Utilisation de méthodes de machine learning pour l’estimation et l’études des précipitations

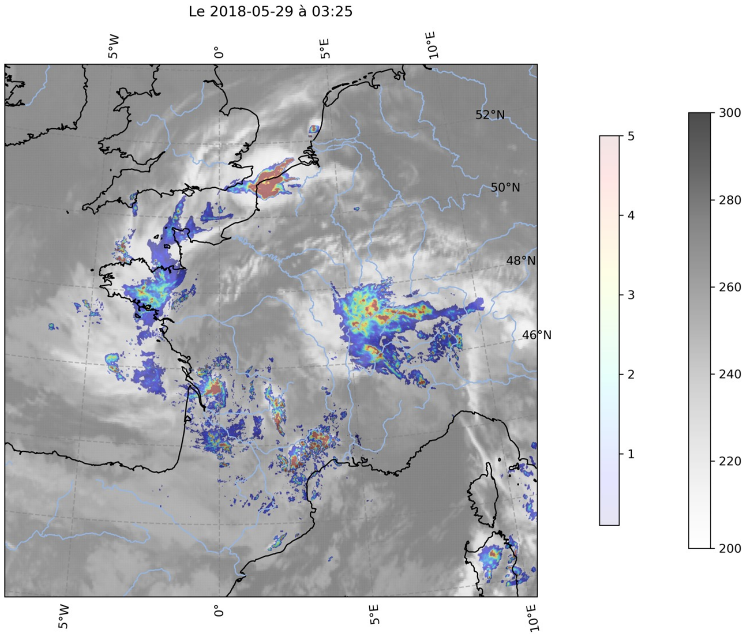
NICOLAS VILTARD, CÉCILE MALLET, LAURENT BARTHÈS,
PIERRE LEPETIT, VIBOLROTH SAMBATH, MATTHIEU MEIGNIN,
VICTOR ENESCU, ASSAAD ZEGHINA
…

Laboratoire ATmosphères, Milieux, Observations Spatiales/Institut Pierre Simon Laplace
CNRS-UVSQ-UPMC

Context

- ❑Rain is a very special variable
  - ❑Spatial intermittence
  - ❑Temporal intermittence
  - ❑Extremely scale dependent
  - ❑Surface rain is dependent on 3-D structure
  - ➔several issues:
    - rain/no rain detection
    - quantitative estimation
    - extreme values

- ❑End users require < 1km, < 5minutes
  - ❑"Direct" rain measurement on LEO
  - ❑Geostationary offer only IR



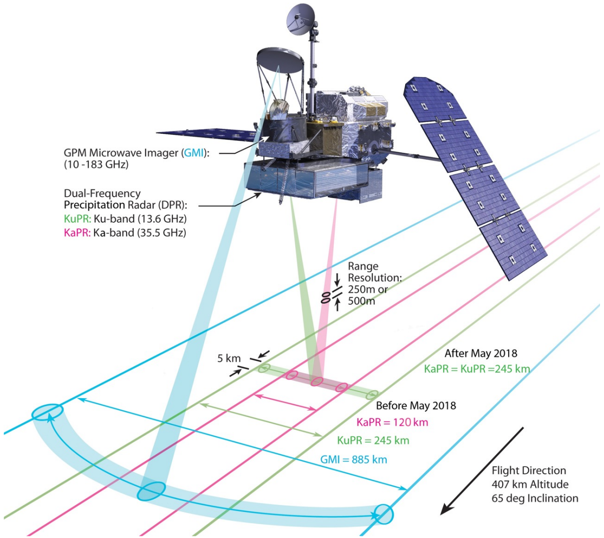
➔ We probably need some sort of fusion in the end…

DRAIN: Deep learning for RAIN

GPM (GMI + DPR) : from BRAIN to… DRAIN

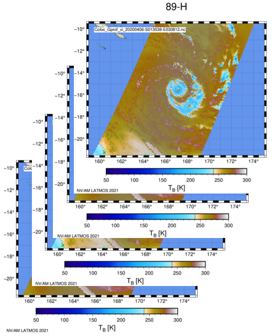
(from Bayesian to Deep-Learning)

- Passive microwave (10 to 100th of GHz)
  - Spatial resolution ~10 km but frequency dependent
  - Measures integrated over the whole column
  - Highly ill-posed problem
- Active microwave (14 and/or 35 GHz)
  - Spatial resolution closer to 5 km
  - Measures a profile
  - Strong dependance on DSD…

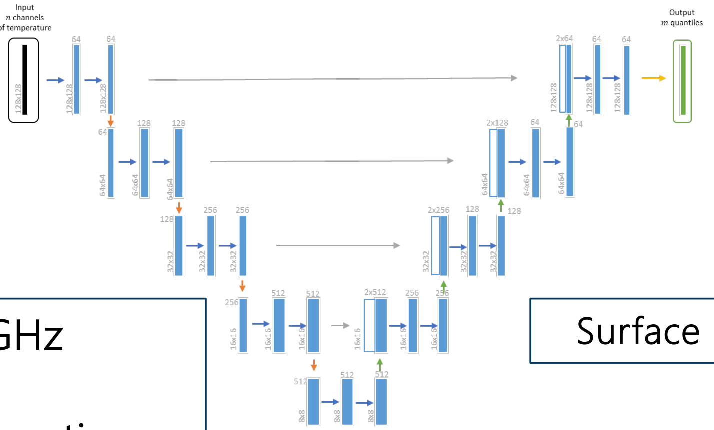


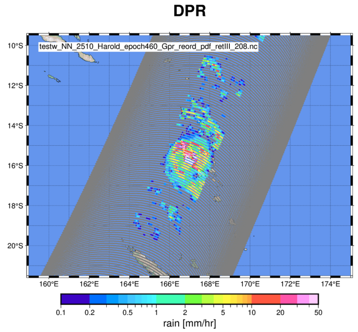
- 1987: DMSP/SSMI; Itératif, RTM
- 1990: Bayesian, RTM
- 1997: TRMM/TMI+PR
- 2000: ML, co-located; Bayésien, co-located
- 2008
- 2014: GPM/GMI+DPR; Constellation
- 2016: Bayésien, co-located, RTM
- 2020: DL, co-located
- 2023: Cycle GAN
- 2025: IR; Diffusion models

Supervised retrieval of rain intensity: DRAIN



Quantile regression
loss function





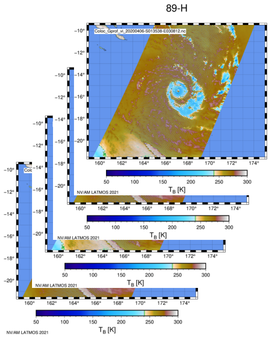
TB of 37 GHz and 89 GHz
channels of GMI
No other a-priori information

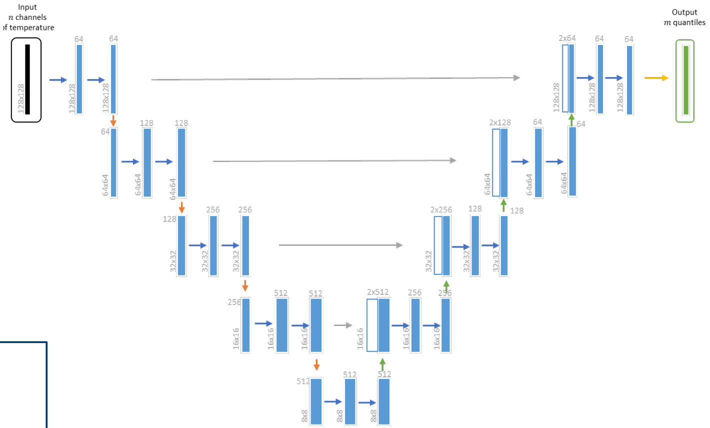
Surface Rain intensity from DPR

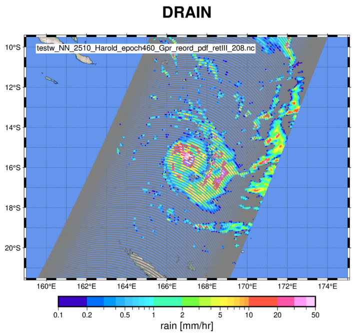
- ❑ Database made of co-located data over 7 years
- ❑ Training ~100 000 images – rainy situations are selected

Supervised retrieval of rain intensity: DRAIN

Quantile regression loss function





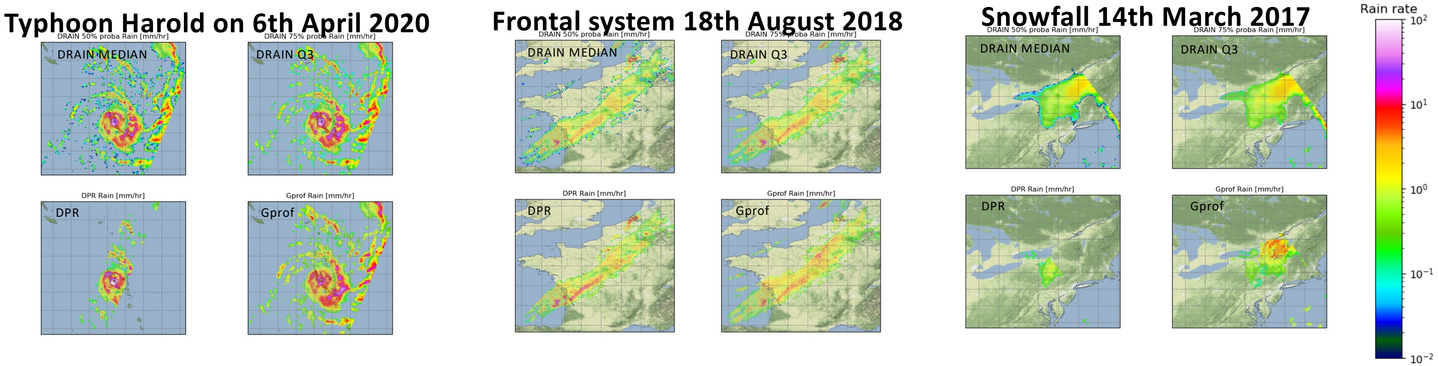


TB of 37 GHz and 89 GHz channels of GMI

Q_i i=1 to 99

DRAIN Q_{50} Results: examples

GPROF: reference algorithm from NASA (with fair amount of aux. data), pixel by pixel

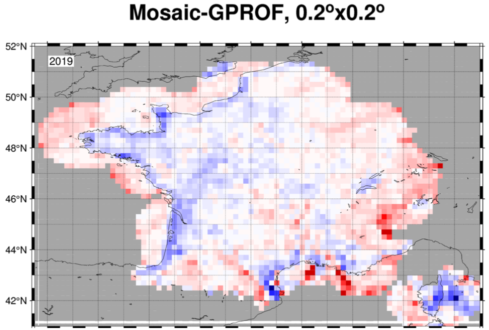


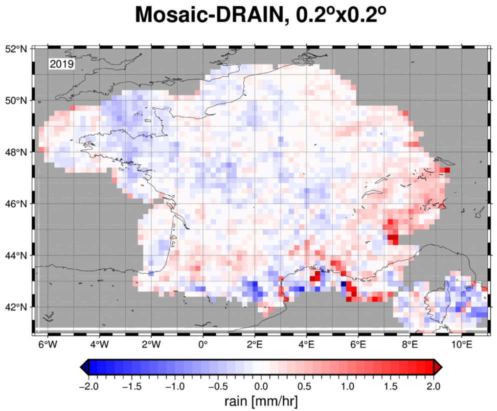
- ❑ Excellent rain/no-rain discrimination w/r to DPR
- ❑ Excellent retrieval of both structure and intensity in very different situations
- ❑ Equivalent to GPROF without any auxiliary data !

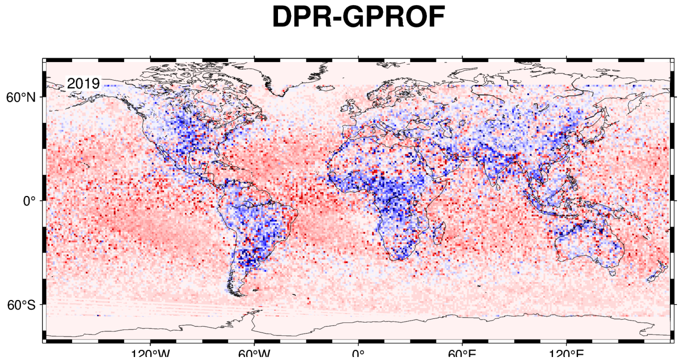
Base de test 2019 comparaison to GPROF

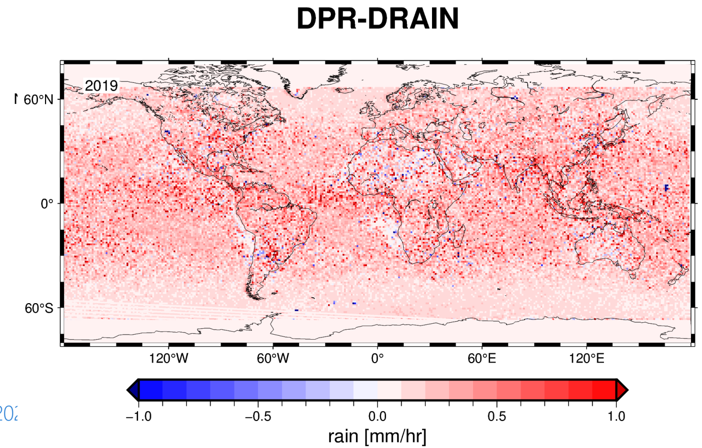
METEO-France MOSAIC
Résolution 0.2°x0.2°

5°x5° resolution





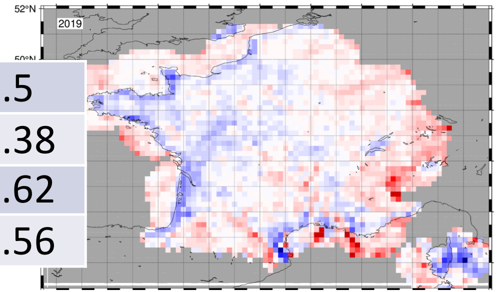




Base de test 2019 comparaison to GPROF

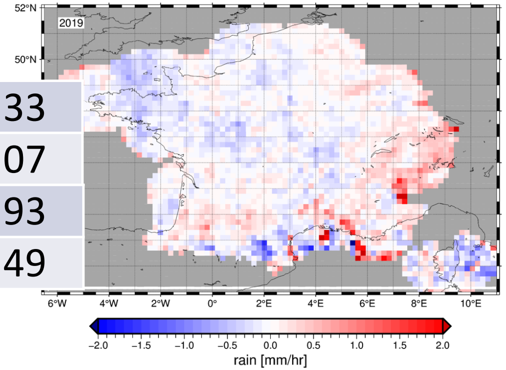
METEO-France MOSAIC Résolution 0.2°x0.2°

Mosaic-GPROF, 0.2°x0.2°



POD	0.5
FAR	0.38
Precision	0.62
F1-score	0.56

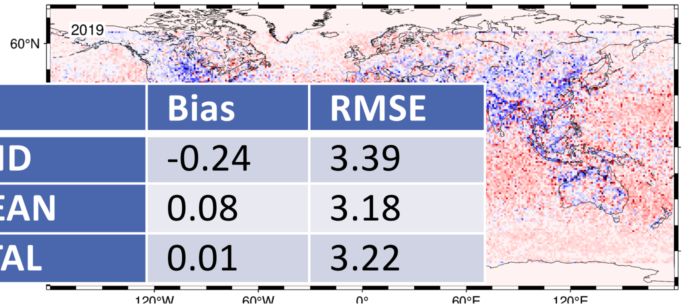
Mosaic-DRAIN, 0.2°x0.2°



POD	0.33
FAR	0.07
Precision	0.93
F1-score	0.49

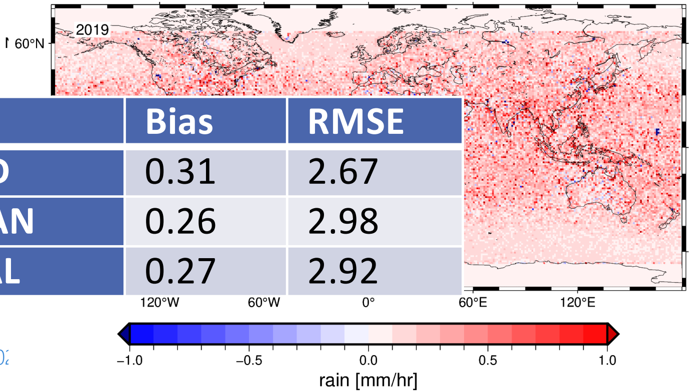
5°x5° resolution

DPR-GPROF



	Bias	RMSE
LAND	-0.24	3.39
OCEAN	0.08	3.18
TOTAL	0.01	3.22

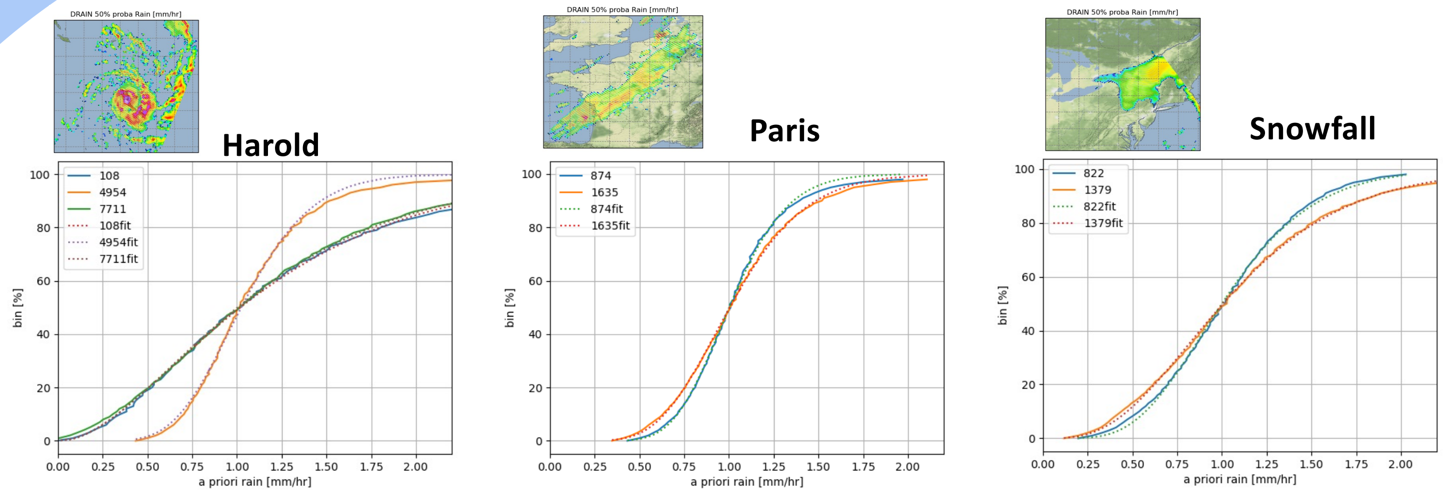
DPR-DRAIN



	Bias	RMSE
LAND	0.31	2.67
OCEAN	0.26	2.98
TOTAL	0.27	2.92

DRAIN Quantile regression
Different estimator

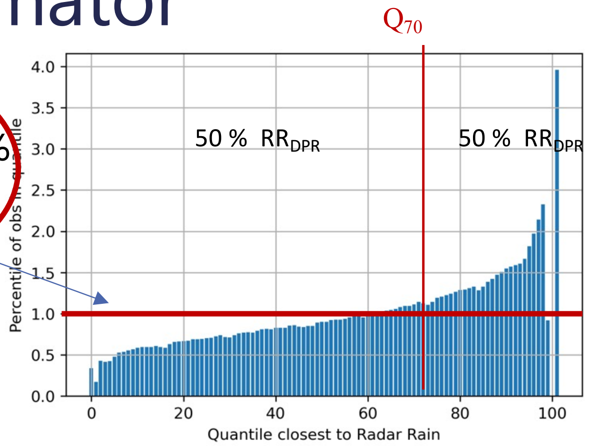
Quantile regression : exploitation ?



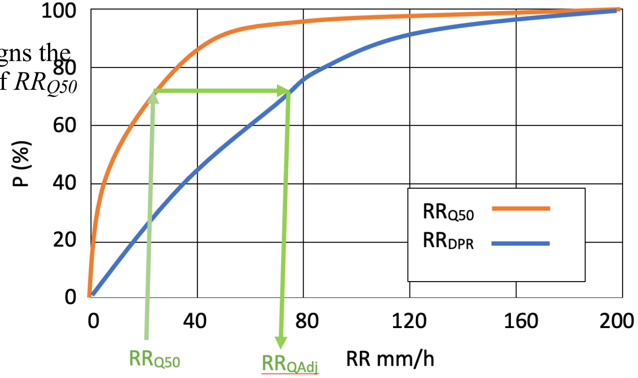
- ❑ CDF for 1 mm/hr (median, random pixels)
- ❑ There seem to be different CDFs for different regimes

Comparison of Different Estimator

Each percentile should contain 1% of the observed values.

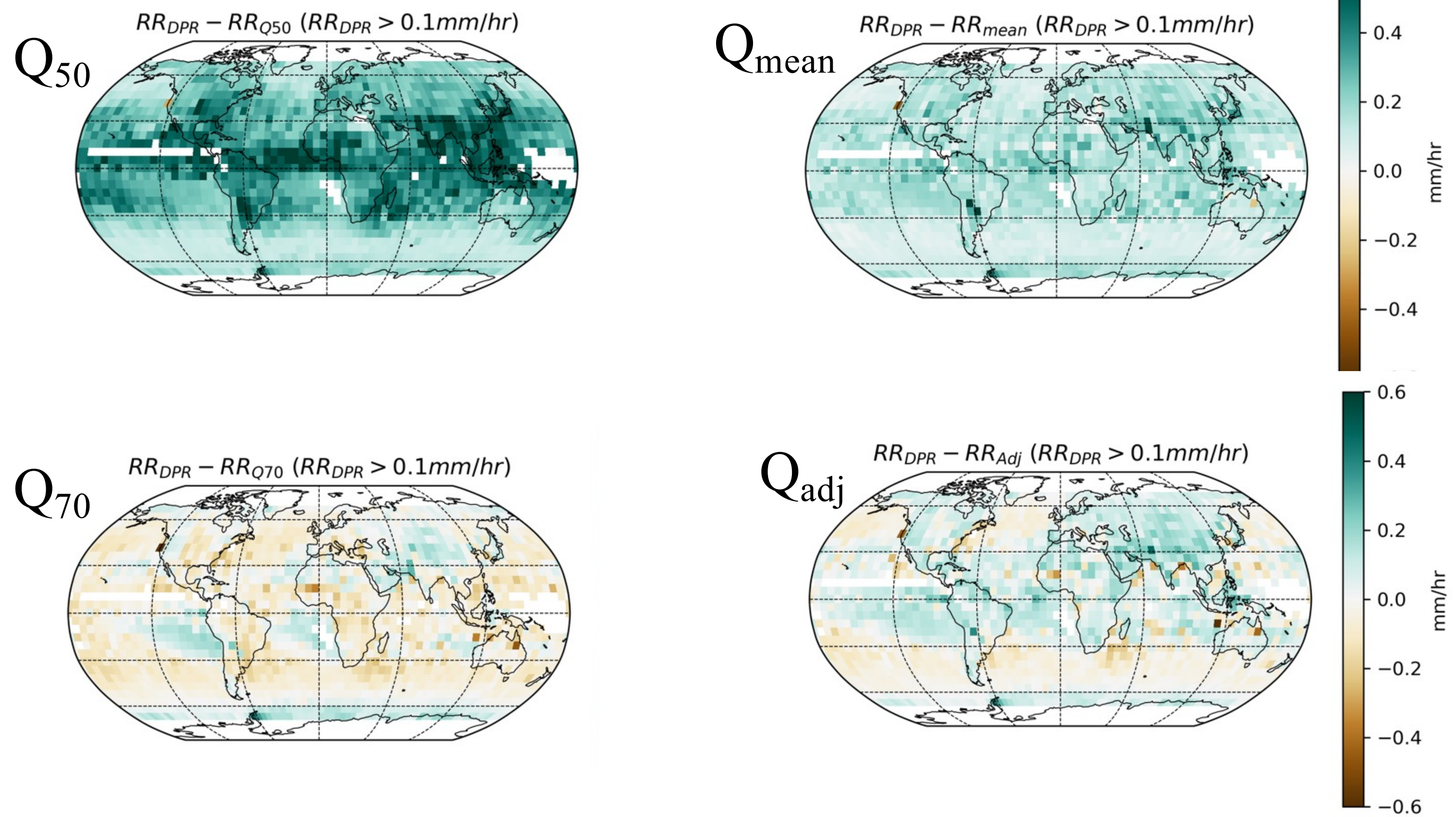


- Q_{50}
- $Q_{mean} = \frac{1}{99} \sum_{i=1}^{99} Qi$
- Q_{70} histogram matching, using a transformation function that aligns the cumulative distribution function (CDF) of RR_{DPR} with that of RR_{Q50}
- Q_{adj} histogram matching



Base de test DPR 2019

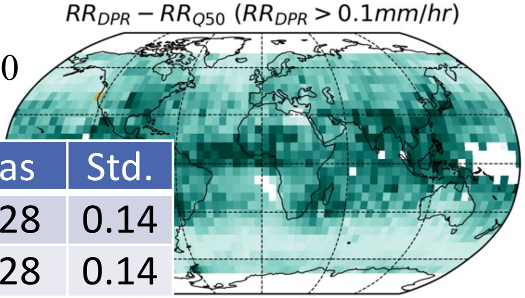
5°x5° resolution



Base de test DPR 2019

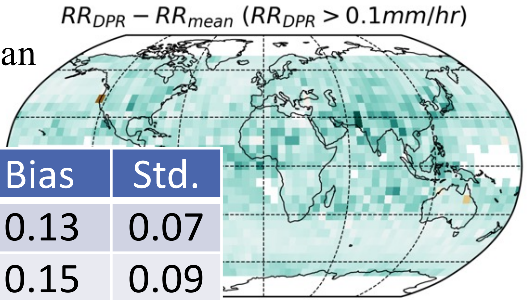
5°x5° resolution

Q_{50}

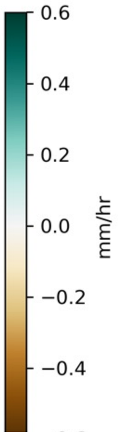


	Bias	Std.
Q_{50}	0.28	0.14
MAE	0.28	0.14

Q_{mean}

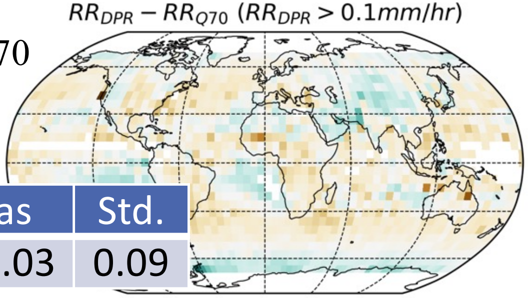


	Bias	Std.
Q_{mean}	0.13	0.07
RMSE	0.15	0.09



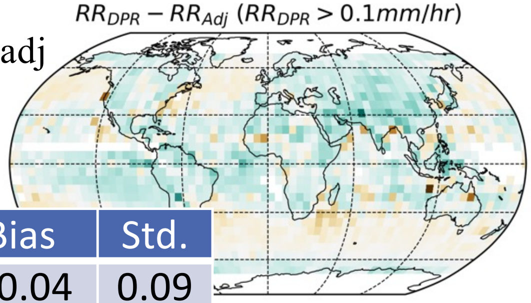
	Bias	Std.
GPROF	0.9	6.67

Q_{70}

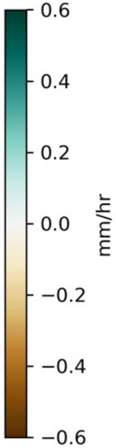


	Bias	Std.
Q_{70}	-0.03	0.09

Q_{adj}



	Bias	Std.
Q_{adj}	0.04	0.09

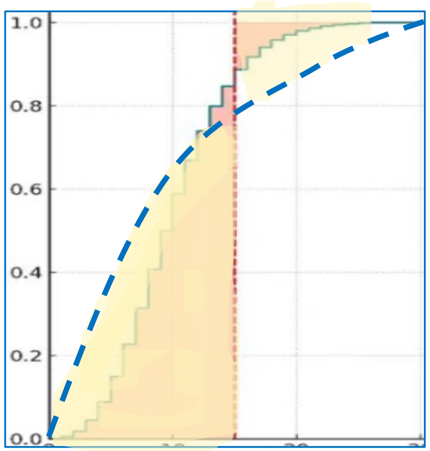


DRAIN Quantile regression
Evaluation of predicted uncertainty

LATMOS

Continuous Ranked Probability Skill Score (CRPSS)

- Continuous Ranked Probability Score $CRPS_{Est}\ (TB)$ and $CRPS_{Ref}\,(TB)$ **for each pixel**



RR$_{DPR}$

CDF$_{Est}$ constructed from DRAIN estimated quantiles RR$_{Qi}$ i=1,...,99

CDFRef constructed from DPR global dataset 2019

- Continuous Ranked Probability Score $CRPS_{Est}$ and $CRPS_{Ref}$ for each 5°x5° box

$$CRPS = \sum_{5°x5°\,box} CRPS\ (TB)$$

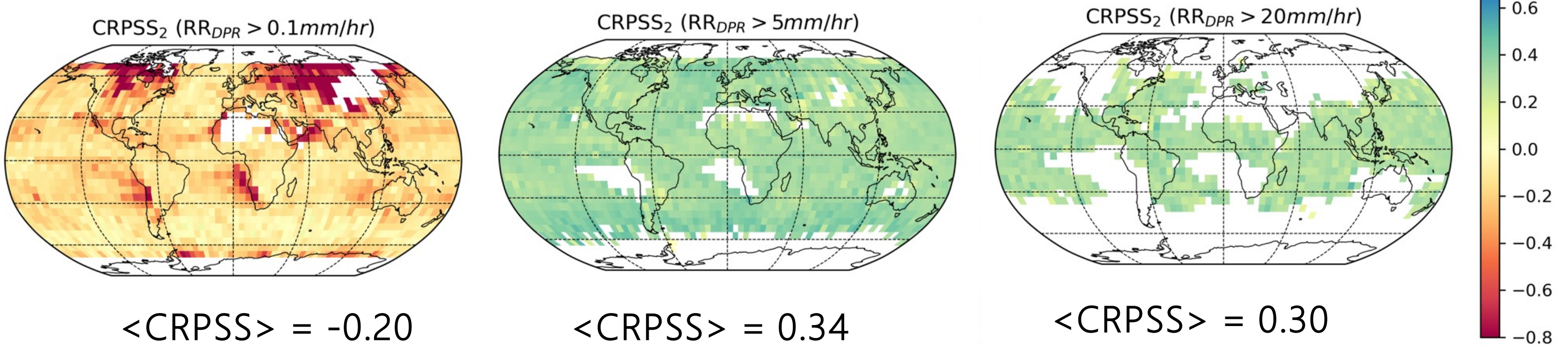
- Continuous Ranked Probability Skill Score $CRPSS$ for each 5°x5° box

$$CRPSS = 1. - CRP\,S_{Est}/C\,RPS_{Ref}$$

Base de test DPR 2019

5°x5° resolution

CRPSS for various thresholds of RR_{DPR}



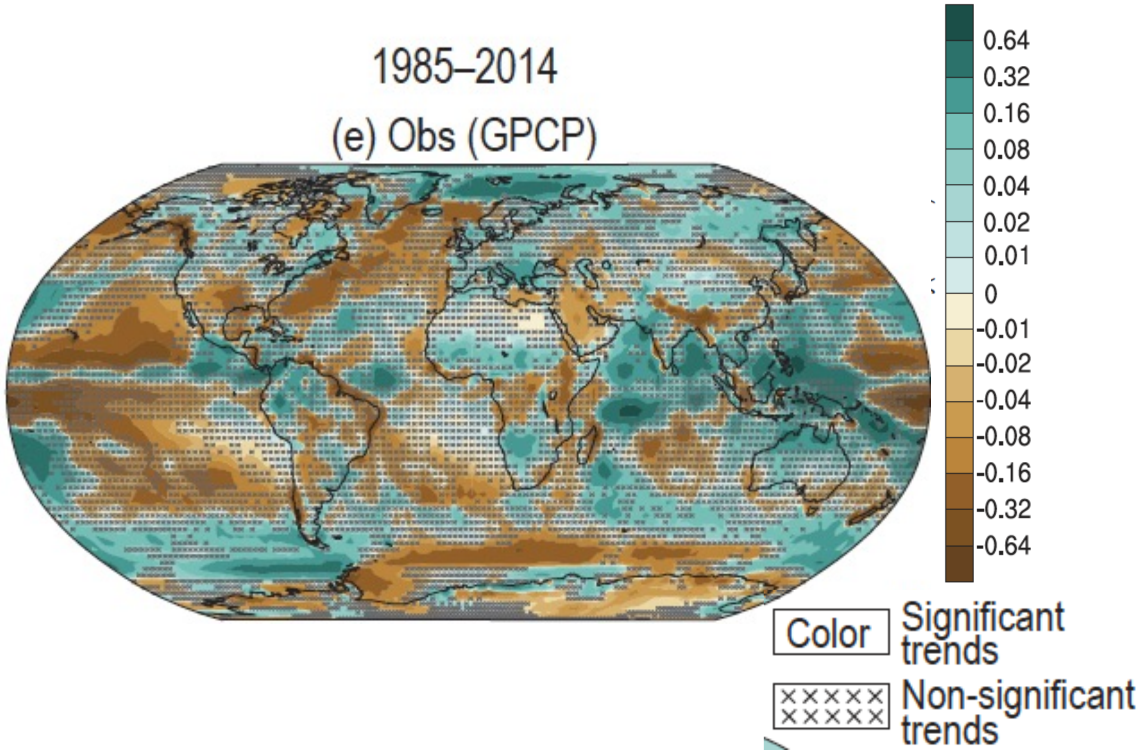
<CRPSS> = -0.20 <CRPSS> = 0.34 <CRPSS> = 0.30

- ❑ RR_{DPR} > 0.1 mm/hr - driven by light rainfall events
  ➔ CDF_{Est} are poorly constrained
- ❑ RR_{DPR} > 5 mm/hr - driven by intense events
  ➔ Model captures the structure of the conditional distribution P(RR/TB)
- ❑ RR_{DPR} > 20 mm/hr - driven by extrem intense events
  ➔ saturation effects in the TB

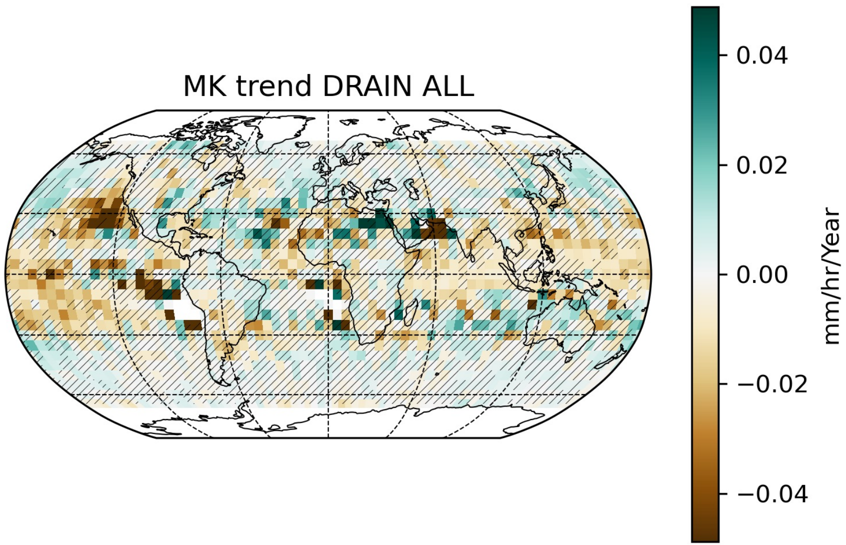
Evaluation of Trends over 2014-2024 period

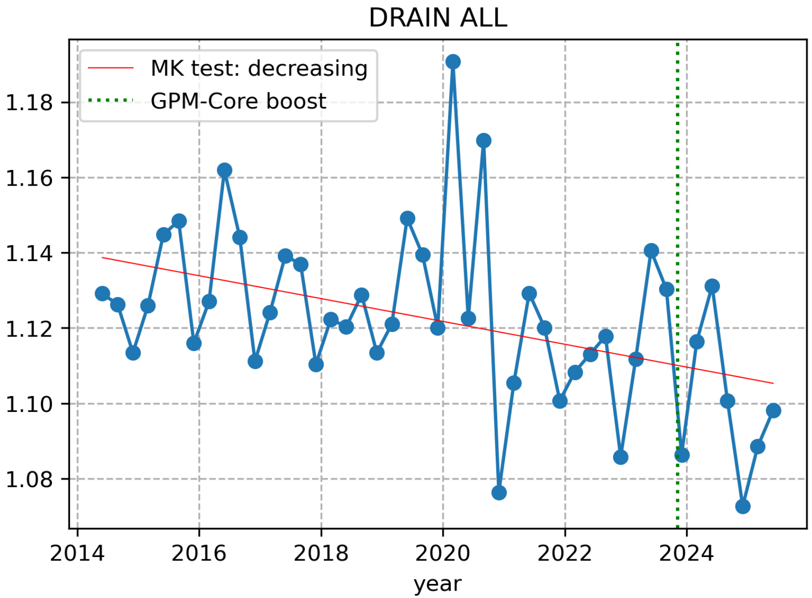
LATMOS

TREND IN ANNUAL MEAN PRECIPITATION



Trends over 2014-2024 period





From: Mischell E, Soden B. Observed Trends in Extreme Precipitation and Convective Intensity under Global Warming. *J. Climate.* *2025;38(24):7543-7556. doi:10.1175/JCLI-D-25-0122.1*

DOI: https://doi.org/10.1175/JCLI-D-25-0122.1

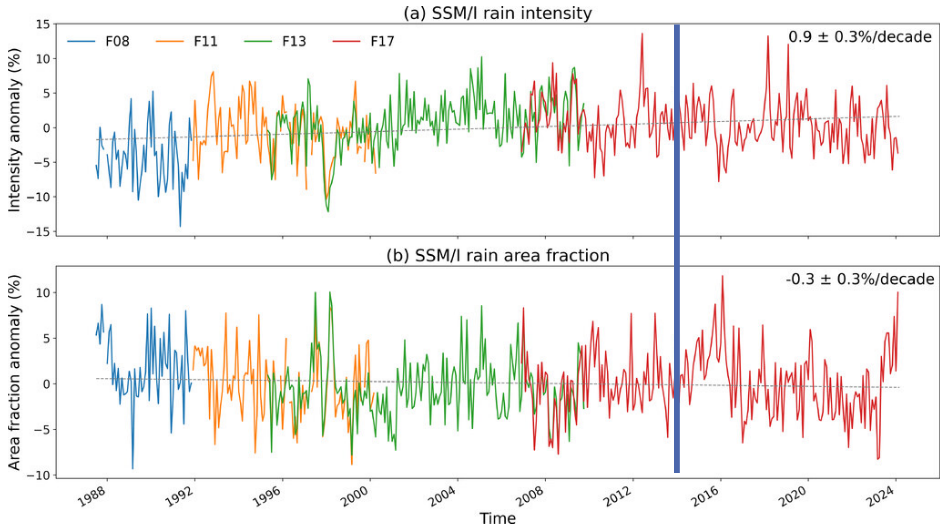
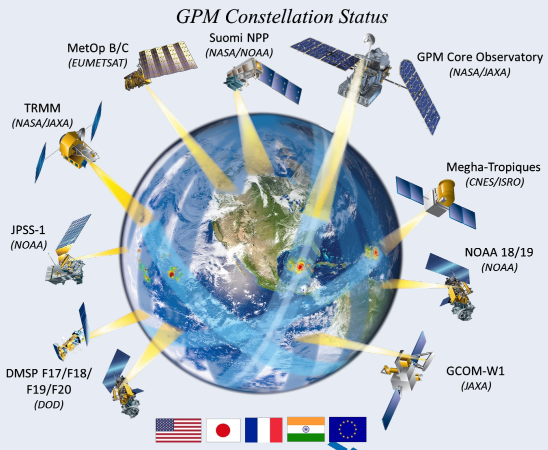


Fig. 5. Time series of the (a) rain intensity and (b) fraction of Earth's surface where the rain rate is nonzero. The trends without F08 are (a) 0.2% ± 0.4% decade−1 and (b) 0.1% ± 0.4% decade−1.

Using unsupervised domain adaptation to go from GPM to constellation



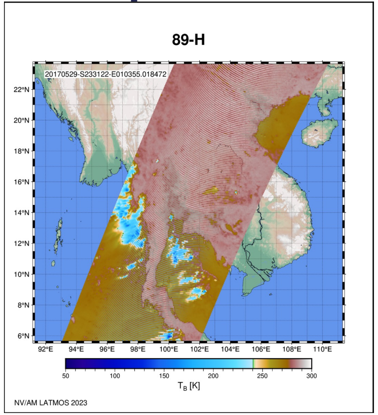
LATMOS

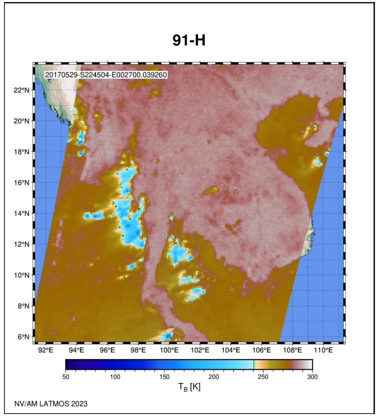
Unsupervised Domain Adaptation

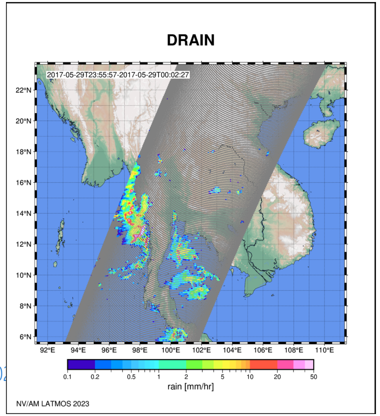
On GPM-core
- GMI
    - 89 GHz, 3x13.4 km^2
    - 36.64 GHz, 6x13.4 km^2

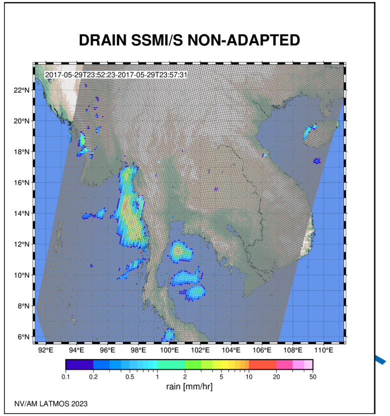
On NOAA-F18, close but different…
- SSMI/S
    - 91.65 GHz, 12.5x12.5 km2
    - 37 GHz, 12.5x25 km2

- DRAIN on GMI
- DRAIN on raw SSMI/S



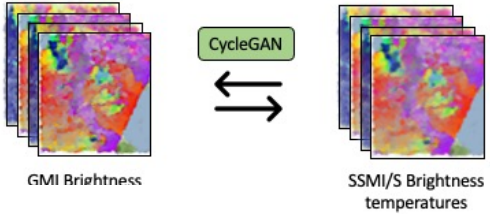


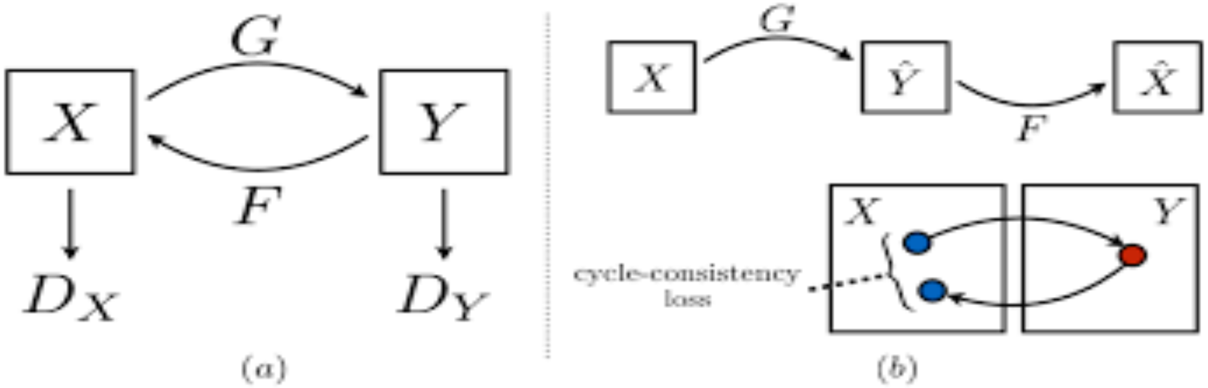




Unsupervised Domain Adaptation

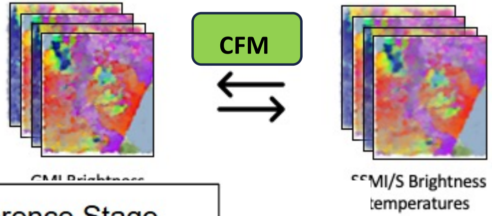
- Cycle Generative Adversarial Nets (Zhu et al. 2017)

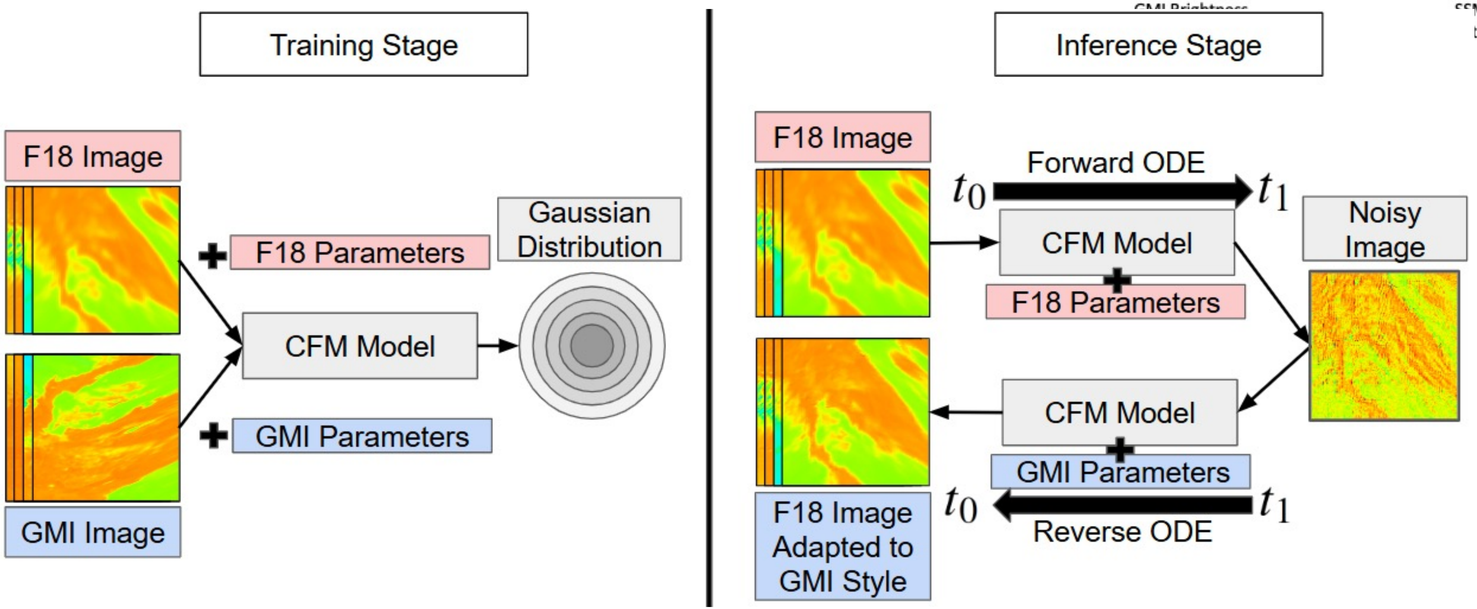


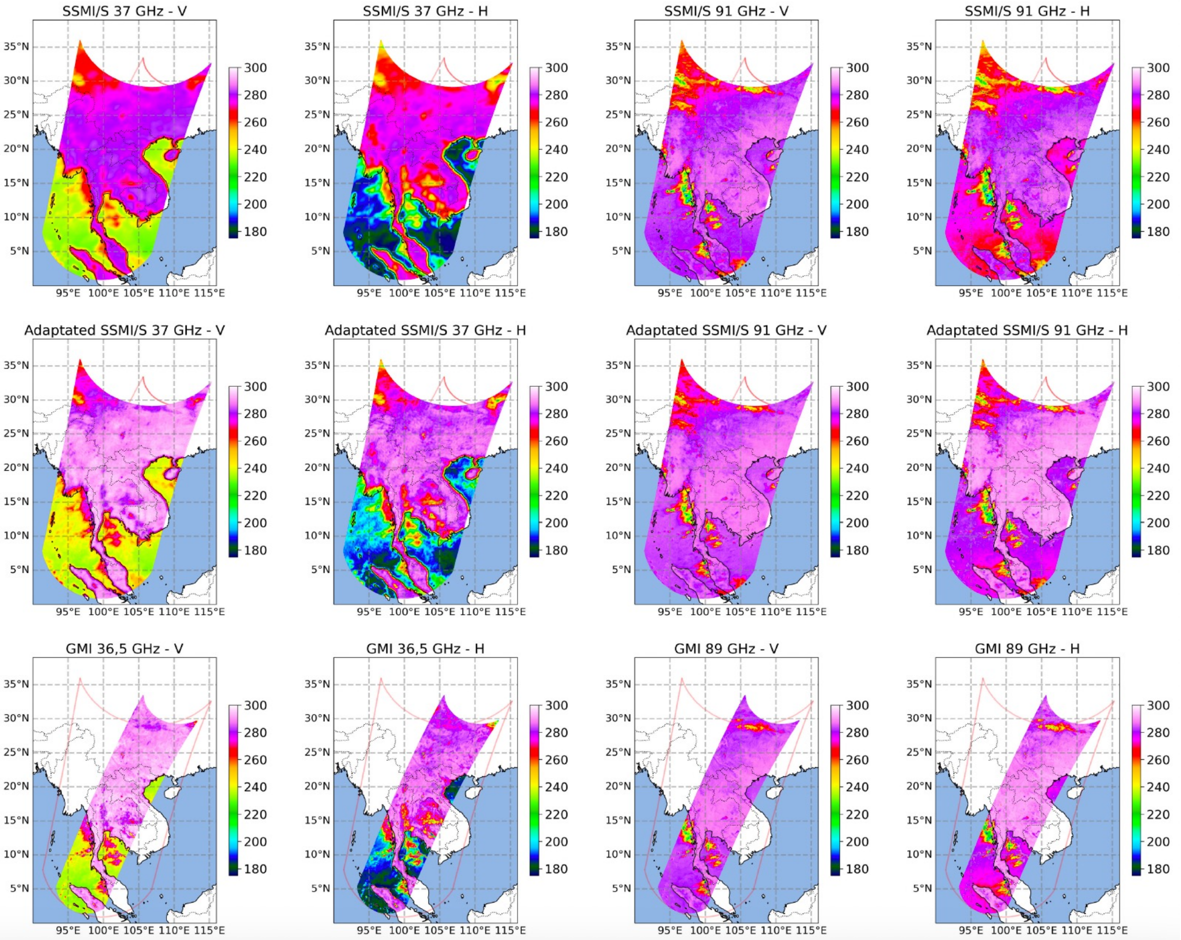


Unsupervised Domain Adaptation

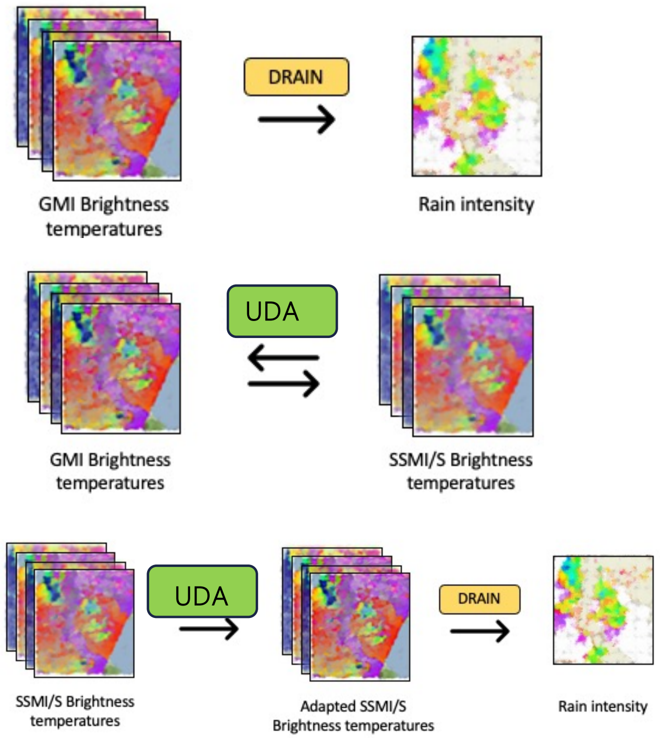
- Conditional Flow Matching

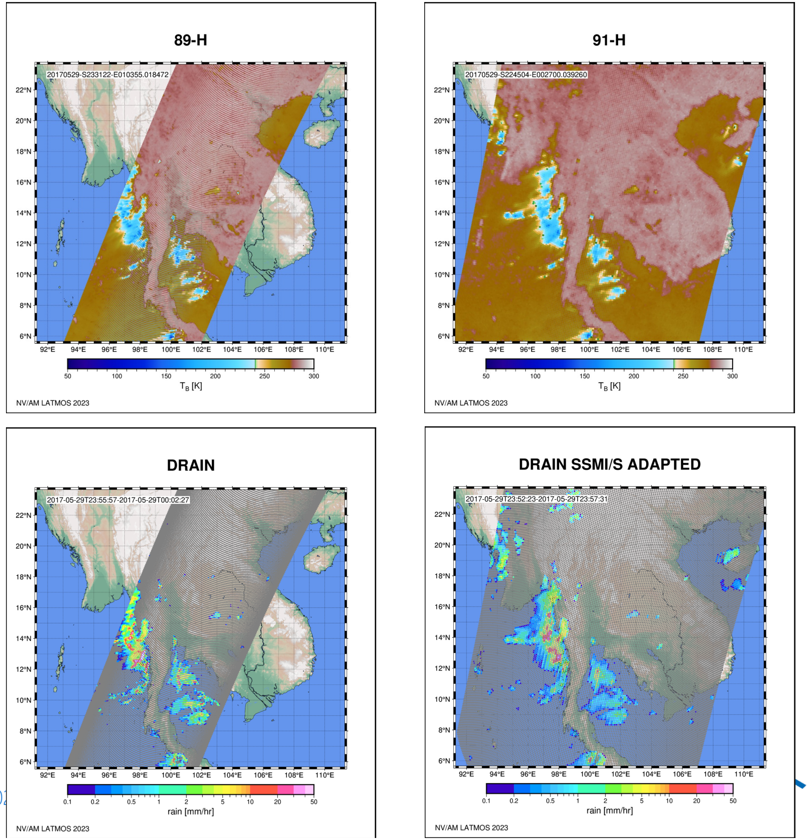




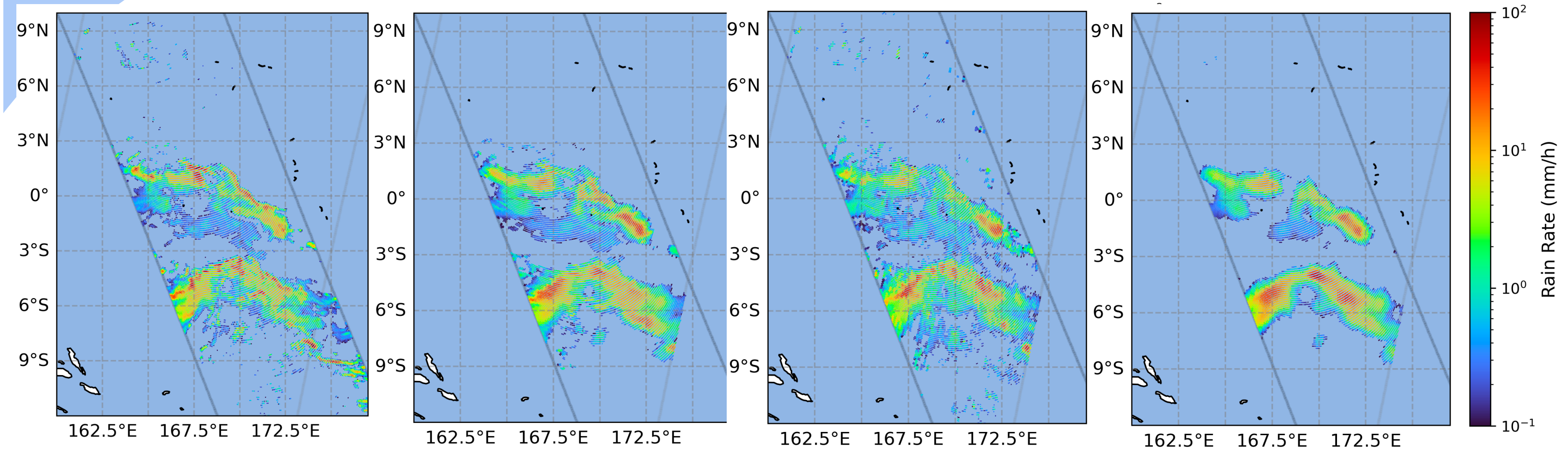

SSMI/S 37 GHz - V
SSMI/S 37 GHz - H
SSMI/S 91 GHz - V
SSMI/S 91 GHz - H
Adaptated SSMI/S 37 GHz - V
Adaptated SSMI/S 37 GHz - H
Adaptated SSMI/S 91 GHz - V
Adaptated SSMI/S 91 GHz - H
GMI 36,5 GHz - V
GMI 36,5 GHz - H
GMI 89 GHz - V
GMI 89 GHz - H
35°N
30°N
25°N
20°N
15°N
10°N
5°N
95°E 100°E 105°E 110°E 115°E
300
280
260
240
220
200
180

Unsupervised Domain Adaptation





Unsupervised Domain Adaptation



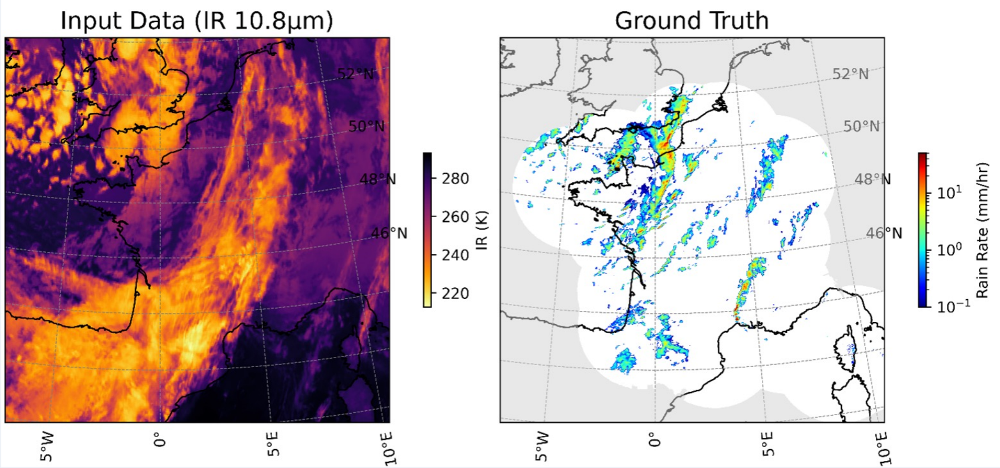
DRAIN on GMI

DRAIN on Adapted F18 CFM

DRAIN on Adapted F18 CycleGAN

DRAIN on F18

Models comparison for precipitation retrievals from IR



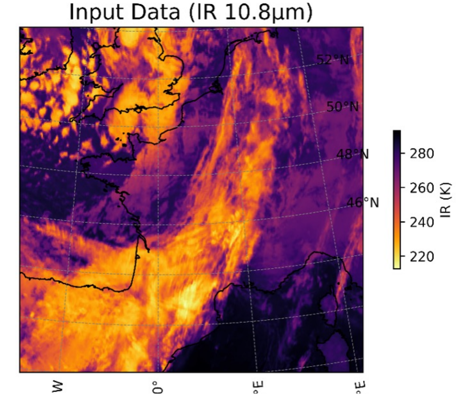
LATMOS

IR DATABASE France 2008–2023

- ❑ Input Data (MSG / EUMETSAT – SEVIRI IR)

  7 infrared channels (6.2 µm – 13.4 µm)

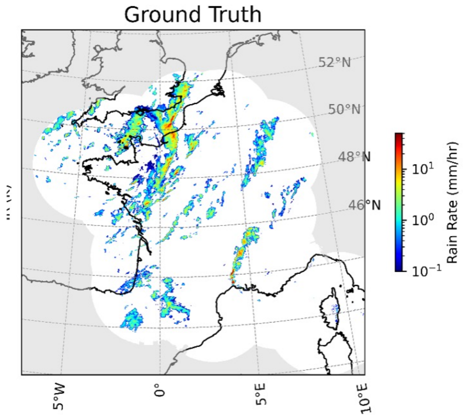
  Every 15-min, 3 km spatial resolution at nadir (~6 km over France), upscaled from 3km to 1km)

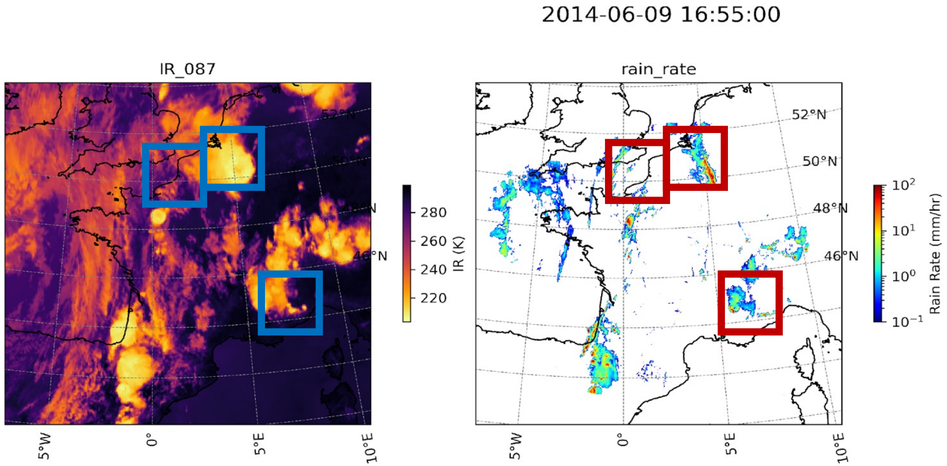


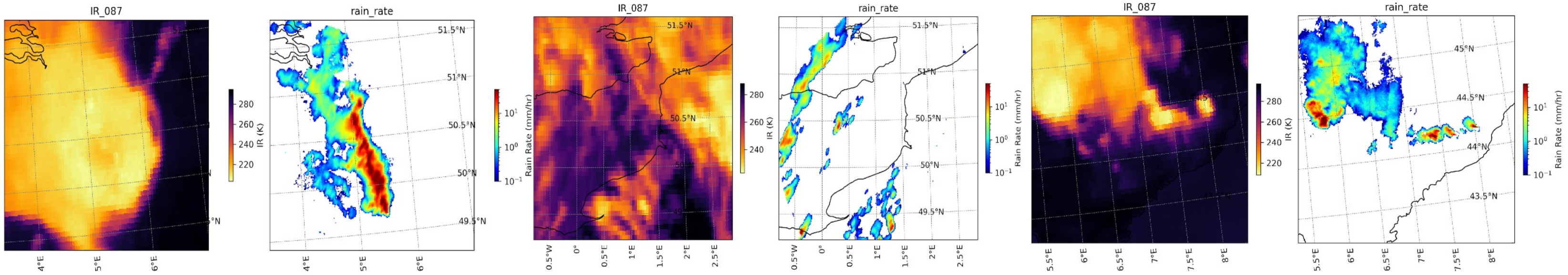
- ❑ Target Data (Météo-France Radar / Ground Truth)

  Radar mosaics

  Every 5 min, 1 km spatial resolution





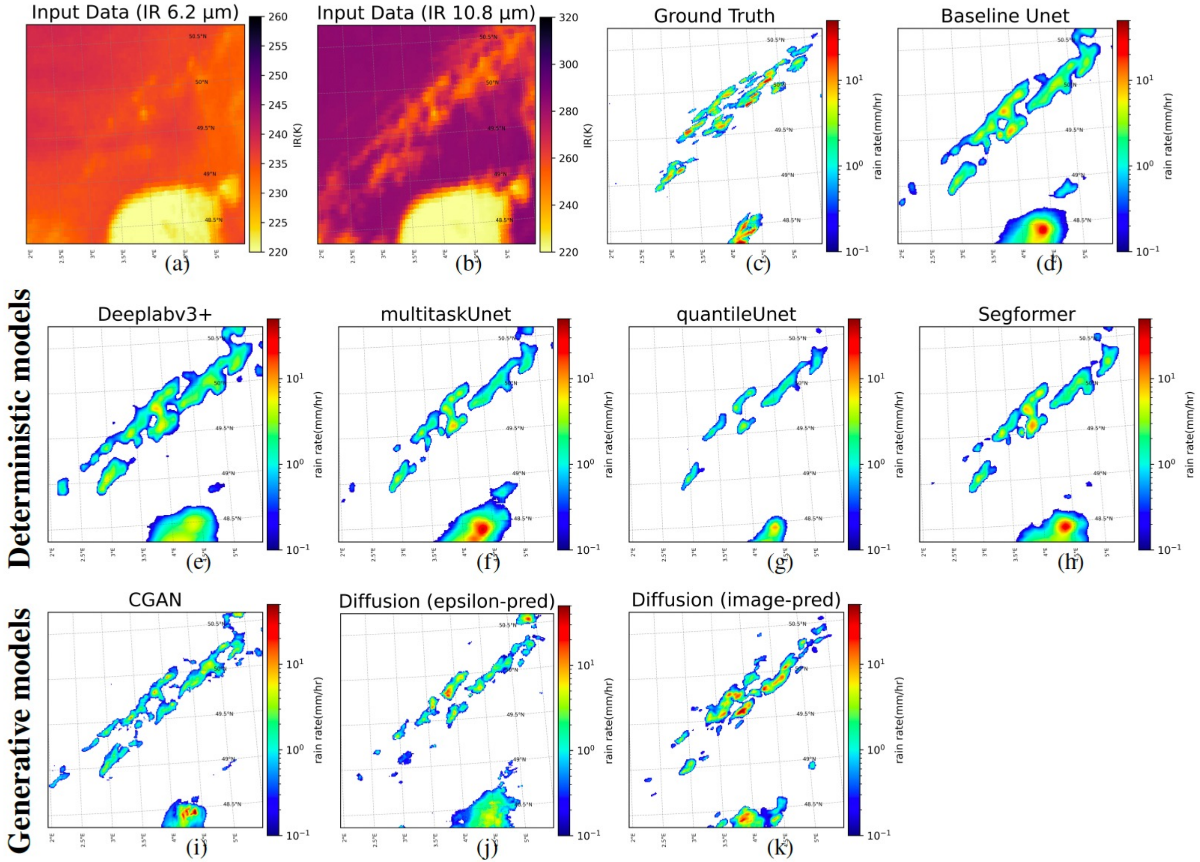


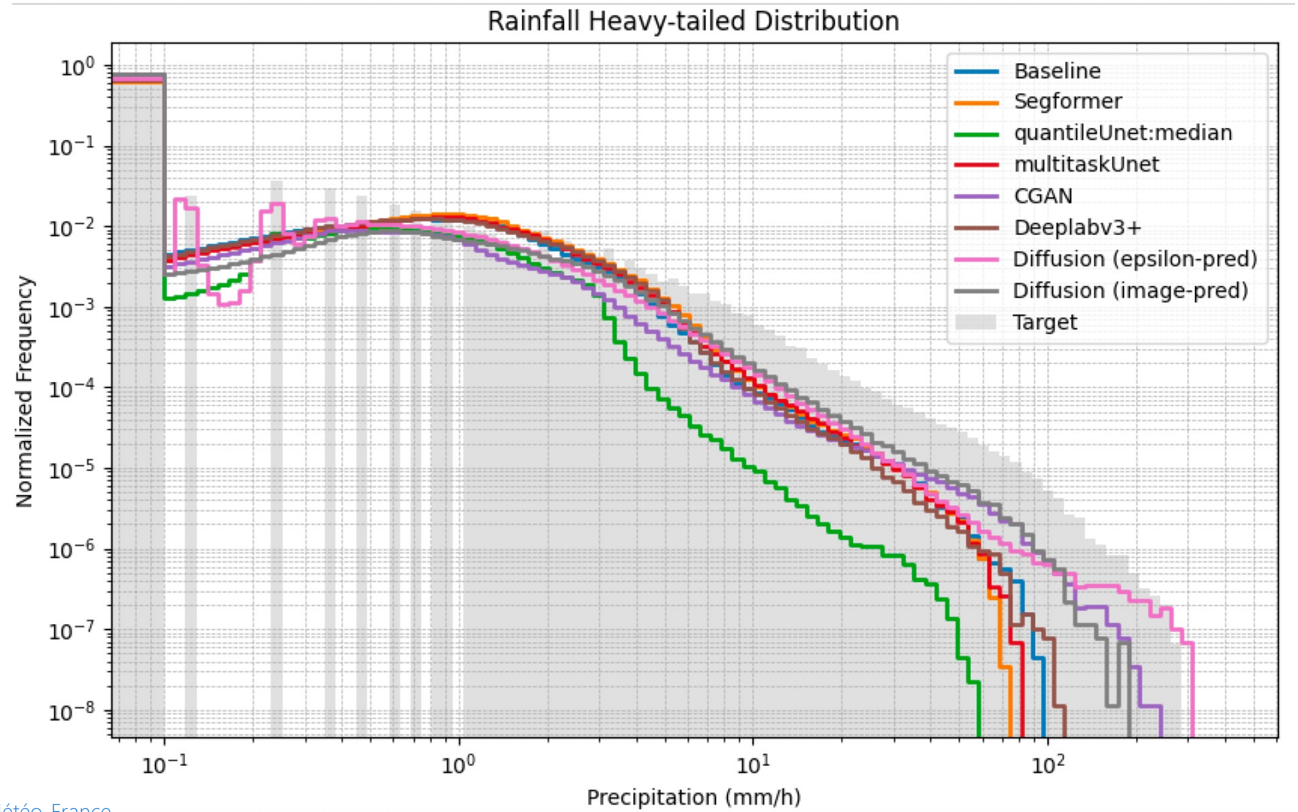
256 km × 256 km tiles used for model input; years 2008–2023 for training/validation, 2019 for testing.

Model inter-comparaison

- Deterministic model
  - U-Net (Ronneberger et al. (2015))
  - Quantile Regression U-Net
  - Multitask U-Net : regression + segmentation
  - DeepLabV3+(Chen et al. (2018))
  - SegFormer (Xie et al. (2021))

- Generative Models
  - Conditional GANs (Hayatbini et al. (2019) and Han et al. (2025))
  - Diffusion Model (Ho et al. (2020),

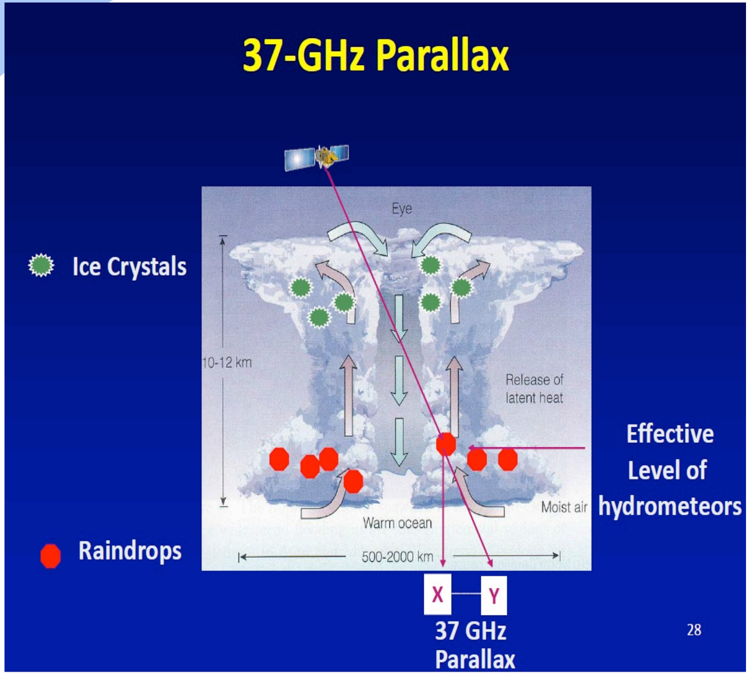

Input Data (IR 6.2 µm)
Input Data (IR 10.8 µm)
Ground Truth
Baseline Unet
IR(K)
rain rate(mm/hr)
(a)
(b)
(c)
(d)
Deterministic models
Deeplabv3+
multitaskUnet
quantileUnet
Segformer
(e)
(f)
(g)
(h)
Generative models
CGAN
Diffusion (epsilon-pred)
Diffusion (image-pred)
(i)
(j)
(k)

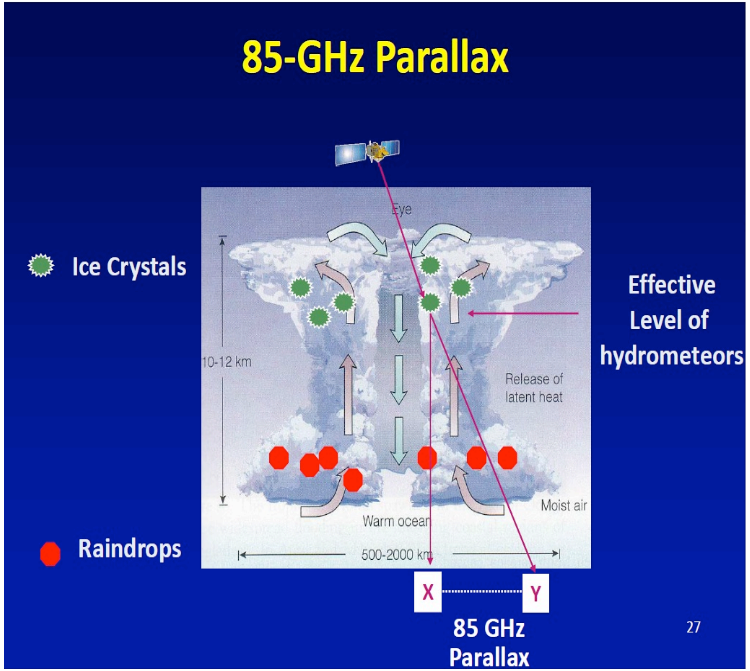

Rainfall Heavy-tailed Distribution
Normalized Frequency
Precipitation (mm/h)
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
10^{-1}
10^{0}
10^{1}
10^{2}
Baseline
Segformer
quantileUnet:median
multitaskUnet
CGAN
Deeplabv3+
Diffusion (epsilon-pred)
Diffusion (image-pred)
Target

Merci

www.latmos.ipsl.fr

Microwave Imager





Sensitivity to liquid rain drops contained in low layers :

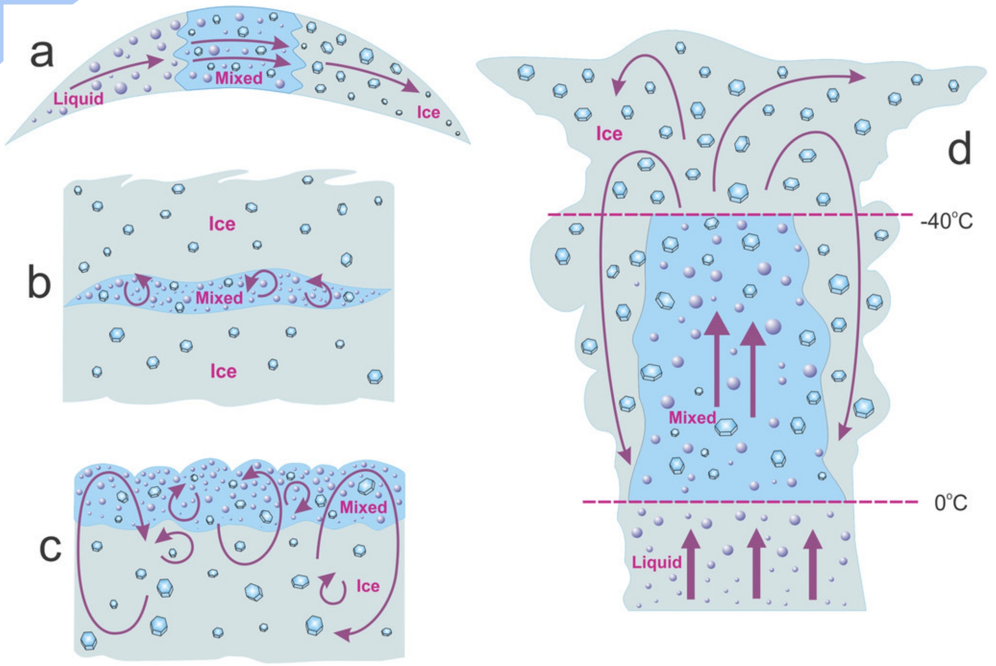
warm temperatures

small parallax error

Sensitivity to ice particles in higher atmospheric layers

Cold temperatures

High parallax error



Conceptual model of the effect of dynamic forcing on the formation of the mixed-phase in different types of cloud:

a) Wave clouds, Ac lenticular
b) frontal cloud, Cs–Ns
c) boundary layer clouds, St–Sc
d) deep convective clouds, Cb, convective storms

Meteorological Monographs 58, 1; 10.1175/AMSMONOGRAPHS-D-17-0001.1

Microwave brightness temperature (TB)

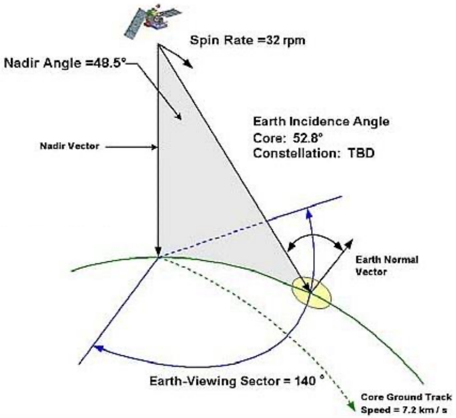


Figure 8. Scanning geometry of GMI (https://www.star.nesdis.noaa.gov/mirs/gpmgmi.php).

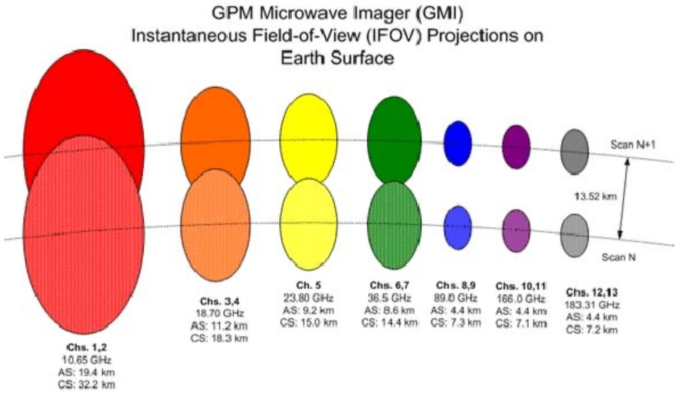


Figure 9. GMI channel pixel size (https://www.star.nesdis.noaa.gov/mirs/gpmgmi.php).

- A passive microwave radiometer measures the microwave radiant energy emitted from the Earth's surface and atmosphere.
- It is equipped with a rotating mirror that receives signals continuously. Then, the resulting signal is integrated over time to create discrete "pixels."

- The signal received is calculated as brightness temperature (TB) in Kelvin.
- A passive microwave radiometer observes the TB through multiple channels designed to be sensitive to different frequencies of microwave energy.