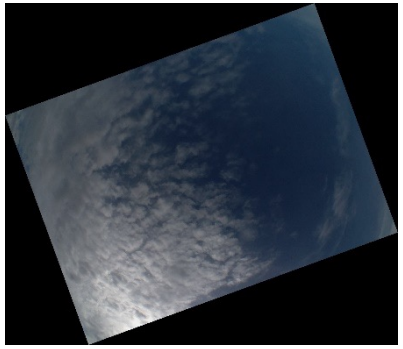
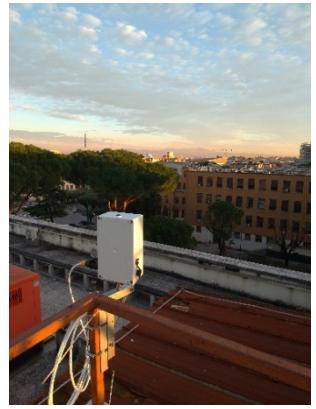
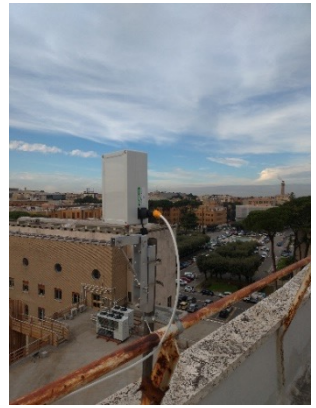


IDEAS-QA4EO

IDEAS-QA4EO Service
Cal/Val Workshop #4
1st March 2023, Potsdam





Ground-based sky-camera observations for cloud-mask validation in support to CMIX

Add-on: Status report CMIX-II

Jan Wevers
(Brockmann Consult GmbH)

Brockmann Consult
Environmental Informatics • Geoinformation Services

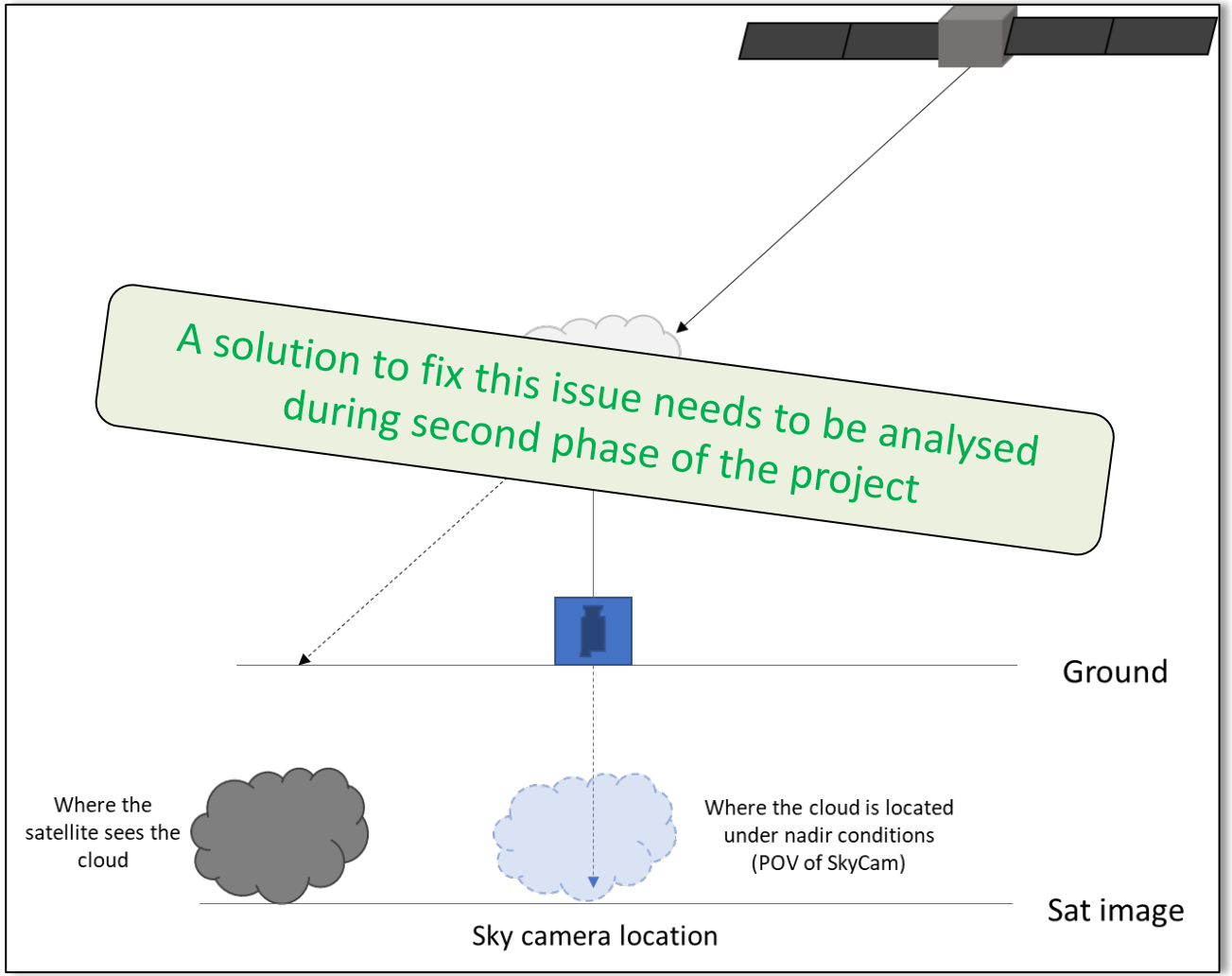
Project Objective – Phase 2

- Develop a global source for independent cloud mask validation of any remote sensing sensor operating in the visual domain (with focus on CMIX).
- Technical tasks of phase 2 – part 1:
  - finding a solution for geometry issues,
  - research on proper rectifying algorithms to correct for lens distortion,
  - more research on SC image classification to train a better classifier,
  - find a better solution to treat sun interferences in the SC images,
  - and to test or even develop an algorithm for stereo SC based cloud height estimation, and
  - finally to compare SC based cloud bottom heights with RAP cloud bottom heights.
- Technical tasks of phase 2 – part 2:
  - transferring the current mixture of cloud- and local desktop-based system to a completely cloud-based system.
  - This task will prepare for a sufficient way of processing SC and satellite data and the final data provision to the users.
  - It also should provide a good solution for the users to bring their own data for validation.

Project Objective – Phase 2

- Management tasks phase 2
  - fostering the good engagement with University of Maryland, to keep up with their developments and to benefit from knowledge exchange.
  - Outreach and communication and this phase should focus on sharing the status of the developments at relevant events and conferences.

Technical tasks part 1



Geometry issues

- The images show that the cloud in the center of SC2 (Fermi) is located northeast of SC2 in the S2 L2A image.
- While the same cloud is located southwest of the center of SC1 (Marconi) and south/over SC1 in the S2 L2A product.
- The cause for this mismatch can be explained by the viewing differences of the three instruments and the location of the cloud above ground.
- The S2 L2A data have been acquired off-nadir with a VAA mean of 130.28053 and a VZA mean of 3.3807745 (purple arrow viewing direction of S2 MSI).
- The parallax between true nadir and the actual S2 location cause the cloud to be projected in north-western direction onto the ground

Technical tasks part 1

SC 1 autom. classif. vs. OLCI L2 LFR Cloud & Ambiguous
Annual results 2022
Sky Camera 1

OLCI L2 LFR \ Class	Clear	Cloud	Sum	U A	E
CLEAR	250	22	272	91.9	8.1
CLOUD	36	87	123	70.7	29.3
Sum	286	109	395		
P A	87.4	79.8		OA:	85.32
E	12.6	20.2		BOA:	83.6

Scotts Pi: 0.646
Krippendorfs alpha: 0.646
Cohens kappa: 0.646

Geometry issues

SC 1 autom. classif. vs. OLCI L2 LFR Cloud & Ambiguous
Annual results 2022 - OZA < 30
Sky Camera 1

OLCI L2 LFR \ Class	Clear	Cloud	Sum	U A	E
CLEAR	146	13	159	91.8	8.2
CLOUD	17	51	68	75.0	25.0
Sum	163	64	227		
P A	89.6	79.7		OA:	86.78
E	10.4	20.3		BOA:	84.65

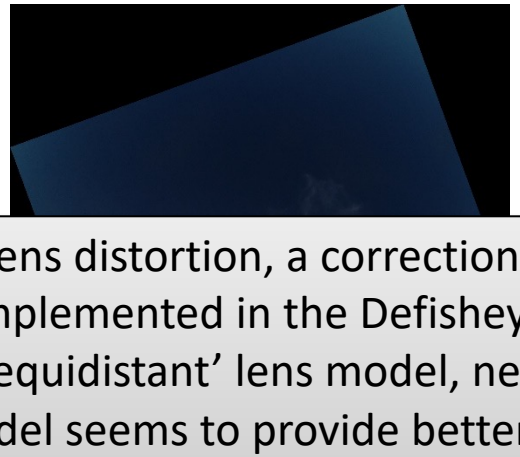
Scotts Pi: 0.679
Krippendorfs alpha: 0.68
Cohens kappa: 0.679

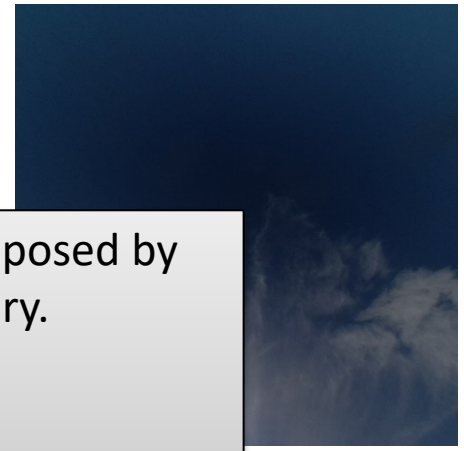
Technical tasks part 1

Old pre-processing

New pre-processing
Reduces geometric distortion

rome-skycam1_2023011

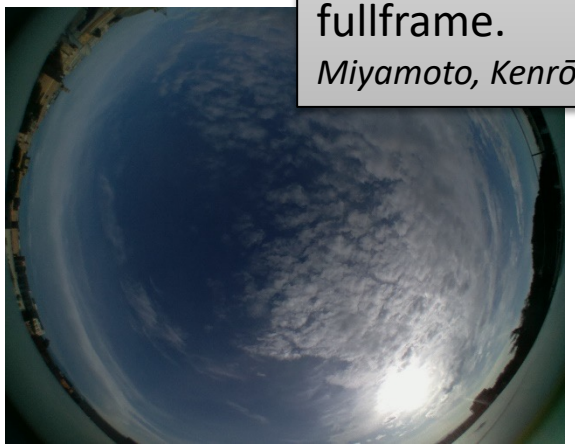




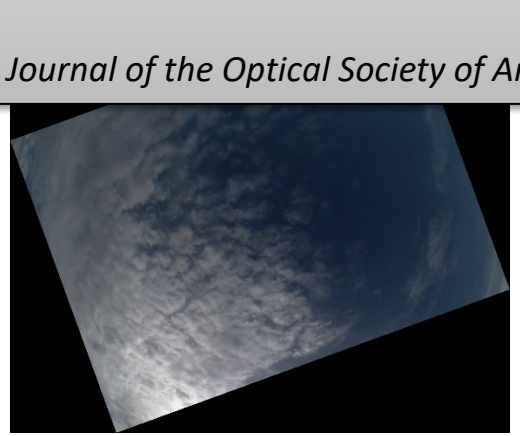
rotate, subset

To correct for the fisheye lens distortion, a correction algorithm proposed by Miyamoto 1964 is used, implemented in the Defisheye Python library.
UoM proposed to use an 'equidistant' lens model, nevertheless, 'equalarea'/'equisolid' model seems to provide better results.
The format of the camera used for correction is circular, compared to fullframe.

Miyamoto, Kenrō. "Fish Eye Lens." Journal of the Optical Society of America 54 (1964): 1060-1061.



rome-skycam1_20230126T090902_0p05.jpeg



Subset, flip, rotate

De-fisheye, flip, rotate, subset

Technical tasks part 1

Sky Camera Sites

- New site location in Valencia, Spain has been added
  - Data is transferred to Brockmann Consult server
  - Development of new classifier on-going (RF, NN, gradient boosting)
- Overall, currently six sites exist
  - GSFC, Greenbelt, Maryland, USA
  - Wisconsin, USA
  - Sapienza University, Rome, Italy
  - Valencia University, Valencia, Spain
  - Sao Paulo University, Sao Paulo, Brazil
  - Antarctica

Technical tasks part 1

Sun interference & Classifier

- Sun interference leads to false classification (clear as cloud)
- Usi n mask cou lly

Development on-going

Technical tasks part 1

Sun interference & Classifier

- Additional training data for the Rome site was collected to make classifier mor robust to seasonal variations -> training on-going
- New training data was collected for the Valencia site -> training on-going
- Three different classifiers will be tested against each other:
  - Random Forest (current classifier)
  - Neural Network
  - Gradient boosting

Technical tasks part 1

Cloud base height estimation – COD estimates

- Cloud base height: Still under development at University of Maryland (UoM)
- UoM is currently developing a COD algorithm for the SC images.
  - Development is done for CMIX-II
  - Valuable information for further validation of satellite-based cloud mask algorithms
  - Shall be finalized within Q1 2023

Technical tasks part 2

Cloud based solution for processing

- Two potential solution have been analysed
  - QA4EO Cal Val platform
  - EuroDataCube Service
- QA4EO CalVal platform
  - Data availability:
    - S2 rolling archive 1PB – older data get lost
    - Landsat: Only L8 and prior, no L9
    - Sentinel-3: Full archive
  - Usability:
    - All routines have to be developed starting with the extraction of data; no predefined modules or API
    - Access through X2Go quite challenging; Jupyter notebook might not be the best solution for operational validation tasks
  - Bring your own data (BYOD):
    - Limited to upload to the VM
    - For the access of other cloud stored data routines have to be developed -> By whom? The user?

Technical tasks part 2

Cloud based solution for processing

- Two potential solution have been analysed
  - QA4EO Cal Val platform
  - EuroDataCube Service
- EuroDataCube Se
  - Data availabilit
    - S2: full arch
    - Landsat: full archive (
    - Sentinel-3: Only L1b
  - Usability:
    - Xcube service: open source Python toolkit for generating, processing, and visualising data cubes. It provides a growing ecosystem of compatible modules supporting users in all steps of data cube generation and handling, while at the same time ensuring maximum compatibility and seamless integration into existing data science workflows by relying on Python's well-established and open source technology stack including packages like xarray, dask, geopandas, and zarr.
  - Bring your own data (BYOD):
    - Access to your own data stored in S3 buckets. Does not require any further data transfer. Methods for S3 data handling already available.

EuroDataCube Service seems to be the better choice
– Less work for integration –
– More user friendly –

Technical tasks part 2

Cloud based solution for processing

- Proposal for a new EDC Service is currently under development
- One of the biggest advantages of using the EDC is that users can get funding from ESA NoR for using any EDC service.

Exchange with UoM and outreach

- Through the joint lead of CMIX-II the Brockmann Consult team is in constant contact and good knowledge exchange with UoM

- The sky camera approach has been presented at international events and working groups
  - Living Planet Symposium, S3QWG, S3VT, OPT-MPC OLCI/SYN ESL & classification working group

Usage of the SC approach outside of QA4EO

Sky Camera validation over Rome
Using 2021 LFR data

SC 1 automatic classification vs. OLCI L2 LFR Cloud & Ambiguous & Margin

Sky Camera 1

OLCI L2 LFR \ Class	Clear	Cloud	Sum	U A	E
CLEAR	136	7	143	95.1	4.9
CLOUD	53	86	139	61.9	38.1
Sum	189	93	282		
P A	72.0	92.5		OA:	78.72
E	28.0	7.5		BOA:	82.25

Scotts Pi: 0.56
Krippendorfs alpha: 0.561
Cohens kappa: 0.572

PixBox validation 2021
using 2018 data over land surfaces

OLCI A+B FR IdePix cloud val. - land surfaces

In-Situ Database

OLCI FR IdePix \ Class	Clear	Cloud	Sum	U A	E
CLEAR	3442	443	3885	88.6	11.4
CLOUD	1039	3183	4222	75.4	24.6
Sum	4481	3626	8107		
P A	76.8	87.8		OA:	81.72
E	23.2	12.2		BOA:	82.3

Scotts Pi: 0.634
Krippendorfs alpha: 0.634
Cohens kappa: 0.635

SC 1 autom. classif. vs. OLCI L2 LFR Cloud & Ambiguous
Annual results 2022

Sky Camera 1

OLCI L2 LFR \ Class	Clear	Cloud	Sum	U A	E
CLEAR	250	22	272	91.9	8.1
CLOUD	36	87	123	70.7	29.3
Sum	286	109	395		
P A	87.4	79.8		OA:	85.32
E	12.6	20.2		BOA:	83.6

Scotts Pi: 0.646
Krippendorfs alpha: 0.646
Cohens kappa: 0.646

IDEAS-QA4EO

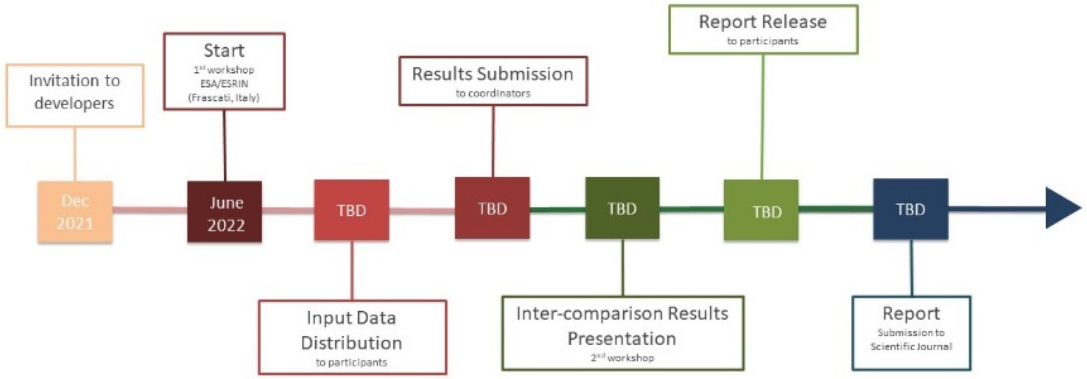
Status report CMIX-II

Cloud Mask Inter-comparison eXercise - II

Jan Wevers
(Brockmann Consult GmbH)

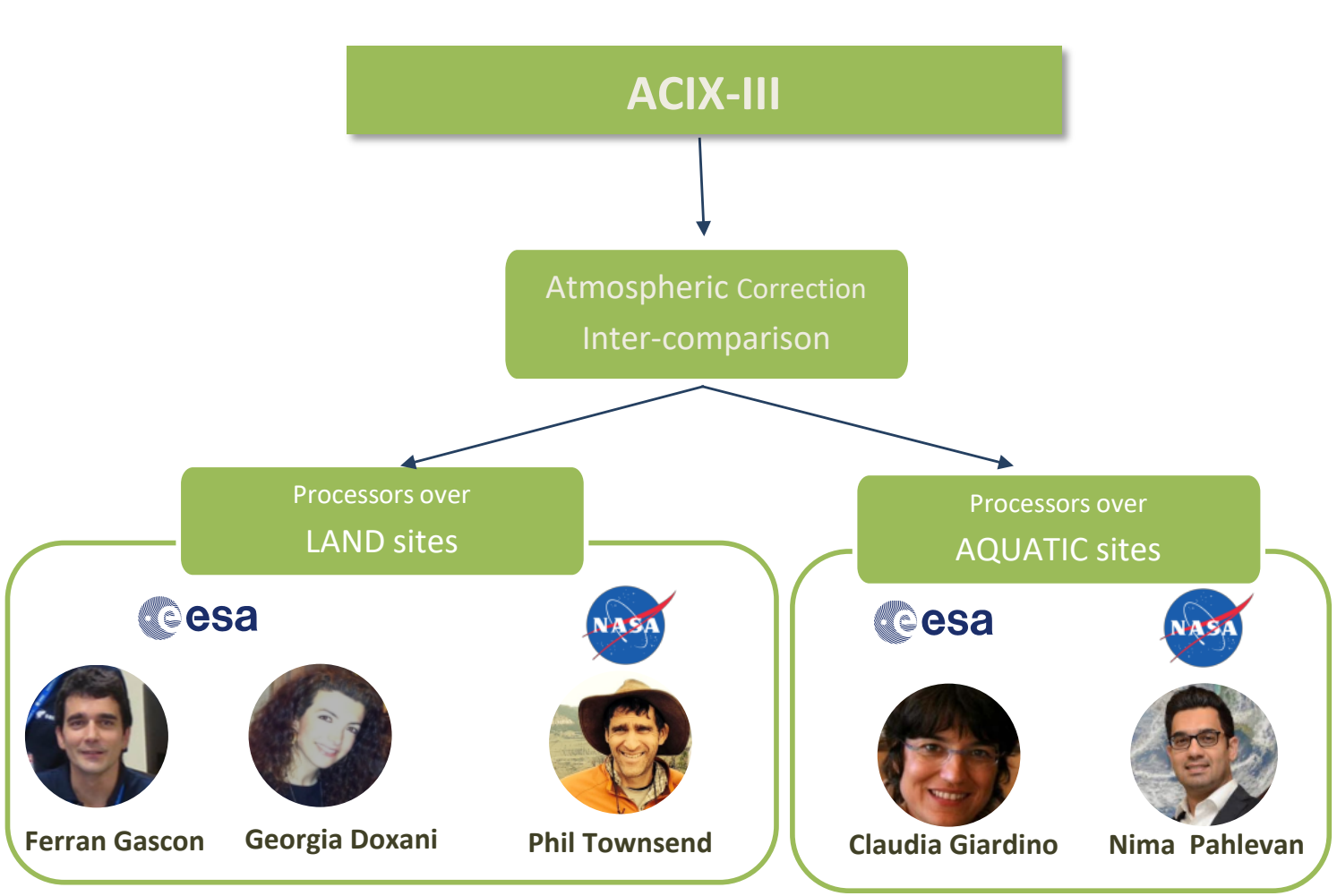
Sergii Skakun
(University of Maryland)

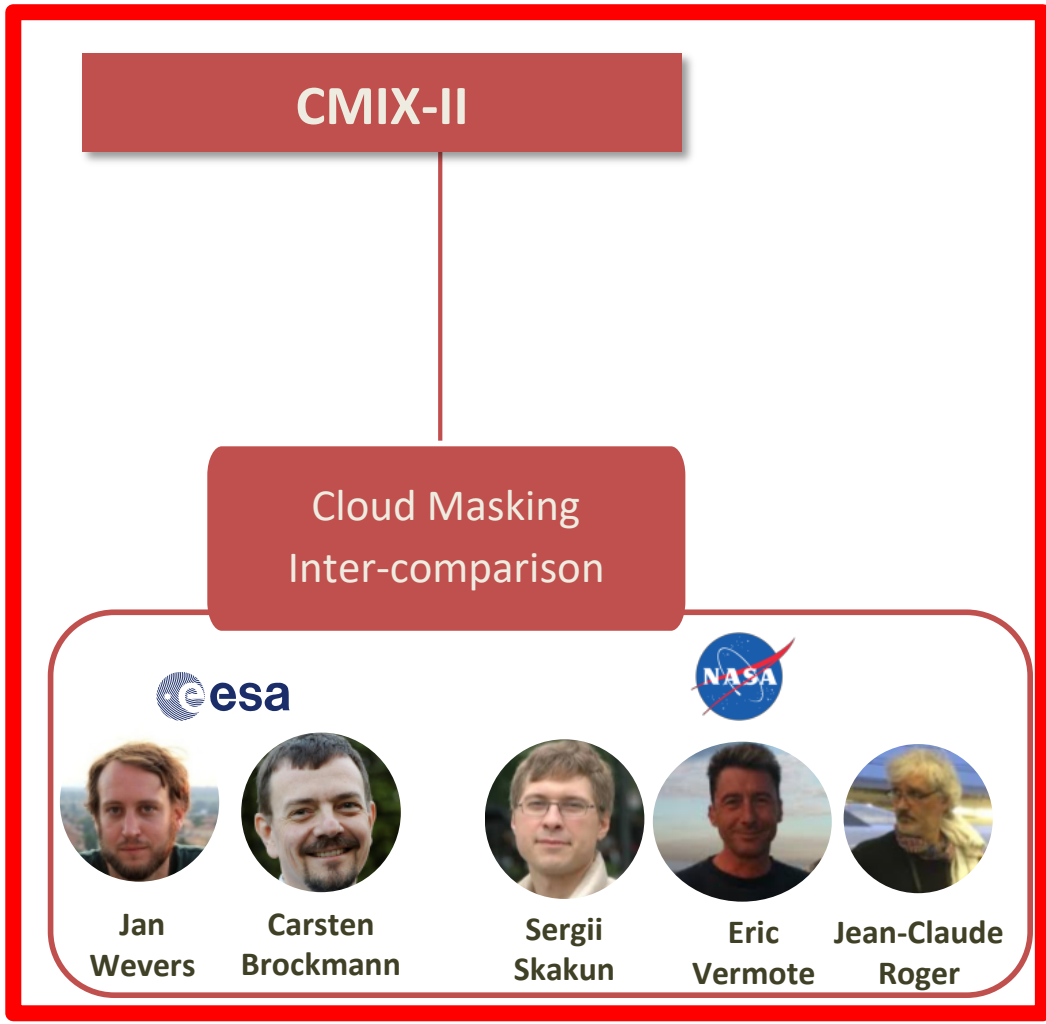
Georgia Doxani
(ESA)



Brockmann Consult
Environmental Informatics • Geoinformation Services

CMIX-II organization



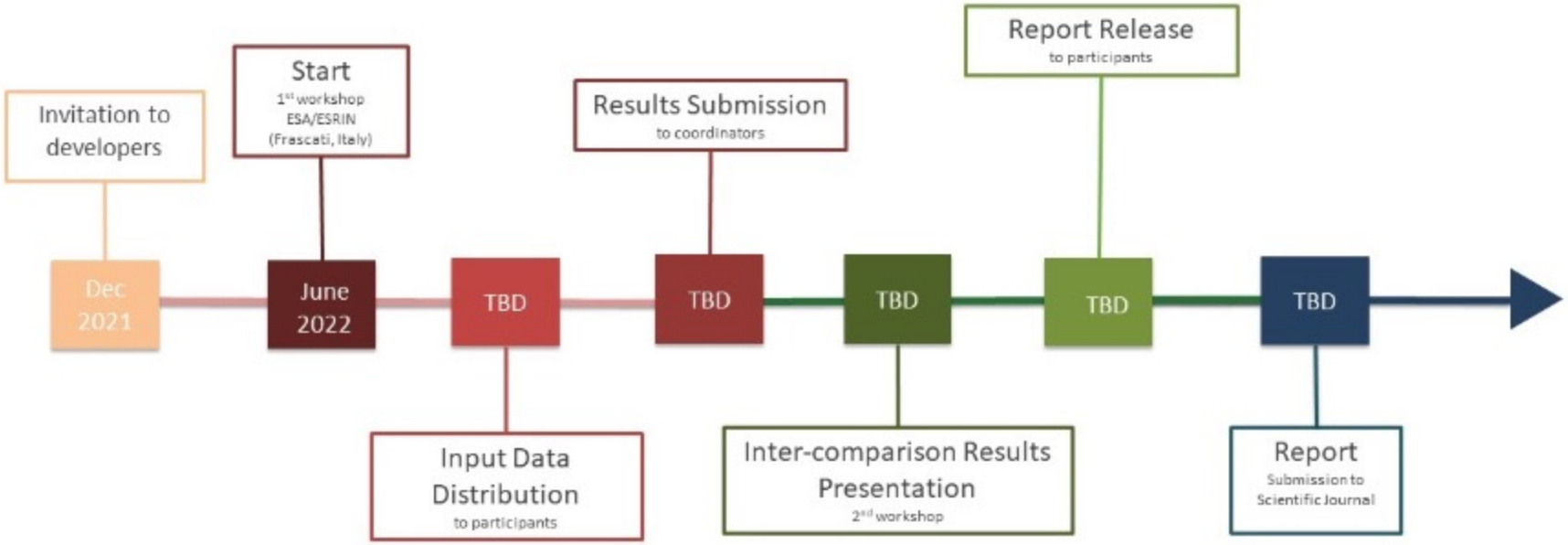


CMIX II - Objective

- To **inter-compare** a set of cloud detection algorithms for space-borne high-spatial resolution (10-30 m) multi- and hyperspectral optical sensors

- Focus on **Sentinel-2**, **Landsat 8/9** ~~and **PRISMA**~~ data

CMIX-II initial schedule



Results from first workshop

- Clear definition of clouds in reference datasets required; more physical based definition
  - For the two main datasets COD estimates will be provided with cloud classification for better understanding of singe classes (e.g. thin clouds, medium thick clouds, totally opaque clouds)
- Compared to the initial schedule a first round with a test dataset (TDS) of all refenrence datasets is planned, to ensure everbody agrees with the datasets.
- Accuracy comparison of cloud shadows is as important as clouds

Major phases

- Definition of the inter-comparison protocol and reference datasets
  - June 2022: This Workshop
- Phase 1 – TDS exercise
  - Provision of reference datasets samples
  - Application of the CM processors
  - Initial analysis of the results
  - Feedback from participants
- Phase 2 – Main exercise
  - Preparation of the reference datasets
    - Preparation of complete datasets: TBD
    - Publication of datasets after analysis: TBD
  - Application of the CM processors

Major phases

- Results Analysis Report (Internally)
  - May 2023 -> November 2023
- 2nd Workshop of CEOS-WGCV ACIX III – CMIX II
  - Approx. June 2023 @ ESA or NASA -> December 2023

Participants

- Number of participants has neraly doubled (CMIX: 9 participants) and has now finally reached 17.
- All of the listed participants have already processed the first results of the TDS or at least gave feedback.

No	Name	Organization	Algorithm
1	Sergii Skakun Eric Vermote Jean-Claude Roger	NASA GSFC/UMD NASA GSFC NASA GSFC	LaSRC
2	Alistair Francis	ESA/ESRIN	SEnSeI
3	Feng Yin	University College London	CNN_cloud_shadow_mask
4	David Frantz	Trier University	FORCE
5	Béatrice Berthelot	MAGELLIUM	MAGCMA
6	Jorrit Scholze	Brockmann Consult GmbH	IdePix
7	Luis Gómez-Chova	University of Valencia	UVDeepCloud
8	Kaupo Voormansik Tetiana Shtym	KappaZeta Ltd	KappaMask
9	Yaokai Liu	CAS.CHINA	Hikerliu
10	Jerome Louis	Telespazio France	Sen2Cor
11	Christopher Brown Valerie Pasquarella	Google, LLC	CloudScore+
12	Hervé Poilvé	Airbus Geo-Intelligence Toulouse (Airbus DS)	Overland
13	Pat Scaramuzza	USGS	Cfmask
14	Bringfried Pflug, Avi Avi Pertiwi Raquel de los Reyes	DLR	PACO
15	Christophe Lerebourg	ACRI	BrightEarth
16	Hankui Zhang	South Dakota State Univ	LANA
17	Zhe Zhu Shi Qiu	University of Connecticut	Fmask 4.6

Datasets

- Reference / validation datasets
  - Sky Camera Network
  - PixBox (expert pixel collection) dataset
    - General collection
    - *Potentially: Cloud border collection*
  - Multi-temporal (time series) critical case collection (identifying potential systematic errors)
  - Collaborative dataset using IRIS (active learning): classification of subsets by the participants

THANK YOU FOR THE ATTENTION