PROGRAMME OF THE EUROPEAN UNION

co-funded with

Discussion on Matchup Protocols for Sentinel-3 Ocean Colour Product Validations in Comparison with in situ Measurements

7th Sentinel-3 Validation Team Meeting 2022

18-20 October 2022 | ESA-ESRIN | Frascati (Rm), Italy

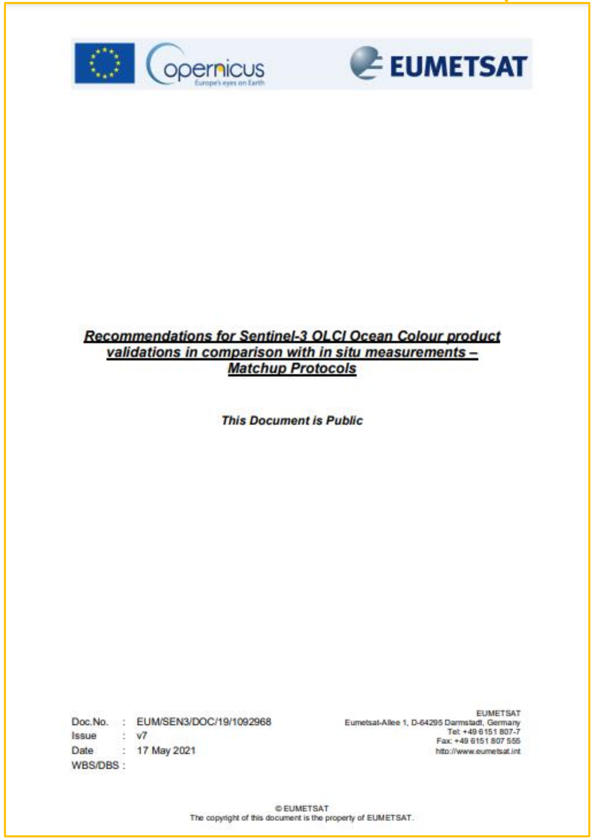
Juan Ignacio Gossn

EUMETSAT

S3 OLCI OC matchup protocols, discussion

Objective of this discussion: **EUMETSAT's Matchup Protocols**

- This protocol arouse as a need to have a common standard so that the results across this team are as comparable as possible
- This document is assumed to respond to a consensus reached from this validation team, so all of us showing matchups should be aware of its existence, follow it or, in case of departing from it, explain the rationale behind your proposed variants.
- The protocol is currently being reviewed at EUMETSAT…



Recommendations for Sentinel-3 OLCI Ocean Colour product validations in comparison with in situ measurements – Matchup Protocols

This Document is Public

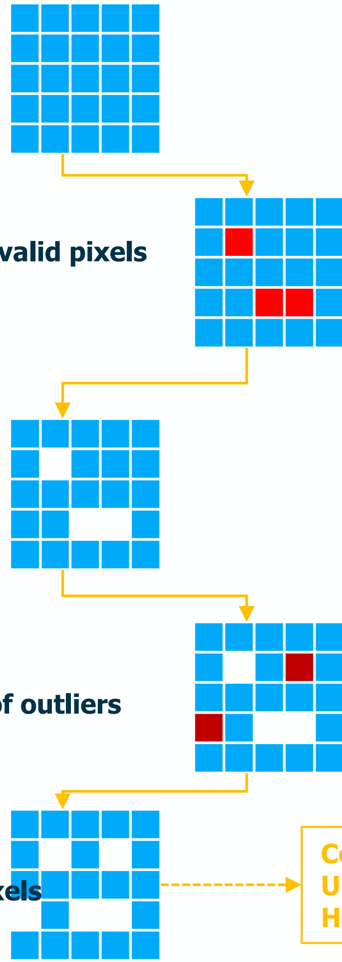
Doc.No. : EUM/SEN3/DOC/19/1092968
Issue : v7
Date : 17 May 2021
WBS/DBS :

EUMETSAT
Eumetsat-Allee 1, D-64295 Darmstadt, Germany
Tel: +49 6151 807-7
Fax: +49 6151 807 555
http://www.eumetsat.int

→ Available at: eumetsat.int/media/44087

S3 OLCI OC matchup protocols, discussion

EUMETSAT's Matchup Protocols: extraction of statistics at macropixel level: **today**



Set BFOR: 5x5 window

Detection of non-valid pixels (flagged pixels)

Pixels are masked/removed if flagged by any of the following:

CLOUD, CLOUD_AMBIGUOUS, CLOUD_MARGIN, INVALID, COSMETIC, SATURATED, SUSPECT, HISOLZEN, HIGHGLINT, SNOW_ICE, AC_FAIL, WHITECAPS, ADJAC, RWNEG_O2, RWNEG_O3, RWNEG_O4, RWNEG_O5, RWNEG_O6, RWNEG_O7, RWNEG_O8
+ product-specific flags e.g. OC4ME_FAIL

Macropixel is discarded if:

$N_{BOR} < 50\% \, N_{BFOR}$

Set BOR: without flagged pixels

Detection of outliers

Pixel 'X' is considered <u>outlier</u> if:

$|value@X - \mu_1| < 1.5\sigma_1$

Macropixel is discarded if:

$CV_{final}(560) > 20\%$

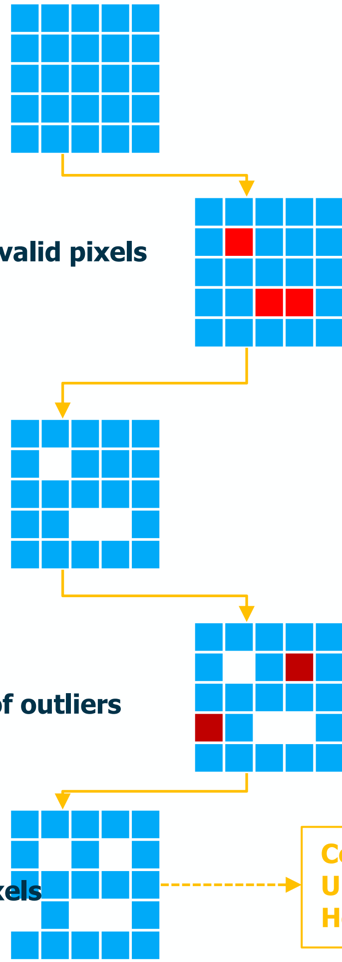
Set final: without flagged & outlier pixels

Central value: $median_{final}$
Uncertainty measure (Type B): σ_{final}
Homogeneity measure: CV_{final}

μ → Mean
σ → Standard deviation

S3 OLCI OC matchup protocols, discussion

EUMETSAT's Matchup Protocols: extraction of statistics at macropixel level: initial changes



Set BFOR: 5x5 window

Detection of non-valid pixels (flagged pixels)

Pixels are masked/removed if flagged by any of the following:

CLOUD, CLOUD_AMBIGUOUS, CLOUD_MARGIN, INVALID, COSMETIC, SATURATED, SUSPECT, HISOLZEN, HIGHGLINT, SNOW_ICE, AC_FAIL, WHITECAPS, ADJAC, RWNEG_O2, RWNEG_O3, RWNEG_O4, RWNEG_O5, RWNEG_O6, RWNEG_O7, RWNEG_O8
+ product-specific flags e.g. OC4ME_FAIL

Macropixel is discarded if:

$N_{BOR} < 50\% \, N_{BFOR}$

Set BOR: without flagged pixels

Detection of outliers

Pixel 'X' is considered outlier if:

|value@X - ~~μ_{BOR}~~ $median_{BOR}$| < ~~$1.5\sigma_{BOR}$~~ $\frac{10}{9} IQR_{BOR}$

Set final: without flagged & outlier pixels

Central value: ~~$median_{final}$~~ μ_{final}
Uncertainty measure (Type A): σ_{final}
Homogeneity measure: CV_{final}

Macropixel is discarded if:

$CV_{final}(560) > 20\%$

μ → Mean
σ → Standard deviation

S3 OLCI OC matchup protocols, discussion

Pixel 'X' is considered outlier if:

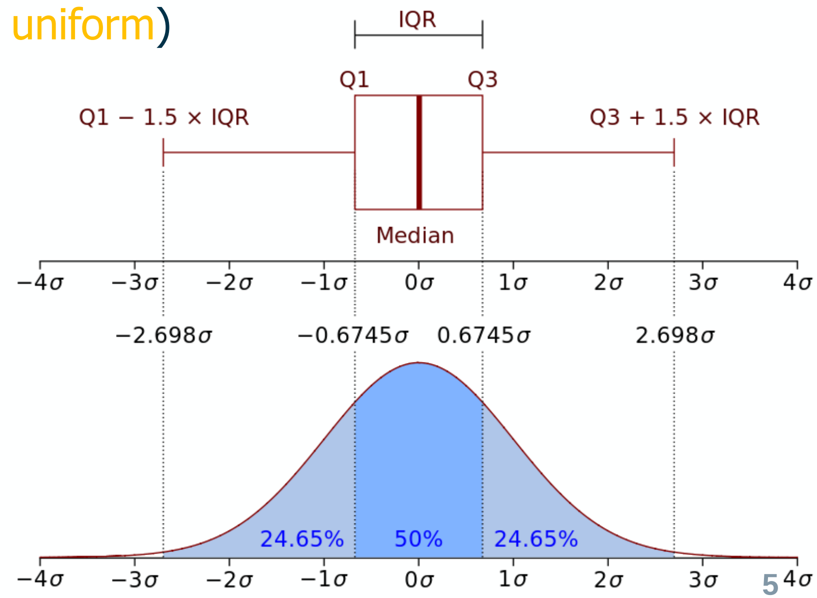
$$|\text{value@X} - \mu| < 1.5 \times \sigma$$

Pixel 'X' is considered outlier if:

$$|\text{value@X} - \text{median}| < \frac{10}{9} \times \text{IQR}$$

Why?

- **Mean** and **standard deviation** are **non-robust statistics**, they are precisely highly affected by outliers
- On the contrary, **median** and **IQR** are **robust** statistics, suitable for outlier detection
- The "tolerance factor" $\frac{10}{9}$ is chosen in a way that $1.5 \times \sigma \cong \frac{10}{9} \times \text{IQR} \cong 1.111 \times \text{IQR}$ (normal distribution)
- This equivalence depends the distribution of the data:
    - e.g. $1.5 \times \sigma \cong \frac{\sqrt{3}}{2} \times \text{IQR} \cong 0.866 \times \text{IQR} =$ (if distribution is uniform)



S3 OLCI OC matchup protocols, discussion

An example

25 random numbers following:

N(μ = 0.025; σ = 0.0025)
+
outliers

Robust

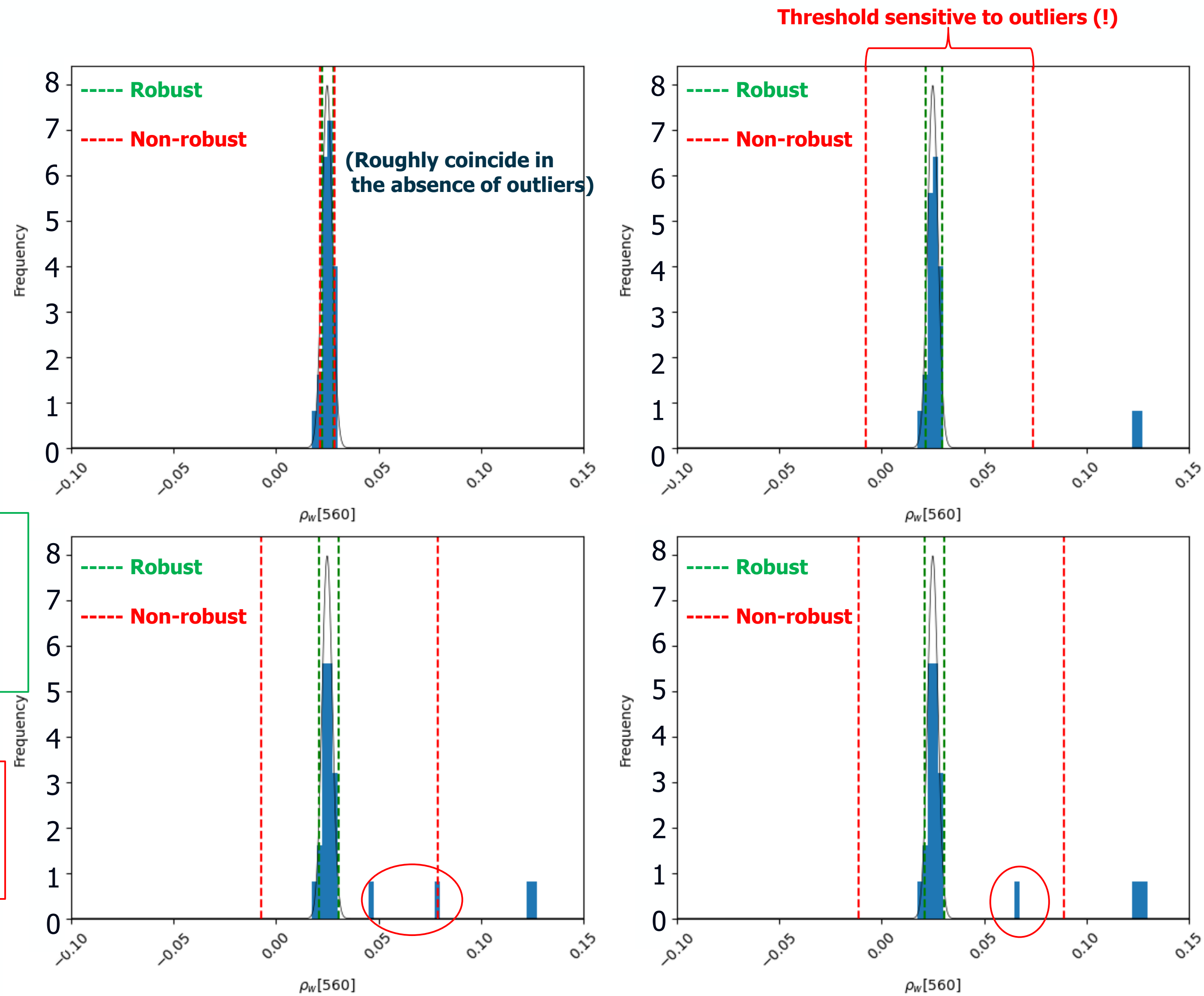
Pixel 'X' is considered outlier if:

$$|\text{value@X} - \text{median}| < \frac{10}{9} \times \text{IQR}$$

Non-robust

Pixel 'X' is considered outlier if:

$$|\text{value@X} - \text{mean}| < 1.5 \times \sigma$$

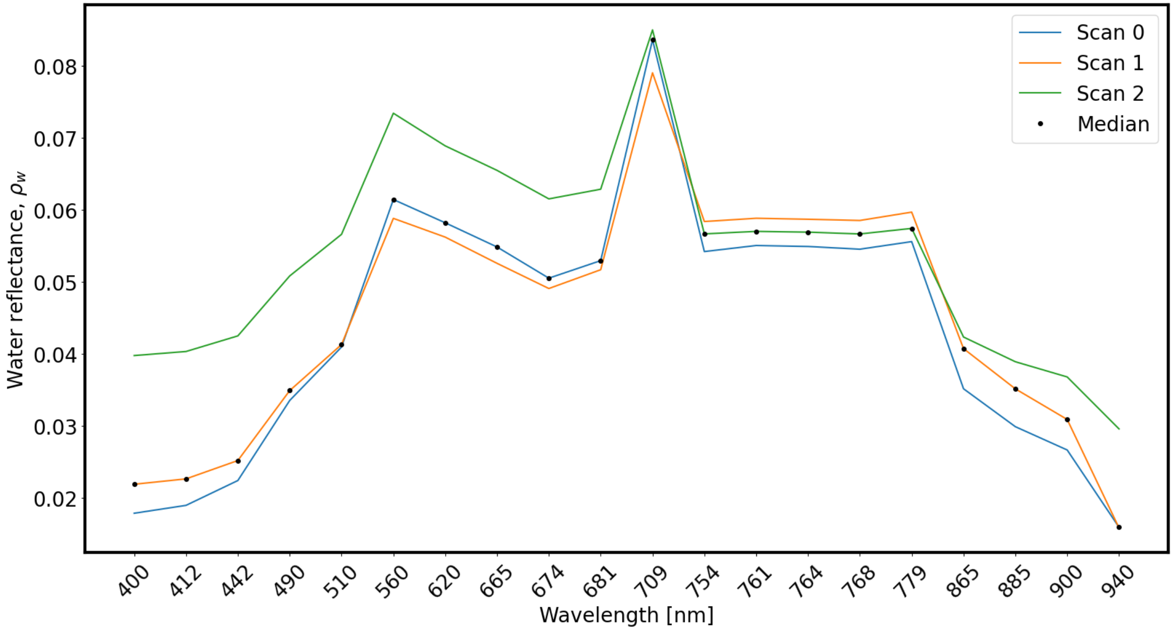


S3 OLCI OC matchup protocols, discussion

Central value: median → **Central value: μ**

Why mean (**μ**) instead of **median**?

- It shouldn't be necessary to report a robust statistic (median): outliers were removed in the previous step!
- Reporting median means choosing the middle values among the set of valid pixels in the 5x5 window.



S3 OLCI OC matchup protocols, discussion

Central value: median → **Central value: μ**

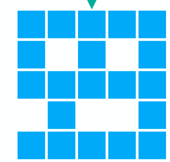
Why mean (**μ**) instead of **median**?

- Mean has a series of well-known properties that median lacks:

S3 OLCI OC matchup protocols, discussion

Central value: median → **Central value: μ**

Extraction window without flagged & outlier pixels



Case 1: Each (valid) pixel is assumed as an **independent realization** of the same measurand

Why mean (**μ**) instead of **median**?

- Mean has a series of well-known properties that median lacks: e.g.:

The **99-95-68** rule applies for normal distribution:

New realization falls in $[\mu - 3\sigma; \mu + 3\sigma]$ with 99.7% confidence (k=3)
New realization falls in $[\mu - 2\sigma; \mu + 2\sigma]$ with 95% confidence (k=2)
New realization falls in $[\mu - \sigma; \mu + \sigma]$ with 68% confidence (k=1)

S3 OLCI OC matchup protocols, discussion

This is the way recommended in the **Guide to the expression of uncertainty in measurement (GUM)**

4.2 Type A evaluation of standard uncertainty

4.2.1 In most cases, the best available estimate of the expectation or expected value μ_q of a quantity q that varies randomly [a **random variable** (C.2.2)], and for which n independent observations q_k have been obtained under the same conditions of measurement (see B.2.15), is the **arithmetic mean** or **average** $\bar{q}$ (C.2.19) of the n observations:

$$\bar{q} = \frac{1}{n}\sum_{k=1}^{n} q_k \tag{3}$$

4.2.2 The individual observations q_k differ in value because of random variations in the influence quantities, or random effects (see 3.2.2). The experimental variance of the observations, which estimates the variance σ^2 of the probability distribution of q, is given by

$$s^2(q_k) = \frac{1}{n-1}\sum_{j=1}^{n}\left(q_j - \bar{q}\right)^2 \tag{4}$$

This estimate of variance and its positive square root $s(q_k)$, termed the **experimental standard deviation** (B.2.17), characterize the variability of the observed values q_k, or more specifically, their dispersion about their mean $\bar{q}$.

4.2.3 The best estimate of $\sigma^2(\bar{q}) = \sigma^2/n$, the variance of the mean, is given by

$$s^2(\bar{q}) = \frac{s^2(q_k)}{n} \tag{5}$$

Thus, for an input quantity X_i determined from n independent repeated observations $X_{i,k}$, the standard uncertainty $u(x_i)$ of its estimate $x_i = \bar{X}_i$ is $u(x_i) = s(\bar{X}_i)$, with $s^2(\bar{X}_i)$ calculated according to Equation (5). For convenience, $u^2(x_i) = s^2(\bar{X}_i)$ and $u(x_i) = s(\bar{X}_i)$ are sometimes called a *Type A variance* and a *Type A standard uncertainty*, respectively.

S3 OLCI OC matchup protocols, discussion

Type A uncertainty when computing average of realizations.
We assume each valid pixel of the window is an **independent realization of the same measurand**

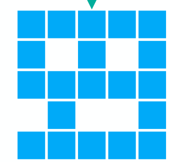
Uncertainty measure (Type A): σ → **Uncertainty measure (Type A): $\sigma/\sqrt{N}$** **?**

→ (Of course, as you know, full uncertainty budget is still missing…)

S3 OLCI OC matchup protocols, discussion

Central value: median → **Central value: μ**

Extraction window without flagged & outlier pixels



Case 2: Each (valid) pixel is assumed as a **realization** of the same measurand (with unimodal distribution)

Type A uncertainty **of the mean** still bounded by standard deviation:

$$\sigma^2(\overline{X}) = \frac{1}{25}\left(\sum_{i=1}^{25}\sigma^2(X_i) + \sum_{i<j} Cov(X_i, X_j)\right) \le \sigma^2_{window}$$

Uncertainty measure (Type A) bounded by: σ

New realisations “X” are still bounded around the mean **μ** (no independence required)

Chebyshev's inequality

$$\Pr(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Vysochanskij–Petunin inequality

$$P(|X-\mu| \geq k\sigma) \leq \frac{4}{9k^2} \quad \text{if} \quad k \geq \sqrt{8/3} = 1.633.$$

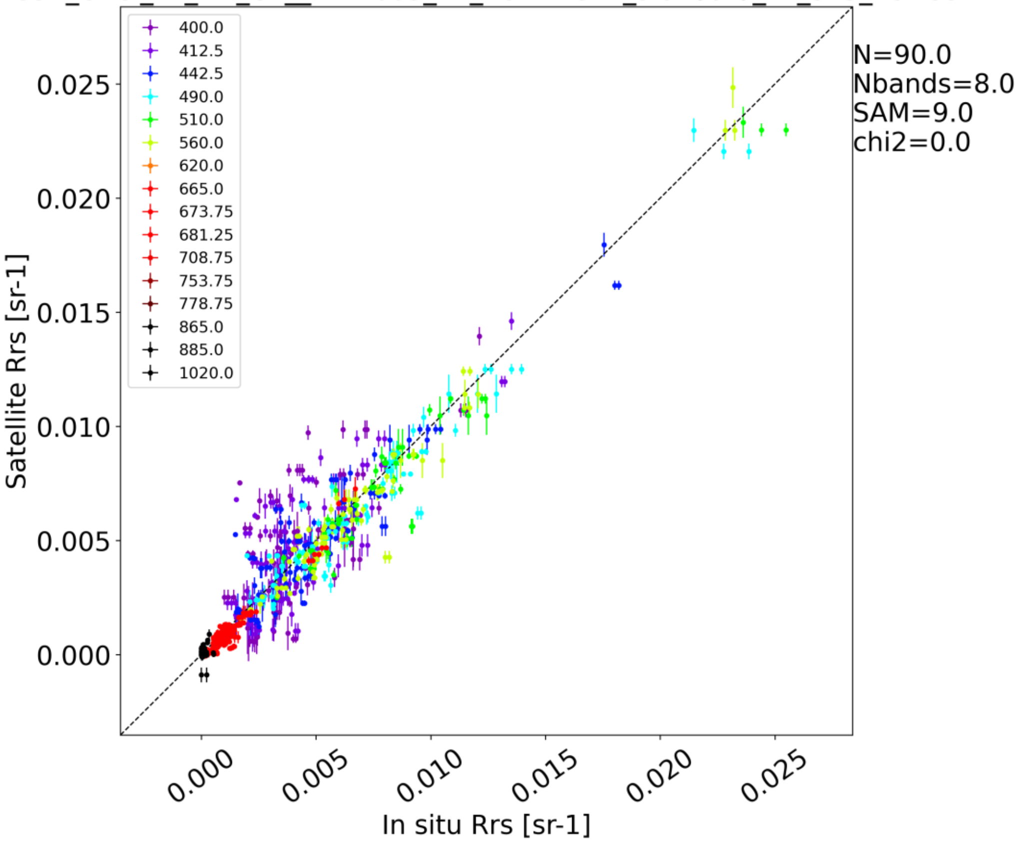
$$P(|X-\mu| \geq k\sigma) \leq \frac{4}{3k^2} - \frac{1}{3} \quad \text{if} \quad k \leq \sqrt{8/3}.$$

S3 OLCI OC matchup protocols, discussion

Current protocol

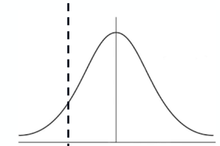
Pixel 'X' is considered outlier if:
|value@X - mean| < 1.5× IQR
Central value = median

S3A_OLCI_L2_IPF_OL__L2M.003_FR_EUMETSAT_standard_L2_5x5_Venise

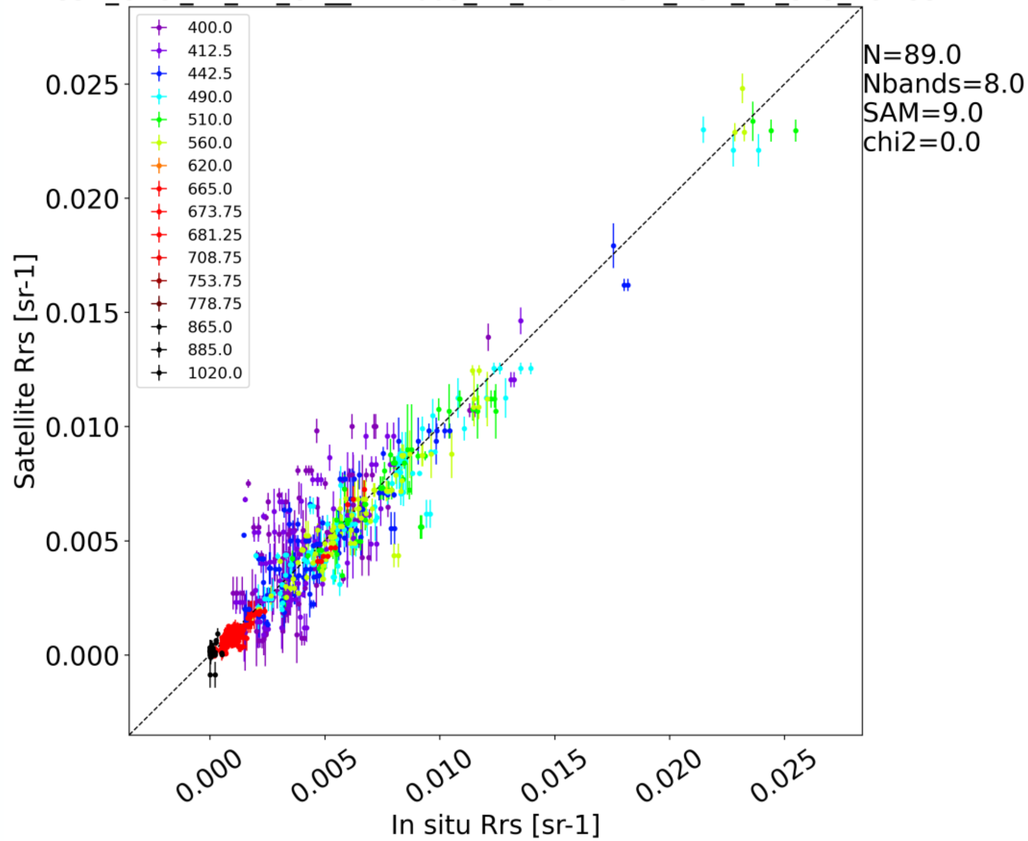


Robust "1"

Pixel 'X' is considered outlier if:
|value@X - median| < $\frac{10}{9}$× IQR
Central value = mean



S3A_OLCI_L2_IPF_OL__L2M.003_FR_EUMETSAT_new_L2_5x5_Venise



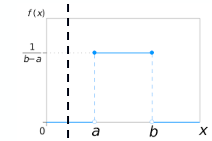
S3 OLCI OC matchup protocols, discussion

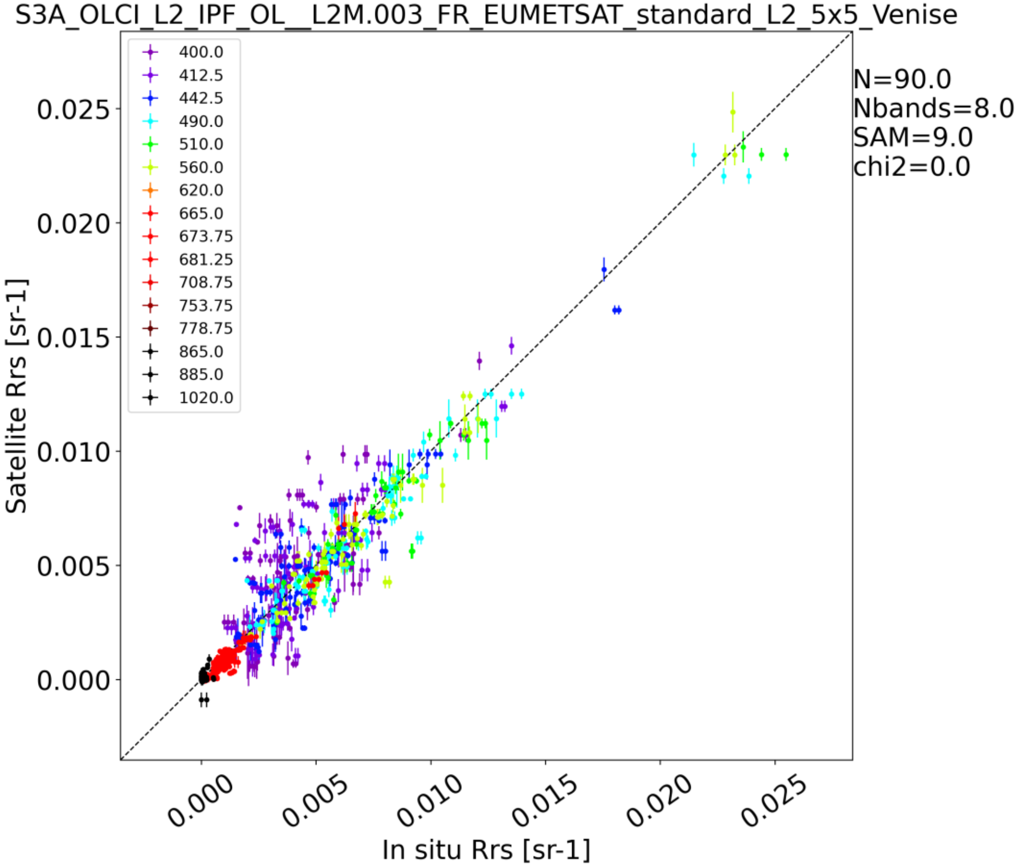
Current protocol

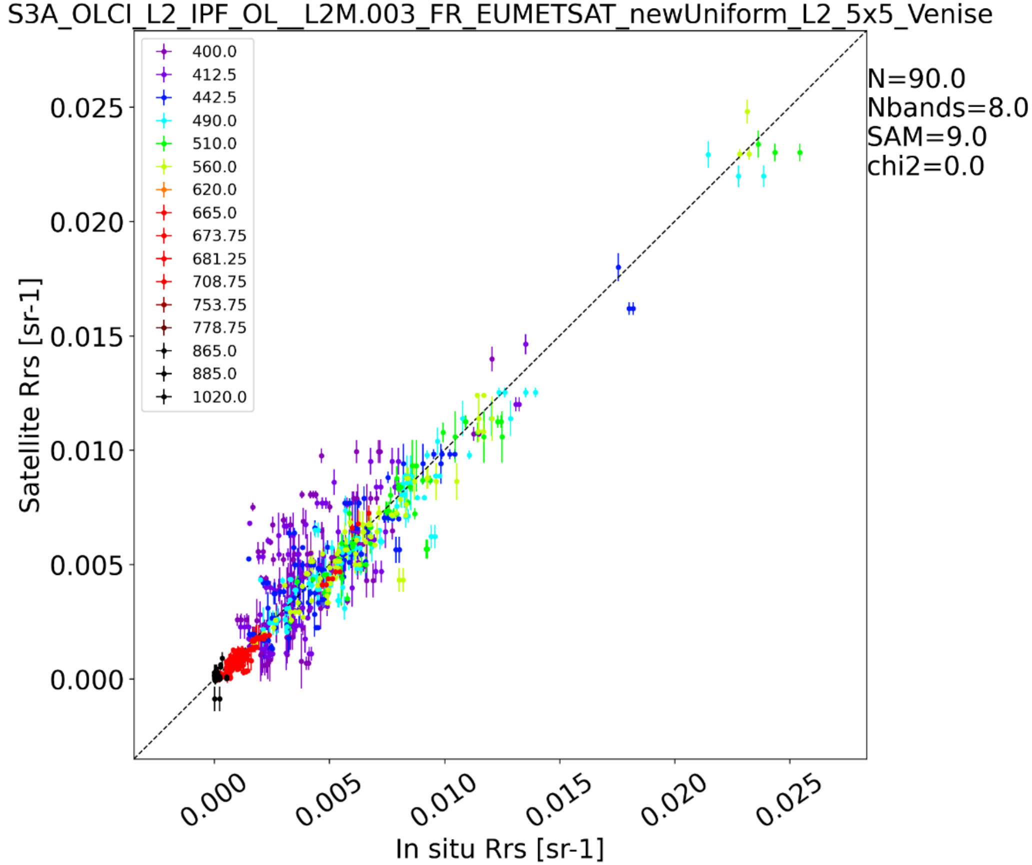
Pixel 'X' is considered outlier if:
$|\text{value@X} - \text{mean}| < 1.5 \times \text{IQR}$
Central value = median

Robust "2"

Pixel 'X' is considered outlier if:
$|\text{value@X} - \text{median}| < \frac{\sqrt{3}}{2} \times \text{IQR}$
Central value = mean





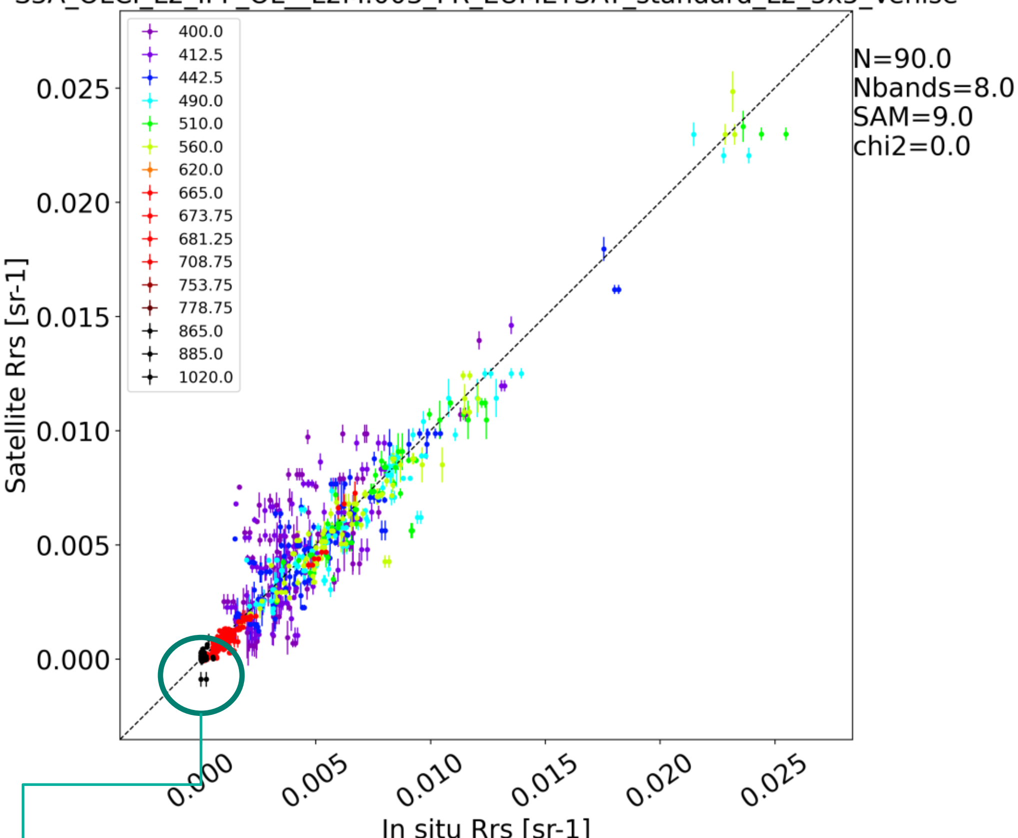


S3 OLCI OC matchup protocols, discussion

Current protocol

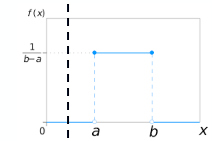
Pixel 'X' is considered outlier if:
$|\text{value@X} - \text{mean}| < 1.5\times \text{IQR}$
Central value = median

S3A_OLCI_L2_IPF_OL__L2M.003_FR_EUMETSAT_standard_L2_5x5_Venise

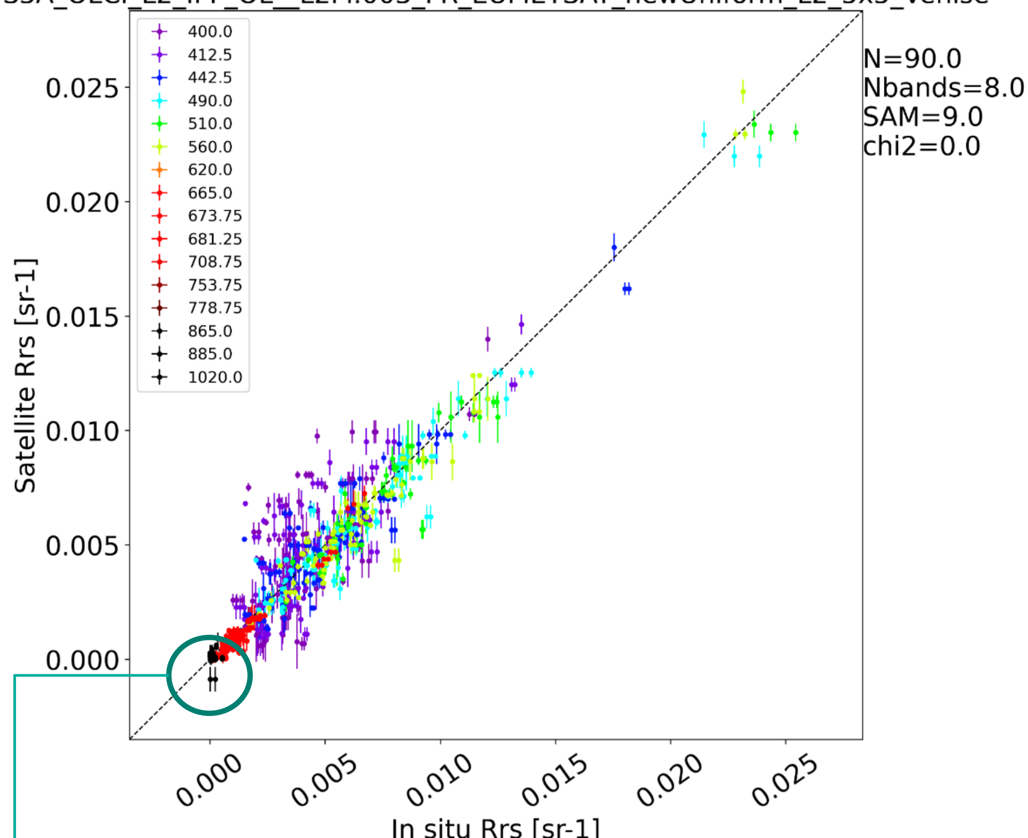


Robust "2"

Pixel 'X' is considered outlier if:
$|\text{value@X} - \text{median}| < \frac{\sqrt{3}}{2}\times \text{IQR}$
Central value = mean



S3A_OLCI_L2_IPF_OL__L2M.003_FR_EUMETSAT_newUniform_L2_5x5_Venise



No recommendation is given regarding space-time interpolation of quasi-simultaneous matchup pairs

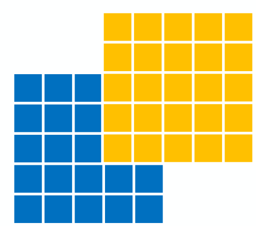
Time-space aggregation of "quasi-simultaneous" matchup pairs?

→ in situ (satellite) measurements corresponding to the **same** satellite (in situ) measurement

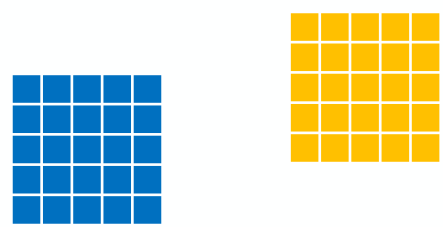
Case 1: Continuous measurements at fix location within valid time difference (±1hr)

Case 2: Continuous measurements on moving vessel within valid time difference (±1hr)

- Extraction windows intersect



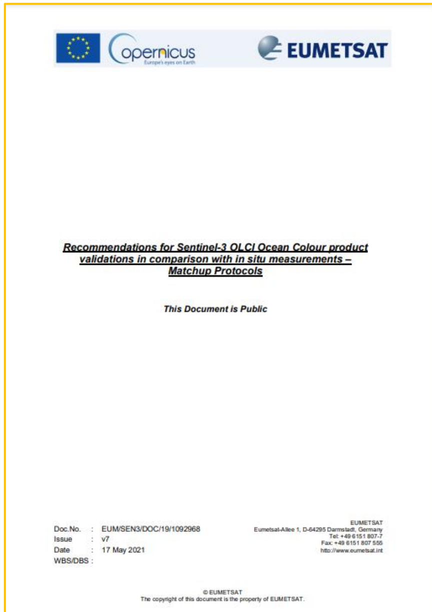
- Extraction windows do not intersect



S3 OLCI OC matchup protocols, discussion

PROGRAMME OF THE EUROPEAN UNION
copernicus Europe's eyes on Earth
EUMETSAT
co-funded with esa

- This document is based on consensus arrived by the S3VT-OC team and is the one that we share with all our users.
- Whenever relevant, the variants that we've seen in the last session (considered flags, time tolerance window, thresholds, spatial-temporal in situ - satellite interpolation) and the rationale behind these variations must be documented.
- We look forward for your feedback on this document (today or in the coming months) over which a updated version will be developed.



Recommendations for Sentinel-3 OLCI Ocean Colour product validations in comparison with in situ measurements – Matchup Protocols

This Document is Public

Doc.No. : EUM/SEN3/DOC/19/1092968
Issue : v7
Date : 17 May 2021
WBS/DBS :

EUMETSAT
Eumetsat-Allee 1, D-64295 Darmstadt, Germany
Tel: +49 6151 807-7
Fax: +49 6151 807 555
http://www.eumetsat.int

→ Available at: eumetsat.int/media/44087