CALIBSUN

SOLAÏS

Identifying statistical behaviors explaining the different performances of site-adaptation of GHI depending on the satellite database

Loïc Yezeguelian

Solaïs
Centre O.I.E., Mines Paris - PSL

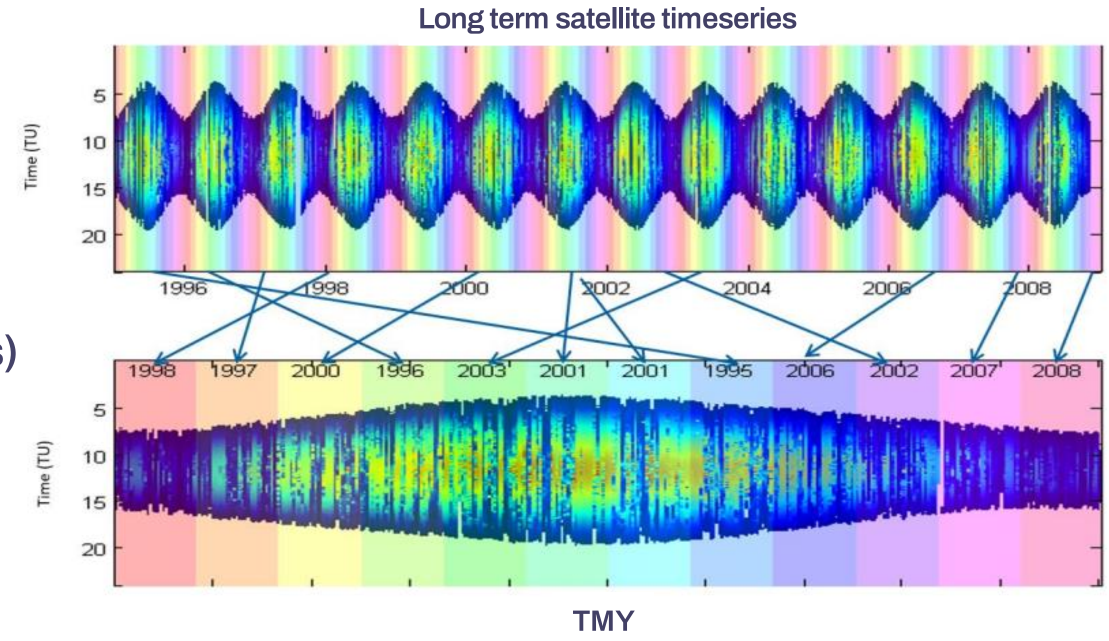
icem 2023
27 - 29 JUNE 2023
PADOVA, ITALY
#icem2023Italy

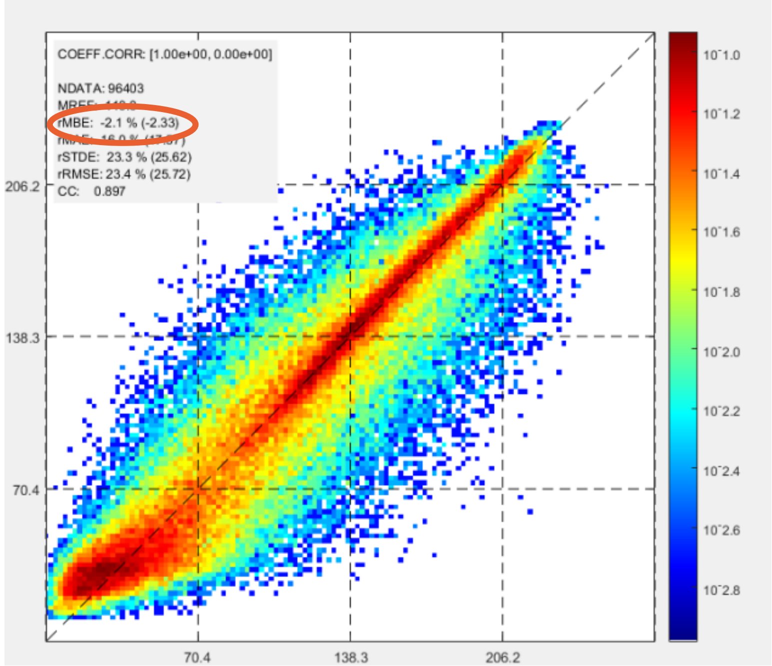
MINES PARIS | PSL

SITE ADAPTATION FOR BANKABLE DATASET

TMY: Typical Meteorological Year

- Paradigm used to build all yield reports
- Based on satellite historical timeseries (10 to 20 years)





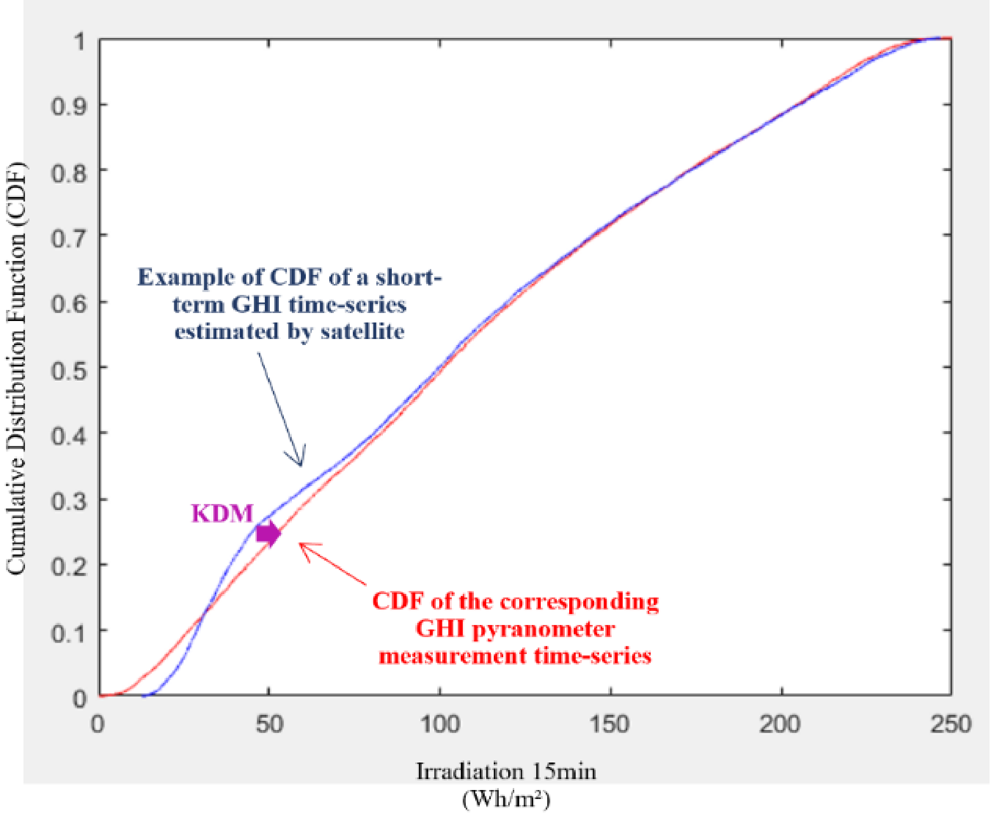
Satellite databases: timeseries may be biased

- 3.5% uncertainty on long term MBE (%) for the best providers
- Less precise P50 yield and greater P50/P90 gap

➢ Need to calibrate using a short-term measurement campaign (minimum 1 year)

KERNEL DENSITY MAPPING CALIBRATION (KDM)

- GOAL: map the cumulative distribution functions on a short term campaign to remove the long term bias
- WHY KDM: benchmarked as one of the most effective bias removal algorithms [1]



CDF are built using a KDE:

$$\forall s \in \mathbb{R}, \quad \begin{cases} PDF(s) = \dfrac{1}{Nh}\sum_{k=1}^{N} \dfrac{1}{\sqrt{2\pi}} e^{-\frac{(s-x_k)^2}{2h^2}} \\ CDF(s) = \int_0^s PDF(t)dt \end{cases}$$

Calibration function:

$$GHI_{LT\,sat_{calib}} = CDF_{ST\,meas}^{-1} \circ CDF_{ST\,sat}(GHI_{LT\,sat_{raw}})$$

[1] Jesus Polo, Carlos Fernández-Peruchena, Vasileios Salamalikis, Luis Mazorra-Aguiar, Mathieu Turpin, Luis Martín-Pomares, Andreas Kazantzidis, Philippe Blanc, Jan Remund, Benchmarking on improvement and site-adaptation techniques for modeled solar radiation datasets, Solar Energy, Volume 201, 2020, Pages 469-479

ASSESS CALIBRATION PERFORMANCE

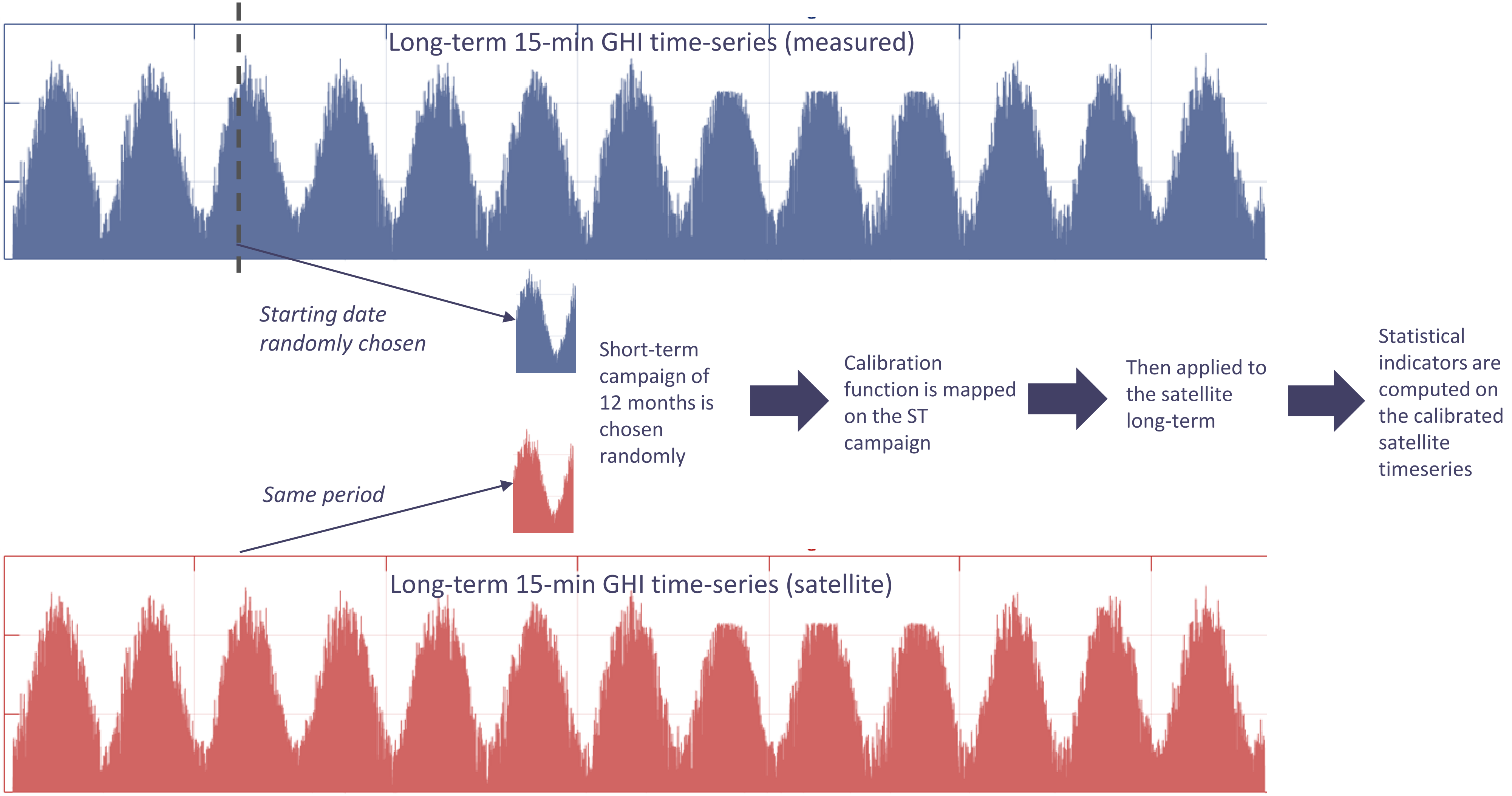
Validation dataset

- Quality long-term pyranometer data were needed
- Seven BSRN stations were selected

BSRN station	Period used
Cabauw	2007-2018
Camborne	2004-2017
Carpentras	2007-2018
Cener	2009-2018
Lindenberg	2004-2017
Palaiseau	2007-2018
Payerne	2004-2013

ASSESS CALIBRATION REMAINING BIAS

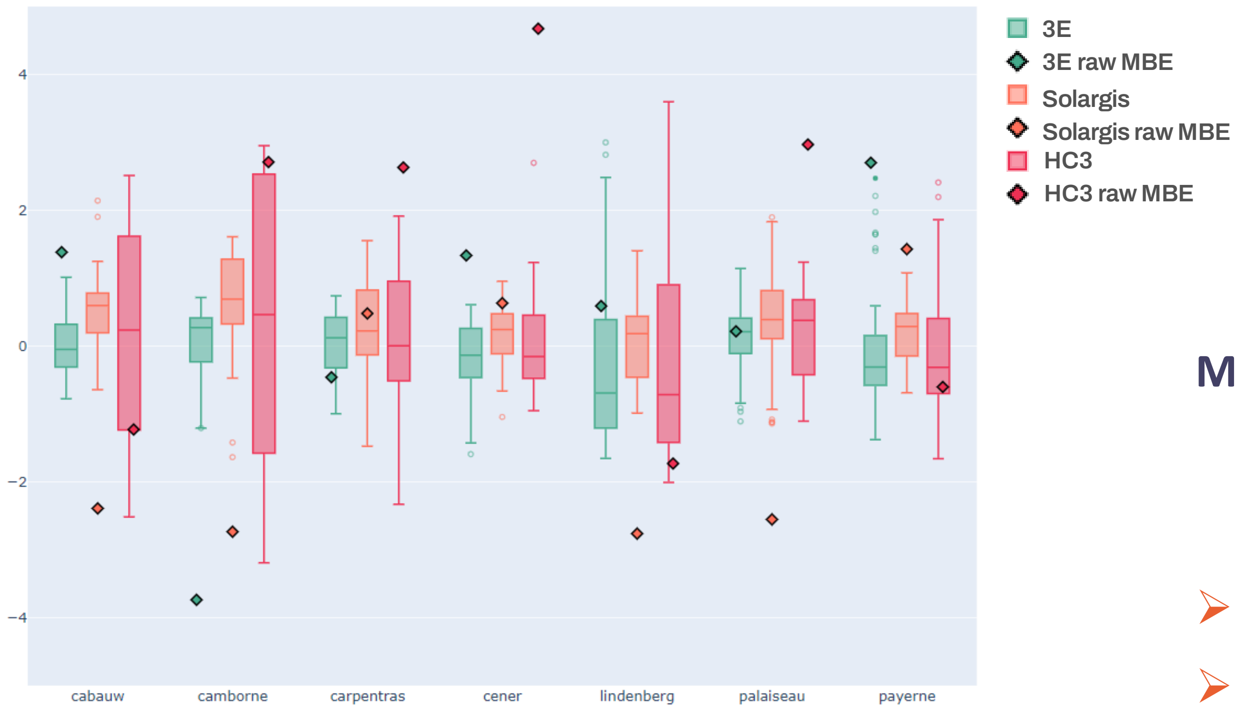
Monte Carlo Framework



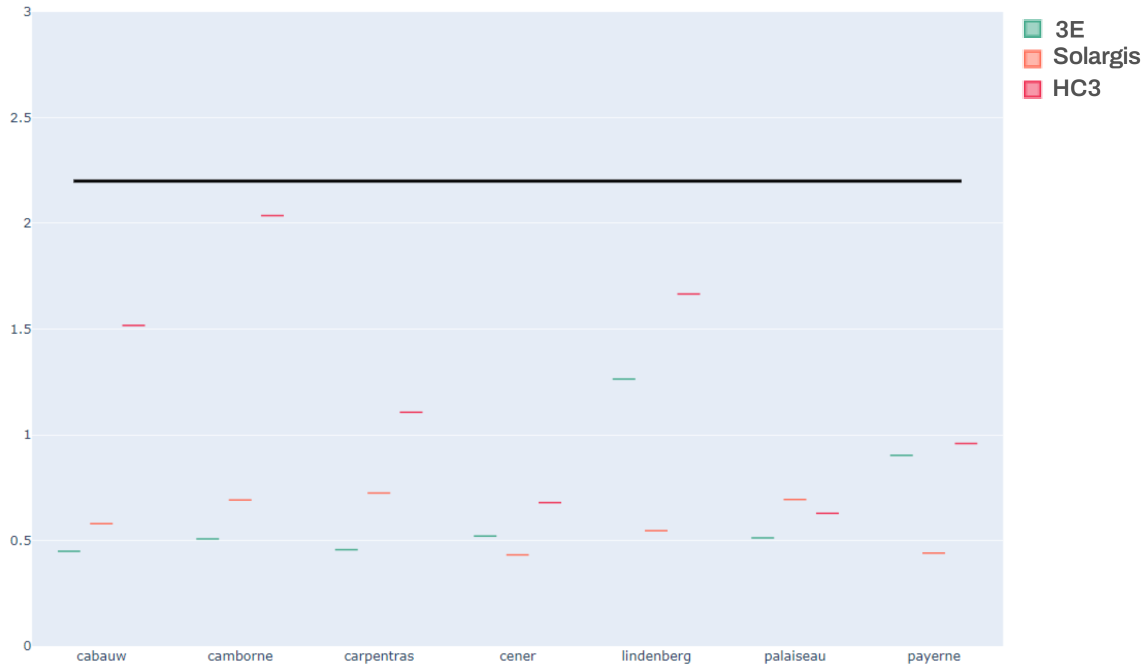
RESULTS

- 12 months-long campaigns
- 100 random campaigns per BSRN station

MBE (%) boxplot - 15min



Standard deviation MBE (%) boxplot - 15min



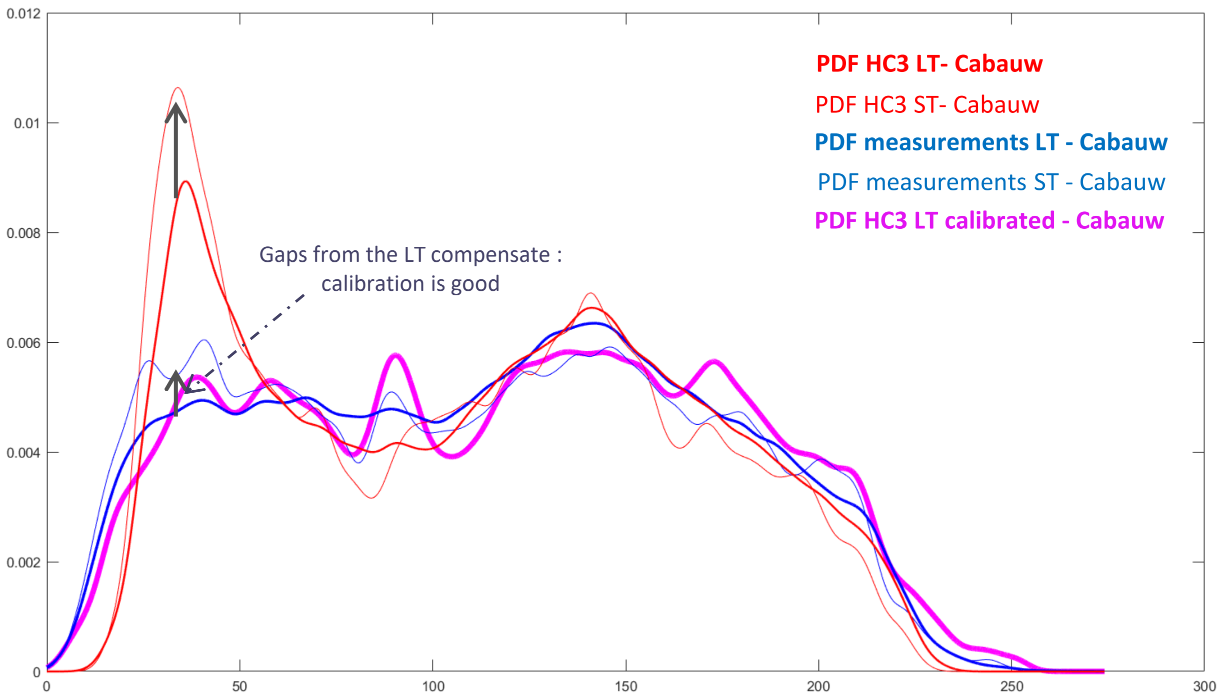
MBE STD (%) :

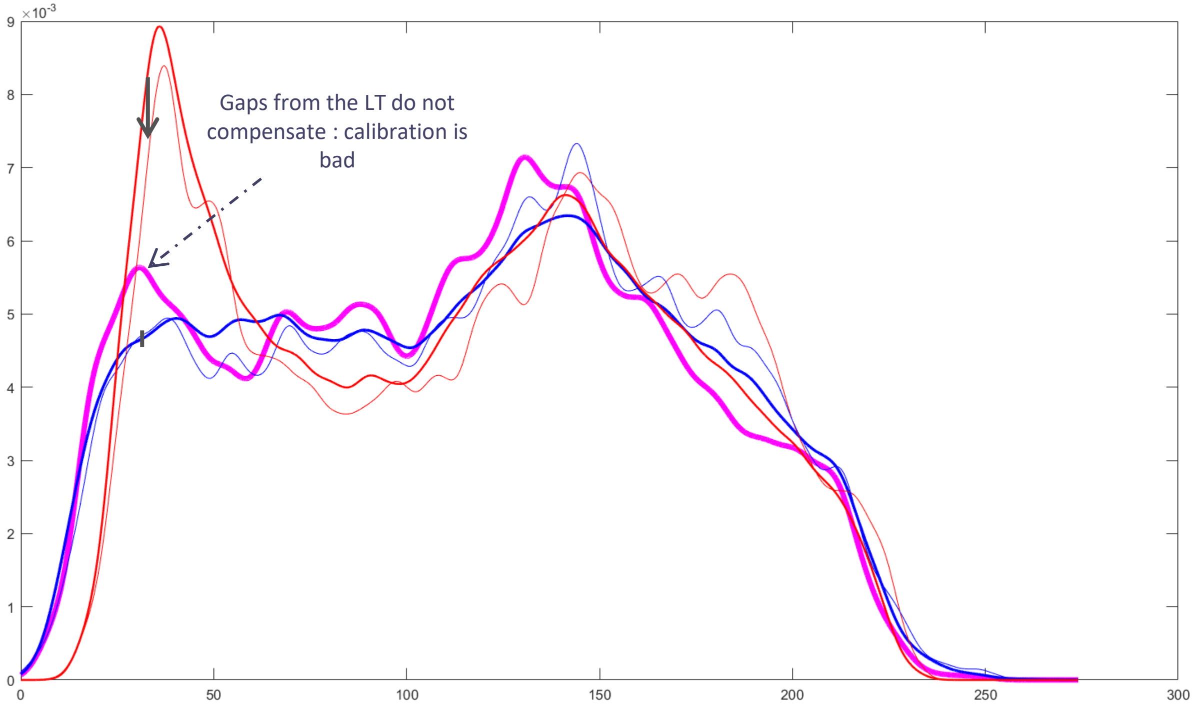
3E: 0.73% Solargis: 0.63% HC3: 1.33%

- ➢ **Why is it non null ?**
- ➢ **Why is it dependent on the database ?**

VISUAL INTERPRETATION

- Calibration errors do not originate from the method, but **from the data !**
- When short-term data (meas and sat) don't have the same distribution « error » compared to their respective long-term, **the satellite distribution is not properly map**





POSSIBLE EXPLANATIONS

- NOT STATISTICALLY STABLE?
  - One year of measurements is not enough?
  - ~96*365/2 = 17520 sample points
  - KDE to smooth the distribution
- NON STATIONNARY?
  - LT and ST do not follow the same distributions
  - ST distribution depends on the starting date
- BSRN?
  - Class A pyrano. 1,5% error at 1min timestep
  - But resampled at 15min
  - And we only consider yearly bias
- SATELLITE?
  - Assumed to be stable over time

➢ Let's consider the i.i.d. case to see if we can determine any persistent bias

CALIBSUN MINES PARIS | PSL SOLAÏS icem 2023

OBSERVED STATISTICAL BEHAVIOR

- We have managed to demonstrate that for quantile mapping calibration in the i.i.d. case:

$$\Delta_{LT}^{*} = \Delta_{LT} - \Delta_{ST} + \varepsilon_{ST}$$

Post-calibration LT bias → Δ_{LT}^{*}; Raw LT bias → Δ_{LT}; Raw ST bias → Δ_{ST}

with ε_{ST} being a known differential indicator

$$\varepsilon_{ST} = h^2 \sum_{l=1}^{b} \delta_{ST,l} \sum_{p=l}^{b} \frac{\delta_p - \eta_{ST}[\{M\}_{ST}]_p}{f_{ST}[\{M\}_{ST}, r]_p}$$

- The visual interpretation lets us think that:

Calibration seems to work well when at any point of the interval $CDF_{ST_{sat}} - CDF_{LT_{sat}} = CDF_{ST_{mes}} - CDF_{LT_{mes}}$

So let's consider the average variable $\hat{\Delta} = \frac{1}{GHI_{max}} \int_0^{GHI_{max}} (CDF_{ST_{sat}} - CDF_{LT_{sat}}) - (CDF_{ST_{mes}} - CDF_{LT_{mes}})$

- We get the following:

$$\hat{\Delta} = \frac{\Delta_{LT} - \Delta_{ST}}{GHI_{max}} = \frac{\Delta_{LT}^{*} - \varepsilon_{ST}}{GHI_{max}}$$

meaning that

$$\Delta_{LT}^{*} = GHI_{max}\hat{\Delta} + \varepsilon_{ST}$$

GAUSSIAN CASE

➢ We tried to understand ε_{ST} by identifying with the gaussian case:

- GAUSSIAN CASE (i.i.d.): if Satellite $\sim N\left(\mu_S, \sigma_S^2\right)$ and Measurements $\sim N\left(\mu_M, \sigma_M^2\right)$

$$\Delta_{LT}^{*} = \frac{\widehat{\sigma_{M_{ST}}}}{\widehat{\sigma_{S_{ST}}}}\left(\widehat{\mu_{S_{LT}}} - \widehat{\mu_{S_{ST}}}\right) - \left(\widehat{\mu_{M_{LT}}} - \widehat{\mu_{M_{ST}}}\right)$$

so in this case

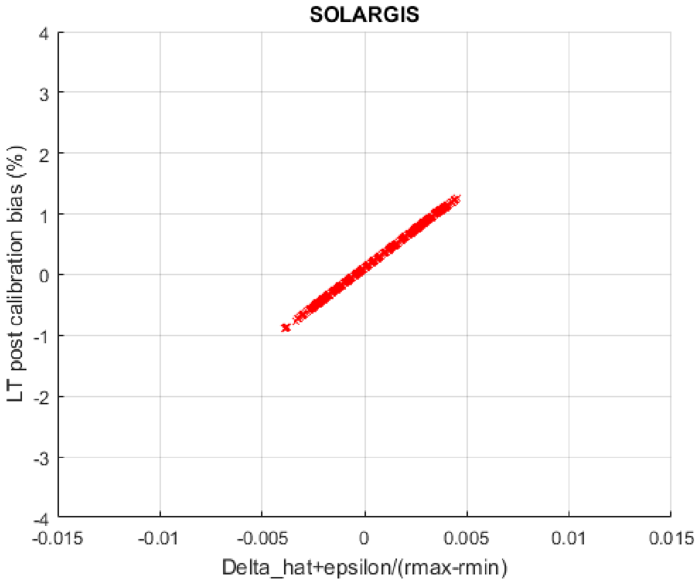
$$\boxed{\varepsilon_{ST} = \frac{\widehat{\sigma_{M,ST}} - \widehat{\sigma_{S,ST}}}{\widehat{\sigma_{S,ST}}}\left(\widehat{\mu_{S,LT}} - \widehat{\mu_{S,ST}}\right)}$$

➢ So that ε_{ST} has good chances to be negligeable

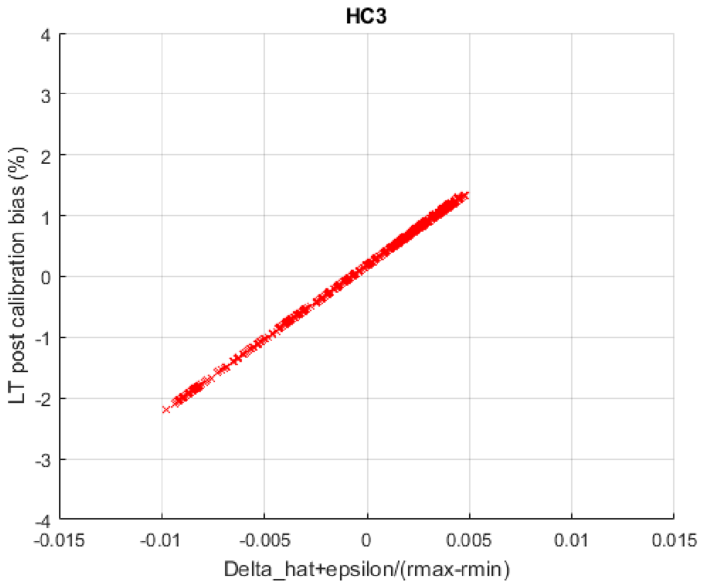
VALIDATION

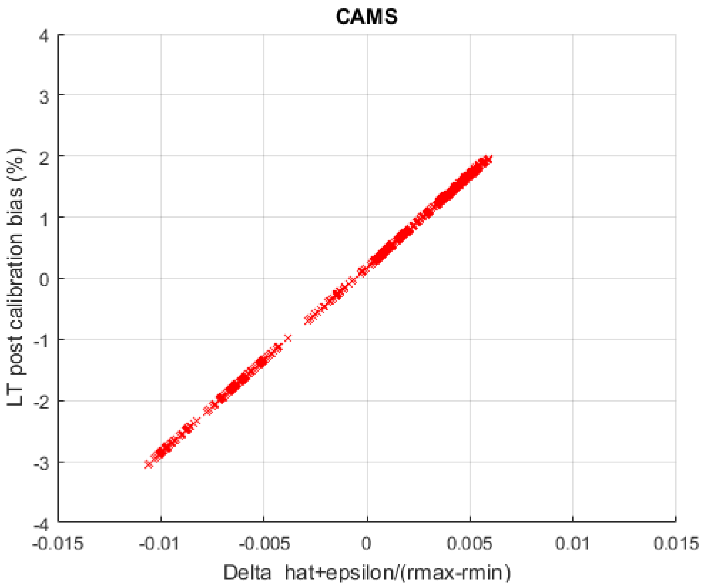
- Previously described Monte-Carlo experiment run on CARPENTRAS:

$\Delta^*_{N\%}$



$$\hat{\Delta} + \frac{\varepsilon_{ST}}{GHI_{max}}$$





VALIDATION

- SCATTERPLOT REGRESSION ON 100 CAMPAIGNS:

BSRN	Database	$\Delta^*_{LT} = \alpha\hat{\Delta} + \beta$			$\Delta^*_{LT} = \alpha(\hat{\Delta} + \frac{\varepsilon_{ST}}{GHI_{max}}) + \beta$			$\widehat{GHI}_{max}$
		α	β	CC	α	β	CC	
CAB	Solargis	285.2	0.272	0.9835	249.1	0.160	0.9999	**249.2**
	HC3	275.6	0.303	0.9969	249.0	0.193	0.9999	**248.1**
	CAMS	213.7	0.166	0.9989	241.6	0.088	0.9999	**245.0**
CAM	Solargis	337.7	0.393	0.9852	255.8	0.115	0.9999	**253.2**
	HC3	284.6	0.278	0.9671	267.2	0.194	0.9998	**259.3**
	CAMS	254.5	0.352	0.9883	250.1	0.150	0.9999	**249.2**
CAR	Solargis	279.5	0.072	0.9967	268.7	0.109	1.0000	**268.0**
	HC3	269.0	0.193	0.9978	265.2	0.196	0.9999	**266.0**
	CAMS	242.3	-0.155	0.9896	266.7	0.154	1.0000	**269.9**
CEN	Solargis	300.4	0.277	0.9598	273.8	0.196	0.9997	**271.6**
	HC3	269.9	0.216	0.9723	265.5	0.199	0.9997	**267.4**
	CAMS	261.5	-0.0005	0.9965	267.7	0.022	0.9999	**271.3**
LIN	Solargis	263.0	0.120	0.9773	246.4	0.103	0.9998	**247.9**
	HC3	274.1	0.265	0.9980	249.1	0.218	1.0000	**249.2**
	CAMS	229.8	0.027	0.9977	243.0	0.086	1.0000	**244.6**
PAL	Solargis	229.5	-0.376	0.9755	255.8	0.059	0.9997	**258.3**
	HC3	254.0	0.178	0.9808	251.3	0.125	0.9999	**251.9**
	CAMS	253.4	0.043	0.9936	254.9	0.041	0.9999	**253.1**
PAY	Solargis	267.8	0.147	0.9864	268.6	0.160	0.9996	**265.3**
	HC3	280.4	0.239	0.9935	264.8	0.199	1.0000	**264.1**
	CAMS	243.1	0.104	0.9959	266.2	0.044	0.9998	**263.2**

➢ ε_{ST} is indeed relatively small

- $\hat{\Delta} + \frac{\varepsilon_{ST}}{GHI_{max}}$ STANDARD DEVIATION ON 100 CAMPAIGNS:

	SOLARGIS	HC3	CAMS
CAB	0,002	0,0048	0,0054
CAM	0,0021	0,0020	0,0044
CAR	0,0022	0,0041	0,0057
CEN	0,0015	0,0027	0,0078
LIN	0,0022	0,0052	0,0051
PAL	0,0028	0,0026	0,0048
PAY	0,0016	0,0025	0,0017

➢ Calibration overall performance is better when systematic errors are consistent from one year to another

- Even for i.i.d. quantile mapping:

$$\Delta_{LT}^{*} = \Delta_{LT} - \Delta_{ST} + \varepsilon_{ST}$$

- For naive bias correction $\varepsilon_{ST} = 0$

- The longer the ST the better: $\Delta_{LT} - \Delta_{ST} \rightarrow 0$

- Satellite databases don't have the same calibration performance, because their $\hat{\Delta}$ distribution (=unconsistency of the error throughout years) are different

- For a same $\hat{\Delta}$ distribution, calibration works better when GHI_{max} is high

- But even for 1 year and no i.i.d. assumption, really decent performance (<1% on most sites and with most providers)

➢ **CONJECTURE: same results for other calibration methods**

CONCLUSION