Copernicus FICE 2024

Training on
In situ Ocean Colour Above-Water Radiometry towards Satellite Validation

ThoMaS - a Tool to generate Matchups of OC products with S-3/OLCI

Juan Ignacio Gossn
EUMETSAT
JuanIgnacio.Gossn@eumetsat.int

PROGRAMME OF THE EUROPEAN UNION
IMPLEMENTED BY EUMETSAT
FRM4SOC Phase-2
fiducial reference measurements for satellite ocean colour
CNR ISMAR
ISTITUTO DI SCIENZE MARINE

6-17 May 2024
Venice, Italy

EUMETSAT

Today's tour with ThoMaS – a Tool to perform Matchups with S3-OLCI

1. What's ThoMaS? Scope
2. Some background
3. Usage
4. Pre-requisites
5. Getting the code
6. Setting the environment
7. Required inputs
8. Run the code: examples

1. What's ThoMaS? Scope

ThoMaS is a toolkit developed to create **matchups** of bio-geophysical **insitu data** with **satellite ocean colour products** from **Sentinel-3 OLCI (S3/OLCI)**.

in SeaBASS format

Standard products from NASA's OBPG also supported
Others easily configurable, if netCDF or series of netCDFs

1. What's ThoMaS? Scope

After running **ThoMaS**, you will get:

→ **Insitu** data "transformed" to **match satellite** (spectral **convolution**, **band-shifting, BRDF…**).

→ **Satellite** data (L1B – TOA radiance - or L2 – BOA water reflectance) from **EUMETSAT Data Store** and **NASA OBPG** (reprocessed/operational) matching spatially/temporally your insitu.

→ **Extractions** of satellite data centred at lat/lon of insitu of user-defined size (3x3, 5x5..).

→ **Statistics of extractions** following EUMETSAT's or any user-defined **matchup protocol**.

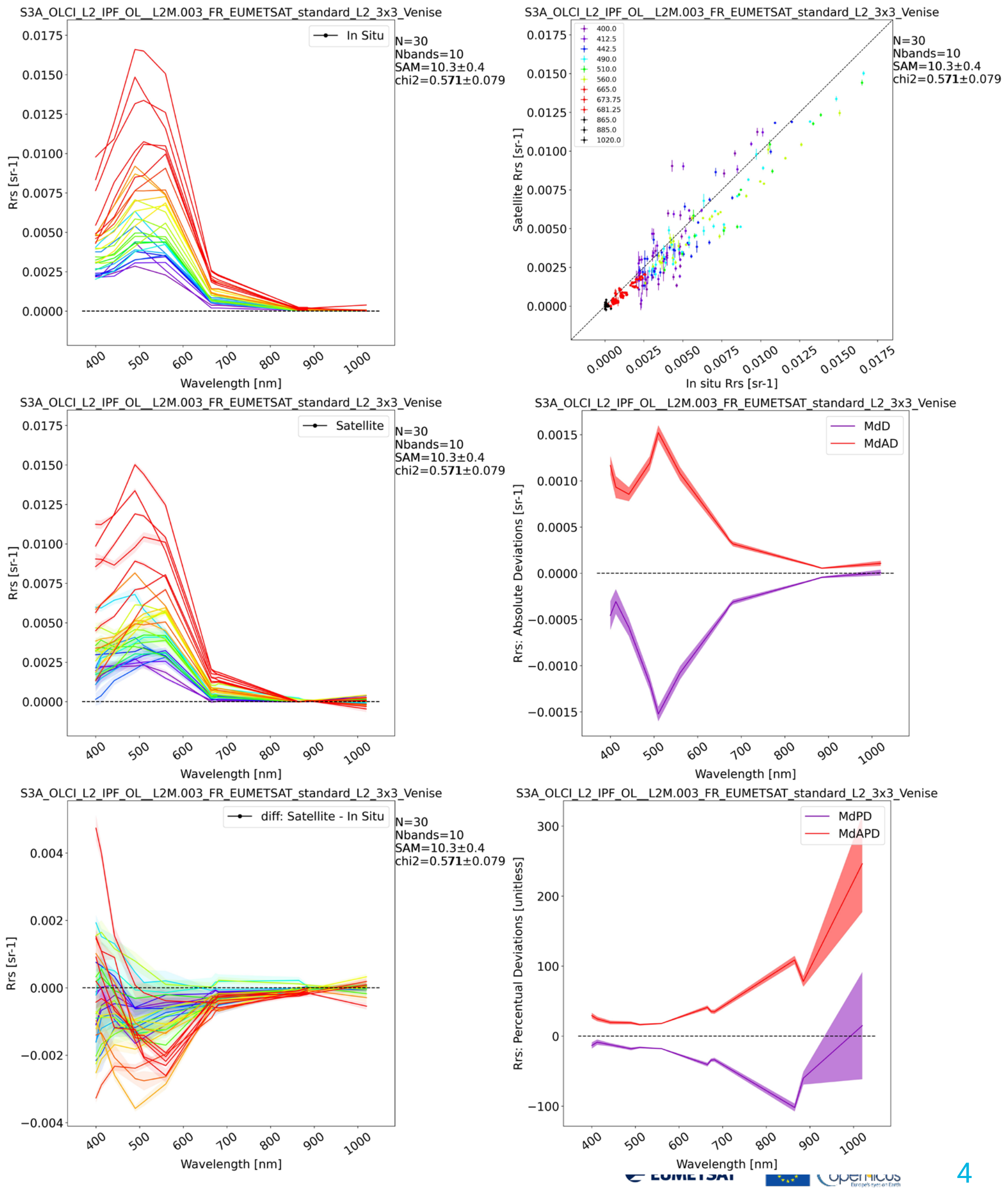
→ **Merging of simultaneous** (spatially-temporally) **insitu-satellite pairs**, **temporal interpolation**, and **statistics of matchups**.

→ Outputs:

→ **NetCDF 4** files: SatData, minifiles, Extraction Data Base files, In situ Data Base file, Matchup Data Base files.

→ **CSV**: summarizing satellite extraction statistics and matchup statistics.

→ **PNG**: Standardised output plots.



1. Motivation for having ThoMaS

1. Well documented, suited also for a first approach to the matchup exercise for those who are new to the intricacies of the matchup steps.
2. It's publicly available, free and open to scrutiny: it serves for the purpose of converging to a standard matchup practice.
3. It supports the most commonly used matchup protocols in the OC community.
   - → e.g. of existing ones: EUMETSAT's, Bailey & Werdell 2006, Zibordi 2009, Copernicus SVC_VIS
   - → Versatile: new matchup protocols can be easily added via configuration files.
   - → It contains an easy syntax to implement quality flags based on simple relations among products.
4. It deals (under some assumptions) with propagation of uncertainties to the performance metrics (using a Monte-Carlo approach).
5. Already supports some of the most commonly used OC satellite missions
   - → Currently supports Sentinel-3 (standard) L1B, L2, MODIS L2 (standard), VIIRS L2 (standard) and SeaHawk L2 (standard).
   - → Versatile: new types of satellite products can be easily added via configuration files (depending on mission, processor and processing baseline).

1. Some disclaimers: ThoMaS is just a baby!

1. The policy of garbage-in garbage-out applies to ThoMaS: you must know well your insitu data, the satellite product, and make sure that you are comparing “apples” to “apples”. Also, that your satellite pixels are sufficiently away from land, and rationally choose your extraction size, time difference tolerance, among many others.
2. ThoMaS does not deal with issues regarding spatial and temporal collocation of insitu and satellite data beyond very standard QC (e.g. a maximum time tolerance window, a choice of window size, and a simple temporal interpolation). This means that the problem of spatial and temporal autocorrelation of the Rrs signal (and any other OC product) is not yet dealt within ThoMaS.
3. ThoMaS cannot still compute match-up statistics of a given insitu-satellite set with varying satellite extraction sizes.
4. The uncertainty of the satellite component is only based on the inter-pixel variability (pixel-by-pixel uncertainties in the satellite component are still ignored in ThoMaS).
5. The uncertainties of the BRDF step are not propagated. Only Morel approach available. Lee 11 coming soon
6. ThoMaS won’t do an A/C of your satellite data!
7. Many other disclaimers (working on many of these ☺), but I hope it still proves useful!
8. Many people use it just for downloading the satellite data matching their in situ and performing the extractions.

- What is a match-up according to **chatGPT**?

j Hi ChatGPT

Hello! How can I assist you today?

j Could you define me "match-up" in the realm of Earth Observation?

In the context of Earth Observation, "match-up" refers to the process of comparing satellite data to ground-based measurements or other reference data sources to assess the accuracy and quality of the satellite observations.

Match-up analysis is typically used to validate satellite measurements of various geophysical parameters, such as surface temperature, vegetation cover, sea surface temperature, and atmospheric aerosols, among others. The ground-based measurements can be obtained from various sources, such as ground-based sensors, buoys, weather stations, or other types of in-situ measurements.

→ Ocean colour 👀 ←

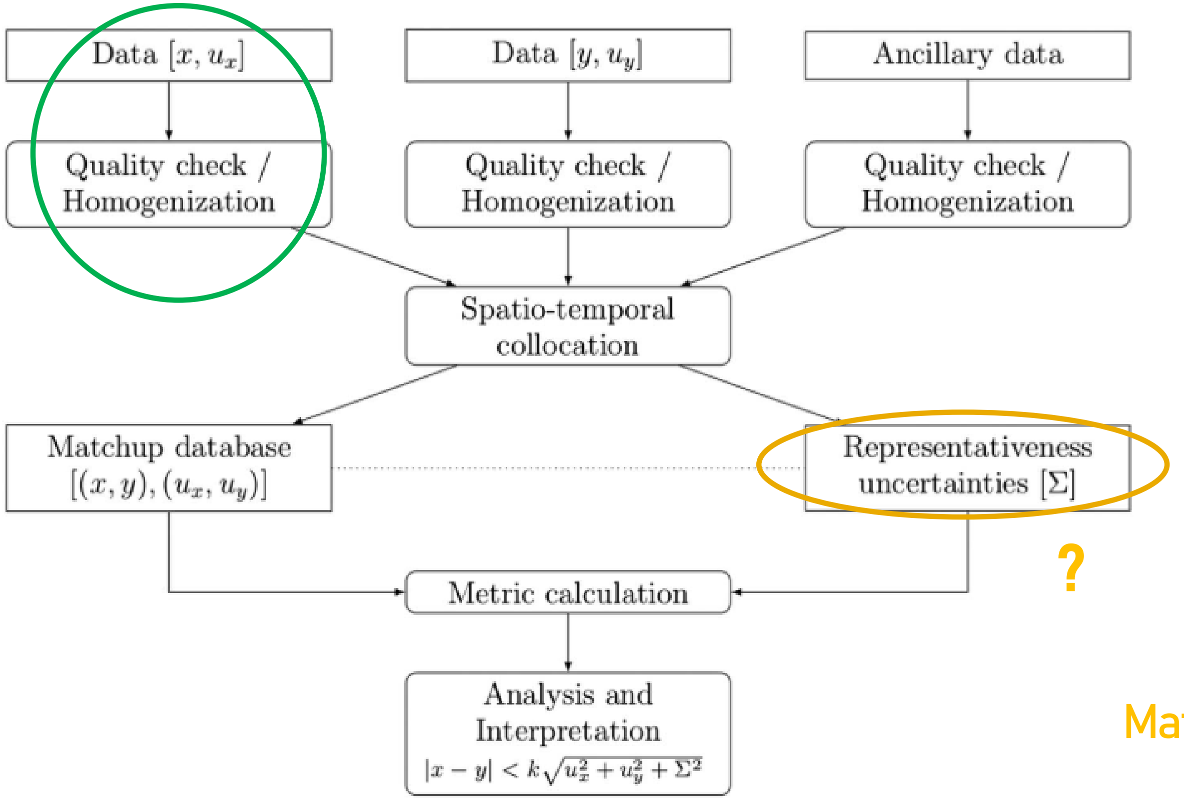
Of course we have much more to define… and take care of…

2. Some background: match-ups

- The EO data are in practice rarely fully traceable, for instance, because fundamental calibrations done in the laboratory prelaunch **cannot be repeated in space**.
- Consequently the **comparison against reference measurements** in a validation exercise is often the only way to link the EO data back to an agreed standard.

[Loew et al. 2017]

Most of our attention in this course is here



x = "reference" measurement (in situ)
ux = insitu uncertainty

y = EO measurement (satellite)
uy = satellite uncertainty

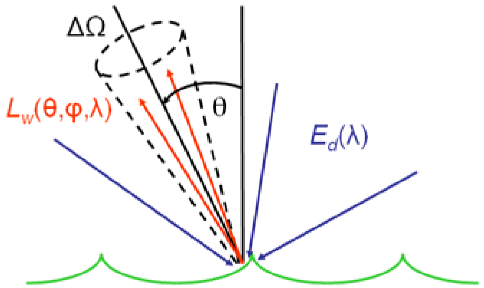
Representativeness
The extent to which a set of measurements taken in a given space-time domain reflect the actual conditions in the same or different space-time domain
[Nappo et al., 1982]

Matchup protocols in OC deal (still sub-optimally) with this...

2. Some background: match-ups

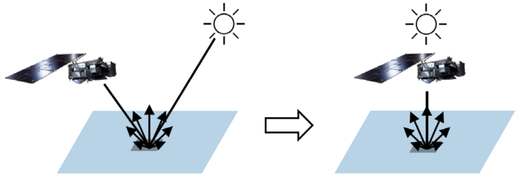
Definition of Rrs

$$R_{rs}(\theta, \phi, \lambda) \equiv \frac{L_w(\text{in air}, \theta, \phi, \lambda)}{E_d(\text{in air}, \lambda)} \qquad (\text{sr}^{-1}) \,.$$



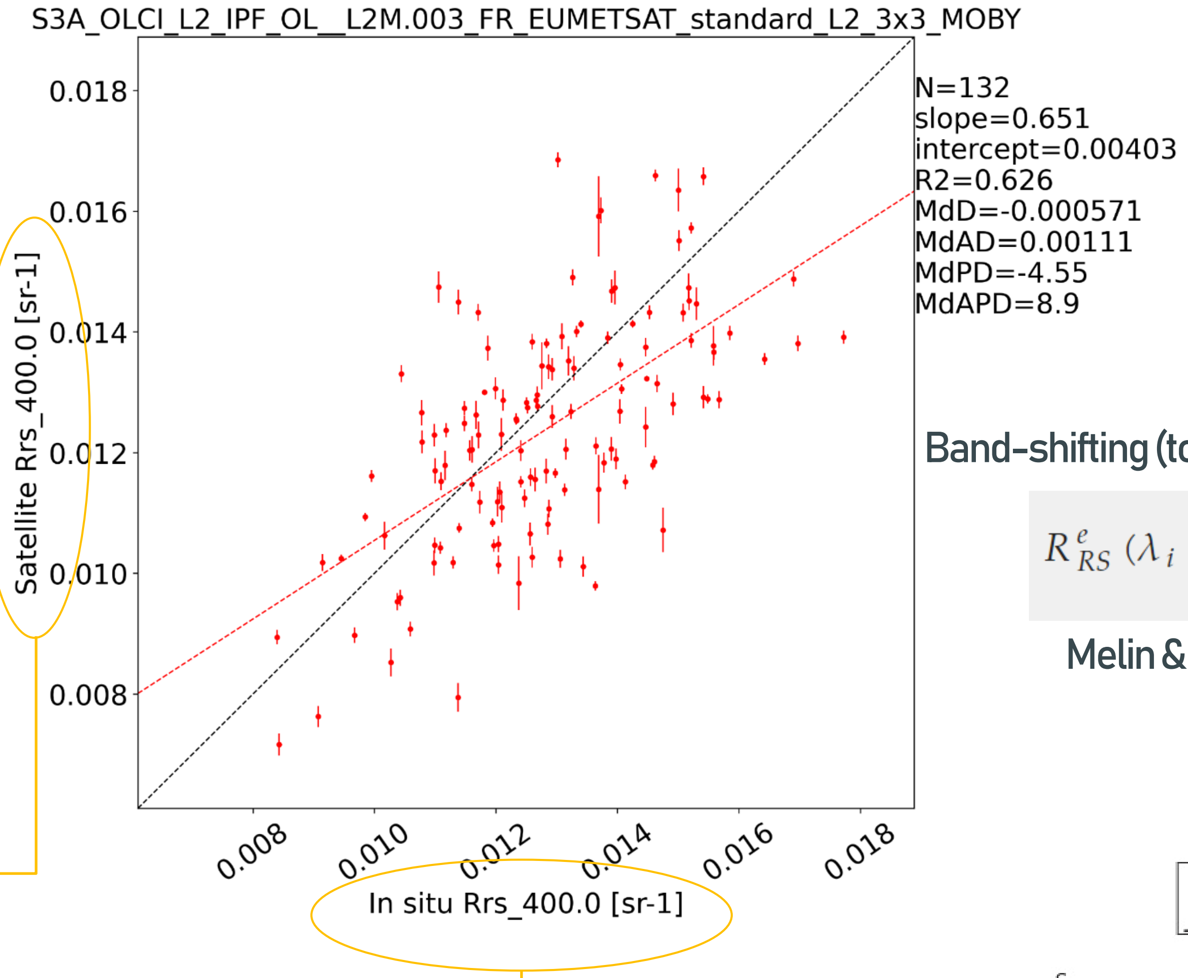
OO Web Book, Mobley, Boss & Roesler

BRDF correction:



D'Alimonte et al.

Morel et al. 2002 supported in ThoMaS

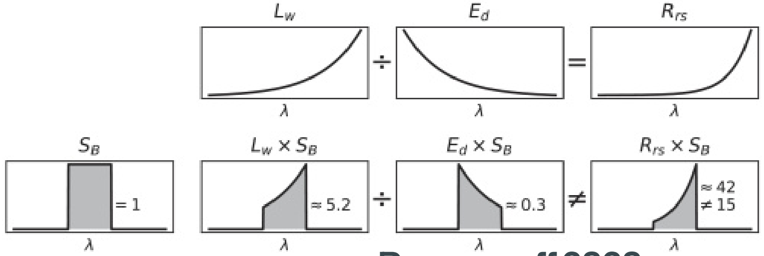


Band-shifting (to pair multispectral to multispectral)

$$R^{e}_{RS}(\lambda_i \rightarrow \lambda_t) = R^{f}_{RS}(\lambda_t)\frac{R_{RS}(\lambda_i)}{R^{f}_{RS}(\lambda_i)}$$

Melin & Sclep 2015 supported in ThoMaS

Spectral convolution



Burggraaff 2020

Are we comparing 🍎 with 🍎?

→ What is the definition of Rrs?

→ Are these two compatible "spectrally"? → convolution/band-shifting

→ Are these two compatible "directionally"? → BRDF correction

Multi-spectral

Hyper-spectral

Multi- or hyper-spectral?

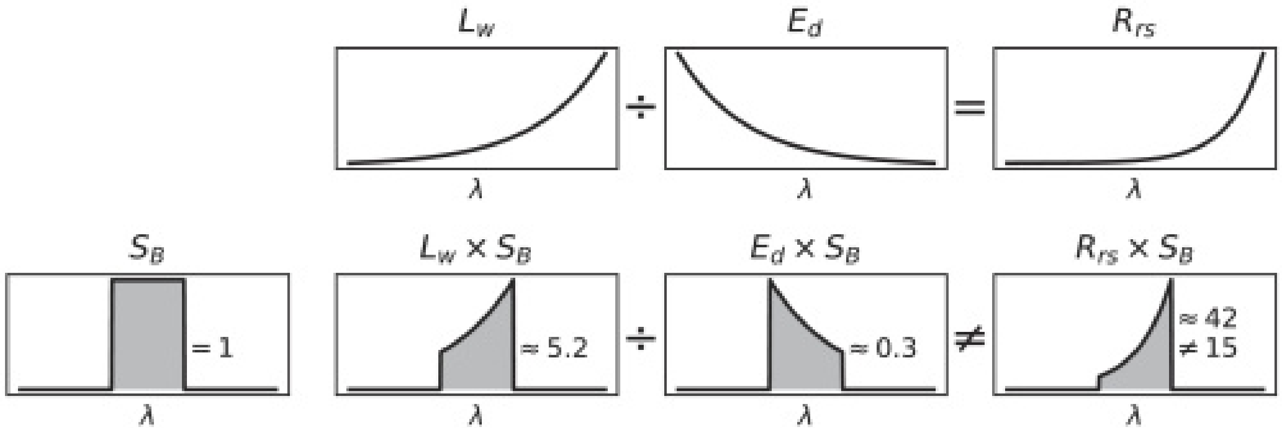
Max. spectral sampling < 8 nm?

Band-shifting (to pair multispectral to multispectral)

$$R^{e}_{RS}(\lambda_i \rightarrow \lambda_t) = R^{f}_{RS}(\lambda_t)\frac{R_{RS}(\lambda_i)}{R^{f}_{RS}(\lambda_i)}$$

Melin & Sclep 2015 supported in ThoMaS

Spectral convolution



Burggraaff 2020 approach supported in ThoMaS

- If band-shifting fails (QAA fails), take correction factor as 1 but flag accordingly (spectralMatching_Fail = ON)
- Uncertainty not propagated through still. Monte-Carlo approach is likely needed

- If convolution fails (SRF covered less than 97.5%) by insitu spectral grid, then NaN and spectralMatching_Fail = ON
- Some assessment of uncertainty supported in ThoMaS

Uncertainty propagation after convolution:

Convolution: measurement equation to obtain Rrs at (multispectral) band B [SRFs and Rrs to a common wavelength grid]

$$Rrs[\lambda_B] = \sum_{i=1}^{N} Rrs[\lambda_i] . \frac{SRF_B[\lambda_i]}{\sum_{i=1}^{N} SRF_B[\lambda_i]} \quad \textbf{[1]}$$

→ **Fully random case, I assume** $u(SRF_B) = 0$ **and** $Cov\left(Rrs[\lambda_i], Rrs[\lambda_j]\right) = 0 \; \forall \; i \neq j$**:**

$$u(Rrs[\lambda_B])^2 = \sum_{i=1}^{N} u(Rrs[\lambda_i])^2 . \left(\frac{SRF_B[\lambda_i]}{\sum_{i=1}^{N} SRF_B[\lambda_i]}\right)^2 \quad \textbf{[2]}$$

...Which is the sum in quadrature.

→ **Fully systematic case: asssume maximum correlation between hyperspectral bands, still** $u(SRF_B) = 0$

$$Cov\left(Rrs[\lambda_i], Rrs[\lambda_j]\right) = u(Rrs[\lambda_i]) . u\left(Rrs[\lambda_j]\right), \; \forall \; i, j \quad \textbf{[3]}$$

Then,

This is what is done in ThoMaS ..... which is like a standard convolution

$$u(Rrs[\lambda_B])^2 = \sum_{i=1}^{N} u(Rrs[\lambda_i])^2 . \left(\frac{SRF_B[\lambda_i]}{\sum_{i=1}^{N} SRF_B[\lambda_i]}\right)^2 + 2. \sum_{i<j} u(Rrs[\lambda_i]) . u\left(Rrs[\lambda_j]\right) . \left(\frac{SRF_B[\lambda_i]}{\sum_{i=1}^{N} SRF_B[\lambda_i]}\right) . \left(\frac{SRF_B[\lambda_i]}{\sum_{i=1}^{N} SRF_B[\lambda_i]}\right) =$$

$$\left(\sum_{i=1}^{N} u(Rrs[\lambda_i]) . \frac{SRF_B[\lambda_i]}{\sum_{i=1}^{N} SRF_B[\lambda_i]}\right)^2 \textbf{[4]}$$

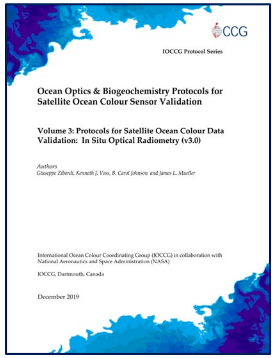
2. Some background: match-ups

copernicus.eumetsat.int

Quality of insitu

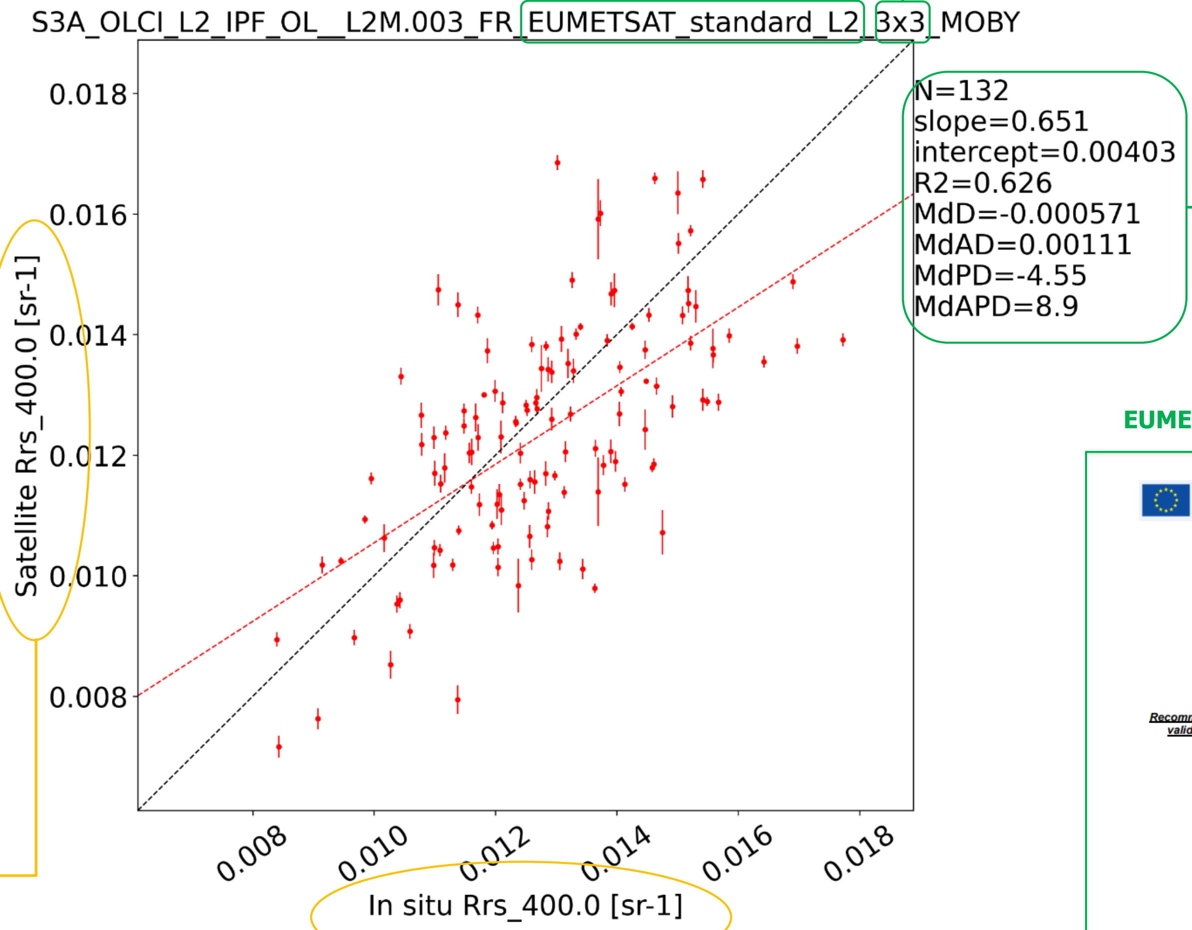
ThoMaS still **<u>does not</u>** consider any quality flag to process insitu data...

fiducial reference measurements for satellite ocean colour



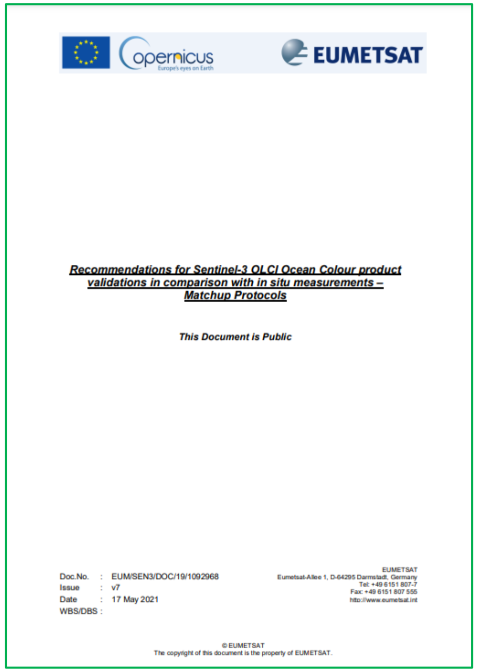
Define your extraction statistics!

Define your extraction window size!



Define the matchup statistics!

EUMETSAT's Matchup Protocols



<u>Are we comparing 🍎 with 🍎?</u>

→ Are the insitu measurements of sufficient quality?

→ Are insitu and satellite measurements temporally-spatially comparable?

→ What value (and uncertainty) shall I extract from the satellite data?

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

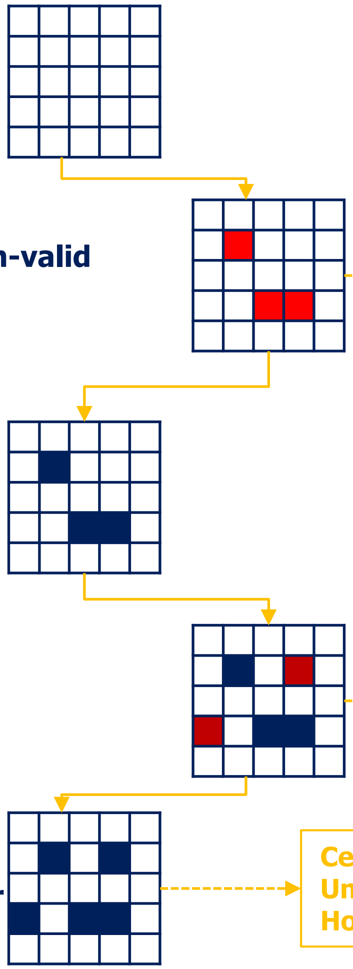
Window size recommended: 5x5 or 3x3
Tolerable insitu-satellite time difference: 1 hr or 3 hrs

Set BFOR: 5x5 window

Detection of non-valid pixels (flagged pixels)

Set BOR: without flagged pixels

Set final: without flagged & outlier



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

Macropixel is discarded if:

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

Pixel 'X' is considered outlier if:

$|value@X - \mu_{BOR}| < 1.5\sigma_{BOR}$

Macropixel is discarded if:

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

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

μ → Mean
σ → Standard deviation

ThoMaS can be run with many other extraction protocols, that you define

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

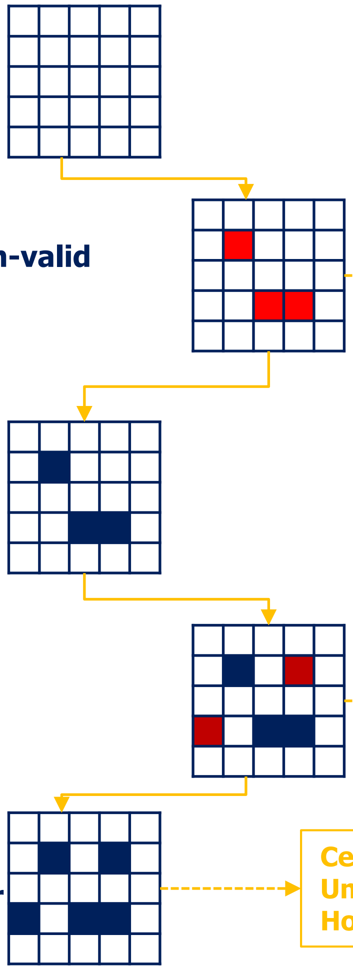
Window size recommended: 5x5 or 3x3
Tolerable insitu-satellite time difference: 1hr or 3hrs

Set BFOR: 5x5 window

Detection of non-valid pixels (flagged pixels)

Set BOR: without flagged pixels

Set final: without flagged & outlier



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

ATMFAIL, LAND, HIGLINT, HILT, HISATZEN, STRAYLIGHT, CLDICE, COCCOLITH, HISOLZEN, LOWLW, CHLFAIL, NAVWARN, MAXAERITER, CHLWARN, ATMWARN, SEAICE, NAVFAIL, ABSAER, MODGLINT

Macropixel is discarded if:

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

Pixel 'X' is considered outlier if:

$|value@X - \mu_{BOR}| < 1.5\sigma_{BOR}$

Macropixel is discarded if:

Median[CV(Rrs(410-551));CV(AOT(869))]>15%

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

μ → Mean
σ → Standard deviation

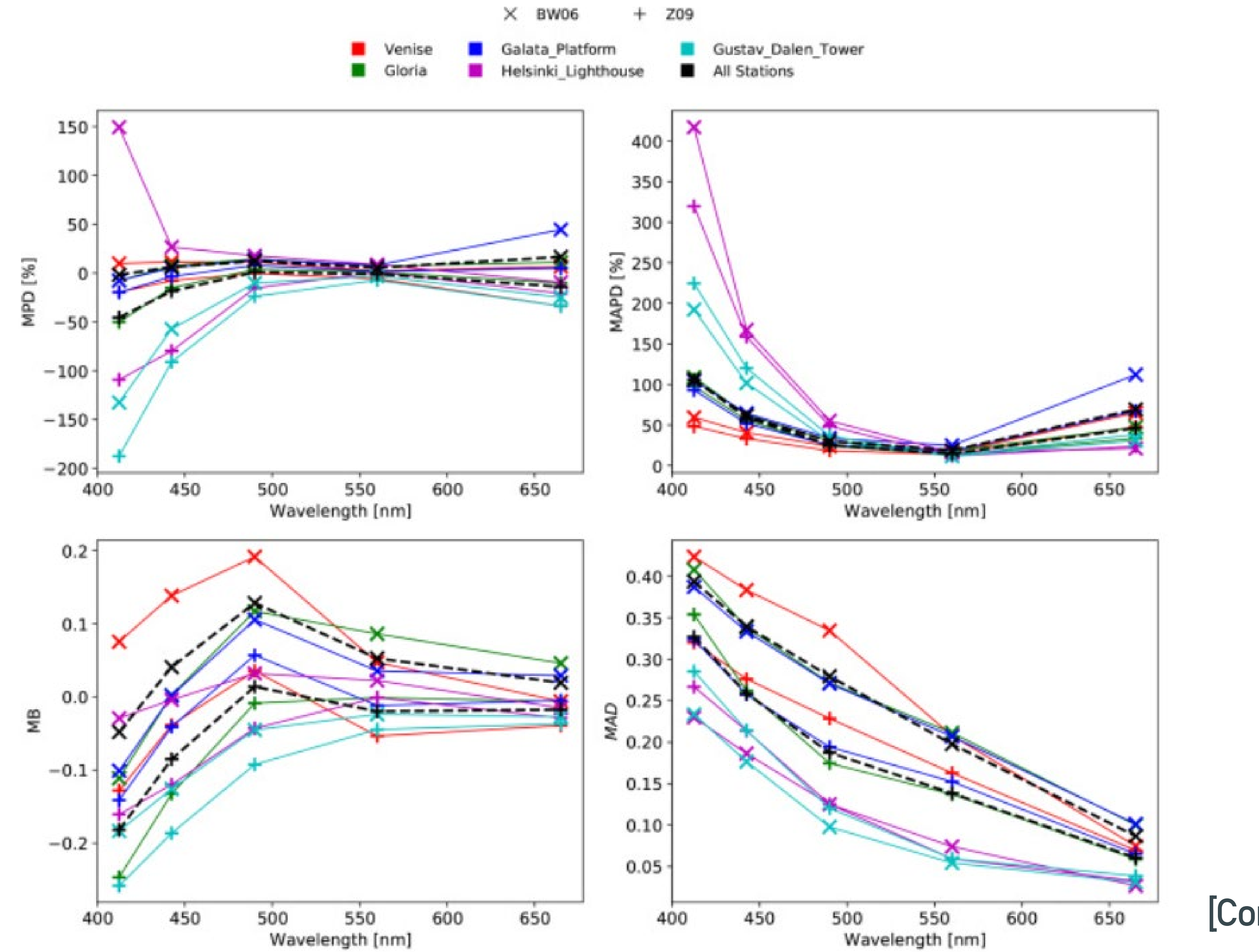
ThoMaS can be run with many other extraction protocols, that you define

2. Some background: match-ups: what protocol to use?

copernicus.eumetsat.int

Should we care abo

YES



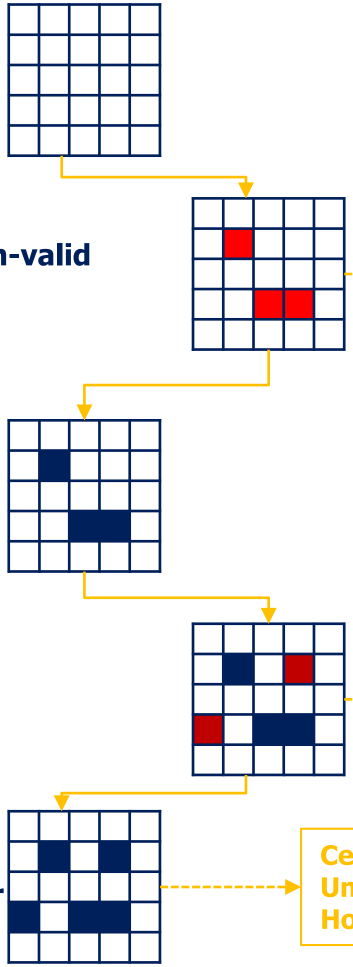
[Concha et al. 2021]

Current EUMETSATs matchup protocol

EUMETSATs Matchup Protocols: extraction of statistics at macropixel level

Window size recommended: 5x5 or 3x3
Tolerable insitu-satellite time difference: 1 hr or 3 hrs

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

Macropixel is discarded if:

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

Set BOR: without flagged pixels

Pixel X is considered outlier if:

$|value@X - \mu_{BOR}| < 1.5\sigma_{BOR}$

Set final: without flagged & outlier

Central value: $median_{final}$
Uncertainty measure: σ_{final}
Homogeneity measure: CV_{final}

Macropixel is discarded if:

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

μ → Mean
σ → Standard deviation

Current EUMETSATs matchup protocol: proposed changes (1)

EUMETSATs Matchup Protocols: extraction of statistics at macropixel level

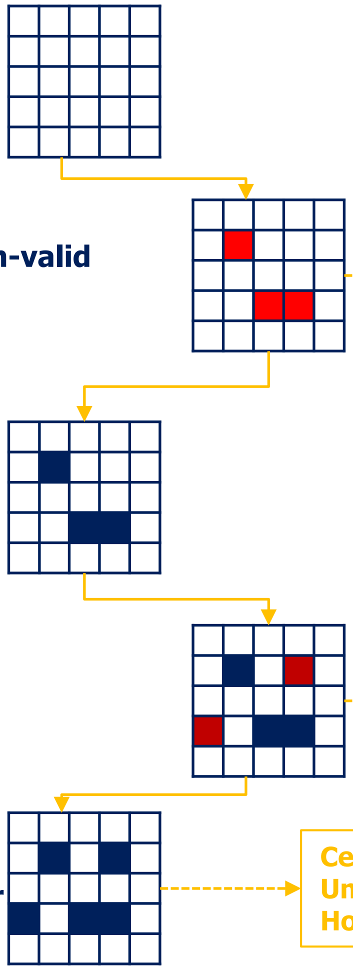
Window size recommended: 5x5 or 3x3
Tolerable insitu-satellite time difference: 1 hr or 3 hrs

Set BFOR: 5x5 window

Detection of non-valid pixels (flagged pixels)

Set BOR: without flagged pixels

Set final: without flagged & outlier



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

Macropixel is discarded if:

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

Pixel X is considered outlier if:

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

Macropixel is discarded if:

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

Central value: ~~$median_{final}$~~ μ_{final}
Uncertainty measure: σ_{final}
Homogeneity measure: CV_{final}

μ → Mean
σ → Standard deviation

Current EUMETSATs matchup protocol: proposed changes (2)

EUMETSATs Matchup Protocols: extraction of statistics at macropixel level

Window size recommended: 5x5 or 3x3
Tolerable insitu-satellite time difference: 1hr or 3hrs

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

Set BOR: without flagged pixels

Macropixel is discarded if:

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

Pixel X is considered outlier if:

|value@X - ~~μ_{BOR}~~ $median_{BOR}$| < ~~$1.5\sigma_{BOR}$~~ $0.86 IQR_{BOR}$

Set final: without flagged & outlier

Central value: ~~$median_{final}$~~ μ_{final}
Uncertainty measure: σ_{final}
Homogeneity measure: CV_{final}

Macropixel is discarded if:

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

μ → Mean
σ → Standard deviation

Testing protocol differences…

- In situ data used: AERONET-OC, Level 2.0
- Stations: Casablanca_Platform, Gloria,Helsinki_Lighthouse, LISCO,Section-7_Platform, USC_SEAPRISM, Galata_Platform, Gustav_Dalen_Tower, Lake_Erie, Palgrunden, Socheongcho, Venise
- A total of 2139 matchups

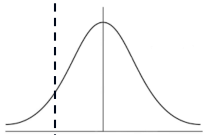
Comparing outlier detection, pixel by pixel

Proposed "1"

Pixel X is considered outlier if:

|value@X - median| < $\frac{\mathbf{10}}{\mathbf{9}}$ × IQR

Central value = mean



Current protocol

Pixel X is considered outlier if:

|value@X - mean| < $\mathbf{1.5}$ × σ

Central value = median

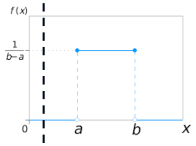
Current protocol \ Proposed "1"	NOT OUTLIER	OUTLIER
NOT OUTLIER	88.4%	3.27%
OUTLIER	1.35%	7.02%

Proposed "2"

Pixel X is considered outlier if:

|value@X - median| < $\frac{\sqrt{3}}{2}$ × IQR

Central value = mean



Current protocol

Pixel X is considered outlier if:

|value@X - mean| < $\mathbf{1.5}$ × σ

Central value = median

Current protocol \ Proposed "2"	NOT OUTLIER	OUTLIER
NOT OUTLIER	83.5%	8.06%
OUTLIER	2.32%	8.14%

Comparing screened macropixels

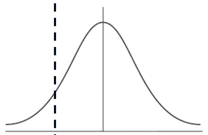
copernicus.eumetsat.int

Proposed "1"

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

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

Central value = mean



Current protocol

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

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

Central value = median

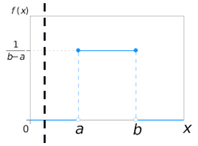
	PASS	FAIL
PASS	1359	9
FAIL	3	768

Proposed "2"

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

$|\text{value@X - median}| < \frac{\sqrt{3}}{2} \times \text{IQR}$

Central value = mean



Current protocol

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

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

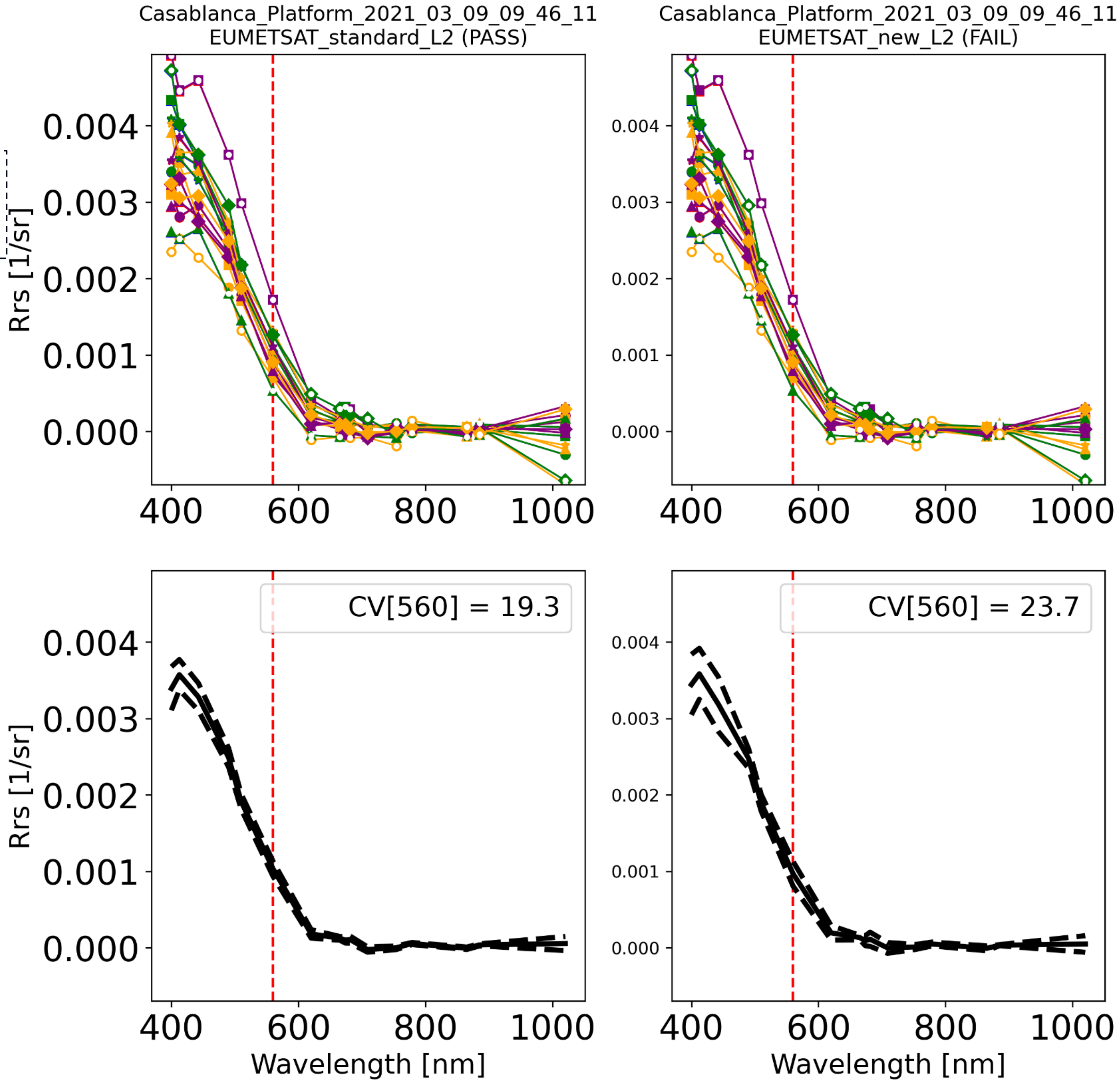
Central value = median

	PASS	FAIL
PASS	1367	1
FAIL	9	762

Current protocol

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

- **White** inner circles: outliers
- **Black** inner circles: flags



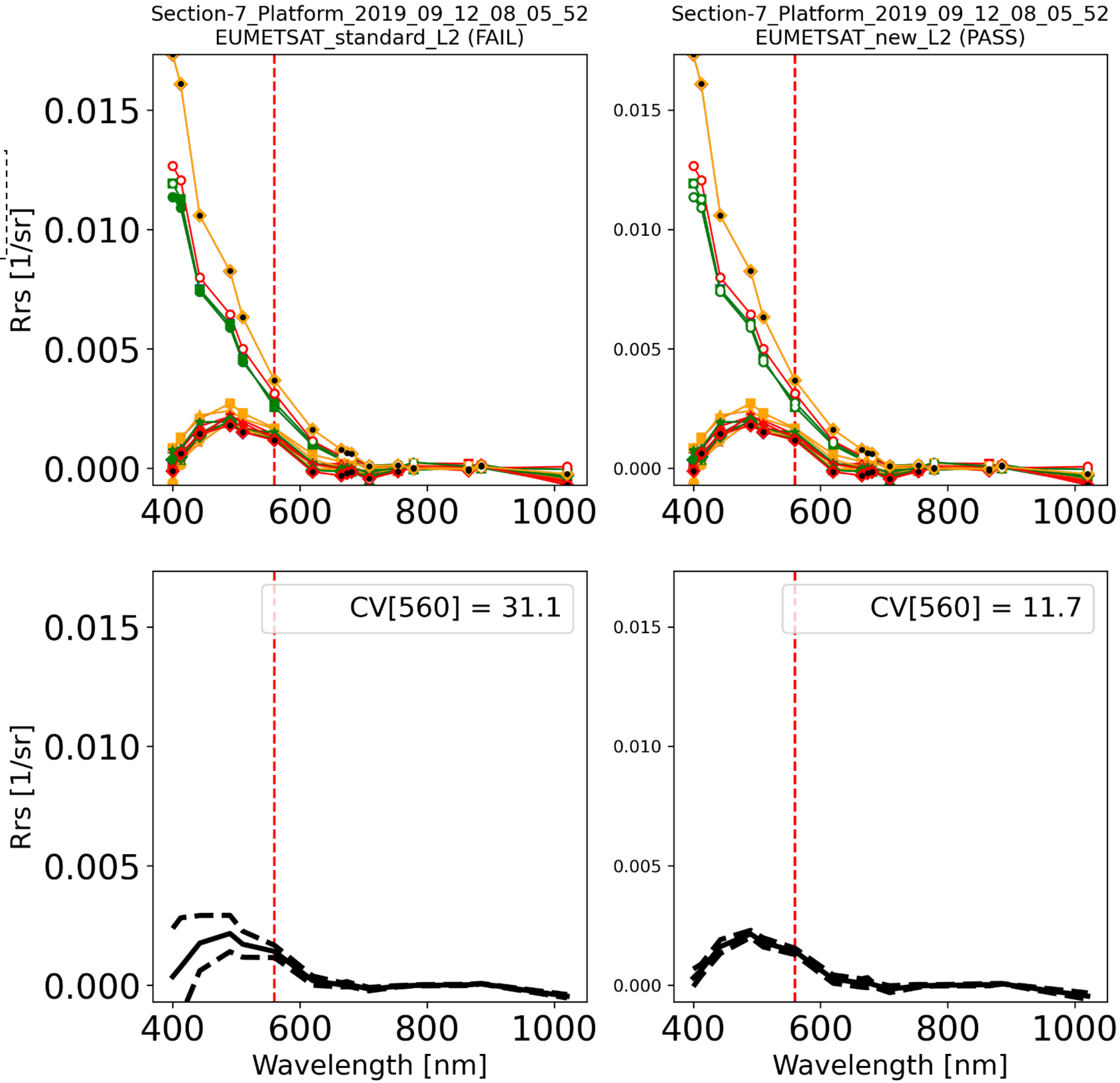
Proposed "1"

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

Current protocol

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

- **White** inner circles: outliers
- **Black** inner circles: flags



Proposed "1"

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

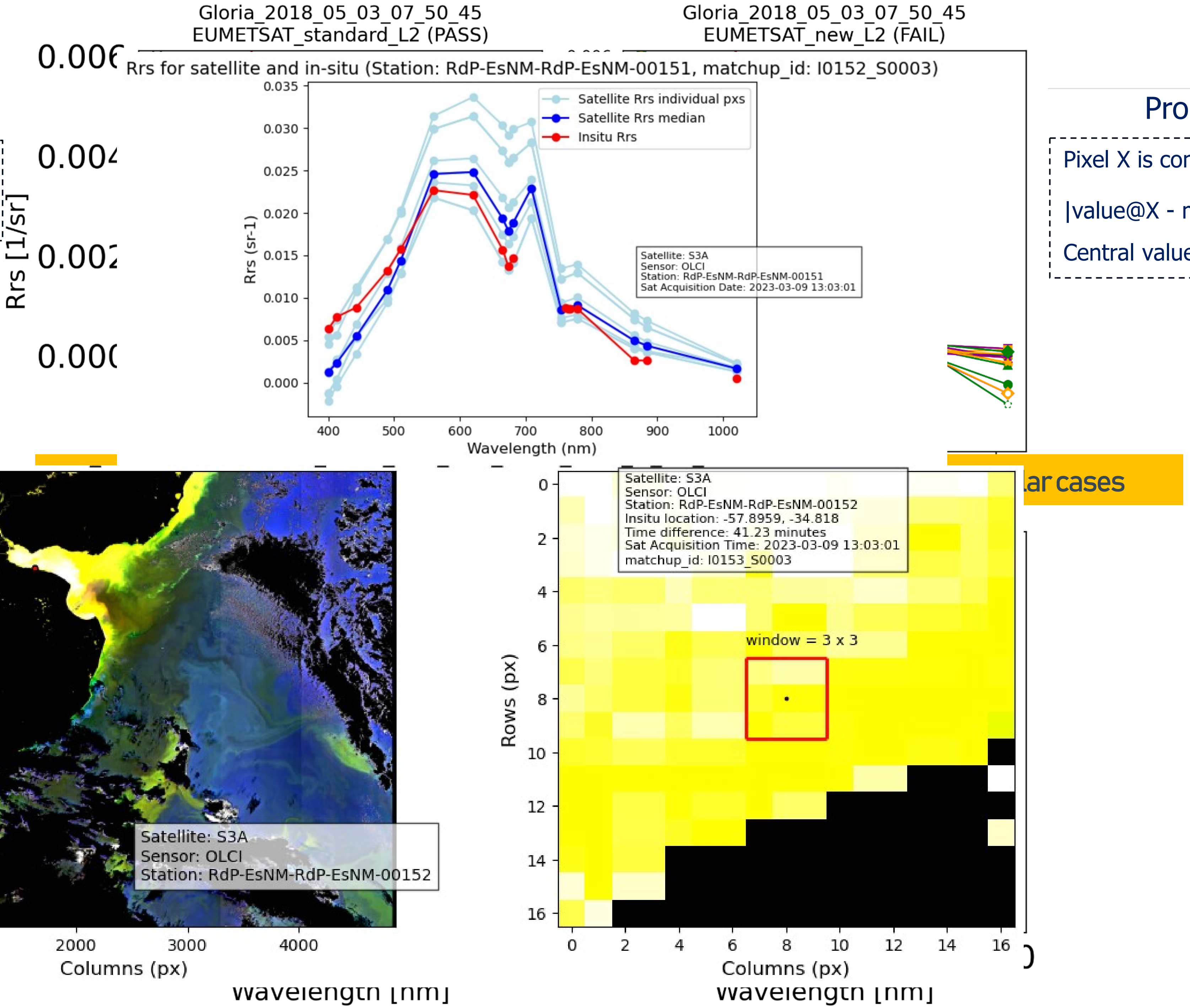
Current protocol

Pixel X is considered outlier if:
|value@X - mean| < $1.5 \times \sigma$
Central value = median

- **White** inner circles: outliers
- **Black** inner circles: flags

Proposed "1"

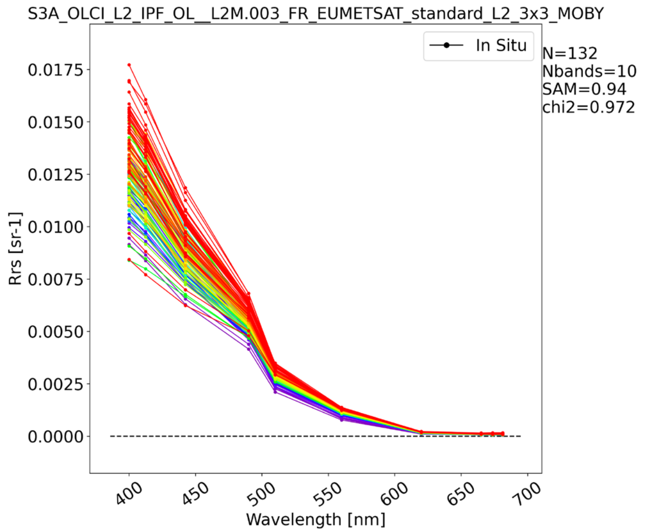
Pixel X is considered outlier if:
$|value@X - median| < \frac{10}{9} \times IQR$
Central value = mean

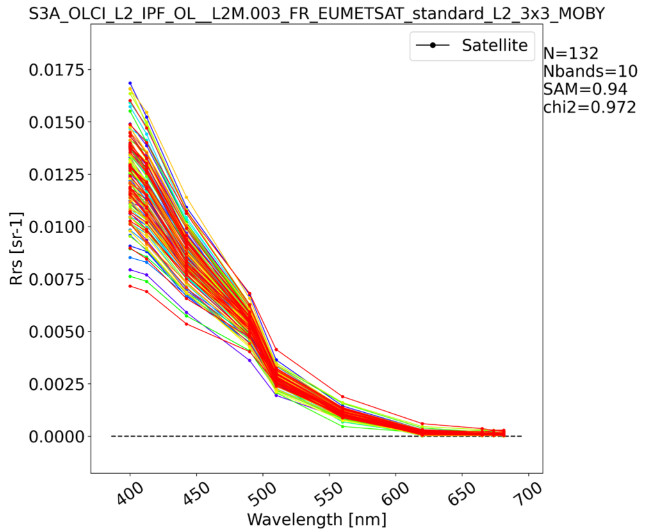


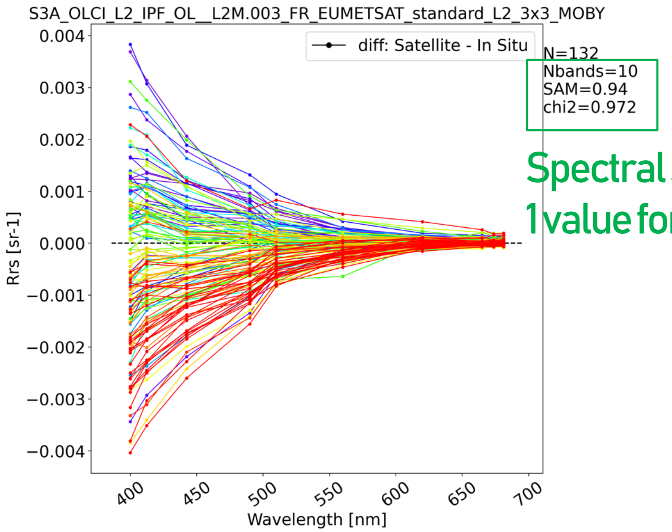
copernicus@eumetsat.int

2. Some background: match-ups

Band-by-band plots and statistics are often not sufficient...

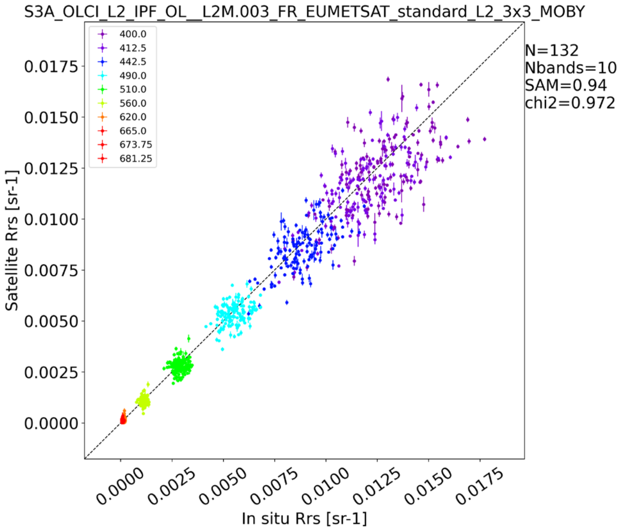


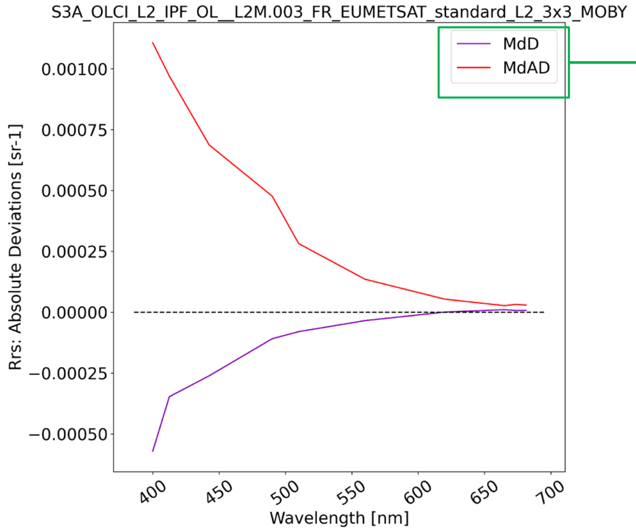


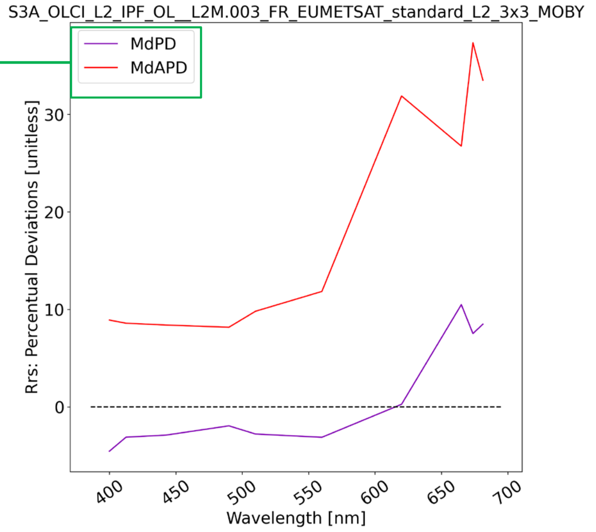


Spectral statistics
1 value for the whole set

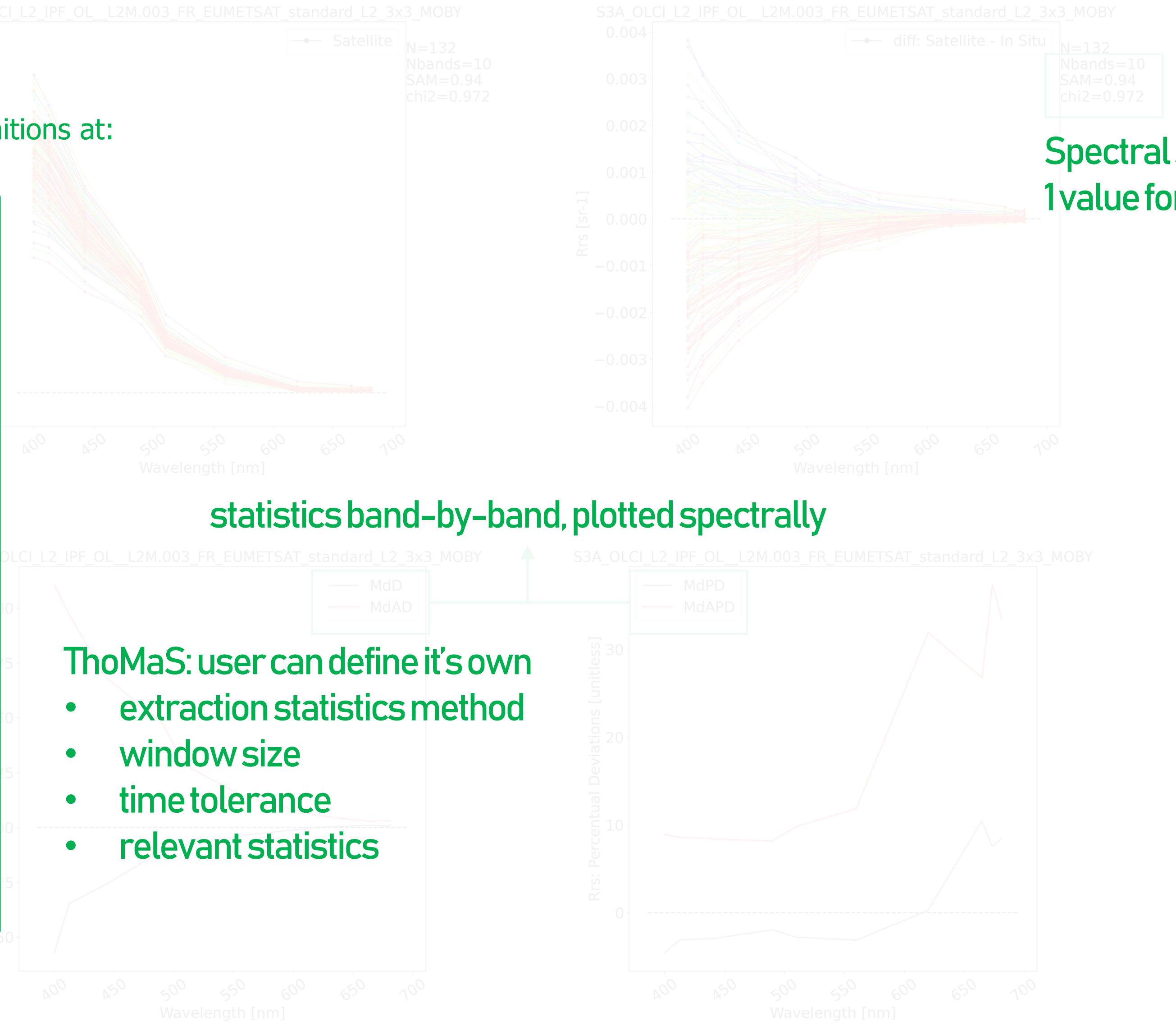
statistics band-by-band, plotted spectrally



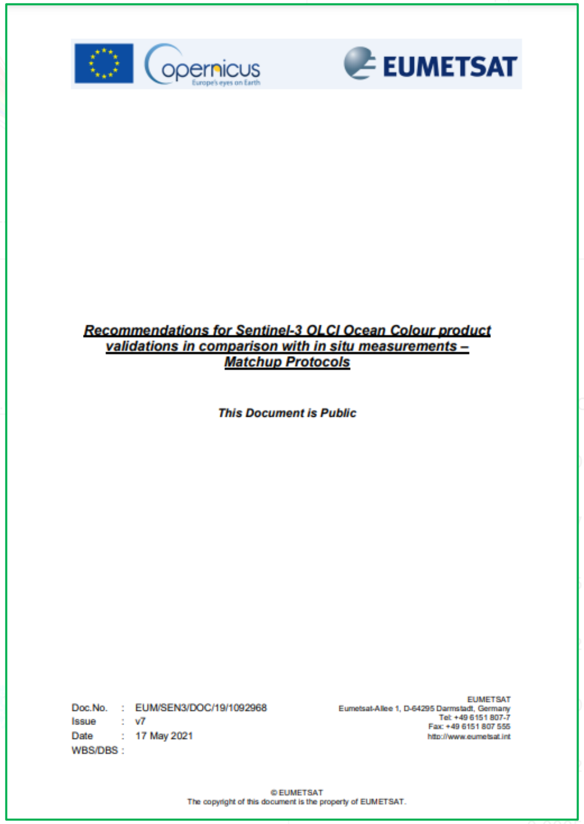




Band-by-band plots and statistics are often not sufficient…



Check recommended statistics and definitions at:
EUMETSAT's Matchup Protocols



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

© EUMETSAT
The copyright of this document is the property of EUMETSAT.

Spectral statistics
1 value for the whole set

statistics band-by-band, plotted spectrally

ThoMaS: user can define it's own
- extraction statistics method
- window size
- time tolerance
- relevant statistics

2. Some background: match-ups, performance metrics used

Mean Difference →

$$\frac{1}{n}\sum_{i=1}^{n}\{Rrs(\lambda)_{OLCI,i} - Rrs(\lambda)_{in\ situ,i}\}$$

Mean Percent Difference →

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{Rrs(\lambda)_{OLCI,i} - Rrs(\lambda)_{in\ situ,i}}{Rrs(\lambda)_{in\ situ,i}}\right\}$$

Mean Absolute Difference →

$$\frac{1}{n}\sum_{i=1}^{n}\{|Rrs(\lambda)_{OLCI,i} - Rrs(\lambda)_{in\ situ,i}|\}$$

Mean Absolute Percent Difference →

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\left|\frac{Rrs(\lambda)_{OLCI,i} - Rrs(\lambda)_{in\ situ,i}}{Rrs(\lambda)_{in\ situ,i}}\right|\right\}$$

Log-based Mean Absolute Difference →

$$LogMAD_{\lambda} = 10^{\sum_{i=1}^{N}|log_{10}(Rrs(\lambda)_{OLCI,i}) - log_{10}(Rrs(\lambda)_{in\ situ,i})|}$$

Following recommendation by Seegers et al. 2018

2. Some background: match-ups, performance metrics used

Median Difference →

$$\underset{1\le i\le N}{\text{median}}\left\{Rrs(\lambda)_{OLCI,i} - Rrs(\lambda)_{in\ situ,i}\right\}$$

Median Percent Difference →

$$\underset{1\le i\le N}{\text{median}}\left\{\frac{Rrs(\lambda)_{OLCI,i} - Rrs(\lambda)_{in\ situ,i}}{Rrs(\lambda)_{in\ situ,i}}\right\}$$

Median Absolute Difference →

$$\underset{1\le i\le N}{\text{median}}\left\{\left|Rrs(\lambda)_{OLCI,i} - Rrs(\lambda)_{in\ situ,i}\right|\right\}$$

Median Absolute Percent Difference →

$$\underset{1\le i\le N}{\text{median}}\left\{\left|\frac{Rrs(\lambda)_{OLCI,i} - Rrs(\lambda)_{in\ situ,i}}{Rrs(\lambda)_{in\ situ,i}}\right|\right\}$$

Spectral Angle Mapper →

$$SAM = \frac{1}{N}\sum_{i=1}^{N}\left(acos\left(\frac{\langle Rrs_{in\ situ,i}, Rrs_{OLCI,i} \rangle}{\left\| Rrs_{in\ situ,i} \right\| \left\| Rrs_{OLCI,i} \right\|} \right) \right)$$

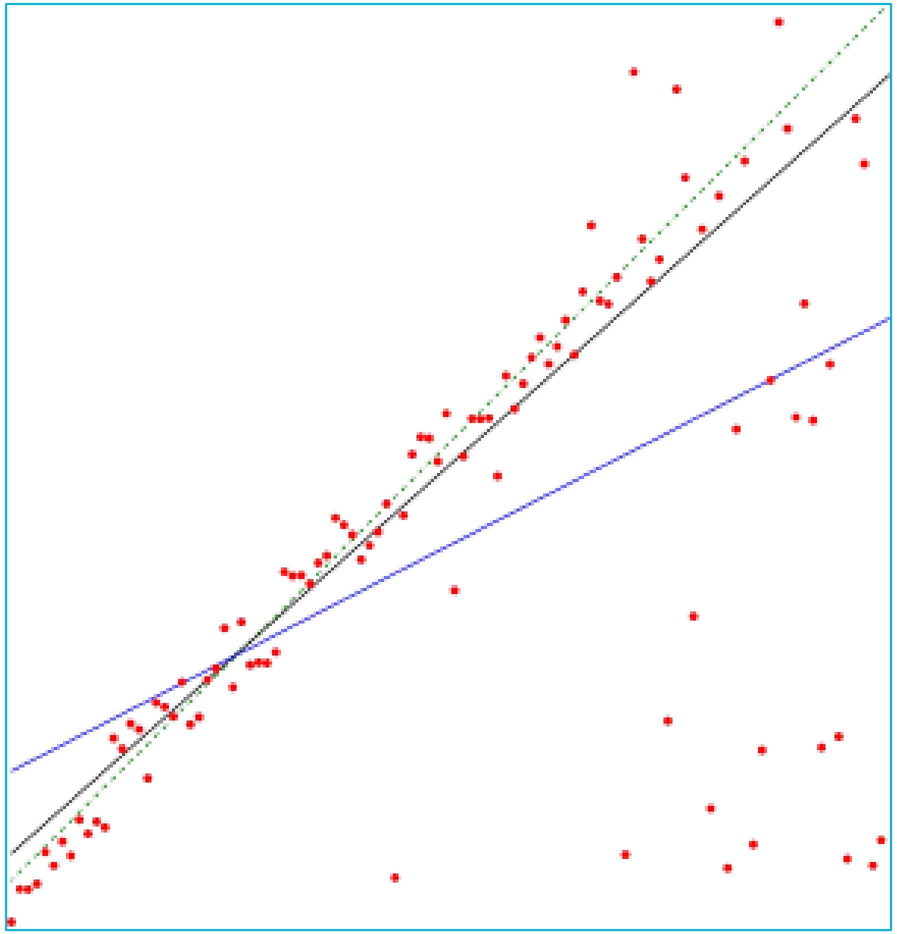
Chi-squared →

$$\chi^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{\lambda} \frac{\left(Y(\lambda)_{in\ situ,i} - Y(\lambda)_{OLCI,i} \right)^2}{Y(\lambda)_{in\ situ,i}} \right)$$

$$Y(\lambda)_i = \frac{Rrs(\lambda)_i}{Rrs(560)_i}$$

+ linear regression of two types…

+ linear regression of two types...



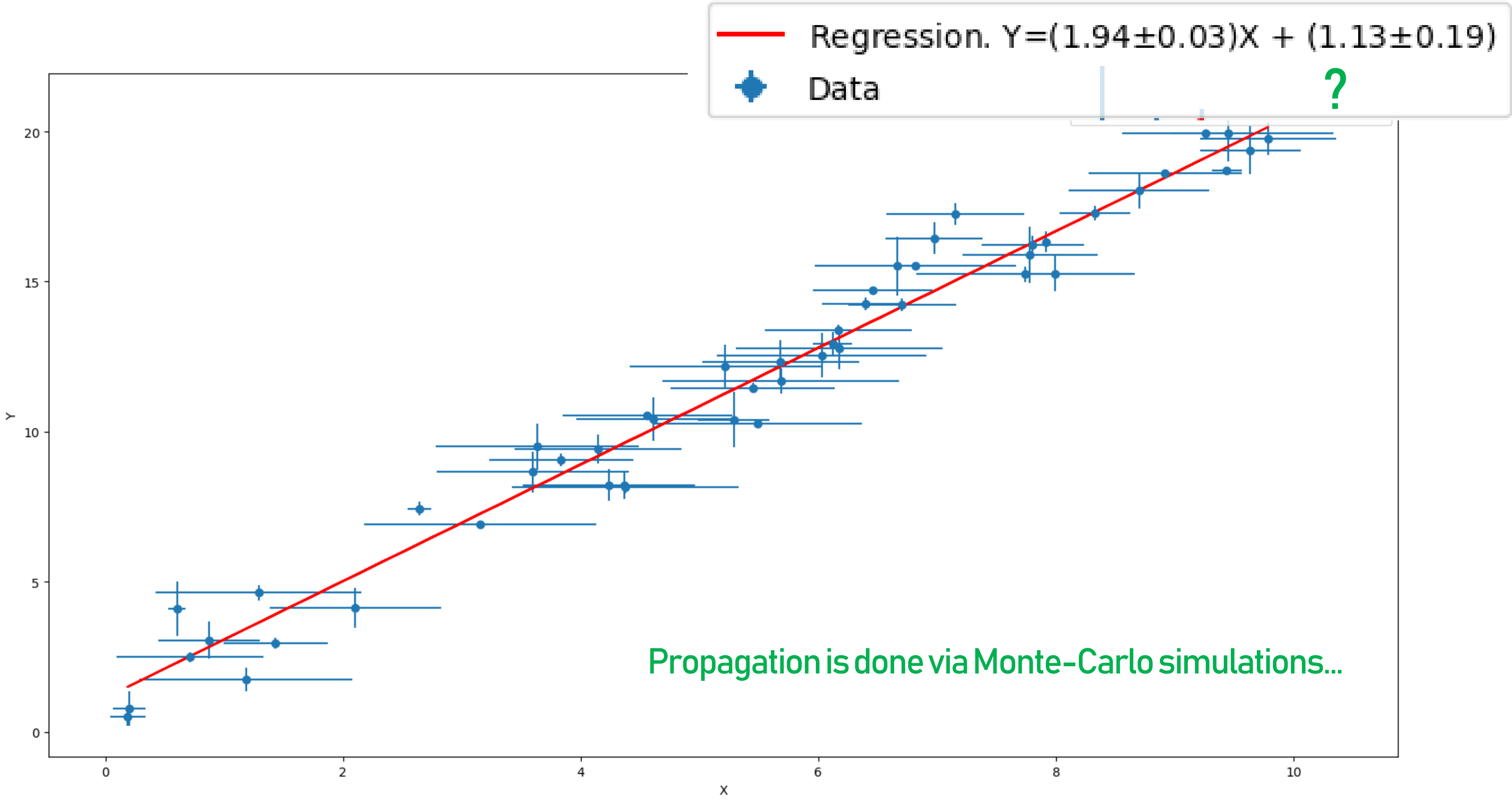
---Standard Least-Squares

---Theil-Sen regressor

- 1:1 line

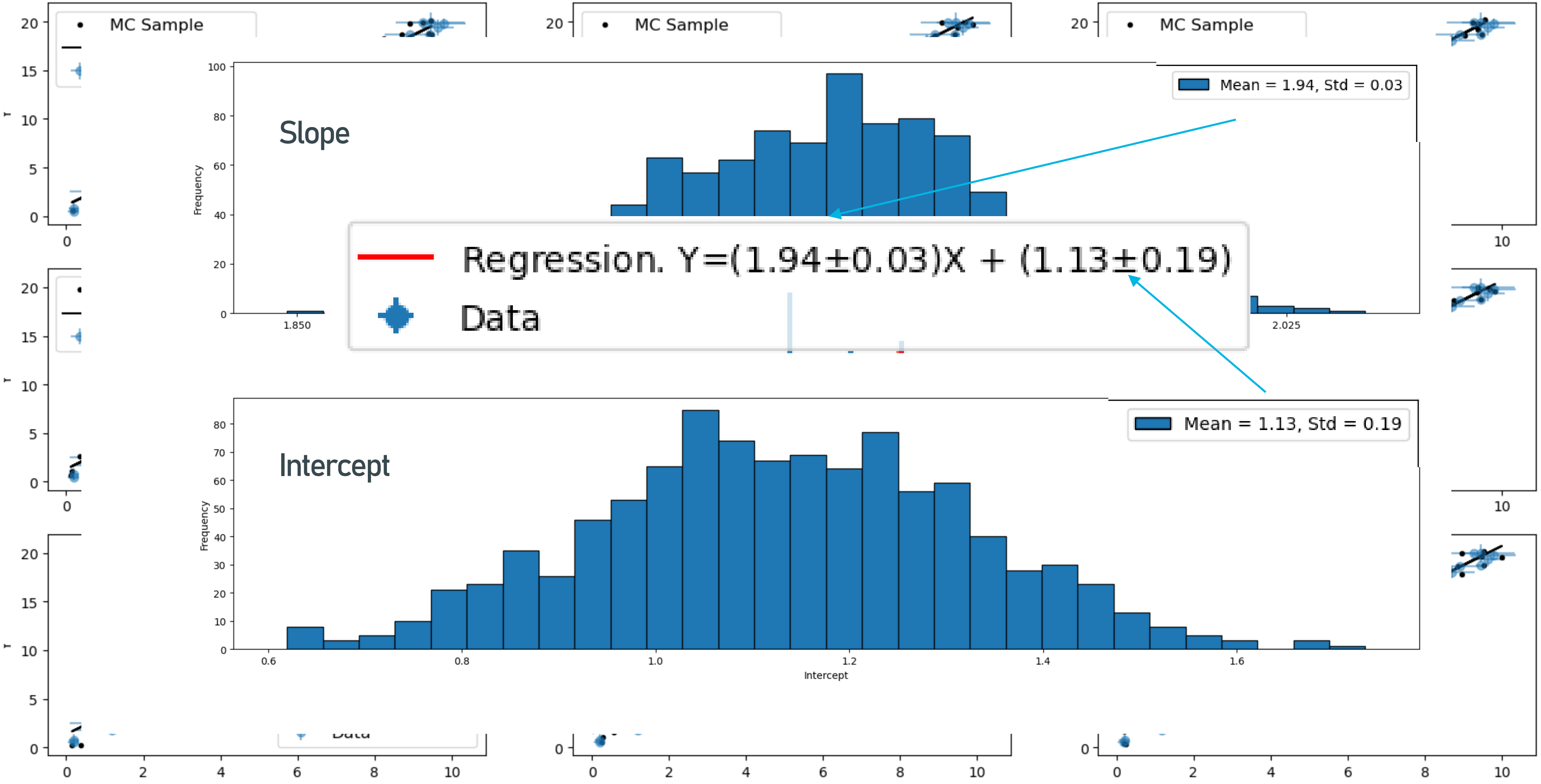
There is no one better than the other. Having both to compare can give a first assessment of the effect of the outliers in your matchup performanaces

2. Some background: Performance metrics: unc. propagation

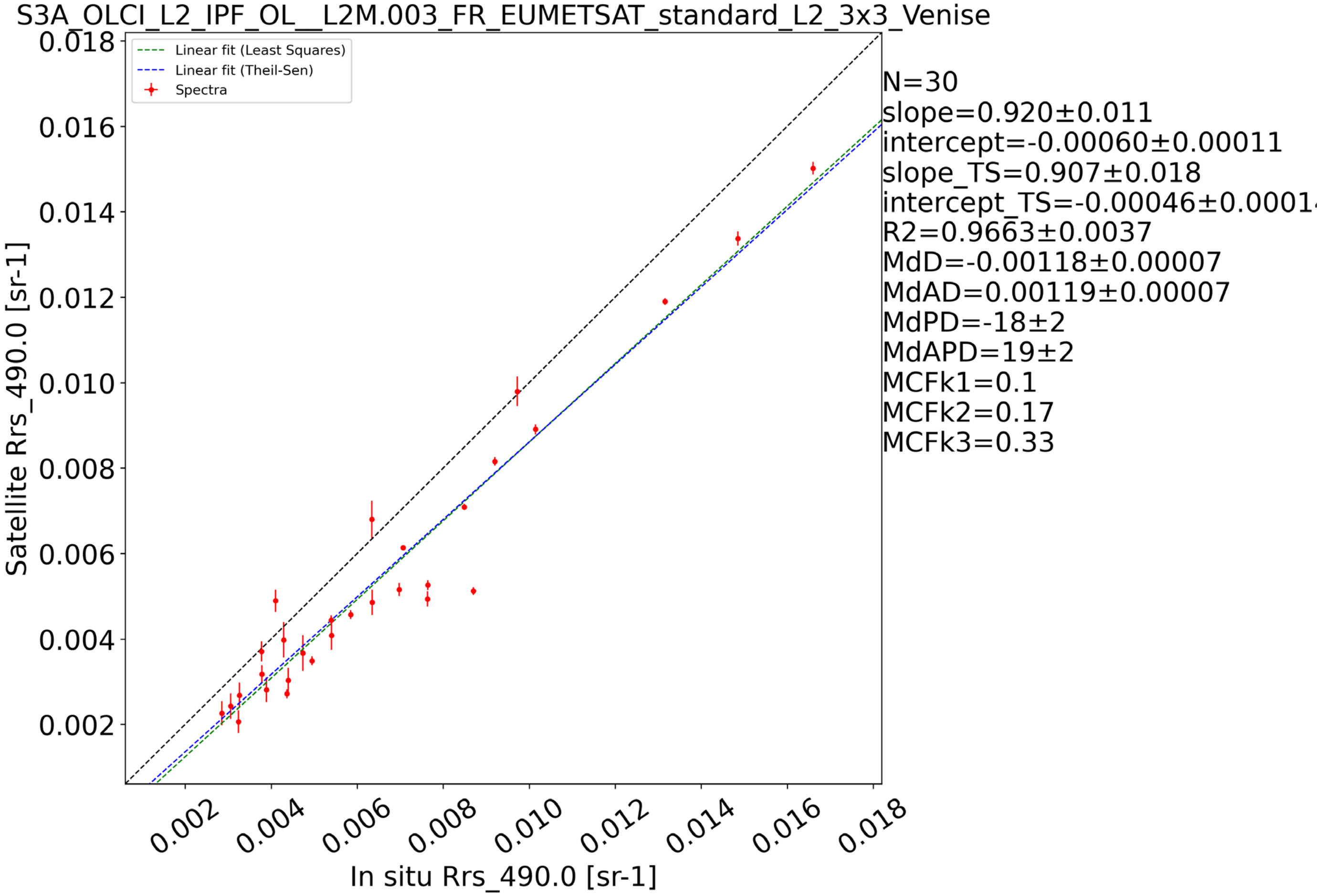


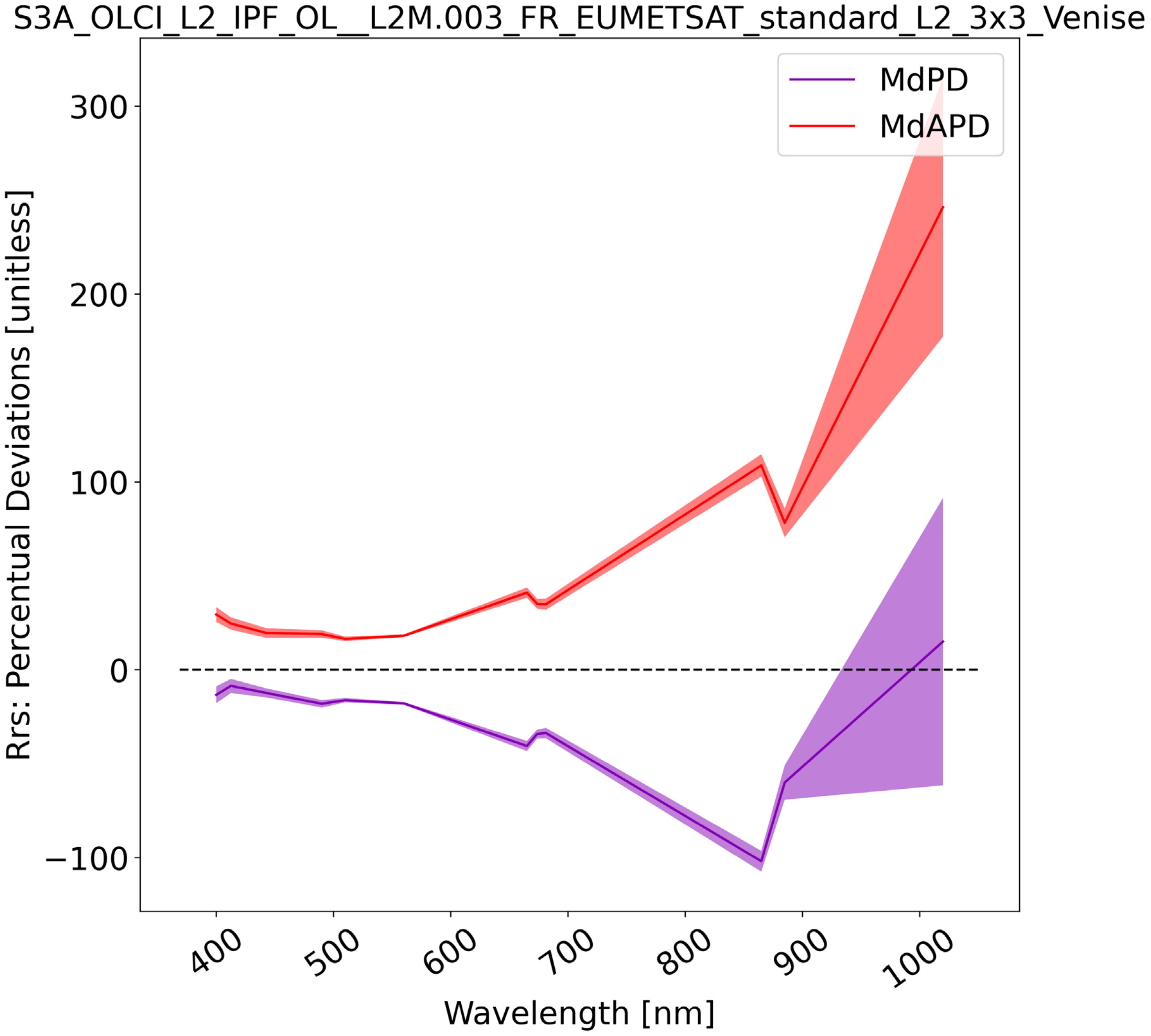
Propagation is done via Monte-Carlo simulations…

1000 random re-samplings of the type $N(x,u_x)$



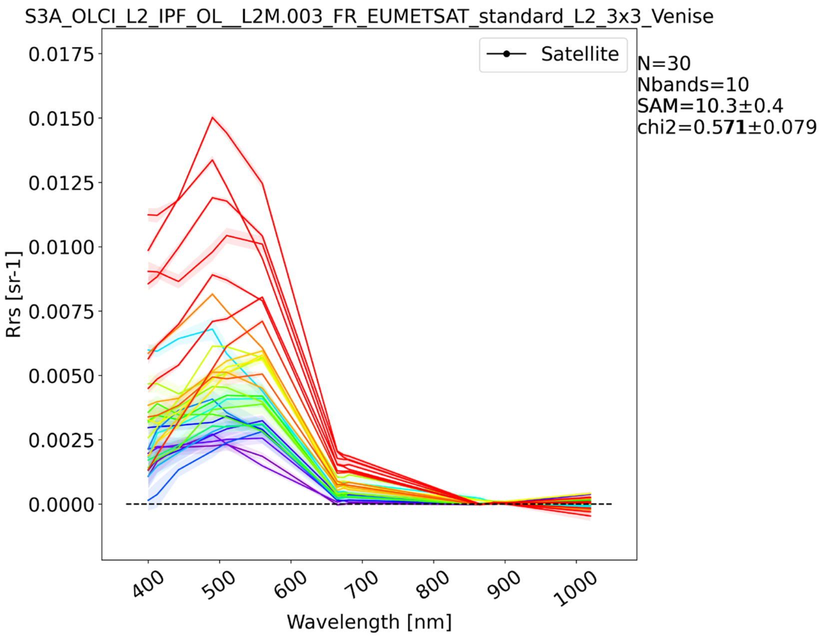
2. Some background: Performance metrics: unc. propagation

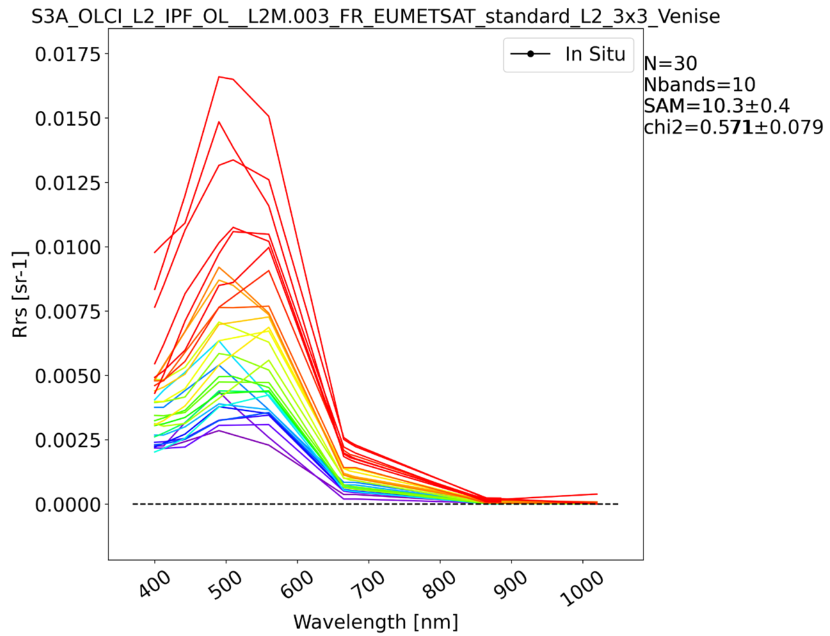


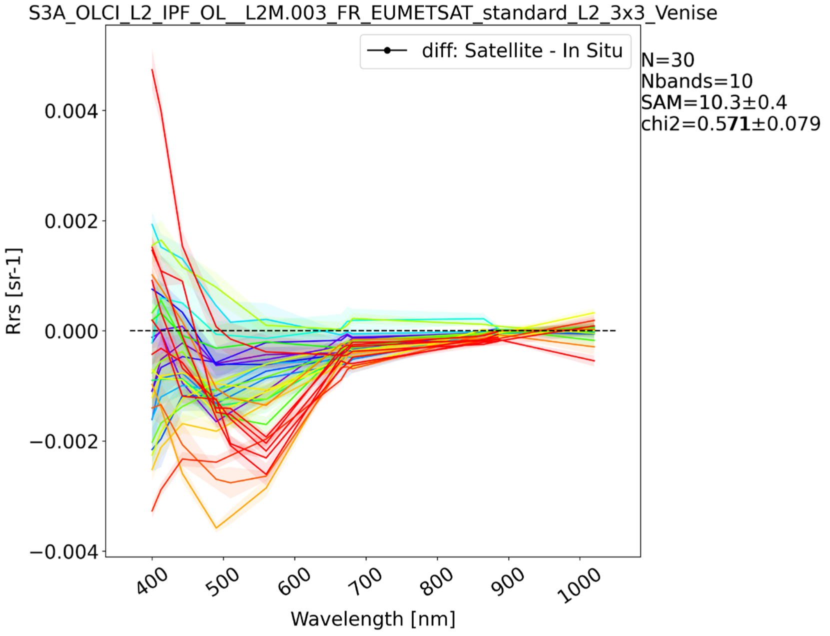


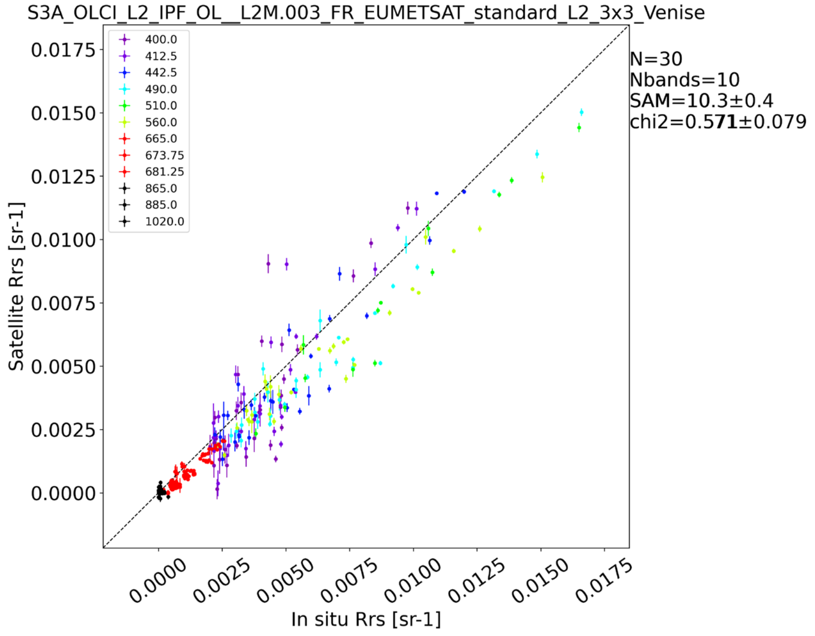
2. Some background: Performance metrics: unc. propagation

copernicus.eumetsat.int

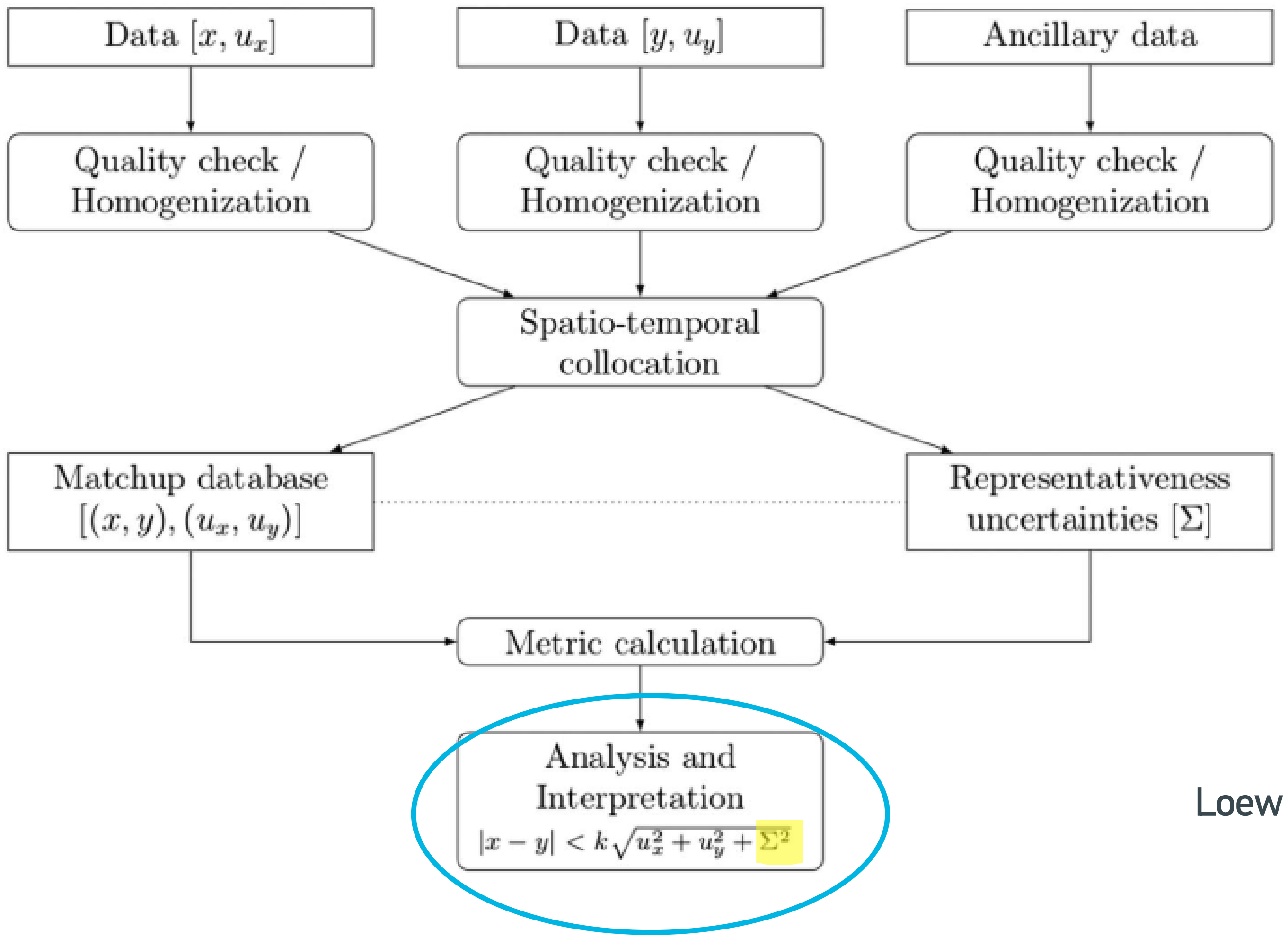








2. Some background: Metrologically-Compatible Fraction

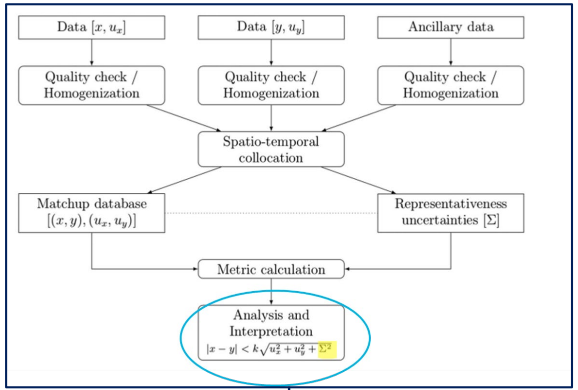


Loew et al. 2017

Difference between measurements fall within uncertainties?

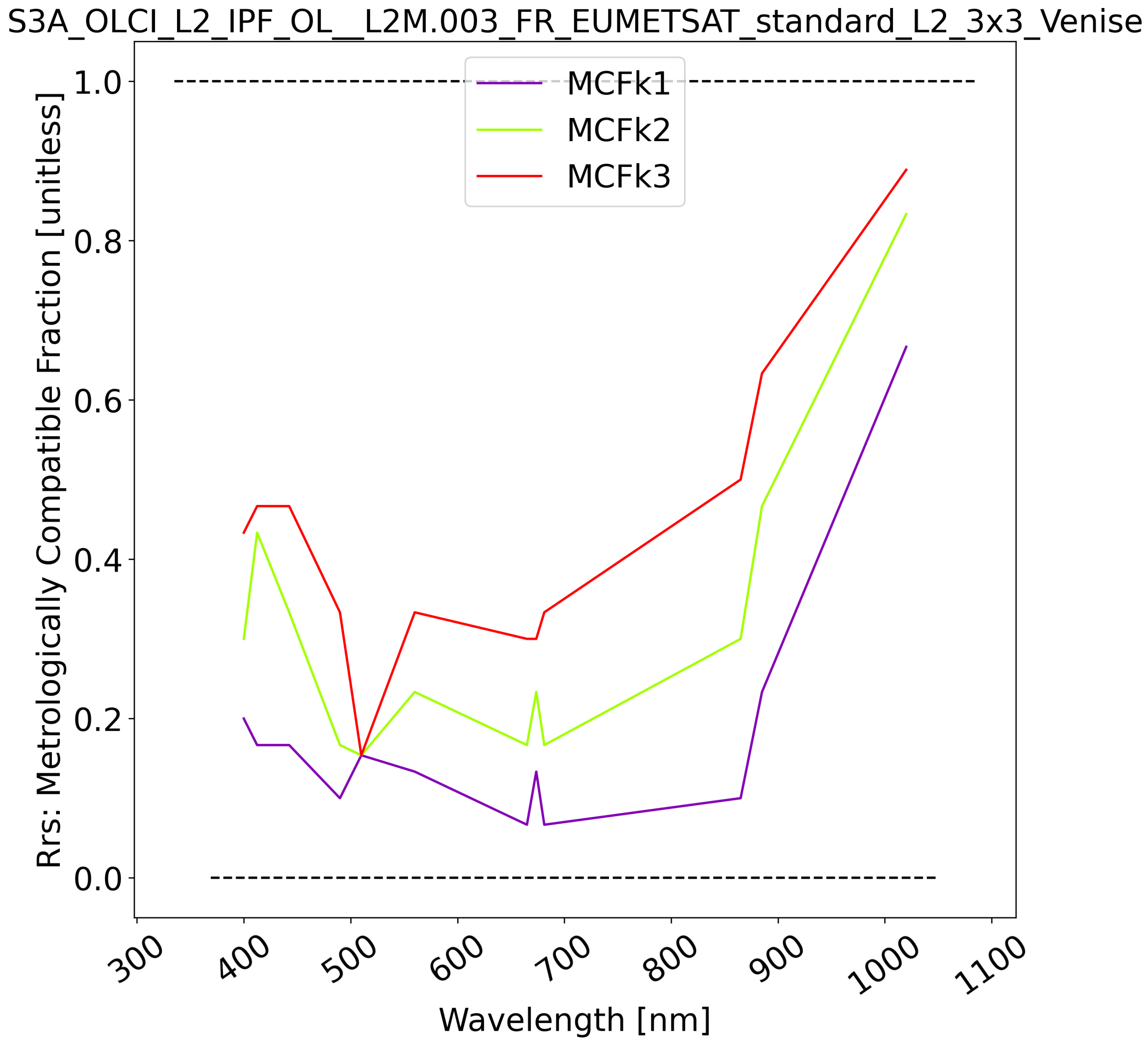
k = coverage factor

2. Some background: Metrologically-Compatible Fraction



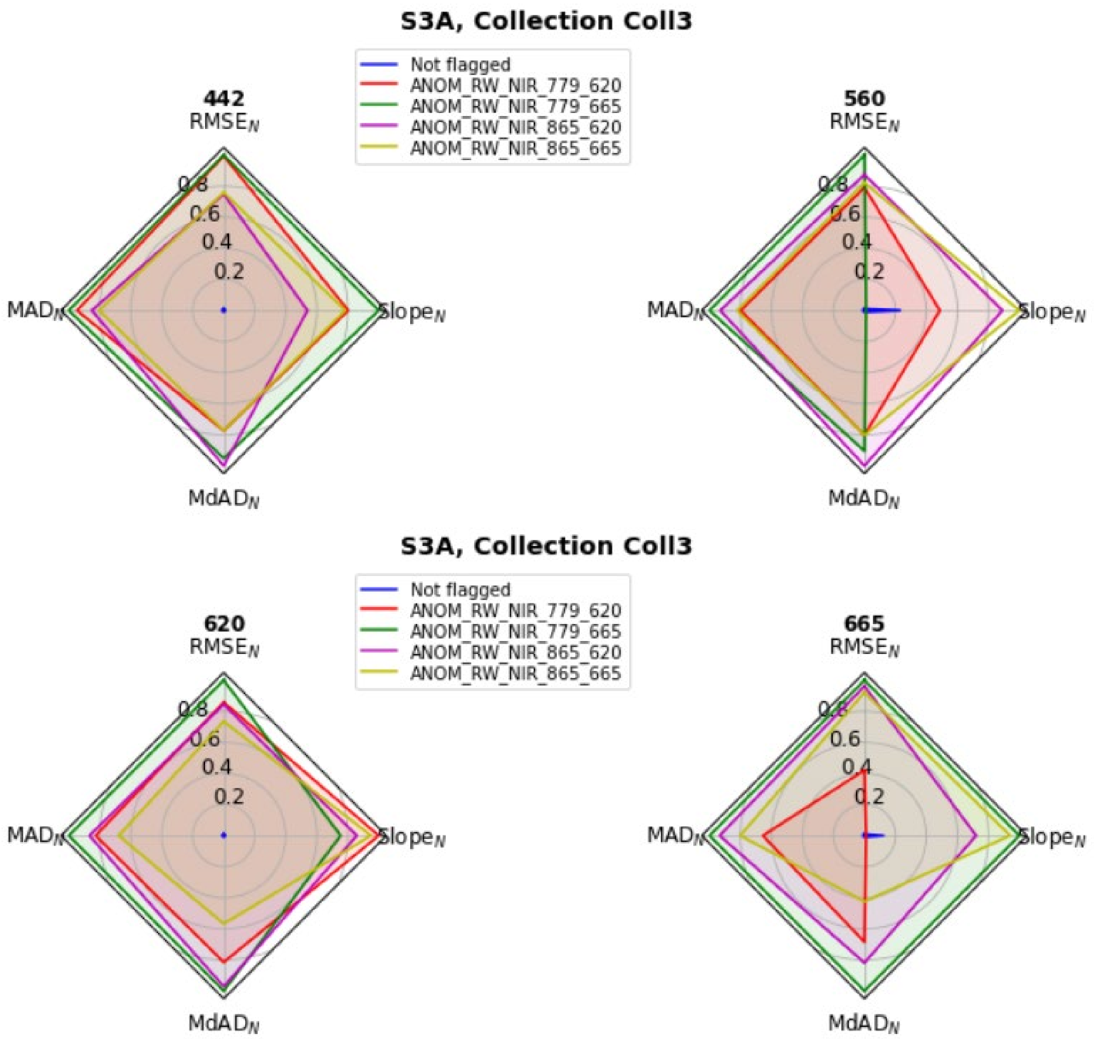
$$MCF(k) = \frac{\#\{|x-y| < k(u_x + u_y)\}}{N}$$

Metrologically-Compatible Fraction
→ Recommended by Kevin Ruddick



2. A discussion slide on radar plots summarising all the statistics

Area of a radar plot, the definite statistics?



3. Usage

ThoMaS workflow is divided into **5** main steps:

The steps can be executed **sequentially** or **independently** in case the needed outputs of the previous steps are available.

1. Step **insitu**
    1. Ingest insitu data from **SeaBASS** input file
    2. Apply several transformations to make **insitu comparable to satellite** data (e.g. **spectral matching** with satellite, **BRDF** correction)
    3. Store them into standard **IDB** (**In situ Data Base**) netCDF4 file.

    → This step can optionally include the acquisition of **ancillary information** from **ECMWF** at the lat-lon-times of your insitu measurements.

2. Step **SatData**: Download and list the **satellite products (L1B and/or L2)** matching spatially-temporally your insitu data.

    → Download for products available in EUMETSAT Data Store and NASA's OBPG.

3. Step **minifiles**: SatData are grouped/unnested into single netCDF4 file, sliced in horizontal dimensions, centred at the desired (in situ) location.

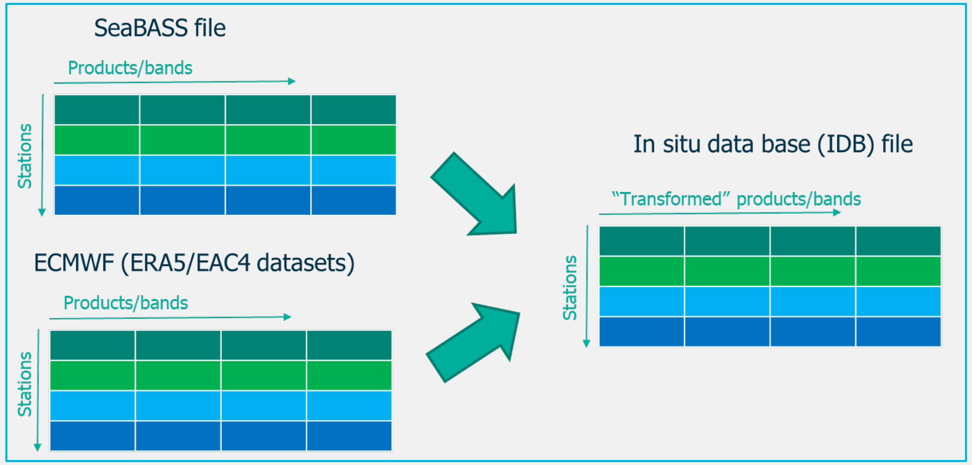
4. Step **EDB**.
    1. Stack minifiles into single netCDF
    2. Apply transformations to SatData to make them comparable to in situ (e.g. **scale/unit conversion**, **BRDF** correction)
    3. Calculate extraction **statistics** over the extraction window following EUMETSAT's or any user-defined Matchup Protocol.
    4. Store into standard **EDB** (**Extraction Data Base**) netCDF4 and CSV files.

5. Step **MDB**.
    1. Combine insitu (**IDB**) and satellite (**EDB**) information indexed into insitu-satellite **matchup pairs**
    2. **Optionally apply time interpolation**
    3. Calculate **matchup statistics**
    4. Store into standard **MDB** (**Matchup Data Base**) netCDF4 and CSV files.

copernicus.eumetsat.int

3. Usage. Step insitu: SeaBASS/OCDB file

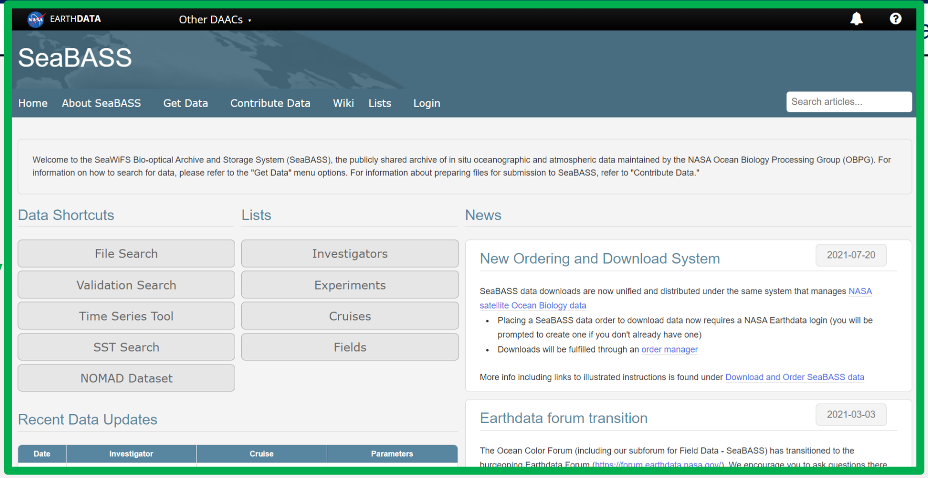

```
/begin_header
/investigators=TestPI
/affiliations=TestAffiliation
/contact=TestContact@TestInstitutution.org
/received=20190101
/processed=20190101
/experiment=TestExperiment
/cruise=TestCruise
/station=NA
/data_file_name=TestOCDBsubmission.txt
/documents=see_comments
/calibration_files=see_comments
/data_type=cast
/data_status=final
/water_depth=NA
/BRDF_correction=None
/instrument_model=RAMSES
/instrument_manufacturer=TriOS
/calibration_date=NA
!
! COMMENTS
!
! Citation: Cite your paper where your data are published
!
/missing=-9999.
/delimiter=comma
/fields=station,date,time,lon,lat,depth,cloud,RelAz,AOT,spm,Chla,Rrs355,Rrs360,Rrs365,Rrs370,Rrs375,Rrs380,Rrs385,Rrs390,Rrs395,Rrs400,Rrs405,Rrs410,…
/units=none,yyyymmdd,hh:mm:ss,degrees,degrees,m,%,degrees,unitless,mg/L,mg/m^3,1/sr,1/sr,1/sr,1/sr,1/sr,1/sr,1/sr,1/sr,1/sr,1/sr,1/sr,1/sr,…
/end_header
TestStation001,20210815,11:30:00,0,0,0.01,0,134.7,0.2444,0.1,0.1,0.001886946,0.002068008,0.002167035,0.002305759,0.002504616,0.002699149,0.002943716,0.003166,…
TestStation002,20201014,11:55:00,-32.6232,32.0859,0.01,0,134.8,0.2388,0.1,0.1,0.002158972,0.002363762,0.00247761,0.00264967,0.002880799,0.003116844,0.0034196,…
TestStation003,20201014,12:16:00,-32.0209,26.9584,0.01,0,134.7,0.2388,0.1,0.1,0.001918189,0.002097384,0.002191946,0.002353893,0.002567901,0.002792078,0.00307,…
TestStation004,20210910,09:40:00,3,-3,0.01,0,134.9,0.2388,0.1,0.1,0.001271305,0.001359205,0.00139314,0.001432681,0.001487187,0.00153414,0.00160502,0.00165816,…
TestStation005,20210910,09:54:00,-4,4,0.01,0,134.8,0.277,0.1,0.1,0.00123858,0.001313013,0.001336158,0.00136599,0.001412448,0.001450581,0.001508596,0.00155277,…
TestStation006,20210910,10:07:00,5,-5,0.01,0,134.8,0.277,0.1,0.1,0.001126482,0.001192422,0.001211226,0.001237373,0.001276362,0.001307301,0.00135953,0.0013975,…
```



3. Usage. Step insitu: SeaBASS/OCDB file

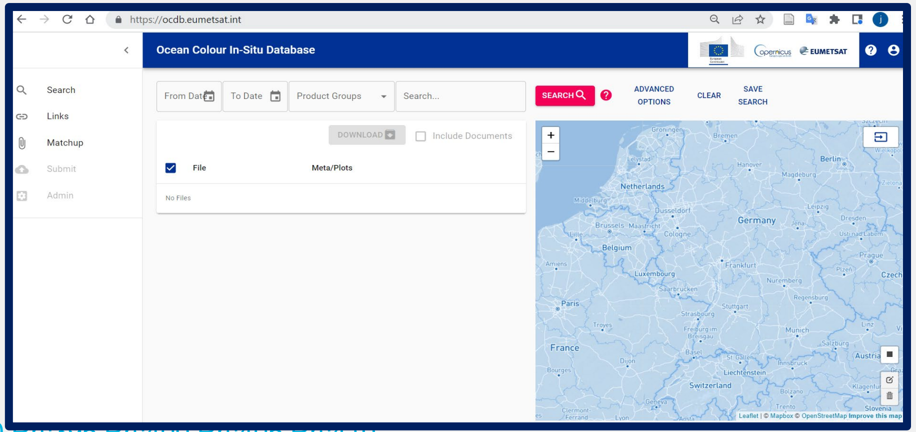
```
/begin_header
/investigators=TestPI
/affiliations=TestAffiliation
/contact=TestContact@TestInstitutution.org
/received=20190101
/processed=20190101
/experiment=TestExperiment
/cruise=TestCruise
/station=NA
/data_file_name=TestOCDBsubmission.txt
/documents=see_comments
/calibration_files=see_comments
/data_type=cast
/data_status=final
/water_depth=NA
/BRDF_correction=None
/instrument_model=RAMSES
/instrument_manufacturer=TriOS
/calibration_date=NA
!
! COMMENTS
!
! Citation: Cite your paper where your data are published
!
/missing=-9999.
/delimiter=comma
/fields=station,date,time,lon,lat,depth,cloud,RelAz,AOT,spm,Chla,Rrs355,Rrs360,Rrs365,Rrs370,Rrs375,Rrs380,Rrs385,Rrs390,Rrs395,Rrs400,Rrs405,Rrs410,…
/units=none,yyyymmdd,hh:mm:ss,degrees,degrees,m,%,degrees,unitless,mg/L,mg/m^3,1/sr,1/sr,1/sr,1/sr,1/sr,1/sr,1/sr,1/sr,1/sr,1/sr,1/sr,1/sr,…
/end_header
TestStation001,20210815,11:30:00,0,0,0.01,0,134.7,0.2444,0.1,0.1,0.001886946,0.002068008,0.002167035,0.002305759,0.002504616,0.002699149,0.002943716,0.003166,…
TestStation002,20201014,11:55:00,-32.6232,32.0859,0.01,0,134.8,0.2388,0.1,0.1,0.002158972,0.002363762,0.00247761,0.00264967,0.002880799,0.003116844,0.0034196,…
TestStation003,20201014,12:16:00,-32.0209,26.9584,0.01,0,134.7,0.2388,0.1,0.1,0.001918189,0.002097384,0.002191946,0.002353893,0.002567901,0.002792078,0.00307,…
TestStation004,20210910,09:40:00,3,-3,0.01,0,134.9,0.2388,0.1,0.1,0.001271305,0.001359205,0.00139314,0.001432681,0.001487187,0.00153414,0.00160502,0.00165816,…
TestStation005,20210910,09:54:00,-4,4,0.01,0,134.8,0.277,0.1,0.1,0.00123858,0.001313013,0.001336158,0.00136599,0.001412448,0.001450581,0.001508596,0.00155277,…
TestStation006,20210910,10:07:00,5,-5,0.01,0,134.8,0.277,0.1,0.1,0.001126482,0.001192422,0.001211226,0.001237373,0.001276362,0.001307301,0.00135953,0.0013975,…
```

SeaBASS
seabass.gsfc.nasa.gov



Inter-operable

OCDB
ocdb.eumetsat.int



3. Usage. Step insitu: SeaBASS/OCDB file

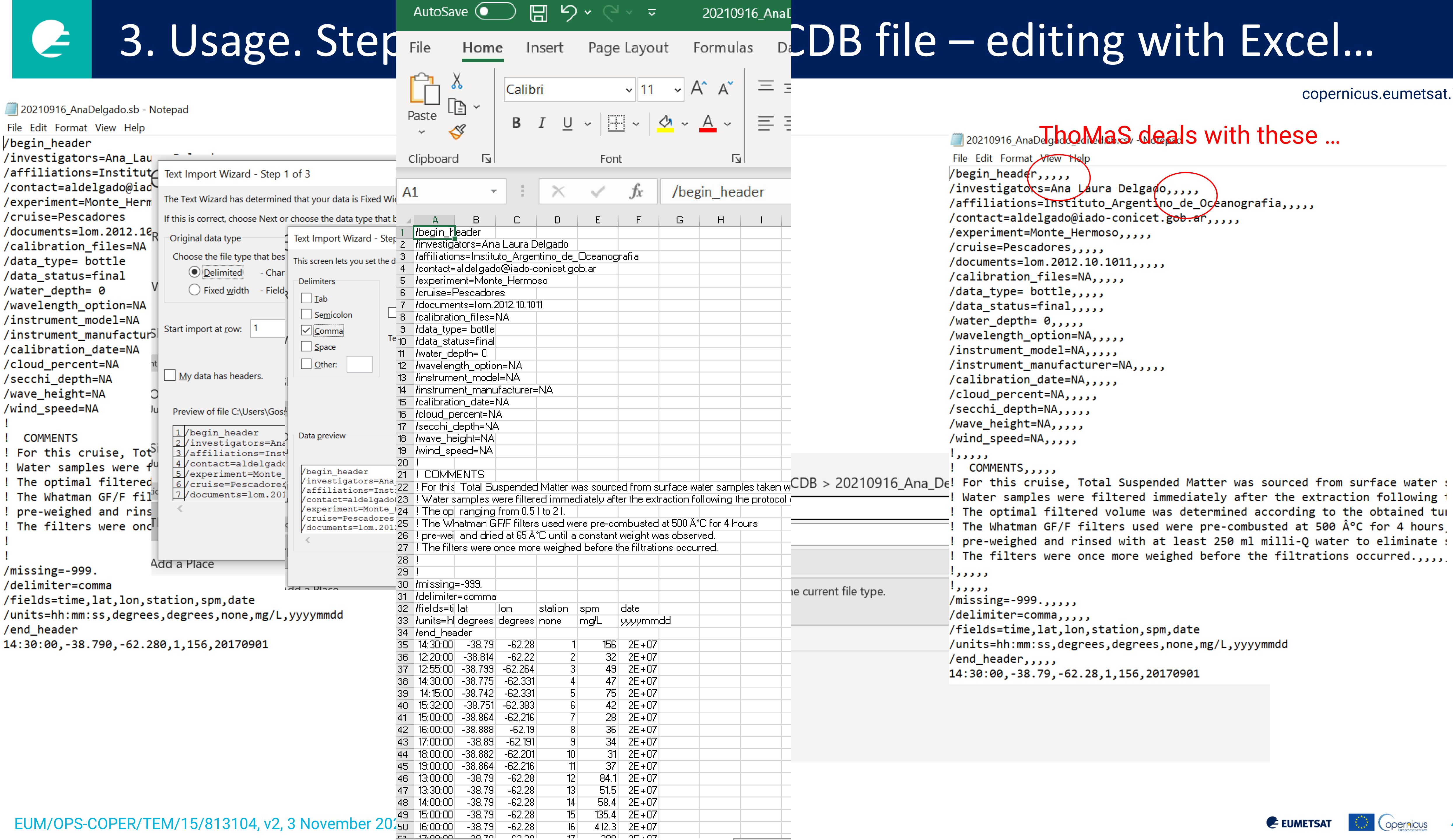

```
/begin_header
/investigators=TestPI
/affiliations=TestAffiliation
/contact=TestContact@TestInstitutution.org
/received=20190101
/processed=20190101
/experiment=TestExperiment
/cruise=TestCruise
/station=NA
/data_file_name=TestOCDBsubmission.txt
/documents=see_comments
/calibration_files=see_comments
/data_type=cast
/data_status=final
/water_depth=NA
/BRDF_correction=None
/instrument_model=RAMSES
/instrument_manufacturer=TriOS
/calibration_date=NA
!
! COMMENTS
!
! Citation: Cite your paper where your data are published
!
/missing=-9999.
/delimiter=comma
/fields=station,date,time,lon,lat,depth,cloud,RelAz,AOT,spm,Chla,Rrs355,Rrs360,Rrs365,Rrs370,Rrs375,Rrs380,Rrs385,Rrs390,Rrs395,Rrs400,Rrs405,Rrs410,…
/units=none,yyyymmdd,hh:mm:ss,degrees,degrees,m,%,degrees,unitless,mg/L,mg/m^3,1/sr,1/sr,1/sr,1/sr,1/sr,1/sr,1/sr,1/sr,1/sr,1/sr,1/sr,1/sr,…
/end_header
TestStation001,20210815,11:30:00,0,0,0.01,0,134.7,0.2444,0.1,0.1,0.001886946,0.002068008,0.002167035,0.002305759,0.002504616,0.002699149,0.002943716,0.003166,…
TestStation002,20201014,11:55:00,-32.6232,32.0859,0.01,0,134.8,0.2388,0.1,0.1,0.002158972,0.002363762,0.00247761,0.00264967,0.002880799,0.003116844,0.0034196,…
TestStation003,20201014,12:16:00,-32.0209,26.9584,0.01,0,134.7,0.2388,0.1,0.1,0.001918189,0.002097384,0.002191946,0.002353893,0.002567901,0.002792078,0.00307,…
TestStation004,20210910,09:40:00,3,-3,0.01,0,134.9,0.2388,0.1,0.1,0.001271305,0.001359205,0.00139314,0.001432681,0.001487187,0.00153414,0.00160502,0.00165816,…
TestStation005,20210910,09:54:00,-4,4,0.01,0,134.8,0.277,0.1,0.1,0.00123858,0.001313013,0.001336158,0.00136599,0.001412448,0.001450581,0.001508596,0.00155277,…
TestStation006,20210910,10:07:00,5,-5,0.01,0,134.8,0.277,0.1,0.1,0.001126482,0.001192422,0.001211226,0.001237373,0.001276362,0.001307301,0.00135953,0.0013975,…
```

Recommendation: use Excel to bring your data to this format…

However…

3. Usage. Step … CDB file – editing with Excel…

copernicus.eumetsat.int

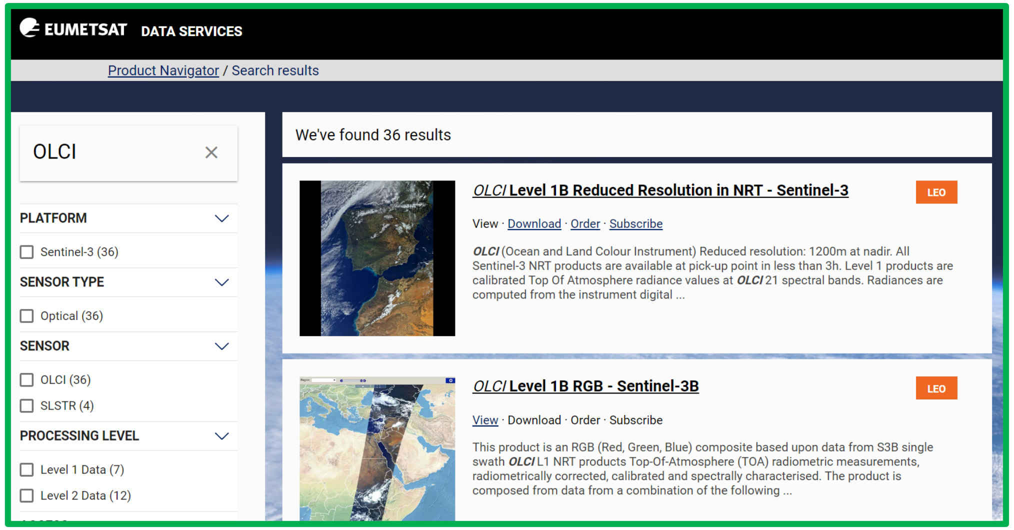


3. Usage. Step SatData

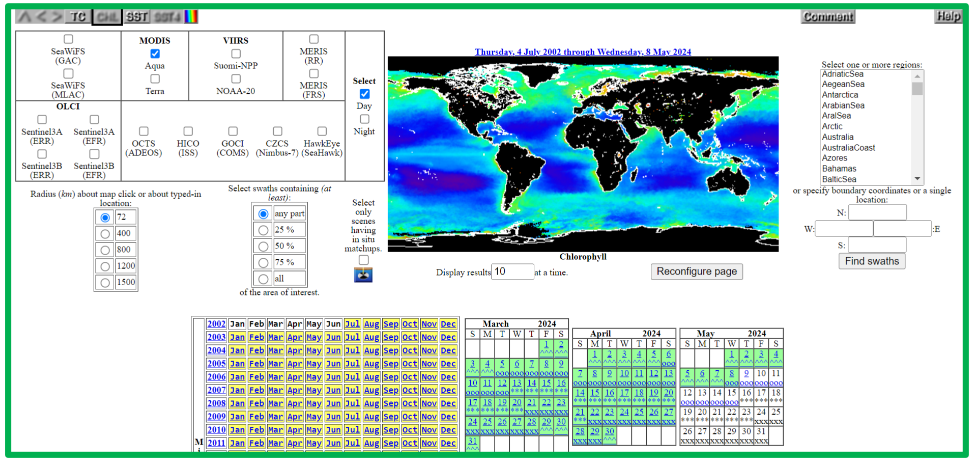
copernicus.eumetsat.int

eumdac

EUMETSAT Data Store



NASA's OBPG L2 standard products

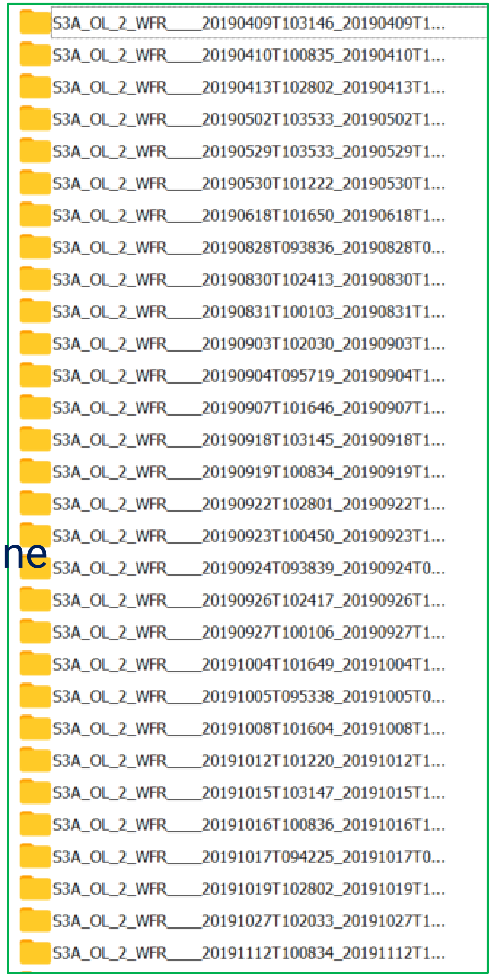


Standard wget request
As in getOC.py from U. Maine

. . .

. . .

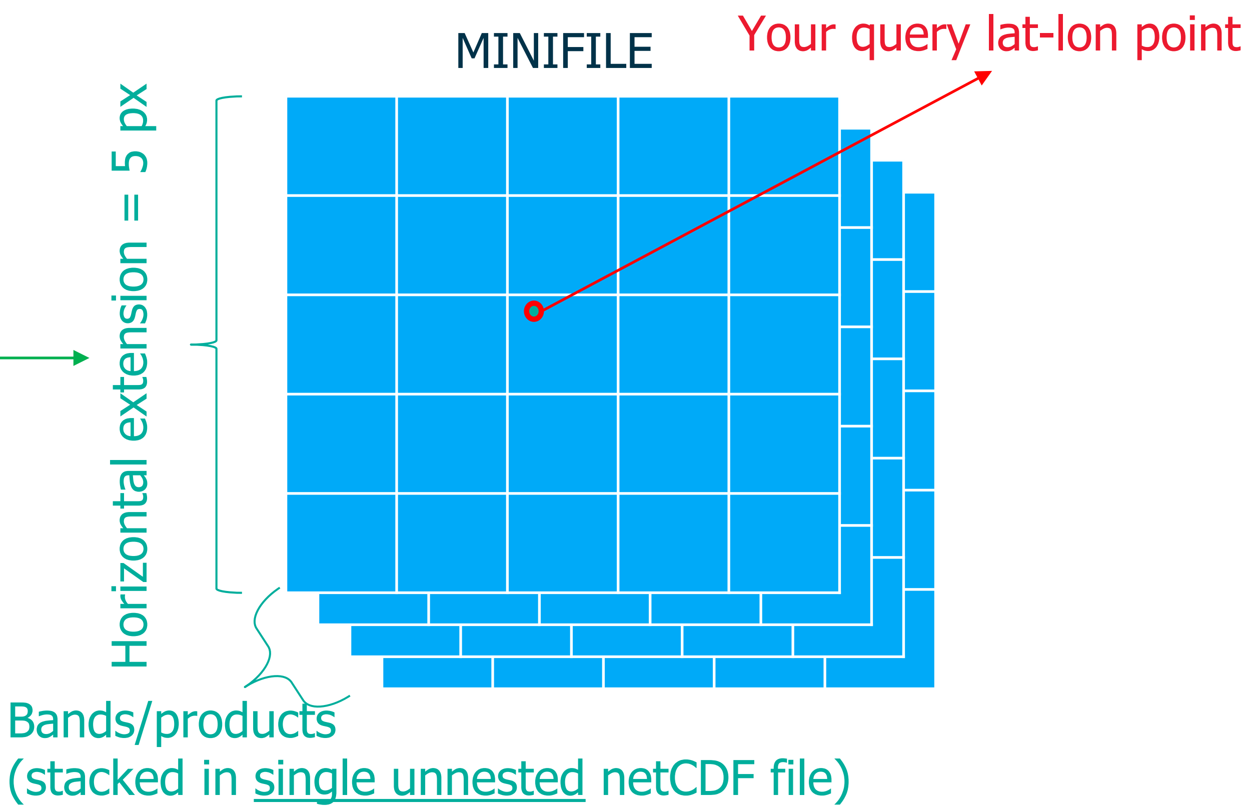
Your local system



+ SatData Lists matching in situ

3. Usage. Step minifiles

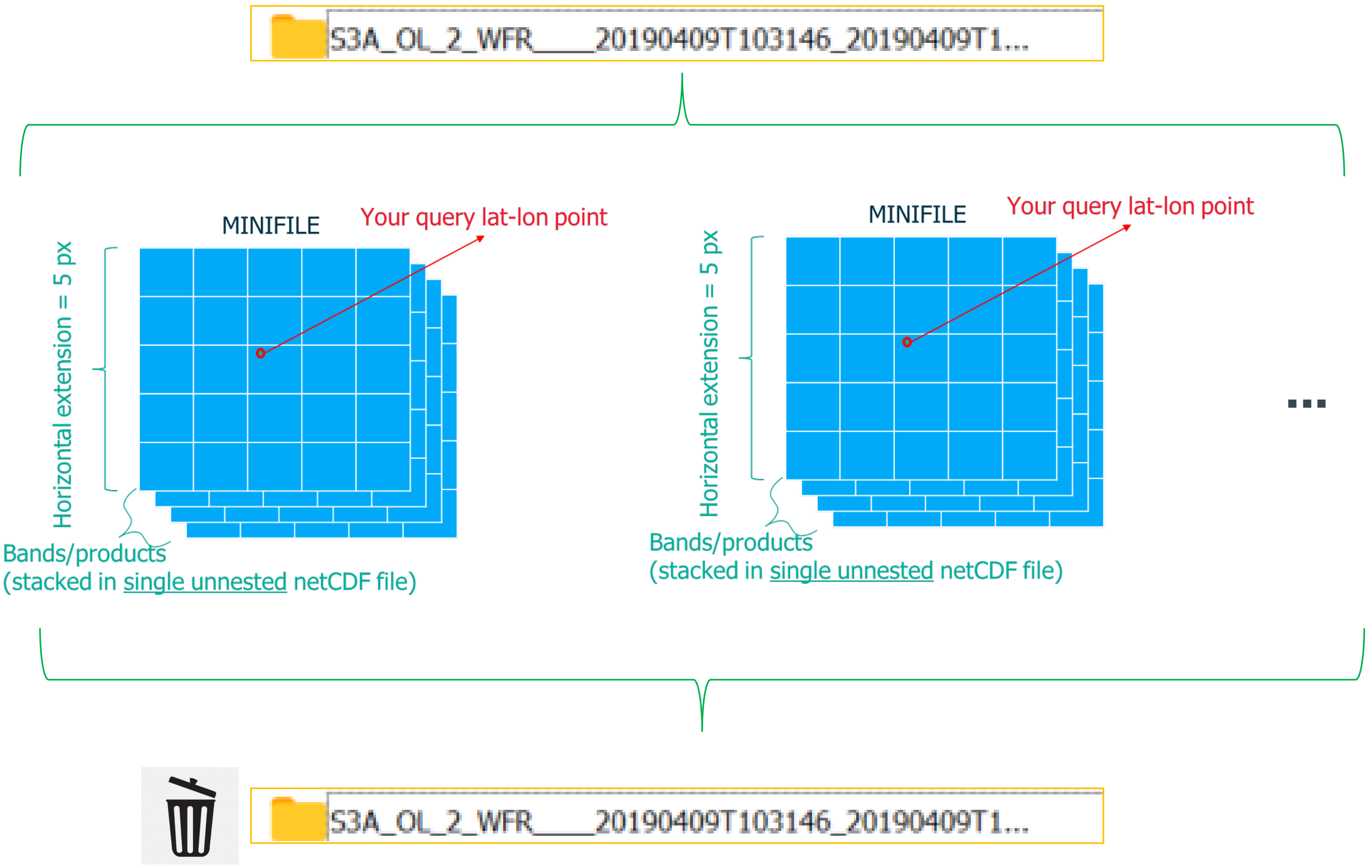
- S3A_OL_2_WFR____20190409T103146_20190409T1...
- S3A_OL_2_WFR____20190410T100835_20190410T1...
- S3A_OL_2_WFR____20190413T102802_20190413T1...
- S3A_OL_2_WFR____20190502T103533_20190502T1...
- S3A_OL_2_WFR____20190529T103533_20190529T1...
- S3A_OL_2_WFR____20190530T101222_20190530T1...
- S3A_OL_2_WFR____20190618T101650_20190618T1...
- S3A_OL_2_WFR____20190828T093836_20190828T0...
- S3A_OL_2_WFR____20190830T102413_20190830T1...
- S3A_OL_2_WFR____20190831T100103_20190831T1...
- S3A_OL_2_WFR____20190903T102030_20190903T1...
- S3A_OL_2_WFR____20190904T095719_20190904T1...
- S3A_OL_2_WFR____20190907T101646_20190907T1...
- S3A_OL_2_WFR____20190918T103145_20190918T1...
- S3A_OL_2_WFR____20190919T100834_20190919T1...
- S3A_OL_2_WFR____20190922T102801_20190922T1...
- S3A_OL_2_WFR____20190923T100450_20190923T1...
- S3A_OL_2_WFR____20190924T093839_20190924T0...
- S3A_OL_2_WFR____20190926T102417_20190926T1...
- S3A_OL_2_WFR____20190927T100106_20190927T1...
- S3A_OL_2_WFR____20191004T101649_20191004T1...
- S3A_OL_2_WFR____20191005T095338_20191005T0...
- S3A_OL_2_WFR____20191008T101604_20191008T1...
- S3A_OL_2_WFR____20191012T101220_20191012T1...
- S3A_OL_2_WFR____20191015T103147_20191015T1...
- S3A_OL_2_WFR____20191016T100836_20191016T1...
- S3A_OL_2_WFR____20191017T094225_20191017T0...
- S3A_OL_2_WFR____20191019T102802_20191019T1...
- S3A_OL_2_WFR____20191027T102033_20191027T1...
- S3A_OL_2_WFR____20191112T100834_20191112T1...



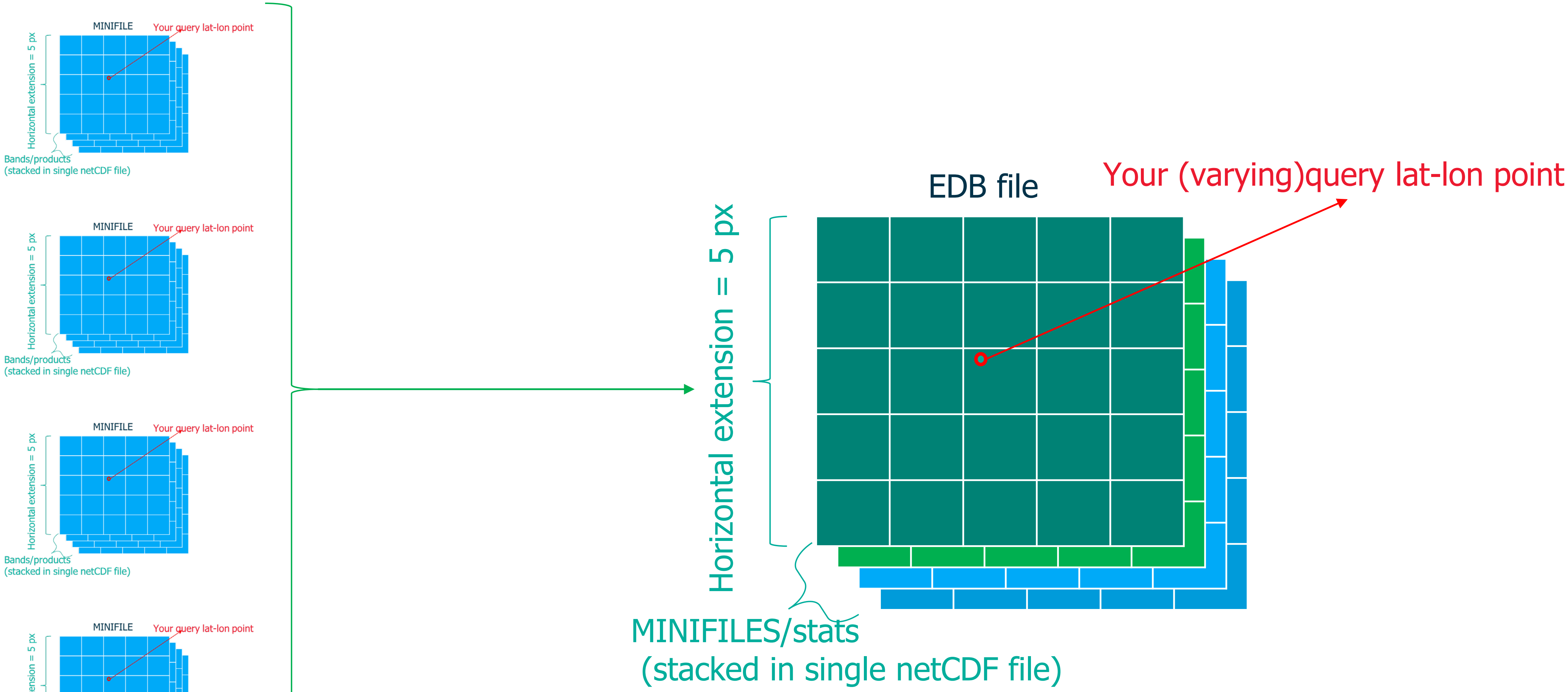
ThoMaS uses the minimisation of the orthodromic distance approach to find the centroid of the window

Satellite products are not altered at all at this step, with the exception of some OLCI geometries...

3. Usage. Option download_extract_delete



3. Usage. Step EDB



+ extraction statistics based on extraction protocol…

3. Usage. Step EDB. Custom quality flags

- [pixel-by-pixel-screening] A threshold on NDVI:
  - `(Rrs{865.0}-Rrs{620.0})/(Rrs{865.0}+Rrs{620.0})<0` ---> Pixels with negative NDVI will be screened.

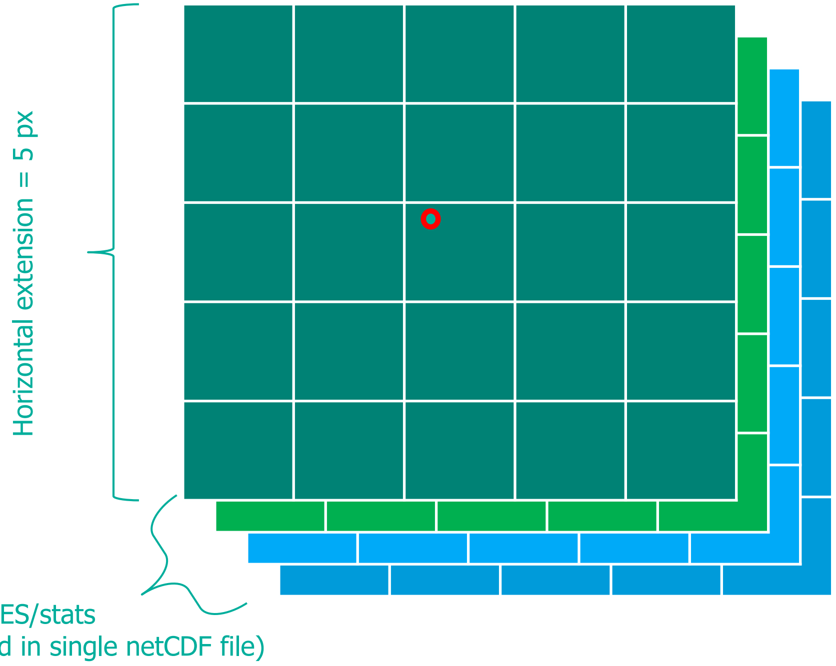
- [macropixel-screening] A threshold on the median of CV[Rrs(412nm-555nm);AOT(865nm)], as in Bailey and Werdell 2006:
  - `median[Rrs:cv{412...555};aot:cv{865}]>15` ---> *Macropixels* whose median coefficient of variation of Rrs in the range 412 nm to 555 nm and coefficient of variation of AOT at 865 nm combined (after pixel-by-pixel and outlier screening) falls above 15% will be screened.
    - `Rrs{412}` --> Specific band: 412 nm
    - `Rrs{...}` --> All bands
    - `Rrs{412...555}` --> All bands whose nominal wavelengths fall in the range 412 - 555 nm
    - `Rrs{...555}` --> All bands whose nominal wavelengths does not exceed 555 nm.
    - `Rrs{412...}` --> All bands whose nominal wavelengths does not go below 412 nm.

MDB file

1. Merging insitu and extractions according to matchup pairs
2. Statistical metrics calculated + scatter/spectral plots

In situ data base (IDB) file

Extraction data base (EDB) file



MDB: indexed by matchup pair ID
Time interpolation (nearest in time) applicable

3. Usage: Summary of the steps

ThoMaS workflow is divided into **5** main steps:

The steps can be executed **sequentially** or **independently** in case the needed outputs of the previous steps are available.

1. Step **insitu**
    1. Ingest insitu data from **SeaBASS** input file
    2. Apply several transformations to make **insitu comparable to satellite** data (e.g. **spectral matching** with satellite, **BRDF** correction)
    3. Store them into standard **IDB** (**In situ Data Base**) netCDF4 file.

    → This step can optionally include the acquisition of **ancillary information** from **ECMWF** at the lat-lon-times of your insitu measurements.

2. Step **SatData**: Download and list the **satellite products (L1B and/or L2)** matching spatially-temporally your insitu data.

    → Download for products available in EUMETSAT Data Store and NASA's OBPG.

3. Step **minifiles**: SatData are grouped/unnested into single netCDF4 file, sliced in horizontal dimensions, centred at the desired (in situ) location.

4. Step **EDB**.
    1. Stack minifiles into single netCDF
    2. Apply transformations to SatData to make them comparable to in situ (e.g. **scale/unit conversion**, **BRDF** correction)
    3. Calculate extraction **statistics** over the extraction window following EUMETSAT's or any user-defined Matchup Protocol.
    4. Store into standard **EDB** (**Extraction Data Base**) netCDF4 and CSV files.

5. Step **MDB**.
    1. Combine insitu (**IDB**) and satellite (**EDB**) information indexed into insitu-satellite **matchup pairs**
    2. **Optionally apply time interpolation**
    3. Calculate **matchup statistics**
    4. Store into standard **MDB** (**Matchup Data Base**) netCDF4 and CSV files.

3. Usage: Summary on the terminology

- **IDB** (In situ Data Base): a netCDF file containing all the information related to the inputted insitu data. + (if requested) ancillary information from **ECMWF reanalysis datasets** at the insitu geographic location and time stamp.

- **SatData**: an image file/directory. In the case of standard L1/L2 OLCI products, it is composed of a directory containing several netCDF files, each containing one/several products + a manifest.xml file.

- **Minifile**: A single netCDF file containing all the relevant L1/L2 products from a single SatData, but only at the required location (and with a predefined window size).

- **EDB** (Extraction Data Base): All the statistical information (pixel-by-pixel flagging, outlier removal, central and dispersion values before/after outlier/mask removal, etc., details of the extraction protocol) is stored for all the extractions in one single netCDF file per extraction set.

- **MDB** (Match-up Data Base): All the information from IDB and EDB combined and re-indexed according to matchup pairs + matchup statistics.

Find **examples** of all these files (except SatData) in the **examples/example_files** directory.

4. Pre-requisites

1. Apart from that background knowledge…
2. Conda: Install the latest Anaconda Python distribution.
3. EUMETSAT Data Store:
   - Create EO Portal user and get API consumer key and secret.
   - Save EO Portal API credentials under **~/.eumdac_credentials.json**
4. NASA OBPG:
   - Create Earthaccess user and get user and password.
   - Save user and password under **~/.obpg_credentials.json**
5. ECMWF: Register to ADS/CDS and get url and key.
6. ECMWF: store ADS/CDS url/keys under **~/.ecmwf_api_config.txt**

Dependencies

item	version	licence	package info
BeautifulSoup	4.6.0	MIT	https://anaconda.org/conda-forge/beautifulsoup4
cdsapi	0.1.6	Apache-2.0	https://anaconda.org/conda-forge/cdsapi
ephem	4.1.3	MIT	https://pypi.org/project/ephem/
eumdac	2.0.1	MIT	https://anaconda.org/eumetsat/eumdac
matplotlib	3.5.2	PSF-based	https://anaconda.org/conda-forge/matplotlib
netcdf4	1.5.8	MIT	https://anaconda.org/conda-forge/netcdf4
numpy	1.23.0	BSD-3-Clause	https://anaconda.org/conda-forge/numpy
pandas	1.4.3	BSD-3-Clause	https://anaconda.org/conda-forge/pandas
python	3.9	PSF	https://docs.python.org/3/license.html
scipy	1.8.1	BSD-3-Clause	https://anaconda.org/conda-forge/scipy
xarray	2022.3.0	Apache-2.0	https://anaconda.org/conda-forge/xarray
jupyter	1.0.0	Unspecified	https://anaconda.org/anaconda/jupyter

Conda will take care of this…

5. Getting the code

- Git way:

```
cd ~
mkdir ThoMaS
cd ThoMaS
git clone --depth 1 https://gitlab.eumetsat.int/eumetlab/oceans/ocean-science-studies/ThoMaS .
```

- Direct download:

https://gitlab.eumetsat.int/eumetlab/oceans/ocean-science-studies/ThoMaS

Recent updates were done on the code

```
git fetch
git pull
```

6. Setting the environment

- Once conda and ThoMaS are installed, create the thomas env:

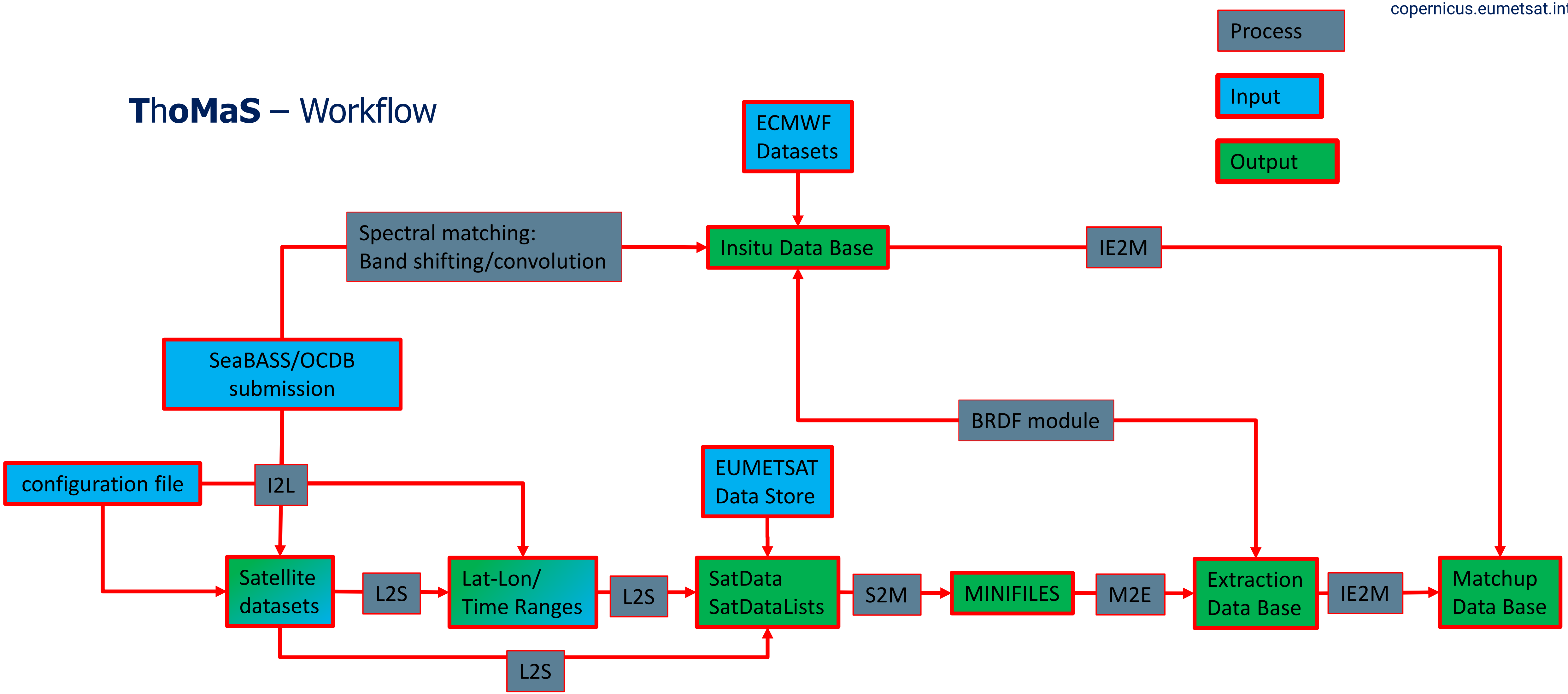
```
cd ~
cd ThoMaS
conda env create –f environment.yml
conda activate thomas
```

libmamba is the best choice for those of you who are stuck in the "**Solving environment step**"

7. Required inputs: A summary

copernicus.eumetsat.int

ThoMaS – Workflow



8. Run the code

Example 1: Indian Ocean

1. You want to download Sen-3A data overpassing the point location **(26 S, 100 E)** in the Indian Ocean, west of Australia, where I deployed an in situ instrument measuring continuously during the first 10 days of June 2022.

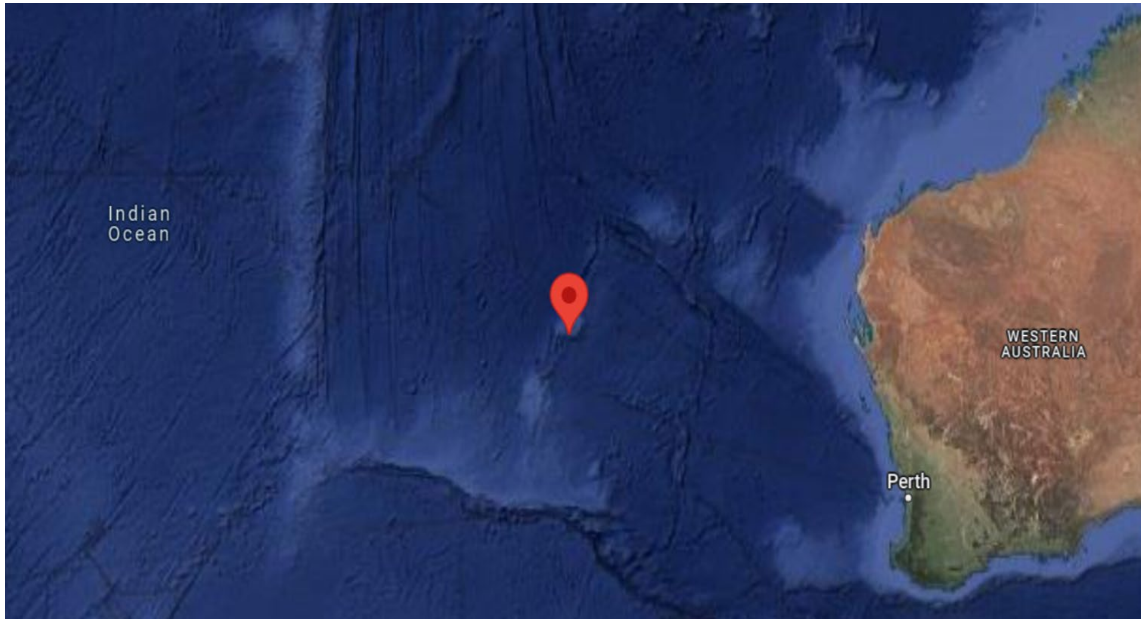
2. You want data from S3A, and L2 of the recent collection OL__L2M.003

3. Only full resolution (FR).

4. You just want to obtain the S3 files (SatData), minifiles and extractions, I have my own scripts to compute the statistics of the comparison with insitu.

5. In particular, you want to test EUMETSAT's standard protocol for window size: 3x3.

6. You want everything related to the run to be stored at /path/to/Indian_Ocean



Since insitu not provided, define **latLonTimeRanges.csv** and store it under ~/Indian_Ocean

StationID	Lat	Lon	time_start	time_stop
Indian_Ocean	-26	100	2022-06-01T00:00:00	2022-06-10T00:00:00

8. Run the code

Example 1: Indian Ocean

1. You want to download Sen-3A data overpassing the point location **(26 S, 100 E)** in the Indian Ocean, west of Australia, where I deployed an in situ instrument measuring continuously during the first 10 days of June 2022.

2. You want data from S3A, and L2 of the recent collection OL__L2M.003

3. Only full resolution (FR).

4. You just want to obtain the S3 files (SatData), minifiles and extractions, I have my own scripts to compute the statistics of the comparison with insitu.

5. In particular, you want to test EUMETSAT's standard protocol for window size: 3x3.

6. You want everything related to the run to be stored at /path/to/Indian_Ocean

Your configuration file must look like:

```
[global]
path_output: /path/to/Indian_Ocean
SetID: Indian_Ocean

[workflow]
workflow: SatData, minifiles, EDB

[satellite]
satellite_path-to-SatData: /path/to/Indian_Ocean/SatData
satellite_source: EUMETSATdataStore
satellite_collections: OL__L2M.003
satellite_platforms: S3A
satellite_resolutions: FR

[minifiles]
minifiles_winSize: 3

[EDB]
EDB_protocols_L2: EUMETSAT_standard_L2
EDB_winSizes: 3
```

Since insitu not provided, define **latLonTimeRanges.csv** and store it under ~/Indian_Ocean

StationID	Lat	Lon	time_start	time_stop
Indian_Ocean	-26	100	2022-06-01T00:00:00	2022-06-10T00:00:00

Example 6: MOBY

1. You have a prepared a set of hyperspectral Rrs insitu measurements from MOBY in SeaBASS format not corrected for BRDF effects.

2.You wish to get matchups between this MOBY subset and
- •S3A/OLCI standard FR L2,
- •From the current collection OL__L2M.003
- •using the standard extraction protocol from EUMETSAT,
- •an extraction window of 5x5,
- •an insitu-satellite time difference threshold of 1 hour (3600 seconds).

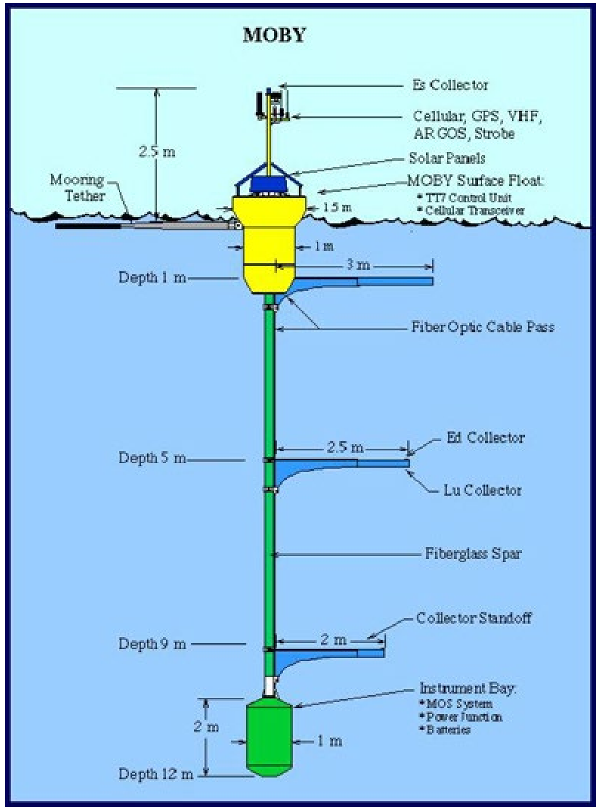
3.You are not interested in getting ancillary data from ECMWF for to the insitu data.

4.You want to apply the Morel et al. 2002 BRDF correction to both satellite and insitu.

5.You may have several insitu measurements corresponding to one single SatData within the time window that you selected, but you wish to keep only the closest in time with the satellite overpass.

6.You wish: SatData to be stored at /path/to/MOBY/SatData

7.all the other outputs (IDB, minifiles, EDB, MDB, etc.) to be stored at /path/to/MOBY



In this case, **latLonTimeRanges.csv** will be generated automatically by ThoMaS (based on your inputted insitu lat-lon and timestamps + your inputted time tolerance in config_file.ini) and stored under path_output

Example 6: MOBY

1. You have a prepared a set of hyperspectral Rrs insitu measurements from MOBY in SeaBASS format not corrected for BRDF effects.

2.You wish to get matchups between this MOBY subset and
- S3A/OLCI standard FR L2,
- From the current collection OL__L2M.003
- using the standard extraction protocol from EUMETSAT,
- an extraction window of 5x5,
- an insitu-satellite time difference threshold of 1 hour (3600 seconds).

3.You are not interested in getting ancillary data from ECMWF for to the insitu data.

4.You want to apply the Morel et al. 2002 BRDF correction to both satellite and insitu.

5.You may have several insitu measurements corresponding to one single SatData within the time window that you selected, but you wish to keep only the closest in time with the satellite overpass.

6.You wish: SatData to be stored at /path/to/MOBY/SatData

7.all the other outputs (IDB, minifiles, EDB, MDB, etc.) to be stored at /path/to/MOBY

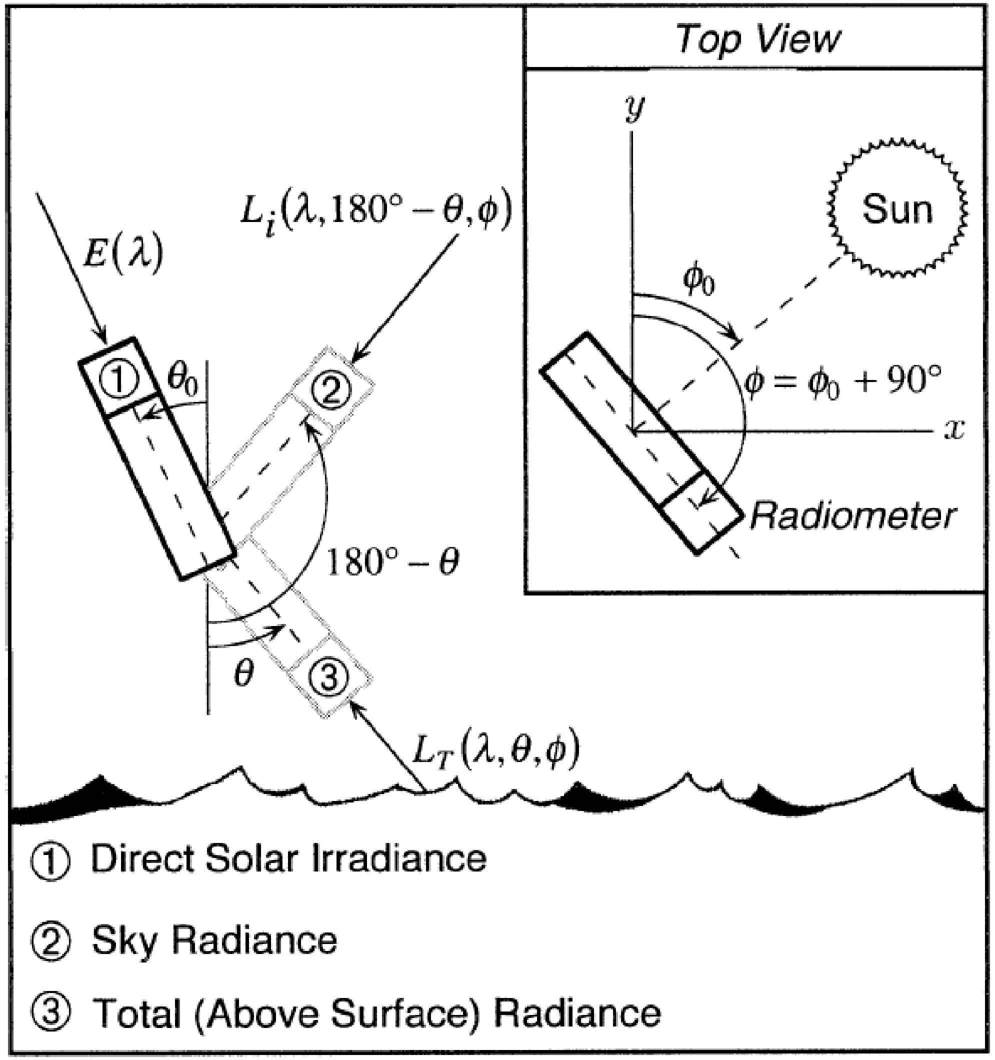
In this case, **latLonTimeRanges.csv** will be generated automatically by ThoMaS (based on your inputted insitu lat-lon and timestamps + your inputted time tolerance in config_file.ini) and stored under path_output

Your configuration file must look like:

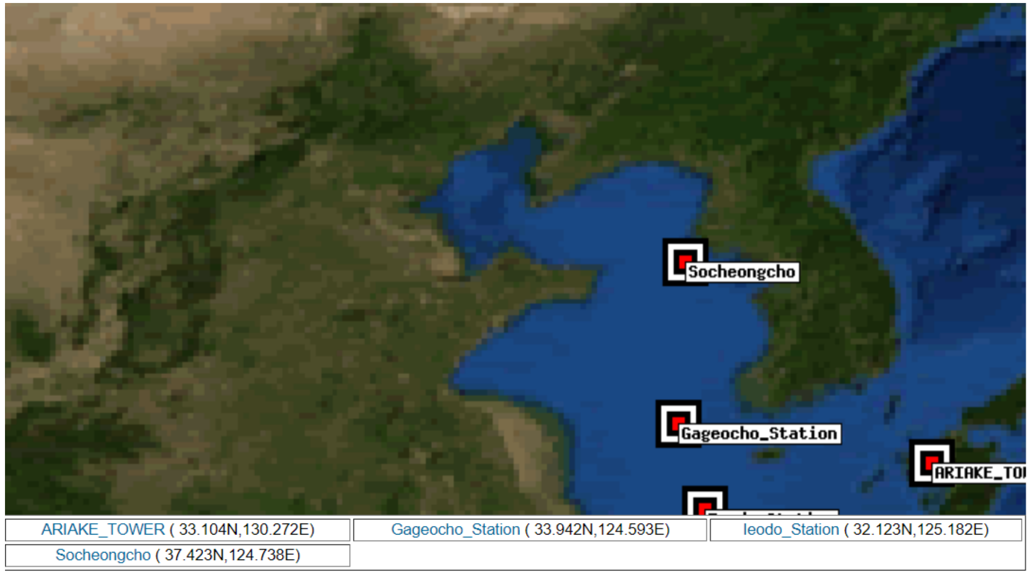
```
[global]
path_output: /path/to/MOBY
SetID: MOBY

[workflow]
workflow: insitu, SatData, minifiles, EDB, MDB

[insitu]
insitu_input: /path/to/MOBY/MOBY_OCDB.csv
insitu_satelliteTimeToleranceSeconds: 3600
insitu_getAncillary: False
insitu_BRDF: M02

[satellite]
satellite_path-to-SatData: /path/to/MOBY/SatData
satellite_source: EUMETSATdataStore
satellite_collections: OL__L2M.003
satellite_platforms: S3A
satellite_resolutions: FR
satellite_BRDF: M02

[minifiles]
minifiles_winSize: 5

[EDB]
EDB_protocols_L2: EUMETSAT_standard_L2
EDB_winSizes: 5

[MDB]
MDB_time-interpolation: insitu2satellite_NN
MDB_stats_plots: True
MDB_stats_protocol: EUMETSAT_standard_L2
```

etsat.int

Example 8: Socheongcho



[Zibordi et al. 2002]



Ongjin Socheongcho Ocean Research Station

8. Run the code

Example 8: Socheongcho

1. You want to test the performance of OLCI at the AERONET-OC station Socheongcho, West of Korea (Yellow Sea) during March 2021.

2. You wish to get matchups between this Socheongcho subset and
   - S3A/OLCI standard FR L2,
   - From the current collection OL__L2M.003
   - using the standard extraction protocol from EUMETSAT,
   - an extraction window of 5x5,
   - an insitu-satellite time difference threshold of 1 hour (3600 seconds).

3. You are not interested in getting ancillary data from ECMWF for to the insitu data.

4. You want to apply the Morel et al. 2002 BRDF correction to both satellite and insitu.

5. You may have several insitu measurements corresponding to one single SatData within the time window that you selected, but you wish to keep only the closest in time with the satellite overpass.

6. You wish: SatData to be stored at /path/to/Socheongcho/SatData

7. all the other outputs (IDB, minifiles, EDB, MDB, etc.) to be stored at /path/to/Socheongcho

In this case, **latLonTimeRanges.csv** will be generated automatically by ThoMaS (based on your inputted insitu lat-lon and timestamps + your inputted time tolerance in config_file.ini) and stored under path_output

```
[global]
path_output: /path/to/Socheongcho
SetID: Socheongcho

[workflow]
workflow: insitu, SatData, minifiles, EDB, MDB

[AERONETOC]
AERONETOC_pathRaw:/path/to/AERONET_OC_raw
AERONETOC_dateStart: 2021-03-01T00:00:00
AERONETOC_dateEnd: 2021-04-01T00:00:00
AERONETOC_dataQuality: 2
AERONETOC_station: Socheongcho


[insitu]
insitu_input: /path/to/Socheongcho/Socheongcho_OCDB.csv
insitu_satelliteTimeToleranceSeconds: 3600
insitu_getAncillary: False
insitu_BRDF: M02

[satellite]
satellite_path-to-SatData: /path/to/Socheongcho/SatData
satellite_source: EUMETSATdataStore
satellite_collections: OL__L2M.003
satellite_platforms: S3A
satellite_resolutions: FR
satellite_BRDF: M02

[minifiles]
minifiles_winSize: 5

[EDB]
EDB_protocols_L2: EUMETSAT_standard_L2
EDB_winSizes: 5

[MDB]
MDB_time-interpolation: insitu2satellite_NN
MDB_stats_plots: True
MDB_stats_protocol: EUMETSAT_standard_L2
```

int

8. Run the code

1. EUMETSAT Data Store credentials obtained and stored?
2. (optional) ECMWF ADS/CDS credentials obtained and stored?
3. ThoMaS code cloned?
4. thomas conda environment set up and activated?
5. Required inputs in place? (config_file.ini, insitu input file?, lat-lon-time ranges file?)

If **YES**.. run by executing this command:
python /path/to/ThoMaS/main.py –cf /path/to/config_file.ini

Hope you enjoyed it! Thank you!

Questions are welcome.

Getting started…

- Installing Anaconda
  - If you don't have Python…

- Cloning repositories
  - OR
  - **On windows:** Open Anaconda, find and launch the powershell.exe application. This will open a command line window.
  - **On Linux or OSx:** Open a terminal window
  - **Clone the learn-olci and ThoMaS repositories:** (you may wish to change directory first)
    - Copy and past the following lines into your terminal and hit "enter"
    - **learn-olci:** *git clone --recurse-submodules --remote-submodules https://gitlab.eumetsat.int/eumetlab/oceans/ocean-training/sensors/learn-olci.git*
    - **ThoMaS:** *git clone https://gitlab.eumetsat.int/eumetlab/oceans/ocean-science-studies/ThoMaS.git*

- Setting up Python environments
  - To create the required Python environments, you should copy and past the following lines into your terminal and hit "enter"
  - **learn-olci:** *conda env create -f learn-olci/environment.yml*
  - **ThoMaS:** *conda env create -f ThoMaS/environment.yml*

Practical session

copernicus.eumetsat.int

- If you haven't yet – clone and set up the git repositories for learn-olci/ThoMaS as per the instructions in the README
  - Note the submoduling for learn-olci
  - Extensions are not needed
- Key elements of both include setting up the credentials for access (file with your consumer key and secret).
  - See 1_OLCI_introductory notebooks on data store access
  - Can also see API_authentication notebook for support
- Explore notebooks for data access depending on your experience level/needs
- Run examples **1**, **6** and **8** from ThoMaS (either command line or jupyter notebook ThoMaS_overview.ipynb)

Thank you!

Questions are welcome.

Contacts and further information

For information on our training programme

training@eumetsat.int

For information/support on EUMETSAT services

ops@eumetsat.int

For our training calendar

https://trainingevents.eumetsat.int/trui/