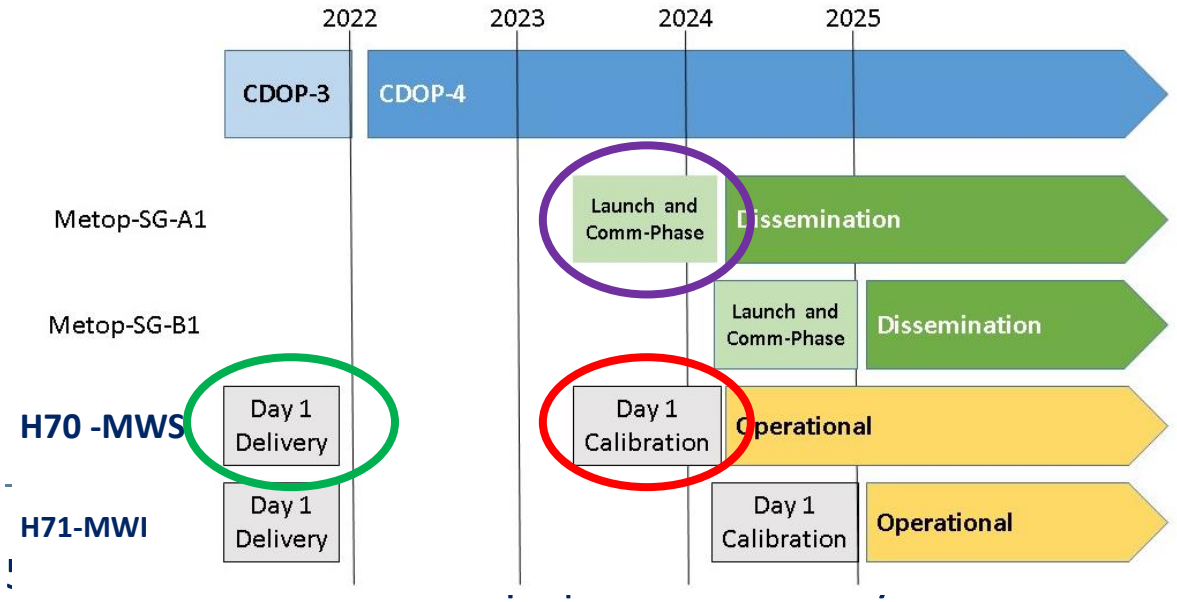
The H SAF EPS-SG MWS Day-1 Machine Learning algorithm for snowfall and rainfall surface precipitation rate retrieval: Intercomparison of Machine Learning Techniques and performance analysis

Paolo Sanò,
Daniele Casella, Andrea Camplani, Leo Pio D'Adderio, Giulia Panegrossi

Institute of Atmospheric Sciences and Climate, National Research Council of Italy (ISAC-CNR), Rome, Italy

H SAF P-IN-MWS (H70)
Precipitation Product by EPS-SG MWS

- **H70 product is a Level 2 (orbital) product providing instantaneous precipitation rate, on a global scale, from the EPS-SG MWS measured brightness temperatures.**
- **H70, designed as the Day 1 operational precipitation product for the Metop-SG-A series, includes different modules specifically designed for the detection and estimate of rainfall and snowfall.**
- **H70 was developed using the ATMS cross-track radiometer, similar in terms of channel frequencies and spatial resolution to MWS which is currently not operational.**
- **The H70 modules are based on machine learning approach to maximize the exploitation of the channels' information.**



H70 was developed within the H SAF CDOP3

Metop-SG-A1 launch and Commissionig Phase

Calibration and tuning of the H70 algorithm

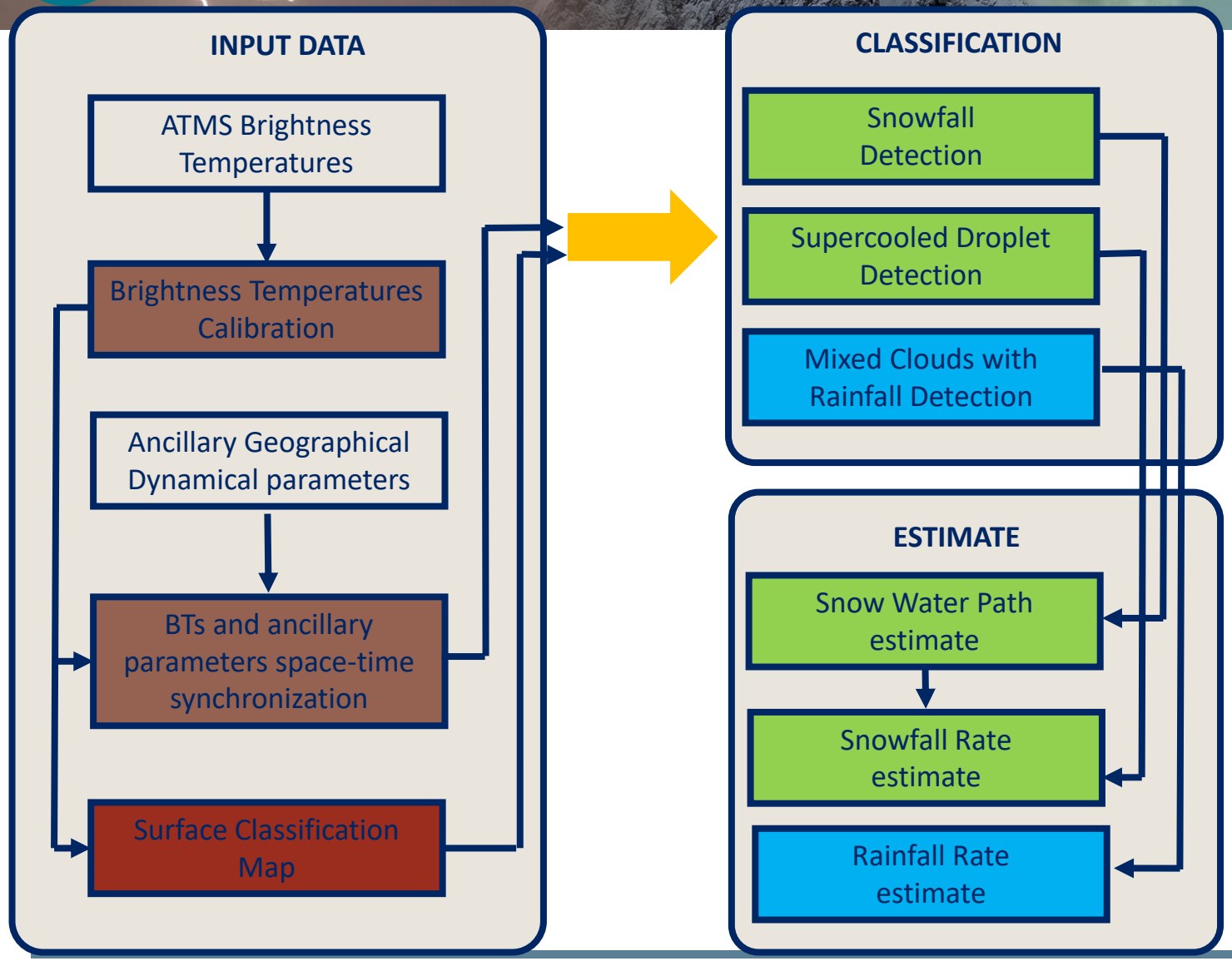
Characteristics of Advanced Technology Microwave Sounder (ATMS) and Micro-Wave Sounder (MWS) MW radiometers

- Both are cross-track scanning
- Similar channel frequencies
- Similar spatial resolution
- Some differences in polarization

For the H70 development the channels in common between MWS and ATMS were used.

N.	ATMS Central frequency (GHz)	Pol.	MWS Central frequency (GHz)	Pol.
	23.800	QV	23.8	QH
	31.400	QV	31.4	QH
	50.300	QH	50.3	QH/QV
	51.760	QH		
	52.800	QH	52.800	QH/QV
			53.246±0.08	QH/QV
	53.596 ± 0.115	QH	53.596±0.115	QH/QV
			53.948±0.081	QH/QV
	54.400	QH	54.4	QH/QV
	54.94	QH	54.94	QH/QV
	55.5	QH	55.5	QH/QV
	57.290344	QH	57.290344	QH/QV
	57.290344±0.217	QH	57.290344±0.217	QH/QV
	57.290344±0.3222±0.048	QH	57.290344±0.3222±0.048	QH/QV
	57.290344±0.3222±0.022	QH	57.290344±0.3222±0.022	QH/QV
	57.290344±0.3222±0.010	QH	57.290344±0.3222±0.010	QH/QV
	57.290344±0.3222±0.0045	QH	57.290344±0.3222±0.0045	QH/QV
	89.5	QV	89	QV
	165.5	QH	165	QH
	183.311±7.0	QH	183.311±7.0	QV
	183.311±4.5	QH	183.311±4.5	QV
	183.311±3.0	QH	183.311±3.0	QV
	183.311±1.8	QH	183.311±1.8	QV
	183.311±1.0	QH	183.311±1.0	QV
			229.0	QV

H70 Algorithm flowchart



- Random Forest
- Gradient Boosting
- Shallow Neural Network

→ Pixel Based

- Convolutional Deep Neural Network

→ Image Based

Reference Products for Rainfall and Snowfall for H70 development

Instrument	CloudSat CPR	GPM DPR	
		KaPR	KuPR
Launch time	18 Apr 2006	27 Feb 2014	27 Feb 2014
Altitude (km)	705	407	407
Inclination angle (°)	98.23	65	65
Frequency (GHz)	**94**	**35.547**	**13.603**
Horizon resolution at nadir (km)	1.4 × 1.7	5	5
Swath width (km)	1.4	120	245
Vertical resolution (m)	500	250/500	250
Minimum detectable Z_e (d BZ)	<–29	12 (KaHS) 18 (KaMS)	18
Measurement accuracy (dBZ)	<2.7	<±1	<±1

Both spaceborne radars were used as a reference for the development of the algorithm.

The GPM Dual-frequency Precipitation Radar (DPR) was used to develop the precipitation modules.

The Cloud Profiling *Radar* (*CPR*) which proved to be very effective in detecting /estimate snow was used for snowfall modules

Spaceborne radars:

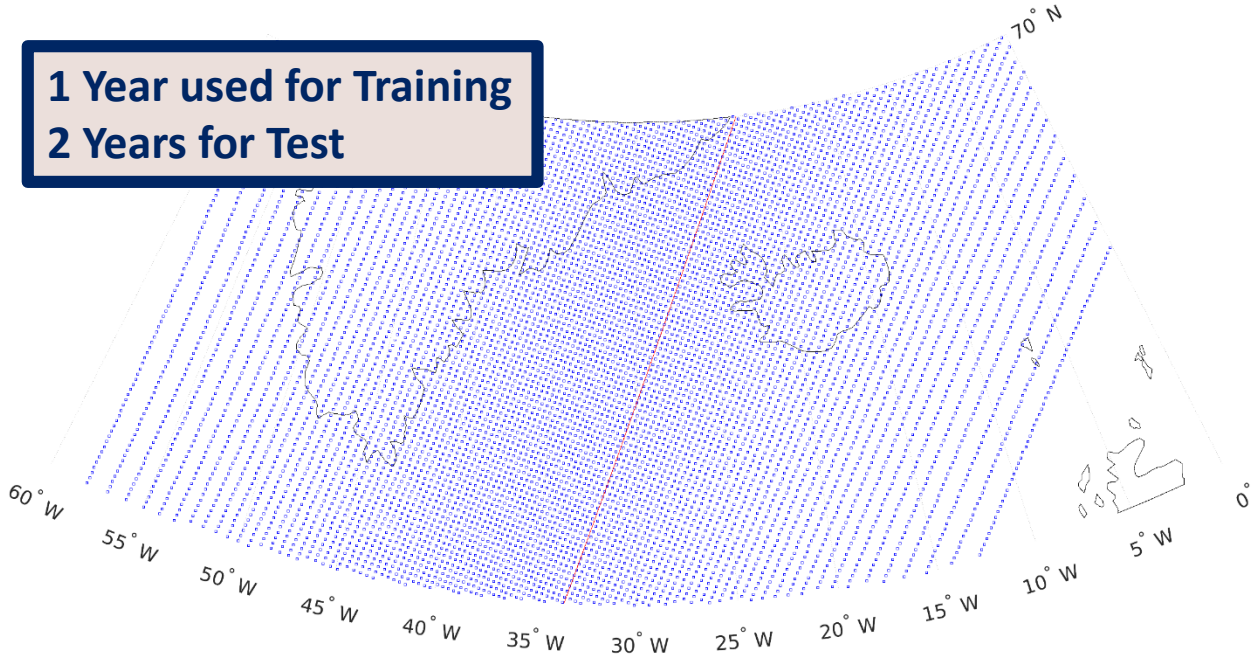
- ***GPM DPR (Ku/Ka band)***: better coverage (large swath but up to 65°N/S) *valuable for medium/heavy precipitation conditions*; low sensitivity hampers detection/quantification capabilities of light precipitation
- ***CloudSat CPR (W band)*** **:** provides by far the most complete view of snow systems (up to 82°N/S) thanks to high sensitivity

ATMS-DPR and ATMS-CPR Coincidence Datasets

ATMS-DPR dataset

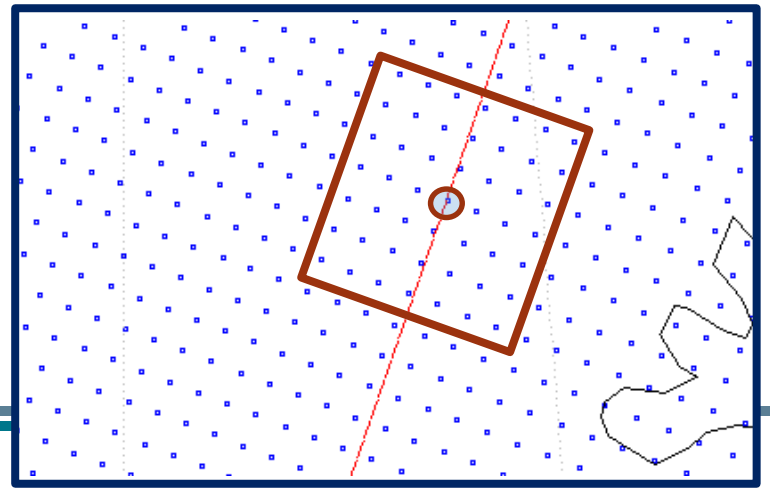
Period	2014–2016
Geographical area	Global
Number of database points	15 M
Number of database points with snowfall	2.2 M
Horizontal resolution (Km)	15.8 x 15.8 (nadir) 30 x 68.4 (scan edge)
Reference Rainfall product	2B-CMB level-2 GMI/DPR combined V06A on Ku-band radar swath (NS)

1 Year used for Training
2 Years for Test



ATMS-CPR dataset

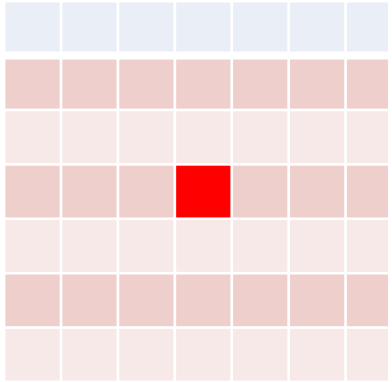
Period	2014 –2016
Geographical area	Global
Number of database points %	6.7 M
Number of database points with snowfall	1.1 M
Horizontal resolution (Km)	15.8 x 15.8 (nadir) 30 x 68.4 (scan edge)
Reference snowfall product	2C-SNOW-PROFILE (Cloudsat CPR derived products)



Input / Output Variables

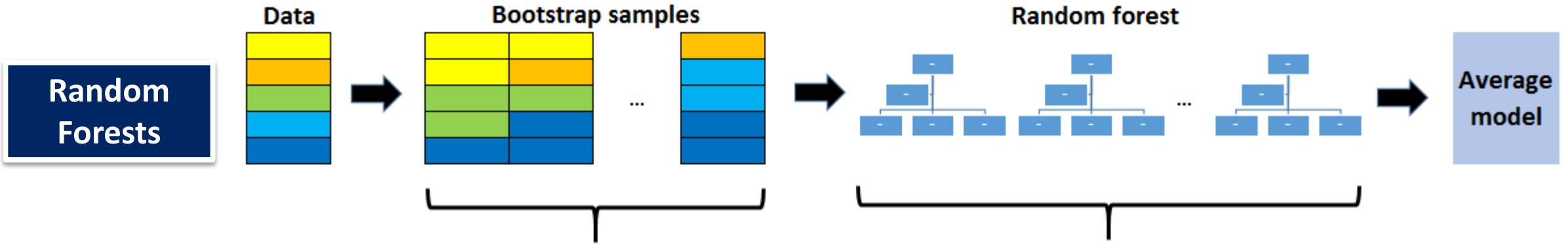
Features	Variable in the database	Data source
1-16	ATMS BTs	NOAA
17	ATMS Scan angle	NOAA
18	PESCA Surface Class	Derived
19	Surface height	DEM
20	2 m temperature	ECMWF-Operational
21	Total column integrated water vapor	ECMWF-Operational
22	Freezing level Height	ECMWF-Operational
23-26	PCA 1-4 Temperature profile	ECMWF-Operational
27-30	PCA 1-4 Relative humidity profile	ECMWF-Operational
31-34	PCA 1-4 Absolute humidity profile	ECMWF-Operational
Target		
1	Supercooled Droplets Fraction	DARDAR (raDAR/liDAR) LATMOS-Reading Univ.
2	Snowfall Rate	2C-SNOW-PROFILE (Cloudsat CPR derived products)
3	Snow Water Path	2C-SNOW-PROFILE (Cloudsat CPR derived products)
Target		
1	Rainfall Rate	2B-CMB Ku (GPM DPR product)

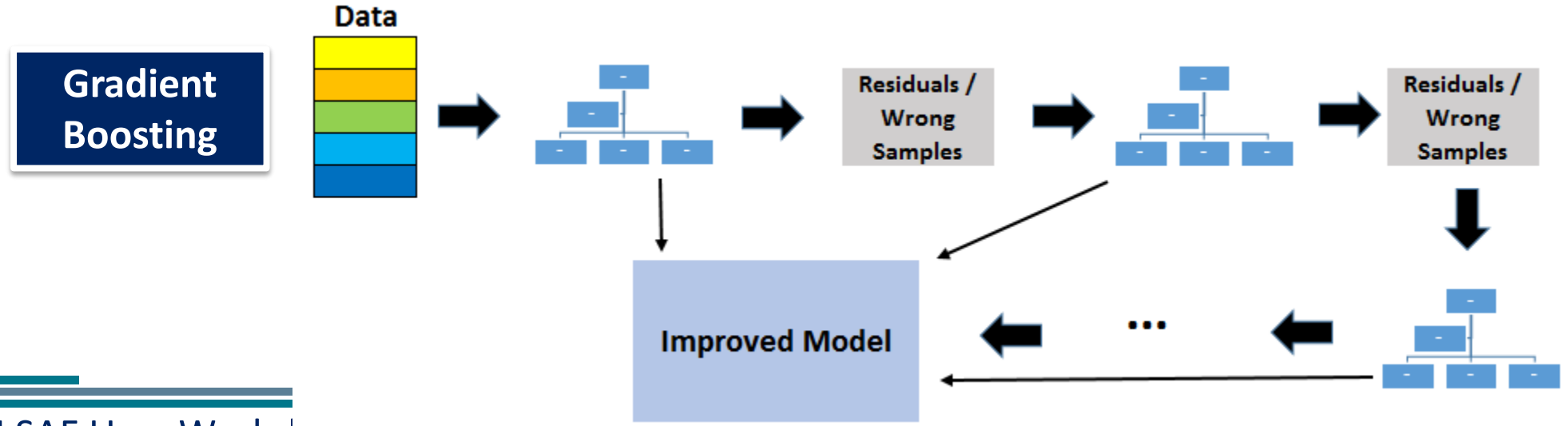
Input N	Central Freq GHz	Pol.
1	23.800	QV
2	31.400	QV
3	50.300	QH
4	51.760	QH
5	52.800	QH
6	53.596 ± 0.115	QH
7	54.400	QH
8	54.94	QH
9	55.5	QH
10	89.5	QV
11	165.5	QH
12	183.311±7.0	QH
13	183.311±4.5	QH
14	183.311±3.0	QH
15	183.311±1.8	QH
16	183.311±1.0	QH



	Pixel Based	Image Based
Features	Central Pixel	7x7 Matrix
Target	Central Pixel	Central Pixel

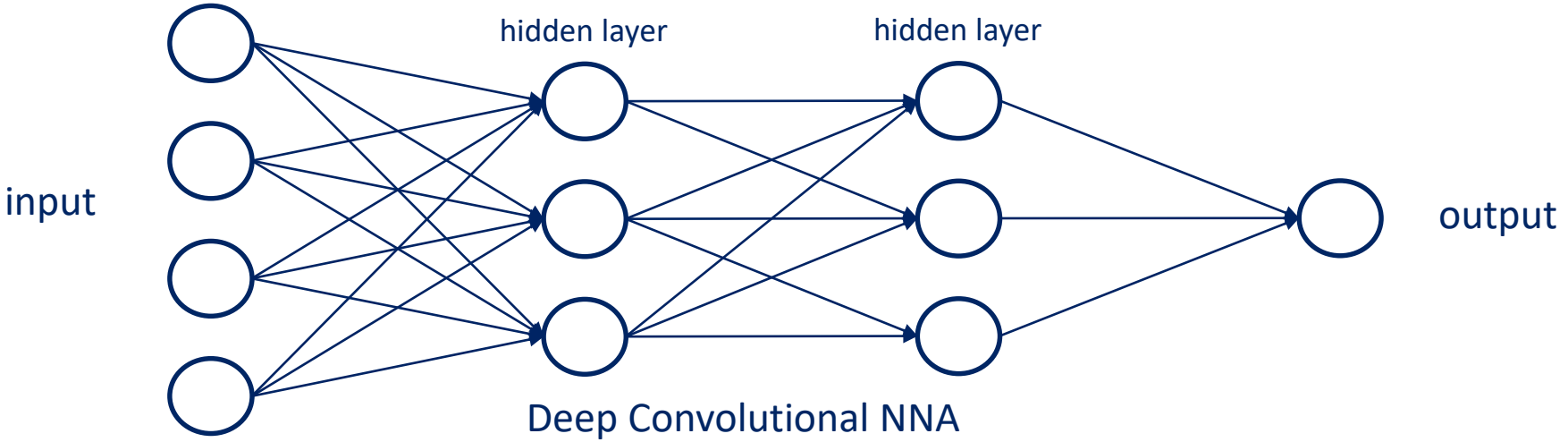
Machine Learning Approches: Random Forests and Gradient Boosting



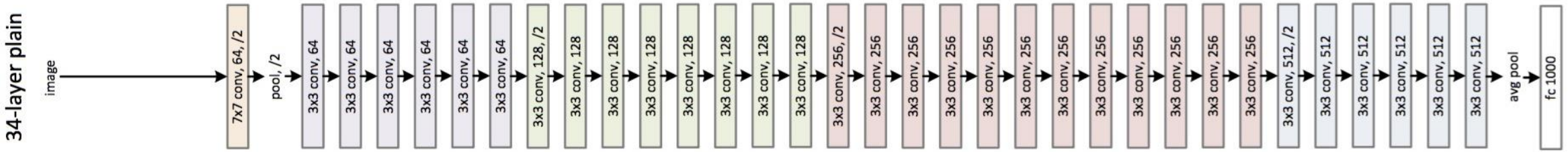


Machine Learning Approches: Shallow vs Deep Neural Networks

Shallow Feed Forward NNA



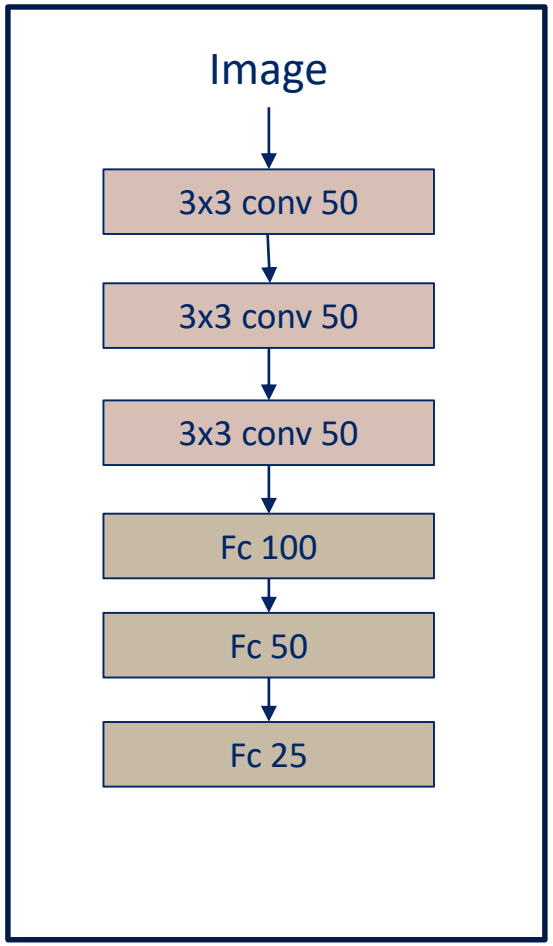
Deep Convolutional NNA



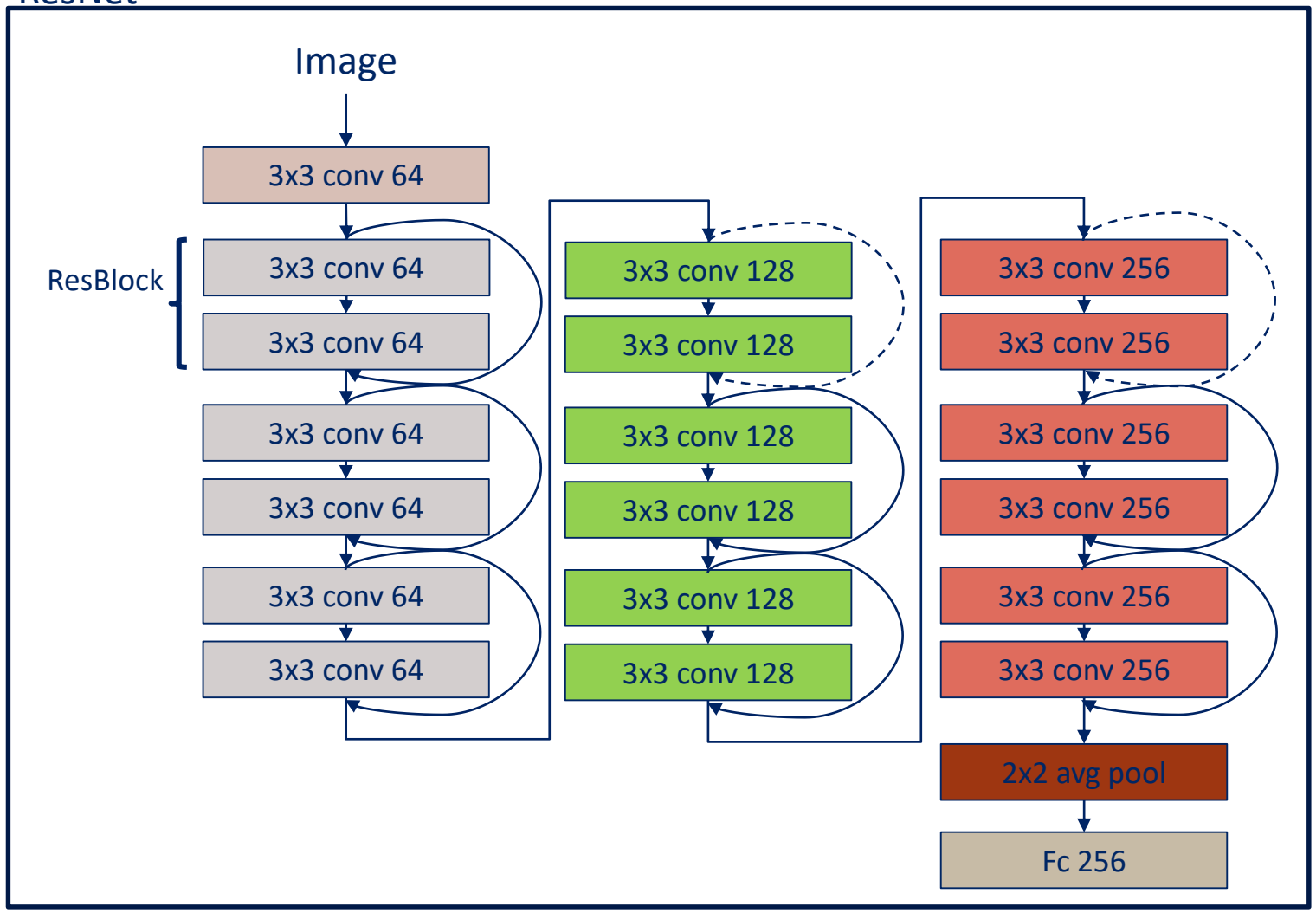
- Main Differences:
  - In Deep Learning Convolutional NN only spatially contiguous inputs are connected
  - N° of Levels and of parameters

Tested Deep Architectures

VGG



ResNet



ML Models Intercomparison – test results

$$R^2 = 1 - \frac{MSE}{VAR}$$

Perfect: $R^2 = 1$

Baseline: $R^2 = 0$

Event forecast	Event observed: Yes	Event observed: No
Yes	a	b
No	c	d

Heidke Skill Score:

fractional improvement of the forecast over the standard forecast.

$$HSS = \frac{2(ad-bc)}{[(a+c)(c+d)+(a+b)(b+d)]}$$

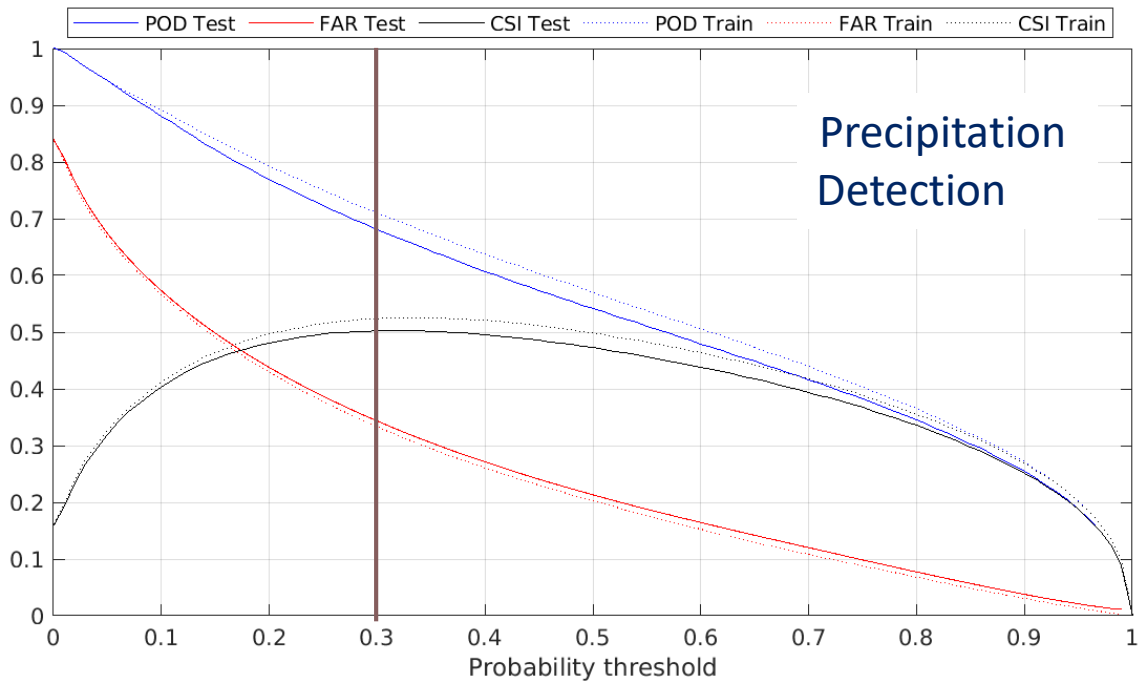
Perfect: HSS =1

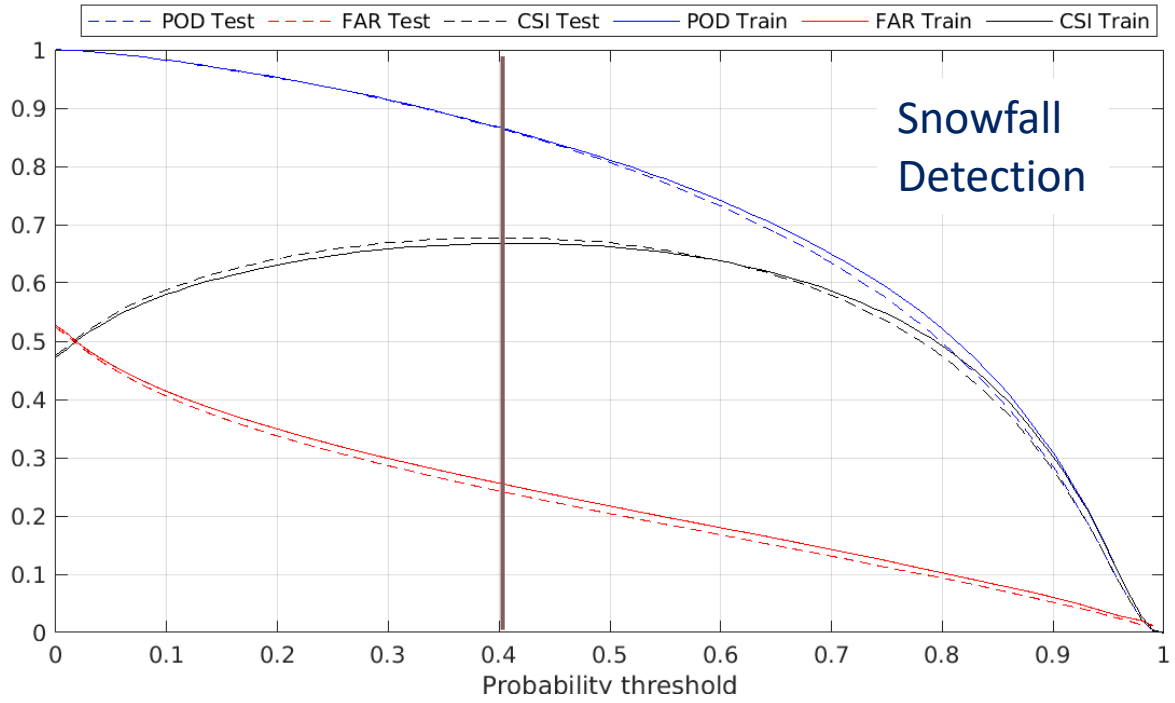
No Skill: HSS=0

	RandomForest	GradientBoosting	ShallowNN	VGG	ResNet
RMSE [kg/m²]	0.078	0.090	0.050	0.055	0.072
R^2	0.955	0.940	0.981	0.978	0.961
ME [kg/m²]	-3.66 E-03	-1.08E-02	-1.59E-05	-5.61E-05	-1.2E-03
Corr	0.86	0.83	0.93	0.92	0.87
Number of parameters	4.00E+06	1.00E+04	3.13E+03	7.16E+04	4.34E+06

	RandomForest	RobustBoost	AdaBoost	ShallowNN	VGG	ResNet
HSS	0.62	0.61	0.61	0.66	0.68	0.64
CSI	0.67	0.66	0.66	0.69	0.70	0.67
POD	0.80	0.79	0.79	0.83	0.83	0.80
FAR	0.20	0.20	0.20	0.19	0.18	0.19
Number of parameters	2.11E+06	1.40E+04	2.20E+04	4.05E+03	5.52E+04	2.75E+05

Detection Modules

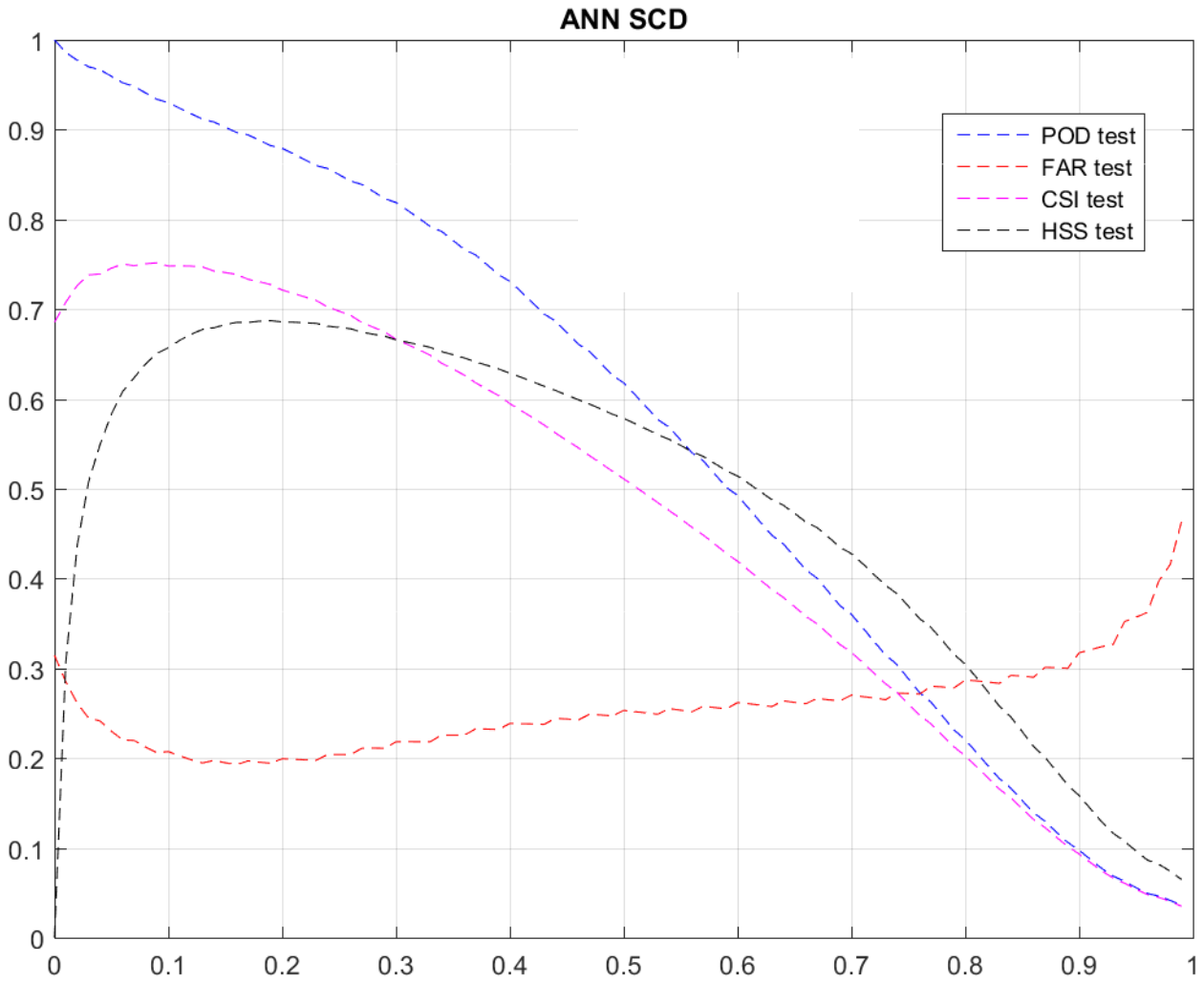




	Precipitation Detection	Snowfall Detection
Prob. Thres.	0.3	0.4
CSI	0.50	0.67
POD	0.70	0.85
FAR	0.30	0.24

The performance was tested using a two year (2014 and 2016) indipendent dataset not used in the training phase.

Supercooled Droplet Detection Module



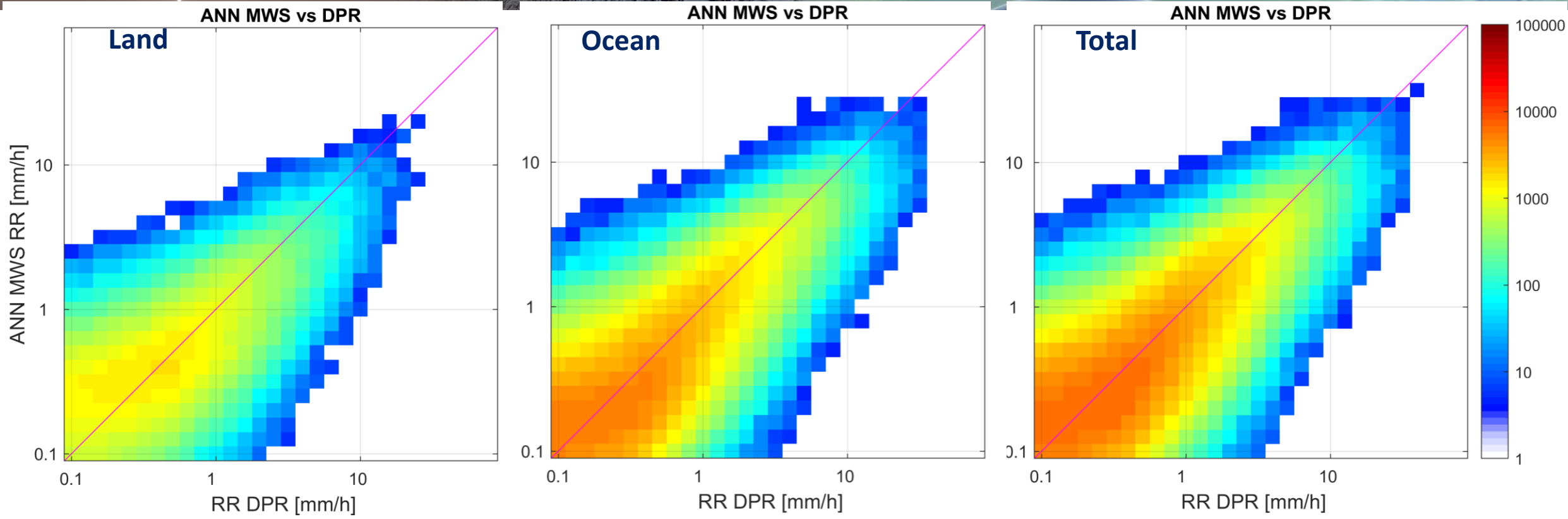
Why Supercooled droplets detection module?

Supercooled droplets tend to partially mask scattering by snow crystals, which complicates the detection of snow using passive microwave radiometers.

Supercooled Fraction Thres.	0.19
HSS	0.70
CSI	0.73
POD	0.88
FAR	0.19

The Supercooled Fraction is the mean value of DARDAR (liDAR + raDAR) binary variable within the ATMS ifov

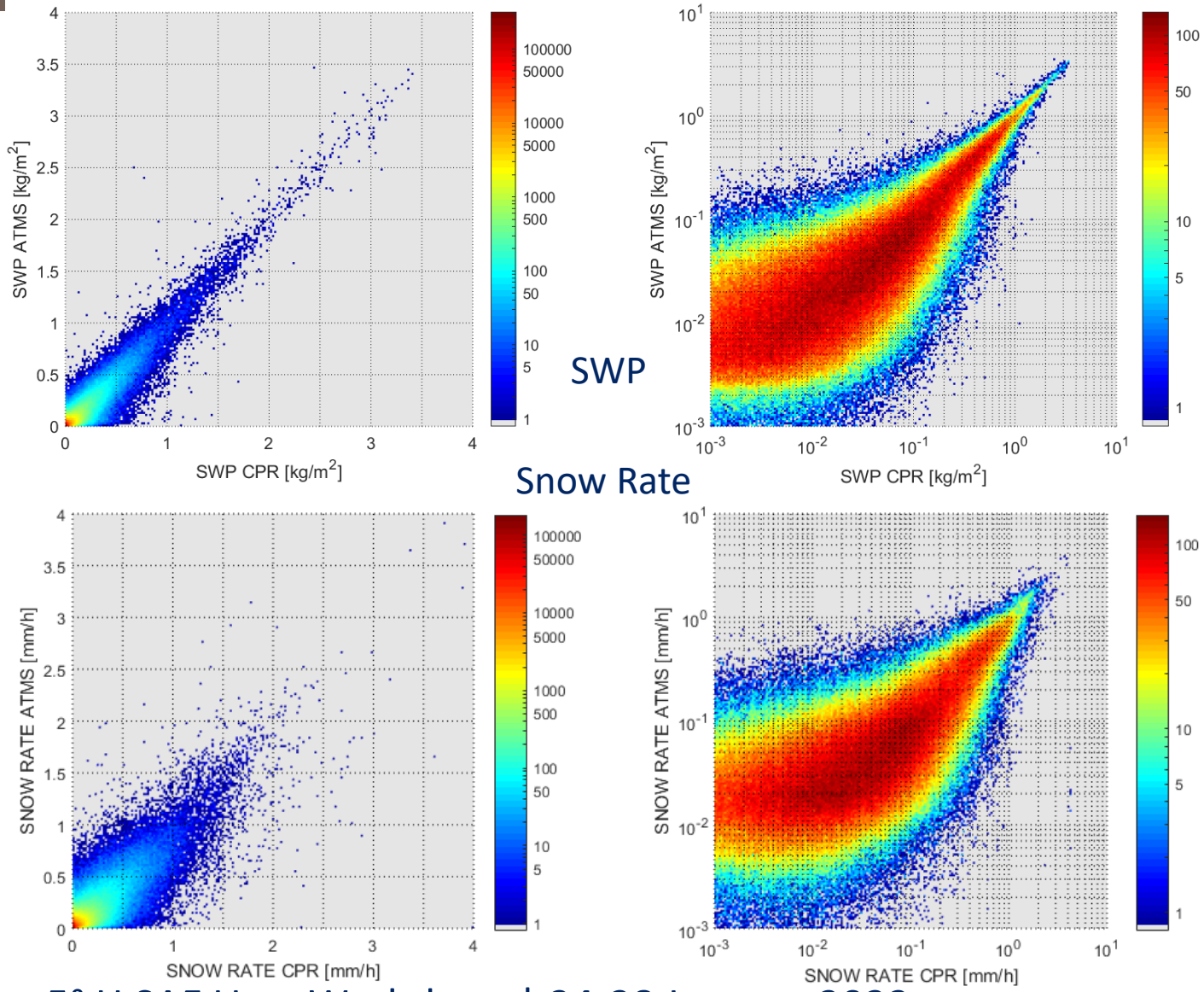
H70 Precipitation Rate Estimate Module



	Land	Ocean	Total
ME (mm/h)	0.03	0.01	0.01
CC	0.71	0.75	0.74
RMSE (mm/h)	0.97	0.98	0.98
N. Point	397429	957058	1354487

Only pixels where both H70 and DPR precipitation rate >0.1 mm/h (hit) were considered

H70 Snow Water Path and Snow Rate Estimate Modules



ME	0.00
CC	0.93
RMSE	0.05
N. Point	678631

The performance was tested using a two year indipendent ATMS-CPR dataset not used in the training phase.

ME	0.00
CC	0.83
RMSE	0.08
N. Point	678631

This result confirms the ability of ML approaches to learn the "truth" from CPR and extend it to the entire ATMS swath

H70 and The Goddard profiling algorithm (GPROF) for ATMS Comparison

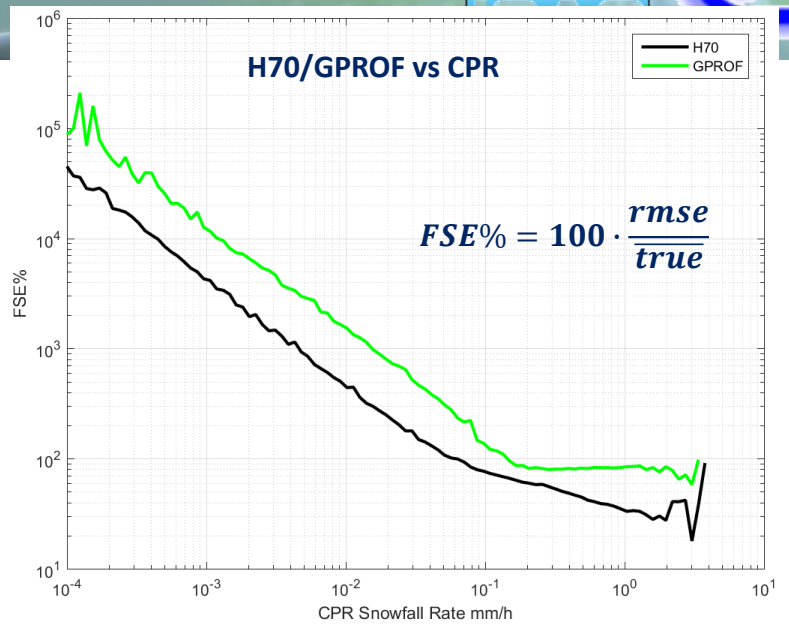
Procedure followed for the analysis

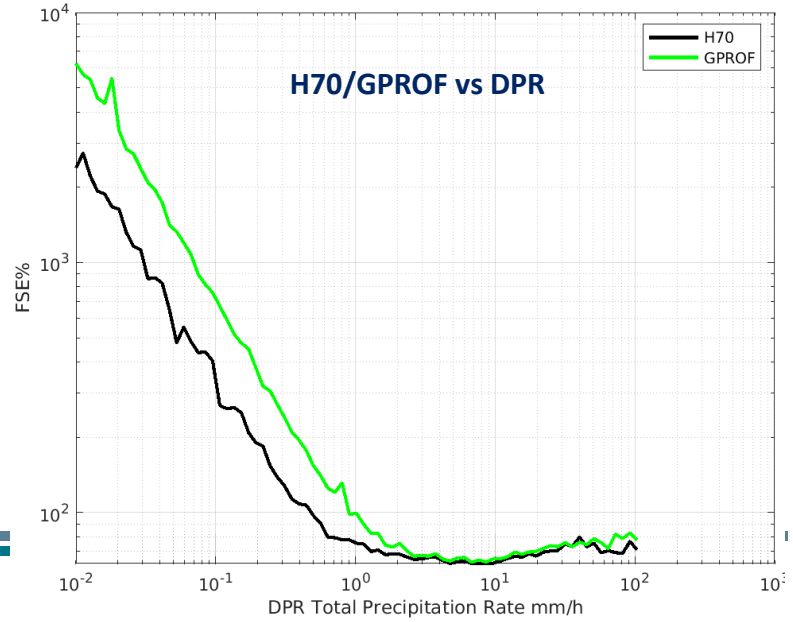
1. The comparison was carried out considering a two-year period (2014 and 2016 – our Test Dataset).
2. The statistics were evaluated according to the surface classes identified by GPROF.
3. For the analysis with CPR (snowfall) the H70 snow rate and the GPROF Frozen Precipitation (FP> 85%) were compared.
4. For the analysis with DPR the H70 and GPROF total precipitation were compared.

It should be pointed out that:

1. In the analysis the reference is for snow 2C-SNOW-PROFILE (CPR derived products used in H70 training) and, for the total precipitation, the 2B-CMB level-2 GMI / DPR combined V06A.
2. GPROF uses for precipitation detection/estimate:
   The Multi-Radar/Multi-Sensor System (MRMS) data over snow covered surfaces
   The 2B-CMB level-2 GMI / DPR combined over ocean and sea ice
   The 2A-DPR (Ku band) and over Land.

	GPROF SNOW	H70 SNOW	GPROF OCEAN	H70 OCEAN	GPROF LAND	H70 LAND
RMSE	0,18	0,10	1,82	1,58	1,65	1,14
ME	0,006	0,002	0,08	0,40	0,15	-0,02
CC	0,55	0,80	0,60	0,71	0,55	0,69
POD	0,20	0,76	0,59	0,64	0,54	0,69
FAR	0,55	0,22	0,56	0,30	0,44	0,33
HSS	0,05	0,63	0,38	0,60	0,48	0,63





Future Developments

Future activities for the MWS day-1 algorithm

- During the MWS commissioning phase the H70 calibration-tuning will be carried out.
- Some algorithm modules based on deep learning techniques could be updated in order to improve the performance. The use of Convolutional Neural Networks (Deep Learning techniques) has shown great capability in precipitation and snowfall detection applications.
- An extensive validation will be carried out to confirm the algorithm performance.

Future activities for the development of MWS day-2 algorithm

- We are also carrying out the activity aimed at the development of the MWS day-2 algorithm (planned in the CDOP4, 2022-2027).
  The main goals of the activity are to improve the light rainfall and the warm rain detection and estimate.
- For this purpose two additional machine learning modules will be developed:
  - The light rainfall detection and estimate modules
  - The warm rain detection and estimate modules