Brockmann Consult

Environmental Informatics • Geoinformation Services

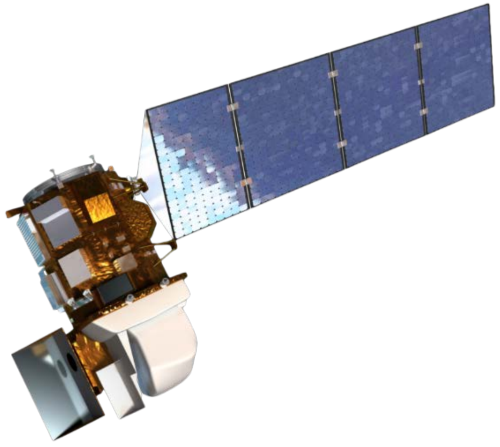
Report from CMIX

Carsten Brockmann, Jan Wevers Brockmann Consult

Sergii Skakun, NASA/Univ Maryland

Overview

- To **inter-compare** a set of cloud detection algorithms for space-borne high-spatial resolution (10-30 m) optical sensors
- Focus on **Landsat 8** and **Sentinel-2** data
- It’s **not a competition**, but **inter-comparison**
- **9 Participants** submitted **10 processor results**
- **7 validation** datasets
  - Sentinel-2: S2 Hollstein dataset, PixBox dataset, GSFC, CESBIO
  - Landsat 8: L8Biome, PixBox dataset, GSFC

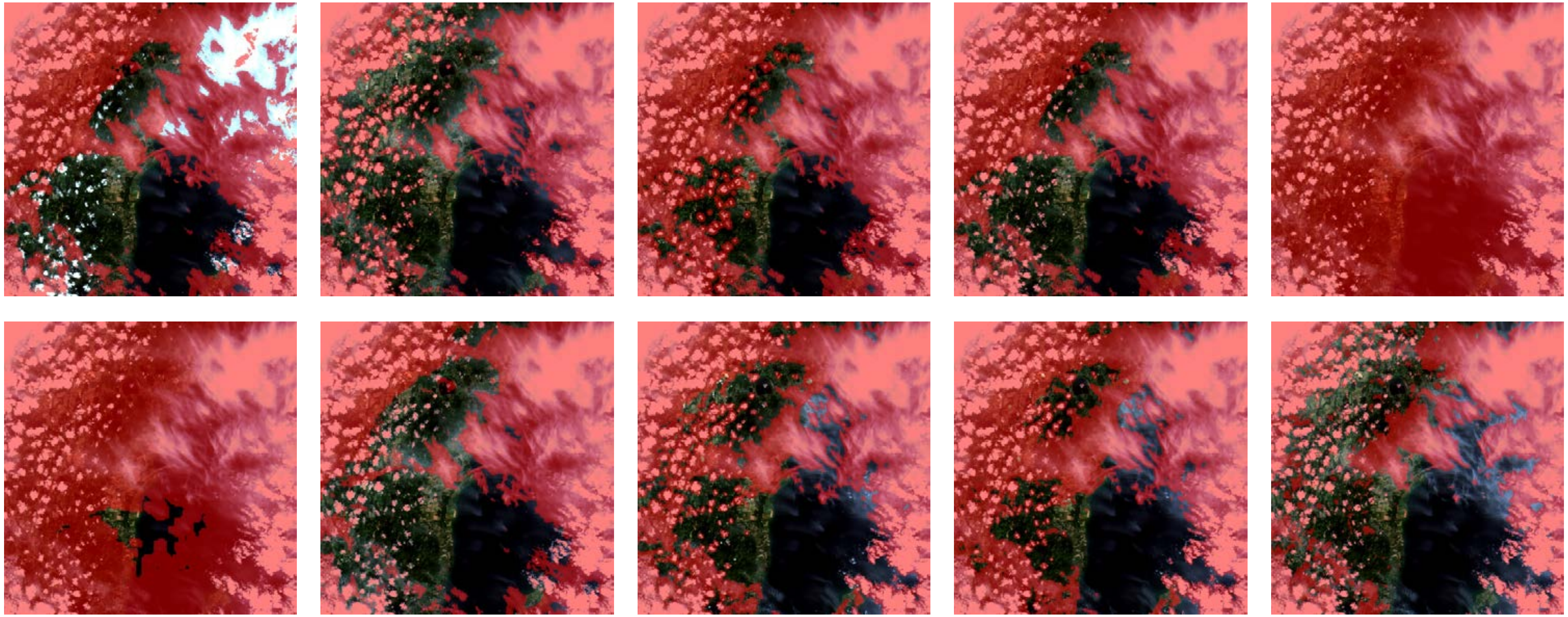


Algorithm Characteristics

Processor	Methodology	Resolution	temporality	Buffer	Fmask based
algo1	Spectral tests	20	Mono	No?	yes
algo2	Spectral tests (+ machine learning)	20	Mono	Small	no
algo3	Spectral tests	240	Multi	Large	
algo4	Spectral tests	20	Mono	Medium	yes
algo5	Spectral test + parallax (S2 only)	10	Mono	Yes	yes
algo6	Machine learning + spatio-temporal context	10	Multi	Medium	no
algo7	Spectral tests	10	Mono	No?	yes
algo8	Machine learning	10	Mono	Medium	no
algo9	Spectral test + auxiliary data	20	Mono	No	no
algo10	Machine learning	10/20/60	Mono	No?	no

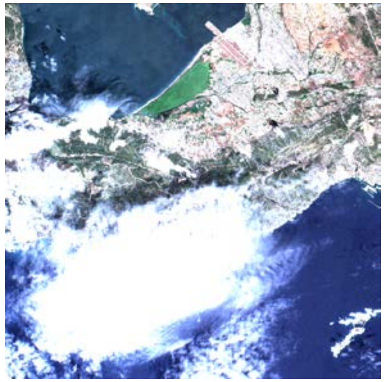
Sentinel 2 MSI – Quality inspections

Algo 1 Algo 2 Algo 3 Algo 4 Algo 5

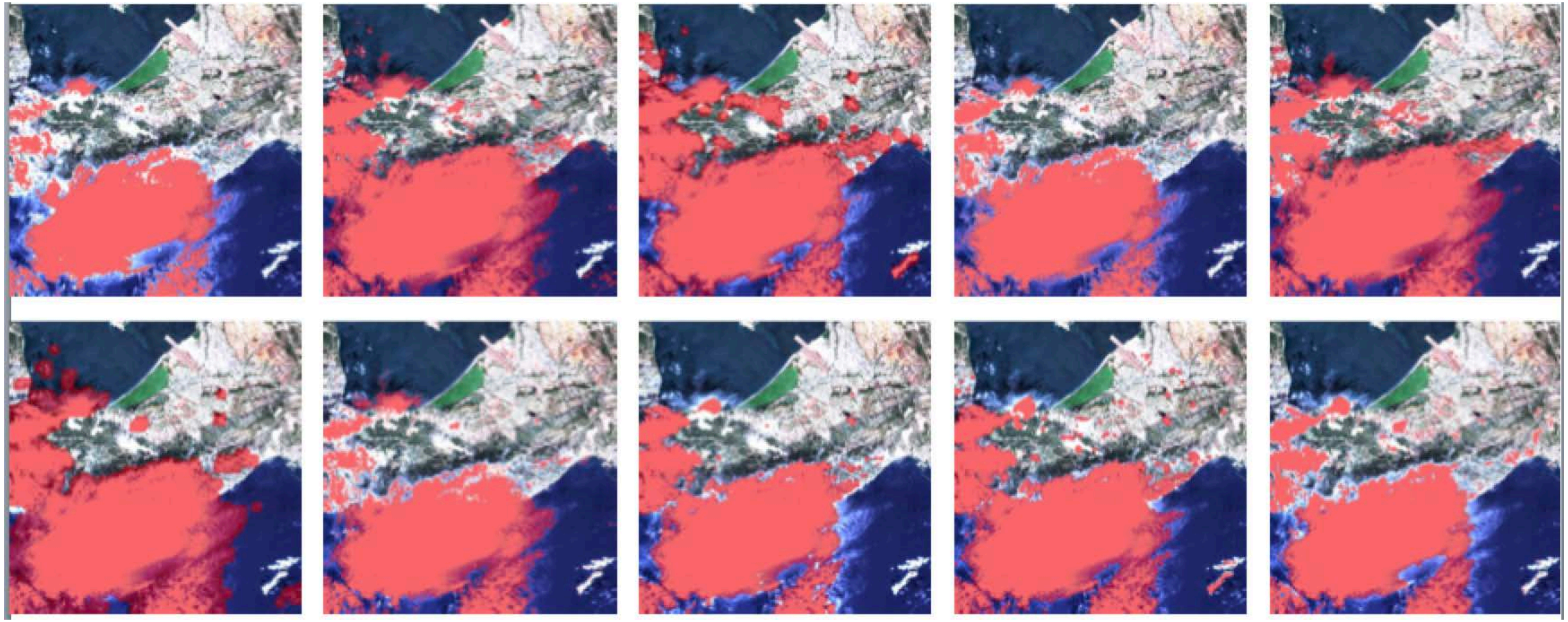


Algo 6 Algo 7 Algo 8 Algo 9 Algo 10

Sentinel 2 MSI – Quality inspections



Algo 1 Algo 2 Algo 3 Algo 4 Algo 5



Algo 6 Algo 7 Algo 8 Algo 9 Algo 10

Metrics (Confusion Matrix)

		Reference	
		Cloud	**No cloud**
Map	**Cloud**	cloud_as_cloud	ncloud_as_cloud
	No cloud	cloud_as_ncloud	ncloud_as_ncloud

$$Cloud\ PA = \frac{\text{cloud\_as\_cloud}}{\text{cloud\_as\_cloud} + \text{cloud\_as\_ncloud}}$$ (complimentary to OE = omission error)

$$Cloud\ UA = \frac{\text{cloud\_as\_cloud}}{\text{cloud\_as\_cloud} + \text{ncloud\_as\_cloud}}$$ (complimentary to CE = commission error)

$$Fscore = 2\,\frac{PA * UA}{PA + UA}$$

Validation Datasets Characteristics

Data	Satellites	Spatial resolution	# scenes	Method
Hollstein	S2	Polygons (20m)	59	Manual classification of polygons
PixBox	S2, L8	10m, 30m	29, 11	Manual classification of single pixels
L8Biome	L8	30 m	L8: 96	Photo-interpretation + ML classification
CESBIO	S2	60 m	S2: 30	Photo-interpretation + ML classification
GSFC	L8, S2	Polygons	L8: 6 S2: 28	Sky images + photo-interpretation

Validation Datasets - **S2** (*Hollstein et al. 2016*)

- There are 108 scenes (59 used for this exercise)
- Photointerpretation
- Selected polygons are labeled manually
- Classes:
  - clear sky, cloud, cloud shadow, cirrus, water, snow

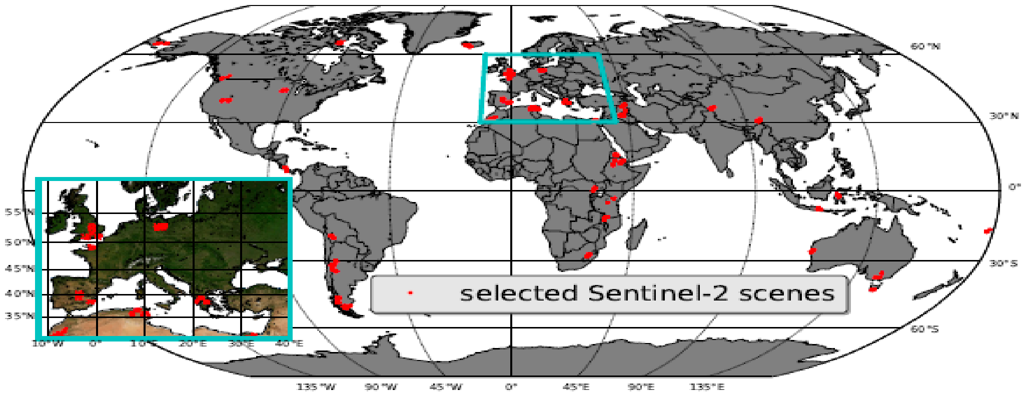
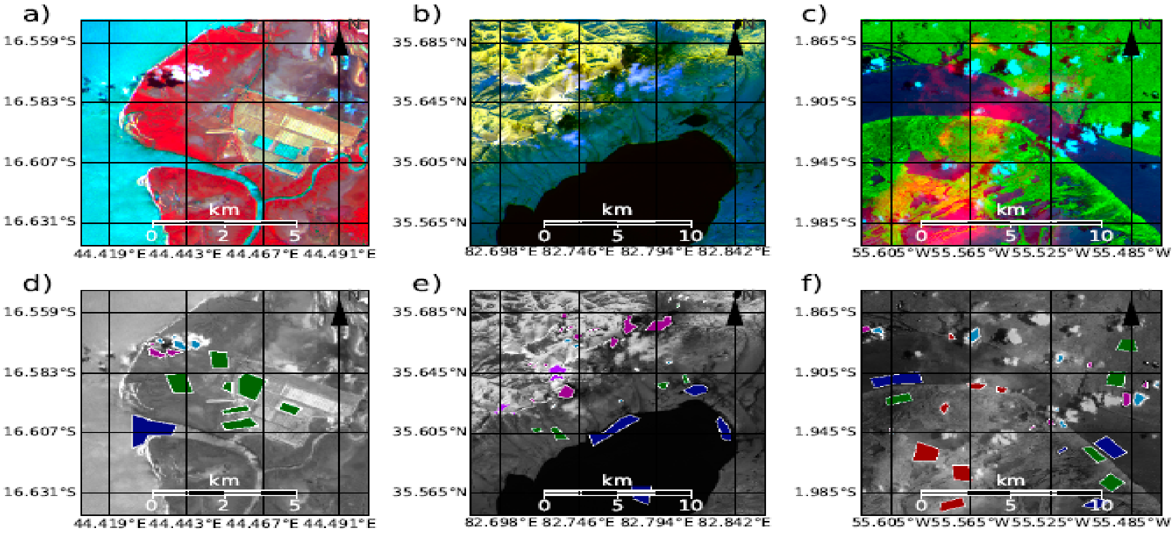


Figure 2. Global distribution of selected Sentinel-2 scenes which are included in the database.

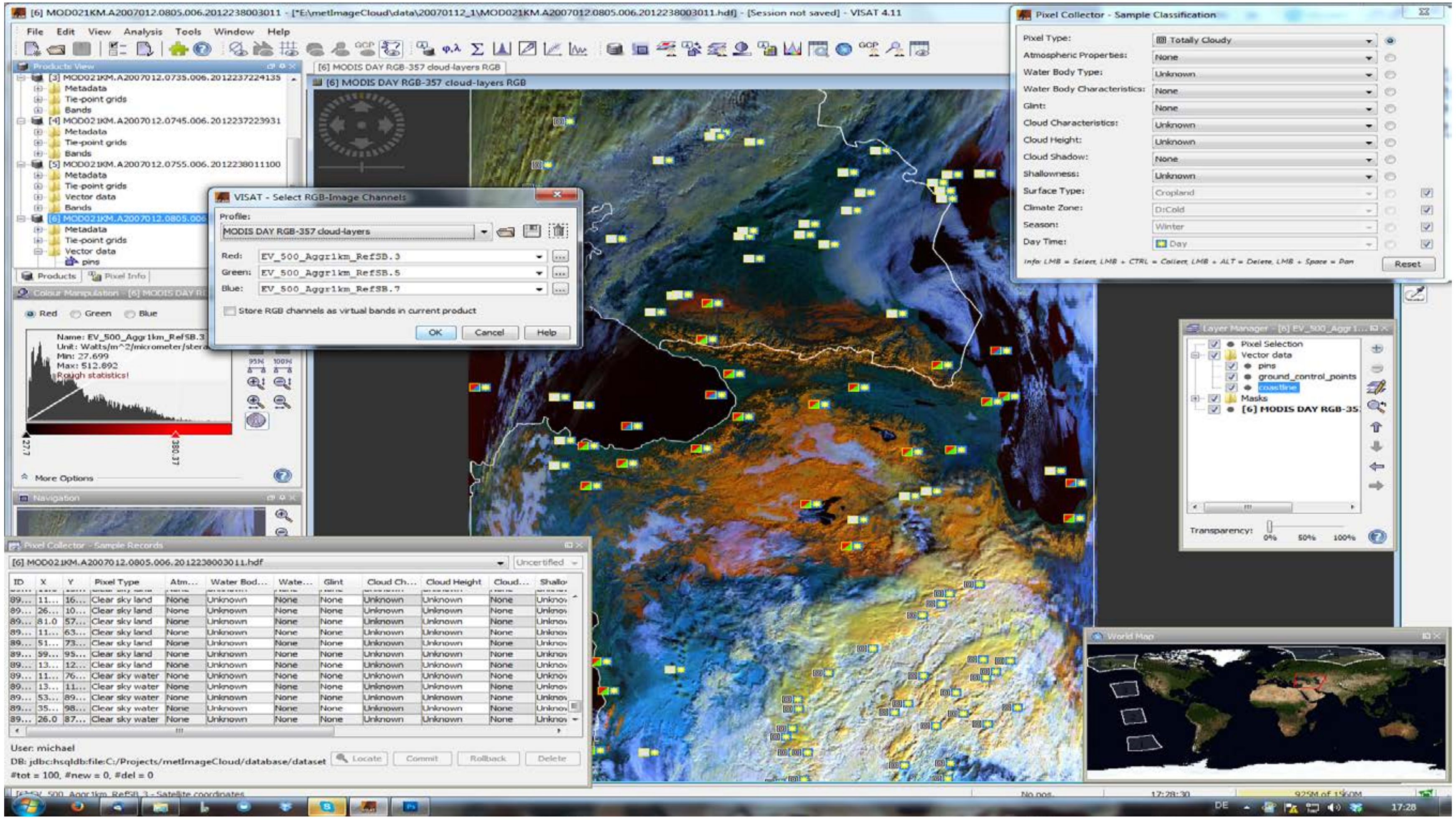


Validation Datasets - S2/L8: PixBox data set

Manually classified pixels single pixels.

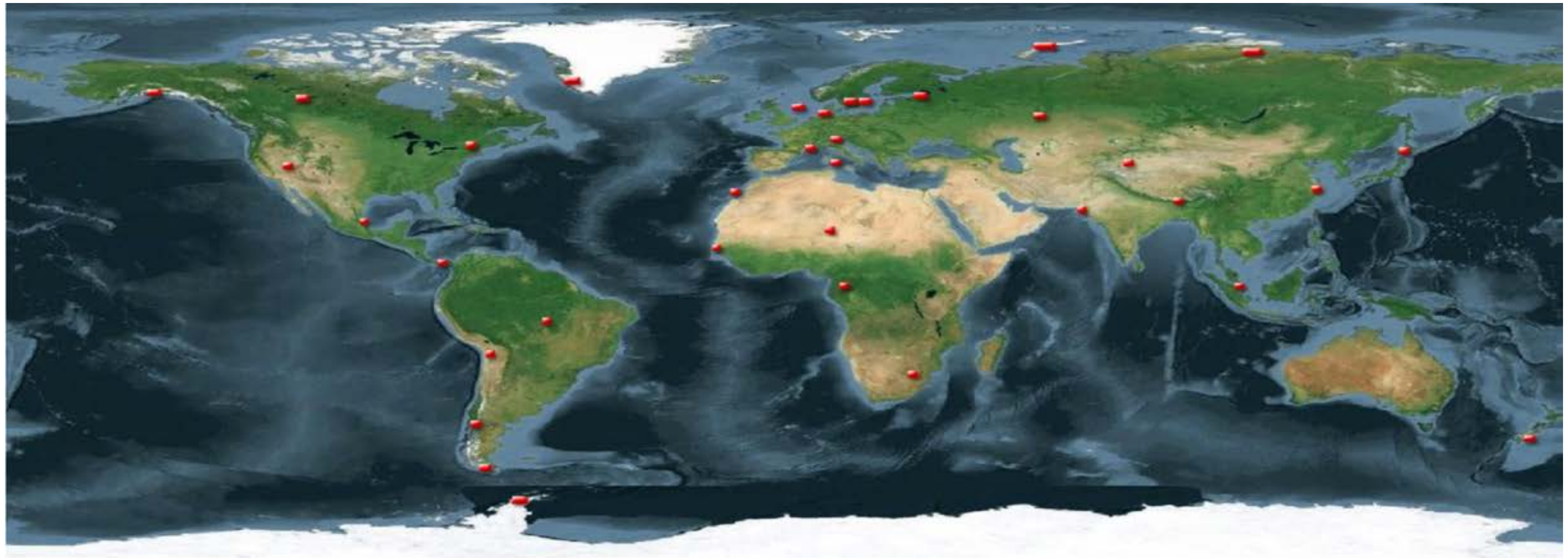
Great variety of surface properties per pixel is collected

Pixel collected using a dedicated SNAP tool.

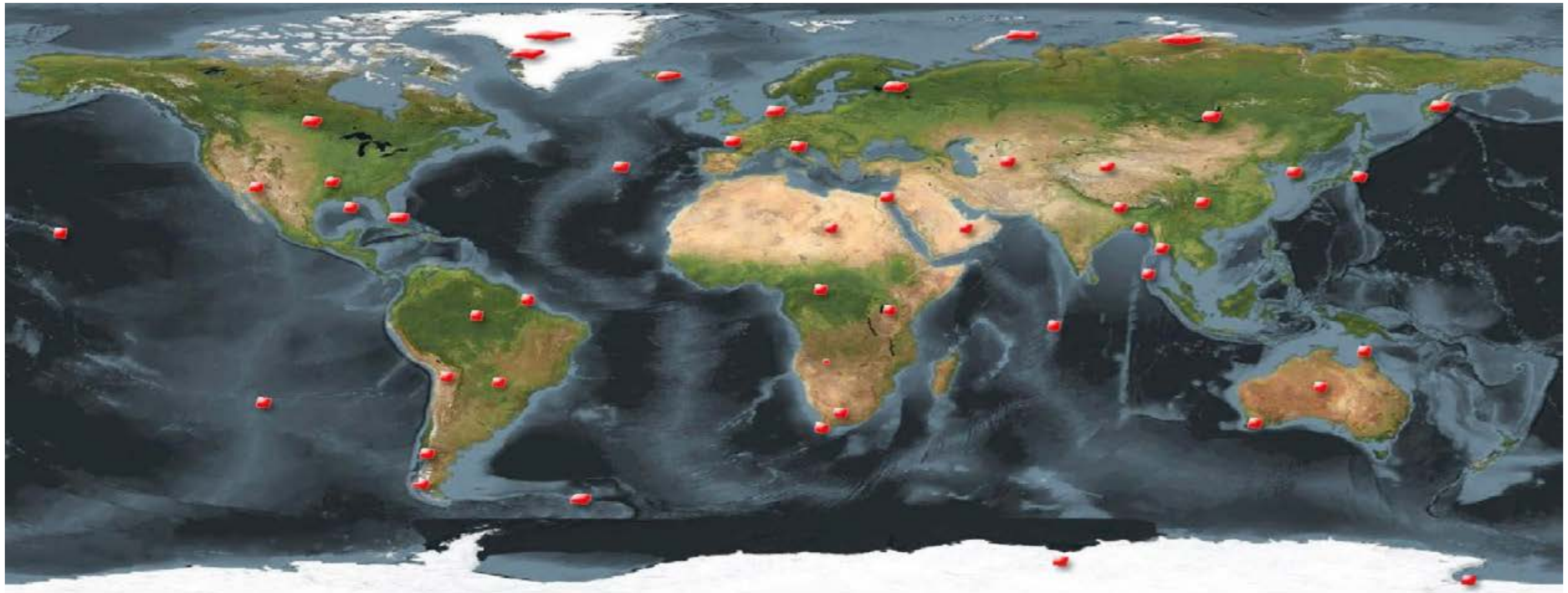


Validation Datasets - S2/L8: PixBox data set

S2 collection:
17351 pixels
From 29 products

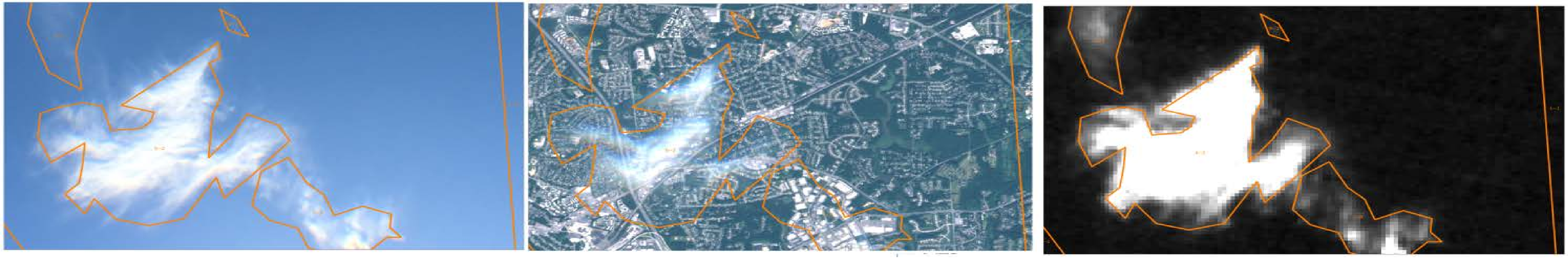


L8 collection:
20500 pixels
From 11 products



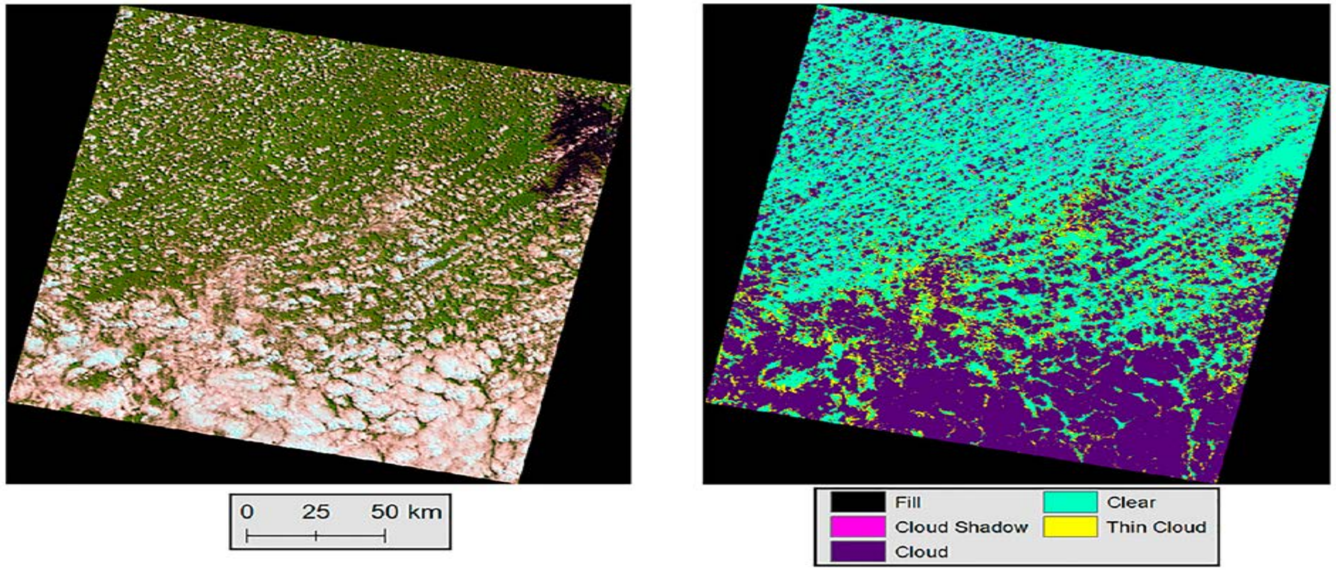
Validation Datasets - GSFC: LC8/S2A/S2B

- Manually labeled polygons assisted by ground photos of sky.
- 6 Landsat 8 and 28 Sentinel-2 scenes labelled.
- The same area over GSFC (also Aeronet measurements available), but varying conditions and time period.



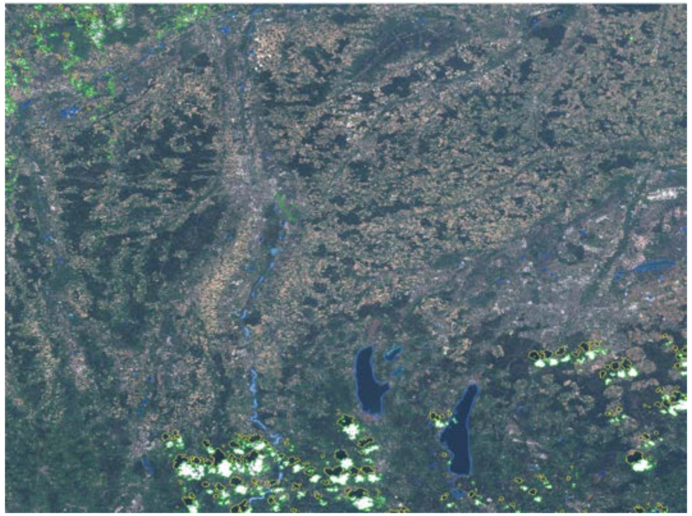
Validation Datasets - **L8Biome** (*Foga et al. 2016*)

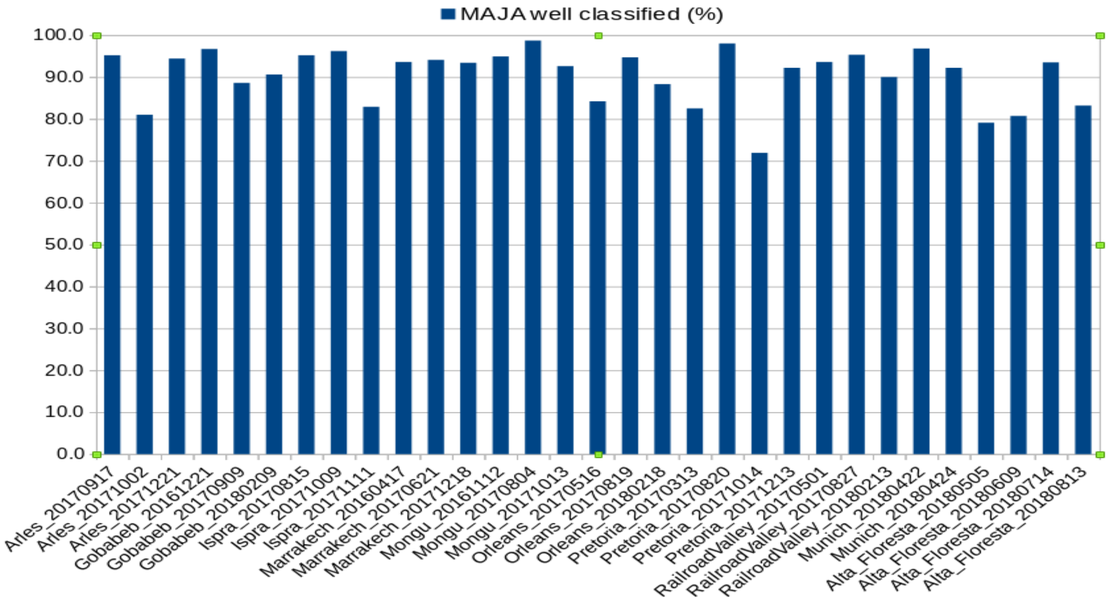
- 96 LC8 scenes, semi-random sampling with Biome stratification
- Photo-interpretation with See5.0
- All pixels are labelled (clear, cloud, cloud shadow, thin cloud).



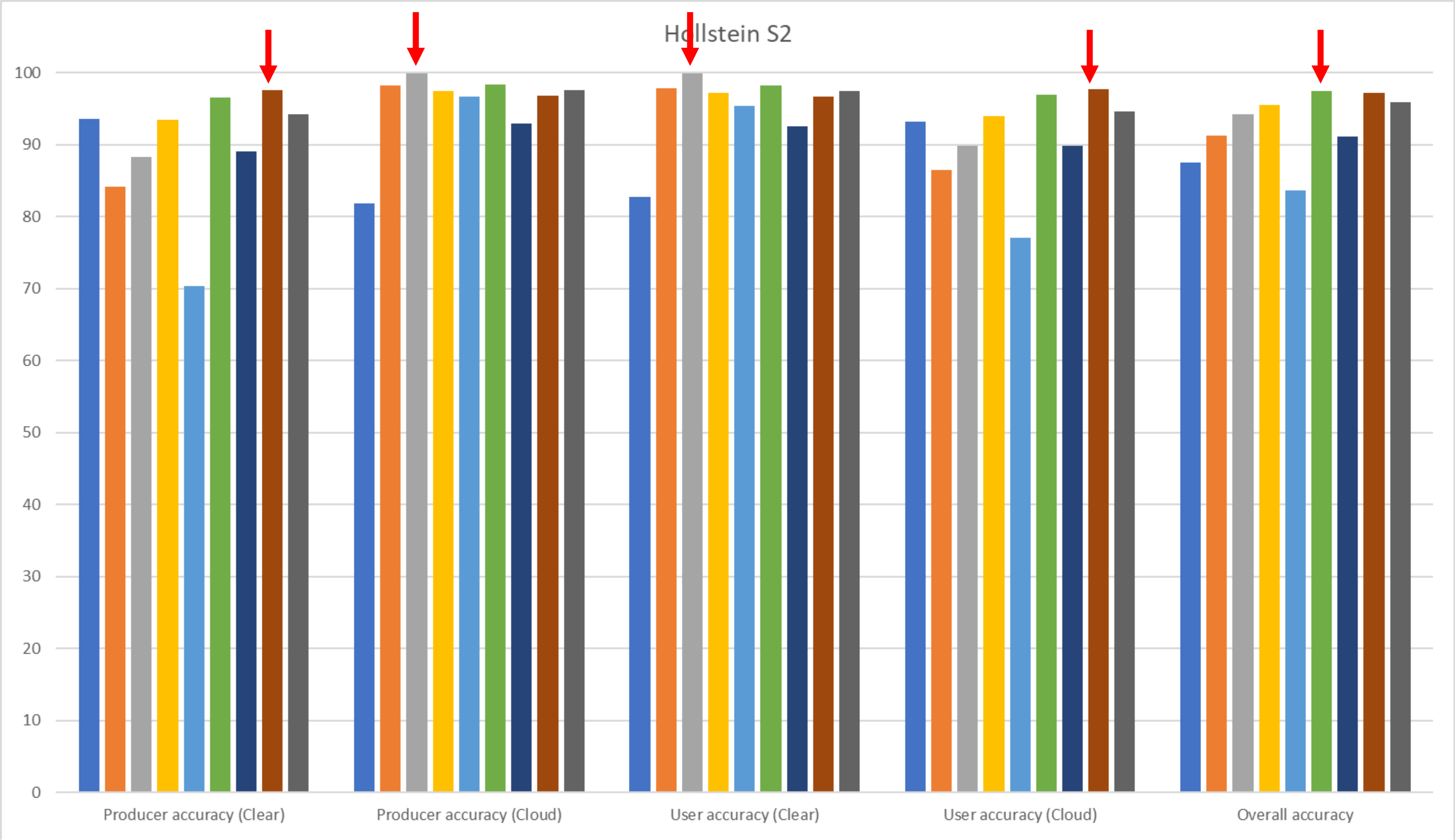
Validation Datasets - **CESBIO**

- 31 fully classified images at 60 m using active learning method
  - Manually supervised and iterative
  - Manual reference points added where first iterations not satisfying
  - Valid/Invalid pixels (an invisible cloud except in cirrus band is valid)
- Data and software are available, can be used to generate reference for ACIX-2 scenes
  - Would save processing for users
  - <2 hours of work per image





Results - Hollstein Dataset / Sentinel-2



Results – CESBIO Dataset

					Cloud			Non-cloud		
Processor	% cloud	Total num valid pixels	Overall accuracy	Balanced OA	PA	UA	F	PA	UA	F
Algo 1	24.26	772086084	88.59	80.38	64.42	84.92	73.26	96.34	89.42	92.75
Algo 2	24.26	772086084	91.7	86.89	77.54	86.86	81.94	96.24	93.04	94.61
Algo 3	25.57	697056918	89.24	90.45	92.92	72.65	81.54	87.98	97.31	92.41
Algo 4	24.26	772086084	93.27	88.88	80.36	90.83	85.27	97.4	93.93	95.63
Algo 5	24.26	3088386349	91.12	88.94	84.72	79.89	82.23	93.17	95.01	94.08
Algo 6	24.26	3088386349	93.23	88	77.84	93.12	84.8	98.16	93.25	95.64
Algo 7	24.26	3088386349	81.22	82.72	85.64	57.6	68.88	79.8	94.55	86.55
Algo 8	24.26	3088386349	93.12	88.79	80.4	90.18	85.01	97.19	93.93	95.53
Algo 9	24.26	772086084	91.03	84.66	72.29	88.65	79.64	97.03	91.62	94.25
Algo 10	24.26	85782723	89.45	79.54	60.28	94.11	73.49	98.79	88.59	93.41

CESBIO reference data are at 60 m

Results – GSFC (without thin clouds)

					Cloud			Non-cloud		
Processor	**% cloud**	**Total num valid pixels**	**Overall accuracy**	**Balanced OA**	**PA**	**UA**	**F**	**PA**	**UA**	**F**
Algo 1	55.47	10236222	86.91	88.19	76.43	99.96	86.63	99.96	77.29	87.18
Algo 2	55.47	10236222	92.52	92.47	92.93	93.55	93.24	92.01	91.27	91.64
Algo 3	40.78	13380304	92.71	92.16	89.13	92.72	90.89	95.18	92.71	93.93
Algo 4	55.47	10236222	96.12	96.47	93.3	99.69	96.39	99.64	92.27	95.81
Algo 5	55.47	40946551	95.91	96.14	94.02	98.54	96.23	98.26	92.95	95.53
Algo 6	55.47	40946551	95.61	96.01	92.37	99.7	95.9	99.65	91.29	95.29
Algo 7	55.46	40937538	97.96	97.89	98.51	97.83	98.17	97.27	98.13	97.7
Algo 8	55.47	40946551	95.74	96.09	92.95	99.33	96.03	99.22	91.87	95.4
Algo 9	55.47	10236222	94.97	95.43	91.17	99.73	95.26	99.7	90.06	94.64
Algo 10	55.47	40946551	92.86	93.55	87.26	99.85	93.13	99.84	86.29	92.57

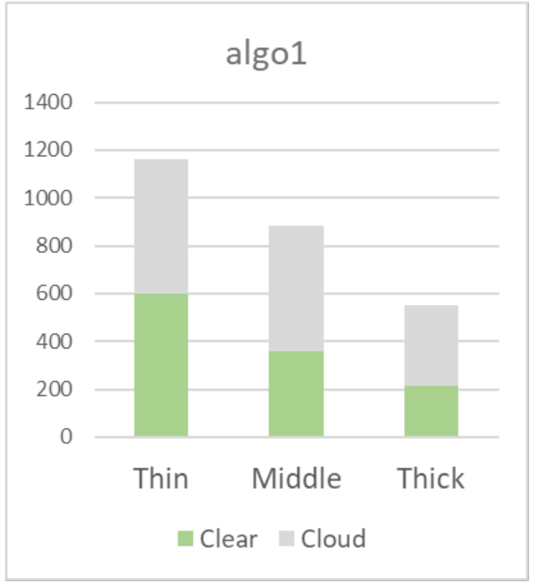
GSFC reference data are vector polygons

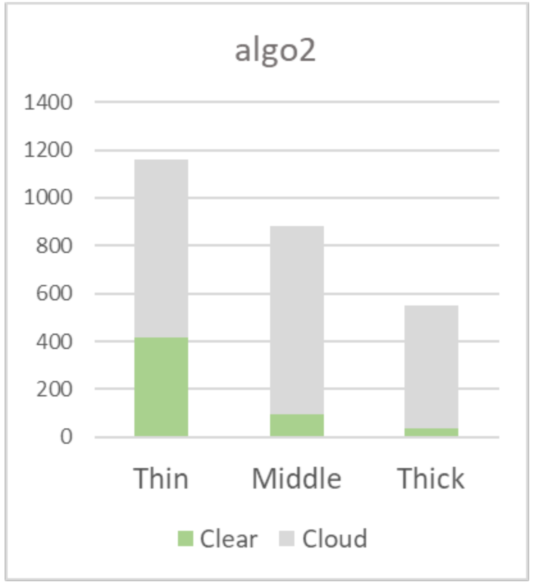
Results – PixBox

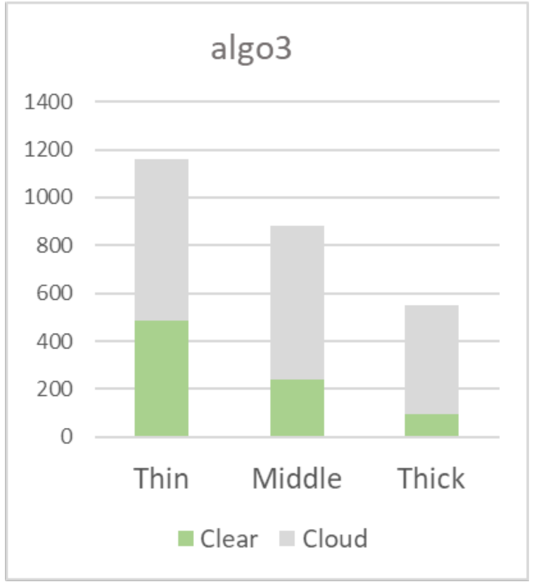
	PA (Clear)	PA (Cloud)	UA (Clear)	UA (Cloud)	OA	Scotts pi	Cohens Kappa	Krippendorf alpha
Algo 1	89.9	68.8	81.2	81.9	81.46	0.601	0.603	0.601
Algo 2	64.7	92.4	92.9	63.2	75.69	0.513	0.529	0.513
Algo 3	82.3	94.5	96.3	75.6	86.78	0.727	0.729	0.727
Algo 4	89	91.2	94.2	83.7	89.83	0.787	0.788	0.787
Algo 5	81.2	90.9	93.5	74.9	84.88	0.69	0.692	0.69
Algo 6	95.2	86.3	91.8	91.8	91.82	0.824	0.824	0.824
Algo 7	48.2	91.1	89.8	52.1	64.58	0.289	0.343	0.289
Algo 8	91.6	91.7	94.7	87.1	91.65	0.824	0.824	0.824
Algo 9	86.9	78.5	86.8	78.8	83.71	0.654	0.654	0.654
Algo 10	93.4	80.6	88.6	88.3	88.5	0.751	0.752	0.751

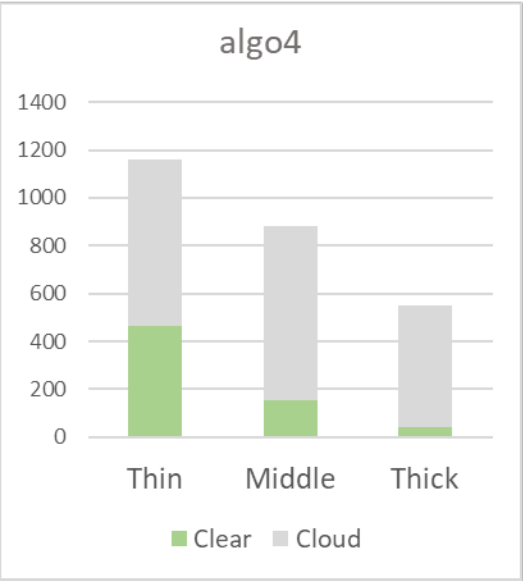
PixBox reference data are single pixels spatially and temporally well distributed with a high detail on pixel information (10m resolution)

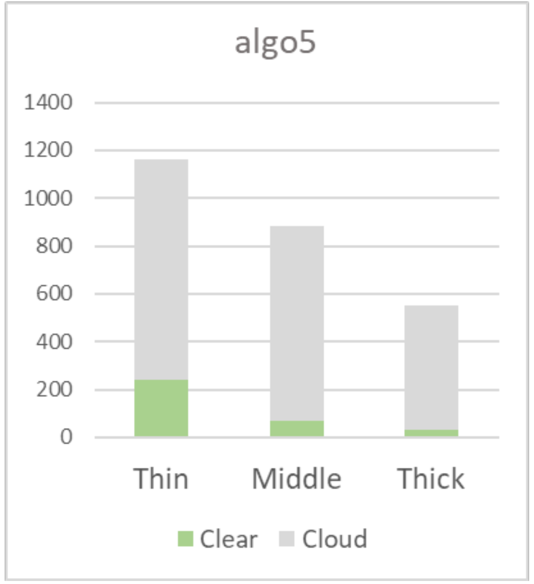
Semi transparent clouds

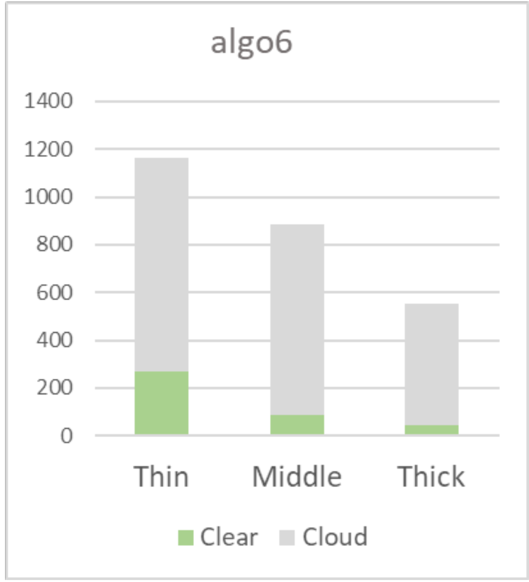












Results – Generalisation

Green: "feature" contributes significantly to detecting "case“

Yellow: "feature" can contribute; should be combined with others

orange : “feature” but is not very sensitive

red : does not help at all

blue-ish: cell not assessed

grey : not applicable

Cases

Processor features

	Opaque clouds	Cloud borders	Thick semi-trans clouds	Thin to medium semi-trans clouds	Cloud/snow w seperation	Bright targets
Spectral test based						
Machine learning						
Mono temporal						
Multitemporal						
No buffer						
Small buffer						
Large buffer						

Summary Results of CMIX I

- All results show high accuracies (> 80%) for all processors -> cloud screening is good but can be improved.
- CMIX had shown that there is no clear superiority of any methodology (Spectral tests vs. AI, mono vs. multitemporal)
- Thin semi-transparent clouds and cloud boundaries are an issue for mostly all algorithms. → How to define a transparent cloud? boundary of a cloud?
- A buffer and its size have a strong influence on the validation results → Bigger buffer = better results
- Current validation datasets and validation methodology is insufficient
- CMIX II is recommended

Summary Results of CMIX I concerning Validation

- All 5 validation datasets (VD) have different strengths and weaknesses -> results vary depending on the validation dataset
  - The combination of 5 VDs allow a more detailed analysis of strengths and weaknesses
  - Subjectivity of detecting/photo-interpreting clouds, especially thin clouds
  - Use of automated cloud detection for VD means that we do an algorithm intercomparison
  - The current VD and method does not allow for detecting systematic errors
  - It was recommended to collect a dedicated validation data set
  - Complement statistical approach with “critical scenes”

- Current cloud masking and validation is lacking objectivity and a rigorous definitions of the target
  - Cloud class should be uniquely defined, and consistent between algorithms‘ flag and validation data
  - Cloud/no cloud too simplistic; more refined (at least slightly) should be used
  - Ongoing EUMETSAT study is developing a physically based approach (obscureness of signal)
  - ESA/OLCI may be moving to “clear sky identification” rather than cloud masking