EUMETSAT

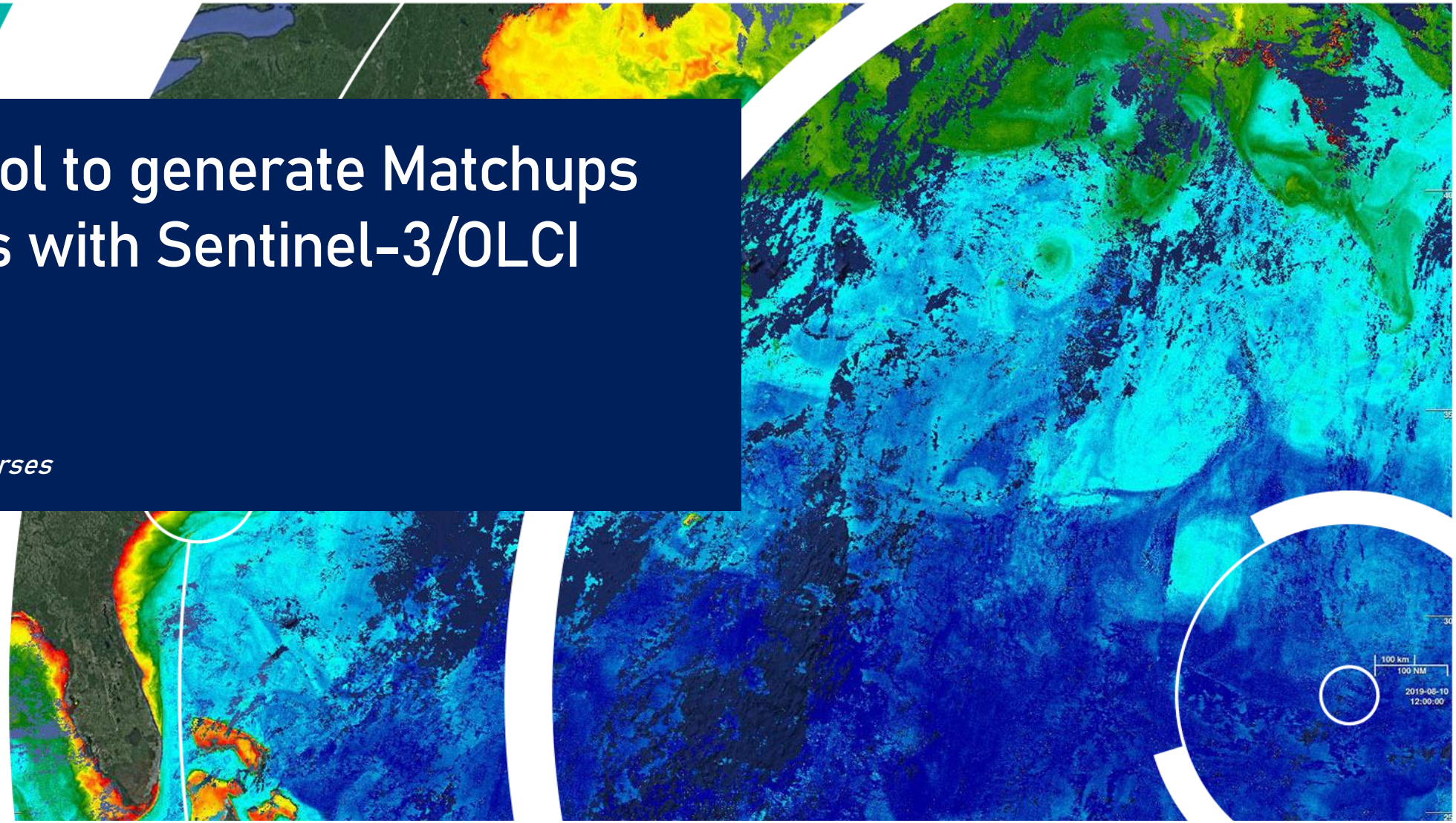
PROGRAMME OF THE EUROPEAN UNION

Copernicus
Europe's eyes on Earth

# ThoMaS – a Tool to generate Matchups of OC products with Sentinel-3/OLCI

Juan Ignacio Gossn
*EUMETSAT*

*EUMETSAT series of short courses*



# Today's tour with ThoMaS



1. What's ThoMaS? Scope
2. Usage
3. Some background
4. Pre-requisites
5. Getting the code
6. Setting the environment
7. Set-up demo
8. Required inputs
9. Run the code
10. Run the code: demo
11. Short tour around ThoMaS

Please interrupt me at any time!
This presentation will be available and most material already available in ThoMaS repository

# 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)**.

# 1. What's ThoMaS? Scope



ThoMaS is a toolkit developed to create **matchups** of bio-geophysical **insitu data (in SeaBASS/OCDB format)** with **satellite ocean colour products** from **Sentinel-3 OLCI (S3/OLCI) (also standard products from NASA's OBPG – l2gen - partially supported)**.

# 1. What's ThoMaS? Scope



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

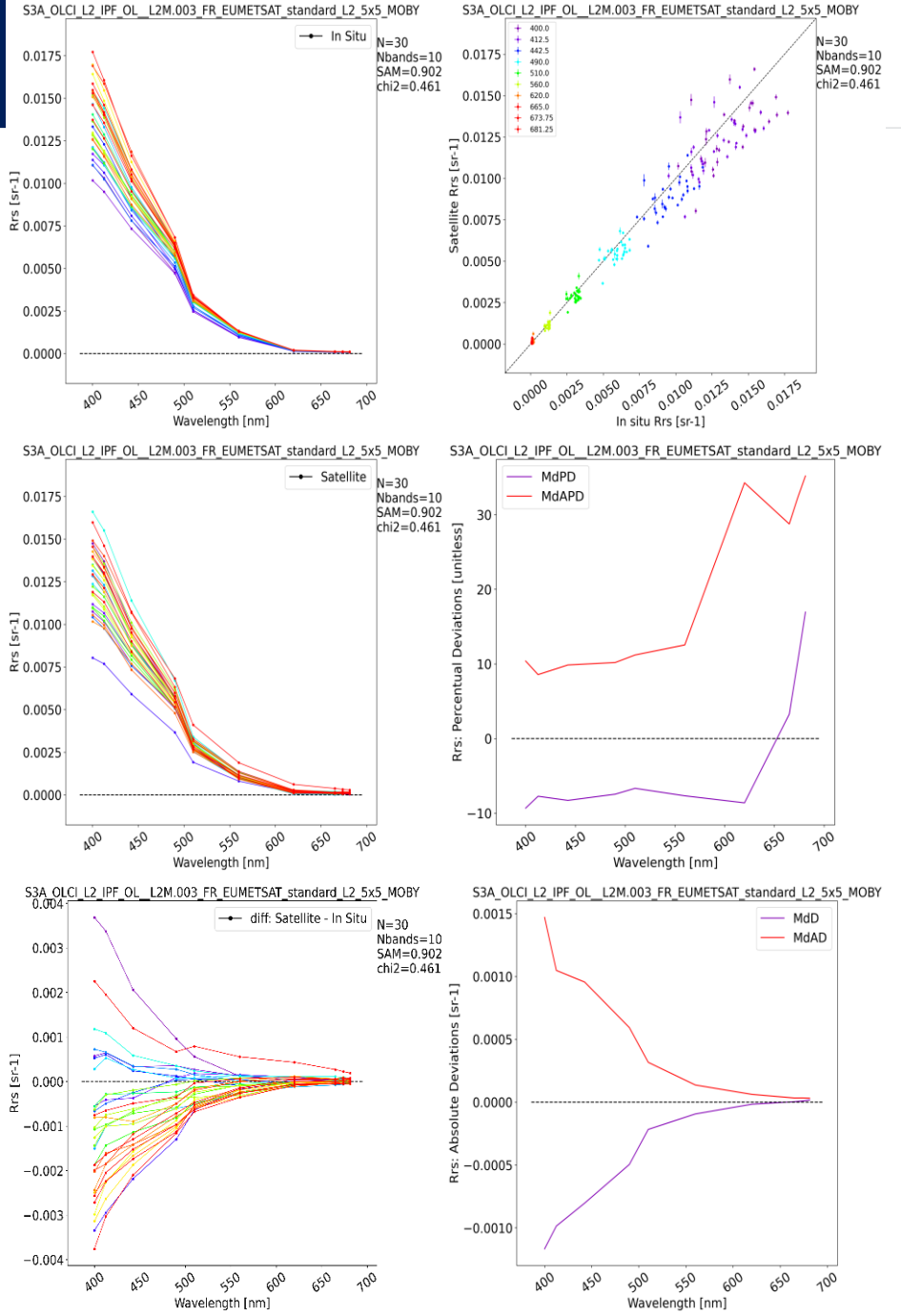
- → **Insitu** data transformed to **spectrally match satellite** (**convolution, spectral matching**).
- → **Satellite** data (L1B – TOA radiance - or L2 – BOA water reflectance) from **EUMETSAT Data Store** (irresp. to whether it's reprocessed or operational) matching your insitu
- → **Extractions** of satellite sata centred at lat/lon/time 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**.

# 1. What's ThoMaS? Scope

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

- → **Insitu** data transformed to **spectrally match satellite (convolution, spectral matching)**.
- → **Satellite** data (L1B – TOA radiance - or L2 – BOA water reflectance) from **EUMETSAT Data Store** (irresp. to whether it's reprocessed or operational) matching your insitu
- → **Extractions** of satellite sata centred at lat/lon/time 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.



# 2. 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.

# 2. 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**: Ingest insitu data from **SeaBASS/OCDB** input file and apply several transformations to make **insitu comparable to satellite** data (e.g. **spectral matching** with satellite, **BRDF** correction) and 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. 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**: Ingest insitu data from **SeaBASS/OCDB** input file and apply several transformations to make **insitu comparable to satellite** data (e.g. **spectral matching** with satellite, **BRDF** correction) and 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**: Obtain/list the **satellite products (L1B and/or L2)** matching spatially-temporally your insitu data.

# 2. 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**: Ingest insitu data from **SeaBASS/OCDB** input file and apply several transformations to make **insitu comparable to satellite** data (e.g. **spectral matching** with satellite, **BRDF** correction) and 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**: Obtain/list the **satellite products (L1B and/or L2)** matching spatially-temporally your insitu data.

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

# 2. 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**: Ingest insitu data from **SeaBASS/OCDB** input file and apply several transformations to make **insitu comparable to satellite** data (e.g. **spectral matching** with satellite, **BRDF** correction) and 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**: Obtain/list the **satellite products (L1B and/or L2)** matching spatially-temporally your insitu data.

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**. Stack minifiles, apply transformations to SatData to make them comparable to in situ (e.g. **scale/unit conversion**, **BRDF** correction) and store into standard netCDF4 and CSV **EDB** (**Extraction Data Base**) files. This step includes calculating **statistics** over the extraction window following EUMETSAT's or any user-defined Matchup Protocol.

# 2. 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**: Ingest insitu data from **SeaBASS/OCDB** input file and apply several transformations to make **insitu comparable to satellite** data (e.g. **spectral matching** with satellite, **BRDF** correction) and 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.

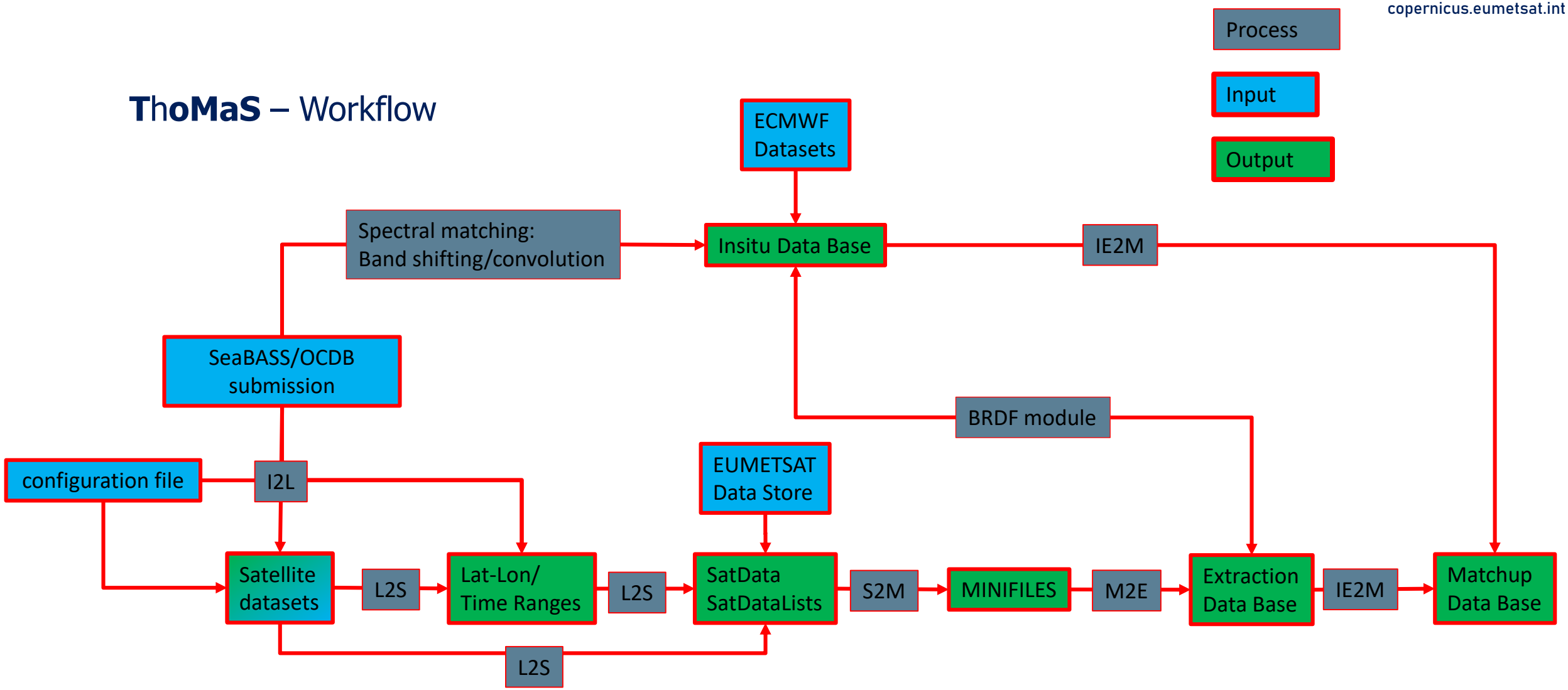
2. Step **SatData**: Obtain/list the **satellite products (L1B and/or L2)** matching spatially-temporally your insitu data.

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**. Stack minifiles, apply transformations to SatData to make them comparable to in situ (e.g. **scale/unit conversion**, **BRDF** correction) and store into standard netCDF4 and CSV **EDB** (**Extraction Data Base**) files. This step includes calculating **statistics** over the extraction window following EUMETSAT's or any user-defined Matchup Protocol.

5. Step **MDB**. Combine insitu (**IDB**) and satellite (**EDB**) information indexed into insitu-satellite **matchup pairs**, **optionally apply time interpolation**, calculate **matchup statistics**, and store into standard **MDB** (**Matchup Data Base**) netCDF4 and CSV files.

## **ThoMaS** – Workflow



# 2. Usage: Summary on the terminology



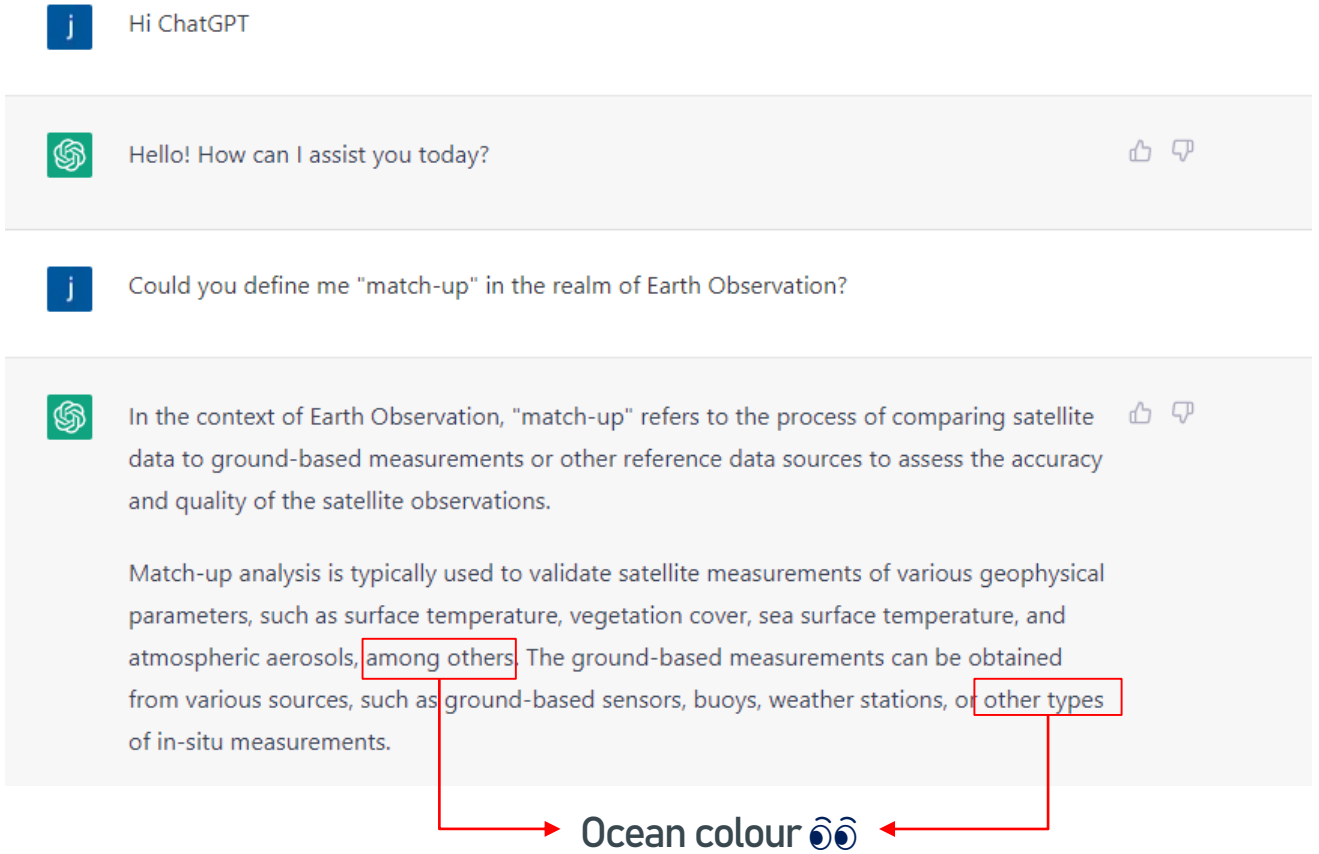
- **IDB** (In situ Data Base): a netCDF file containing all the information related to the inputted insitu data. + (if requested) ancillary information from **ECMWF reanalysis datasets** at the insitu geographic location and time stamp.
- **SatData**: an image file/directory. In the case of standard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

# 3. Some background: match-ups



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

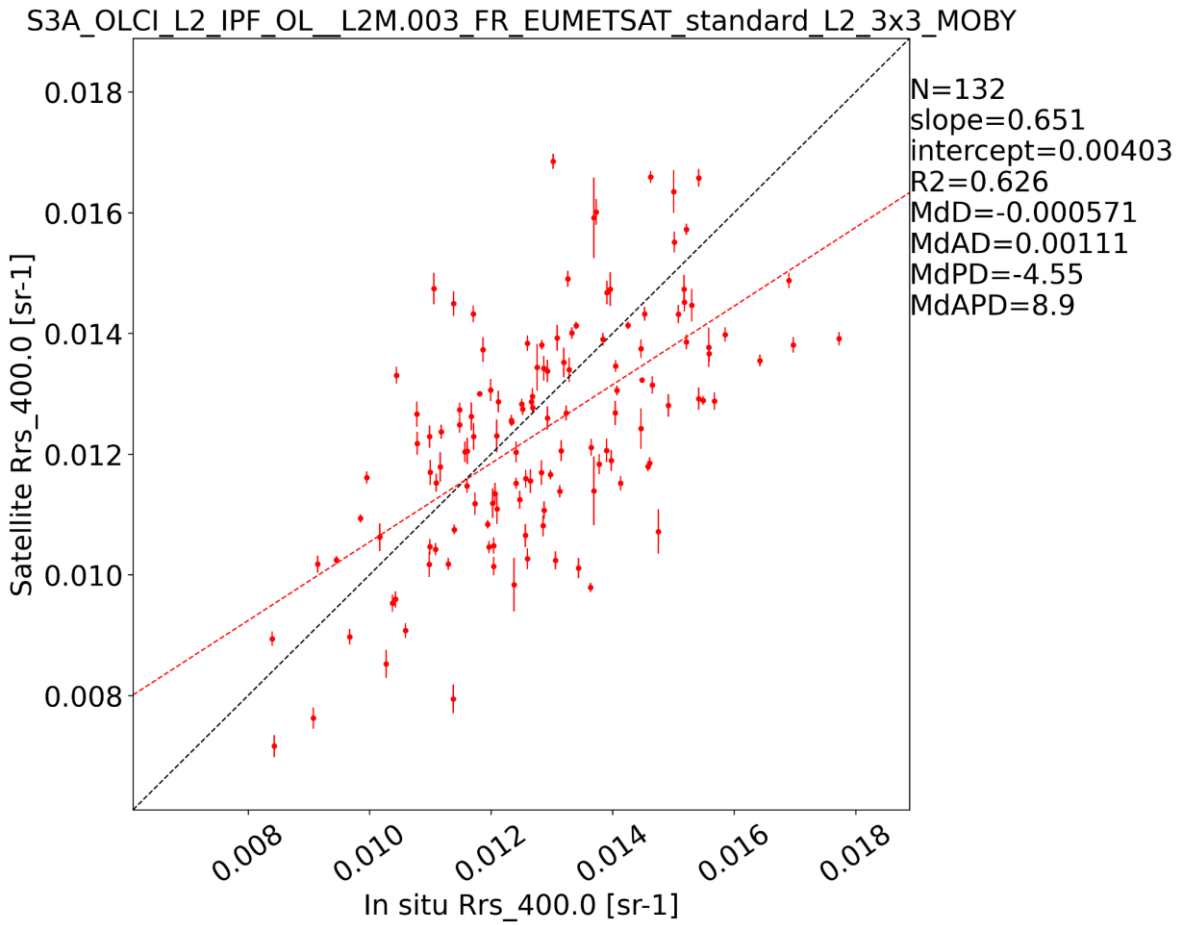




- Of course we have much more to define… and take care of…

# 3. Some background: match-ups

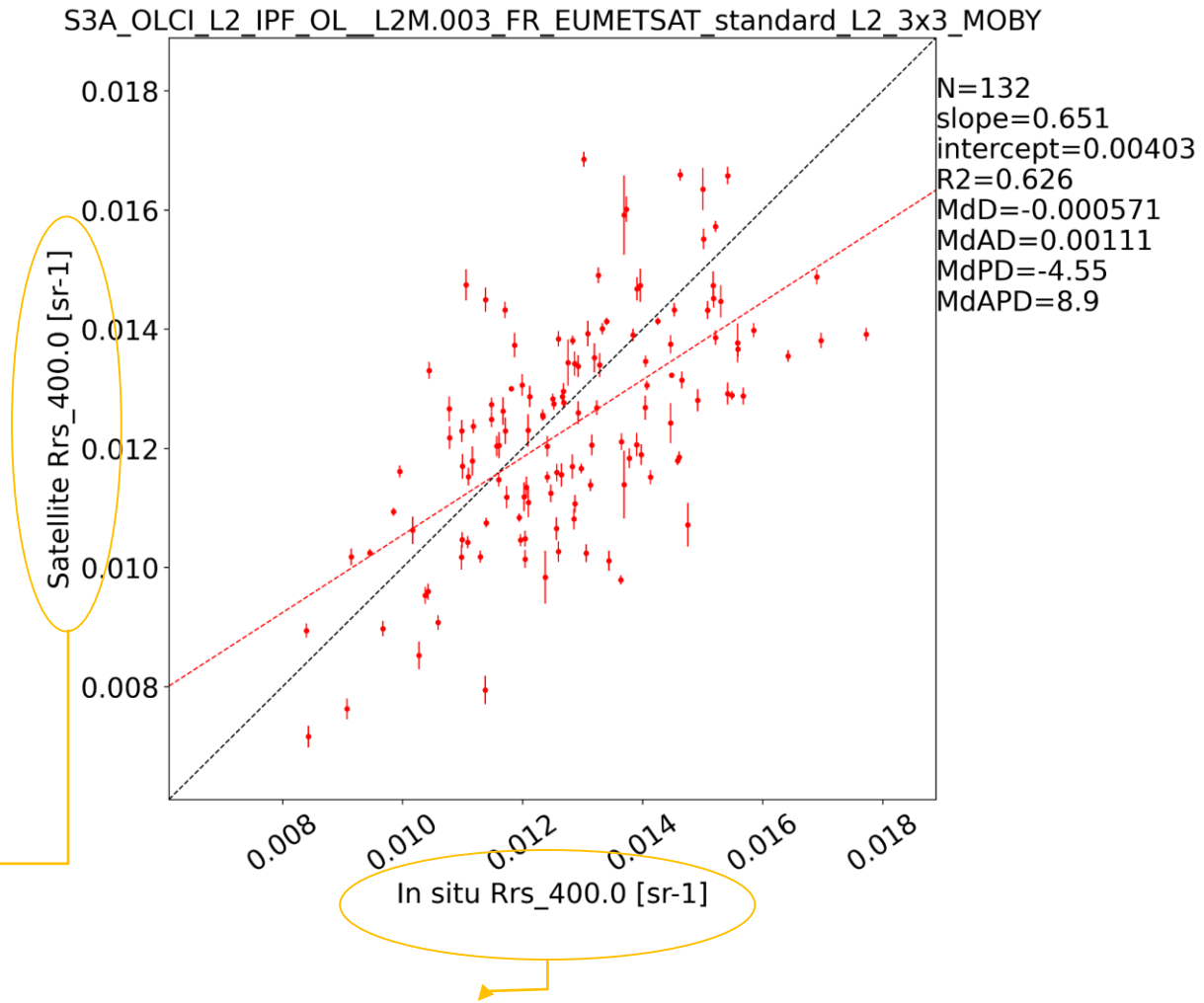





- Of course we have much more to define… and take care of…

# 3. Some background: match-ups






**Are we comparing 🍎 with 🍎?**

→ What is the definition of Rrs?

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

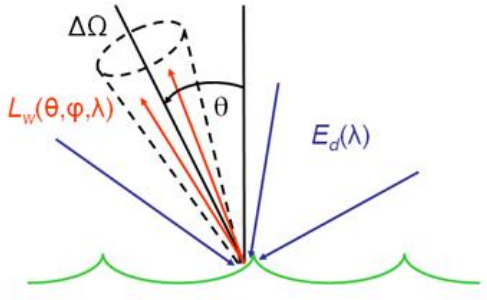
→ Are these two compatible "directionally"? → BRDF correction

# 3. Some background: match-ups

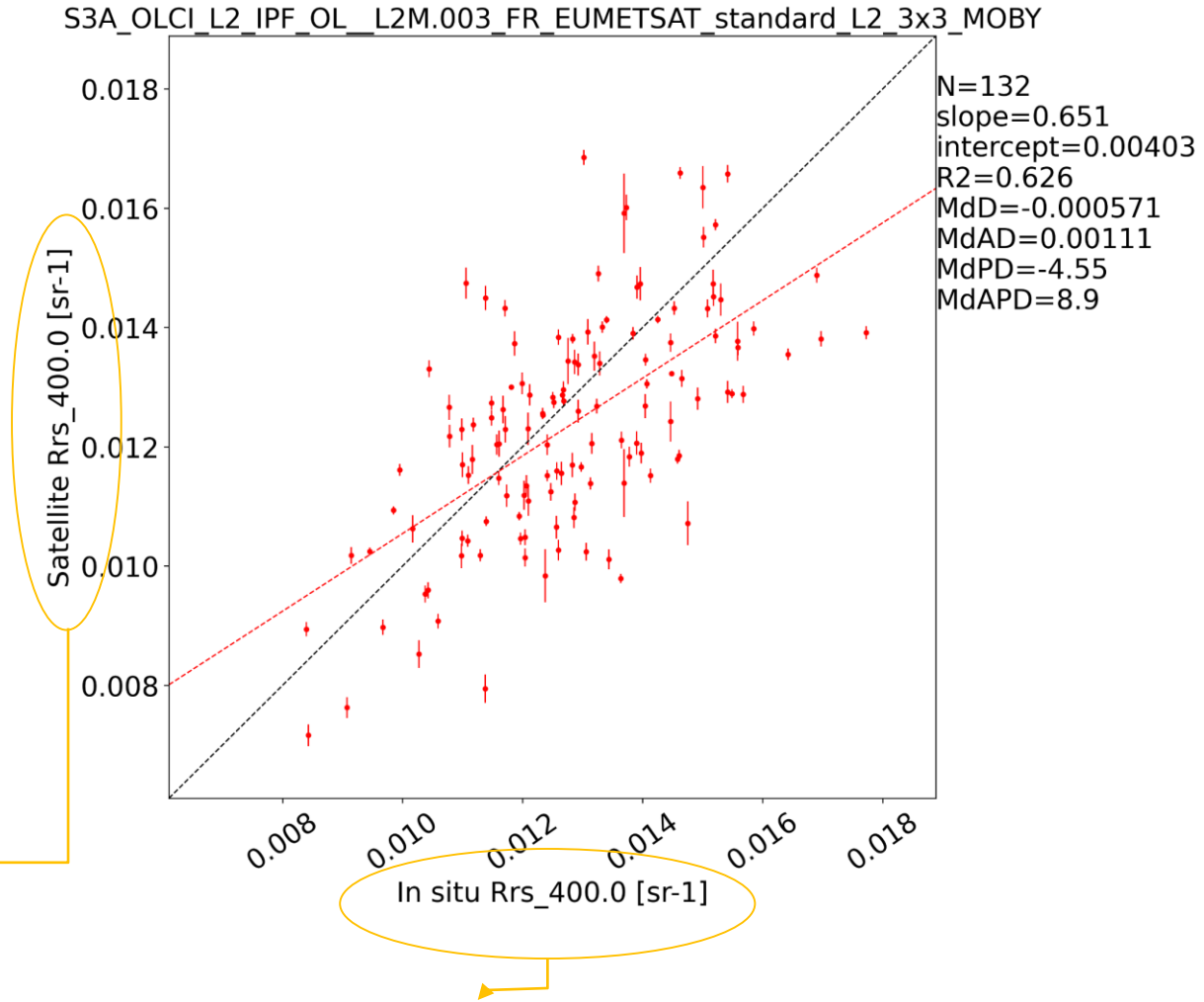


Definition of Rrs

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




OO Web Book, Mobley, Boss & Roesler




**Are we comparing 🍎 with 🍎?**

→ What is the definition of Rrs?

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

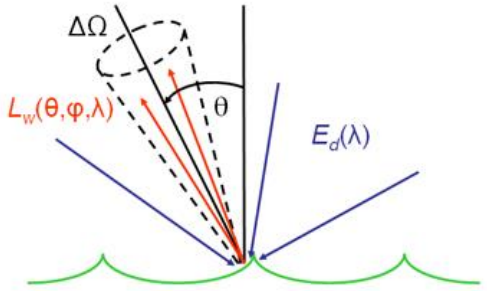
→ Are these two compatible "directionally"? → BRDF correction

# 3. Some background: match-ups

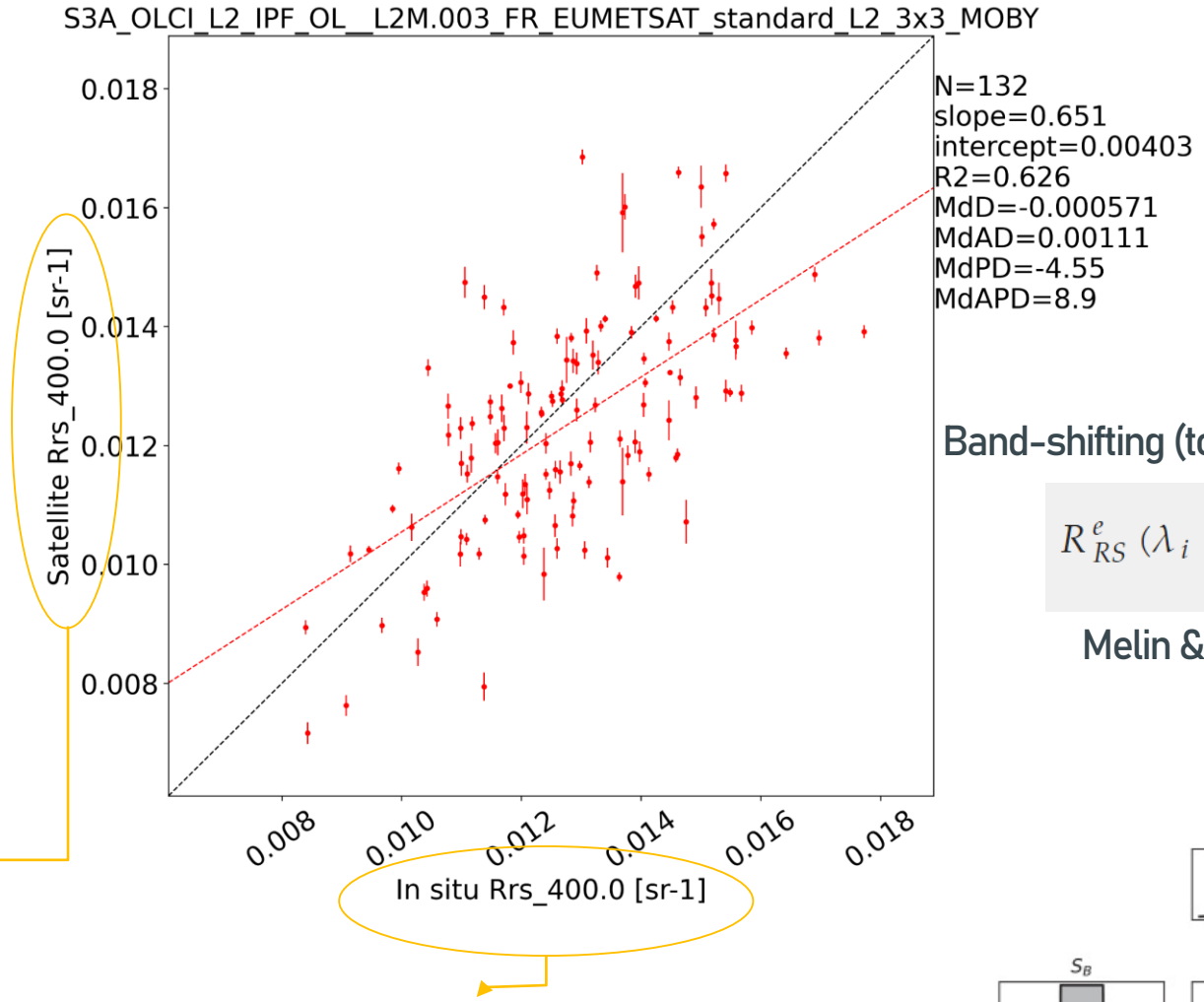


## Definition of Rrs

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




OO Web Book, Mobley, Boss & Roesler

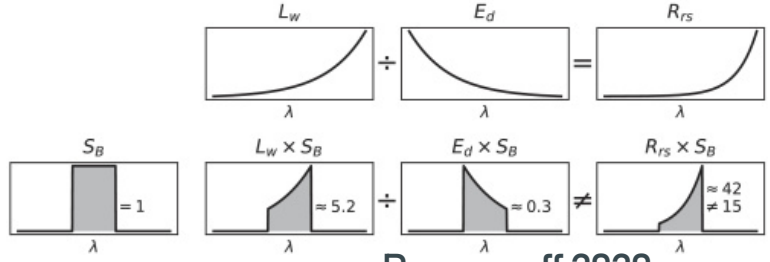



## Band-shifting (to pair multispectral to multispectral)

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

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

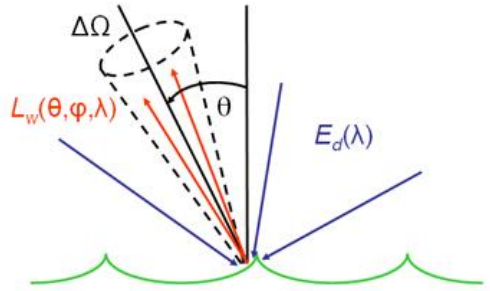
→ Are these two compatible "directionally"? → BRDF correction

# 3. Some background: match-ups



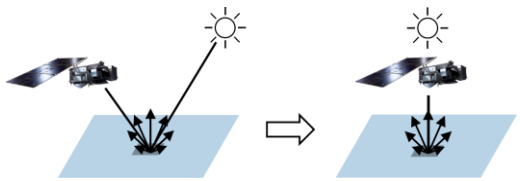
Definition of Rrs

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



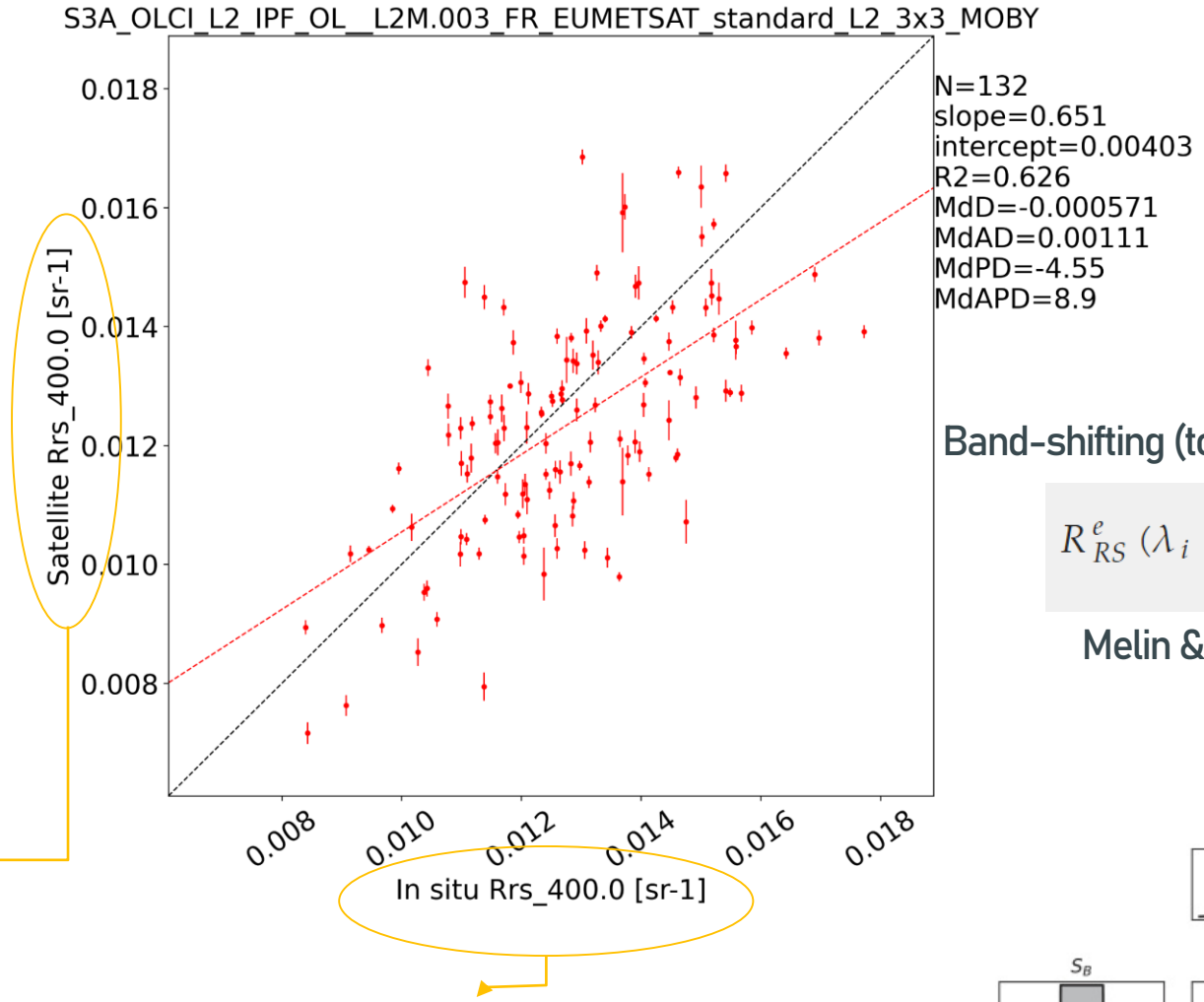

OO Web Book, Mobley, Boss & Roesler

BRDF correction:



D'Alimonte et al.

**Morel et al. 2002 supported in ThoMaS**

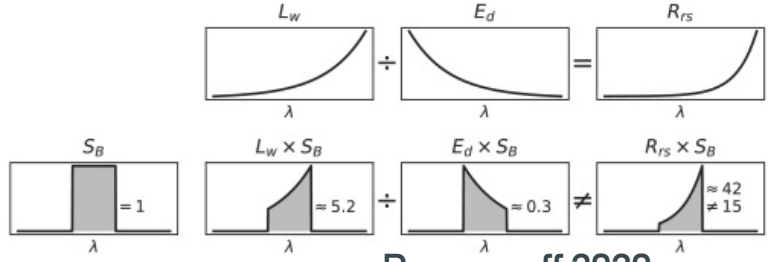



Band-shifting (to pair multispectral to multispectral)

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

Melin & Sclep 2015 supported in ThoMaS

Spectral convolution




Burggraaff 2020

**Are we comparing 🍎 with 🍎?**

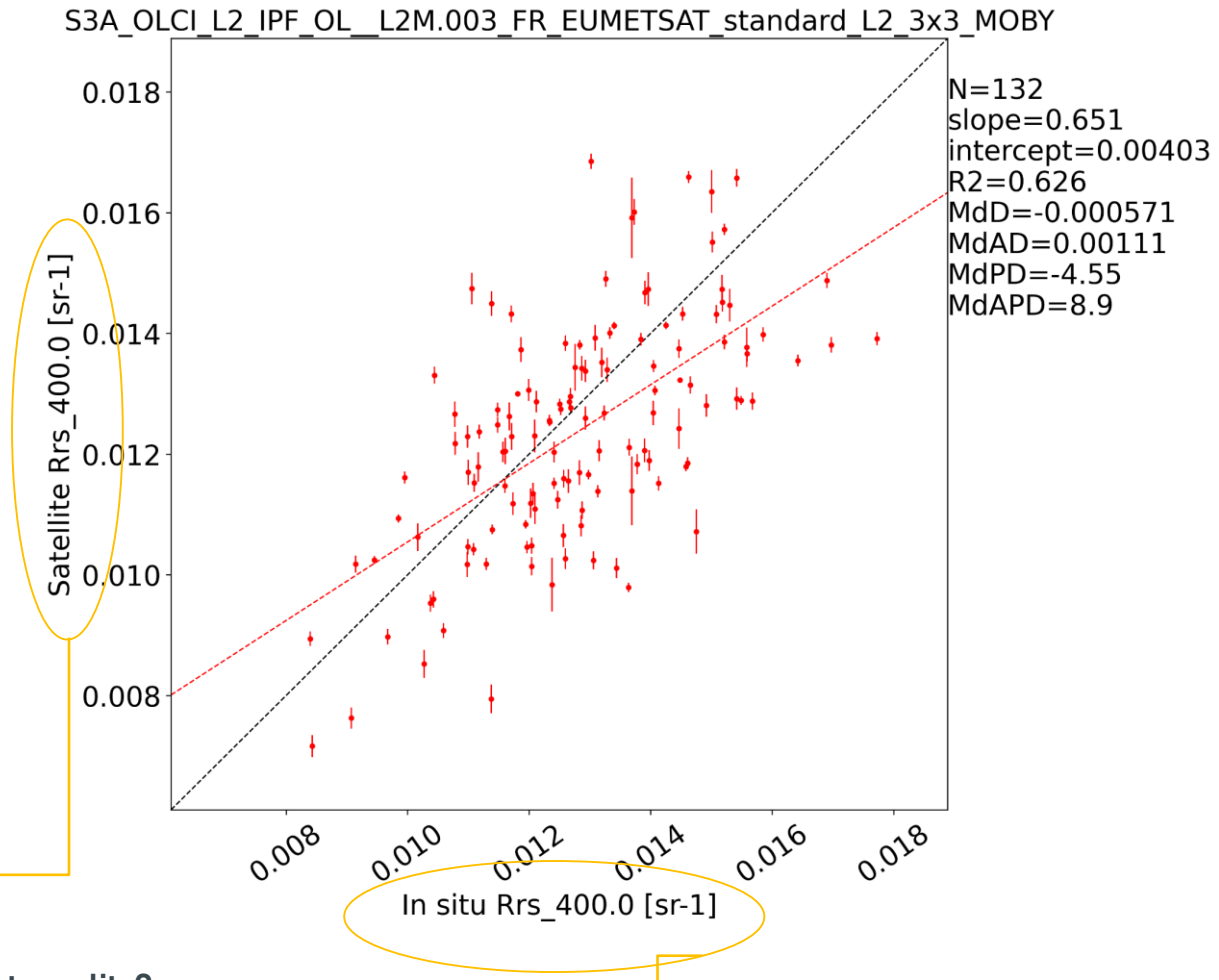
→ What is the definition of Rrs?

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

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

# 3. Some background: match-ups






**Are we comparing 🍎 with 🍎?**

→ 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?

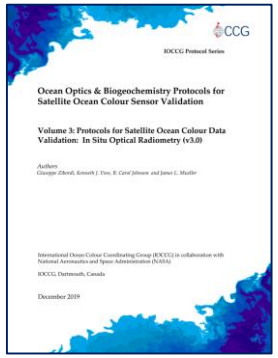
# 3. Some background: match-ups

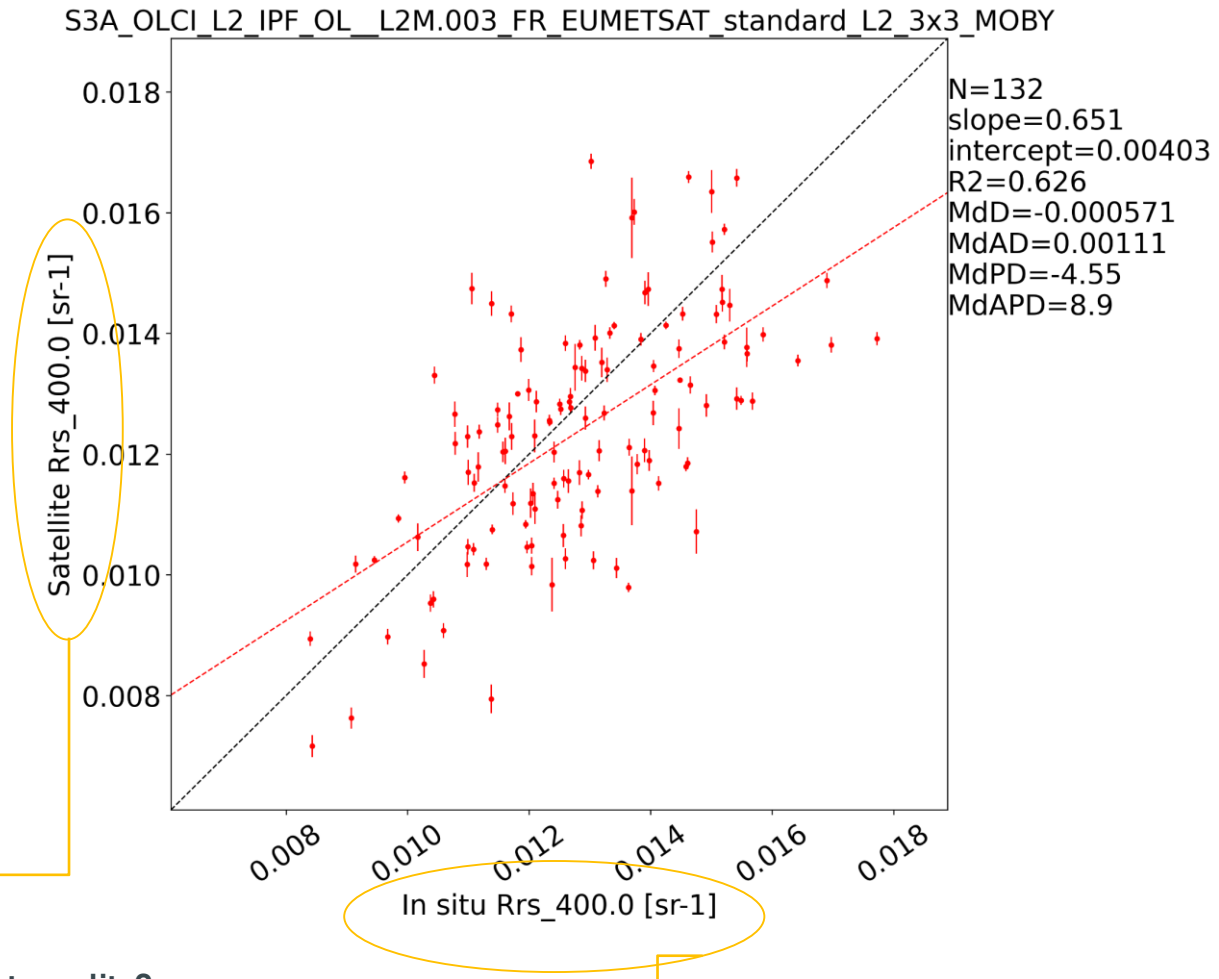


## Quality of insitu

ThoMaS still does not consider any quality flag to process insitu data... but stay tunned ☺...


fiducial reference measurements for satellite ocean colour




**Are we comparing 🍎 with 🍎?**

→ 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?

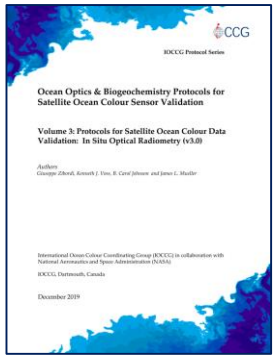
# 3. Some background: match-ups



## Quality of insitu

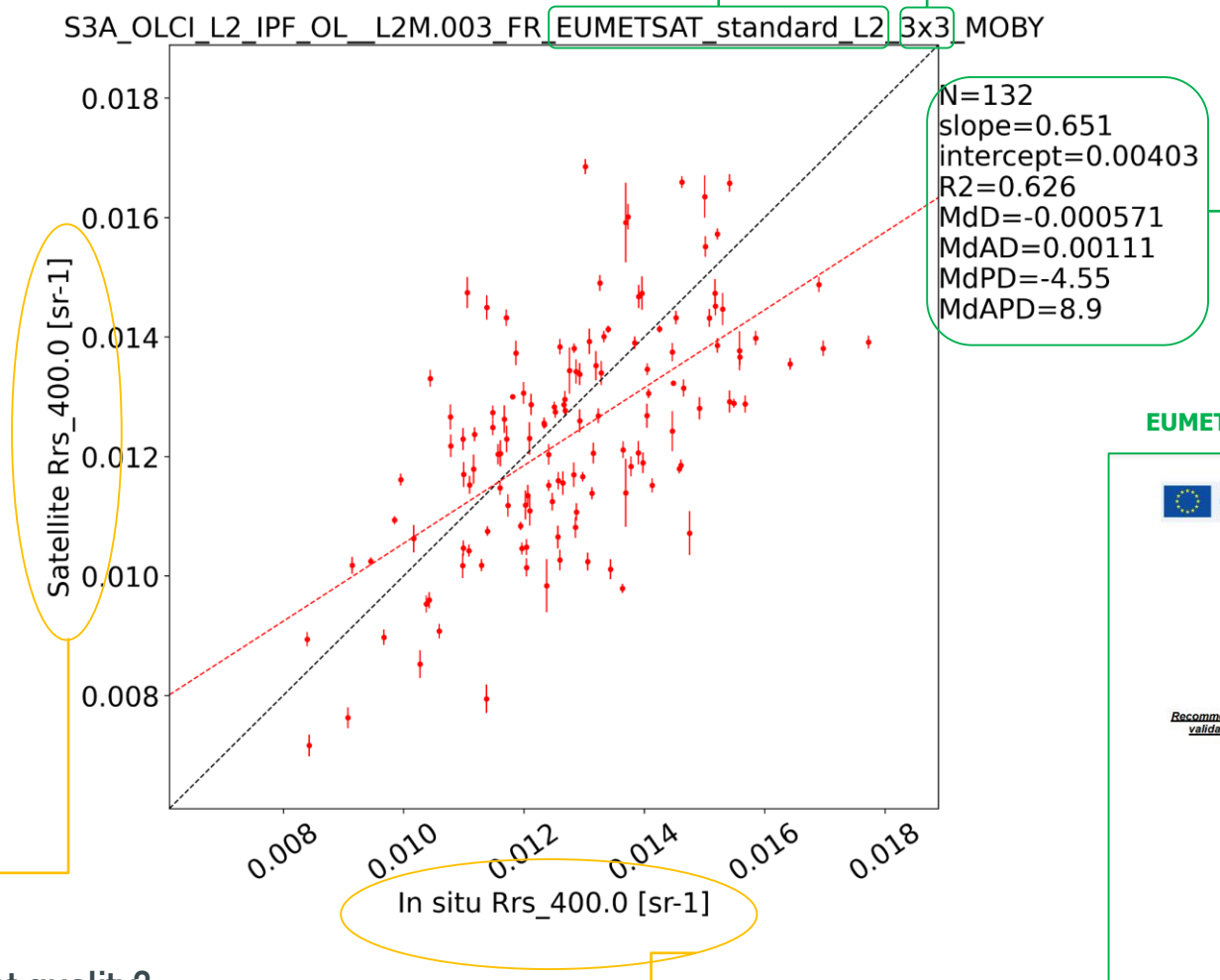
ThoMaS still **does not** consider any quality flag to process insitu data… but stay tunned ☺…


fiducial reference measurements for satellite ocean colour



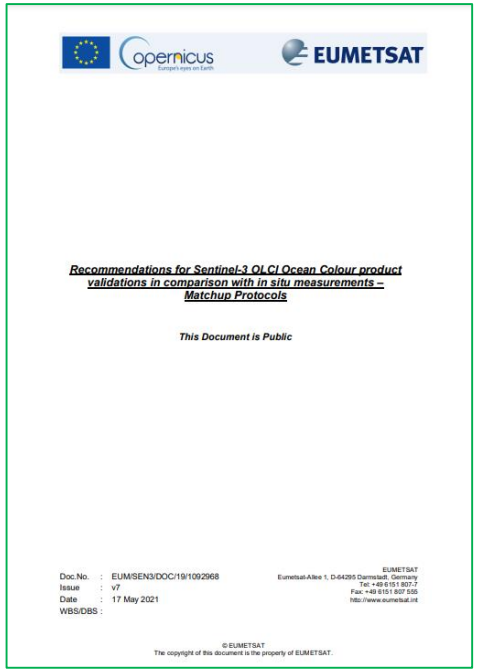

## Define your extraction statistics!

## Define your extraction window size!




## Define the matchup statistics!

## EUMETSAT's Matchup Protocols




## Are we comparing 🍎 with 🍎?

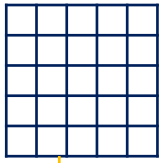
- → Are the insitu measurements of sufficient quality?
- → Are insitu and satellite measurements temporally-spatially comparable?
- → What value (and uncertainty) shall I extract from the satellite data?

# 3. Some background: match-ups: EUMETSAT extraction protocol

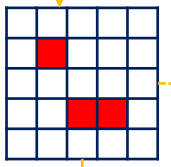
**EUMETSAT's Matchup Protocols:** extraction of statistics at macropixel level

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

**Set BFOR: 5x5 window**



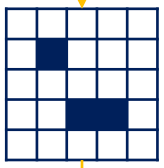
**Detection of non-valid pixels (flagged pixels)**

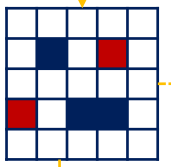


**Pixels are masked/removed if flagged by any of the following:**

*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**

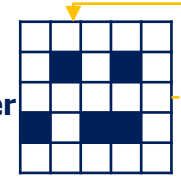




**Pixel 'X' is considered outlier if:**

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

**Set final: without flagged & outlier**



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

**Macropixel is discarded if:**

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

**Macropixel is discarded if:**

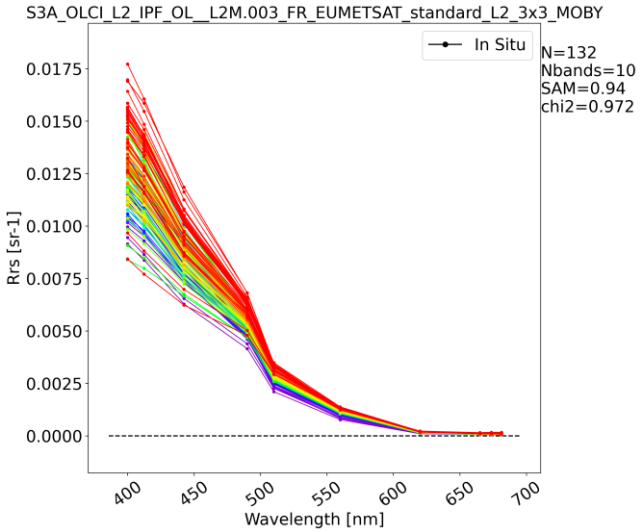
$CV_{final}(560) > 20\%$

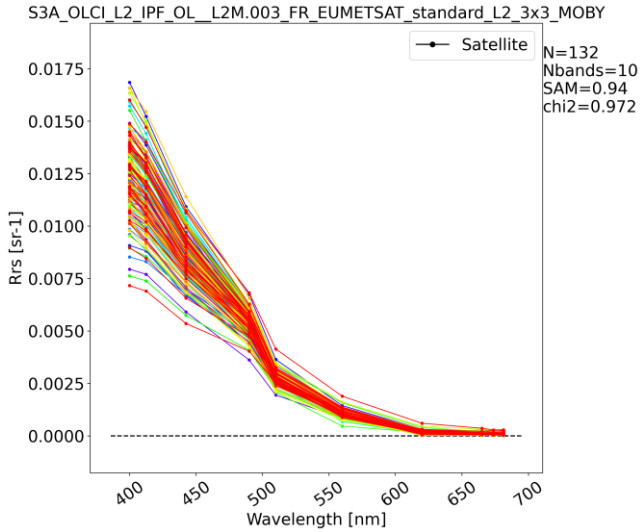
μ → Mean
σ → Standard deviation

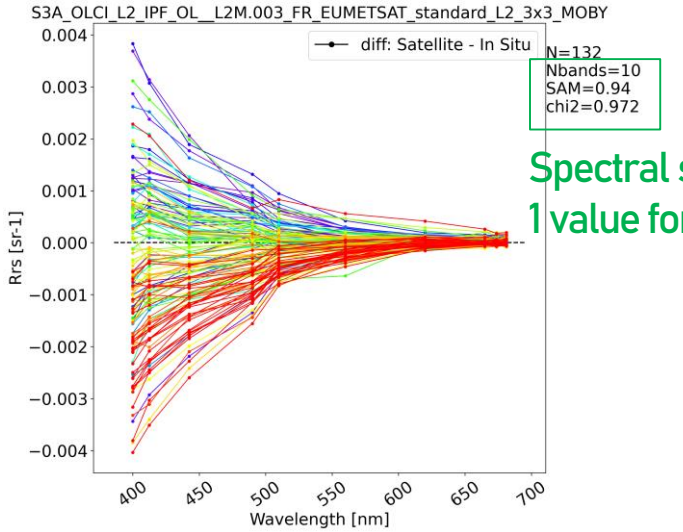
# 3. Some background: match-ups



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

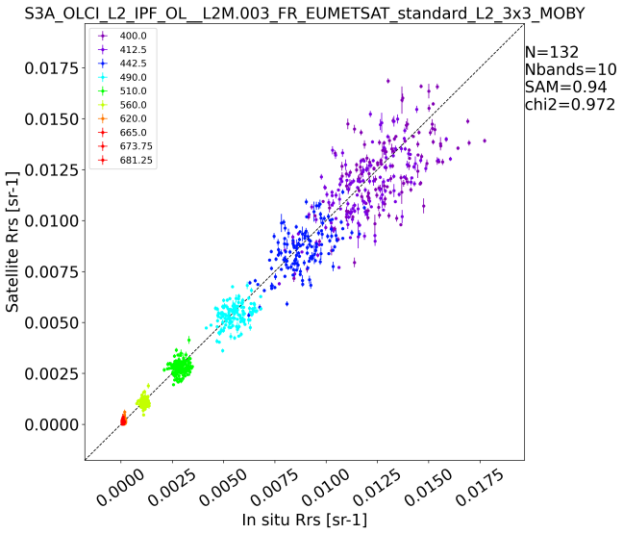


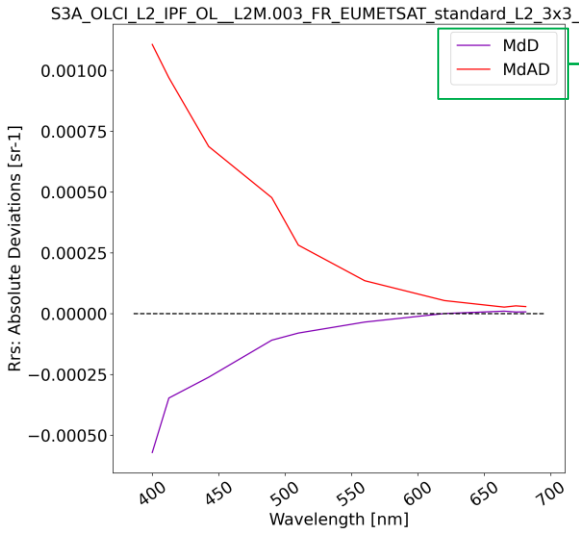



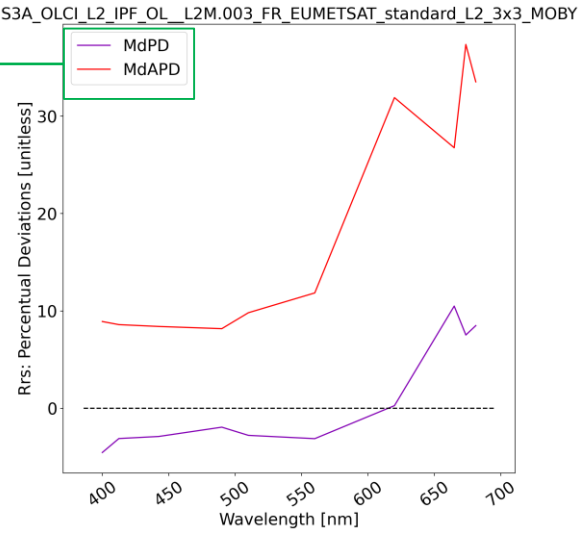




Spectral statistics
1 value for the whole set

statistics band-by-band, plotted spectrally



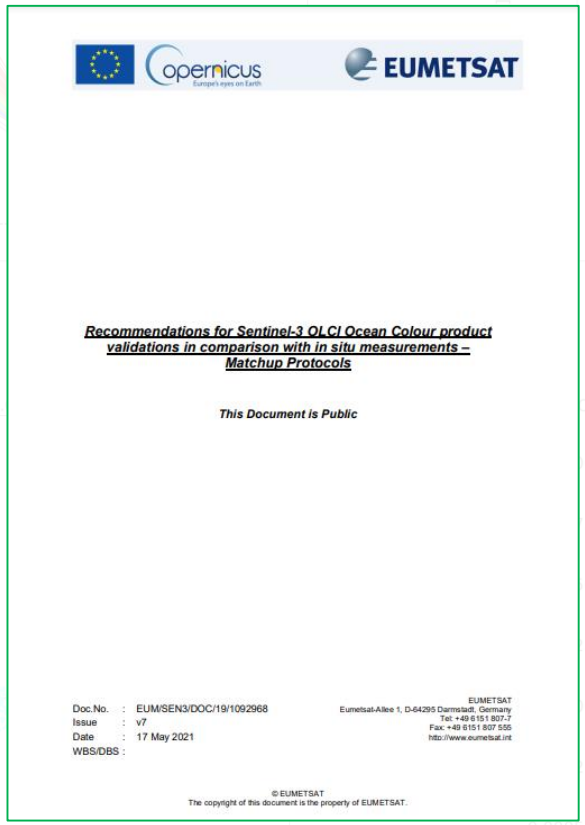



# 3. Some background: match-ups



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



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

# 3. Some background: basic (Linux) terminal commands



| Command | Description |
|---|---|
| `cd dir_name` | Change directory to “dir_name” |
| `cd ~` | Change directory to home dir |
| `mkdir dir_name` | Make new directory “dir_name” |
| `ls` | List files in current directory |
| `ls -a` | List files including hidden ones |
| `conda` | Check conda is installed |
| `conda env create –f env_file –n env_name` | Create environment “env_name” from “env_file” file |
| `conda activate env_name` | Activate environment “env_name” |
| `git` | Check git is installed |
| `git clone url .` | Clone repository from URL to current directory |
| `nano filename` | Check/edit file “filename” |

# 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.
4. EUMETSAT Data Store: Save EO Portal API credentials under **~/.eumdac/credentials**
5. ECMWF: Register to ADS/CDS and get url and key.
6. ECMWF: store ADS/CDS url/keys under **~/.ecmwg_api_config**

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 |

Conda will take care of this…

# 5. Getting the code



- Git way:

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

- Or direct download from https://gitlab.eumetsat.int/eumetlab/oceans/ocean-science-studies/ThoMaS

# 6. Setting the environment

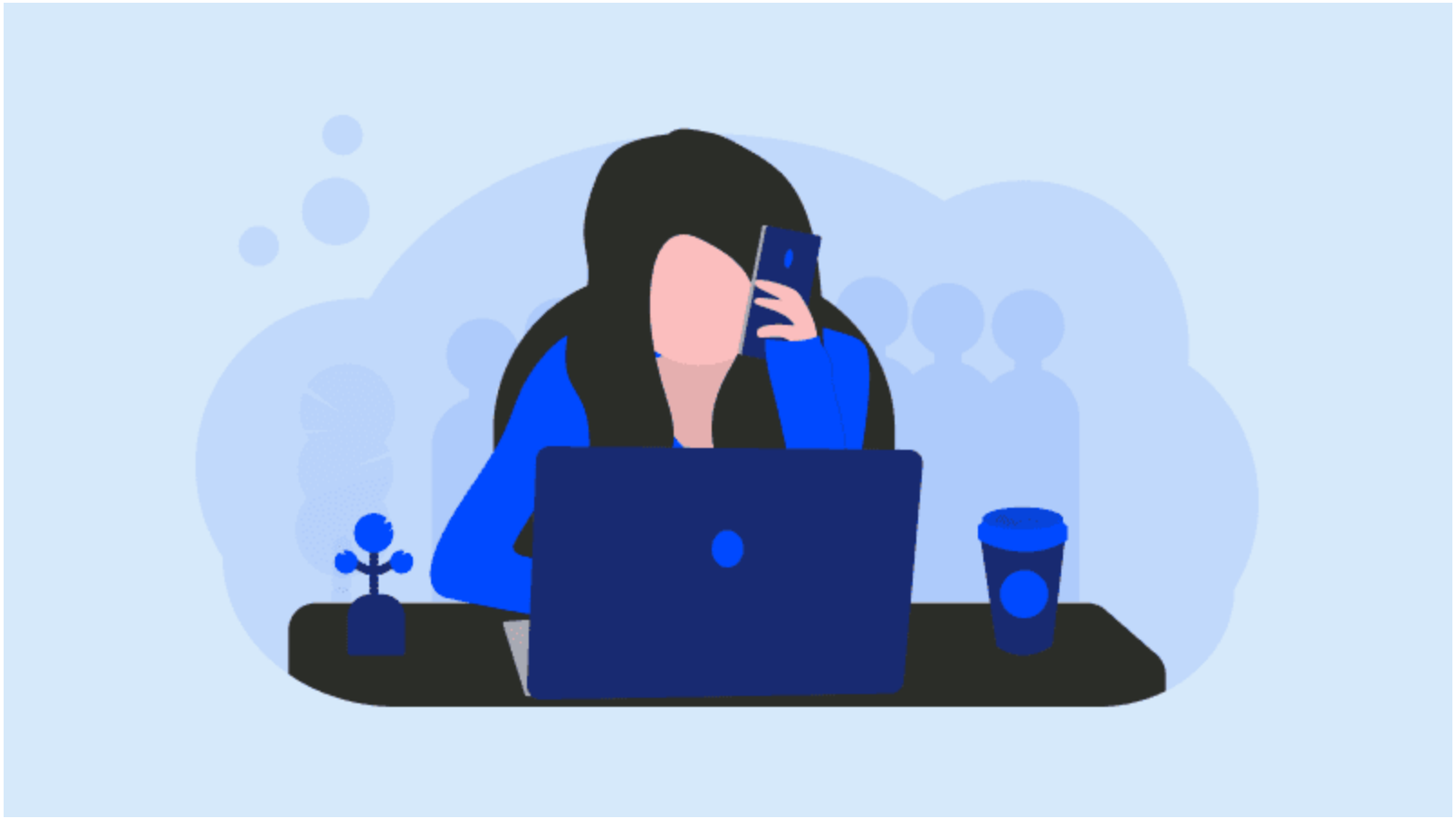


- Once conda and ThoMaS are installed:

```
conda
cd ~
cd ThoMaS
ls
nano environment.yml
conda env create -f environment.yml -n thomas
conda activate thomas
```

# 7. Set-up demo

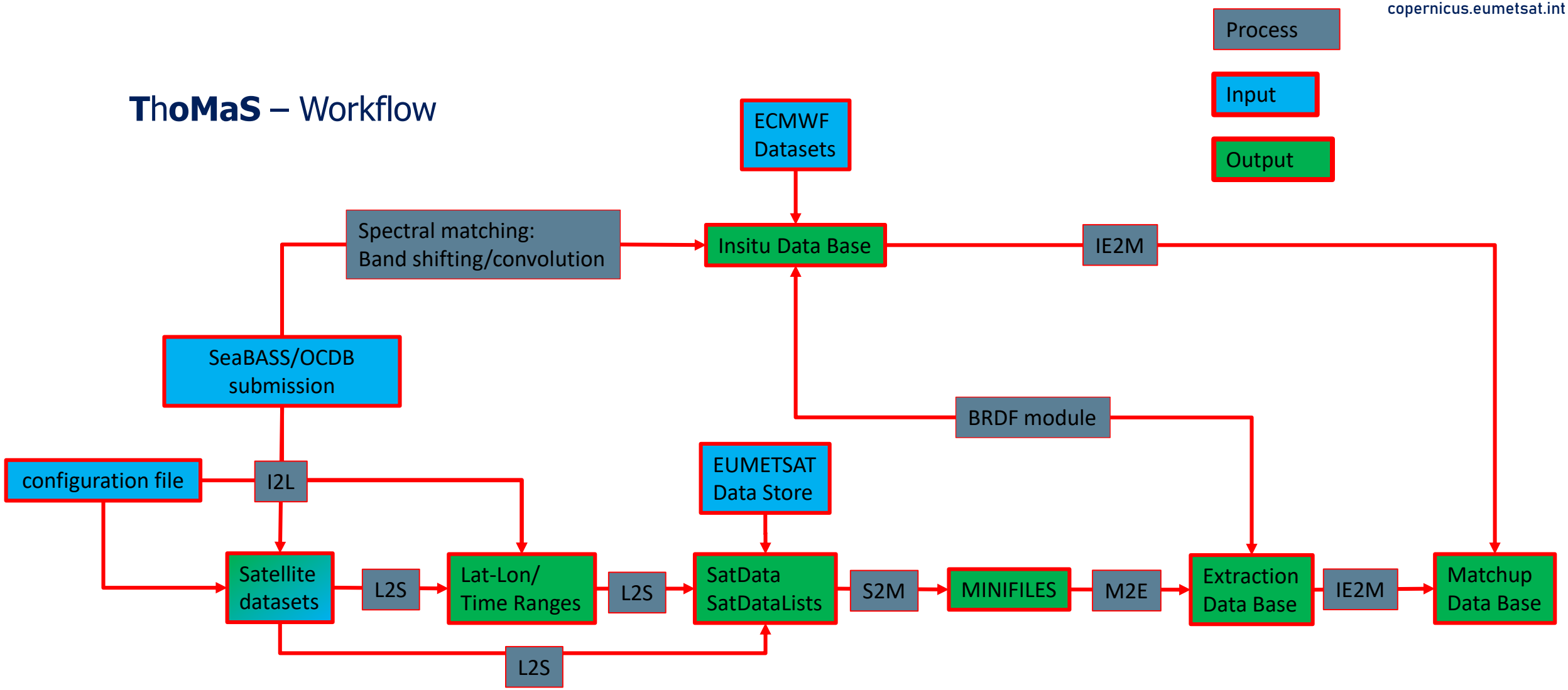




# 8. Required inputs



## **ThoMaS** – Workflow



# 8. Required inputs: 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
/wavelength_option=hyperspectral
/BRDF_option=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,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,…
```

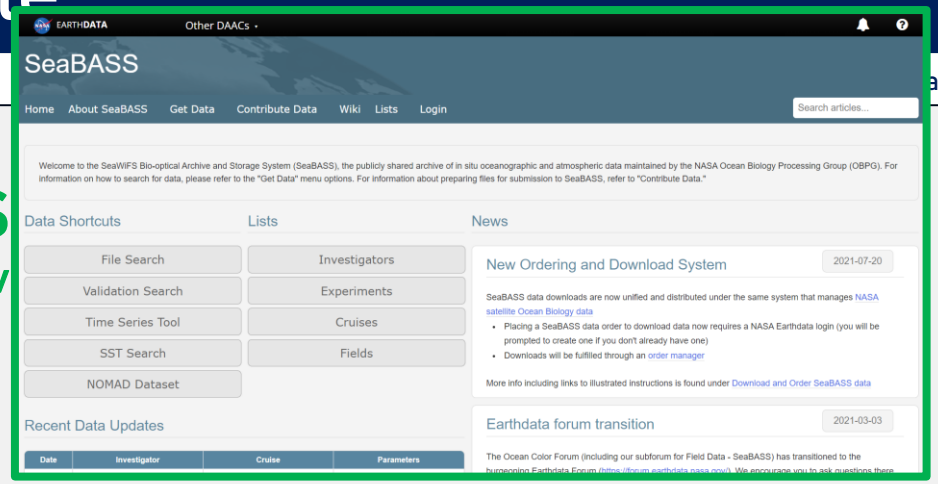
# 8. Required inputs: SeaBASS/OCDB file

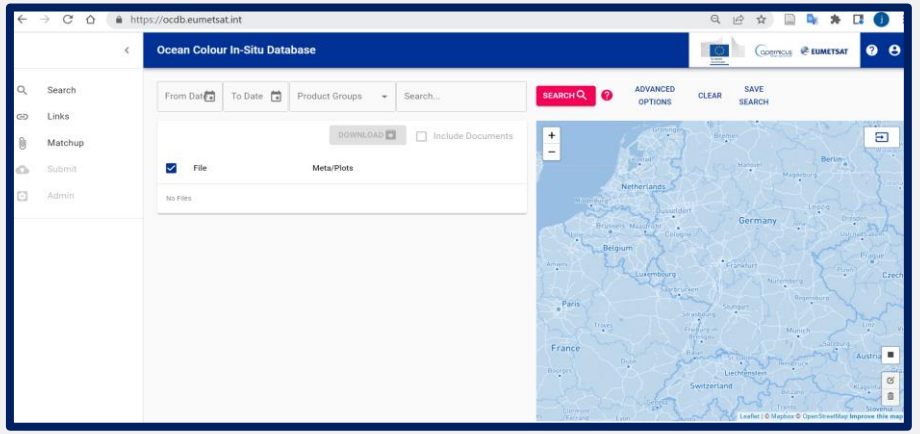
```
/begin_header
/investigators=TestPI
/affiliations=TestAffiliation
/contact=TestContact@TestInstitutution.org
/received=20190101
/processed=20190101
/experiment=TestExperiment
/cruise=TestCruise
/station=NA
/data_file_name=TestOCDBsubmission.txt
/documents=see_comments
/calibration_files=see_comments
/data_type=cast
/data_status=final
/water_depth=NA
/wavelength_option=hyperspectral
/BRDF_option=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,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



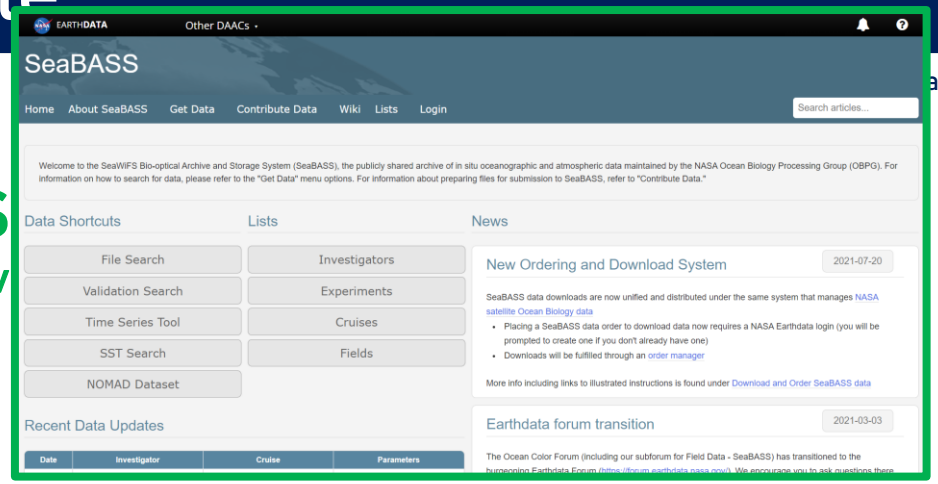
# 8. Required inputs: SeaBASS/OCDB file

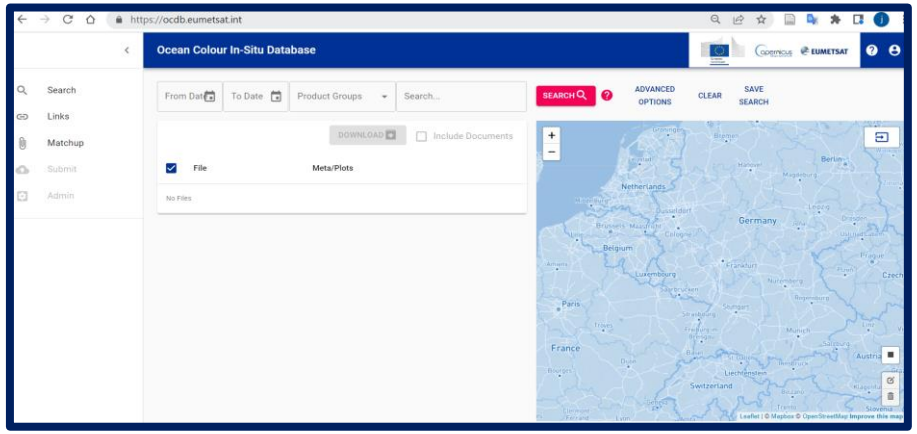
1. OCDB & SeaBASS offer documentation on how to get your in situ data in the correct format.
2. When submitting your data to OCDB, OCDB will guide you on the format.
3. ThoMaS repository contains example in situ files in this format and links to all the necessary resources

**SeaBASS**
seabass.gsfc.nasa.gov

Inter-operable

**OCDB**
ocdb.eumetsat.int




```
/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
/wavelength_option=hyperspectral
/BRDF_option=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,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,…
```

# 8. Required inputs: config_file.ini and satellite_datasets.csv



## Example 1:

1. I want to download Sen-3 data from the outer Gironde estuary at location (45.6N, 1.6W), where I deployed an in situ instrument measuring continuously during the first 5 days of September 2021.

2. I want data from both S3A and S3B, and L2 of the recent collection OL_L2M.003

3. Only full resolution (FR).

4. I 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, I want to test EUMETSAT's standard protocol, but I want to test several window sizes: 3x3, 5x5, 7x7.

6. I want everything related to the run to be stored at ~/Gironde

# 8. Required inputs: config_file.ini and satellite_datasets.csv



## Example 1:

1. I want to download Sen-3 data from the outer Gironde estuary at location (45.6N, 1.6W), where I deployed an in situ instrument measuring continuously during the first 5 days of September 2021.

2. I want data from both S3A and S3B, and L2 of the recent collection OL_L2M.003

3. Only full resolution (FR).

4. I 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, I want to test EUMETSAT's standard protocol, but I want to test several window sizes: 3x3, 5x5, 7x7.

6. I want everything related to the run to be stored at ~/Gironde

My **config_file.ini** should look like:

```
[global]
path_output: /tcenas/home/gossn/Gironde
SetID: Gironde

[workflow]
workflow: SatData, minifiles, EDB

[minifiles]
minifiles_winSize: 7

[EDB]
EDB_protocols_L2: EUMETSAT_standard_L2
EDB_winSizes: 3, 5, 7
```

My **satellite_datasets.csv** should look like:

| SetID | Region | Country | Lat | Lon | time_start | time_stop | Platform | Sensor | Level | Processor | Collection | Resolution | path_to_SatData |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Gironde | Gironde | France | 45.6 | -1.6 | 2021-09-01T00:00:00 | 2021-09-05T23:59:59 | S3A | OLCI | L2 | IPF | OL__L2M.003 | FR | EUMETSATdataStore:~/Gironde/SatData |
| Gironde | Gironde | France | 45.6 | -1.6 | 2021-09-01T00:00:00 | 2021-09-05T23:59:59 | S3B | OLCI | L2 | IPF | OL__L2M.003 | FR | EUMETSATdataStore:~/Gironde/SatData |

# 8. Required inputs: config_file.ini



## Example 2:

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 3x3,
- •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 /home/myName/MOBY/SatData
•all the other outputs (IDB, minifiles, EDB, MDB, etc.) to be stored at /home/myName/MOBY

# 8. Required inputs: config_file.ini



## Example 2:

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 3x3,
   - •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 /home/myName/MOBY/SatData
•all the other outputs (IDB, minifiles, EDB, MDB, etc.) to be stored at /home/myName/MOBY

In this case, **satellite_datasets.csv** will be generated automatically by ThoMaS and stored under path_output

```
[global]
path_output: /tcenas/home/gossn/MOBY
SetID: MOBY

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

[insitu]
insitu_input: /tcenas/home/gossn/MOBY/MOBY_OCDB.csv
insitu_satelliteTimeToleranceSeconds: 3600
insitu_getAncillary: False
insitu_BRDF: M02

[satellite]
satellite_path-to-SatData: /tcenas/home/gossn/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
```

# 9. Run the code

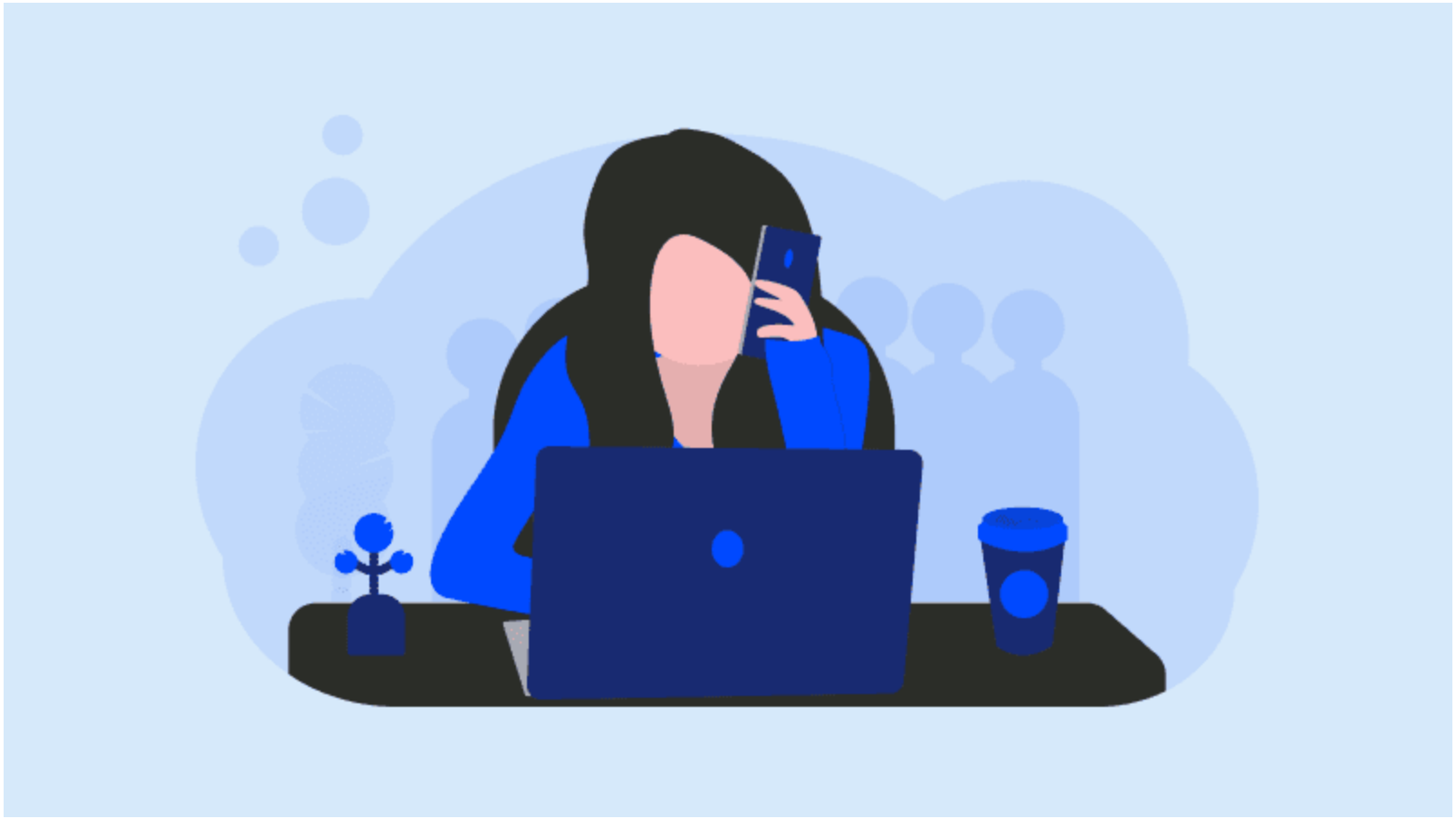


1. EUMETSAT Data Store credentials obtained and stored?
2. ECMWF ADS/CDS credentials obtained and stored?
3. ThoMaS code cloned?
4. thomas conda environment set up?
5. Required inputs in place? (config_file.ini, satellite_datasets.csv, SeaBASS/OCDB input file?

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

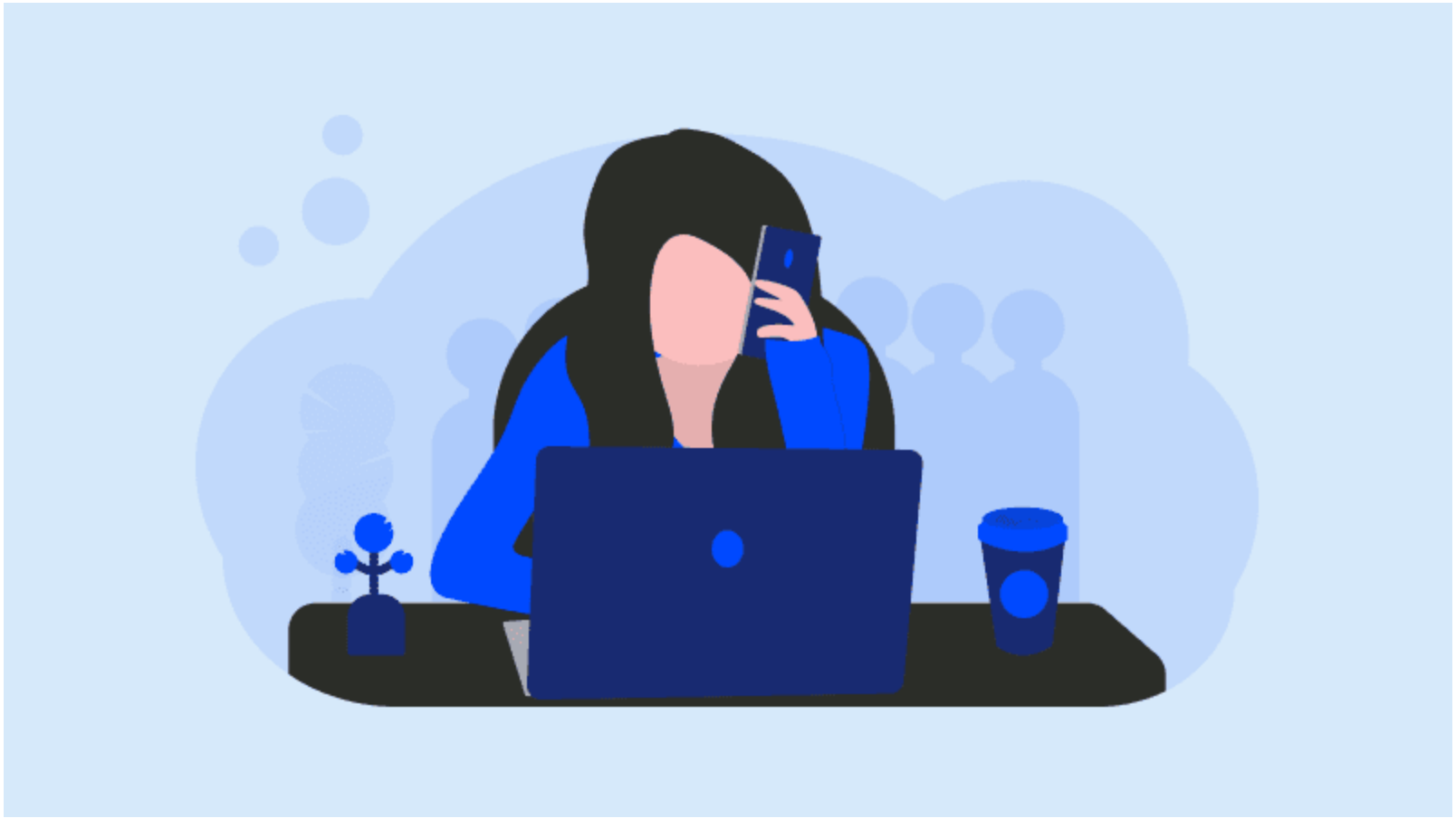
# 10. Run the code: Demo





# 11. Short tour around ThoMaS





# Today's tour with ThoMaS



**Hope you enjoyed it! Thank you!**
**Questions are welcome.**

1. What's ThoMaS? Scope
2. Usage
3. Some background
4. Pre-requisites
5. Getting the code
6. Setting the environment
7. Set-up demo
8. Required inputs
9. Run the code
10. Run the code: demo
11. Short tour around ThoMaS