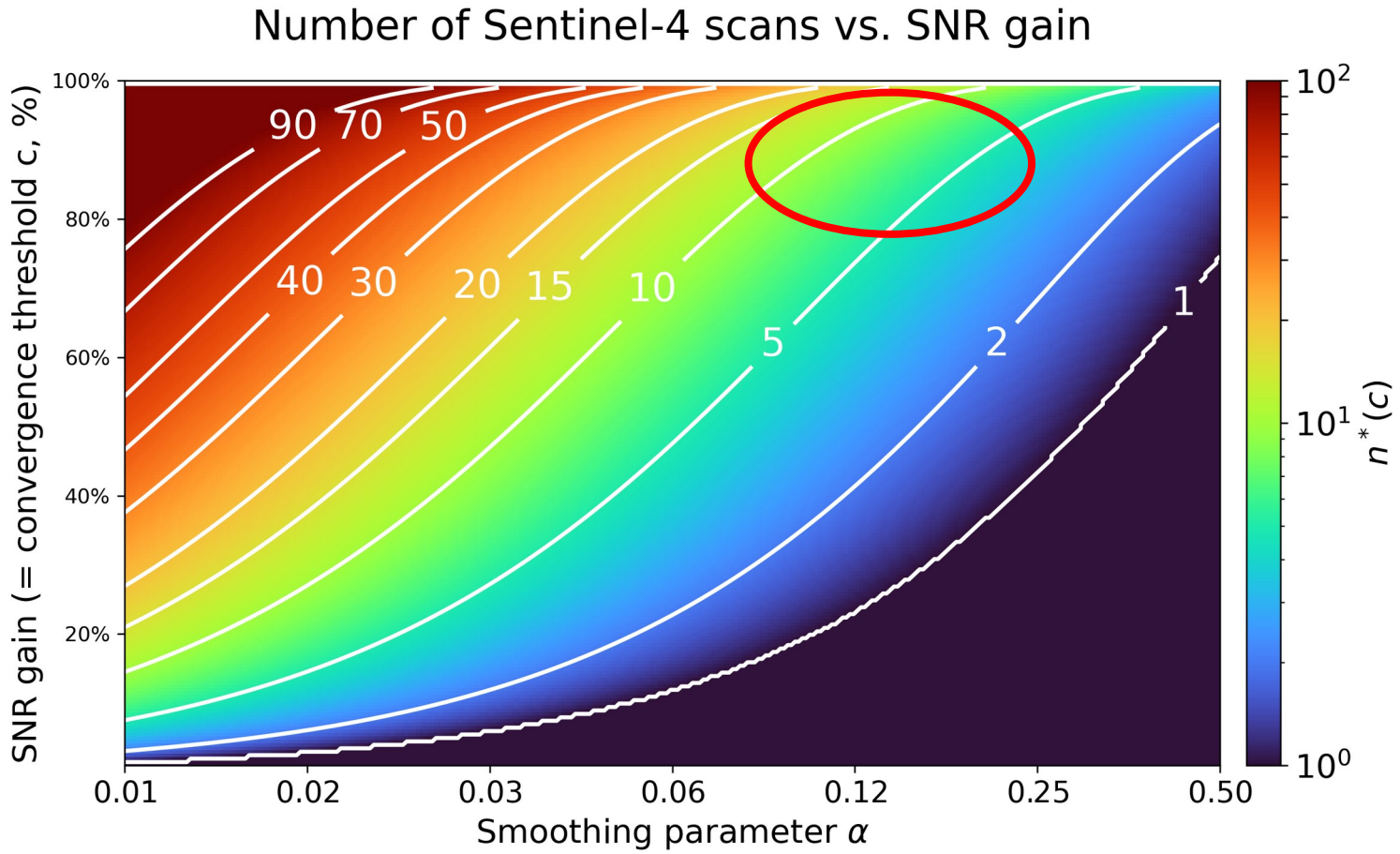
# Algorithm: destriping Sentinel-4 L1



# Algorithm: destriping Sentinel-4 L1 – the sweet spot

$$r_n = (1-\alpha)^n r_0 \; + \; \alpha \sum_{k=0}^{n-1} (1-\alpha)^k \, x_{n-k}.$$

$$w_k = \alpha \sum_{k=0}^{n-1} (1-\alpha)^k = \alpha \sum_{k=0}^{n-1} e^{\,k \ln(1-\alpha)}$$




- ***r_n*** is the running estimate after observing **x_n**
- ***w_k*** applied weights on ***k*** past scans for a given parameter ***alpha***

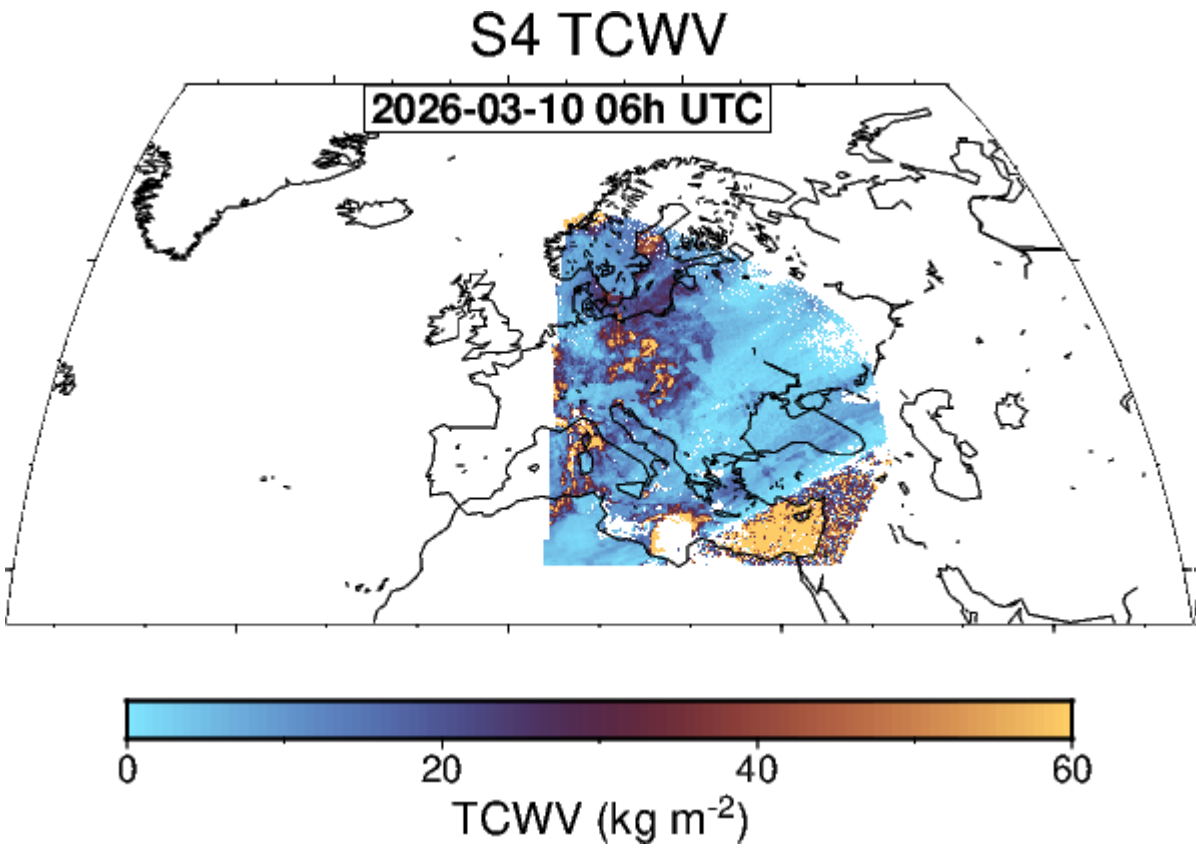
$$\text{SNR gain} = \sqrt{\frac{2-\alpha}{\alpha}} \approx \frac{1}{\sqrt{\alpha}} \quad \text{for small } \alpha.$$

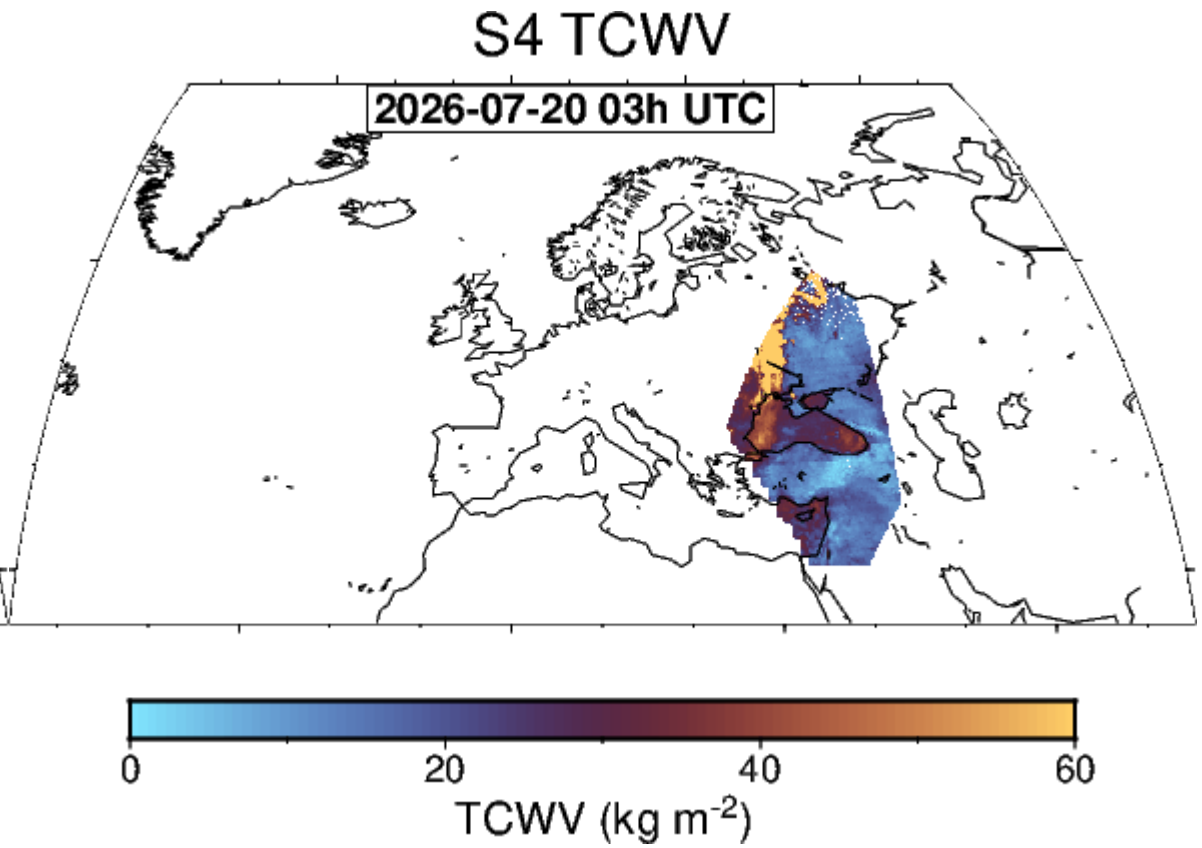
- Without prior state, roughly half of the actual day is required before the correction is reliable.

# Sentinel 4: CalVal datasets

March 05-11

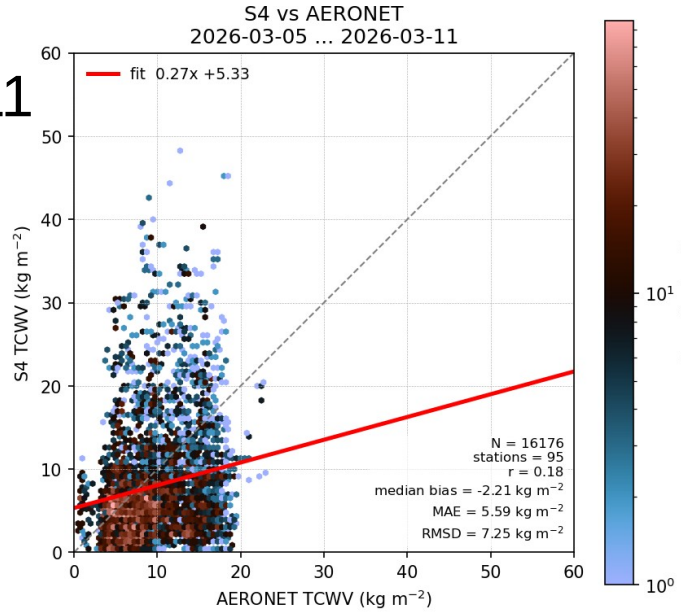
July 20-30

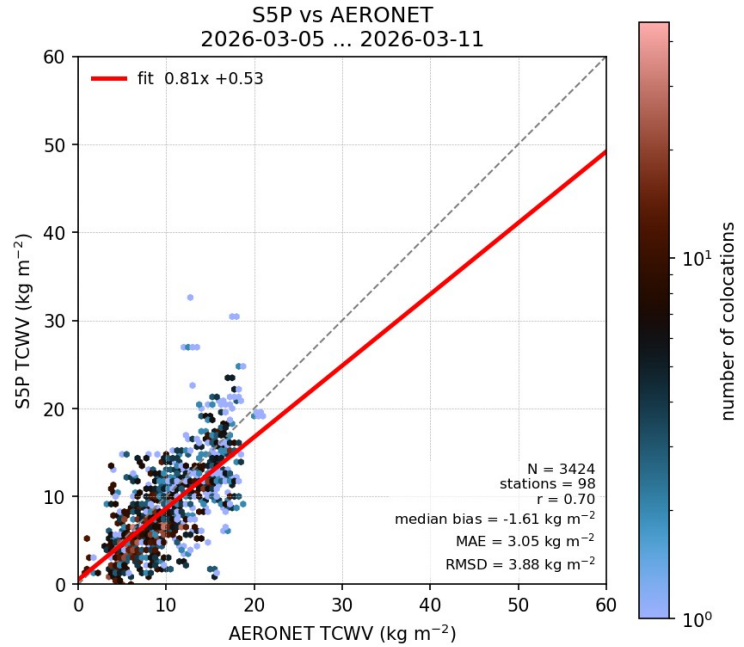


# Sentinel 4 vs S5P vs AERONET

March 05-11

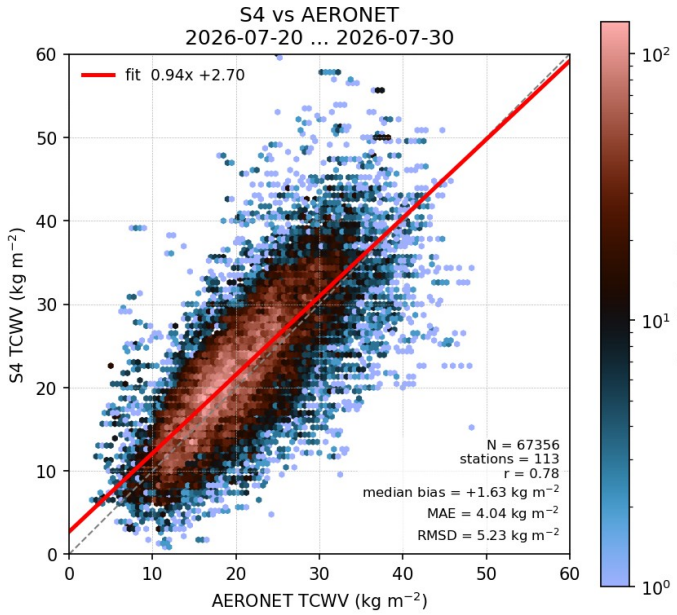


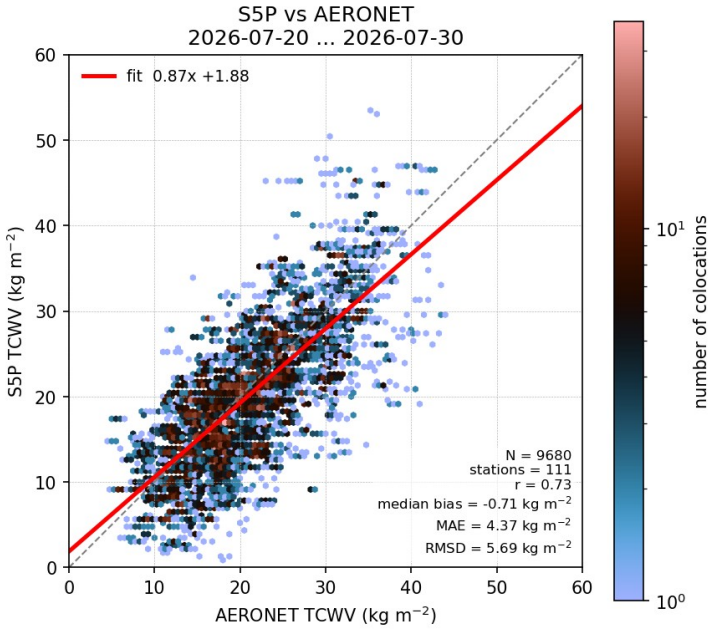




**March**
S5P is fine (r 0.70)
S4 is poor (r 0.16)

*Sampling is ruled out*
restricting S4 to S5P's exact midday overpass window barely moves S4
(r 0.16 → 0.21
RMSD 7.79 → 7.30)

July 20-30



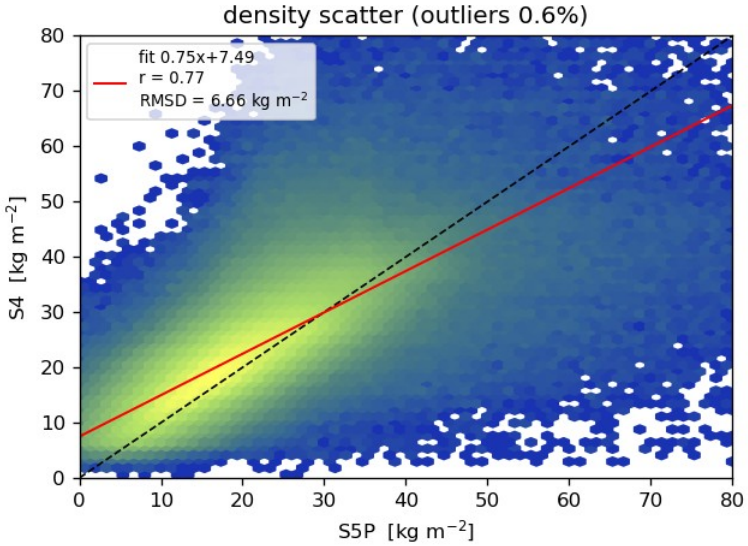

*S5P sees the same dry scene at the same times and gets r=0.70 with half the RMSD.*

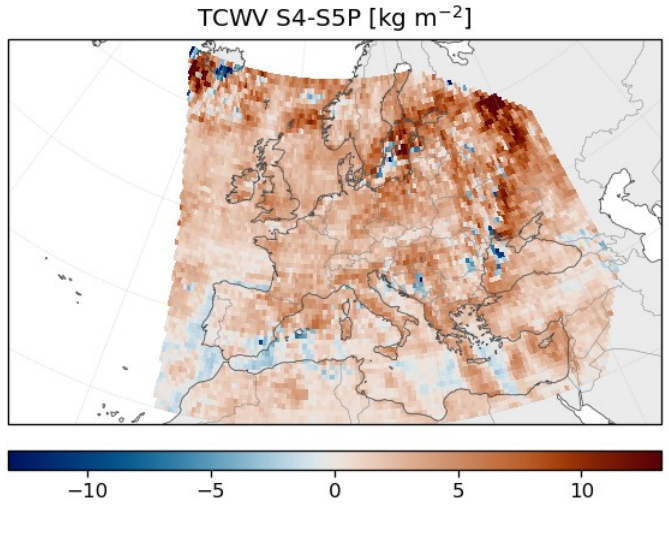
**July**
S4 and S5P agree with AERONET about equally well (r 0.74 vs 0.73).
*Both good. ✓ as expected.*

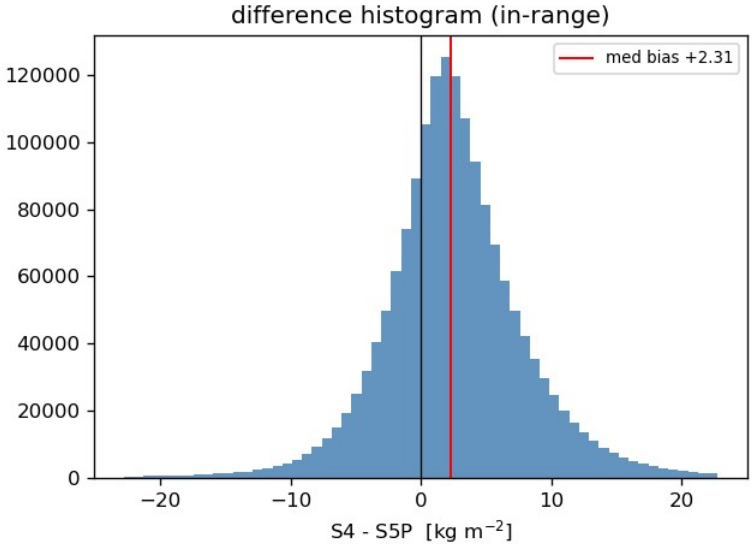
# Sentinel 4 validation: S5P and ERA5 for July 20-30

## vs S5P

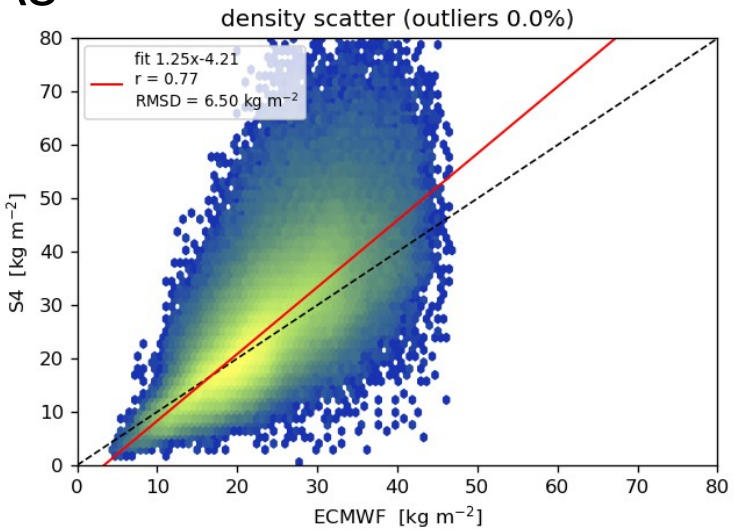
S4 vs S5P -- TCWV [N=1663905 medbias=+2.33 meanbias=+2.53 RMSD=7.22 | in-range(99%): medbias=+2.31 RMSD=6.66 r=0.765 slope=0.75]

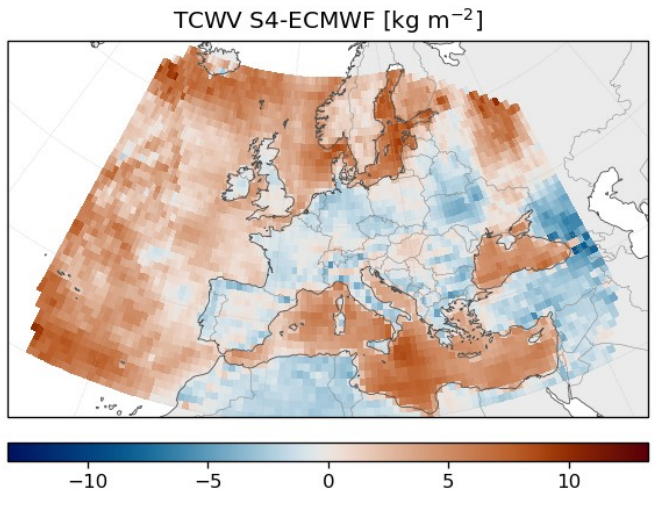


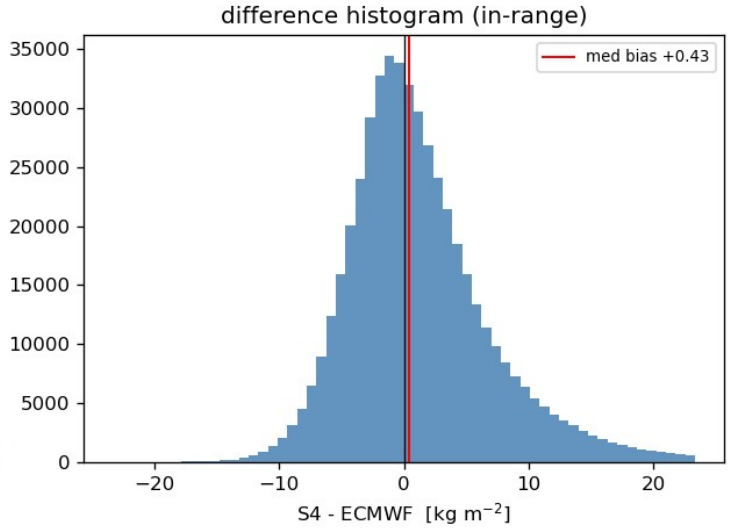



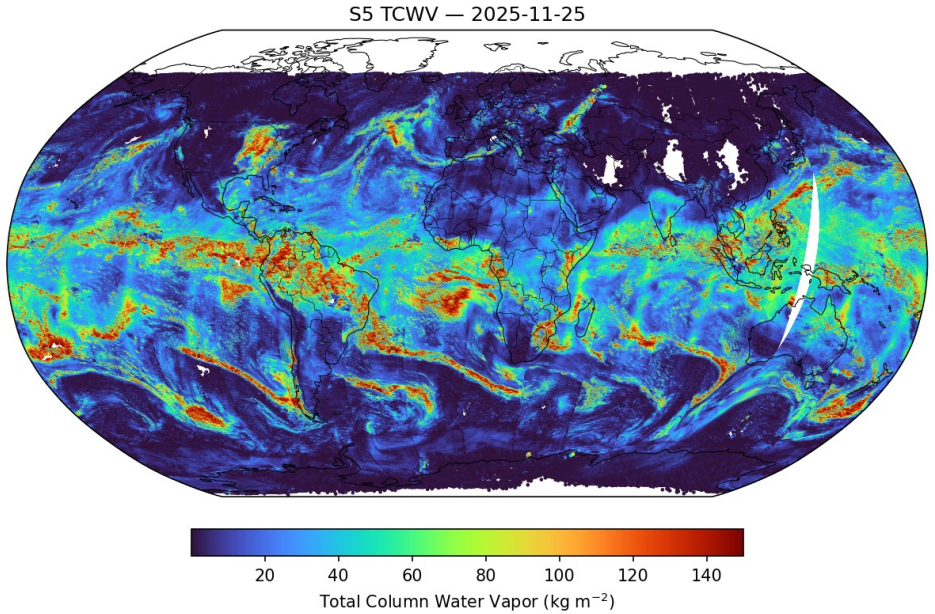

## vs ERA5

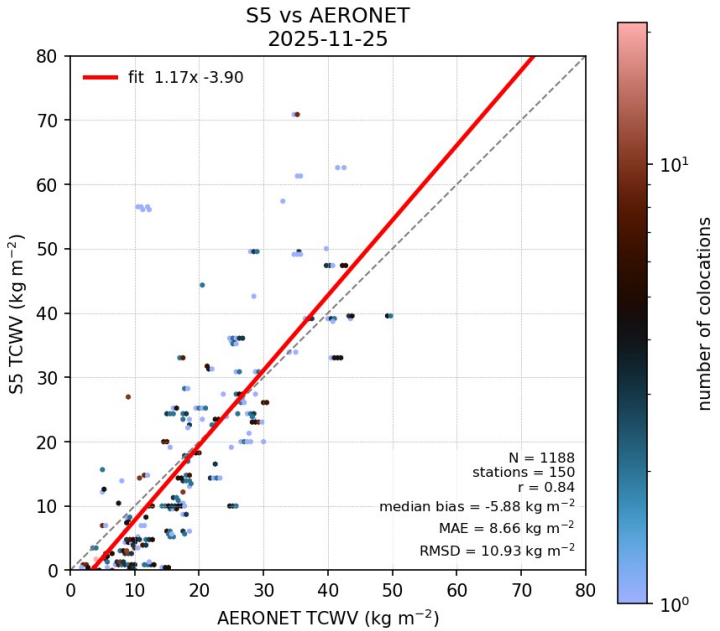


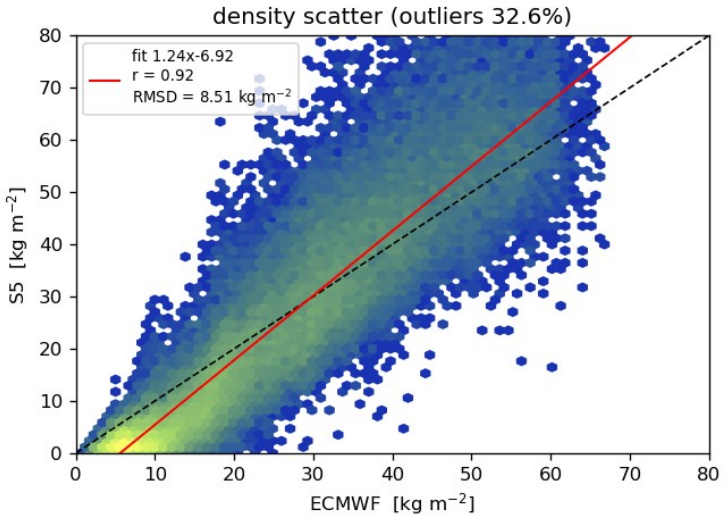



# Sentinel 5: initial results

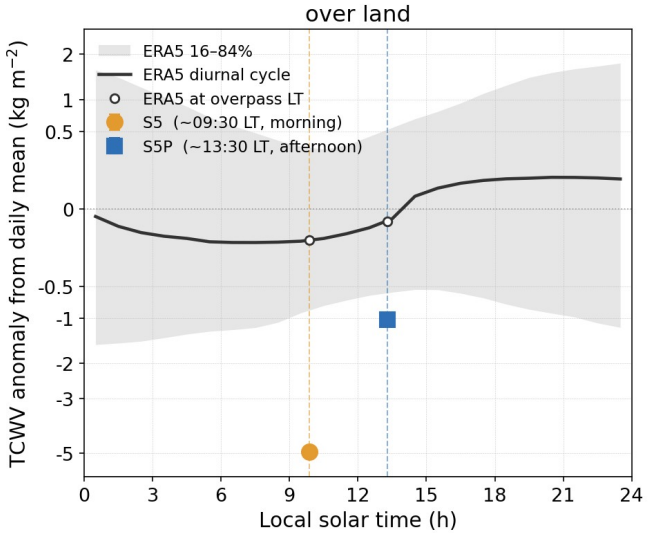


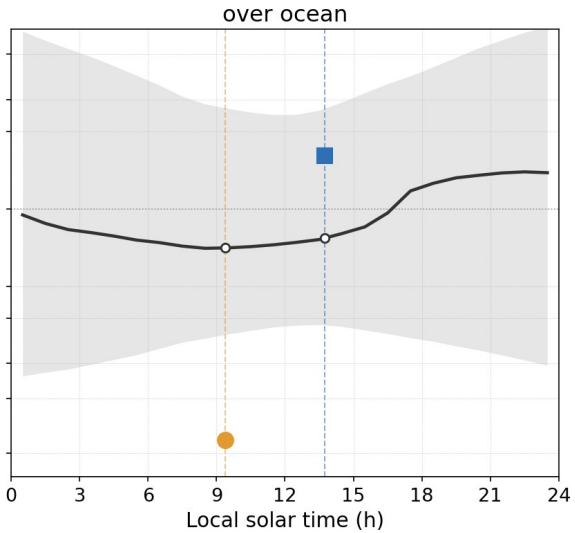




- Not enough data yet
- Status of the TCWV processor
  **Drier in dry**, **wetter in moist**
- Higher than usual negative SCD
- **Recalibrate fit window first**




S5 (9:30 LT) and S5P (13:30 LT) along ERA5 diurnal cycle



# Sentinel 4: application – hygroscopic growth at Ispra

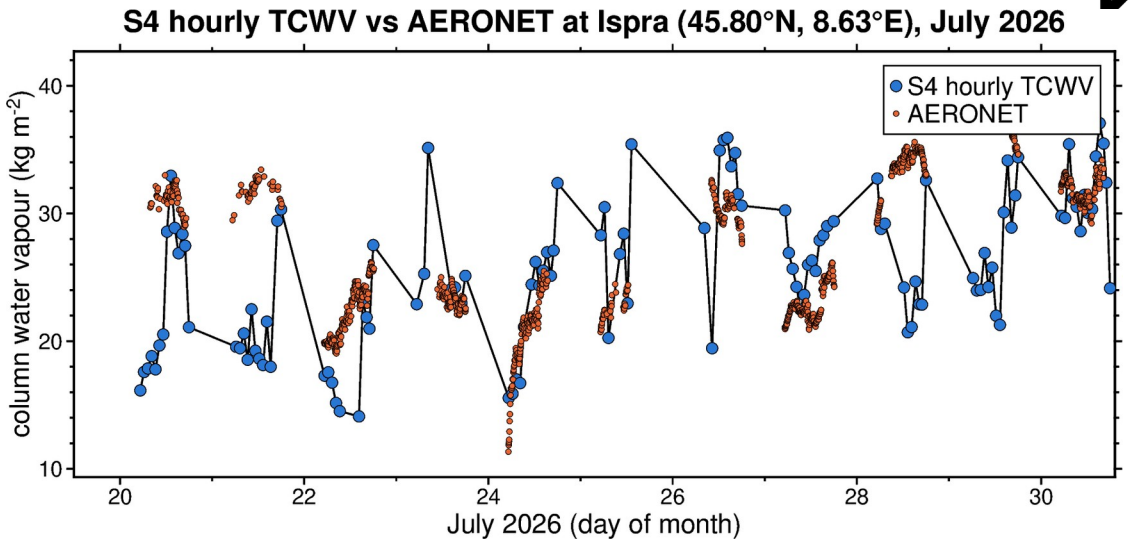
$$\mathrm{AOD}(t) \;=\; \mathrm{AOD}_{\mathrm{dry}}\left[1 - \frac{\mathrm{RH}_{\mathrm{eff}}(t;\tau)}{100}\right]^{-\gamma}$$

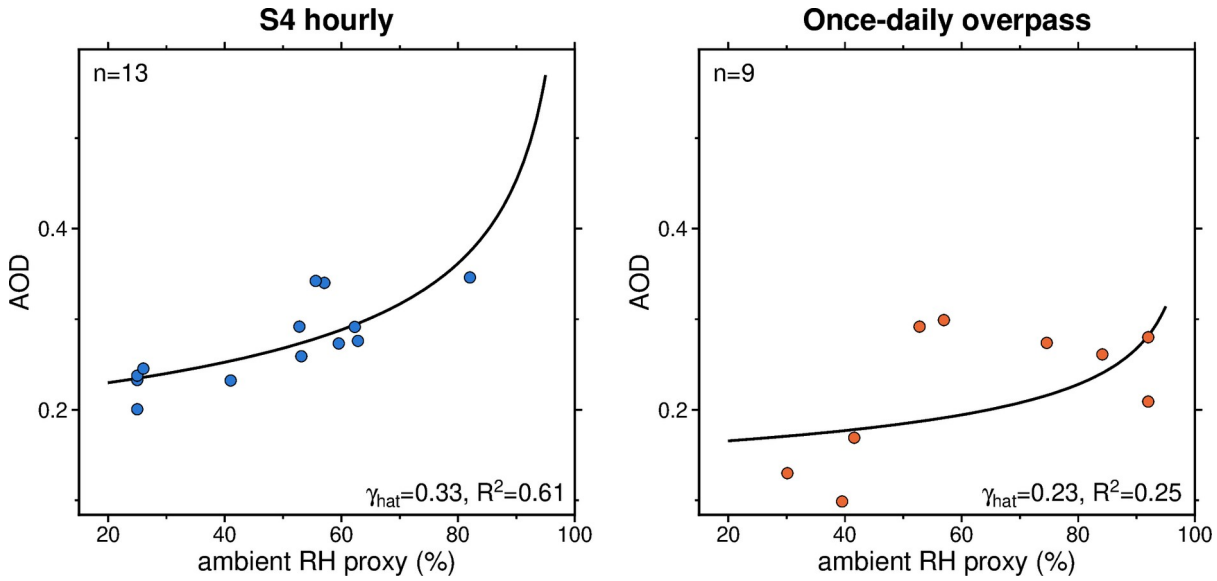
$$\hat{\tau} \;=\; \arg\max_{\tau} R^2\Big[\,\log \mathrm{AOD}(t) \text{ vs. } \log\big(1 - \mathrm{RH}_{\mathrm{eff}}(t;\tau)/100\big)\Big]$$

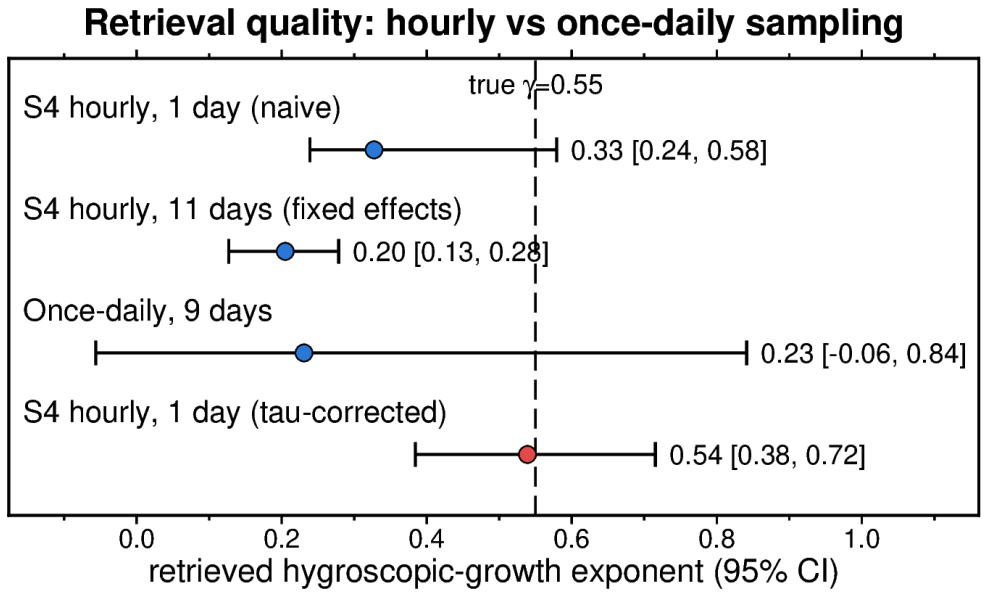
Kasten, F., Tellus, 1969

We solve the equations above for:

***1)Tau*** – lagged water uptake (relaxation time)
***2)Gamma*** – hygroscopic growth factor



# Imprint

Topic: **Total column water vapor for Sentinel-4 and Sentinel-5 measurements: inital results**

Date: 2026-09-25

Author: Luca Lelli

Institute: ACS – IMF - EOC