VALIDATING SATELLITE PRECIPITATION PRODUCTS

Viviana Maggioni, Ph.D.

Department of Civil, Environmental & Infrastructure Engineering

https://maggioni.vse.gmu.edu/

GEORGE MASON UNIVERSITY

INTRO

Precipitation products that are based on satellite retrievals are affected by errors and uncertainty

<u>Errors</u>: unintended, generally small and known problems that can (and should) be fixed

<u>Uncertainties</u>: those we do not (fully) understand

Quantifying such errors and uncertainties is essential for the appropriate use of satellite precipitation products in any applications

VALIDATION

Defined by the Committee on Earth Observing Satellites Working Group on Calibration and Validation (CEOS-WGCV) as "the process of assessing by independent means the quality of the data products derived from the system outputs"

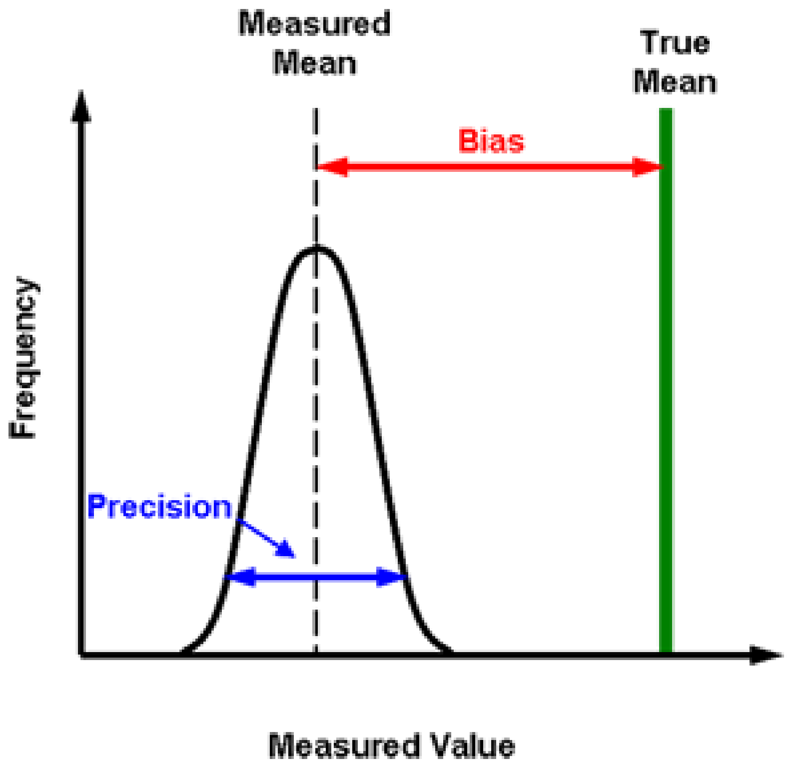
WHAT TYPE OF ERRORS/UNCERTAINTIES?

- **Sensor errors and uncertainties**: physical limitations of engineering and knowledge
- **Retrieval scheme errors and uncertainties**: assumptions, information utilization, and the mechanisms of the retrieval algorithm itself
- **Product errors and uncertainties**: progression from instantaneous to daily/monthly products, temporal and spatial sampling, and inheritance of errors and uncertainties

ERROR vs UNCERTAINTY

Error:

- Commonly defined as the difference between the satellite product and a reference considered to be the "truth"
- Characterized by:
  - a *systematic component* – a shift of the mean from the reference
  - a *random component* – which varies in an unpredictable way



Uncertainty: Represents the range of values within which the true value lies with some level of confidence

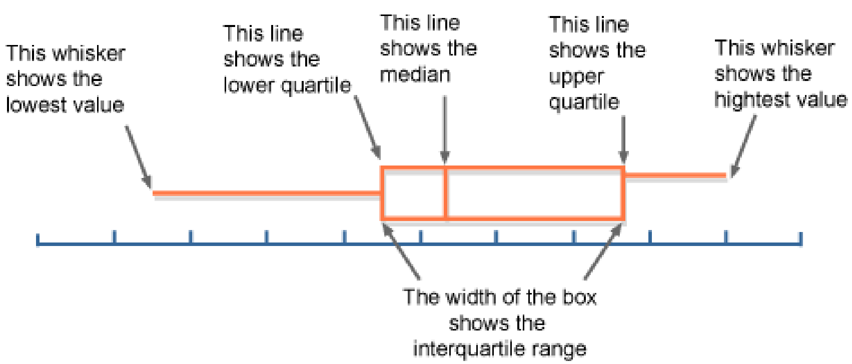
SYSTEMATIC ERRORS

- Systematic errors are not determined by chance but are introduced by an inaccuracy inherent in the system (e.g., instruments bias).

- Systematic errors are difficult to detect because all data are off in the same direction (either too high or too low).

RANDOM ERRORS

- Random errors are caused by unknown and unpredictable causes.
  - Random errors often have a known distribution (e.g., Gaussian). In such cases statistical methods may be used to analyze the data.
  - For Gaussian distributions, the mean *m* of a number of measurements of the same quantity is the best estimate of that quantity, and the standard deviation *s (or sigma)* of the measurements shows the accuracy of the estimate.

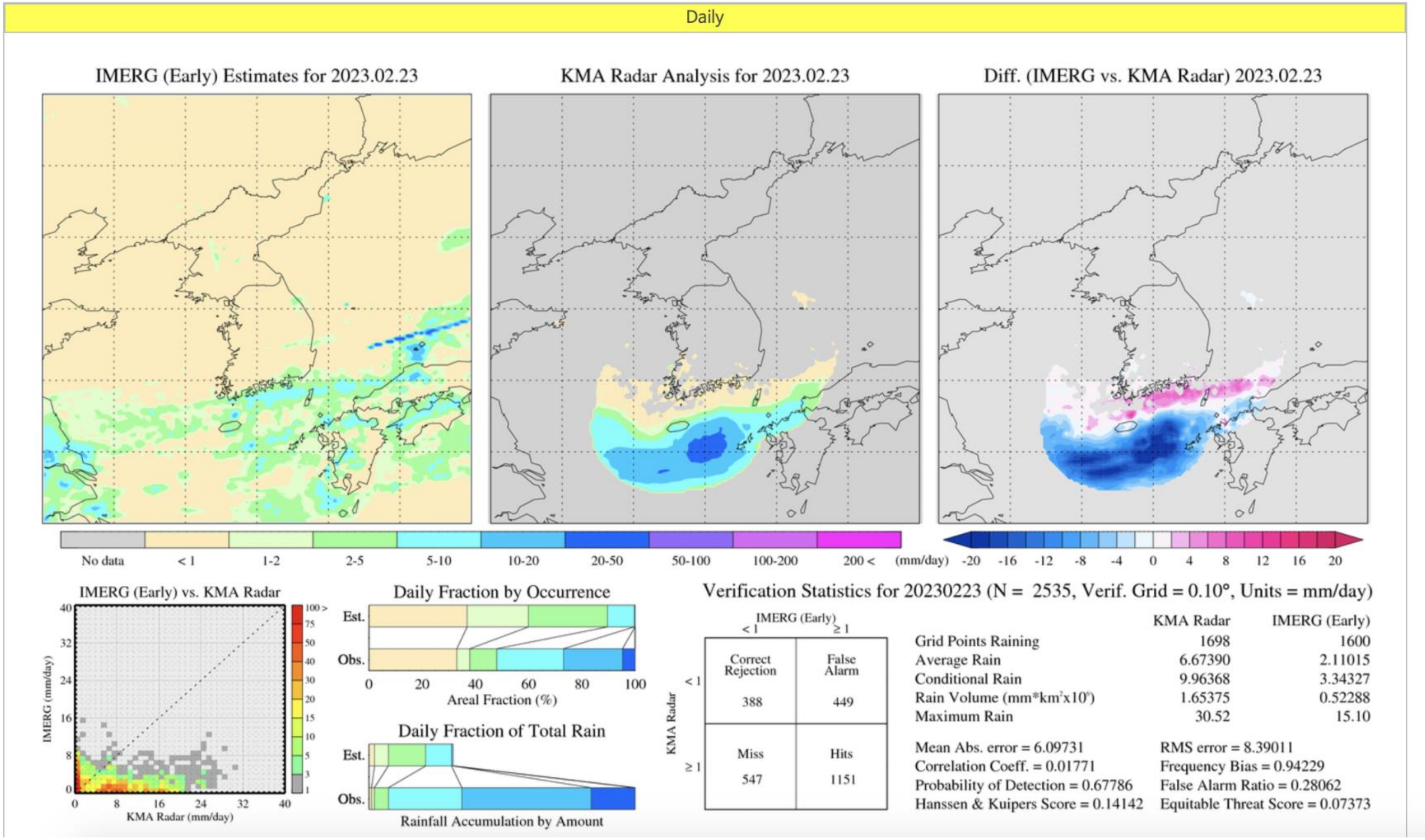


VALIDATION STEPS

1. **Sampling design**, indicating how many spots must be sampled.
2. **Collecting reference data** over the selected spots to be sampled.
3. **Extracting information** from the collected data.
4. **Comparing** reference and estimate (i.e., remote sensing product) using appropriate statistical techniques
5. **Analyzing the causes** of uncertainty and error distributions.

HOW DO WE ASSESS SATELLITE PRODUCTS?

Quantifying the "distance" between the satellite estimate and the true precipitation → VALIDATION



http://nmsc.kma.go.kr/enhome/html/ipwg/viewer/selectIpwg.do

HOW DO WE ASSESS SATELLITE PRODUCTS?

- **Categorical verification metrics** measure the correspondence between the estimated and observed occurrence of events through a contingency table
- **Continuous metrics** measure the accuracy of a continuous variable, such as rain amount or intensity
- Both types of scores provide a measure of the precipitation error/uncertainty in a conceptually different manner therefore any meaningful validation should assess both.

CATEGORICAL METRICS

- Often based on the use of a contingency table (also known as confusion matrix) that displays the (multivariate) frequency distribution of two variables, reference and satellite SPP).

		Satellite	
Reference		RSat $\geq$ th	RSat < th
	RRef $\geq$ th	H	M
	RRef < th	F	Z

- **N: sample size,** i.e., the total number of observed events

$$N = H + M + F + Z$$

- **H: hit cases**, when both SPP and reference $\geq$ rain/no-rain threshold (th)
- **F: false alarms**, when SPP $\geq$ th, but reference < th
- **M: missed events**, when the reference $\geq$ th, but SPP < th
- **Z: correct no-rain detection**, when both SPP and reference < th

CATEGORICAL METRICS (CONT'D)

- **Hit rate**: HR = (H + Z)/N
  - it credits correct detection (whether rain or no rain) cases equally
  - not recommended for extreme event analysis, where the number of correct zeroes is high, resulting in a high score, even if the probability of correctly detecting an extreme event (e.g., tornado) is low
  - perfect H value is 1

- **Critical success index or threat score**: CSI =TS = H/(H + M + F)
  - it is a better metric than HR when events are rare, since it removes the effect of Z
  - perfect value of CSI/TS is 1

- **Probability of detection**: POD = H/(H + M)
  - it measures the likelihood of SPP to detect an event when it in fact occurs
  - perfect value of CSI/TS is 1

		Satellite	
Reference		RSat $\geq$ th	RSat $<$ th
	RRef $\geq$ th	H	M
	RRef $<$ th	F	Z

CATEGORICAL METRICS (CONT'D)

- **<u>Success Ratio</u>**: SR = H/(H + F)
  - it measures the likelihood of SPP being correct, when detecting rain
  - perfect SR value is 1

- **<u>False alarm ratio</u>**: FAR = F/(F + H)
  - it measures the likelihood that a precipitation event does not occur when SPP estimates rain
  - perfect FAR value is zero

- **<u>Frequency Bias</u>**: FB = (H + F)/(H + M)
  - it measures the total number of events estimated by SPP divided by the total number of events observed by the reference
  - The perfect FB value is 1

		Satellite	
Reference		RSat $\geq$ th	RSat $<$ th
	RRef $\geq$ th	H	M
	RRef $<$ th	F	Z

CONTINUOUS METRICS

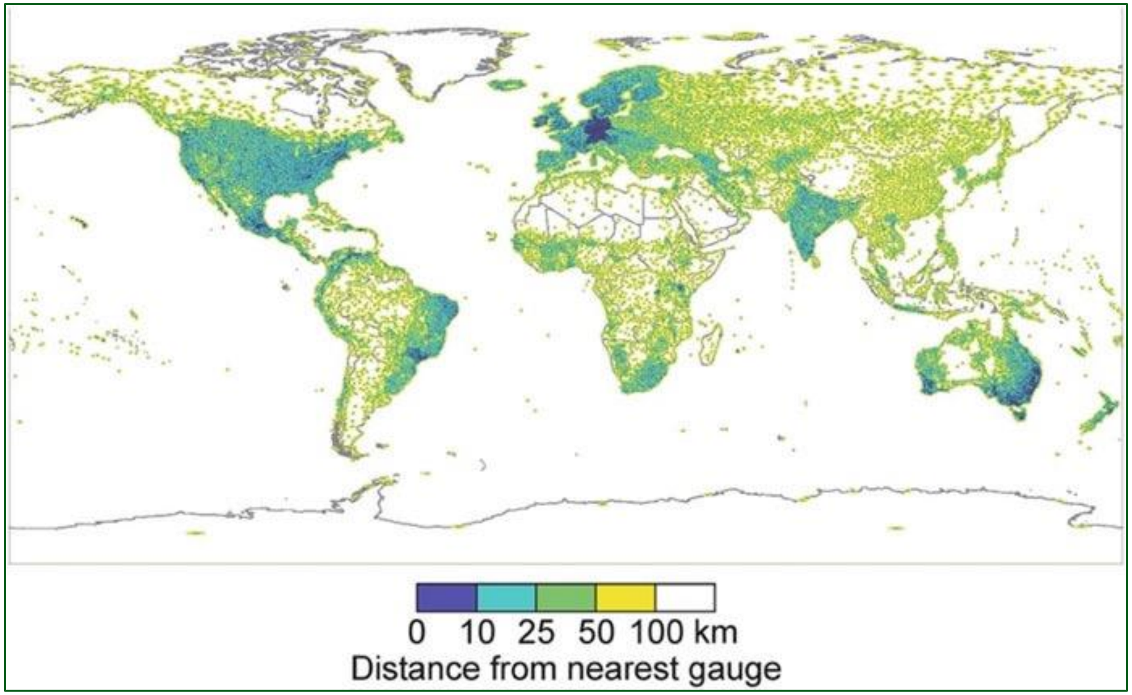
- **Bias**: The difference between the SPP mean and the mean of the reference observations. Also known as overall bias, systematic bias, or unconditional bias.
  - perfect score is zero
  - units of precipitation
- **Bias ratio**: The ratio between the SPP mean and the mean of the reference observations.
  - perfect score is 1
  - unitless
- **Correlation Coefficient**: A measure of the linear association between SPP and reference independent of the mean and variance of the marginal distributions. Pearson Correlation Coefficient and Spearman Rank Correlation are the most widely used ones.
  - perfect score is 1
  - unitless

CONTINUOUS METRICS (CONT'D)

- **Mean Absolute Error (MAE)**: The average of the absolute differences between SPP and reference
  - considered a more robust measure than MSE that is sensitive to large outlier errors
  - perfect score is 0
  - units of precipitation
- **Mean Error (ME)**: The average difference between SPP and reference.
  - perfect score is zero
  - units of precipitation
- **Root Mean Square Error (RMSE)**: The square root of the average of the squared differences between SPP and reference.
  - puts a greater influence on large errors than smaller errors, which may be good if large errors are especially undesirable, but may also encourage conservative precipitation estimation.
  - perfect score is zero
  - units of precipitation

REFERENCE? WHAT REFERENCE?

- We do not have knowledge of the "true" precipitation field → validation is commonly carried out using an independent reference or benchmark, such as observations from rain gauges and/or ground radars, assuming that they are characterized by a much lower error than the satellite-based products.

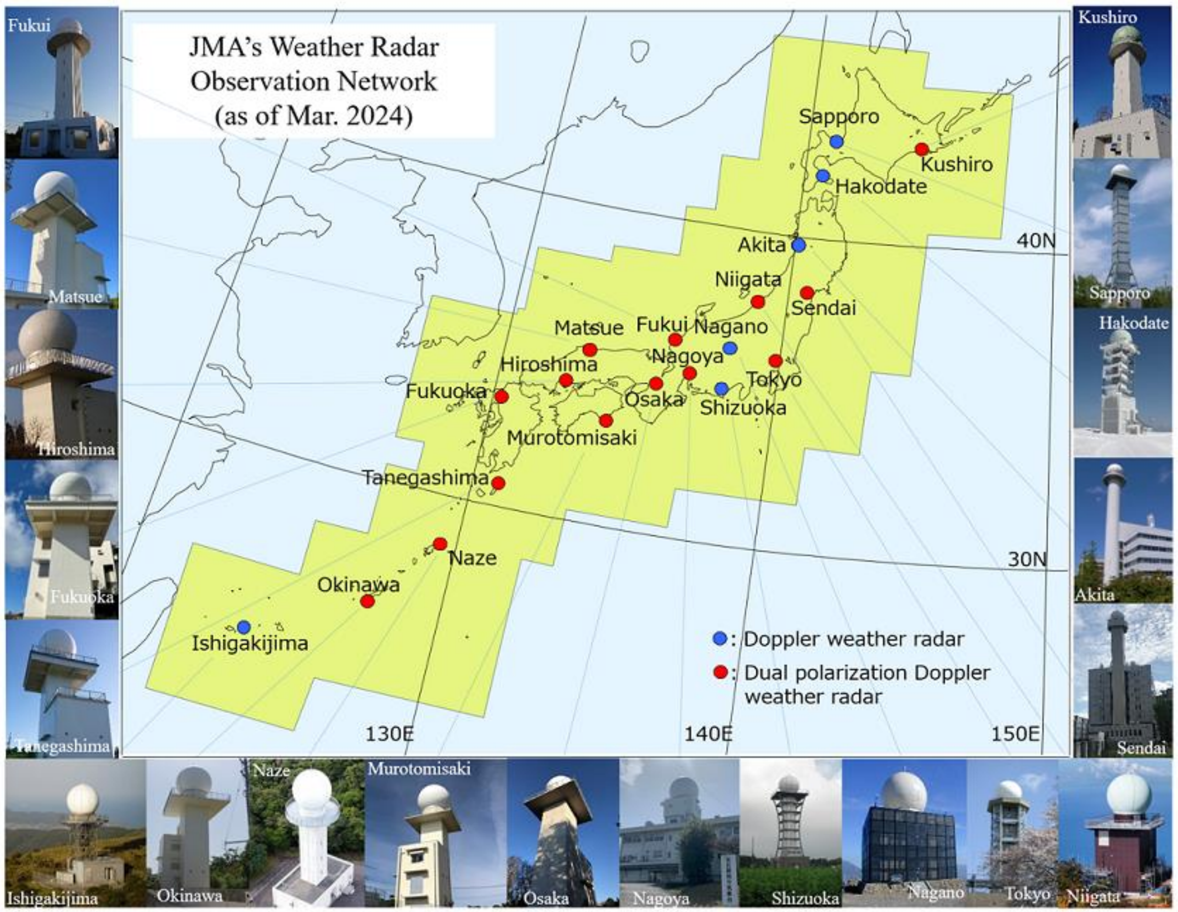


Map showing the distance to nearest GPCC gauges; blank areas are beyond 100 km from the nearest gauge. (Adapted from Kidd et al. 2017)

- **Rain gauges**: directly provide a cumulative estimate, typically unrepresentative of the areal and instantaneous precipitation observed by satellites.

REFERENCE? WHAT REFERENCE?

- **<u>Ground radars</u>**: provide a "snapshot" type of measurements with spatial resolution more similar to satellites. However, radars also provide an indirect rainfall estimate and are prone to errors and significant biases.

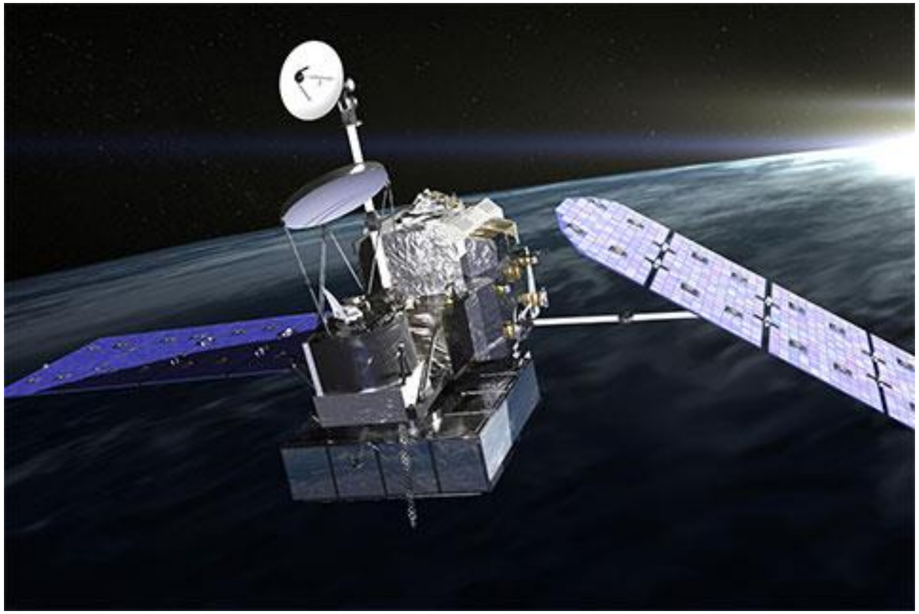


JMA's Weather Radar Observation Network (as of March 2024)

https://www.jma.go.jp/jma/en/Activities/radar/radar.html

REFERENCE? WHAT REFERENCE?

- **Satellite-based radars**: global and high resolution. Issues:
  - As calibrators of the PMW sensors that equip most of the satellite platforms in the GPM constellation, they are used to populate the retrieval databases and train the PMW retrieval algorithms.
  - *Spatial discrepancies*. Radars offer orbital precipitation estimates at a spatial resolution from 1.5-km to 5-km (footprint size), which needs to be spatially re-gridded for comparison with multi-satellite merged products like IMERG (10km).
  - Radar observations are near-instantaneous and therefore would have to be aggregated to the temporal scale of the merged products.



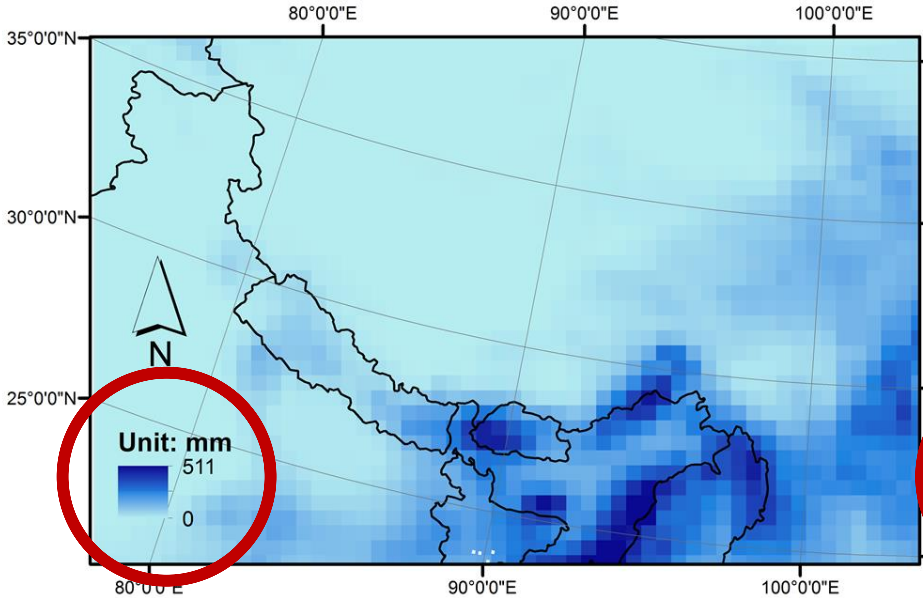
Global Precipitation Measurement/Dual-frequency Precipitation Radar (GPM/DPR)

https://global.jaxa.jp/projects/sat/gpm/

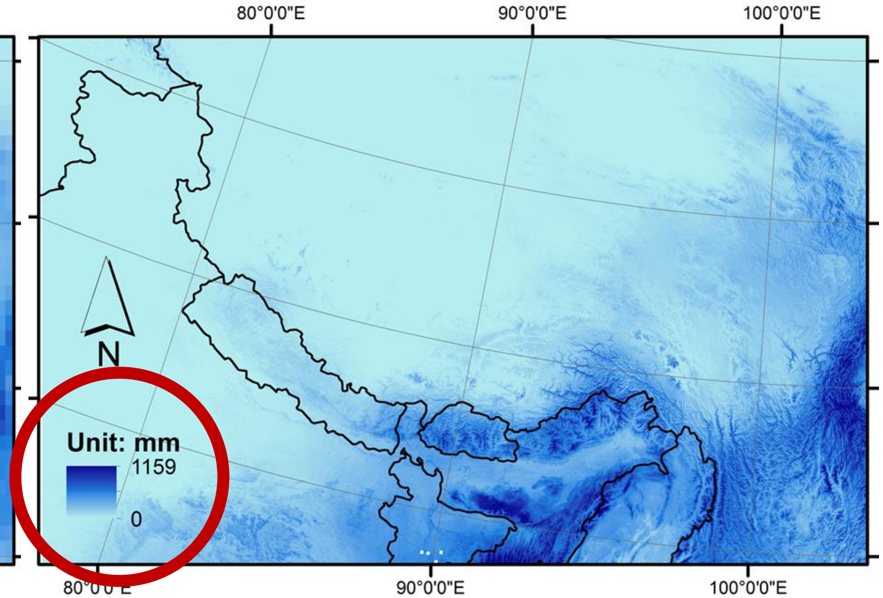
MORE PROBLEMS?

- Uncertainty is added to the validation process by the need to *map* and *aggregate* the datasets onto a common grid.
- This may result in reduced spatial detail and reduced maximum rain rates (i.e., smoothing of extreme events).

MERRA2 Original Resolution (~50km) – Oct 2007



Downscaled MERRA2 (1km) – Oct 2007

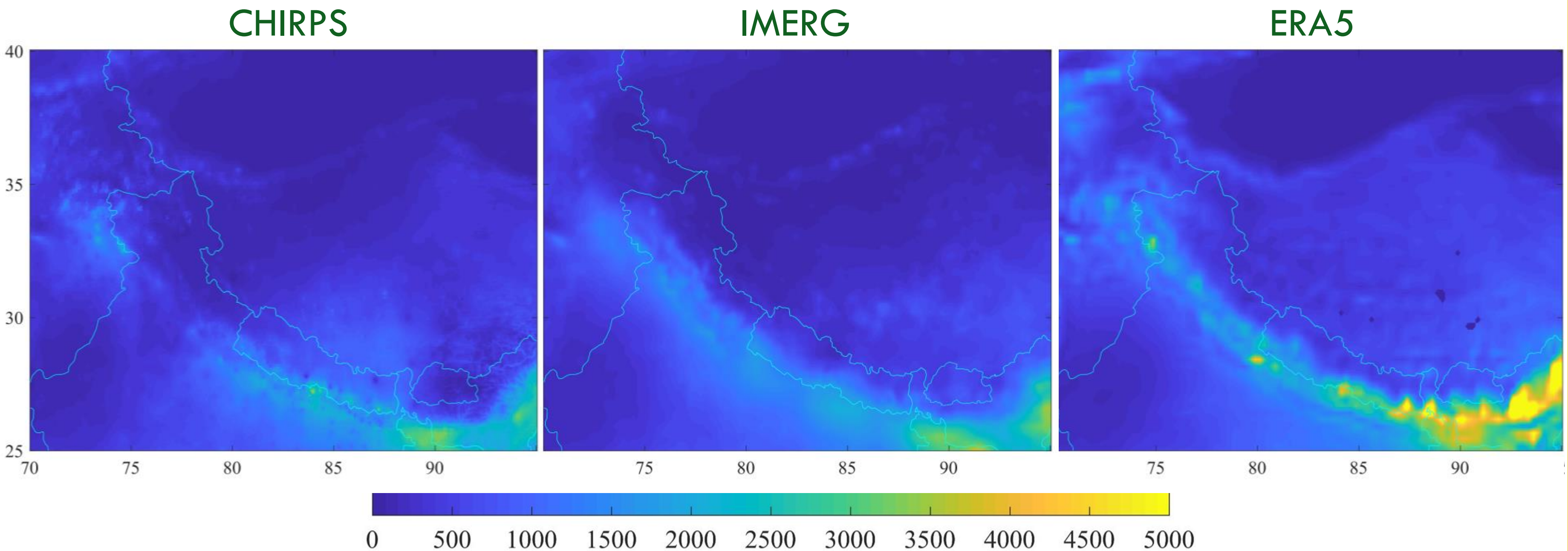


- Important: validation results depend on the spatial scale at which validation is performed, with coarser grids generally producing better results.

VALIDATING/INTERCOMPARING PRECIPITATION

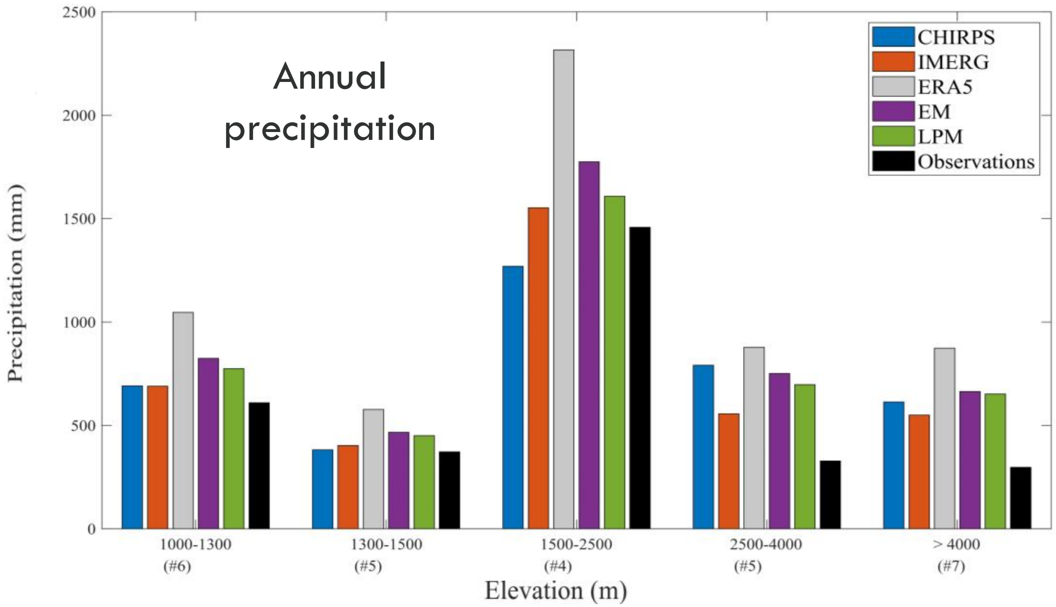
- Products:
  - CHIRPS
  - IMERG
  - ERA5
  - Ground-based observations

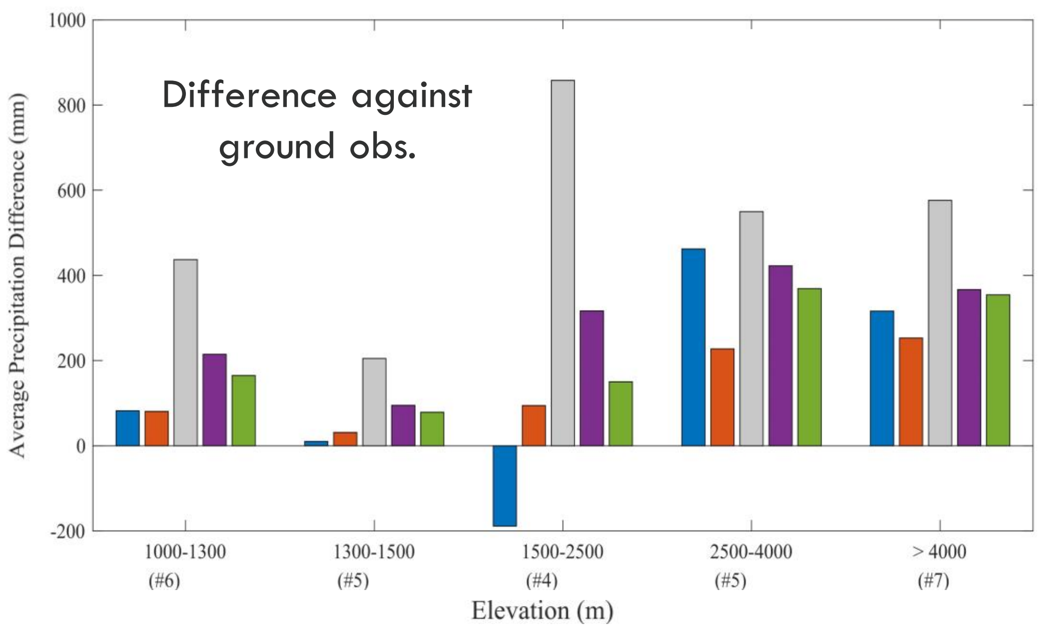
- 2001-2008; Daily/5km



Annual accumulation
(average during 2001-2008) in mm

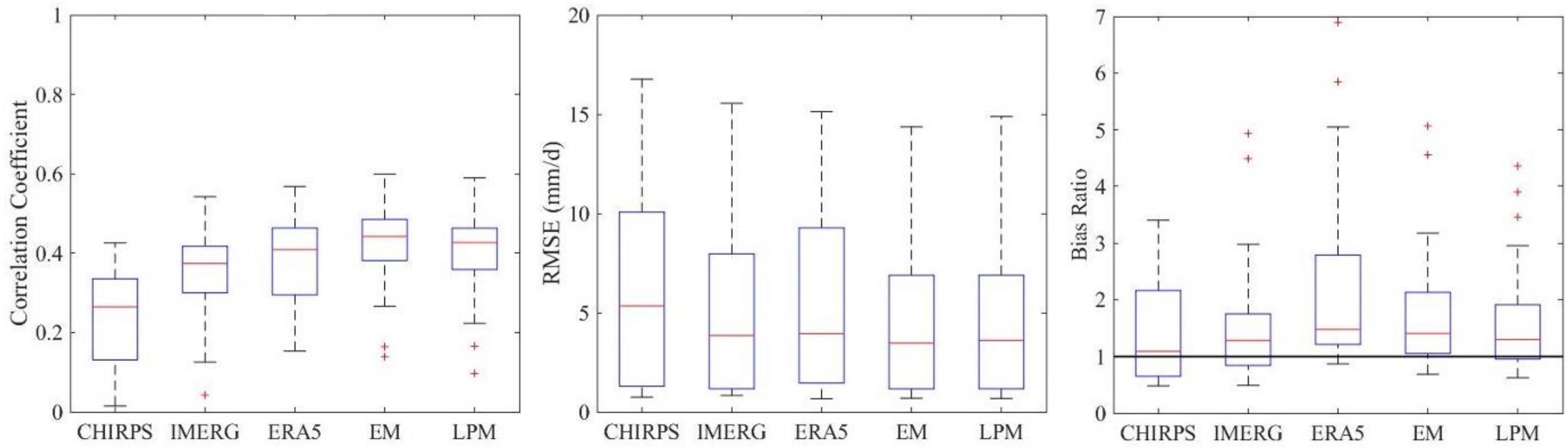
VALIDATING/INTERCOMPARING PRECIPITATION



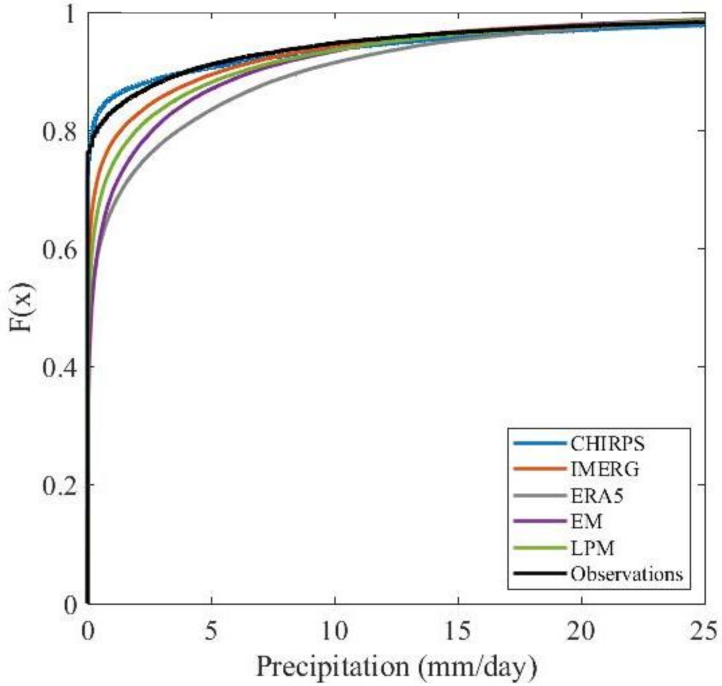


- 27ground stations
- Higher values at mid-elevations (1500-2500m)
- Overestimation of annual average is particularly pronounced at elevation ranges higher than 2500m

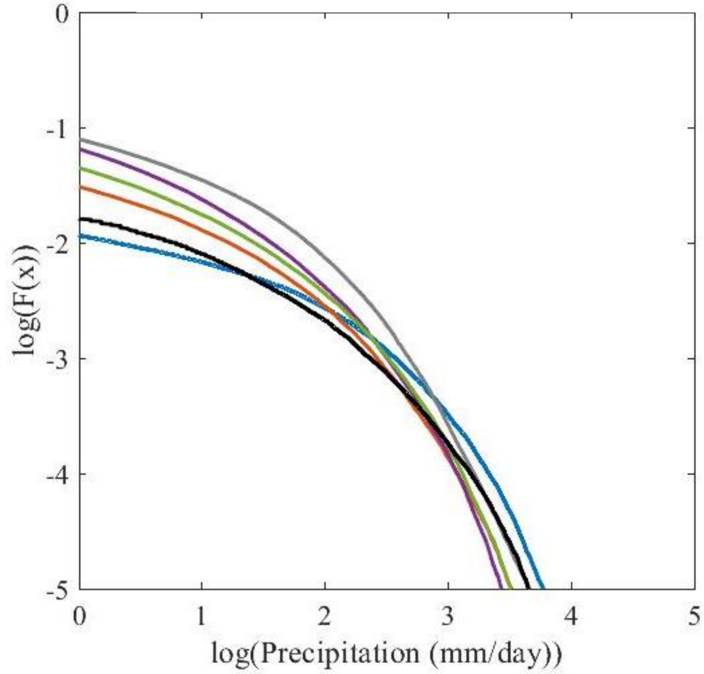
VALIDATING/INTERCOMPARING PRECIPITATION



CDF of daily rainfall

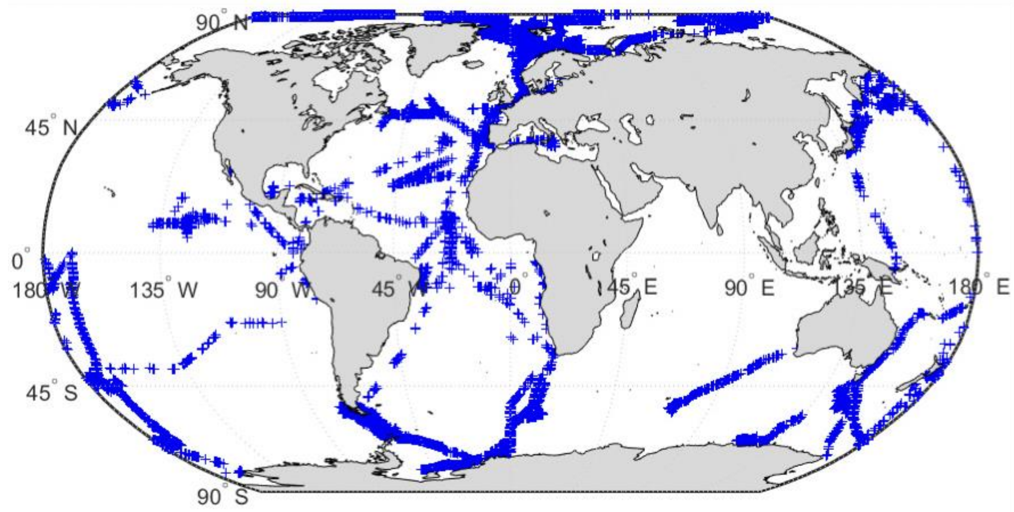


Complementary CDF



OCEANS

- Validation is way more difficult because of their inaccessibility and extent.
- Available benchmark data:
  - weather radars located on islands and coastlines
  - rain gauges onboard cruise, merchant, and research ships
  - buoy gauge arrays
- These observations are affected by deficiencies due to high wind speeds and snowfall.
- Ocean Rainfall and Ice-Phase Precipitation Measurement Network (OceanRAIN), which has been sampling precipitation from optical disdrometers carried by various research vessels since 2010.



Ship tracks of OceanRAIN during 2014–2017 (Klepp et al. 2018)

SO? WHAT REFERENCE?

- The choice of one type of the benchmark depends on:
  - data availability
  - the type of products to be validated
  - the specific objective of the validation study

For instantaneous and high spatial resolution estimates → gauge-corrected radar estimates are generally preferable

For larger regions and timescales (6h to daily) → rain-gauge analyses or combined gauge/radar analyses should be preferred to raw gauge or radar observations

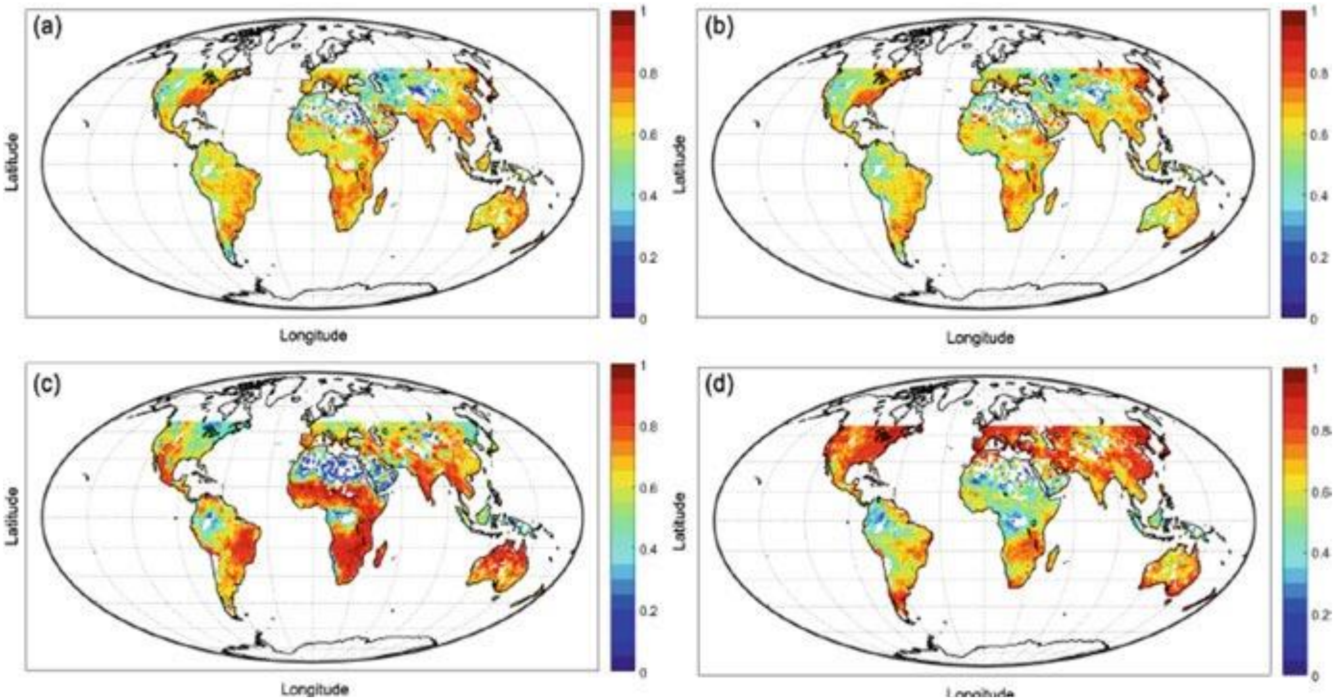
- The presence of *errors in the benchmark* dataset increases the apparent error of the satellite estimates and thus must be considered when validating satellite products.
- A common assumption is that as long as the observational error is random and is *much smaller* than the satellite error, then the reference can be reliably used to intercompare estimates from different products.

ANY ALTERNATIVE?

- Alternative techniques in absence of ground-based observations, e.g., Triple Collocation Analysis (TCA).
- Given 3 estimates of the same variable characterized by (i) stationarity of the statistics, (ii) linearity between the 3 estimates across all timescales, and (iii) existence of uncorrelated error among the three estimates, TCA is able to provide error and correlation of each of the 3 datasets.

3B42RT

CMORPH

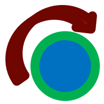


SM2RAIN

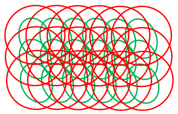
ERA INTERIM

Global correlation obtained by TCA (Adapted from Massari et al. 2017)

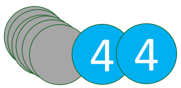
IN SUMMARY



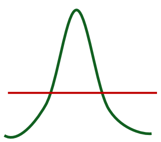
- Orbits, many drifting → non-uniform sampling (esp. non-sun-sync)



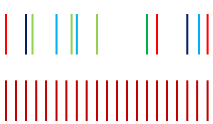
- Different sensors, channels, resolutions



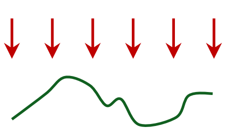
- Observations vs 'samples' vs independent samples



- Retrieval sampling period is small - long-exposure 'snapshots' (c.20 mins), but are deemed representative of precipitation (and processes) over a certain temporal/spatial domain



- Different products using different techniques & different sampling (only IR provides 'regular & frequent' sampling)



- Accumulation period vs natural cycles (e.g. monthly vs MJO) leading to rainfall events split over sampling periods

Courtesy of C. Kidd

THANK YOU!
QUESTIONS?

https://maggioni.vse.gmu.edu/

vmaggion@gmu.edu