ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Antarctic cloud detection and classification from far and mid infrared downwelling radiance spectra: performances optimization and results

Davide Magurno[1], Tiziano Maestri[1], William Cossich[1], Gianluca di Natale[2], Luca Palchetti[2], Giovanni Bianchini[2], Massimo Del Guasta[2]

(1) University of Bologna, DIFA Department of Physics and Astronomy, Bologna, Italy;

(2) Istituto Nazionale di Ottica, CNR, Via Nello Carrara, 1, Sesto Fiorentino, Italy.

EGU2020: Sharing Geoscience Online 4–8 May 2020

Session AS5.14 «Remote Sensing of Clouds and Aerosols: Techniques and Applications»

Introduction

The machine learning Cloud Identification and Classification algorithm (CIC, Maestri et al., 2019) is applied to spectra collected by the Radiation Explorer in the Far InfraRed-Prototype for Applications and Development (REFIR-PAD, Palchetti et al., 2015) based at Dome C, Antarctic Plateau.

The CIC algorithm is trained on a sample of REFIR-PAD spectra (test set) comprising downwelling radiances collected in clear conditions or in presence of ice and water clouds. CIC is then applied to classify a larger REFIR-PAD dataset of about 90,000 spectra collected from 2012 to 2015.

Two main goals are pursued:

- To provide a solid statistic of clear sky, liquid water clouds and ice clouds occurrence on the Antarctic Plateau. Monthly and seasonal mean occurrences are also provided.
- To evaluate the information content within the observed spectra by applying CIC to multiple spectral intervals covering the mid- and far-infrared (MIR and FIR) part of the spectrum.

The present work contributes to the preparatory studies for the Far-infrared Outgoing Radiation Understanding and Monitoring (FORUM) mission that has been selected as ESA's 9th Earth Explorer mission.

REFIR-PAD@DomeC

The dataset comprise spectral measurements collected at the Antarctica Concordia station by the **Radiation Explorer in the Far-InfraRed – Prototype for Applications and Development (REFIR-PAD)**

Four years period analysed data (time resolution ~14 min):

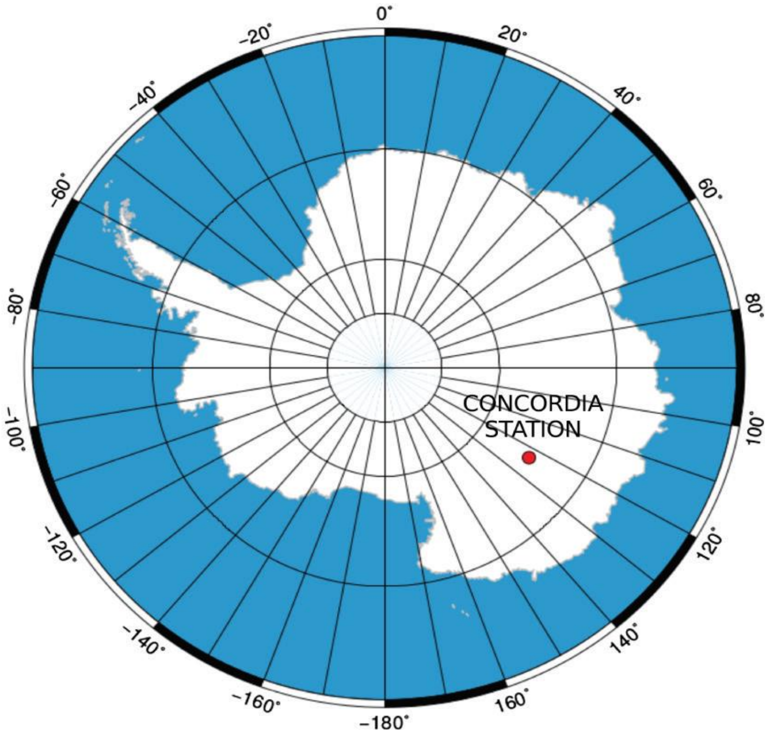
2012: 16177 spectra (155.2 to 1080.0 cm^{-1} ; resolution 0.4 cm^{-1})

2013: 19298 spectra (100.4 to 1499.6 cm^{-1} ; resolution 0.4 cm^{-1})

2014: 25089 spectra (155.2 to 1080.0 cm^{-1} ; resolution 0.4 cm^{-1})

2015: 27396 spectra (100.4 to 2499.6 cm^{-1} ; resolution 0.4 cm^{-1})

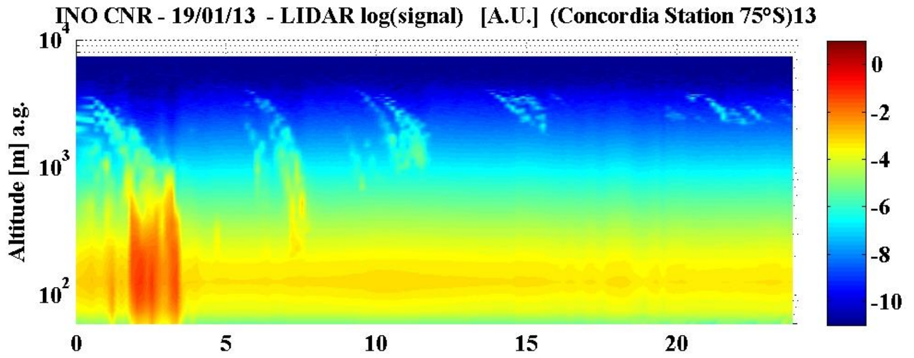
TOTAL: 87960 spectra

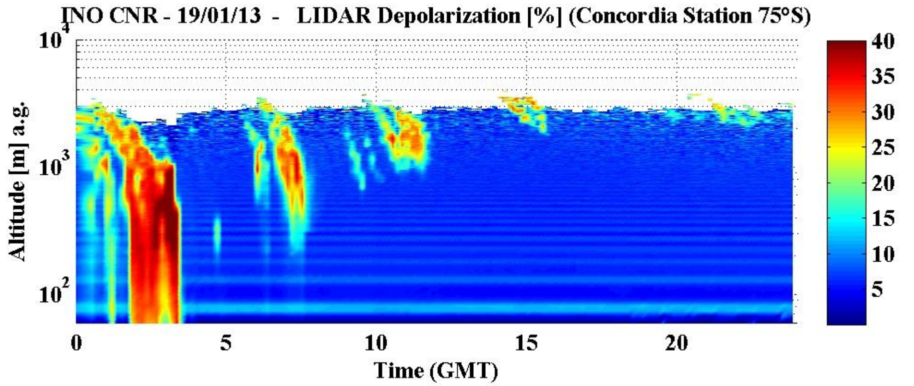


ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

LIDAR@DomeC

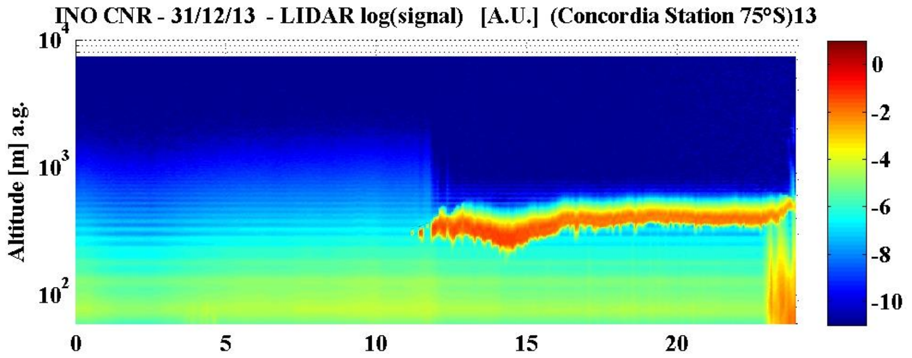
Backscatter and depolarization profiles derived from a tropospheric **LIDAR system** (λ = 532 nm, vertical resolution = 7.5 m, time resolution = ~13 min; Ricaud et al., 2017) are **used for defining the training sets and for results verification**.

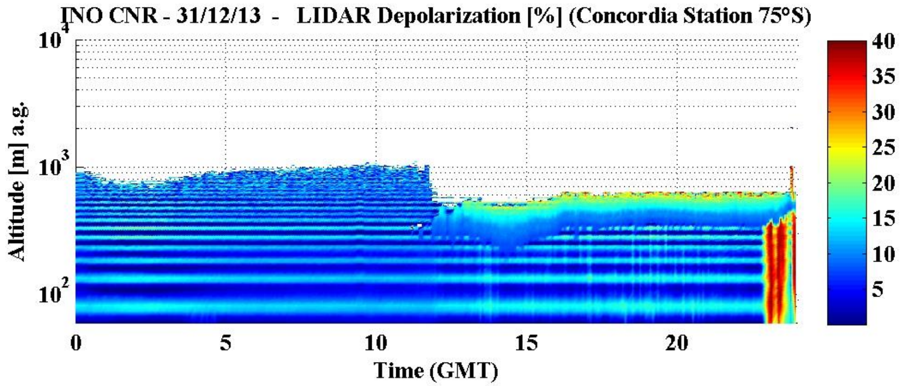




Ice clouds: high reflectivity and high depolarization

Liquid water/Mixed phase clouds: high reflectivity and low depolarization



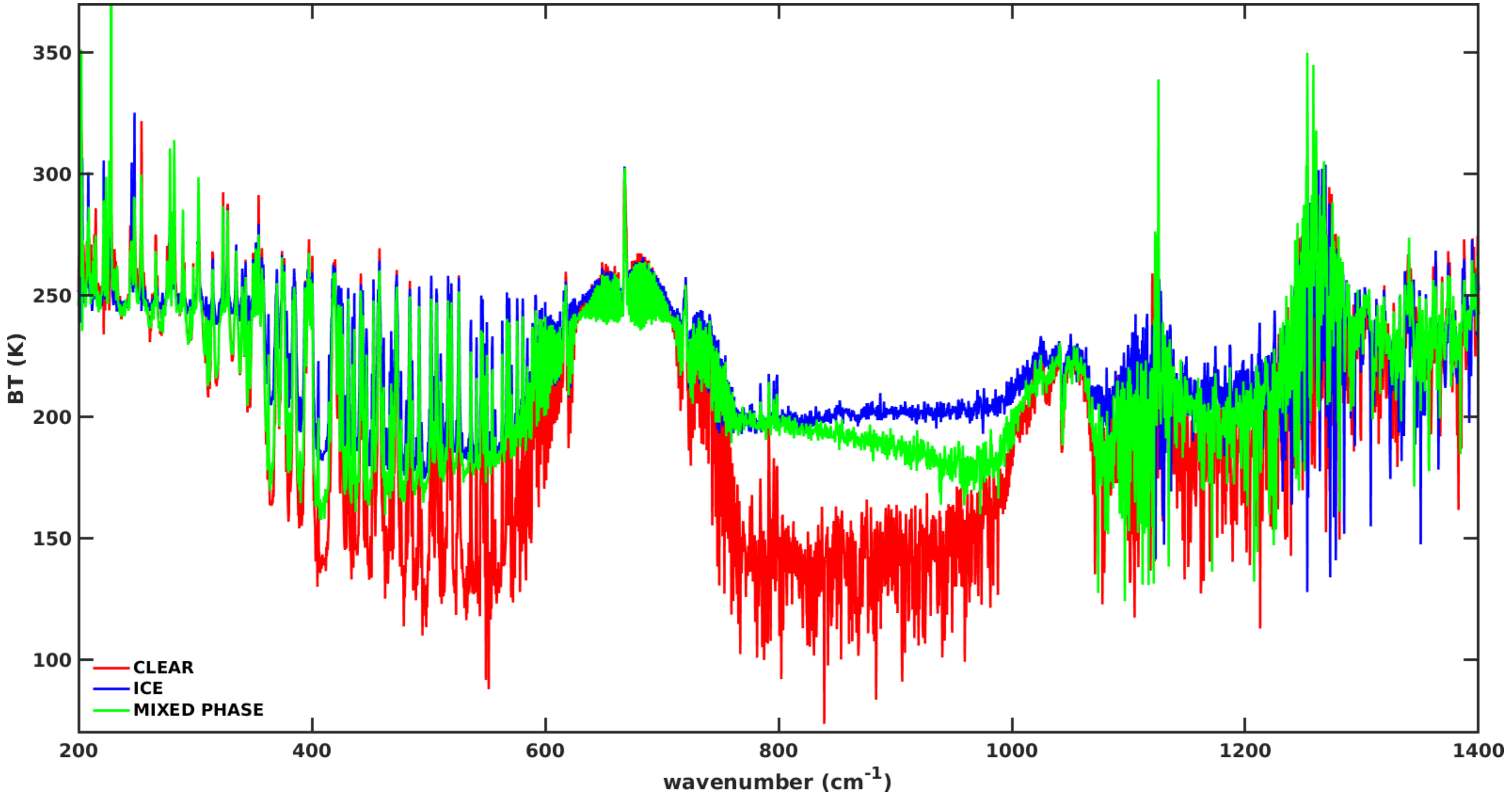


Training Sets

A **machine learning algorithm** performs a **principal component analysis** (PCA) of the analysed spectra to **classify the scene** (see Maestri et al., 2019 for details).

2013 LIDAR data are used to select **representative spectra for the three classes** of interest (**Clear sky, Ice cloud, Liquid/Mixed phase cloud**) and to build proper **training sets (TS)** for the algorithm.

BT spectra for three representative elements of the selected training sets



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

The test set

The **algorithm capabilities** are tested on **1726 spectra (test set)** for which we **defined** the **membership a-priori** (clear sky, ice cloud, mixed phase cloud) by visual inspection of the LIDAR data.

The CIC algorithm is applied to the test set (multiple **spectral intervals of applications** are used). **Results** are evaluated in terms of the **threat score**.

Threat Score: Statistical index to evaluate the success of a multi-class classification algorithm.

TP = True Positive
FP = False Positive
FN = False Negative

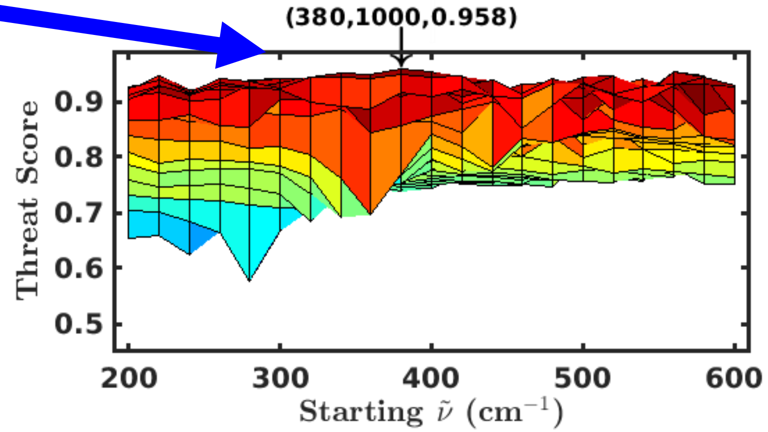
$$\text{Threat Score} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}$$

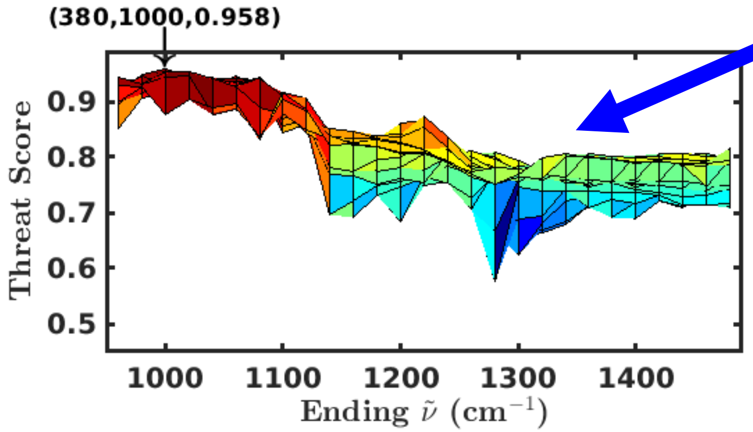
Spectral interval optimization

Highest threat score: 0.958 → Optimal spectral interval: 380-1000 cm^{-1}

Exploiting part of the far-infrared spectrum allows high CIC performances

(starting ν, ending ν, threat score)

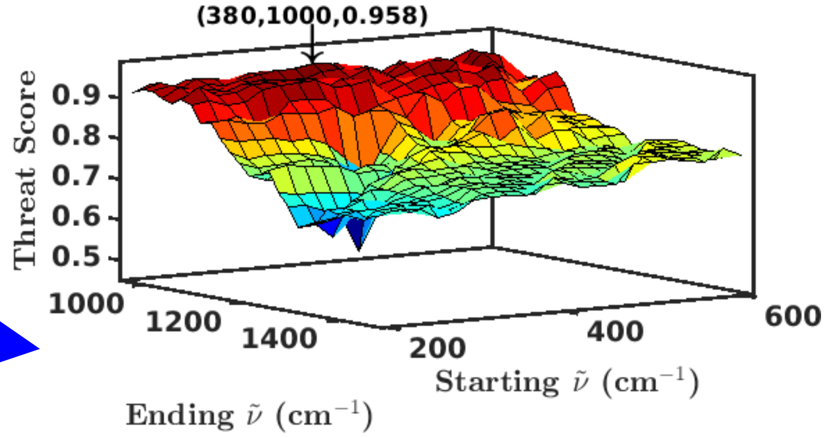




Highest wavenumbers reduce the CIC performances due to the instrument characteristics.

The algorithm is calibrated in accordance with the sensors characteristics.

The CIC applied on the test set in the spectral interval from «starting ν» to «ending ν» gives a specific «threat score»



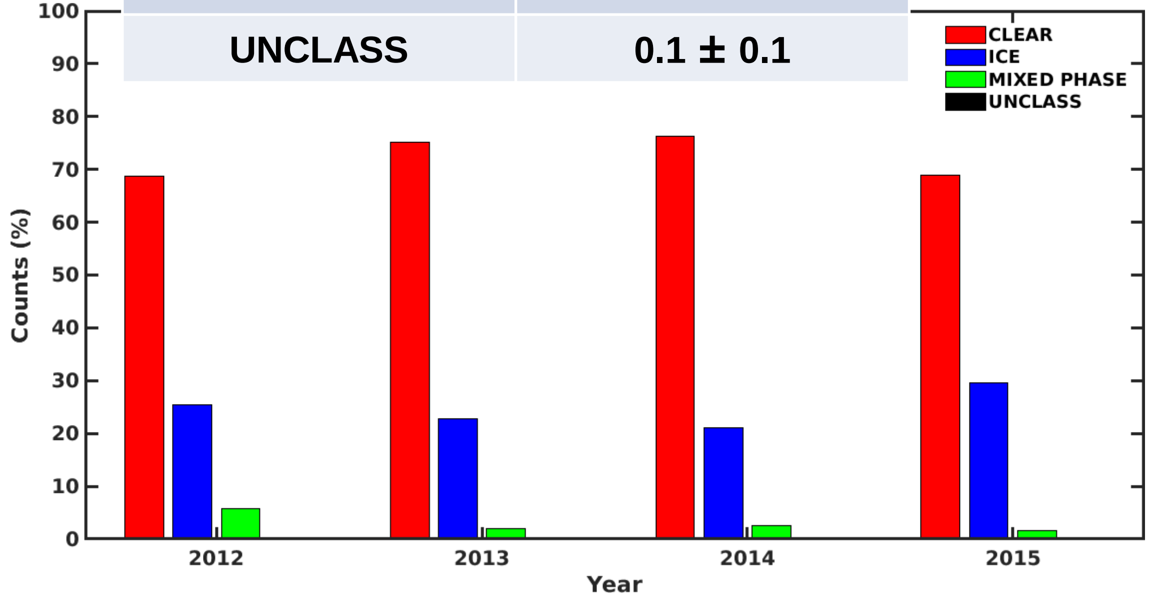
Test set: evaluation of performances

TEST SET CLASS	# SPECTRA	CIC RESULTS	THREAT SCORE
CLEAR	**559**	**98.0% - CLEAR** 2.0% - ICE 0.0% - MIXED PHASE	**0.96**
ICE	**1022**	**98.7% - ICE** 0.9% - CLEAR 0.4% - MIXED PHASE	**0.97**
LIQUID/MIXED PHASE	**145**	**91.0% - MIXED PHASE** 8.3% - ICE 0.7% - CLEAR	**0.89**
TOTAL	**1726**	**97.9% - CORRECT** 2.1% - MISCLASSIFIED	**weighted mean** **0.96**

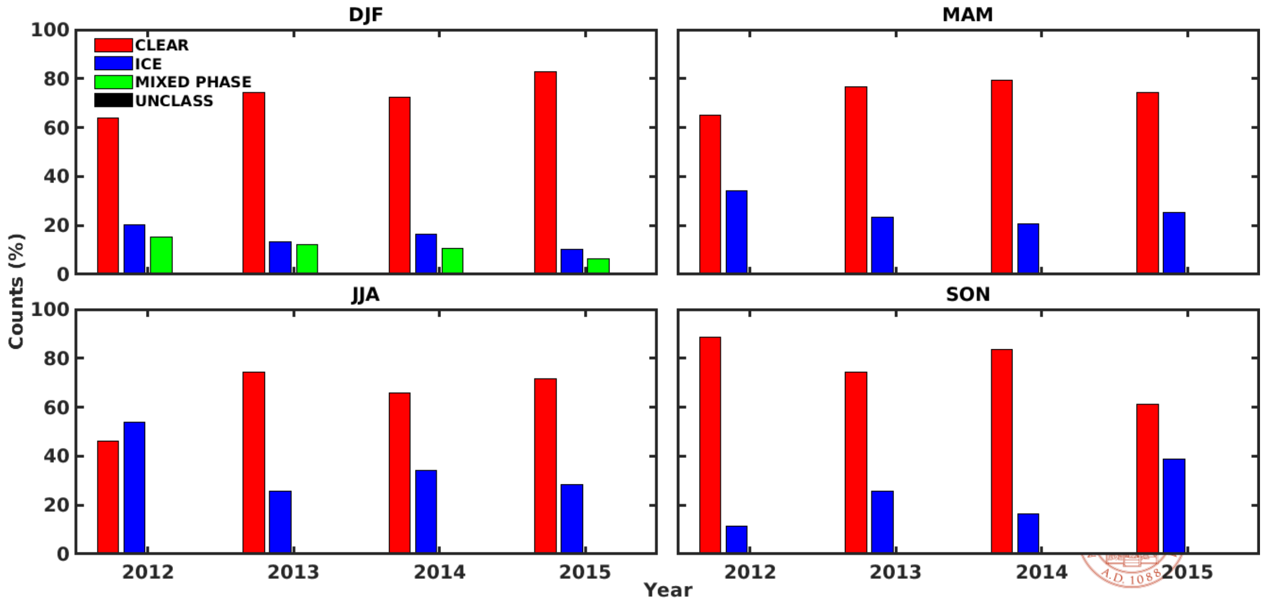
The full dataset

The CIC algorithm is applied to the **entire available dataset (87758 spectra, 4 years)** with the optimal selected interval (380-1000 cm^{-1}). The majority of the observations (>70%) are in clear sky conditions; liquid/mixed phase clouds are observed mainly in summer season.

CIC CLASSIFICATION	TOTAL %
CLEAR	72.3 ± 2.8
ICE	24.9 ± 0.9
MIXED PHASE	2.7 ± 0.3
UNCLASS	0.1 ± 0.1

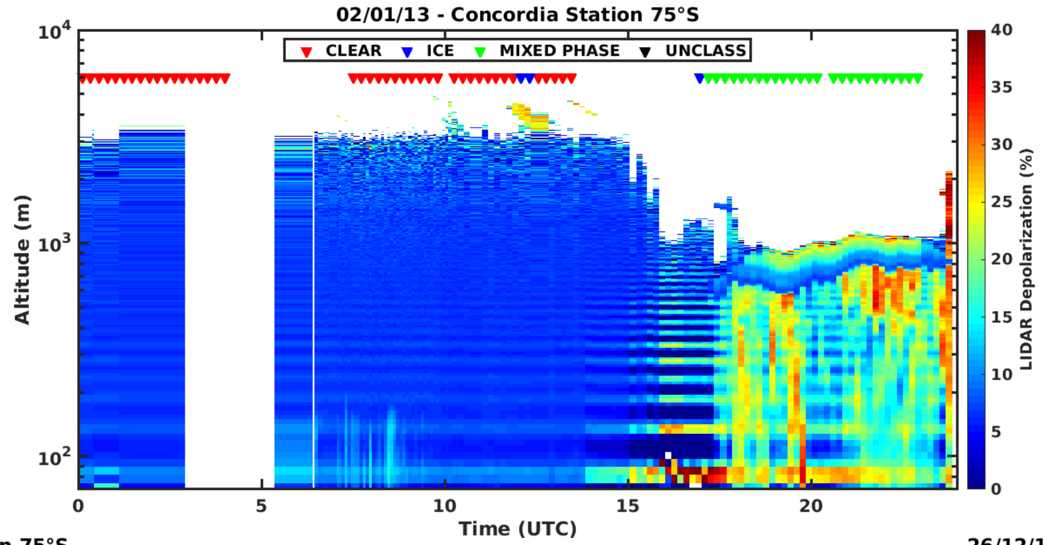


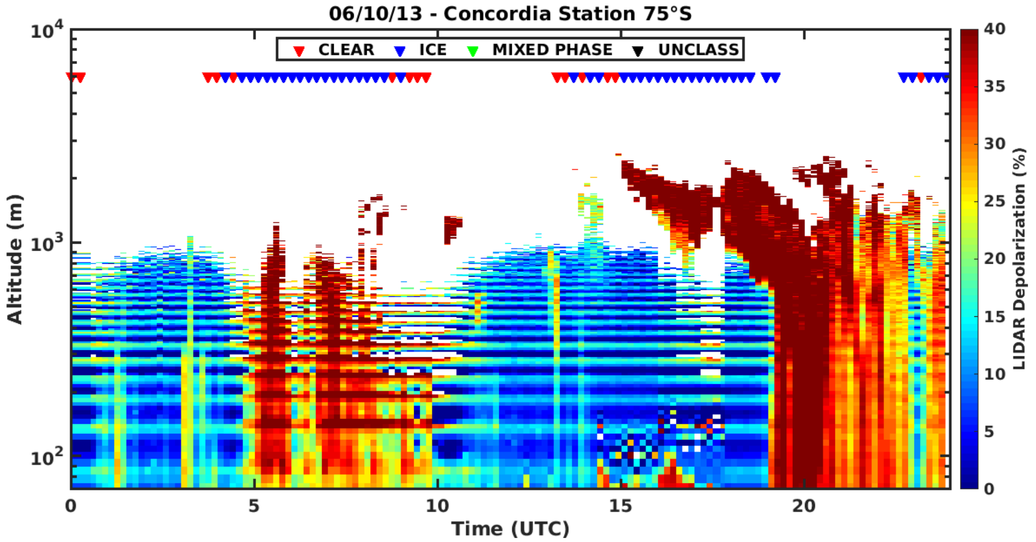
Seasonal occurrences

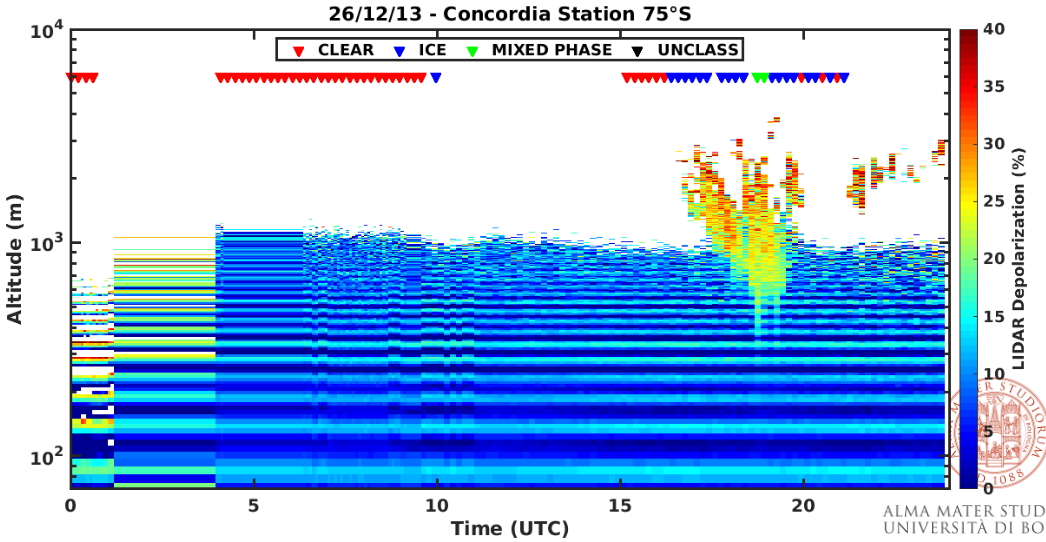


Example of CIC results on LIDAR data

Triangles on top show the available REFIR-PAD spectra. The CIC classification is color coded.







Conclusions

- The CIC algorithm is applied to the entire four year REFIR-PAD dataset. Results show that:
  - clear sky conditions occurrence is of the order of ~72%, with limited variations within the 4 years of observations
  - Ice clouds have their maximum occurrence during the JJA season
  - Liquid/Mixed phase clouds are observed mainly during summer time
- The CIC performances span from 58% to 96% in accordance with the selected spectral interval. The algorithm optimization selected the interval 380-1000 cm^{-1} as the best performing one.
- The result highlights the importance of a proper algorithm calibration in accordance with the charcateristics of the analysed sensor and the fundamental role of the FIR part of the spectrum in cloud identification.

References:

Maestri et al. 2019, Atmospheric Measurement Techniques, 12, 3521
Palchetti et al. 2015, Bulletin of the American Meteorological Society 96, 1505
Ricaud et al. 2017, Atmospheric Chemistry and Physics, 17, 5221

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Davide Magurno
University of Bologna, DIFA Department of Physics and Astronomy, Bologna, Italy
davide.magurno2@unibo.it