Reconceiving the synergy between LEO and GEO observations for satellite precipitation retrievals in the era of deep learning

Clément Guilloteau, Gavin Kerrigan, Kai Tyrus Nelson, Giosue Migliorini, Padhraic Smyth, Runze Li, Efi Foufoula-Georgiou

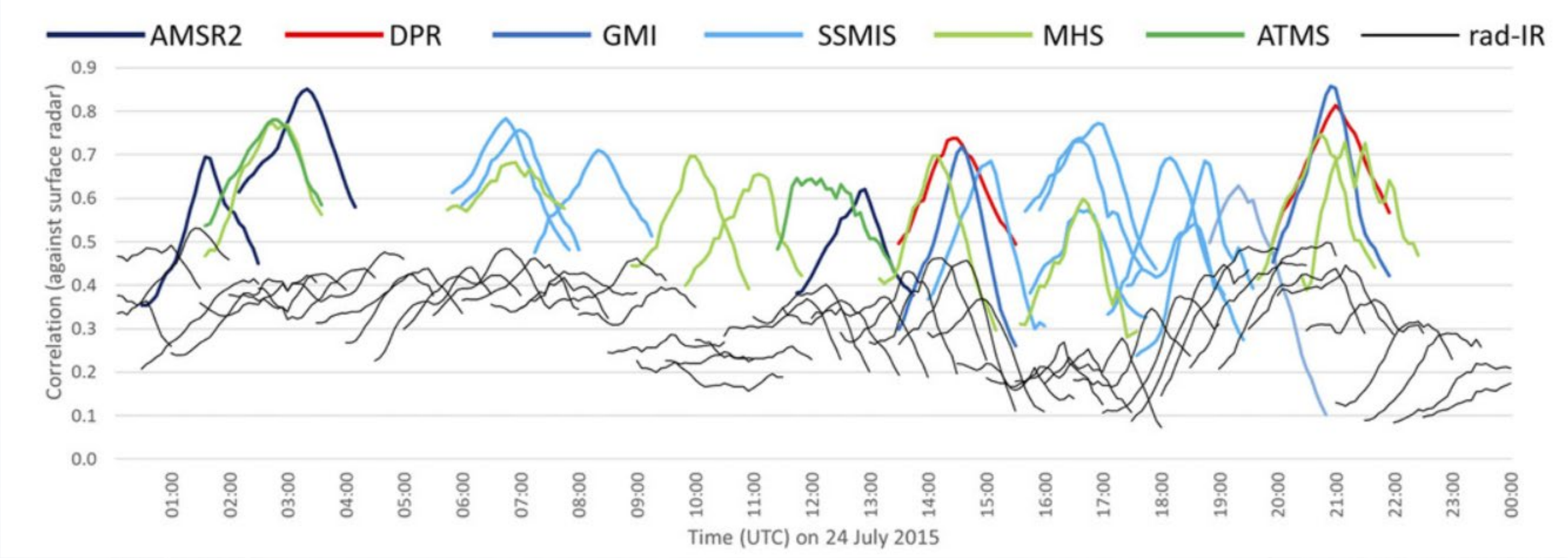
IPWG International Precipitation Working Group

2024-07-15

UCIrvine

GEO IR and LEO MW for precipitation monitoring

- In terms of surface precipitation intensity, the information content of **instantaneous** IR radiances from GEO imagers is low compared to that of passive MW radiances from LEO instruments.



Lagged correlations of MW and IR satellite estimates of precipitation rate against ground radar measurements.

From Kidd et al. 2021
"The global satellite precipitation constellation: Current status and future requirements", *BAMS*

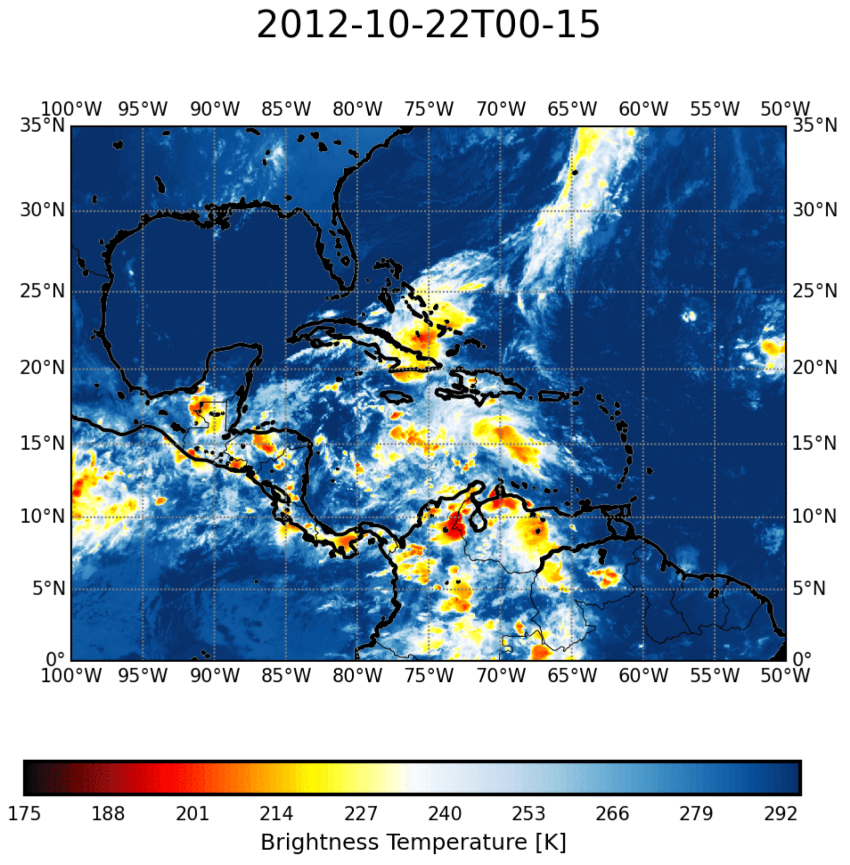
=> In state-of-the-art global precipitation mapping algorithms, the GEO information is mostly used to "fill the gaps" between LEO overpasses.

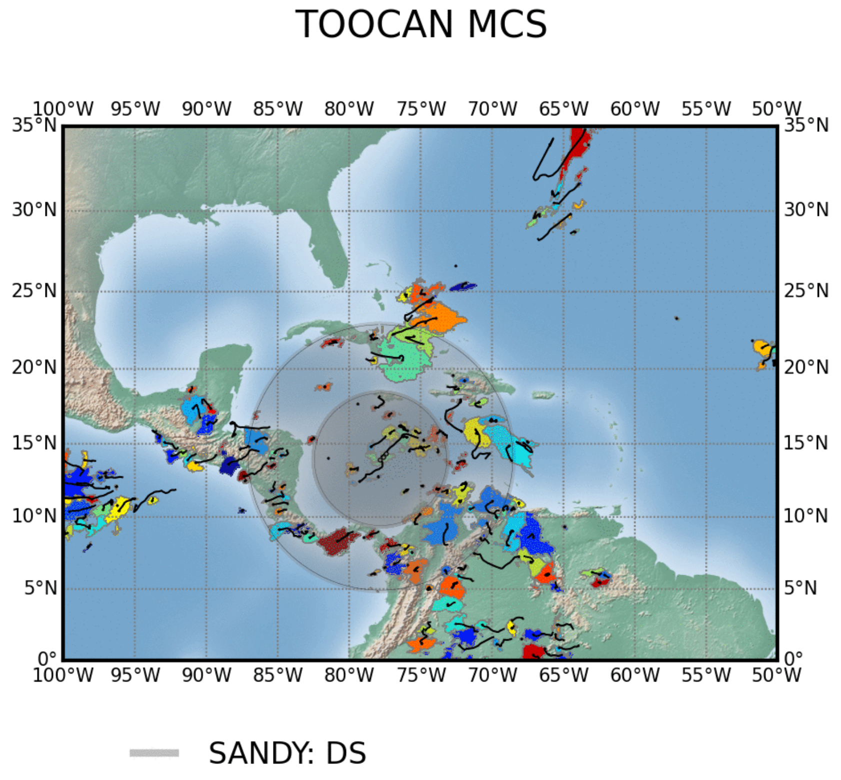
GEO-IR cloud tracking

- **TOOCAN algorithm**, Fiolleau and Roca 2013, continuous (every 30 min), global (40° N to 40° S) tracking of cloud systems.

=> Location, size, and other geometrical and dynamical properties of cloud systems can be monitored from initiation to dissipation.

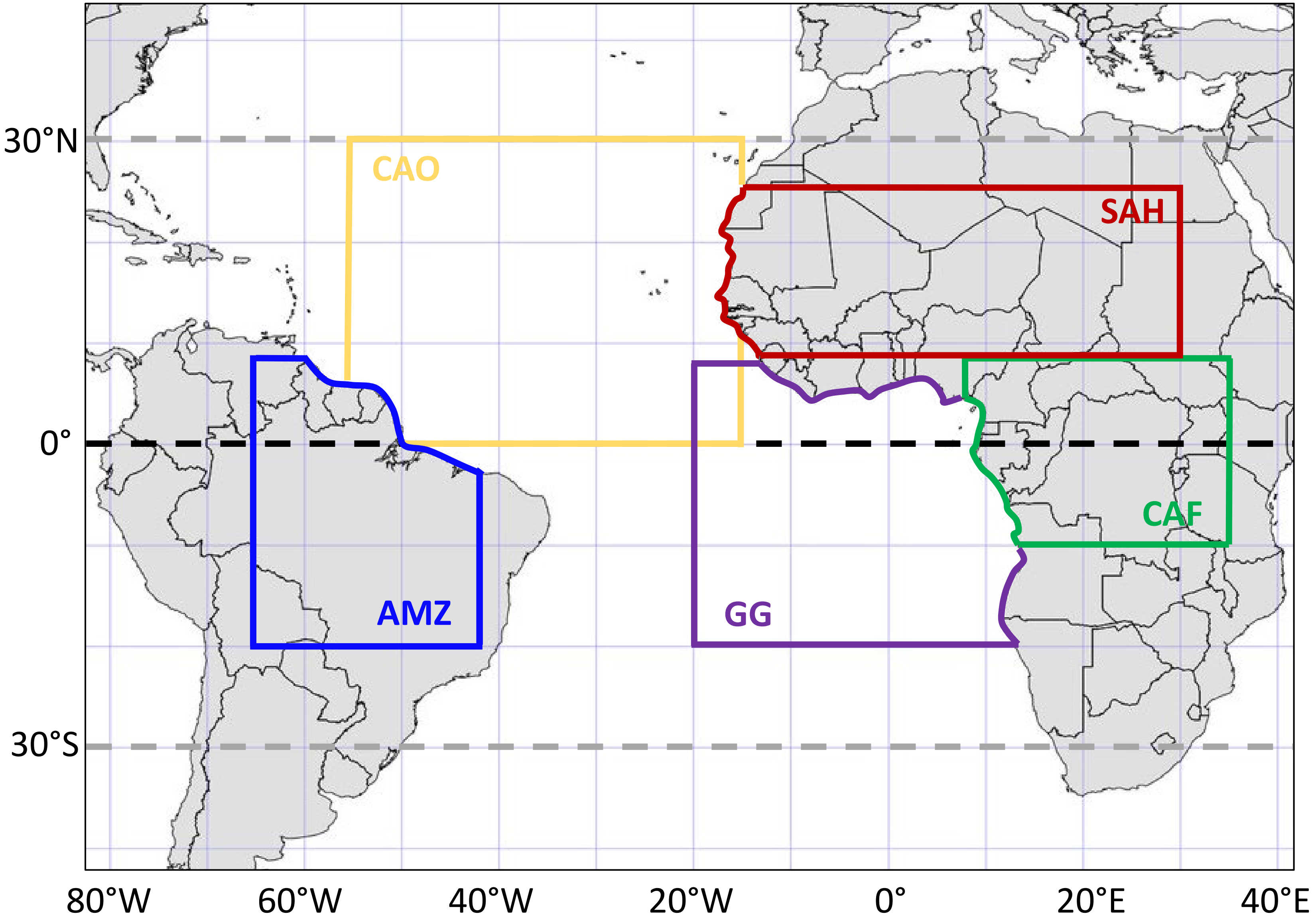
=> We have access to the whole **history** of any cloud system.





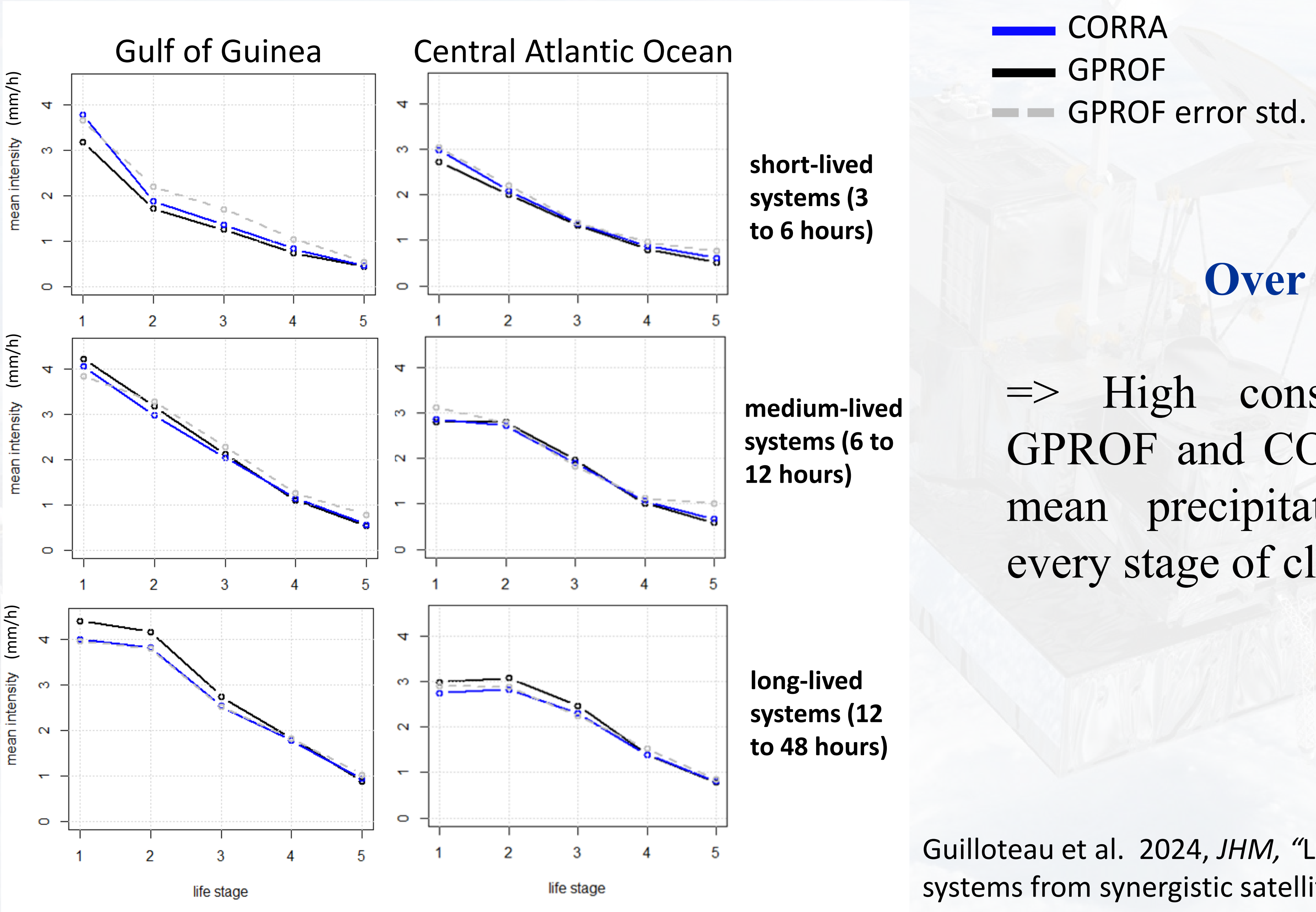
credit:
Thomas Fiolleau
toocan.ipsl.fr

Collocation between TOOCAN and GPM overpasses



- 5 study areas: Amazonia, Central Africa, Sahel, Gulf of Guinea, and Central Atlantic Ocean.
- All GPM overpasses for the 2015-2016 period.
- Around **100 000 unique cloud systems sampled** by GPM DPR and GMI.

Biases and errors of GMI-GPROF vs CORRA conditioned on cloud system information from IR GEO tracking

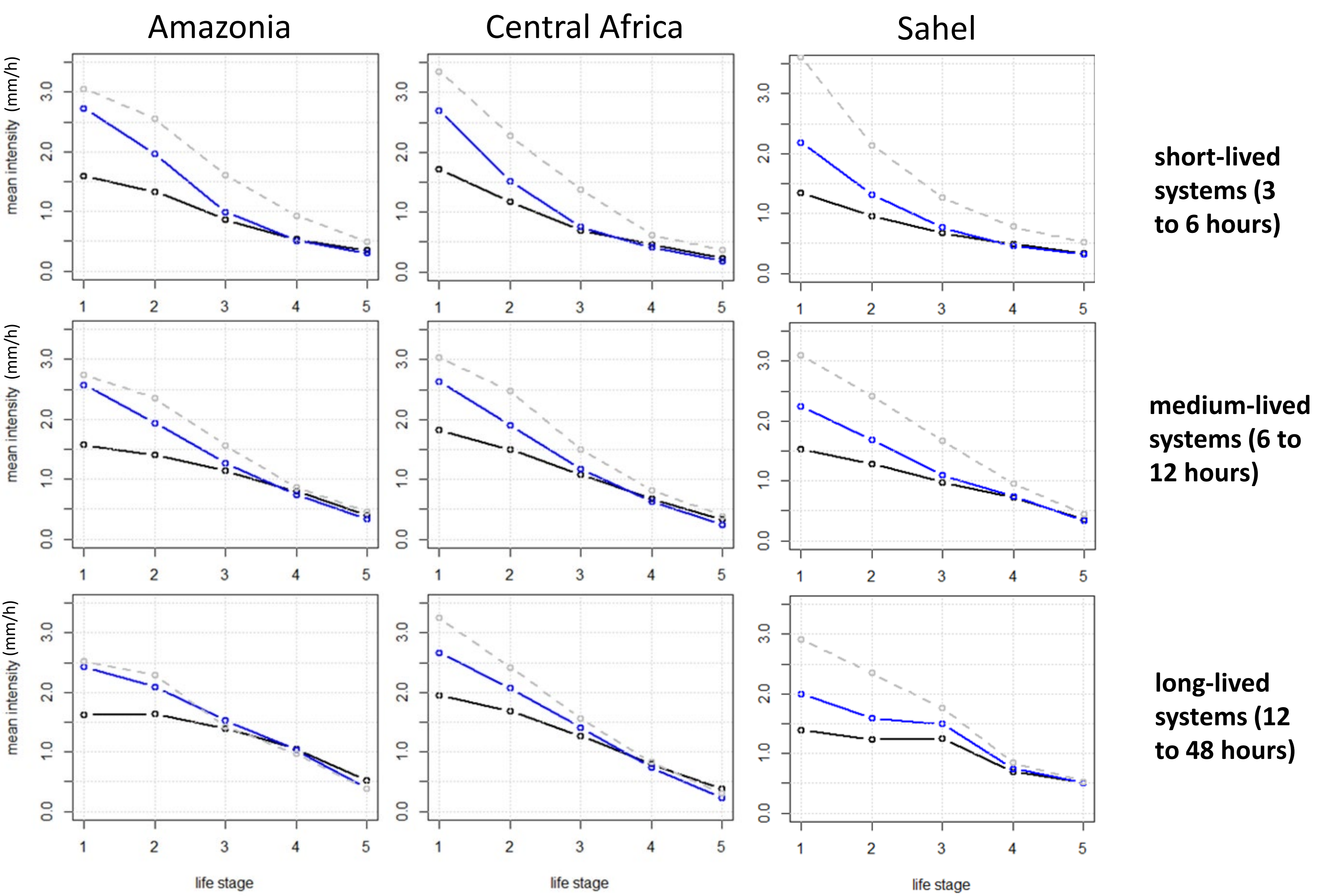


Over Ocean:

=> High consistency between GPROF and CORRA in terms of mean precipitation intensity at every stage of clouds' life cycle.

Guilloteau et al. 2024, *JHM,* "Life cycle of precipitating cloud systems from synergistic satellite observations"

Biases and errors of GMI-GPROF vs CORRA conditioned on cloud system information from IR GEO tracking



Over Land:

=> Consistent underestimation of GPROF against CORRA in the early stages of the clouds' life cycle.

Guilloteau et al. 2024, *JHM,* "Life cycle of precipitating cloud systems from synergistic satellite observations"

Reconceiving the LEO-GEO synergy

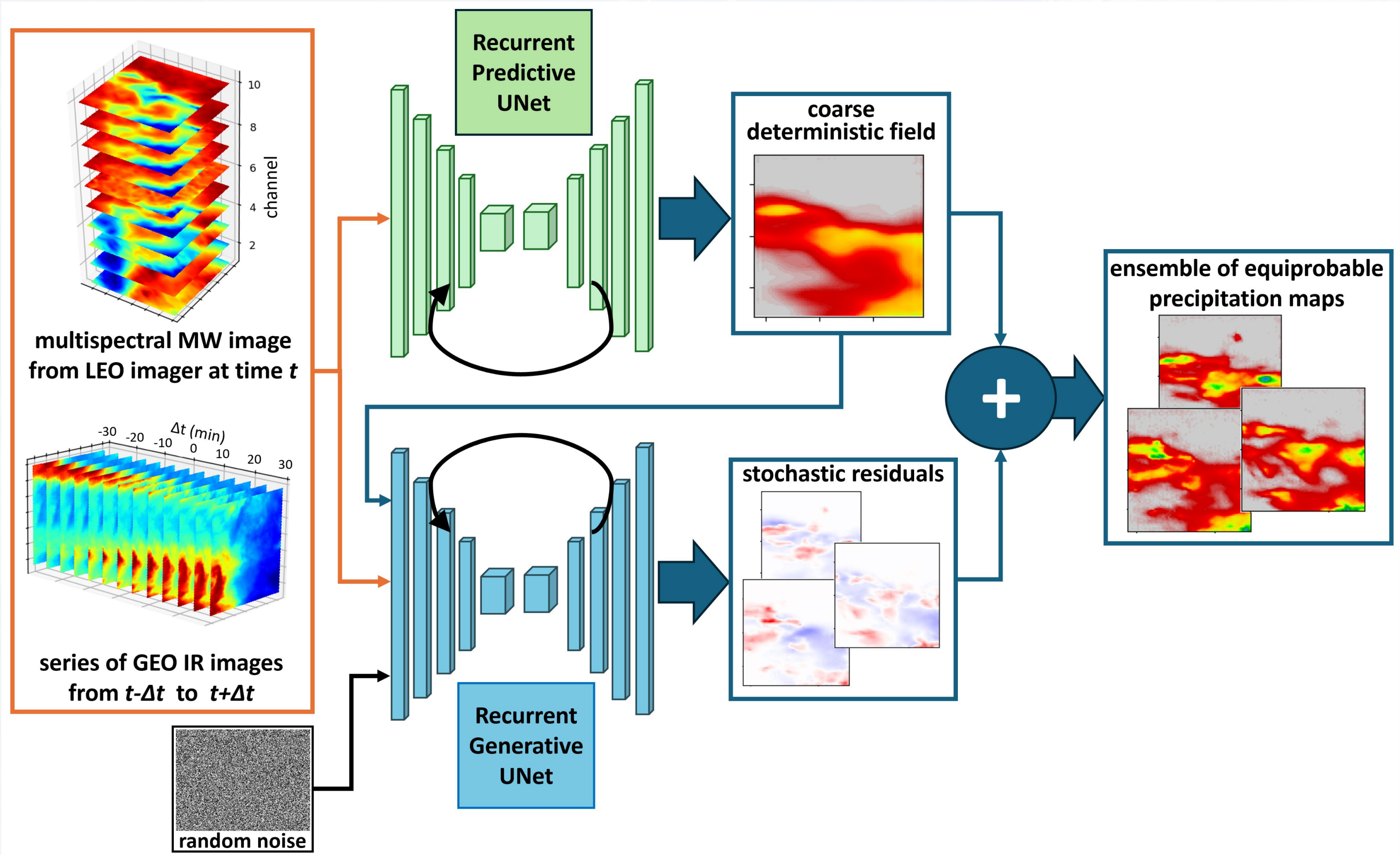
- GEO IR can to a certain degree predict some of the biases and errors of LEO MW estimates of precipitation.

⇒ Adding IR-derived information to MW TBs before inversion into precipitation intensity shall increase retrieval accuracy.

- The **temporal information in GEO-IR** can characterize the “history” of a cloud system. The GEO-IR signal shall be considered as a dynamical series rather than a set of independent instantaneous snapshots.

- The advent of the **deep learning** era offers new possibilities and a framework to “easily” combine information from various sources.

A deep generative neural network combing GEO and LEO observations



INPUTS:

- MW radiances at time *t*
- series of IR images from *t-Δt* to *t+Δt*

OUTPUTS:

ensemble of accumulated precipitation maps from *t-Δt* to *t+Δt*

ENSEMBLE GENERATION:

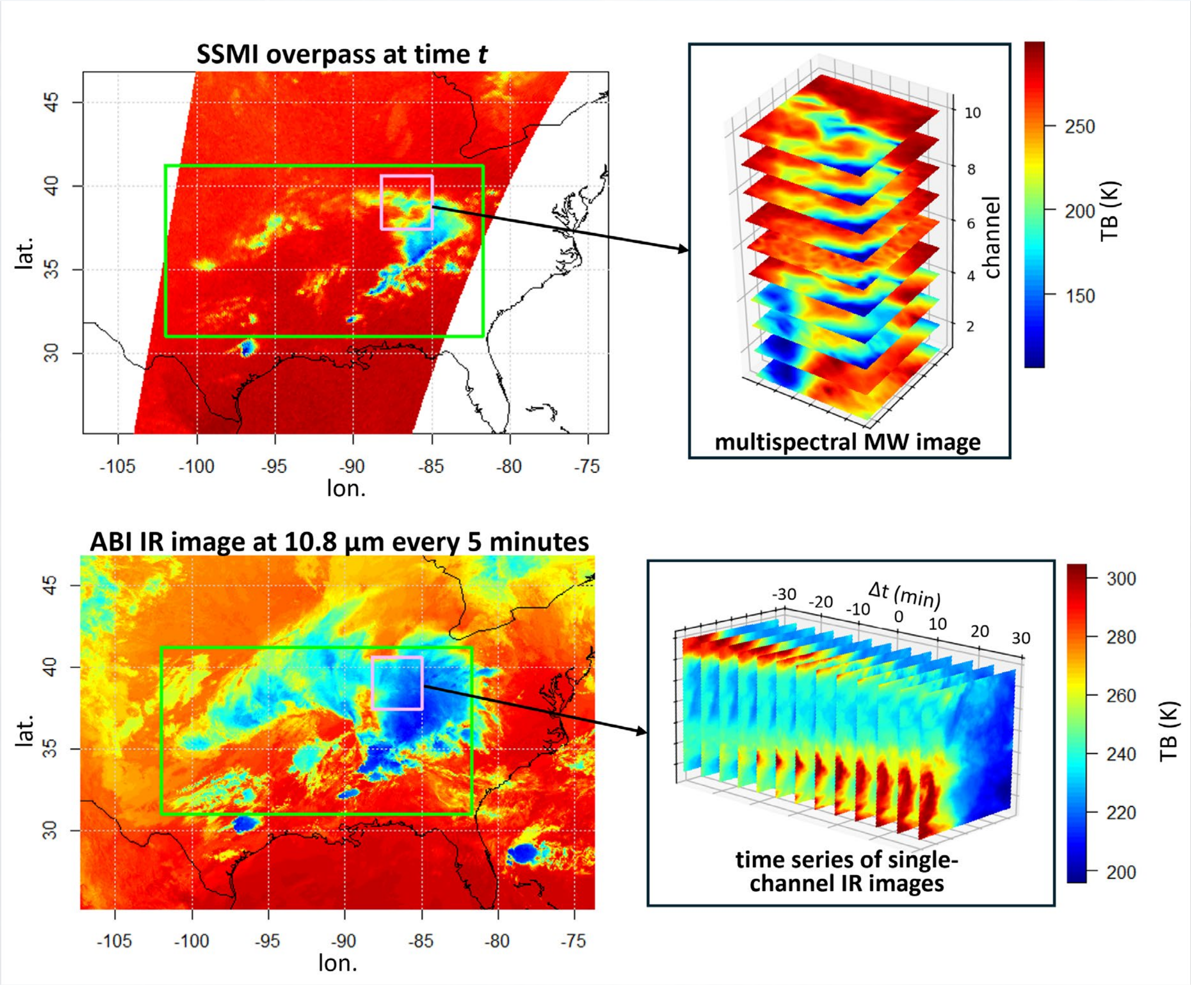
A **conditional generative diffusion model** is used in an ensemblist approach.

The ensemble members are considered as equiprobable possible realizations of the “true” precipitation field.

The ensemble mean (EM) is considered as the *a posteriori mean* Bayesian estimator (MMSE).

The ensemble dispersion is a measure of uncertainty.

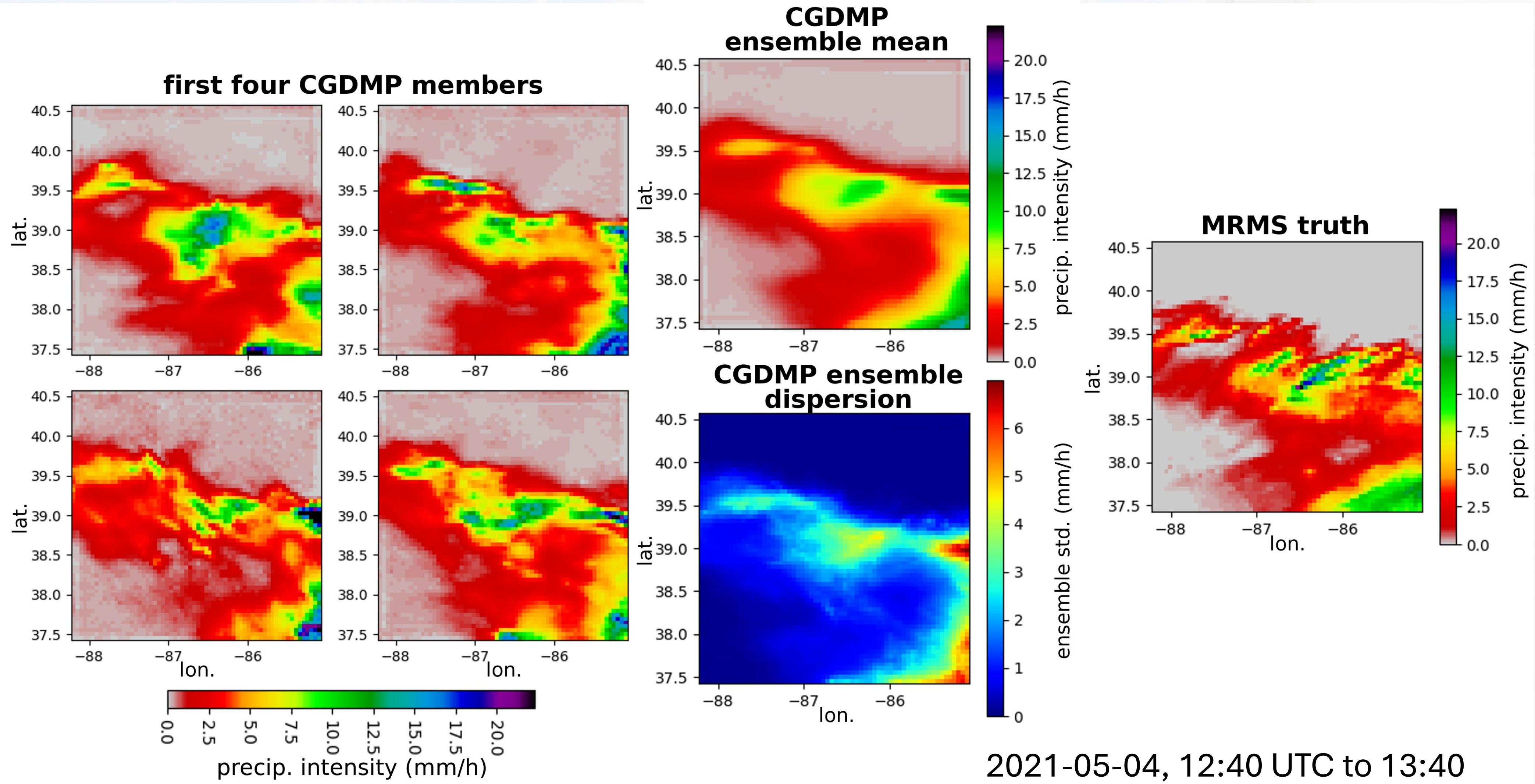
Experimental setup



- MW at time *t* from DMSP-F17 SSMI/S
- IR 10.3 µm every 5 min from *t* – 30min to *t* + 30min from GOES-16 ABI
- Target: one-hour accumulated precipitation field from *t* – 30min to *t* + 30min at 5.5 km resolution
- All inputs projected on a 5.5-km resolution grid
- Retrieval performed on 64 by 64 pixels patches (350 km by 350 km)
- Trained and tested against MRMS gauge-radar fields over southeastern US
- 3 years of data (2021-2023), every fifth SSMI/S overpass used for testing and the rest for training
- 128-member ensemble generation

Examples of generated precipitation fields

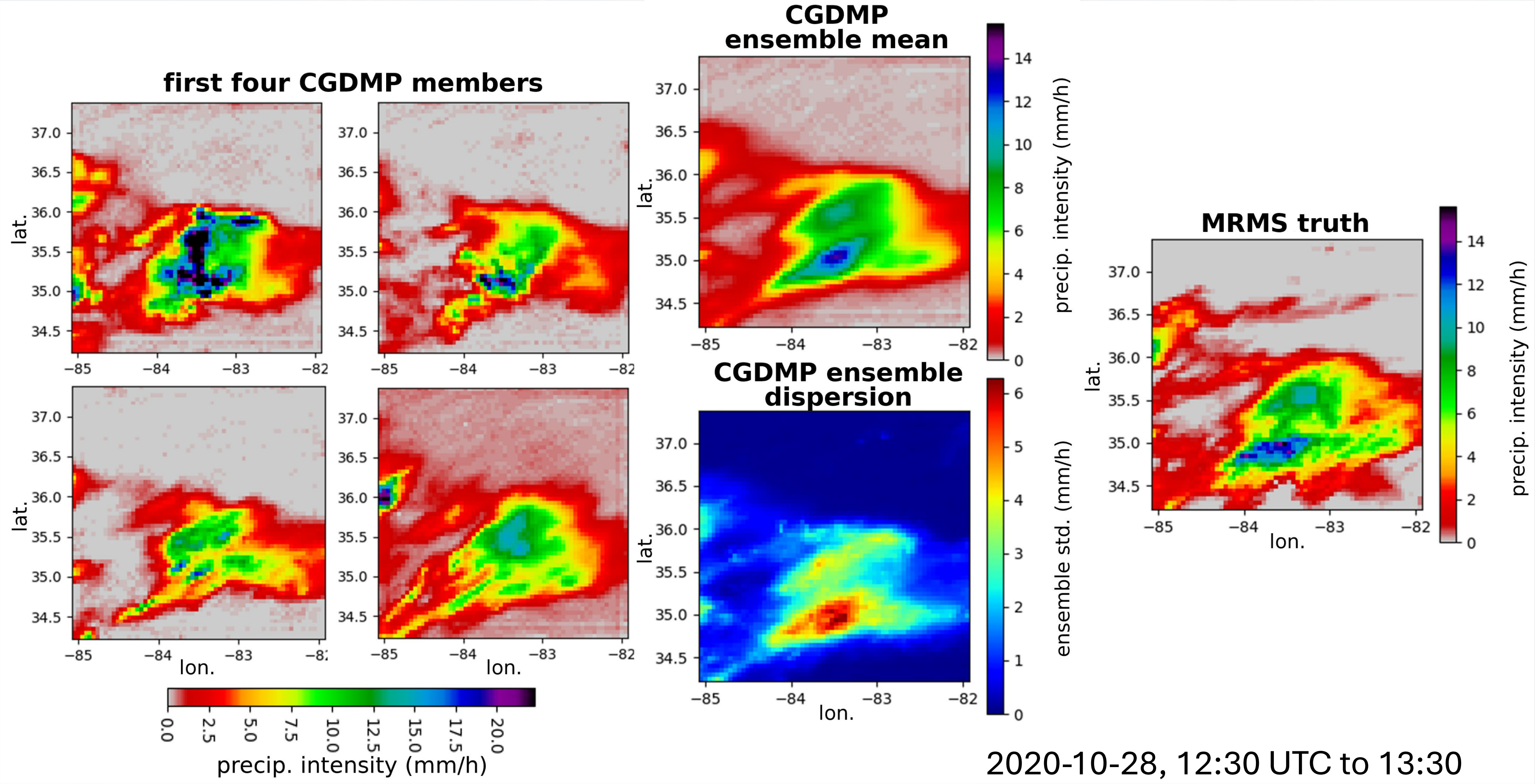
Conditional Generative Diffusion Model for mapping Precipitation intensity (CGDMP) over southeastern US (testing dataset separate from the training dataset)



2021-05-04, 12:40 UTC to 13:40 UTC

Examples of generated precipitation fields

Conditional Generative Diffusion Model for mapping Precipitation intensity (CGDMP) over southeastern US (testing dataset separate from the training dataset)

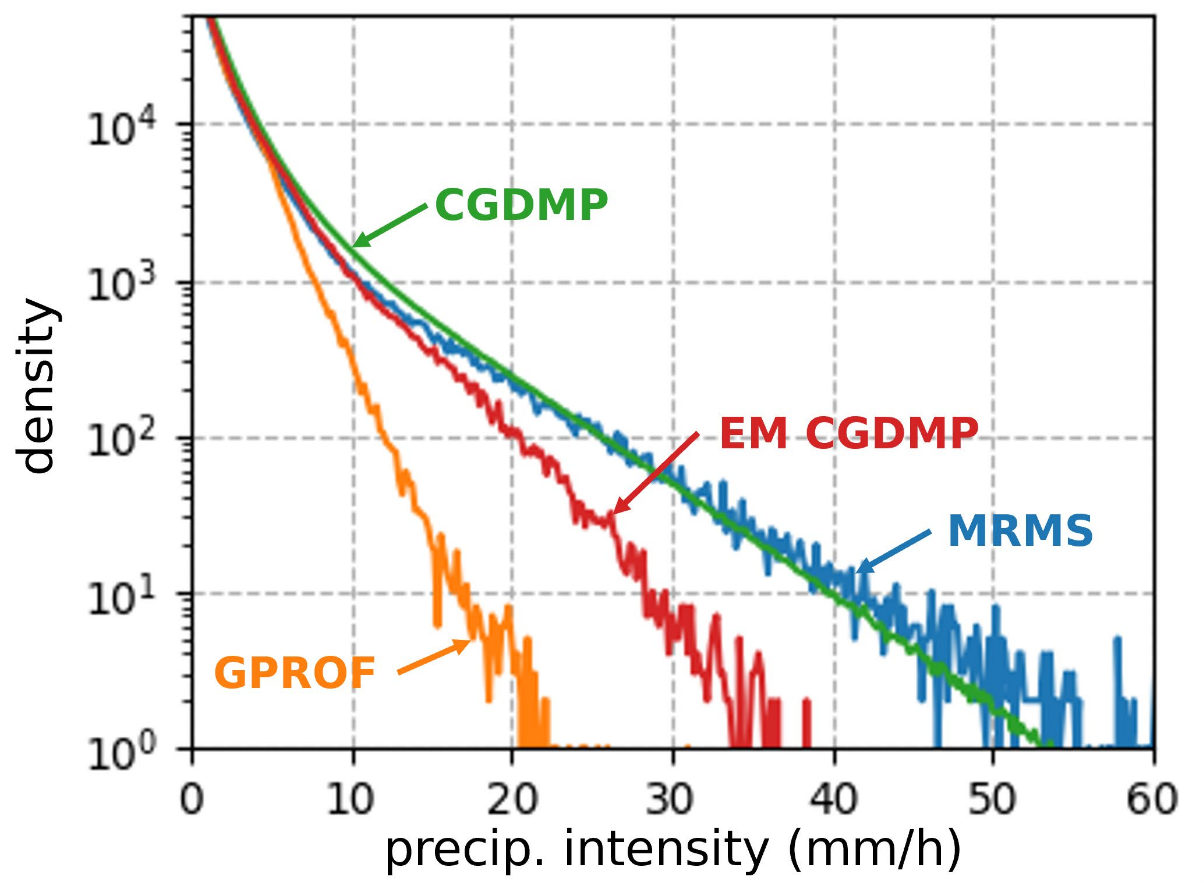


2020-10-28, 12:30 UTC to 13:30 UTC

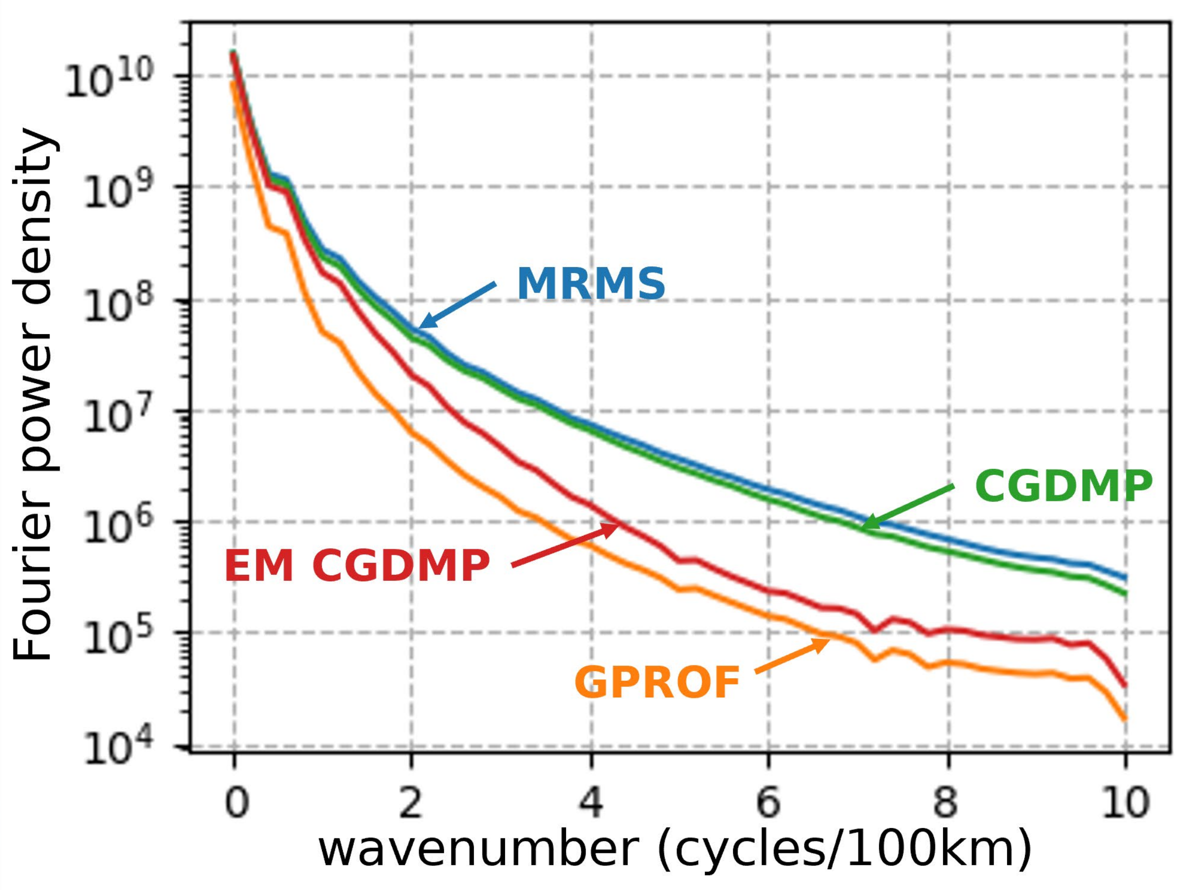
How realistic looking are the generated fields?

Conditional Generative Diffusion Model for mapping Precipitation intensity (CGDMP) over southeastern US (testing dataset separate from the training dataset)

distribution of hourly 5-km precipitation intensities



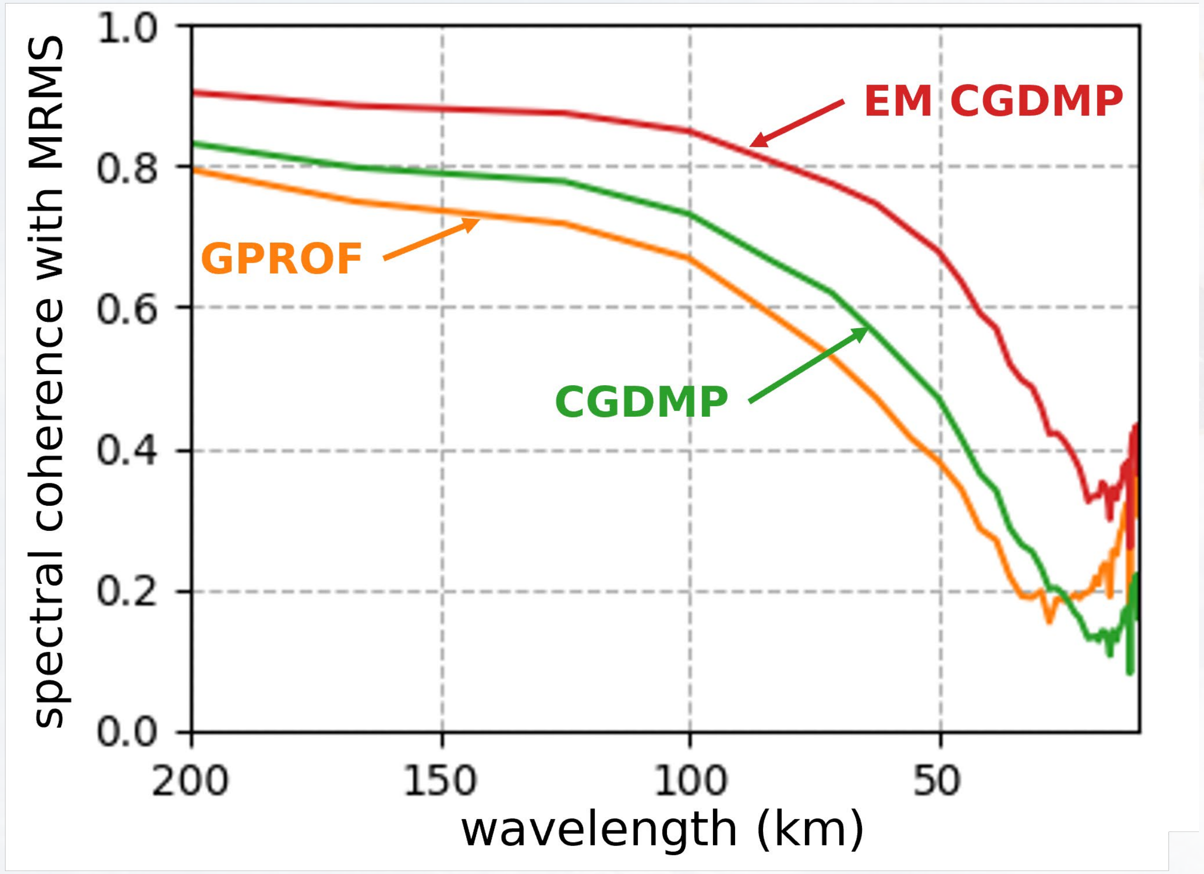
spatial Fourier power spectra of the intensity fields



How coherent are the generated fields with the MRMS "truth"?

Conditional Generative Diffusion Model for mapping Precipitation intensity (CGDMP) over southeastern US (testing dataset separate from the training dataset)

spatial Fourier coherence with MRMS

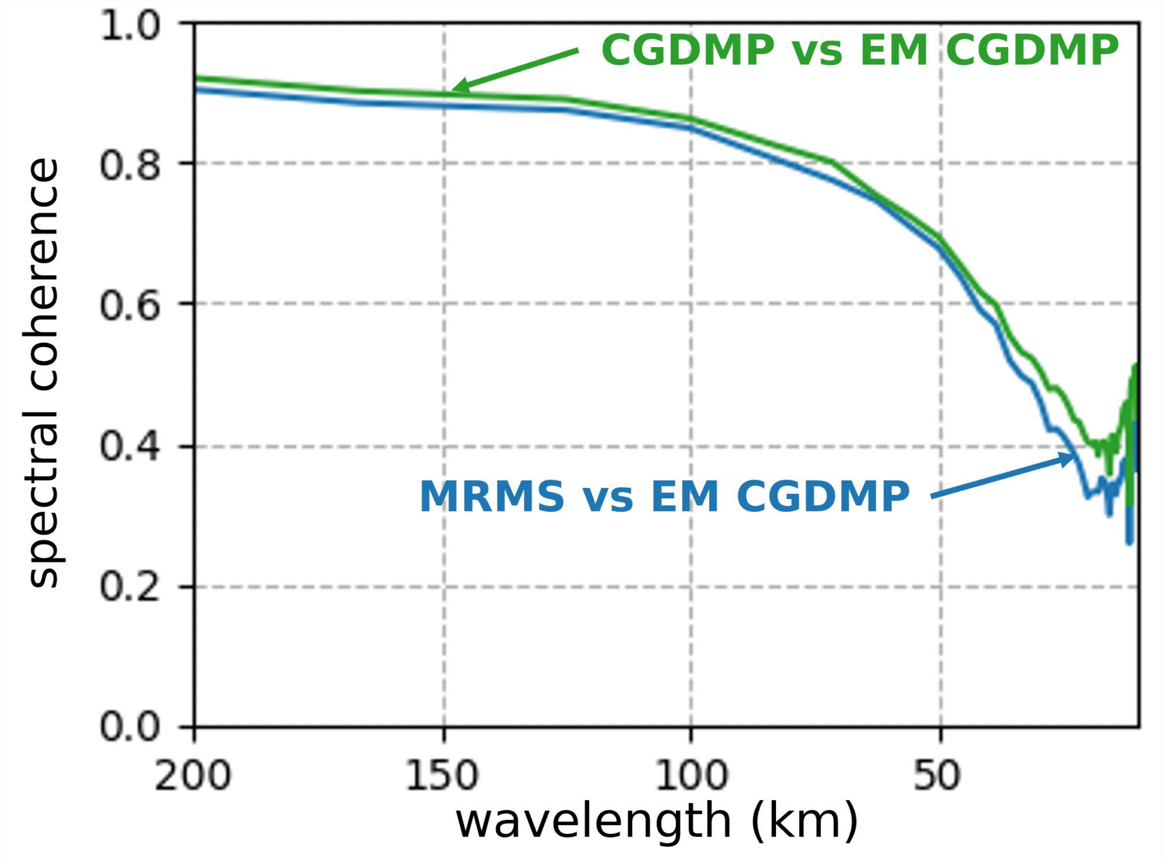


	EM CGDMP	GPROF
CC with MRMS	0.82	0.67
MSE vs MRMS	2.48 mmh^{-2}	4.31 mmh^{-2}

While the individual members are more realistic looking and preserve the statistical properties of the precipitation fields, the smooth ensemble mean is closer to the truth in terms of MSE, coherence and correlation.

Estimating retrieval uncertainty from the ensemble members

spatial Fourier coherence of the individual members with the ensemble mean



- The coherence between any ensemble member and the ensemble mean is similar to the coherence between the MRMS truth and the ensemble mean.

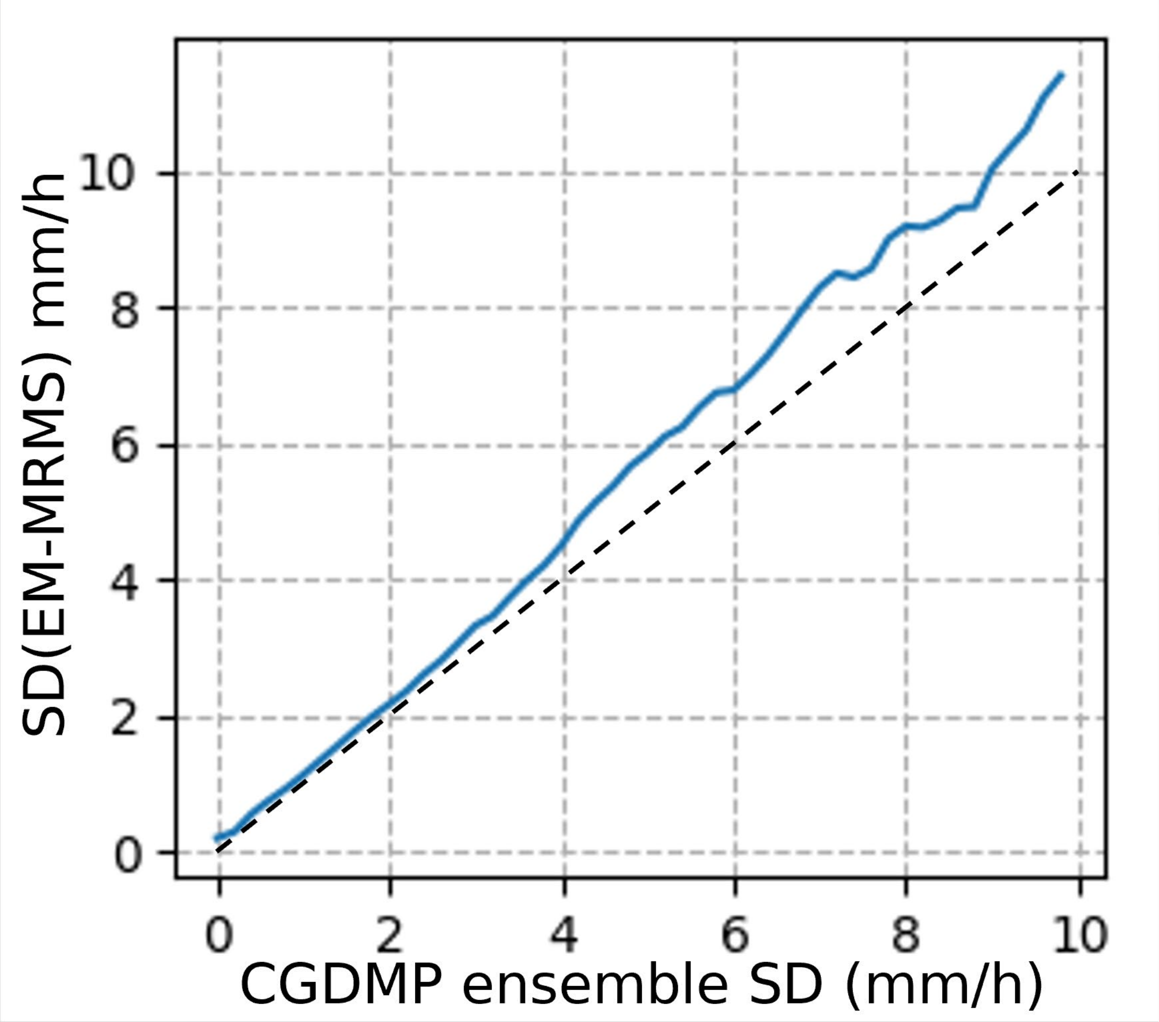
⇒ If we were to "hide" the MRMS truth among the ensemble members, it would be statistically indistinguishable from them.

⇒ We can use the dispersion of the ensemble members as a measure of uncertainty.

Estimating retrieval uncertainty from the ensemble members

- We sort the CGDMP pixels into bins depending on their ensemble SD (50 bins between 0 and 10 mm/h).
- For each bin we compute the empirical SD of the error (difference between CGDMP-EM and MRMS).

empirical SD of the error against CGDMP ensemble dispersion



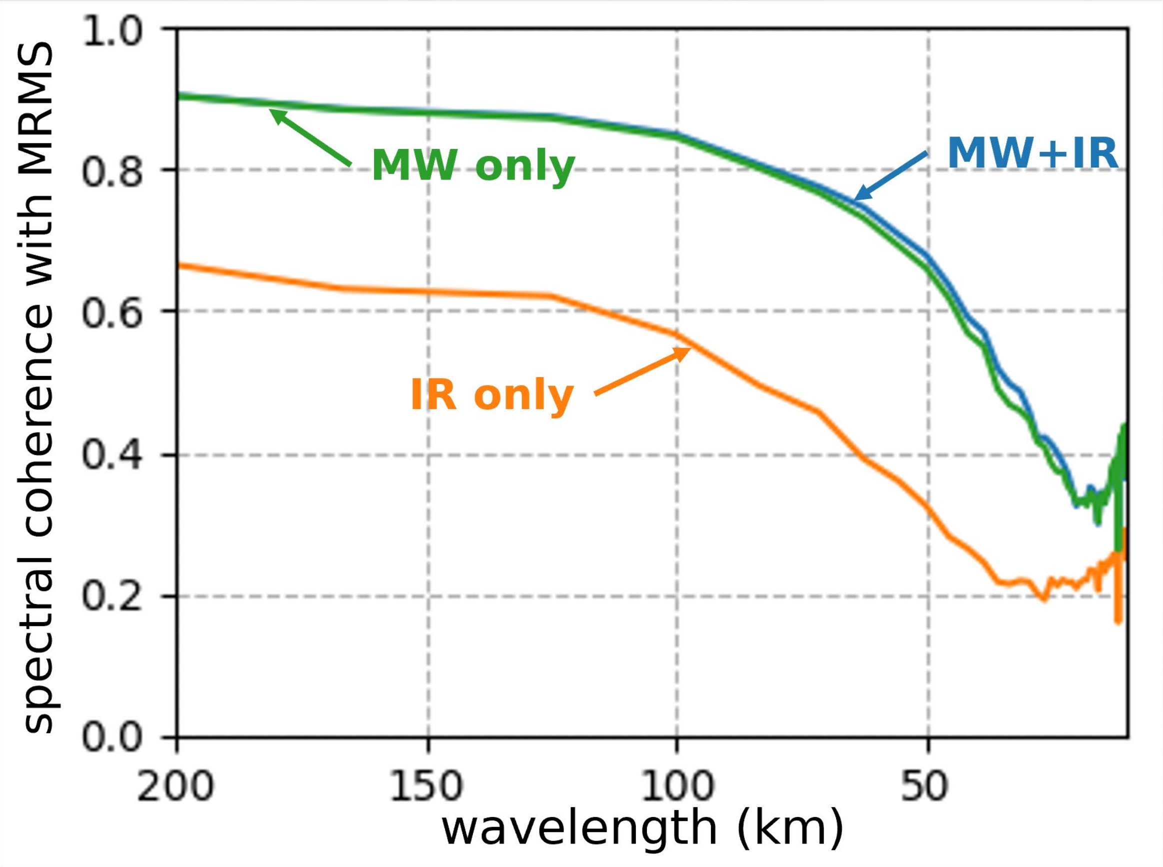
Is the IR really useful?

The CGDMP was also trained and tested using the MW data only and using the IR data only.

	CC with MRMS	MSE vs MRMS
MW+IR	0.82	2.48
MW only	0.81	2.57
IR only	0.54	5.43

statistics for the ensemble mean

spatial Fourier coherence of CGDMP ensemble mean with MRMS



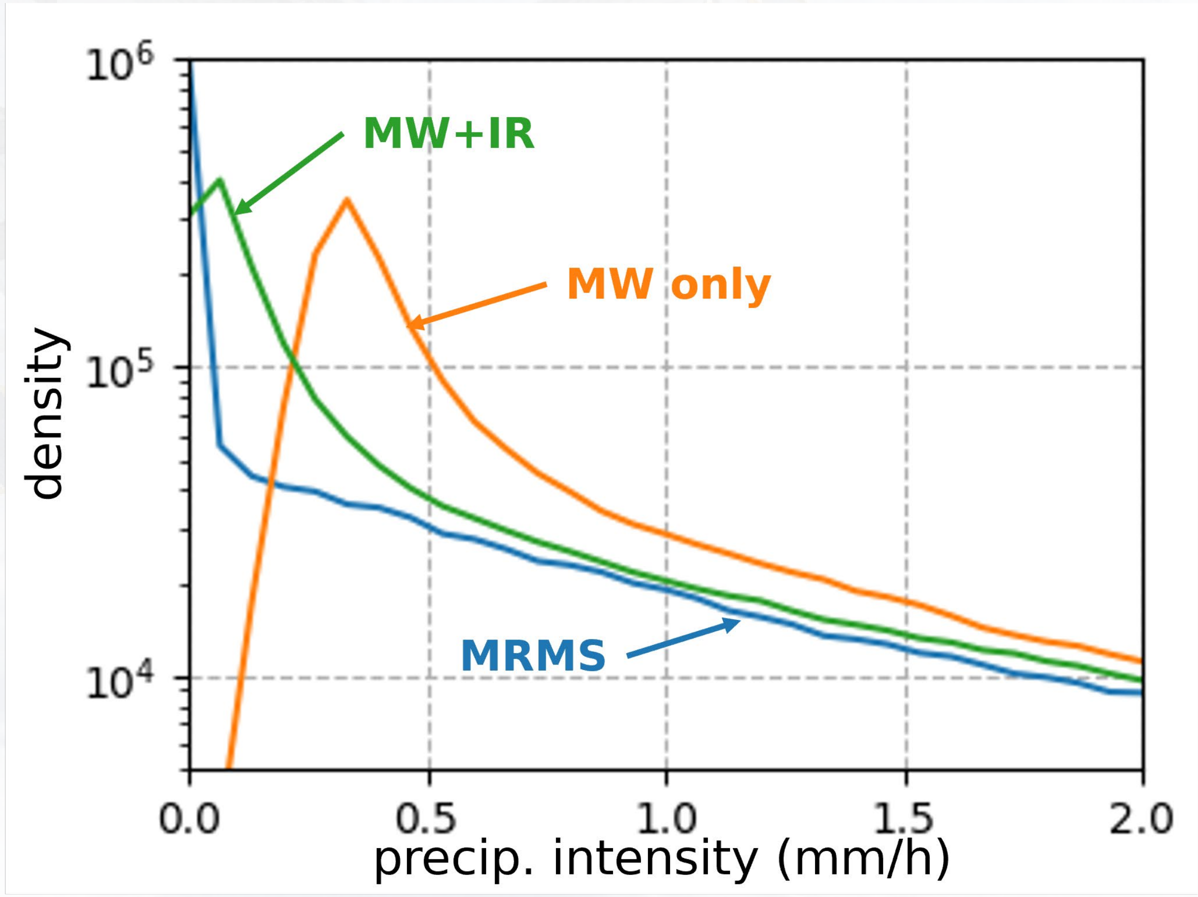
Is the IR really useful?

The CGDMP was also trained and tested using the MW data only and using the IR data only.

	CC with MRMS	MSE vs MRMS
MW+IR	0.82	2.48
MW only	0.81	2.57
IR only	0.54	5.43

statistics for the ensemble mean

distribution of low-intensity precipitation rates in the ensemble mean fields



Conclusion

- The **deep learning revolution** offers new possibilities and a framework to "easily" combine information from various sources; even asynchronous observations across different scales.
- Generative deep learning allows to handle uncertainty through **ensemble generation**.
- Deep learning can produce impressive results when adequate data is available for training.