INDRA

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

GSICS web meeting on Bootstrap approach to uncertainty estimation

Jorge Gil: Indra
Miriam Pablos: UPC & ICM-CSIC

27 August 2026

INDRA

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Summary

INDRA

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

01

Recapitulation

An Evaluation of the Uncertainty of the GSICS SEVIRI-IASI Intercalibration Products (Hewison, 2013)

Recapitulation: An Evaluation of the Uncertainty of the GSICS SEVIRI-IASI Intercalibration Products (Hewison, 2013)

https://ieeexplore.ieee.org/document/6450080

- A previous GSICS SEVIRI–IASI **inter-calibration** algorithm produced **calibration corrections** together with **uncertainty estimates**
- This work is an **uncertainty analysis** through a **measurement model** of the inter-calibration processes, following the guidance provided by **QA4EO** which is based on the GUM
- **Validate** the quality indicators provided as uncertainty estimates with the GSICS correction

Intercalibration product

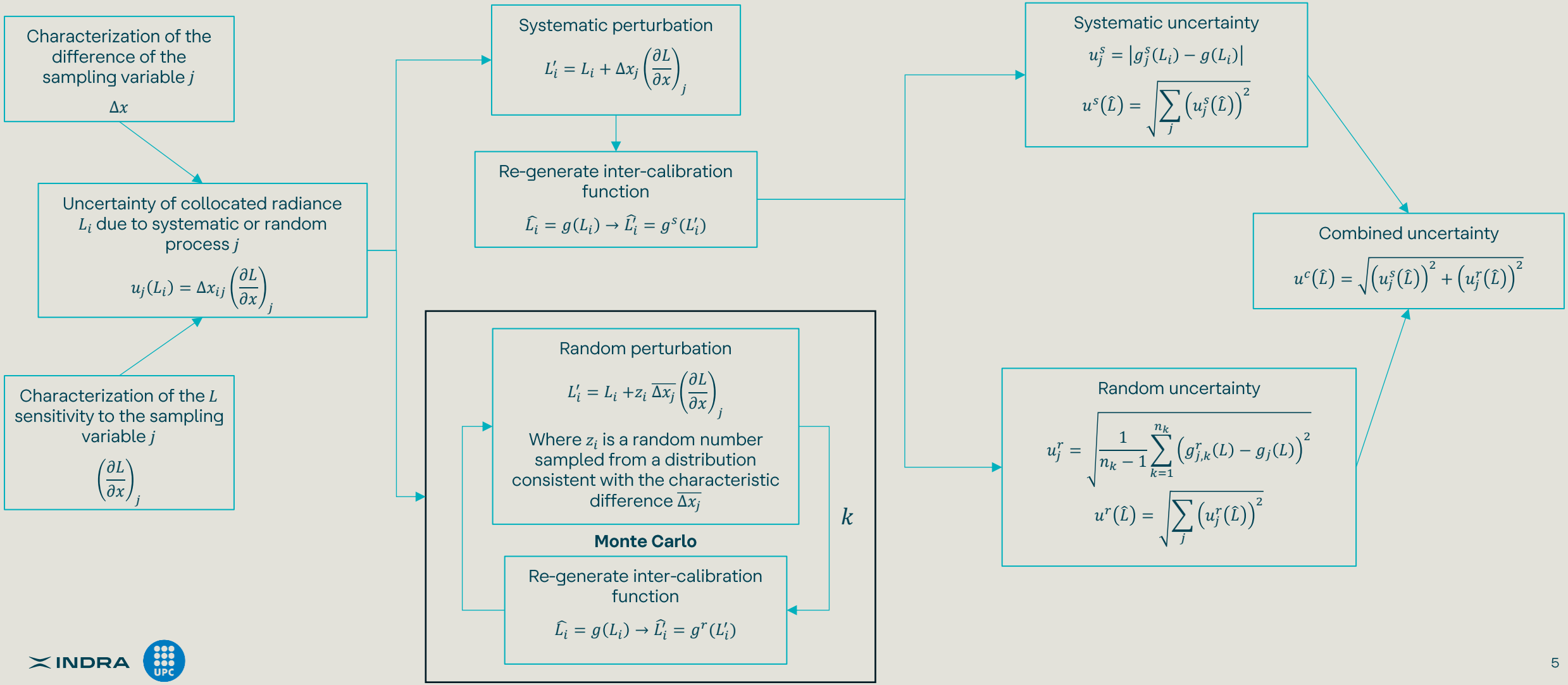
- Collocated observations are transformed to be **comparable** on spatial scales and spectral coverage
- Each pair of collocated observations is allocated a weight based on its measured spatial variance and the specified radiometric noise of each channel to apply a **weighted regression**
- The regression **propagates the variances** to estimate the uncertainty on the corrected radiance, which is provided as a **quality indicator** for the intercalibration product

Evaluation

- Uncertainties are analysed through a **measurement model** of the algorithm's processes
- Each process is considered from both its **random** and **systematic** contributions
- The uncertainties are then combined to produce ***Type B* evaluation**
- The random component is **compared** to the statistics of time series of results from the intercalibration algorithm
- **Recommendations** are be made for adjustments of the intercalibration algorithm to produce more consistent uncertainty estimates

Recapitulation: An Evaluation of the Uncertainty [...] - **General Process**

The uncertainty is estimated by perturbing the collocated radiances according to each uncertainty source, re-fitting the GSICS correction, and quantifying the resulting change or spread in the correction function.



Recapitulation: An Evaluation of the Uncertainty [...] – **Systematic errors**

Temporal Mismatch

- Collocation criteria: $\pm 300\ s$
- Systematic mismatch: $\overline{\Delta t} = 30\ s$
  - Obtained from the collocation ensemble
  - Uniform distribution assumed
- Sensitivity: Derived from sequence of SEVIRI observations

Longitudinal and Latitudinal Mismatches

- Collocation criteria: Not described
- Systematic mismatch:
  - Derived from linear combination of:
    - Reported SEVIRI L1.5 navigation error: $1.2\ km$
    - Worst reported IASI L1C error: $2\ km$
  - Assumptions
    - Partitioned equally in latitude and longitude
    - Uniform distribution
- Sensitivity: Derived from adjacent pixels of a Meteosat-9 image

Geometric mismatch

- Collocation criteria: Ratio of atmospheric path difference $< 1\%$
- Systematic mismatch: $\frac{\overline{\Delta \sec \theta}}{\sec \theta} = -0.00069$
  - Obtained from the collocation ensemble
  - Assumed uniform distribution
- Sensitivity: Derived from an RTM simulating
  - Radiances at $\theta = 30°$ and $\theta = 29°$
  - A range of atmospheric and surface conditions representing clear skies or uniform cloud at different heights
- Cloud induced uncertainty due to parallax errors
  - Spatial offset $\Delta x = \sqrt{2(z_{CTH}\ \sin\theta)^2(1-\cos\Delta\varphi)}$
  - Mean cloud top height estimated: $z_{CTH} = 2\ km$
  - $\Delta x \approx 1.4\ km$ treated as equivalent as spatial mismatch

Recapitulation: An Evaluation of the Uncertainty [...] – **Systematic errors**

Spectral Mismatch

- Direct treatment (no magnitude/sensitivity terms)
- **Deficient SEVIRI IR3.9 coverage by IASI**
  - Truncated and not truncated IR3.9 SRFs applied to several RTM simulations
  - The correction, modelled linearly, was $0.005\ K$ RMS
- **IASI spectral calibration accuracy**
  - Pre-launch spectral-shift specification Δv/v=2 ppm
  - Method: Shift and repeat convolution
  - Negligible effect ($< 1\ mK$)
- **SEVIRI spectral-response-function interpretation**
  - Provided at **irregular** wavelength intervals
  - Method: Conversion to regular IASI SRFs using linear (generally recommended) and quadratic interpolations
  - Errors considered small ($\sim 0.01\ K$)

Recapitulation: An Evaluation of the Uncertainty [...] – **Random errors**

Temporal Variability

- Collocation criteria: $|\Delta t|_{max} = 300\ s$
- Distribution: Uniform $\Longrightarrow \Delta t = |\Delta t|_{max}/\sqrt{3} \approx 173\ s\ RMS$
- Sensitivity: From variogram

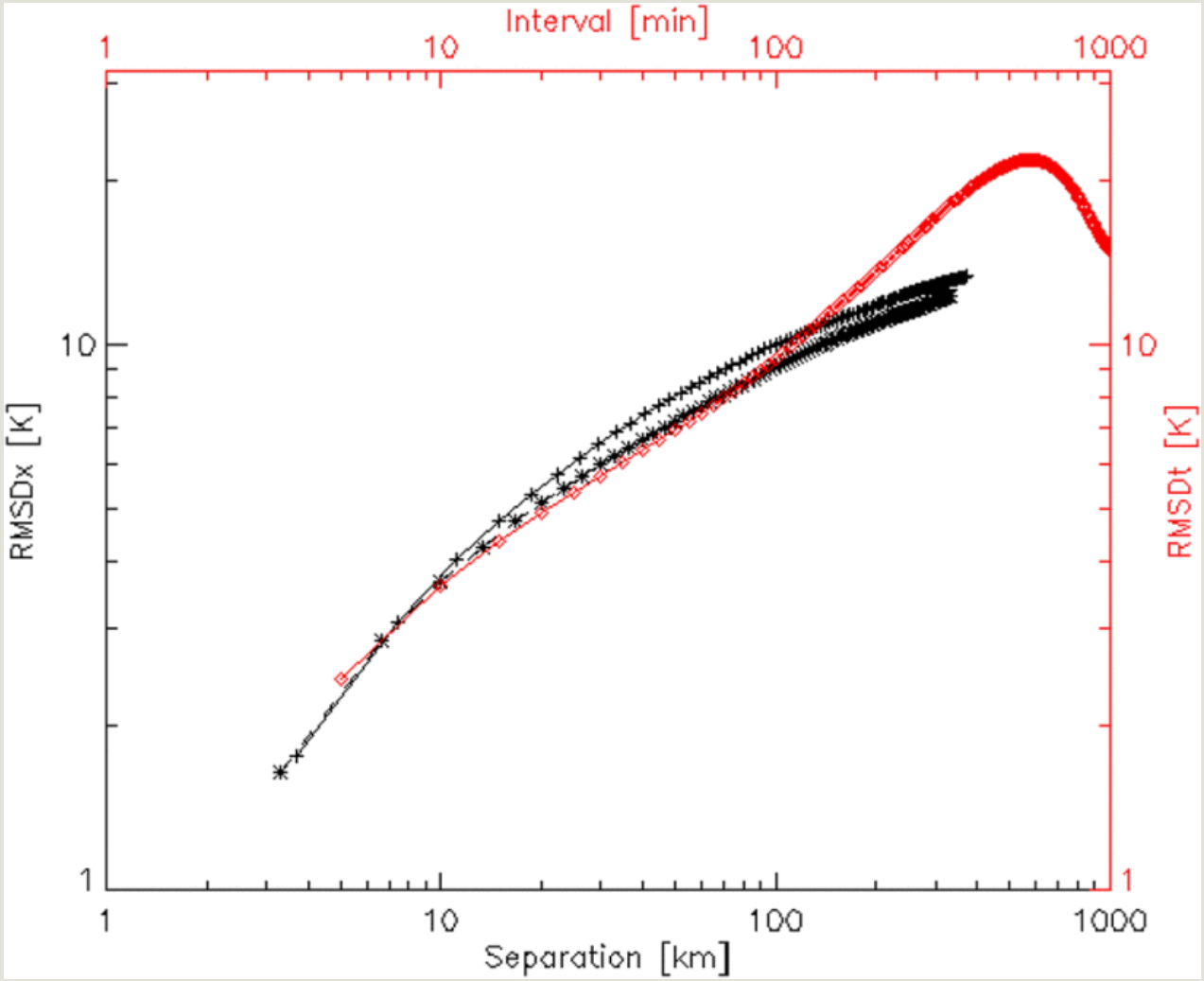
Longitudinal and Latitudinal Variability

- SEVIRI L1.5 grid:
  - Approximately uniform over the target domain of the collocations, where the median distances between adjacent pixels are $\Delta lon_{max} = 3.41\ km$ and $\Delta lat_{max} = 3.38\ km$
  - Considered the dominant geolocation error effect
- Distribution: Uniform over $\Delta lon_{max}, \Delta lat_{max}$
- Sensitivity: From variogram

Random perturbation

$$L_i' = L_i + z_i\,\overline{\Delta x_j}\left(\frac{\partial L}{\partial x}\right)_j$$

Where z_i is a random number sampled from a distribution consistent with the characteristic difference $\overline{\Delta x_j}$



Variograms calculated as rms differences in Meteosat-8/SEVIRI 10.8- μm brightness temperatures (red diamonds, upper x-axis) with time intervals from rapid scanning data and with spatial separation in (black pluses, lower x-axis) the north–south direction and (black stars, lower x-axis) the west–east direction.

Hewison, (2012)

Recapitulation: An Evaluation of the Uncertainty [...] – **Random errors**

Geometric Variability

- Collocation criteria: Ratio of atmospheric path difference $< 1\%$
- Distribution: Uniform; $\frac{\Delta \sec \theta}{\sec \theta} = 0.01$
- Sensitivity: Re-used from the systematic case
- Random cloud parallax is acknowledged and approximately sized, but it is not implemented as an independent Monte Carlo uncertainty term

Spectral Variability

- IASI spectral calibration uncertainty: $\Delta\nu/\nu = 2\ ppm$
- Distribution: Normal
- Radiance effect re-used from the systematic spectral-shift calculation. Evaluated as negligible.

Radiometric Noise

- Assumed white
- Already included in the temporal and spatial variograms because they are derived from real observations
- Also evaluated separately in the random uncertainty budget

Recapitulation: An Evaluation of the Uncertainty [...] – **Results**

Meteosat/ SEVIRI Channel	*IR3.9*	*IR6.2*	*IR7.3*	*IR8.7*	*IR9.7*	*IR10.8*	*IR12.0*	*IR13.4*	*Unit*
Standard Radiance (as Tb)	284	236	255	284	261	286	285	267	K
Typical Standard Correction	0.071	-0.130	0.204	-0.002	-0.048	0.002	0.095	-1.136	K
Total Random Uncertainty from this analysis	0.009	0.004	0.009	0.011	0.012	0.013	0.011	0.006	K
Median Quoted Uncertainty from ATBD	0.004	0.004	0.004	0.003	0.006	0.003	0.004	0.006	K
Rolling SD of Standard Bias from observations	0.013	0.016	0.017	0.022	0.022	0.020	0.016	0.021	K
Total Systematic Uncertainty	0.008	0.003	0.002	0.002	0.002	0.003	0.003	0.004	K
Total Combined Uncertainty	0.012	0.005	0.009	0.012	0.012	0.013	0.012	0.007	K

Recapitulation: An Evaluation of the Uncertainty [...] – **Questions**

Questions?

INDRA

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

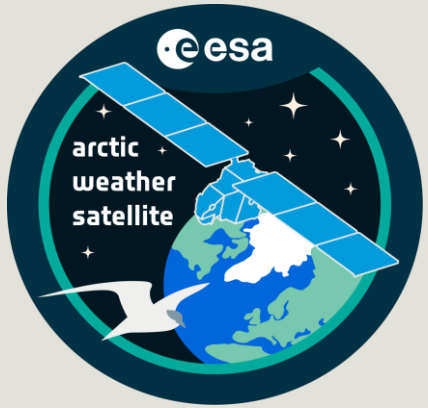
02

The Sterna Inter-calibration Study

The Sterna Inter-calibration Study: Introduction

EUMETSAT Polar System (EPS) - Sterna mission

- Future constellation of **sun-synchronous polar-orbiting** microsatellites
- 2 satellites per plane in 3 orbital planes, scheduled to launch in 2029
- High-temporal-resolution **microwave sounding**
- Complementing MetOp-SG, NOAA JPSS and FengYun-3
- Prototype: Arctic Weather Satellite (AWS), a 19-channel cross-track microwave radiometer



Sterna Inter-calibration Study

- Promoted and funded by EUMETSAT
- Inter-calibration methods:
  - **Simultaneous Nadir Overpasses (SNOs)**
  - **Opportunistic Constant Target Matching (OCTM)**
  - **Vicarious Pseudo-Invariant Calibration Targets**

- With corresponding FIDUCEO-based uncertainty frameworks

EUMETSAT

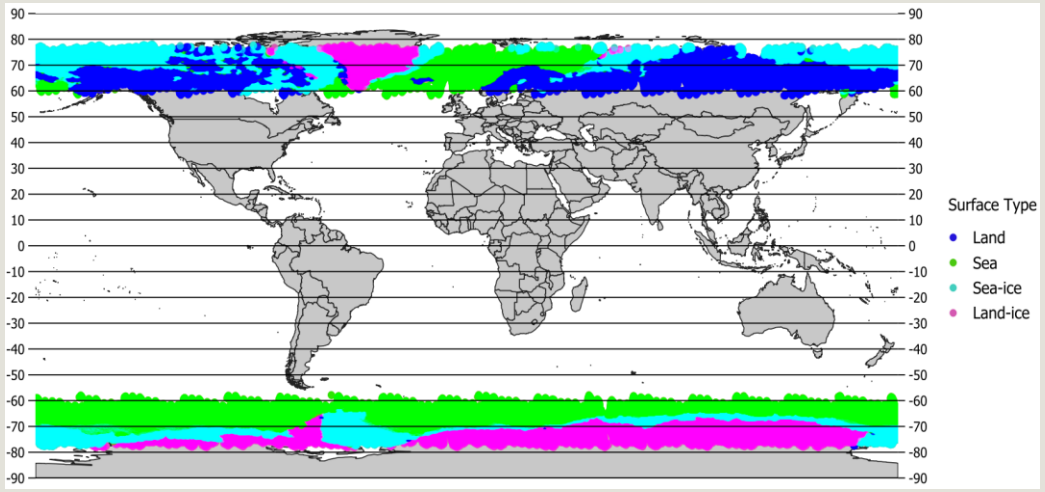
INDRA

UPC UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

The Sterna Inter-calibration Study: Intercalibration methods

Simultaneous Nadir Overpass (SNO)

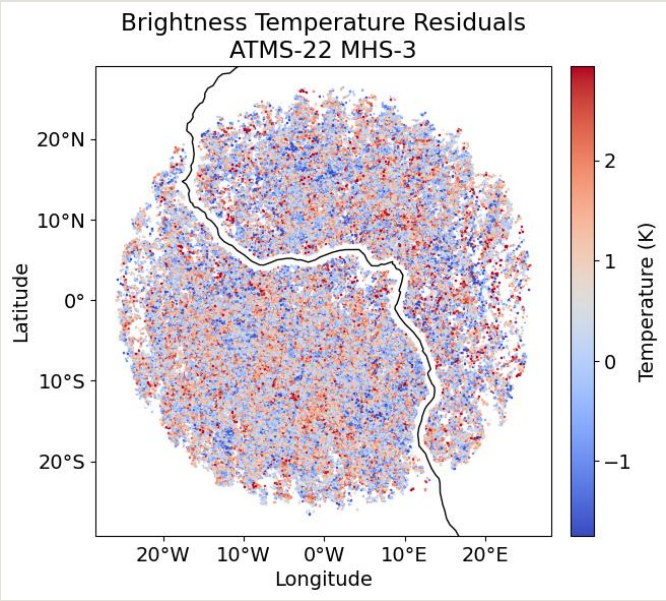
- Occur mostly in polar regions
- Biased towards cold scenes
- Limited dynamic range observed



Global distribution of matched AWS channel 6 pairs ATMS for 2025-04

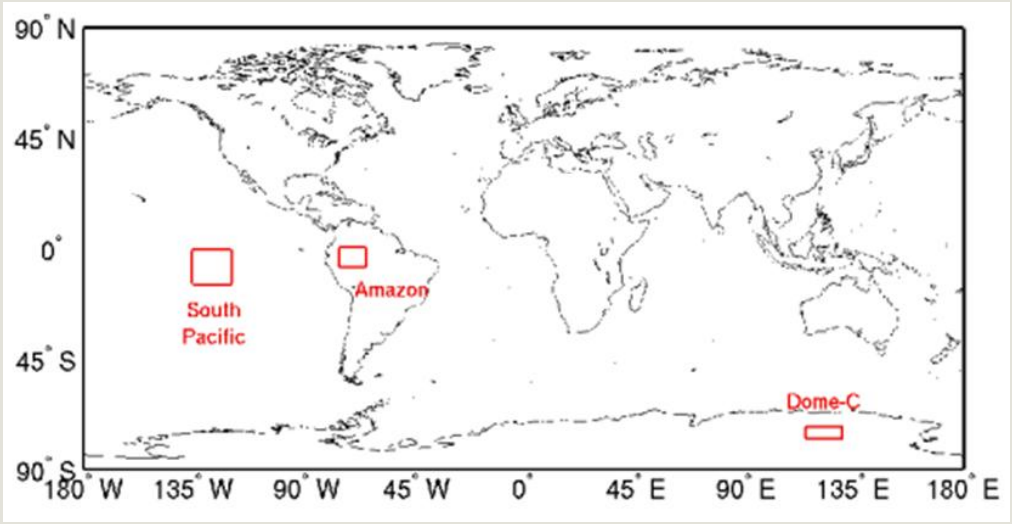
Vicarious

- Comparison of near-nadir observations over pseudo-invariant calibration targets.
- Three evaluated targets: South Pacific, Amazon rainforest and Dome C.



OCTM

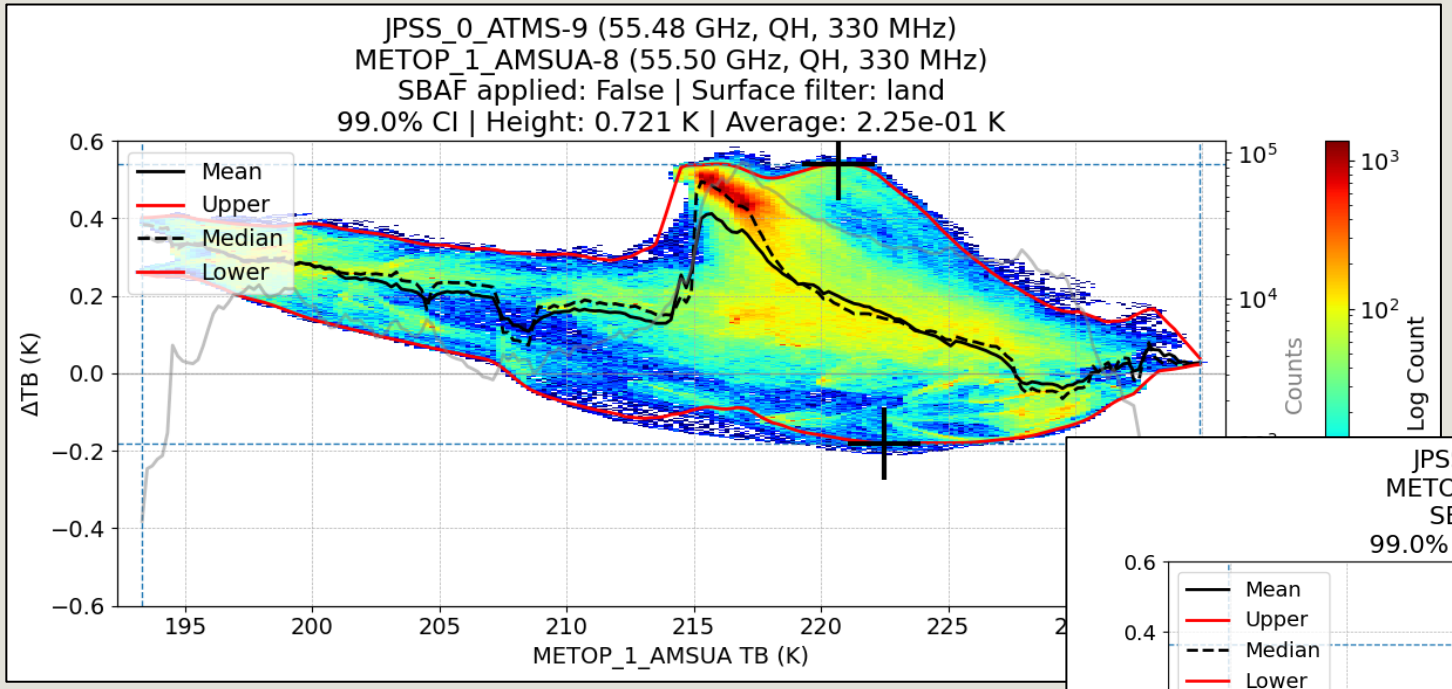
- Relax SNO “simultaneity” with “stability”
- Geostationary data as indicator of scene stability
- Matchups can potentially be hours apart
- Enlarges geographic coverage (polar & tropical regions)
- Increases dynamic range



[1] Buehler et al. (2020) Opportunistic Constant Target Matching—A New Method for Satellite Intercalibration DOI:10.1029/2019EA000856

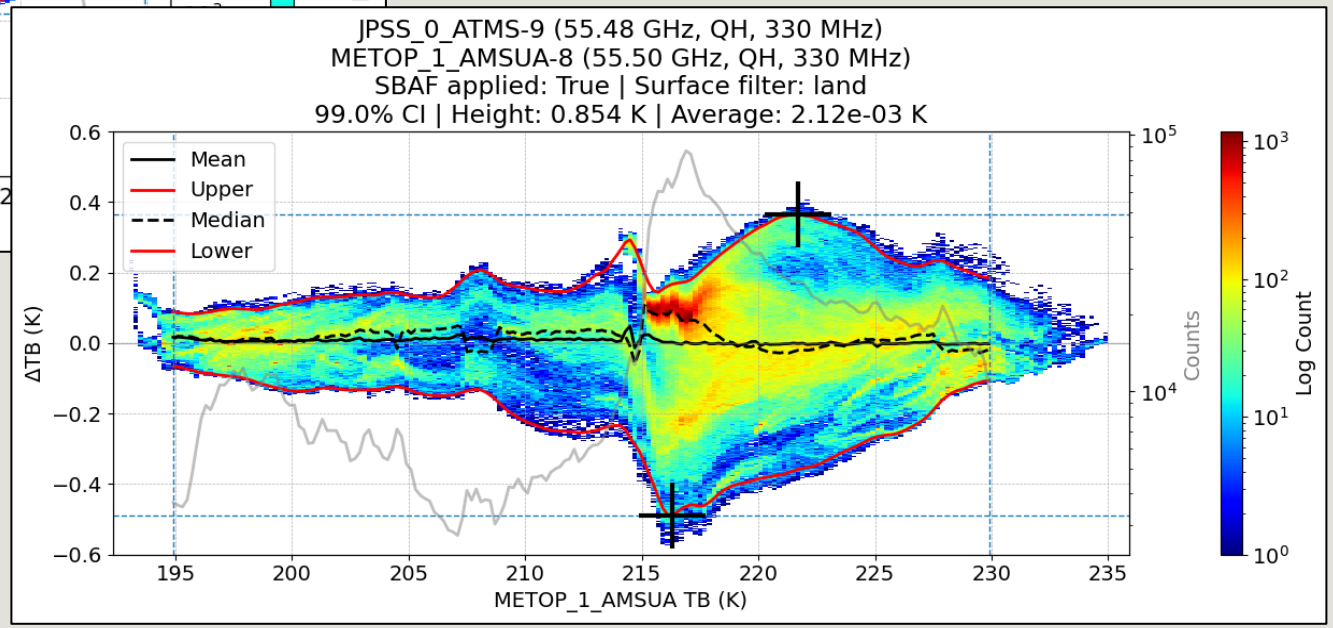
The Sterna Inter-calibration Study: SBAF

- SBAF correction of systematic differences caused by non-identical sensor spectral-response functions
- CDF-matching derivation from RTTOV simulations, with separate land, sea, sea-ice and land-ice classes

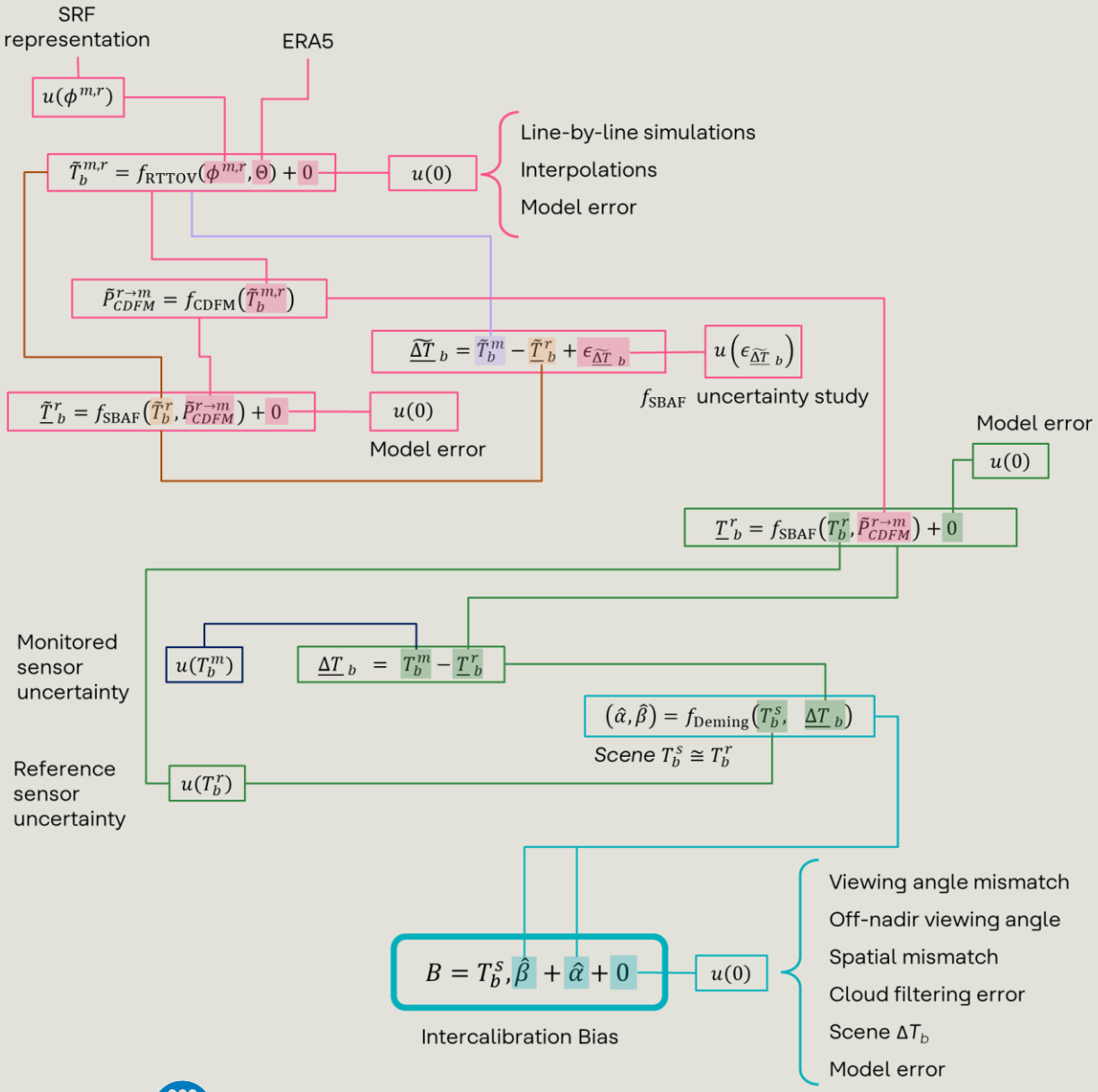


$$\Delta TB = \overbrace{CDF_m^{-1}\left(CDF_r(T_b^r)\right)}^{f_{SBAF}} \Rightarrow \mathbb{E}[\Delta TB] = 0$$

f_{SBAF}



The Sterna Inter-calibration Study: FIDUCEO uncertainty tree



- Bias modelled as a function of scene temperature
- The tree distinguishes between
  - Simulation-based components (pink): SBAF, derived from RTTOV simulations
  - Observation-based components (blue): Measured brightness temperatures
  - Model-based elements (green): Represent the regression formulation and parameters
- Uncertainty is propagated to the fitted coefficients with multiple draws of the dataset with replacement (**bootstrap**)
- **MC** is applied to account for NEΔT or PICT variability and SBAF uncertainty

The Sterna Inter-calibration Study: Questions

Questions?

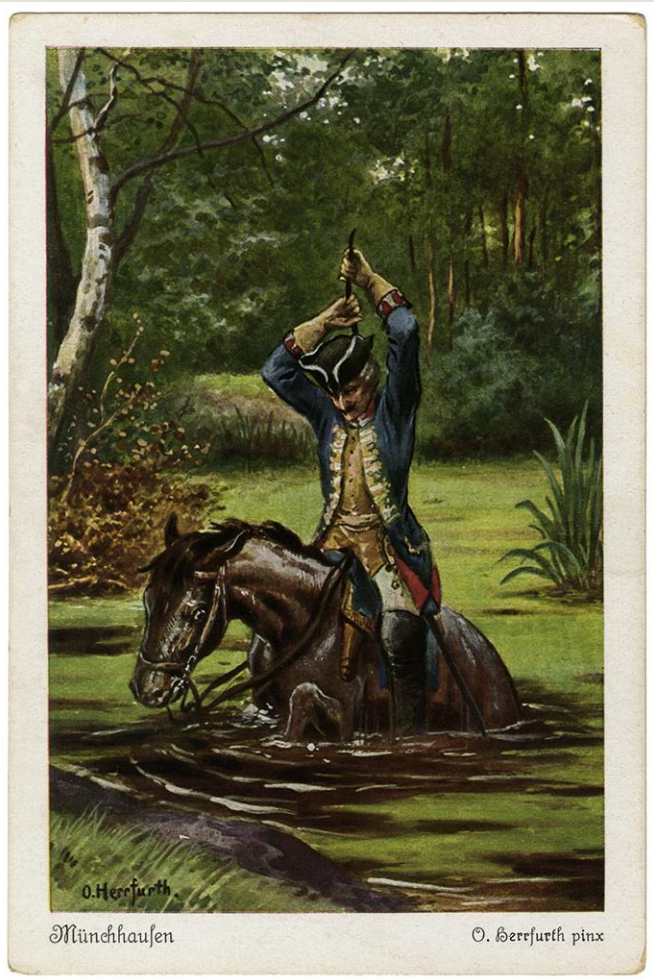
INDRA

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

03

Bootstrap resampling

Bootstrap: From ironic metaphor to technical term



Baron von Munchausen pulling himself out of a swamp by his own hair

Herrfurth, Oskar (German, 1862-1934) -The Jack Zipes Historic Fairy Tale Postcard Collection/Minneapolis College of Art and Design

From the expression **"to pull oneself up by one's bootstraps"**

- Originally, an impossible achievement
- *to improve your situation without any help from other people* (Cambridge Dictionary)

From a 1953 article by Werner Buchholz on the IBM 701 computer. The bootstrap technique allows the computer to start loading itself from a very small initial capability:

> *Self-Loading Procedure*
>
> Another area where the programming facilities of the computer have successfully replaced physical hardware is in the loading of a new program into the computer. There is a load button and a selector switch on the machine, but they do just barely enough to get the process started. The rest is accomplished by a technique sometimes called the "**bootstrap technique**" The switch determines whether the program is to be loaded from tape, cards, or drum. Pushing the load button then causes one full word to be loaded into a memory address previously set up on the address entry keys on the operator's panel, after which the program control is directed to that memory address and the computer starts automatically.

Bradley Efron introduced the statistical bootstrap in "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics*, 7(1), 1–26 (1979)

Bootstrap: Motivation

Motivation: one sample is limited

- One sample gives **one estimate**
- But the estimate depends on **which observations happened to be sampled**
- A different sample would give a different estimate
- The sampling distribution is what we want - but cannot be observed directly

Ideally

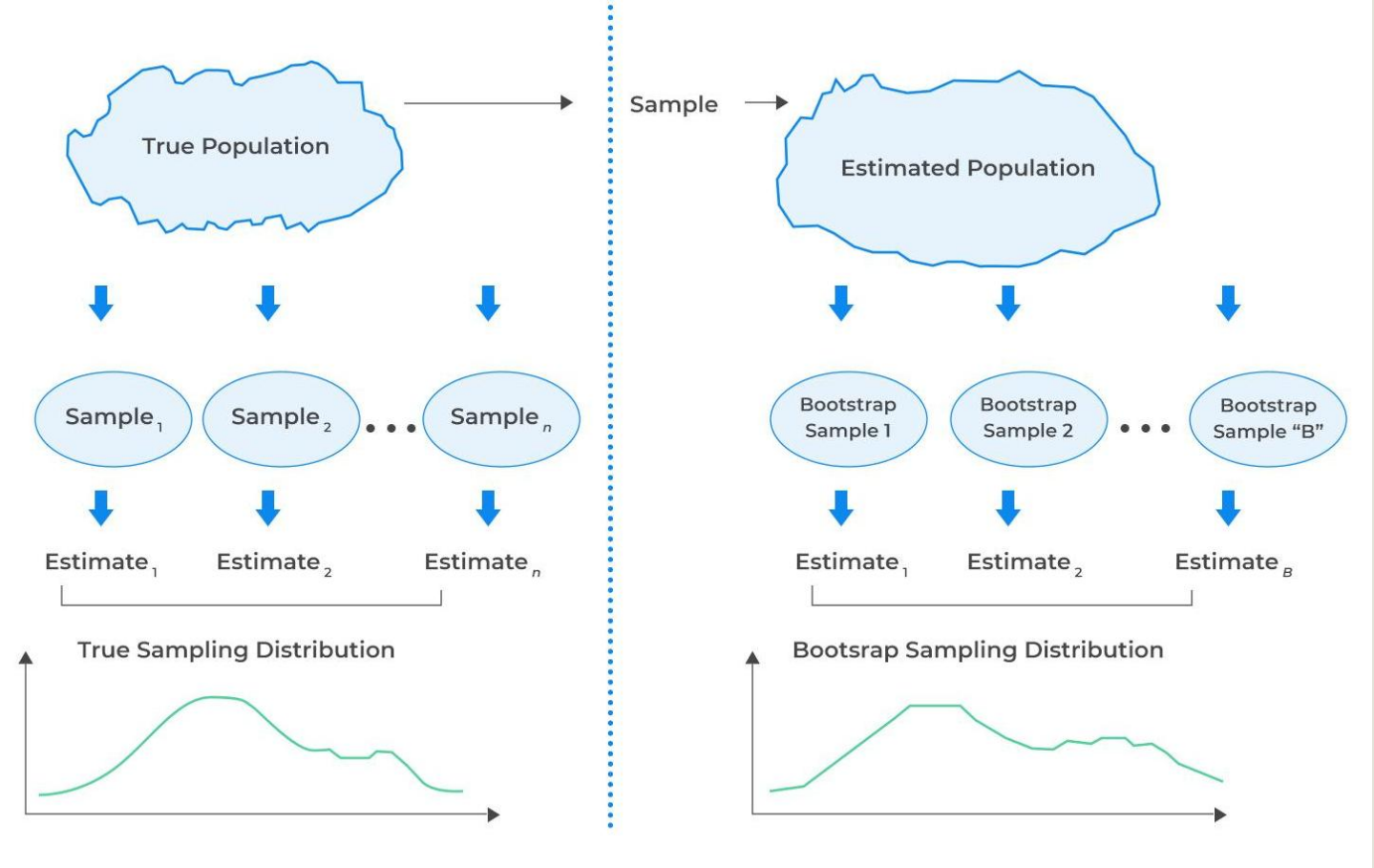
- Repeat the experiment many times.
- Observe the variability of the estimate

In practice

- We usually have **only one sample**.
- Repetition may be impractical

Use the observed data to learn how variable a result from those data is

Resample the sample itself to approximate repeated sampling.

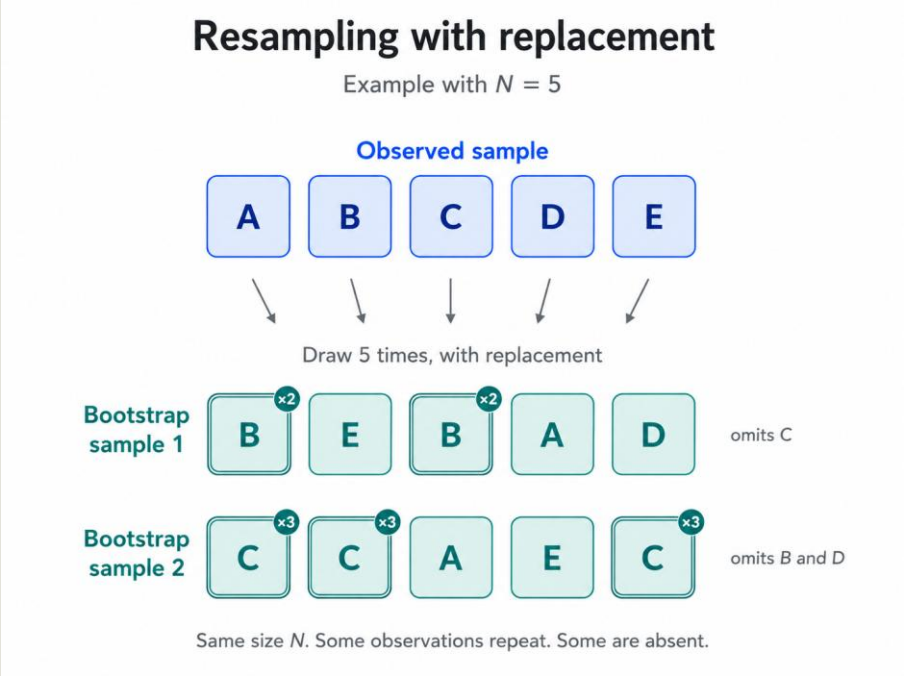


AnalystPrep - Resampling Methods in Statistics

Bootstrap: Resampling produces a distribution of the estimator

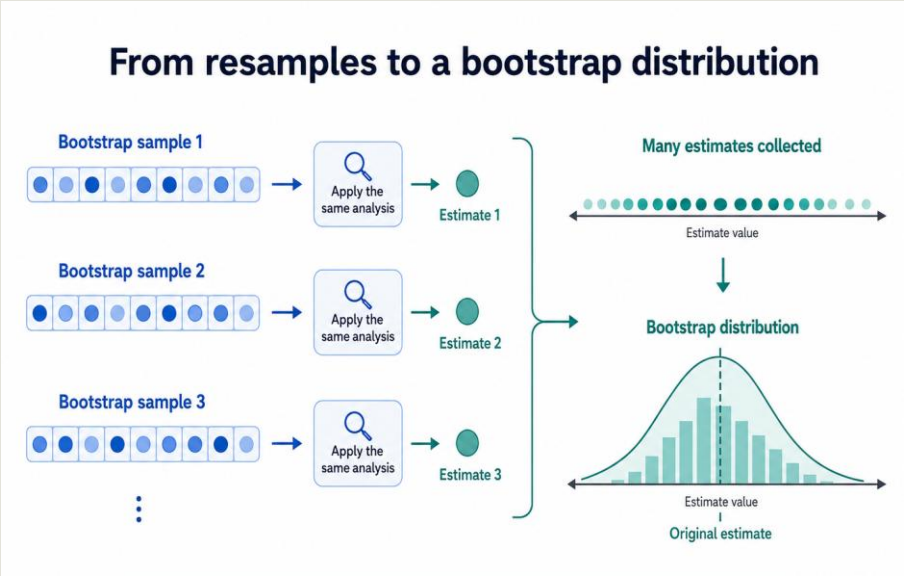
Generate new samples from the empirical distribution
- Draw N observations from the empirical distribution
  - **with replacement**
  - keeping the original sample size
- Each observation has probability $1/N$ at every draw

As an obvious consequence
- Some observations appear more than once
- Some observations are not selected
- What was not observed does not appear
- Each resample is a different plausible realization of the empirical distribution



For each bootstrap sample
- Apply the **same estimator** as to the original data.
- **The estimates vary** because each bootstrap sample is different
- Their empirical distribution is the **bootstrap distribution**

Which represents
- An approximation to how the estimator would vary under repeated sampling
- Its spread estimates the **sampling variability** of the estimator



Bootstrap: Assumptions, limitations and variants

Assumptions:

- The observed sample is sufficiently representative of the population of interest.
- For the standard bootstrap:
  - Observations are treated as **independent and identically distributed**
  - The empirical distribution reasonably represents the population

It does not:

- Create conditions that are **absent from the observed sample**.
- Reproduce structures that are ignored by the resampling scheme
- Compensate over/under-represented samples

Variants:

- **Stratified bootstrap**
  - Split the data into representative groups and resample within each group, so important parts of the population remain represented (e.g., brightness temperature ranges, surface types)
  - Mitigates poor or unstable representation of important subgroups in the resamples
- **Block bootstrap**
  - Resample groups of related observations together, so dependence between observations is retained (e.g., consecutive nearby observations)
  - Mitigates bias from temporal or spatial autocorrelation, by preserving short-range dependence.
- **Cluster bootstrap**
  - Resample *natural* groups (e.g. orbit, pass, SNO event)
  - Preserves event correlation.

Bootstrap: Questions

Questions?

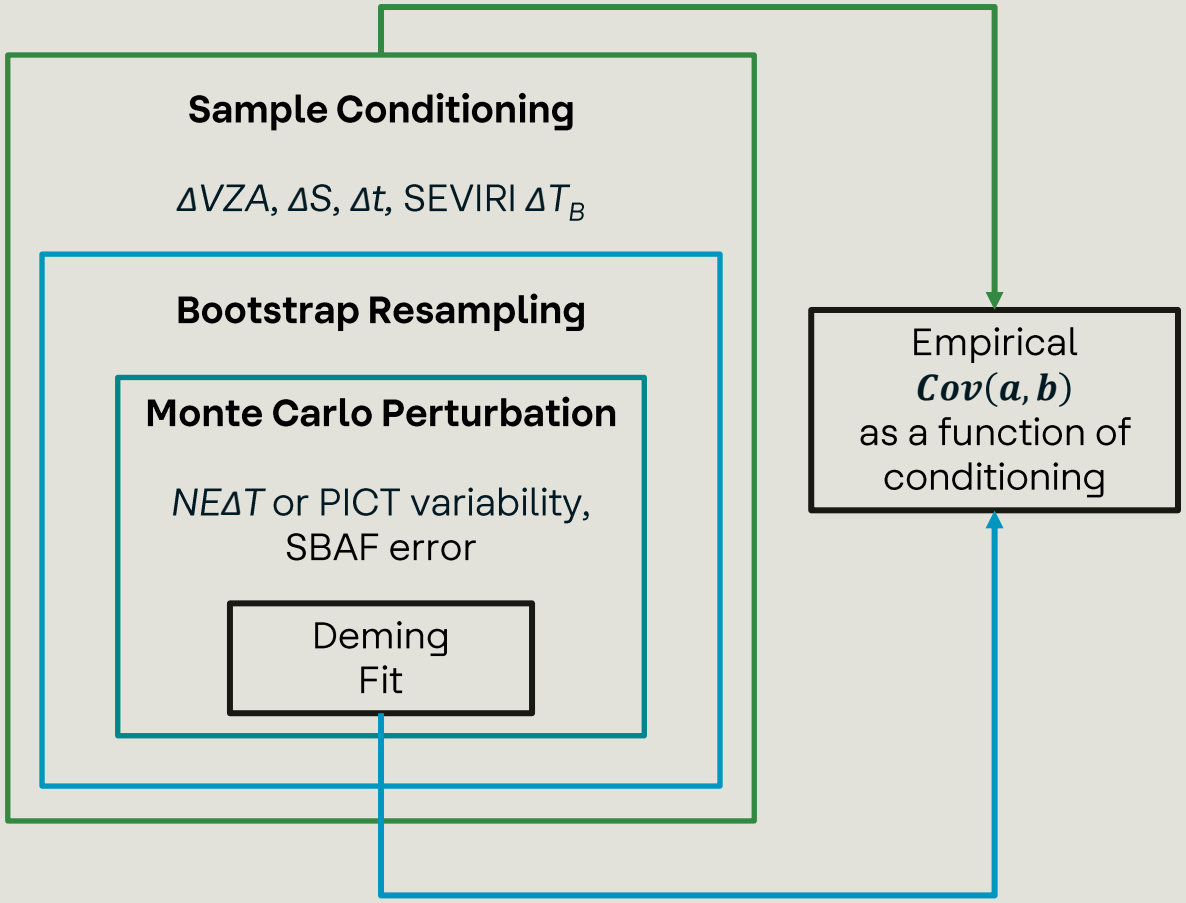
INDRA

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

04

Bootstrap in the Sterna intercalibration study

The Sterna Inter-calibration Study: Uncertainty as a function of conditioning



- Systematic effects are not explicitly treated
  - The systematic spectral mismatch is minimized by the SBAF
  - Others are mitigated through conditioning or considered negligible
  - **Bootstrap naturally propagates fluctuations in sample-dependent systematic effects** (e.g., time mismatch)
- NEΔT, PICT variability and SBAF error contributions are random and perturbed with MC
- Bootstrap is also applied to obtain an estimate of how stable the regression is, given filtered samples
- The framework empirically estimates the uncertainty of the bias correction, capturing sampling variability, measurement effects, and the impact of mismatch effects

Bootstrap in the Sterna intercalibration study: Sampling conditioning

Adjust mismatch-induced variability

Restricting the matchup dataset using a selected variable

$$|\Delta X| < \Delta X_{max}$$

changes which observations are admitted to the regression

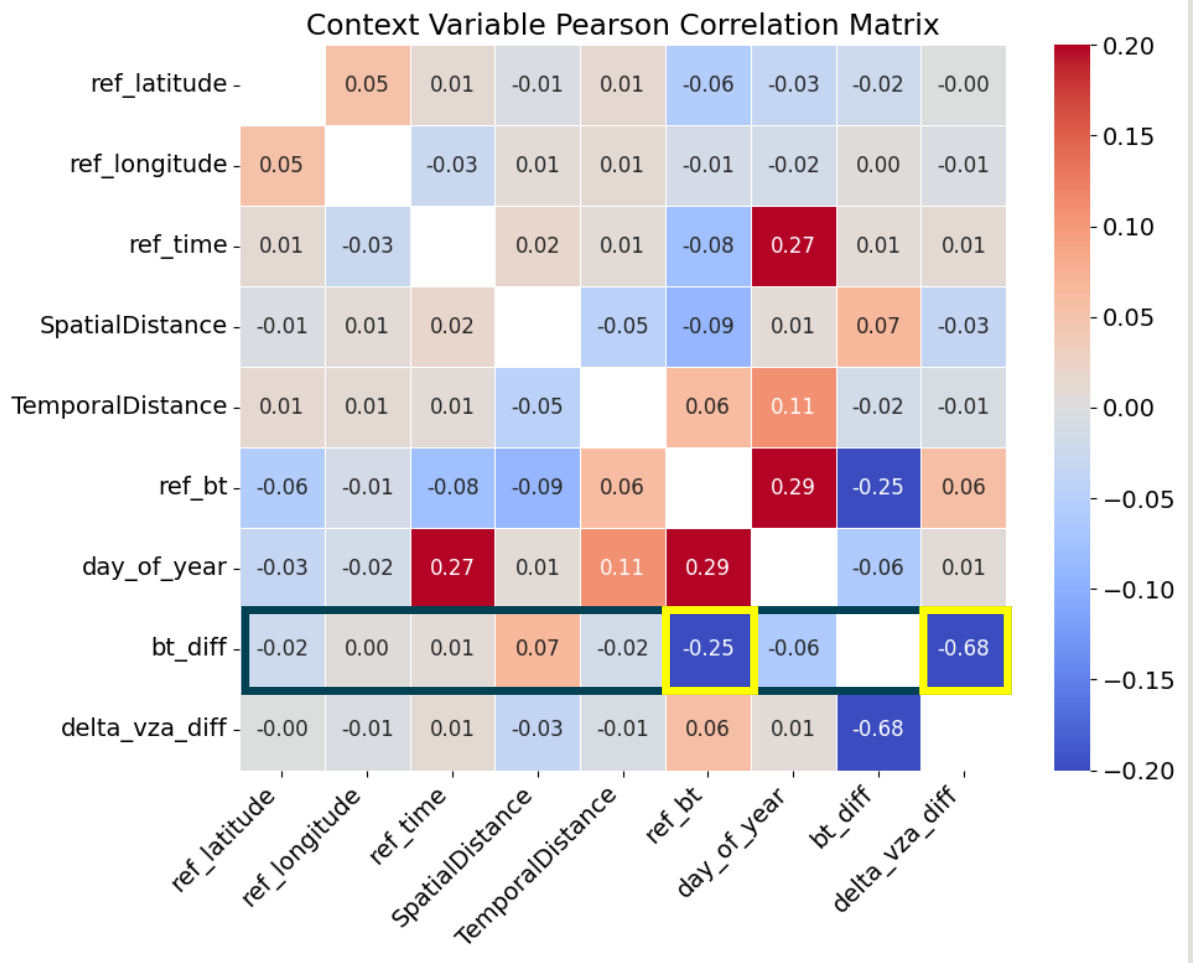
Relaxed conditioning → more observations, but potentially **greater scene mismatch and uncertainty**.

The threshold is selected by looking for the best trade-off between

- **Uncertainty**
- **Sample size**
- **Radiometric dynamic range**.

In this study, conditioning variables include:

- |ΔVZA|
- Spatial distance |ΔS|
- Temporal distance |Δt|
- Maximum VZA
- SEVIRI |ΔTB| (OCTM method only)



Example of Pearson correlation with environment and observation variables.

ATMS Ch9 vs. AMSU-A Ch8 for the OCTM method

55.5 GHz - temperature sounding channel, peaking around the tropopause, with no surface sensitivity

ref_bt and *delta_vza_diff* are candidates. Note that these two variables are uncorrelated

Bootstrap in the Sterna intercalibration study: Represented variability

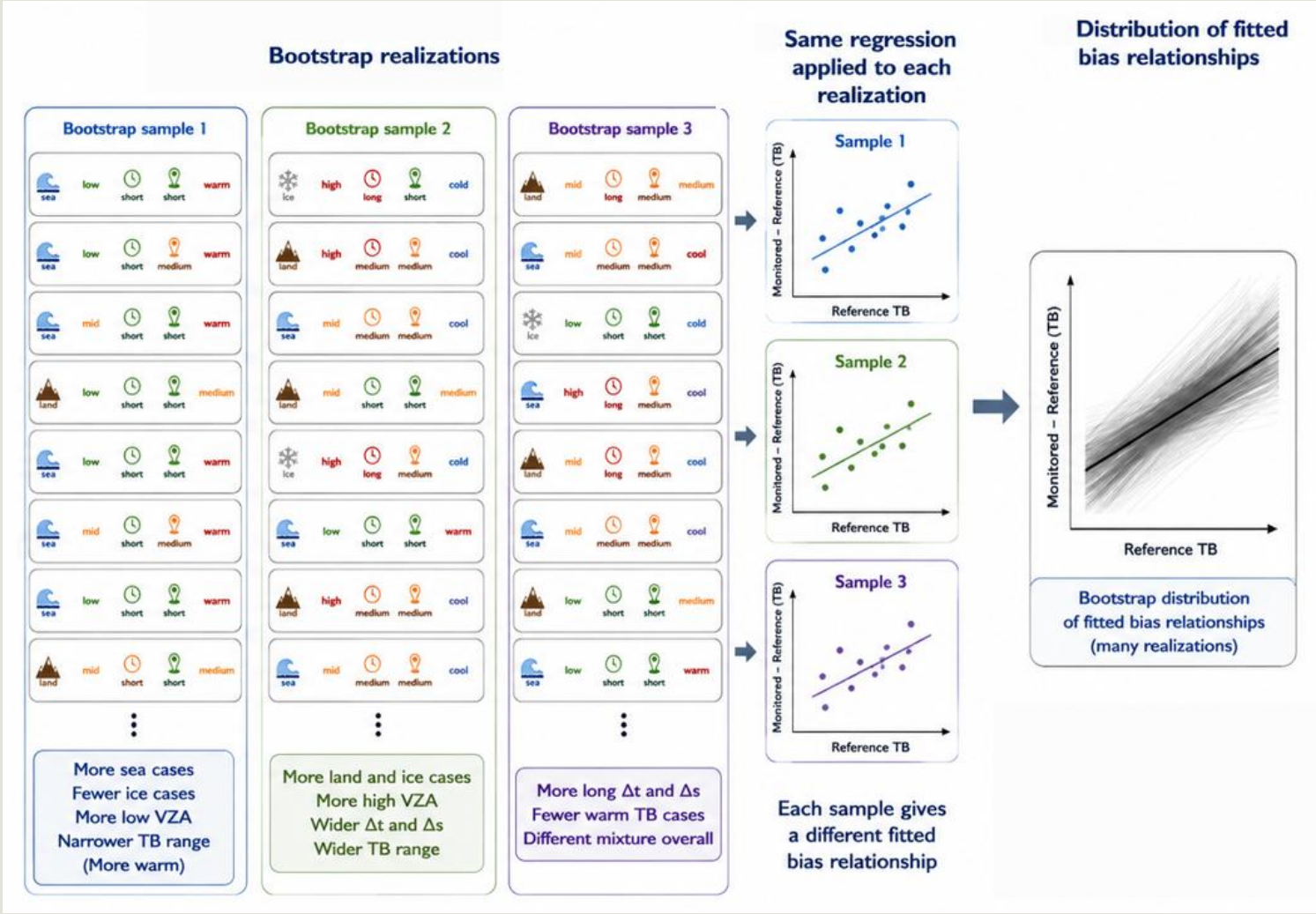
Each matchup carries a set of observing conditions
- Surface type.
- Viewing geometry.
- Time separation.
- Spatial separation.
- …

Bootstrap resampling
- Resample **complete matchups**
- Keep the observed combinations of conditions together
- Change how frequently each matchup appears

Each bootstrap realization therefore has a different empirical distribution of observing conditions

Then
- Apply Monte Carlo
- Calculate the regression to obtain a new set of model parameters
- Repeat

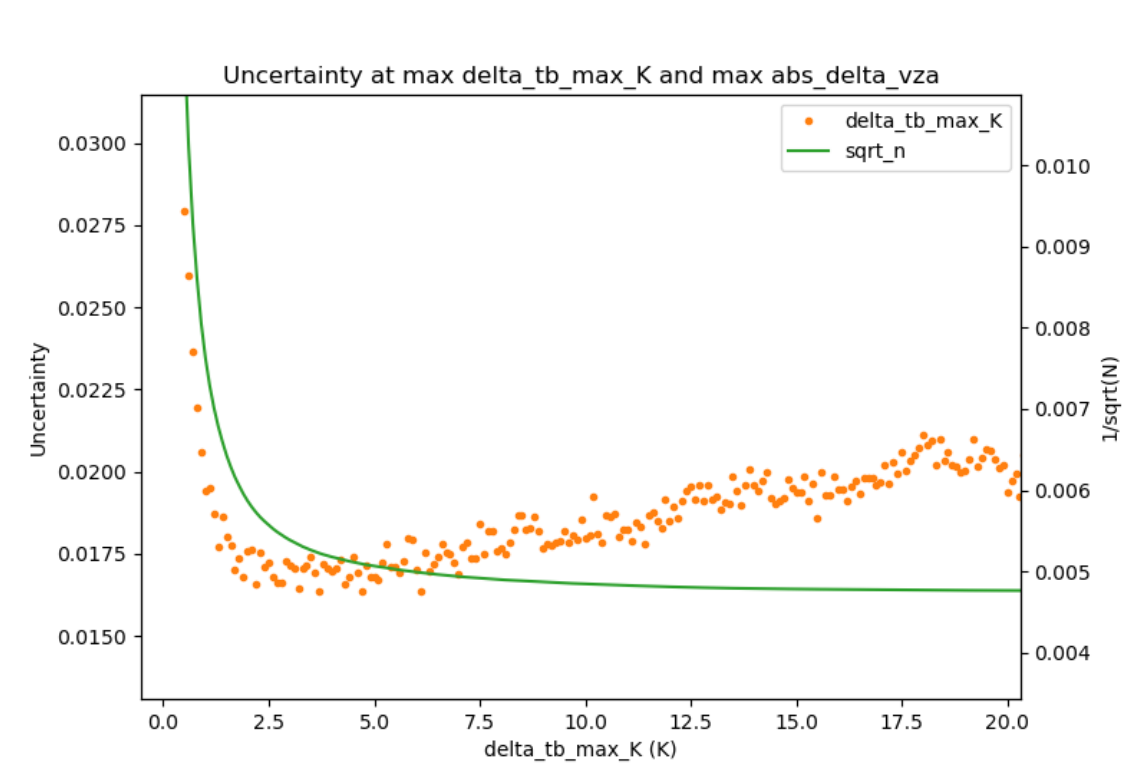
Finally
- Obtain the covariance matrix of the model parameters



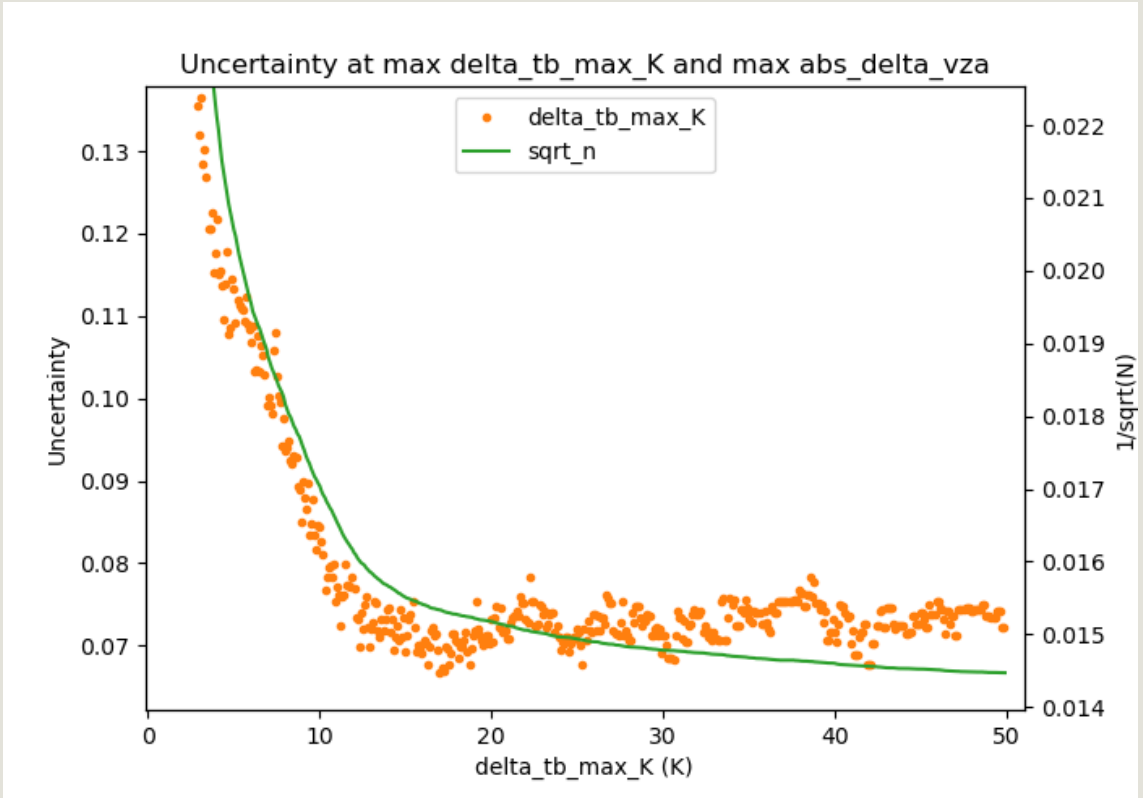
Bootstrap probes how the fitted bias responds to finite variability in the distribution of the conditioned sets of observations.

Bootstrap in the Sterna intercalibration study: Examples

scenario_1_MHS-3_ATMS-22_sev5

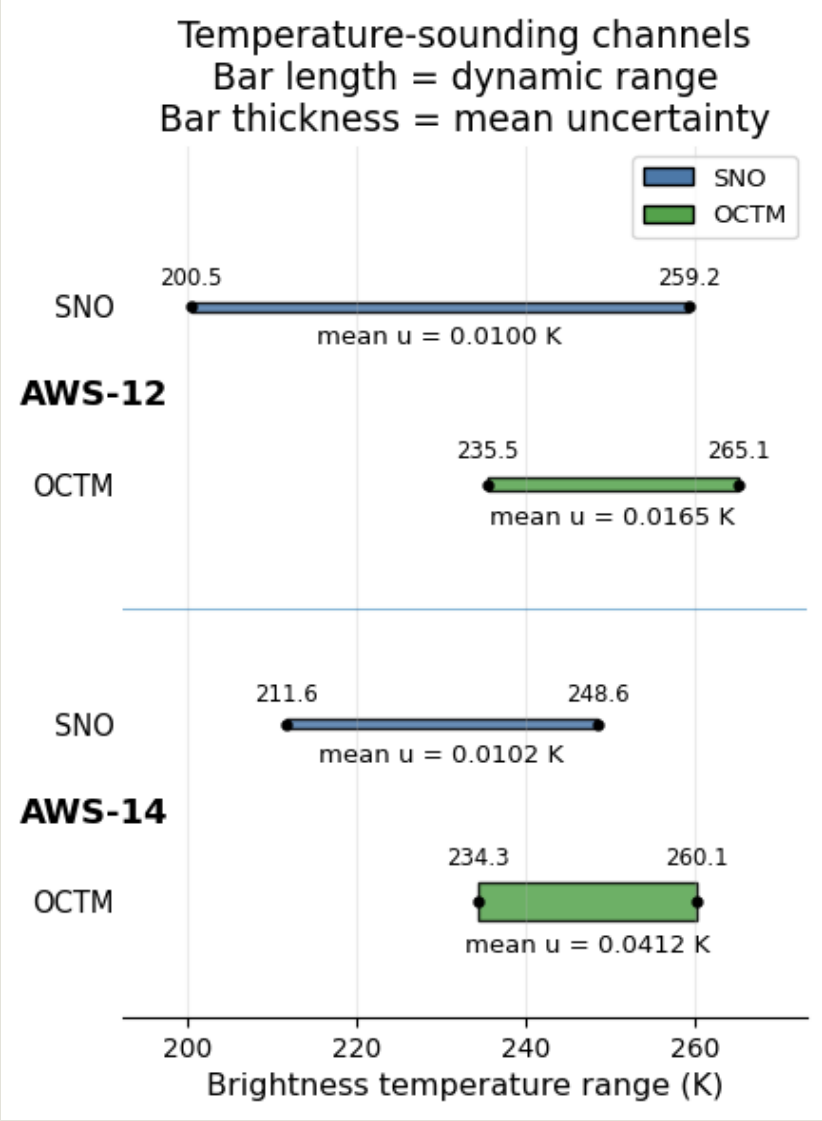


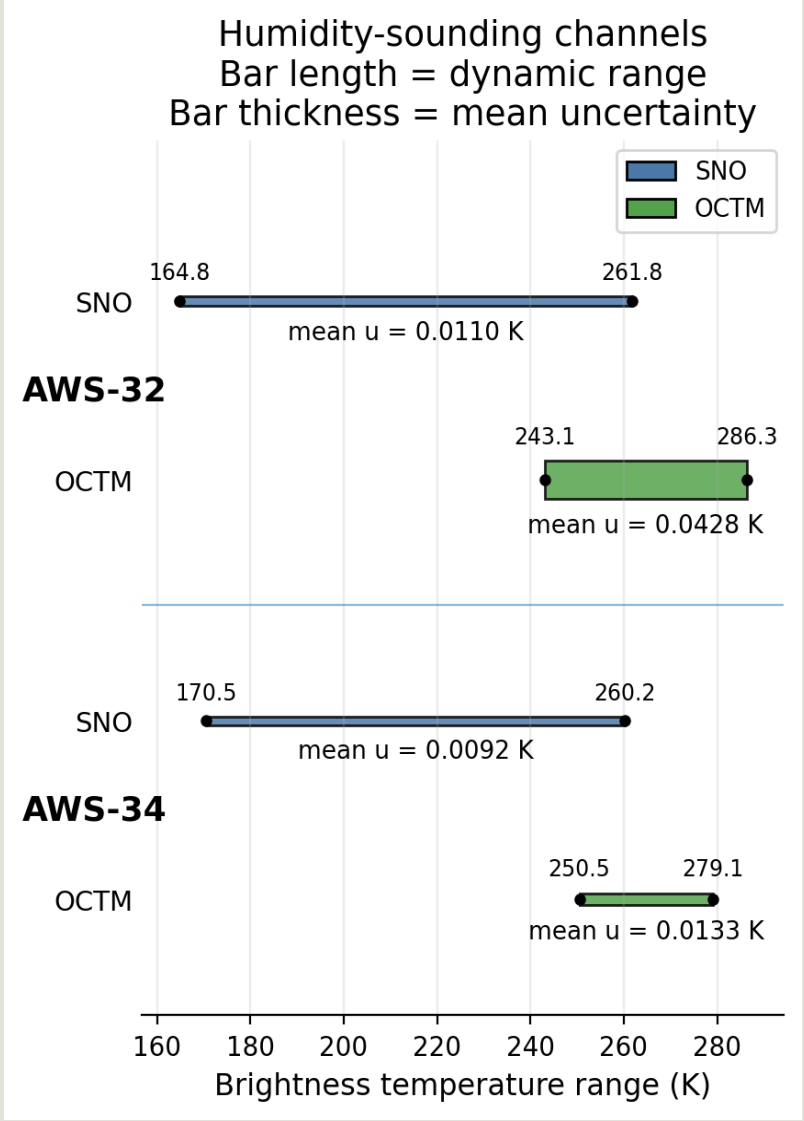
scenario_3_AMSUA-4_ATMS-5_sev9

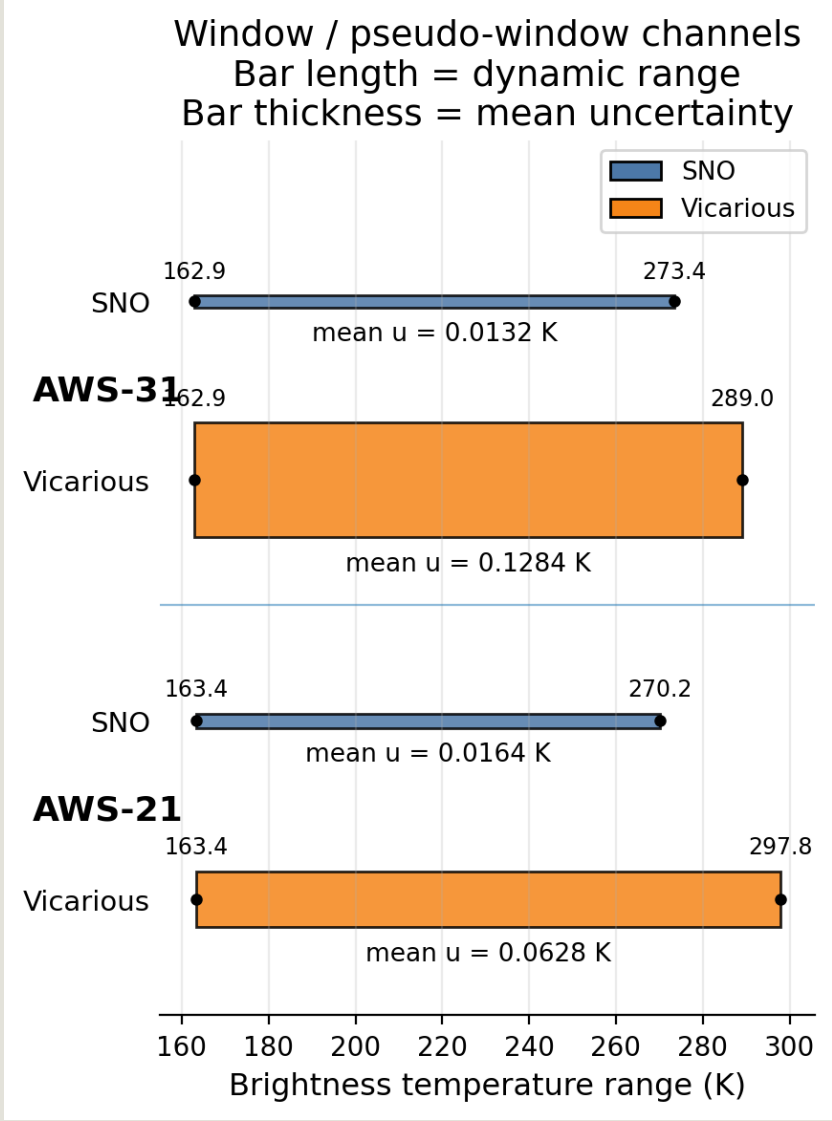


Uncertainty (K) under |ΔTB_SEVIRI| conditioning for two cases. 1/√N is shown at a different scale for reference

Bootstrap in the Sterna intercalibration study: Examples of uncertainties and coverages by method







Bootstrap in the Sterna intercalibration study: Conclusions

- **One common uncertainty framework** is applied to OCTM, vicarious inter-calibration and SNO, making them consistently comparable.
- **The same bias model and propagation procedures** are used across methods
- **Conditioning is method-specific:** Different variables control the corresponding mismatch effects for each intercalibration method
- **Bootstrap provides the finite-sample contribution**, while Monte Carlo propagation represents measurement/model uncertainty; together they quantify how accurately the bias can be established from a one-month dataset
- **The framework is broadly transferable to other sample-based inter-calibration methods**. Block or cluster bootstrap is recommended where the matchup data exhibit dependence or natural grouping.

Bootstrap in the Sterna intercalibration study: Questions

Thank you

Questions?


INDRA
UPC
UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH