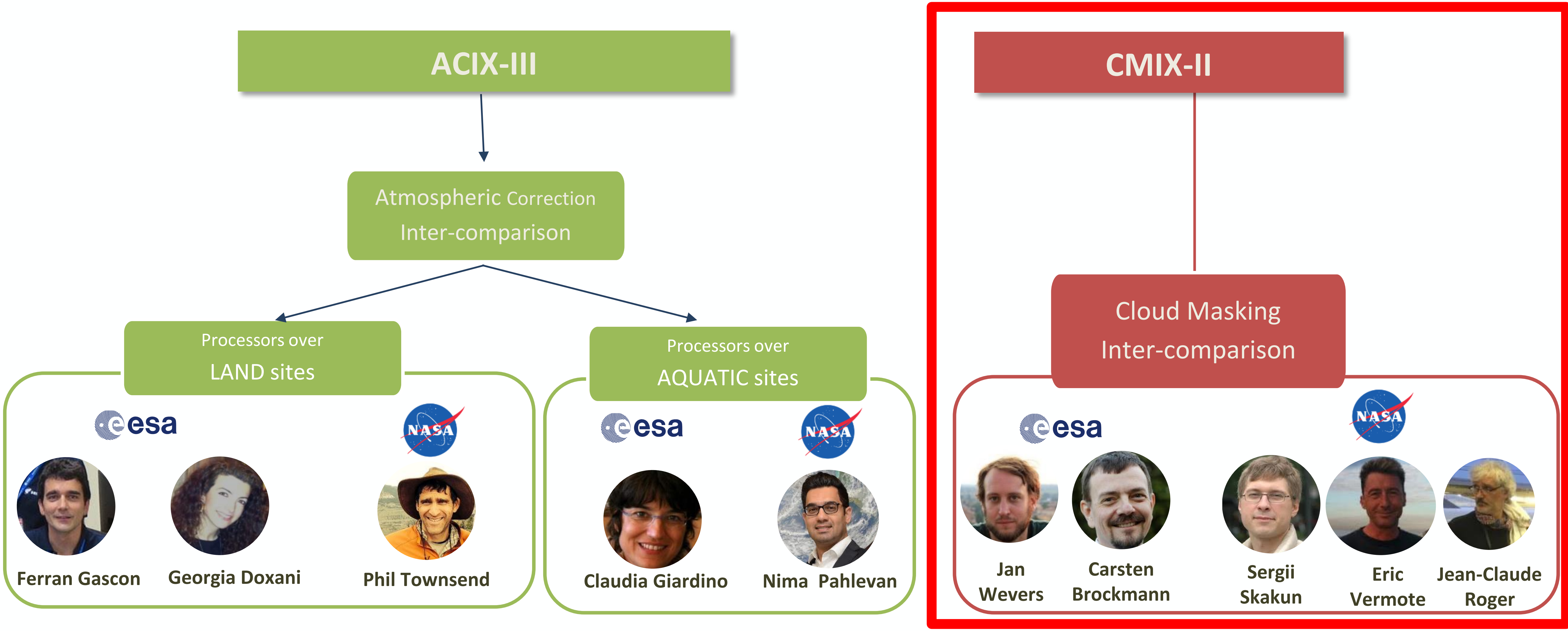
PROGRAMME OF THE EUROPEAN UNION

co-funded with

CMIX-II: Second edition of the Cloud Mask Intercomparison eXercise

Jan Wevers, Carsten Brockmann – Brockmann Consult GmbH Germany
Sergii Skakun, Eric Vermote, University of Maryland / NASA GSFC

Organization



CMIX - Recap and lessons learned

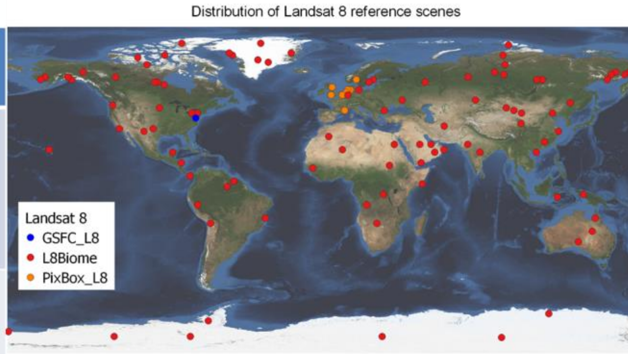
First CMIX in Numbers

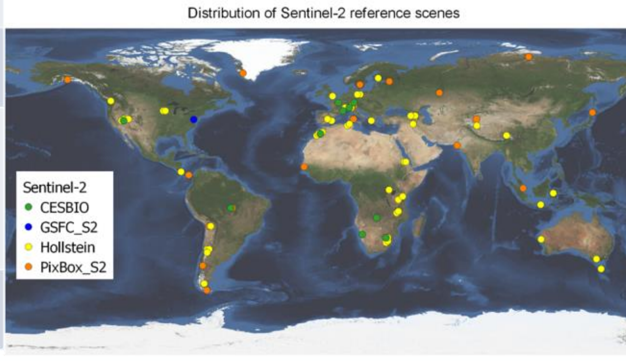
In 2019 European Space Agency (ESA) and National Aeronautics and Space Administration (NASA) initiated the first Cloud Masking Inter-comparison Exercise (CMIX) in the frame of CEOS WGCV (Committee on Earth Observation Satellites, Working Group on Calibration and Validation)

9 participants - 10 algorithms - 5 Reference datasets - 146 Sentinel-2 products - 113 Landsat 8 products

Processor	Organization	Methodology
ATCOR	DLR	Spectral tests (L8, S2)
CD-FCNN	University of Valencia	Machine learning (L8, S2)
Fmask 4.0 CCA	USGS	Spectral tests (L8, S2)
FORCE	Humboldt-Universität zu Berlin / Trier University	Spectral test + parallax (for S2) (L8, S2)
IdePix	Brockmann Consult	Spectral tests (S2)
InterSSIM	Sinergise	Machine learning + spatiotemporal context (S2)
LaSRC	NASA / University of Maryland	Spectral tests (L8, S2)
MAJA	CNES / CESBIO	Multi-temporal and spectral tests (S2)
s2cloudless	Sinergise	Machine learning (S2)
sen2cor	ESA / Telespazio France	Spectral test + auxiliary data (S2)

Dataset	Spatial domain	Spatial resolution	# scenes
CESBIO	Fully classified Sentinel-2 scenes	60 m	S2: 30
GSFC	Sample polygons	Polygons (vector)	L8: 6 S2: 28
Hollstein	Sample polygons	Polygons (at 20 m)	S2: 59
L8Biome	Fully classified Landsat 8 scenes	30 m	L8: 96
PixBox	Sample pixels	S2: 10 m L8: 30 m	S2: 29 L8: 11





CMIX - Recap and lessons learned

Major findings

- Performance varied depending on the reference data
- Average OA for Sentinel-2: 80% to 89%
- Average OA for Landsat 8: 80% to 98%
- Performance improved when thin/semi-transparent clouds not considered
- Usage of a buffer improves performance of cloud detection at cloud edges

Remote Sensing of Environment 274 (2022) 112990

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

ELSEVIER

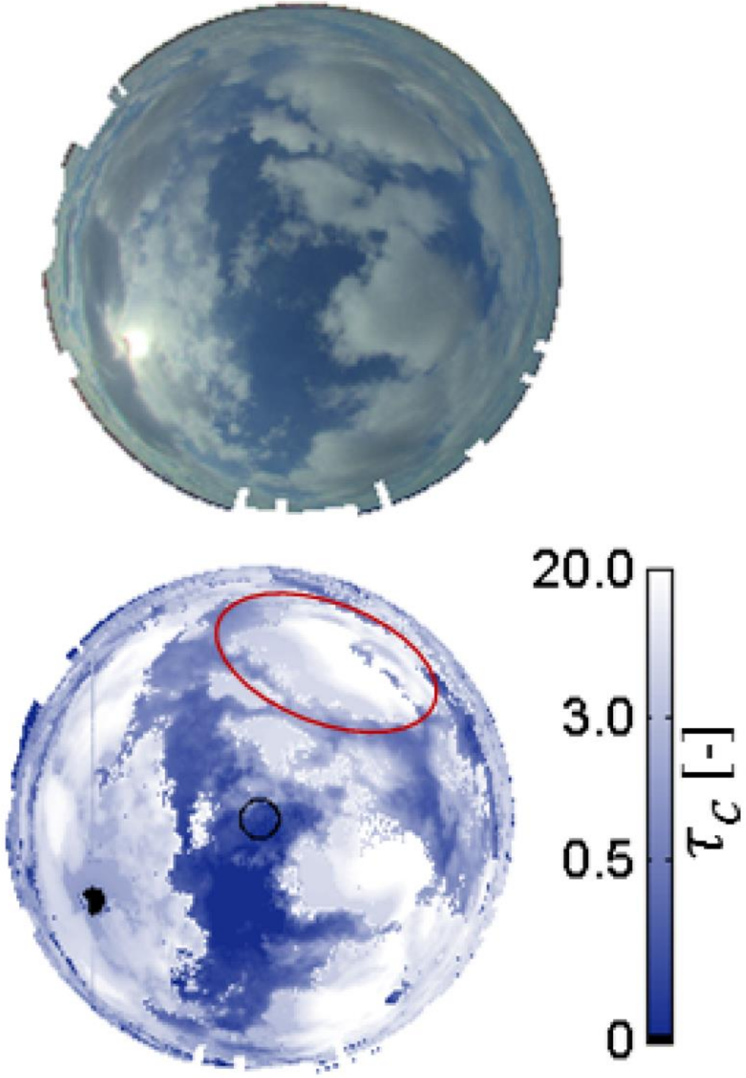
Cloud Mask Intercomparison eXercise (CMIX): An evaluation of cloud masking algorithms for Landsat 8 and Sentinel-2

https://doi.org/10.1016/j.rse.2022.112990

CMIX - Recap and lessons learned

Recommendations

- Definition of cloud should be based on a more physical approach
  - Cloud Optical Depth (COD)
- Validation datasets need to be produced specifically for the exercise
  - Consistent cloud definition
  - Cloud boundary
  - Time series
- The analysis framework should be designed to address different questions
  - Sample-based vs. area based
  - Temporal analysis
  - Application-based



Cloud optical depth retrieval from ground-based cloud imager (Mejia et al., 2016)

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

Participants responsibilities

- Contribute to the definition the protocol
- Provide a detailed **description of the CM processor**
- Run CM processor with **its standard parameters** according to the protocol described and on data provided by the CMIX organizers
- Submit the results according to the **recommended formats,** and upload the results to the secure CMIX FTP within the scheduled period
- **Contribute** with preparation of the papers, presentations and other official communications to be published in the scientific literature at the end of the exercise

Major phases

- **Definition of the inter-comparison protocol and reference datasets**
  - June 2022: First Workshop
- **Phase 1 – TDS exercise**
  - Provision of reference datasets samples (2 of 4 datasets)
  - Application of the CM processors (2 of 4 datasets)
  - Initial analysis of the results
  - Feedback from participants
- **Phase 2 – Main exercise**
  - Preparation of the reference datasets
  - Provision of input datasets (to participants)
  - Application of the CM processors
  - Analysis of the results
  - Publication of complete datasets (reference + input data) after analysis
- **Results Analysis Report** (Internally to participants)
  - Q1/Q2- 2024
- **2nd Workshop of CEOS-WGCV CMIX II**
  - Approx. mid 2024 @ ESA or NASA

Results from first workshop

- Provide definition of cloud ✓
  - Physically based, to be included in reference datasets
  - 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, include a first round with a test dataset (TDS) of all reference datasets, to ensure everbody agrees with the datasets. ✓
- Accuracy comparison of cloud shadows is as important as clouds ✓

CMIX-II Participants

- Number of participants has doubled (CMIX: 9 participants) and has now finally reached 18.
- All of the listed participants have already processed the first results of the TDS or at least gave feedback.

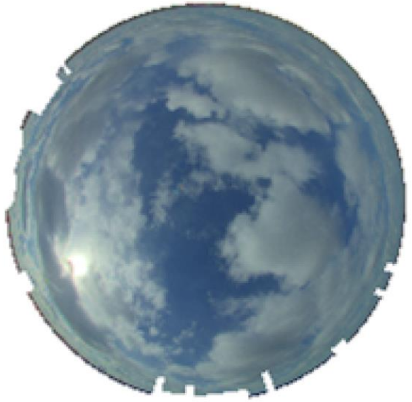
No	Algorithm	Organization	Name
1	LaSRC	NASA GSFC/UMD	Sergii Skakun
		NASA GSFC	Eric Vermote
		NASA GSFC	Jean-Claude Roger
2	SEnSeI	ESA/ESRIN	Alistair Francis
3	CNN cloud shadow mask	University College London	Feng Yin
4	FORCE	Trier University	David Frantz
5	MAGCMA	MAGELLIUM	Béatrice Berthelot
6	IdePix	Brockmann Consult GmbH	Jorrit Scholze
			Jan Wevers
7	UVDeepCloud	University of Valencia	Luis Gómez-Chova
8	KappaMask	KappaZeta Ltd	Kaupo Voormansik
			Tetiana Shtym
9	Hikerliu	CAS.CHINA	Yaokai Liu
10	Sen2Cor	Telespazio France	Jerome Louis
11	CloudScore+	Google, LLC	Christopher Brown
			Valerie Pasquarella
12	Overland	Airbus Geo-Intelligence Toulouse (Airbus DS)	Hervé Poilvé
13	Cfmask	USGS	Pat Scaramuzza
14	PACO	DLR	Bringfried Pflug, Avi
			Avi Pertiwi
			Raquel de los Reyes
15	BrightEarth	ACRI	Christophe Lerebourg
16	LANA	South Dakota State Univ	Hankui Zhang
17	Fmask 4.6	University of Connecticut	Zhe Zhu Shi Qiu
18	*Not provided*	EO Consultancy	Ute Gangkofner

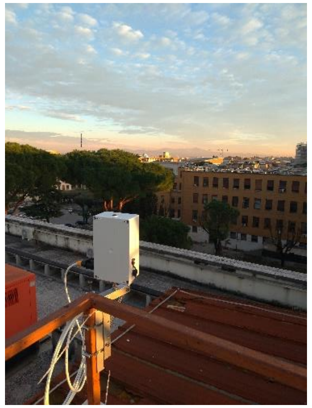
Datasets

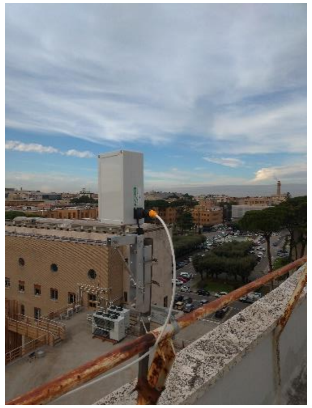
- Reference datasets
  - Sky Camera Network
  - PixBox (expert pixel collection) dataset
    - General collection
    - Cloud border collection
  - Multi-temporal (time series) critical case collection (identifying potential systematic errors)
  - Collaborative dataset using IRIS (active learning): classification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

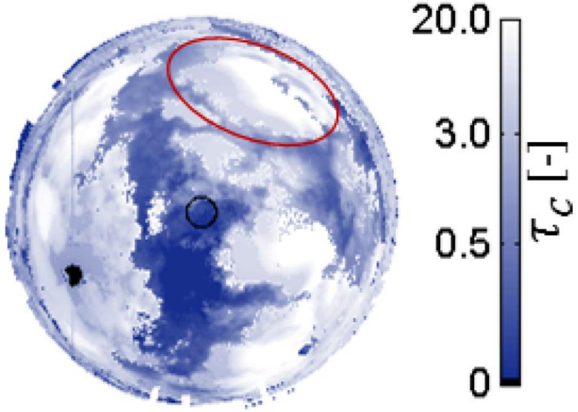
Sky Camera

- Affordable sky camera system based on Raspberry Pi 4 and Omnivision OV5647 sensor (HFOV: 194, VFOV: 142)
- Stereo setup for future cloud-base height estimations
- Providing cloud optical depth estimates
- Growing sky camera network with 6 sites:
  - GSFC, Greenbelt, Maryland, USA
  - Wisconsin, USA
  - Sapienza University, Rome, Italy
  - Valencia University, Valencia, Spain
  - Sao Paulo University, Sao Paulo, Brazil
  - Antarctica https://geog.umd.edu/feature/umd-skycam-project-deploys-sky-imagers-antarctica





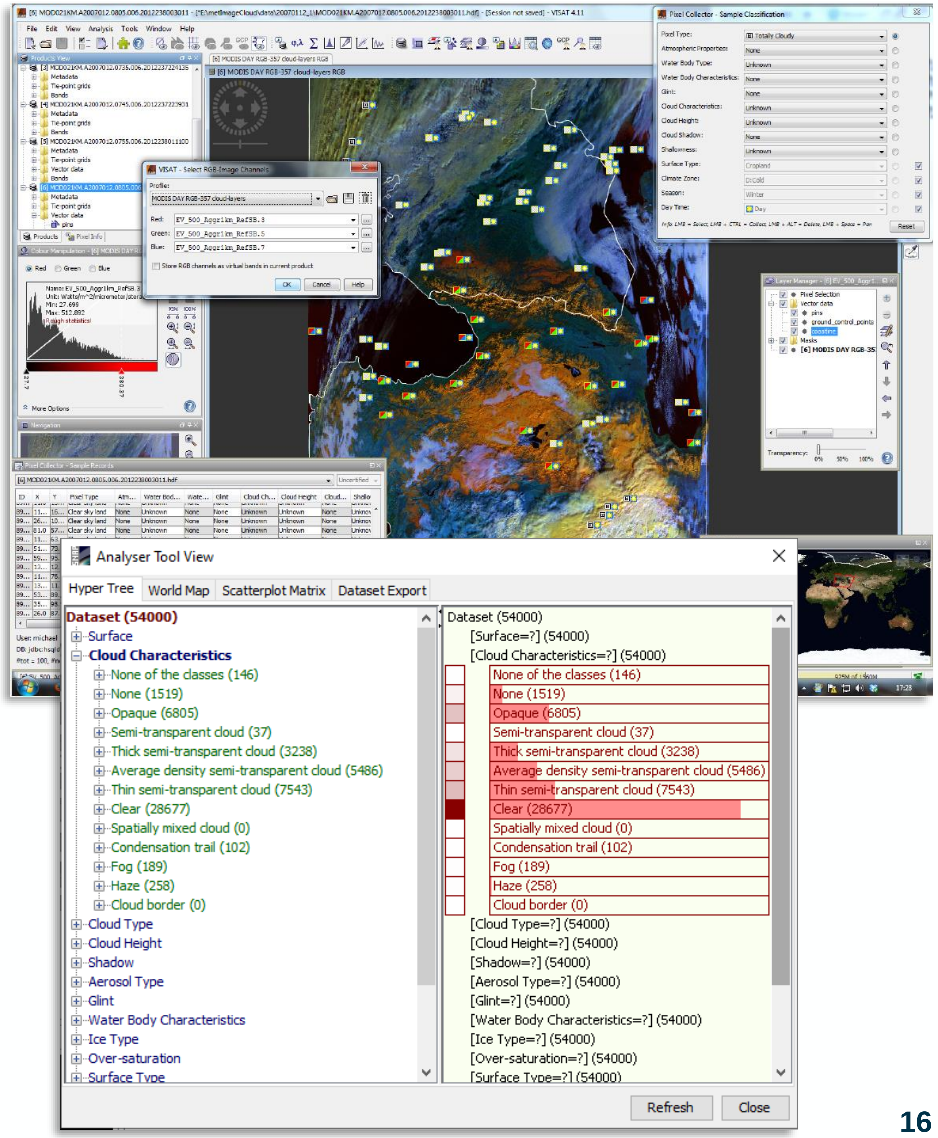




Cloud optical depth retrieval from ground-based cloud imager (Mejia et al., 2016)

PixBox – expert pixel collection

- Trained expert(s) manually labels pixels of an image sensor into a detailed set of pre-defined classes (E.g., different cloud transparencies, cloud shadow, and condition of the underlying surface)
- Collected dataset includes 10's of thousands of pixels because it necessitates representation for all classes, and for various observation and environmental conditions such as climate zones, solar illumination, viewing angles, etc.

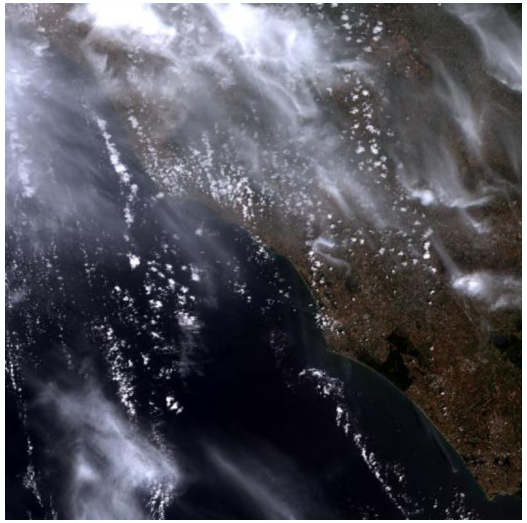


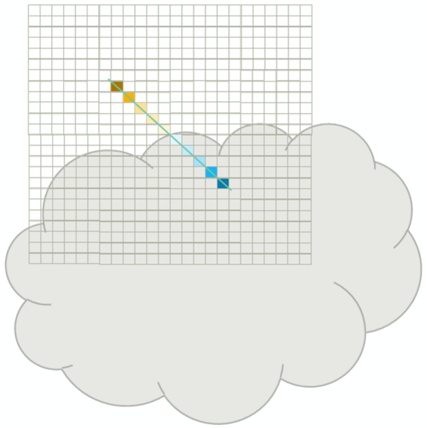
PixBox – expert pixel collection

- Major improvement compared to CMIX: COD estimate is provided for each collected pixel.

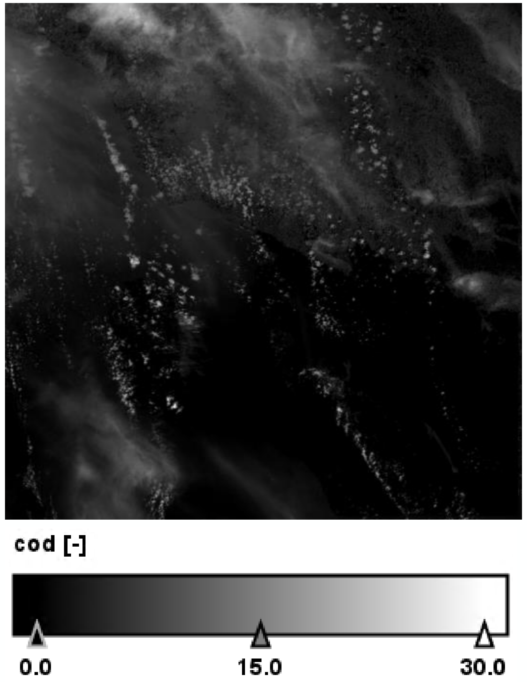
- Two pixel collections per sensor
  - General collection – comparable to CMIX but better distributed, including cloud shadow and COD
  - Dedicated cloud border collection
    - Might be skipped due to time constraints

S2 L1C





COD



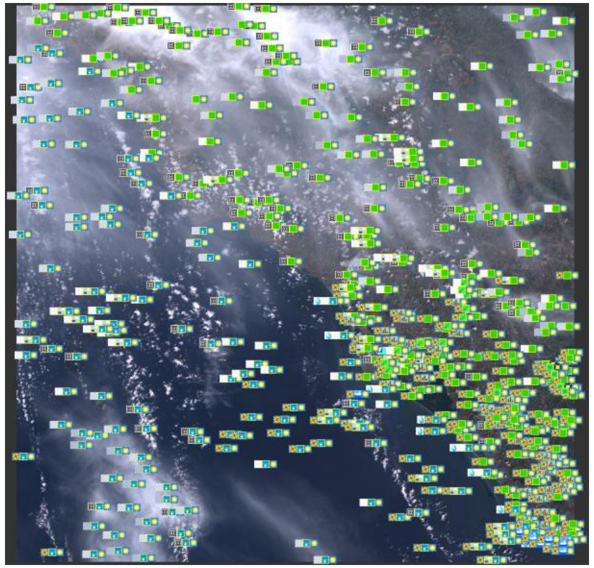
Metrics

- Per pixel: confusion matrix and estimated OA, PA, UA
- Visual inspection (potentially study an impact on SR, especially with transparent/cirrus clouds)
- Comparisons to sky imagery outputs:
  - Per-pixel comparison, when a precise co-registration between sky and satellite images will be feasible.
  - Cloud fraction: cloud fractions derived from sky and satellite imagery in a specified ROI will be estimated and will be compared. A separate analysis of completely clear and fully cloudy scenes will be performed. Analysis will be performed using a varying threshold for defining clouds (fuzzy logic analysis).
- Time series analysis: a temporal analysis of metrics calculated for each match (between ground- and satellite-based imagery) will be performed for each location to identify consistent errors produced by algorithms.

Results from test datasets

Validation results for 16/18 algorithms for the PixBox S2 TDS

Test product:
S2B_MSIL1C_20210704T100029_N0301_R122_T32TQM_20210704T121226



Algo 1
PixBox test validation - Sentine-2

PixBox

Class	Clear	Cloud	Sum	U A	E
CLEAR	194	11	205	94.6	5.4
CLOUD	33	262	295	88.8	11.2
Sum	227	273	500		
P A	85.5	96.0		OA:	91.2
E	14.5	4.0		BOA:	90.75

Scotts Pi: 0.82
Krippendorfs alpha: 0.82
Cohens kappa: 0.821

Algo 2
PixBox test validation - Sentine-2

PixBox

Class	Clear	Cloud	Sum	U A	E
CLEAR	181	8	189	95.8	4.2
CLOUD	46	265	311	85.2	14.8
Sum	227	273	500		
P A	79.7	97.1		OA:	89.2
E	20.3	2.9		BOA:	88.4

Scotts Pi: 0.777
Krippendorfs alpha: 0.777
Cohens kappa: 0.779

Algo 3
PixBox test validation - Sentine-2

PixBox

Class	Clear	Cloud	Sum	U A	E
CLEAR	198	24	222	89.2	10.8
CLOUD	29	249	278	89.6	10.4
Sum	227	273	500		
P A	87.2	91.2		OA:	89.4
E	12.8	8.8		BOA:	89.2

Scotts Pi: 0.785
Krippendorfs alpha: 0.785
Cohens kappa: 0.785

Algo 4
PixBox test validation - Sentine-2

PixBox

Class	Clear	Cloud	Sum	U A	E
CLEAR	215	29	244	88.1	11.9
CLOUD	12	244	256	95.3	4.7
Sum	227	273	500		
P A	94.7	89.4		OA:	91.8
E	5.3	10.6		BOA:	92.05

Scotts Pi: 0.835
Krippendorfs alpha: 0.835
Cohens kappa: 0.835

Algo 5
PixBox test validation - Sentine-2

PixBox

Class	Clear	Cloud	Sum	U A	E
CLEAR	168	7	175	96.0	4.0
CLOUD	59	266	325	81.8	18.2
Sum	227	273	500		
P A	74.0	97.4		OA:	86.8
E	26.0	2.6		BOA:	85.7

Scotts Pi: 0.725
Krippendorfs alpha: 0.725
Cohens kappa: 0.728

Algo 6
PixBox test validation - Sentine-2

PixBox

Class	Clear	Cloud	Sum	U A	E
CLEAR	201	3	204	98.5	1.5
CLOUD	26	270	296	91.2	8.8
Sum	227	273	500		
P A	88.5	98.9		OA:	94.2
E	11.5	1.1		BOA:	93.7

Scotts Pi: 0.881
Krippendorfs alpha: 0.881
Cohens kappa: 0.882

Algo 7
PixBox test validation - Sentine-2

PixBox

Class	Clear	Cloud	Sum	U A	E
CLEAR	159	4	163	97.5	2.5
CLOUD	68	269	337	79.8	20.2
Sum	227	273	500		
P A	70.0	98.5		OA:	85.6
E	30.0	1.5		BOA:	84.25

Scotts Pi: 0.697
Krippendorfs alpha: 0.697
Cohens kappa: 0.702

Algo 8
PixBox test validation - Sentine-2

PixBox

Class	Clear	Cloud	Sum	U A	E
CLEAR	174	3	177	98.3	1.7
CLOUD	53	270	323	83.6	16.4
Sum	227	273	500		
P A	76.7	98.9		OA:	88.8
E	23.3	1.1		BOA:	87.8

Scotts Pi: 0.767
Krippendorfs alpha: 0.767
Cohens kappa: 0.769

Algo 9
PixBox test validation - Sentine-2

PixBox

Class	Clear	Cloud	Sum	U A	E
CLEAR	196	8	204	96.1	3.9
CLOUD	31	265	296	89.5	10.5
Sum	227	273	500		
P A	86.3	97.1		OA:	92.2
E	13.7	2.9		BOA:	91.7

Scotts Pi: 0.84
Krippendorfs alpha: 0.841
Cohens kappa: 0.841

Algo 10
PixBox test validation - Sentine-2

PixBox

Class	Clear	Cloud	Sum	U A	E
CLEAR	221	15	236	93.6	6.4
CLOUD	6	258	264	97.7	2.3
Sum	227	273	500		
P A	97.4	94.5		OA:	95.8
E	2.6	5.5		BOA:	95.95

Scotts Pi: 0.915
Krippendorfs alpha: 0.915
Cohens kappa: 0.915

Algo 11
PixBox test validation - Sentine-2

PixBox

Class	Clear	Cloud	Sum	U A	E
CLEAR	189	14	203	93.1	6.9
CLOUD	38	259	297	87.2	12.8
Sum	227	273	500		
P A	83.3	94.9		OA:	89.6
E	16.7	5.1		BOA:	89.1

Scotts Pi: 0.787
Krippendorfs alpha: 0.788
Cohens kappa: 0.788

Algo 12
PixBox test validation - Sentine-2

PixBox

Class	Clear	Cloud	Sum	U A	E
CLEAR	213	25	238	89.5	10.5
CLOUD	14	248	262	94.7	5.3
Sum	227	273	500		
P A	93.8	90.8		OA:	92.2
E	6.2	9.2		BOA:	92.3

Scotts Pi: 0.843
Krippendorfs alpha: 0.843
Cohens kappa: 0.843

Algo 13
PixBox test validation - Sentine-2

PixBox

Class	Clear	Cloud	Sum	U A	E
CLEAR	196	8	204	96.1	3.9
CLOUD	31	265	296	89.5	10.5
Sum	227	273	500		
P A	86.3	97.1		OA:	92.2
E	13.7	2.9		BOA:	91.7

Scotts Pi: 0.84
Krippendorfs alpha: 0.841
Cohens kappa: 0.841

Algo 14
PixBox test validation - Sentine-2

PixBox

Class	Clear	Cloud	Sum	U A	E
CLEAR	128	21	149	85.9	14.1
CLOUD	99	252	351	71.8	28.2
Sum	227	273	500		
P A	56.4	92.3		OA:	76.0
E	43.6	7.7		BOA:	74.35

Scotts Pi: 0.488
Krippendorfs alpha: 0.489
Cohens kappa: 0.501

Algo 15
PixBox test validation - Sentine-2

PixBox

Class	Clear	Cloud	Sum	U A	E
CLEAR	196	110	306	64.1	35.9
CLOUD	31	163	194	84.0	16.0
Sum	227	273	500		
P A	86.3	59.7		OA:	71.8
E	13.7	40.3		BOA:	73.0

Scotts Pi: 0.433
Krippendorfs alpha: 0.434
Cohens kappa: 0.447

Algo 16
PixBox test validation - Sentine-2

PixBox

Class	Clear	Cloud	Sum	U A	E
CLEAR	217	26	243	89.3	10.7
CLOUD	10	247	257	96.1	3.9
Sum	227	273	500		
P A	95.6	90.5		OA:	92.8
E	4.4	9.5		BOA:	93.05

Scotts Pi: 0.855
Krippendorfs alpha: 0.855
Cohens kappa: 0.855

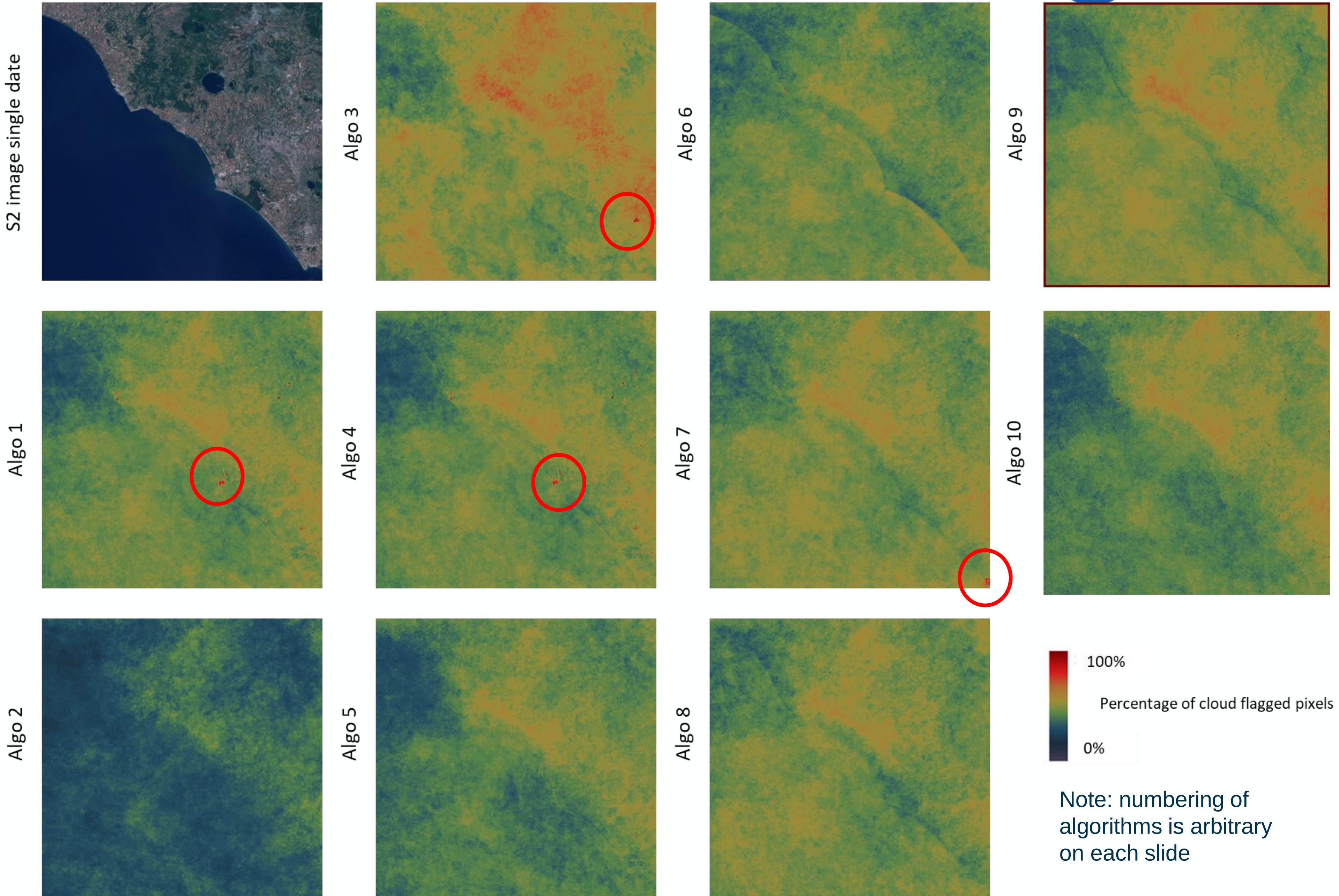
Results from test datasets

Results for multi-temporal TDS

10/18 algorithms provided results

Calculate percentage of cloud flagging per pixel on annual data stack

This method helps to identify semi-systematic and systematic false detections



Note: numbering of algorithms is arbitrary on each slide

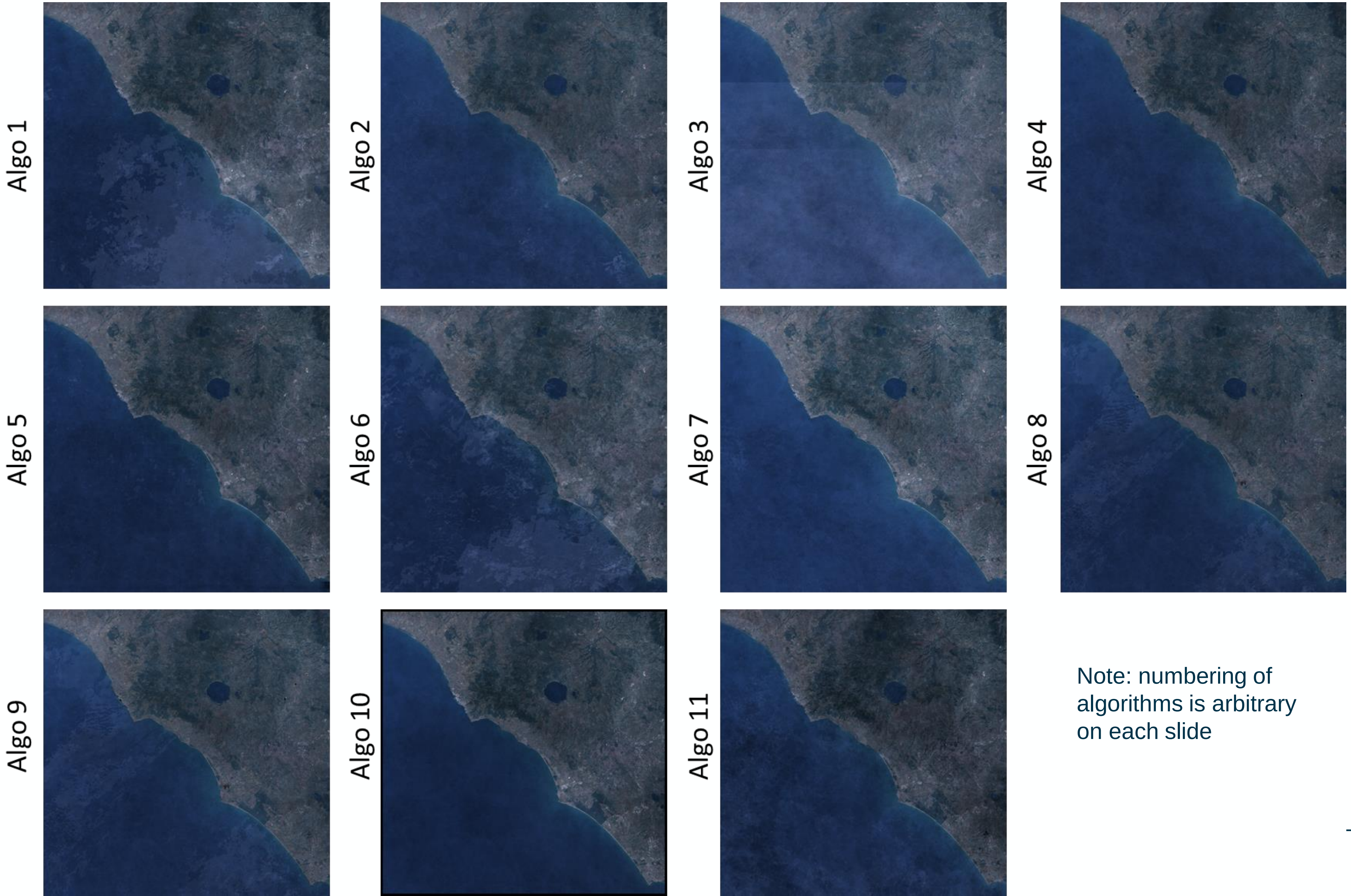
Results from test datasets

Results for multi-temporal TDS

11/18 algorithms provided results

Using the cloud flags to create mean annual cloud free composites

This method helps to identify algorithm problems (e.g., striping Algo 3) or higher cloud residuals (e.g., Algos 1,3,6)



Note: numbering of algorithms is arbitrary on each slide

Outlook

- **Definition of the inter-comparison protocol and reference datasets**
  - June 2022: First Workshop
- **Phase 1 – TDS exercise**
  - Provision of reference datasets samples
    - PixBox delivered to participants
    - Multi-temporal delivered to participants
    - Sky-Camera TDS currently being uploaded to FTP
    - Workshop for the collaborative TDS planned end Sep. early October
  - Application of the CM processors
    - Results for PixBox and multi-temporal received from the participants
  - Initial analysis of the results
    - PixBox and multi-temporal delivered to participants
  - Feedback from participants
    - No feedback received yet

Outlook

- **Phase 2 – Main exercise starting January 2024**
    - Preparation of the reference datasets
    - Provision of input datasets (to participants)
    - Application of the CM processors
    - Analysis of the results
    - Publication of complete datasets (reference + input data) after analysis
- **Results Analysis Report** (Internally)
    - Q1/Q2 2024
- **2nd Workshop of CEOS-WGCV CMIX II**
    - Approx. mid 2024 @ ESA or NASA