ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Sharing is encouraged

EGU

Cloud Identification and Classification from Ground Based and Satellite Sensors on the Antarctic Plateau

Authors: Michele Martinazzo, Viviana Volonnino, Tiziano Maestri, Fabrizio Masin, Gianluca Di Natale, Luca Palchetti, Giovanni Bianchini, and Massimo Del Guasta.

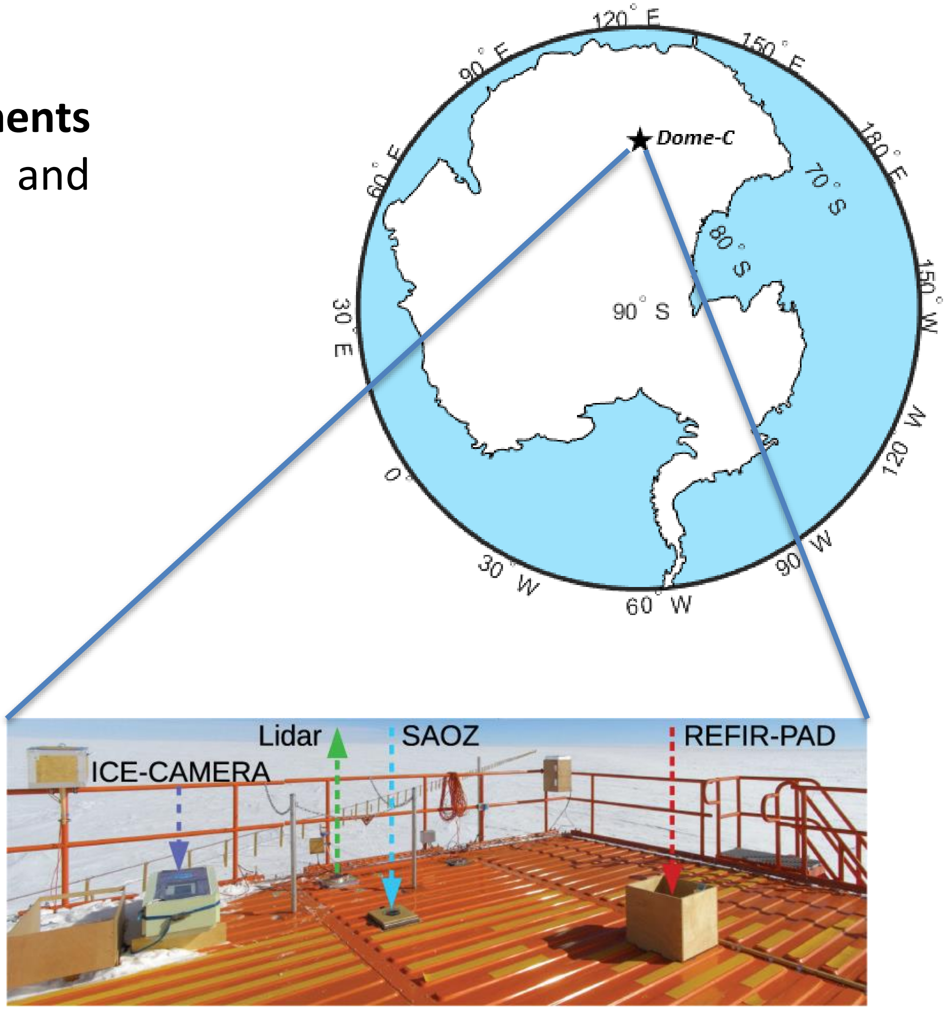
Outline

Problem:

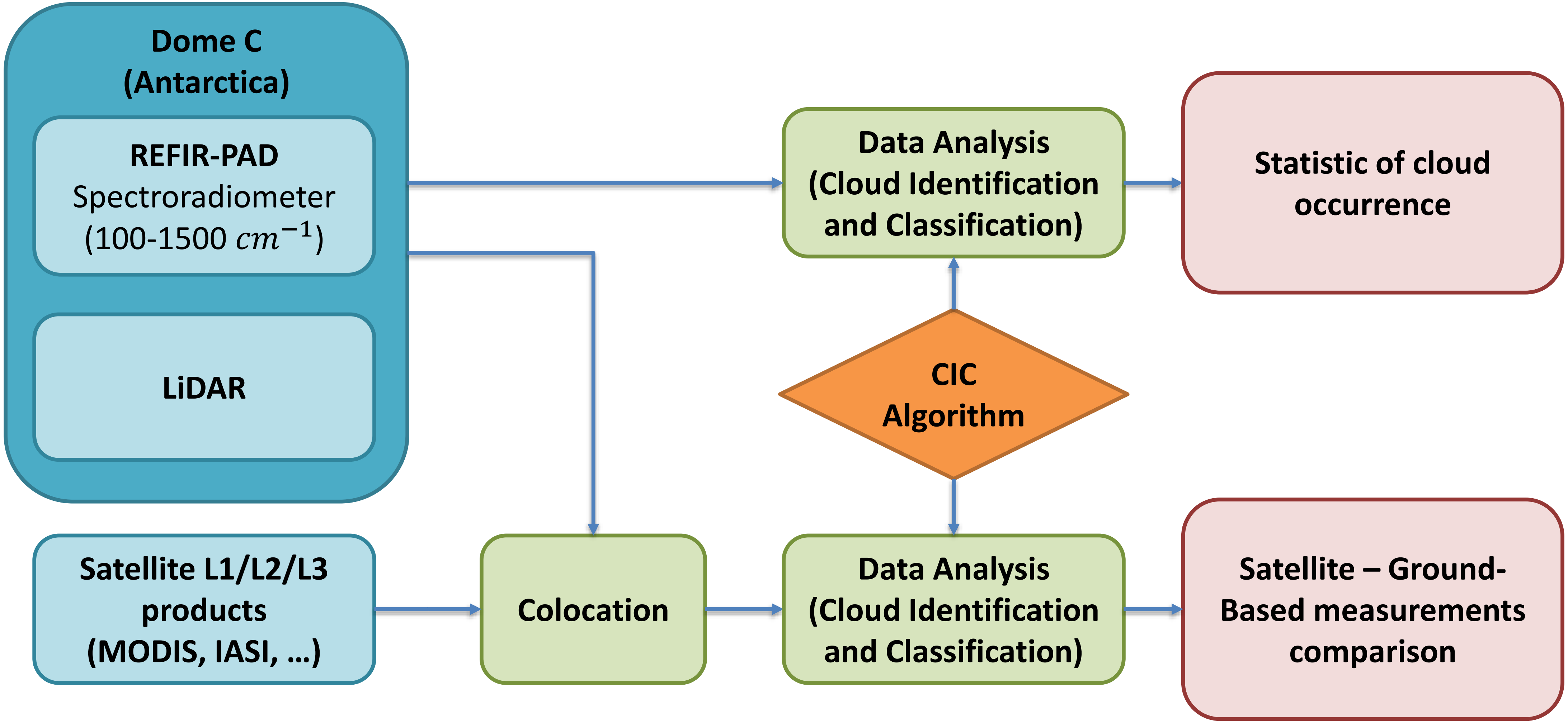
Cloud identification from satellites is challenging **in polar environments** (especially for thin ice clouds). This is due to a similarity in thermal and radiative properties of the surface and the cloud layer.

- Uncertainties about cloud radiative and **microphysical properties** and radiation budget contribution.
- Uncertainties on retrieved atmospheric and **surface products** (when incorrectly unidentified).
- Introduction of parametric **errors in climate models**.

We aim at providing a **reference statistic of cloud occurrence** and testing the performances of satellite sensors in detecting clouds at Polar latitudes.



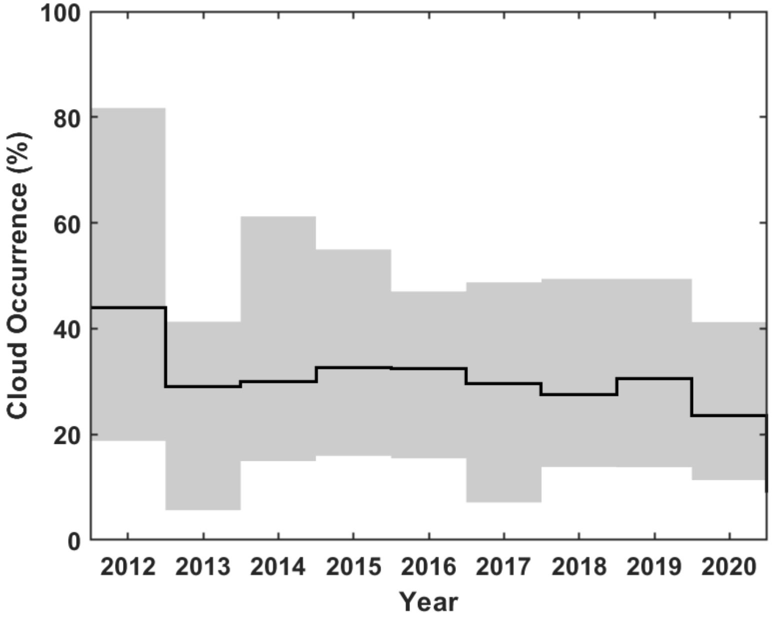
Data analysis at Dome-C (Antarctic Plateau)

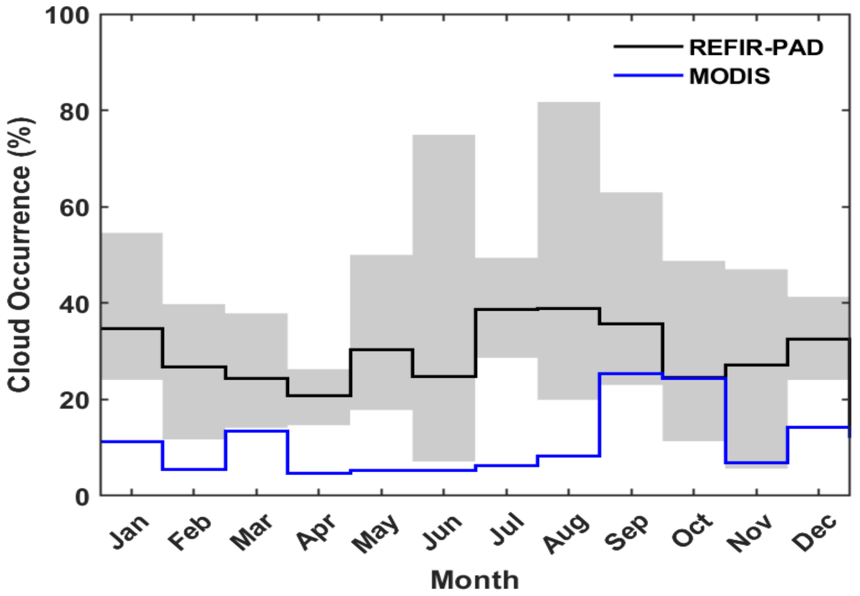


Main Results

Almost **70%** of the full dataset is composed of **clear-sky** elements, the rest 30% is divided between **ice clouds (almost 28%) and mixed-phase clouds (2%)**.

Different satellite products are considered, and the analysis are produced both for space and time collocated data:
Important discrepancies between the ground based and satellite data are found.





Year	Clear-sky (%)	Ice cloud (%)	Mixed-phase cloud (%)	Observation Time (%)
2012	56.07 ± 4.70	38.25 ± 0.66	5.68 ± 0.92	27.38
2013	71.01 ± 5.96	27.51 ± 0.47	1.48 ± 0.24	29.61
2014	69.97 ± 5.87	27.97 ± 0.48	2.06 ± 0.33	67.36
2015	67.46 ± 5.66	30.70 ± 0.53	1.84 ± 0.30	65.27
2016	67.50 ± 5.66	30.55 ± 0.53	1.96 ± 0.32	82.32
2017	70.42 ± 5.90	27.83 ± 0.48	1.74 ± 0.28	79.86
2018	72.48 ± 6.08	24.35 ± 0.42	3.17 ± 0.51	87.14
2019	69.52 ± 5.83	28.24 ± 0.49	2.25 ± 0.36	90.75
2020	76.39 ± 6.41	22.13 ± 0.38	1.48 ± 0.24	90.57
Total	**70.09 ± 5.88**	**27.70 ± 0.48**	**2.25 ± 0.36**	69.11

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Thank you for your attention.

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Cloud Identification and Classification from Ground Based and Satellite Sensors on the Antarctic Plateau

(Long Version)

Authors: Michele Martinazzo, Viviana Volonnino, Tiziano Maestri, Fabrizio Masin, Gianluca Di Natale, Luca Palchetti, Giovanni Bianchini, and Massimo Del Guasta.

Data analysis at Dome-C (Antarctic Plateau)

Within the PNRA project **FIRCLOUDS** a remote ground-based station on the Antarctic Plateau is used to test ability to observe and derive cloud properties by exploiting the far infrared (FIR, 100-600 cm^{-1}) and mid infrared spectral regions.

We aim at:

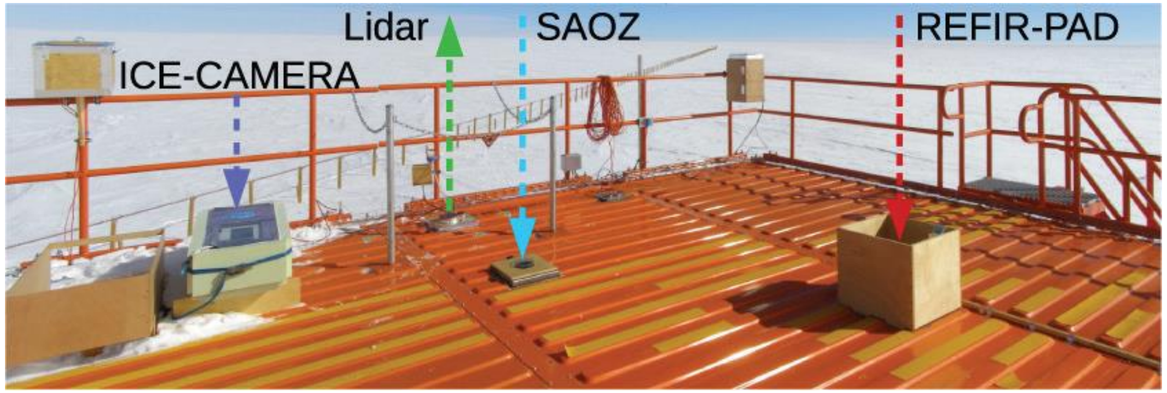
- Testing **identification/classification** algorithms for high spectral resolution FIR-MIR remote sensing data
- Providing a reference **statistic of cloud occurrence** at multiple timescales and investigating for possible correlations with atmospheric parameters
- Comparing **cloud occurrence statistic** on Dome-C from ground-based measurements with **satellite L2 products**
- Testing satellite performances in **Cloud identification** on the Antarctic Plateau

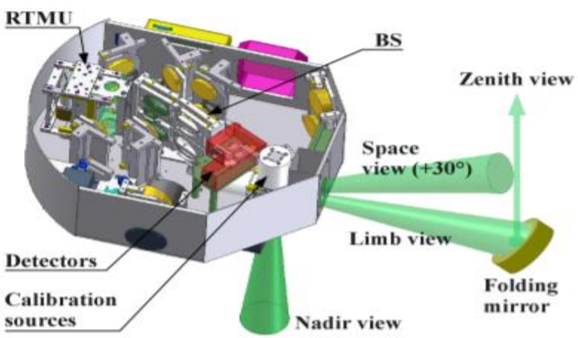
Concordia Station – Dome C (75.1°S, 123.3°E, 3233m)

REFIR-PAD

Parameter	Values
Spectral bandwidth	**100–1500 cm^{-1}** (100–6.7 μm)
Spectral resolution	0.4 cm^{-1} (double-sided interferograms)
Optical throughput	0.01 cm^2 sr
Line of sight	Zenith looking with a field of view of about 100 mrad
Single-spectrum integration time	80 s
Measurement	∽ 5.5 min (average of four observations), Repetition rate ∽ **14 min** (sequence duration)
Measurement cycle	5-6 hours of measurements 1-3 hours of analysis
NESR	About 10^{-3} W/(m^2 sr cm^{-1}) at 400 cm^{-1}

Year	2012-2020
# Spectra	**233508**



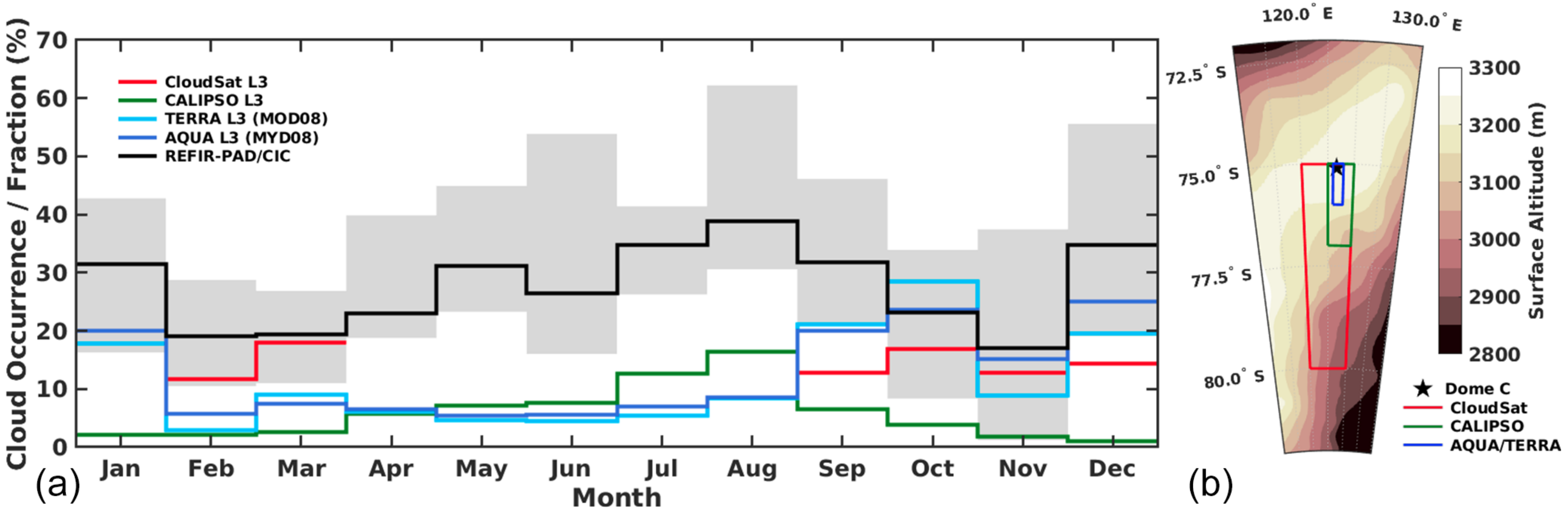


LiDAR

Parameter	Values
Channels	Backscatter and depolarization channels
Wavelength	532 nm (linear polarization)
Measurement range	30–7000 m
Vertical resolution	7.5 m
Line of sight	Zenith looking through a window all weather
Telescope	10 cm diameter, f = 30 cm refractive optics
Filter	0.15-nm interference filter
Laser	Quantel (Brio)

Cloud occurrence on the Antarctic Plateau

- **Motivation of the work:** Cloud identification from satellites is challenging in polar environments due to a similarity in thermal and radiative properties of the surface and the cloud layer. Both active and passive measurements suffer from various issues which affect cloud detection in Artic and Antarctic regions.

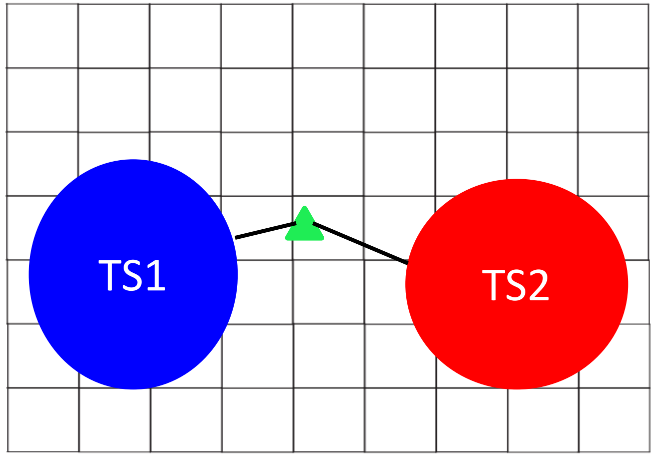


Sensor	Product (area)	Notes
MODIS	Cloud fraction (3 000 km^2)	Low efficiency at IR – Dependent on solar zenith angle
CALIOP	Vol. cloud occurrence (15 000 km^2)	Low efficiency to detect thin clouds close to the ice surface
CLOUDSAT	Vol. cloud occurrence (75 000 km^2)	Unable to detect thin cirrus (vertical res 500 m)

Analysis Methods: The CIC Algorithm

CIC is a machine learning algorithm that performs a **principal component analysis** (PCA) to classify the scene as clear or cloudy, and to identify the type of cloud – multi-class comparison

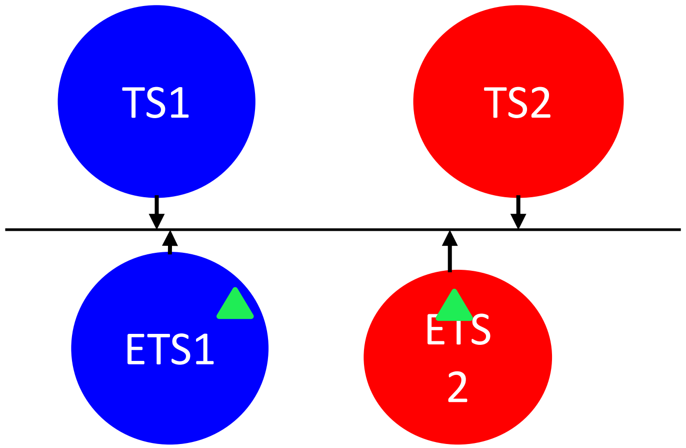
Support Vector Machine with linear kernel analyses the key features of two classes (defined by a linear discriminant analysis).
A supervised learning algorithm defines the features of 2 reference training sets and a metric to establish the distance between the analyzed element and the TSs



The metric is defined by the comparison of the features of the TSs and of the element

Spectrum to be classified

CIC exploits the PCA analysis to extract the information content of 2 reference TSs and their change due to the addition of the analyzed element. This allows to rely on small TSs and to easily implement the algorithm to different systems



The metric is defined by the information content of the TS

CIC - Algorithm

The number of **significant principal components** (P_0) for each TS, that bring the **information about the TS variability**, is computed by minimizing the *IND* factor (Turner et al., 2006)

$$IND(p) = \frac{RE(p)}{(P-p)^2} \quad \text{where} \quad RE(p) = \sqrt{\frac{\sum_{i=p+1}^{P} \lambda_i}{T_X(P-p)}}$$

A **similarity index (SI)** is defined as the comparison of the significant eigenvalues of TS and ETS **for each class**

$$SI_X = 1 - \frac{1}{2P_0} \sum_{p=1}^{P_0} \sum_{\nu=1}^{\nu_{tot}} |ETSEM_X(\nu,p)^2 - TSEM_X(\nu,p)^2|$$

The SIs are compared in couples, for all the classes. **The highest SI in each couple defines a «score point» for that class – Elementary approach**.

$$SID = SI_{CLOUD} - SI_{CLEAR}$$

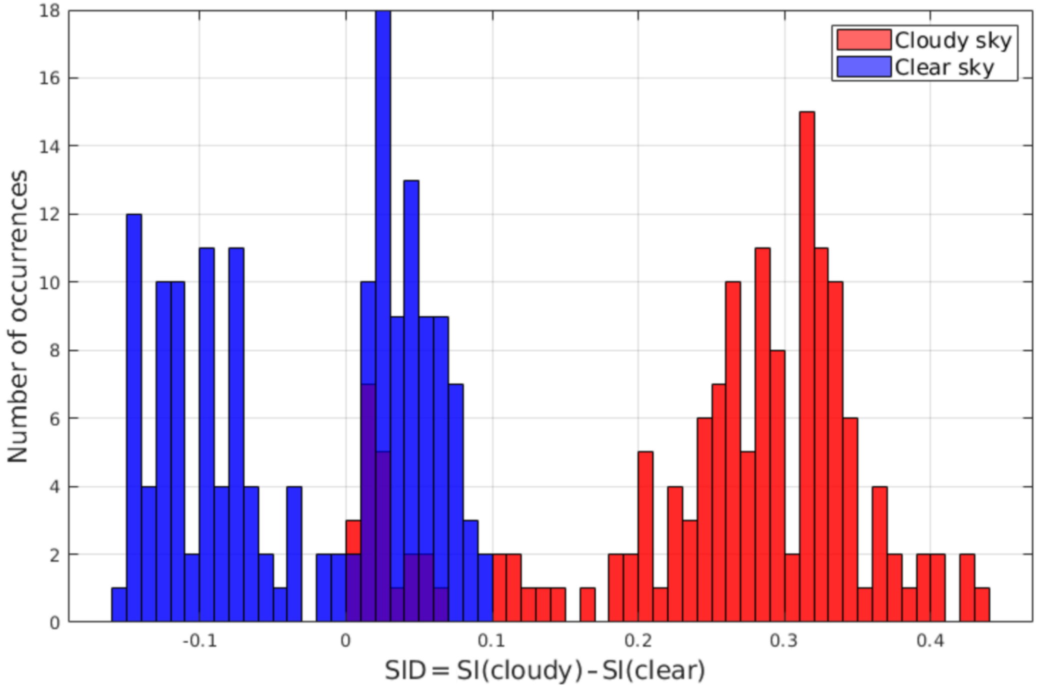
$SID < 0$ → CLEAR SKY or $SID > 0$ → CLOUDY SKY

CIC - Algorithm

Results might be affected (and biased) if one of the training sets is not well populated by spectra that are representative of the variability within the class. In this case, a distributional approach can be adopted for the classification in which the distribution of the SIDs of the training set is analyzed before performing the classification

$$CSID = SID - shift_{opt}$$

The $shift_{opt}$ is computed maximizing the number of TS spectra correctly classified (True Positive)



Maestri et al., 2019

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

CIC: optimization and error definition

Training Sets

Spectra co-located with lidar observations (warm and cold season)

A **metric of the information content** is defined by applying the CIC to the elements of the TSs

CIC

Verification Set

$$\mathrm{ThS} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}}$$

lidar co-located spectra

CIC is optimized by finding the highest Threat Scores for all the classes → **380-1000 cm^{-1}**

Test Set

$$\epsilon_A\% = \frac{N_A^{CIC}}{N_{TOT}} \times \left(\frac{1}{HR_A} - 1\right) \times 100$$

CIC performances are evaluated on a test set on which an Hit Rate is defined that is used to provide **classification percentual error**

Dataset
233508 spectra

The 11 years dataset is analyzed and statistics are provided on multiple time scales

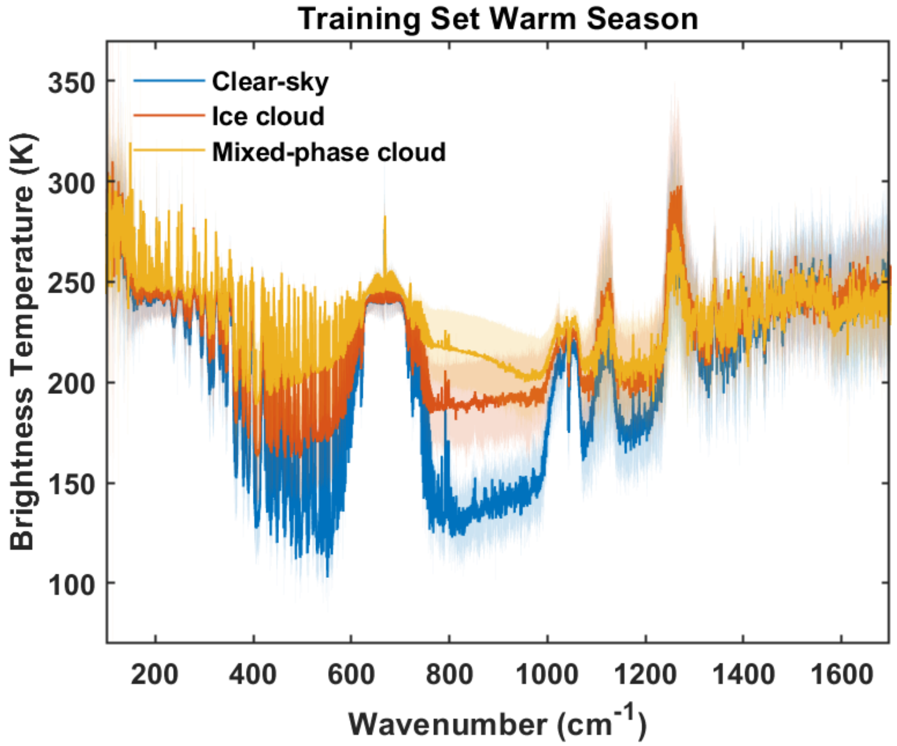
CIC applied to REFIR-PAD spectra – TRAINING SET

Spectra composing the Training Sets are selected from a subset of manually classified observations. This pre-classification is performed using the **LiDAR** instrument.

LiDAR *backscatter* signal increases in presence of cloud layers

- *Depolarization* < 0.15 → mixed-phase cloud
- *Depolarization* > 0.15 → ice cloud

Training Set || 119 spectra			
	Clear-sky	**Ice cloud**	**Mixed-phase cloud**
Warm Season	23	22	14
Cold Season	40	20	-



CIC applied to REFIR-PAD spectra – TEST SET

Indices used to evaluate the algorithm performance:

Threat Score $$ThS = \frac{TP}{TP + FN + FP}$$

Hit Rate $$HR = \frac{TP}{TP + FN}$$

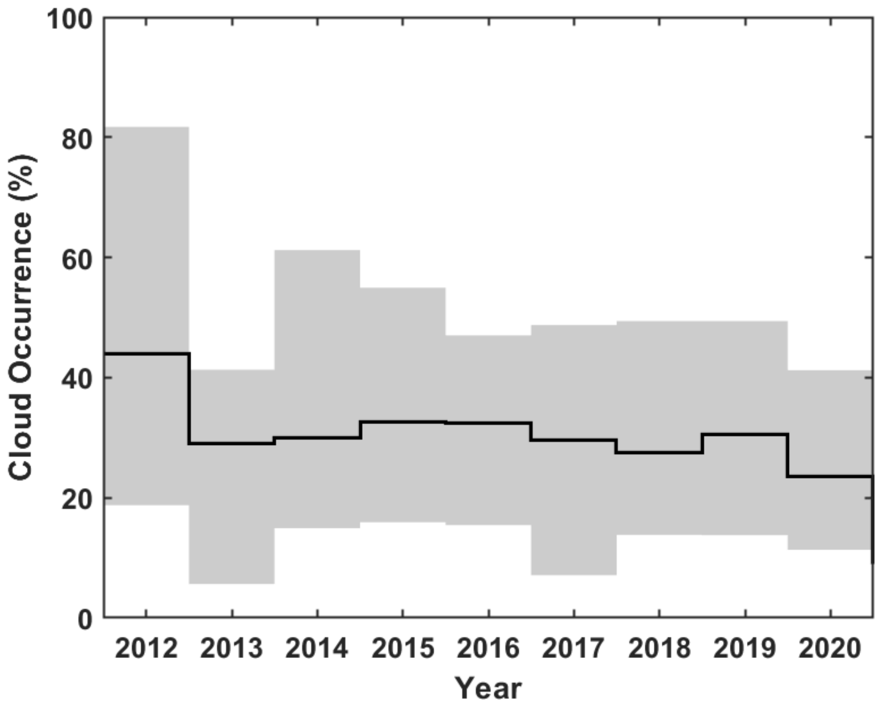
Positive Predictive Value $$PPV = \frac{TP}{TP + FP}$$

Field	N. Spectra	ThS	HR	N. Misclass spectra	Misclassification	PPV
Clear-sky	323	**0.91**	**0.92**	25	7.8% ice cloud 0% mixed-phase cloud	**0.99**
Ice cloud	590	**0.93**	**0.98**	10	0.68% clear-sky 1.02% mixed-phase cloud	**0.94**
Mixed-phase cloud	79	**0.80**	**0.86**	11	0% clear-sky 13.92% ice cloud	**0.92**
Tot	**992**	**0.91**	**0.95**	**46**	**4.6%**	**0.95**

95% of spectra are correctly classified.

CIC applied to REFIR-PAD spectra – TEST SET

The CIC is finally run over the full dataset (2012-2020). Results are provided in terms of percentages, with an associated error. Results refer to the spectral interval [380-1000]cm^{-1} and the maximum num. of PCs.



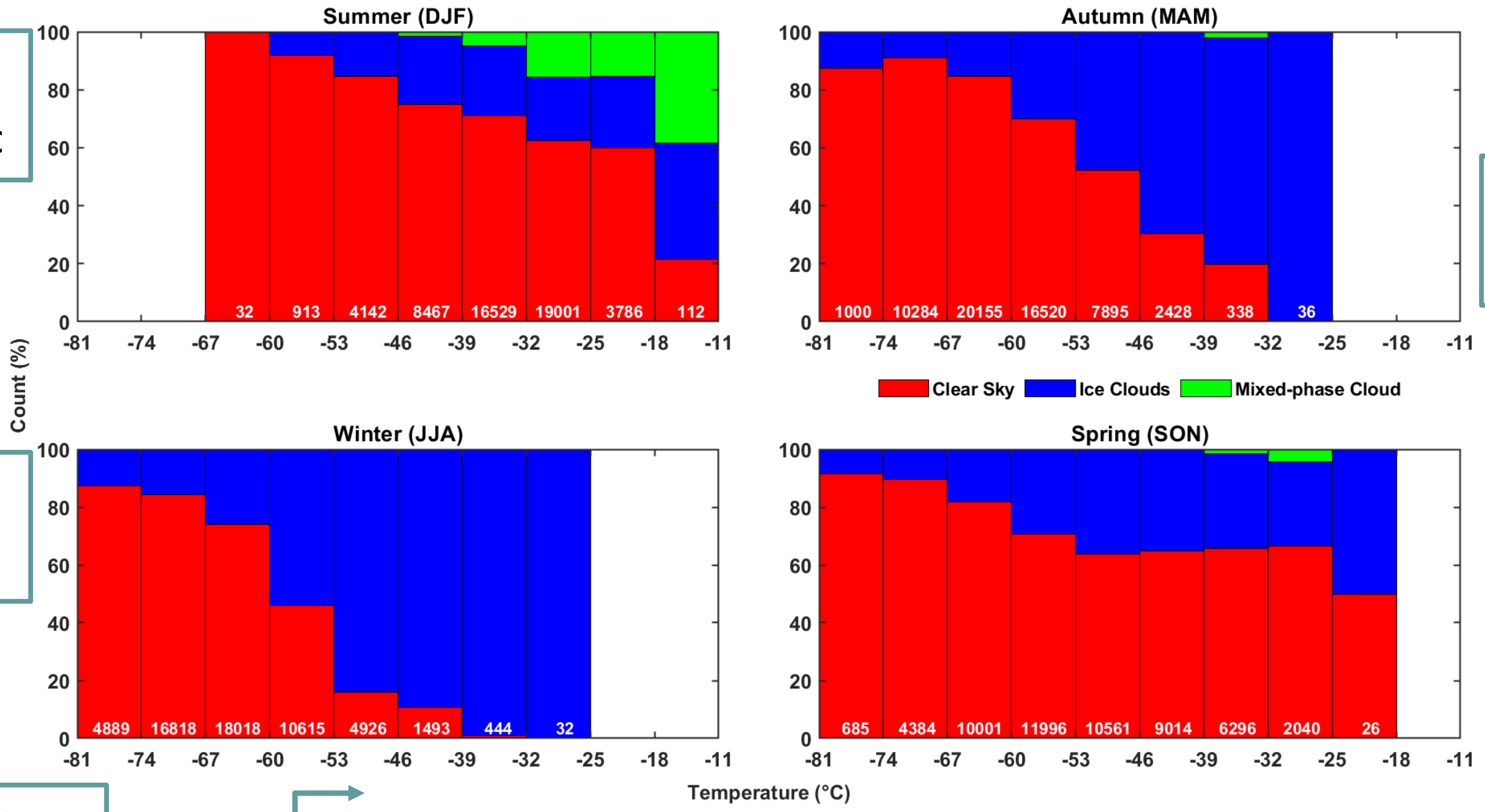
Year	Clear-sky (%)	Ice cloud (%)	Mixed-phase cloud (%)	Observation Time (%)
2012	56.07 ± 4.70	38.25 ± 0.66	5.68 ± 0.92	27.38
2013	71.01 ± 5.96	27.51 ± 0.47	1.48 ± 0.24	29.61
2014	69.97 ± 5.87	27.97 ± 0.48	2.06 ± 0.33	67.36
2015	67.46 ± 5.66	30.70 ± 0.53	1.84 ± 0.30	65.27
2016	67.50 ± 5.66	30.55 ± 0.53	1.96 ± 0.32	82.32
2017	70.42 ± 5.90	27.83 ± 0.48	1.74 ± 0.28	79.86
2018	72.48 ± 6.08	24.35 ± 0.42	3.17 ± 0.51	87.14
2019	69.52 ± 5.83	28.24 ± 0.49	2.25 ± 0.36	90.75
2020	76.39 ± 6.41	22.13 ± 0.38	1.48 ± 0.24	90.57
Total	70.09 ± 5.88	27.70 ± 0.48	2.25 ± 0.36	69.11

Almost **70%** of the full dataset is composed of **clear-sky** elements, the rest 30% is divided between **ice clouds (almost 28%) and mixed-phase clouds (2%)**.

Correlation with Surface Temperature

The detected cloudy sky occurrence increases (clear skies decrease) as surface air temperature increases (except that for the spring -SON)

$\langle T_{clear} \rangle$ = -35.64°C
$\langle T_{cloud} \rangle$ = -32.79°C
$\langle T_{all-sky} \rangle$ = -34.76°C



$\langle T_{clear} \rangle$ = -62.04°C
$\langle T_{cloud} \rangle$ = -55.41°C
$\langle T_{all-sky} \rangle$ = -60.36°C

$\langle T_{clear} \rangle$ = -66.62°C
$\langle T_{cloud} \rangle$ = -57.97°C
$\langle T_{all-sky} \rangle$ = -63.65°C

$\langle T_{clear} \rangle$ = -53.44°C
$\langle T_{cloud} \rangle$ = -49.41°C
$\langle T_{all-sky} \rangle$ = -52.29°C

$\langle T_{cloud} \rangle$ - $\langle T_{clear} \rangle$ = 8.65°C
$\langle T_{all-sky} \rangle$ - $\langle T_{clear} \rangle$ = 2.97°C

REFIR-PAD vs MODIS cloud occurrence

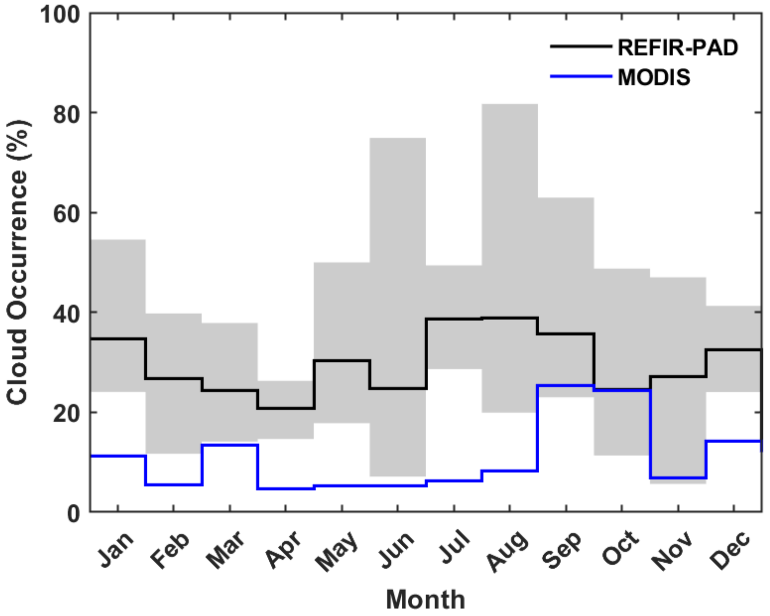
In collaboration with the
University of Wisconsin-Madison
Dr. R.E. Holz
Dr. P. Veglio

MODIS data available from **2012 to 2020**.
Types of collocation:

→ **Spatial collocation**: Dome-C is inside MODIS field of view (max distance *1km*). MODIS cloud mask on Dome-C for the same time interval (2012-2015) is compared to CIC monthly mean statistic.
→ **Spatial-temporal collocation**: each MODIS observation corresponds to a REFIR-PAD measurement (satellite passing maximum *15 minutes* before the observation time of the REFIR-PAD)

Statistical comparison

MODIS **spatially** collocated data: 2052
Nadir observations (zenith angle < 8°)



One-to-one comparison

MODIS **spatially and temporally** collocated data: 1118
Nadir observations (zenith angle < 8°)

Total		**Clear-sky**	**Ice cloud**	**Mixed-phase cloud**
1118	**REFIR-PAD (CIC)**	766 (68.5%)	337 (30.1%)	10 (0.9%)
	MODIS	1002 (89.6%)	108 (9.7%)	3 (0.3%)

More than 70% of not identified clouds by MODIS occur in winter months (March to August).

*MODIS products are derived from **MOD35** cloud mask*

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

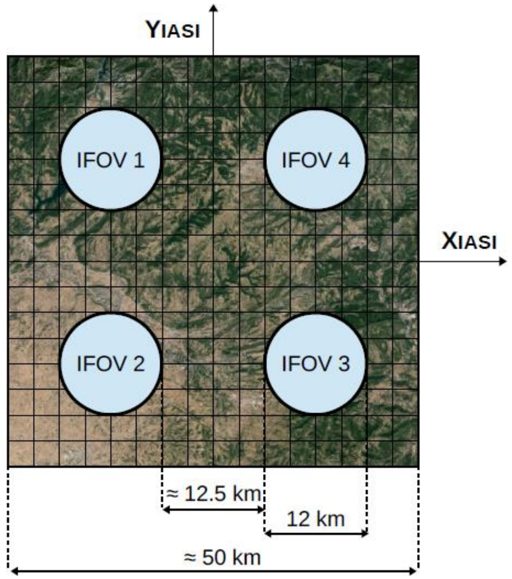
REFIR-PAD spectra vs IASI cloud occurrence

IASI data available from **2012 to 2015**

Spatial-temporal collocation for L1 and L2 IASI products:

- the IASI measurement is carried out within **15 minutes** off the observation time of the REFIR-PAD
- the maximum distance of the IASI pixel center from the REFIR-PAD instrument is less than **6 km** (the ground observation is within the IASI IFOV)

A filter for satellite **zenith angles below 6.7°** has been introduced to avoid geometric distortions (only "almost-nadir observations").



Total		**Clear-sky**	**Ice cloud**	**Mixed-phase cloud**	**Unclass**
167	**REFIR-PAD (CIC)**	117 (70.06%)	44 (26.35 %)	6 (3.59 %)	-
	IASI	43 (25.75 %)	121 (72.45 %)	3 (1.80 %)	-

IASI spatially and temporally collocated data: 167

Comparison of MYD06 and IASI L2 cloud products

Is the satellite footprint size the cause for the different cloud occurrence statistic between REFIR-PAD (or MODIS) and IASI?

Total		Clear-sky	Ice cloud	Mixed-phase cloud	Unclass
167	REFIR-PAD (CIC)	117 (70.06%)	44 (26.35 %)	6 (3.59 %)	-
	IASI	43 (25.75 %)	121 (72.45 %)	3 (1.80 %)	-
40	REFIR-PAD (CIC)	28 (70%)	9 (22.5%)	3 (7.5%)	-
	MODIS AQUA (1 km)	35 (87.5%)	4 (10%)	0	1 (2.5%)
	CLOUDSAT-CALIPSO	38 (95%)	0 (0%)	2 (5%)	
	MODIS AQUA (12 km) (CF>0)	**31 (77.5%)**	**8 (20%)**	**0**	1 (2.5%)

NOTES:

- CIC: Highly sensible to thin clouds
- IASI: extremely clear conservative
- CALIOP: Low efficiency to detect thin clouds close to the ice surface
- CPR: Unable to detect low level and thin cirri (vertical resolution = 500 m)
- MODIS: Low efficiency at IR – Dependent on solar zenith angle

Main Conclusions

- **L3 products** are representative of grid areas. Monthly means L3 cloud occurrence (fraction) show significant differences with respect to ground-based derived products likely due to the large dimension of the grid and to detection performances (seasonal dependent).

- **Long records from ground-based measurements are required to test L2 satellite products** when strict collocation constrains are set.

- **Significant differences** in cloud occurrences are found among multiple satellite sensors both passive (i.e. MODIS vs IASI) or active (CPR and CALIOP) and ground based derived products.

- The main cause affecting the satellite detection performances seems to be the difficulties in identifying of **low level thin ice layers**. The CIC algorithm is sensible to very thin ice layers and diamond dust

- CIC is a reliable tool for cloud identification and classification, demonstrated by 95% of REFIR-PAD spectra correctly classified**.**

- **When CIC is applied to IASI spectra, 65% of scenes match the ground observations** (64% for clear-sky and 72% for cloudy sky). On the limited dataset, the IASI **cloud occurrence** is 74% for IASI L2 product and **37%** for CIC classification.

- An improvement in cloud identification from passive sensors is expected by exploiting the FIR part of the spectrium (FORUM)

References and Documentations

Main References

Cossich, W., Maestri, T., Magurno, D., Martinazzo, M., Di Natale, G., Palchetti, L., Bianchini, G., and Del Guasta, M.: Ice and Mixed-Phase Cloud Statistics on Antarctic Plateau, Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2021-97, in review, 2021.

Di Natale, G.; Bianchini, G.; Del Guasta, M.; Ridolfi, M.; Maestri, T.; Cossich, W.; Magurno, D.; Palchetti, L. Characterization of the Far Infrared Properties and Radiative Forcing of Antarctic Ice and Water Clouds Exploiting the Spectrometer-LiDAR Synergy. *Remote Sens.* 2020, *12 (21)*, 3574. https://doi.org/10.3390/rs12213574

Maestri, T., Arosio, C., Rizzi, R., Palchetti, L., Bianchini, G., & Del Guasta, M. (2019). Antarctic ice cloud identification and properties using downwelling spectral radiance from 100 to 1400 cm^{-1}. Journal of Geophysical Research: Atmospheres, 124. https://doi.org/10.1029/2018JD029205

Data Documentation (L2)

MODIS: https://atmosphere-imager.gsfc.nasa.gov/sites/default/files/ModAtmo/MOD06-ATBD_2015_05_01_2.pdf

CLOUDSAT-CALIPSO: CloudSat 2B-CLDCLASS-LIDAR Product Process Description and Interface Control Document (2019)

IASI: IASI Level 2: Product Generation Specification, EPS.SYS.SPE.990013 v8E e-signed (2017)