The statistical approach to the validation and intercomparison of reference measurements

Alessandro Fassò

ICM-16 - Tenerife - 18 November 2025

UNIVERSITÀ
DEGLI STUDI
DI BERGAMO

Dipartimento
di Scienze Economiche

Reference measurement comparison

In this talk, I will consider some statistical issues when comparing two GDPs or other ref measurement products.
In particular, I will consider mainly two issues:

- Comparing the ability to understand climatology
- Assessing measurement compatibility

I will focus primarily in RSs which provide high resolution profiles,

but other products and issues with a limited number of layers and/or co-location issues will be also considered.

For simplicity, I will use RS examples on T and U.

Reference measurement comparison

In this talk, I will consider some statistical issues when comparing two GDPs or other ref measurement products.
In particular, I will consider mainly two issues:

- Comparing the ability to understand climatology
- Assessing measurement compatibility

I will focus primarily in RSs which provide high resolution profiles,

but other products and issues with a limited number of layers and/or co-location issues will be also considered.

For simplicity, I will use RS examples on T and U.

Reference measurement profiles

The underlying physical process is denoted by

$$z(s, t)$$

- $s = (lat, lon, z)$ moves around the atmospheric shell
- t is time
- z may be an array of variables, e.g. $z = (T, U, p, W, ...)$

Reference measurement profiles

- We have partial and noisy observations of z.
- In particular, we consider atmospheric profiles (soundings or other).
- Namely the i-th sounding is represented as

$$y_{t_{i,j}} = z(s_{i,j}, t_{i,j}) + \varepsilon(s_{i,j}, t_{i,j})$$

  - $j = 1, 2, ..., B_i$ is seconds from launch
  - B_i is the top time (burst)
  - $t_{i,j} = LaunchTime_i + j$
  - $^*s_j = s^0(t_j) + \varepsilon^s(t_j)$ is not error free and s^0 can be part of z.

*I will omit i wherever possible.

Uncertainty

Considering a single component of y, its uncertainty is given by

$$u_j^2 = E[\varepsilon^2(t_j)]$$

If y and ε are vectors, should we have a matrix uncertainty

$$U_j = E[\varepsilon(t_j)\varepsilon'(t_j)] \quad ?$$

Uncertainty

Considering a single component of y, its uncertainty is given by

$$u_j^2 = E[\varepsilon^2(t_j)]$$

If y and ε are vectors, should we have a matrix uncertainty

$$U_j = E[\varepsilon(t_j)\varepsilon'(t_j)] \quad ?$$

Comparison type

We have two instruments, "A" and "B" and profiles

$$y^A_{t_{A,j}} \text{ and } y^B_{t_{B,j}}$$

If the two instruments are synchronised (dual launches) $t_{A,j} = t_{B,j} = t_j$ the difference

$$\Delta_j = y^A_{t_j} - y^B_{t_j}$$

is immediately computed.

Otherwise, some rilocation/collocation technique must be used.

Type A approach

Recent literature considered the comparisons between ensembles of profiles, focussing on bias:

$$\hat{\Delta}(S,P) = \frac{1}{N_{S,P}} \sum_{i,j \in S,P} \Delta_{i,j}$$

and Standard Deviation $\sigma(S,P)$ where

$$\sigma^2(S,P) = \frac{1}{N_{S,P}} \sum_{i,j \in S,P} (\Delta_{i,j} - \Delta_{S,P})^2$$

Comparison of iMS-100 GDP and RS92 GDP

Hoshino et al (2022) considered $n = 99$ dual soundings based on Meisei iMS-100 and Vaisala RS92 radiosondes at Tateno, Japan.

Type A - Statistical (ensamble based)

iMS - RS92

Atmos. Meas. Tech., 15, 5917–5948, 2022
https://doi.org/10.5194/amt-15-5917-2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Atmospheric Measurement Techniques
Open Access
EGU

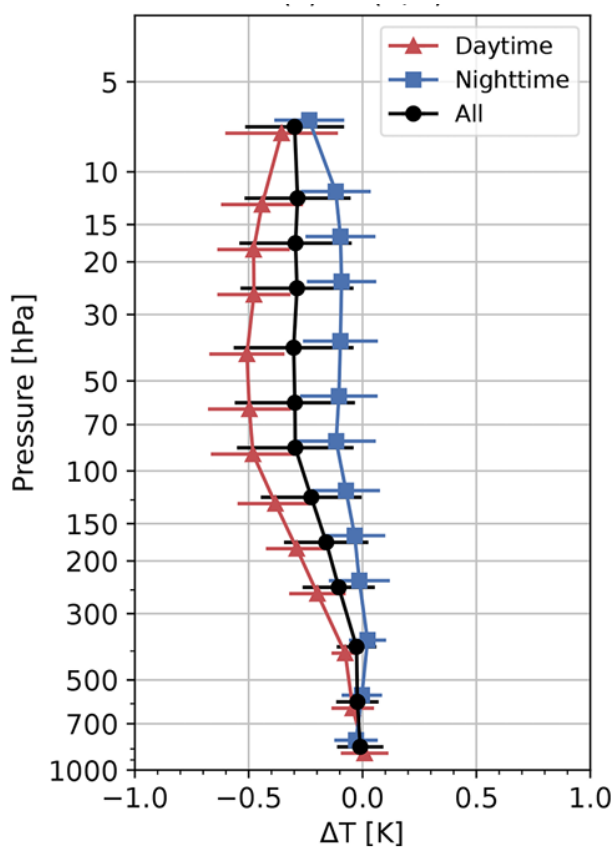
Comparison of GRUAN data products for Meisei iMS-100 and Vaisala RS92 radiosondes at Tateno, Japan

Shunsuke Hoshino[1], **Takuji Sugidachi**[2], **Kensaku Shimizu**[2], **Eriko Kobayashi**[3], **Masatomo Fujiwara**[4], **and Masami Iwabuchi**[3]

[1]Office of Numerical Prediction Modeling Fundamental Technology, Numerical Prediction Division, Information Infrastructure Department, Japan Meteorological Agency, 1–2 Nagamine, Tsukuba, Ibaraki 305-0052, Japan
[2]Meisei Electric Co., Ltd., 2223 Naganumamachi, Isesaki, Gunma 372-8585, Japan
[3]Aerological Observatory, 1–2 Nagamine, Tsukuba, Ibaraki 305-0052, Japan
[4]Faculty of Environmental Earth Science, Hokkaido University, Kita 10 Nishi 5, Kita-ku, Sapporo 060-0810, Japan

Correspondence: Shunsuke Hoshino (shoshino@met.kishou.go.jp)

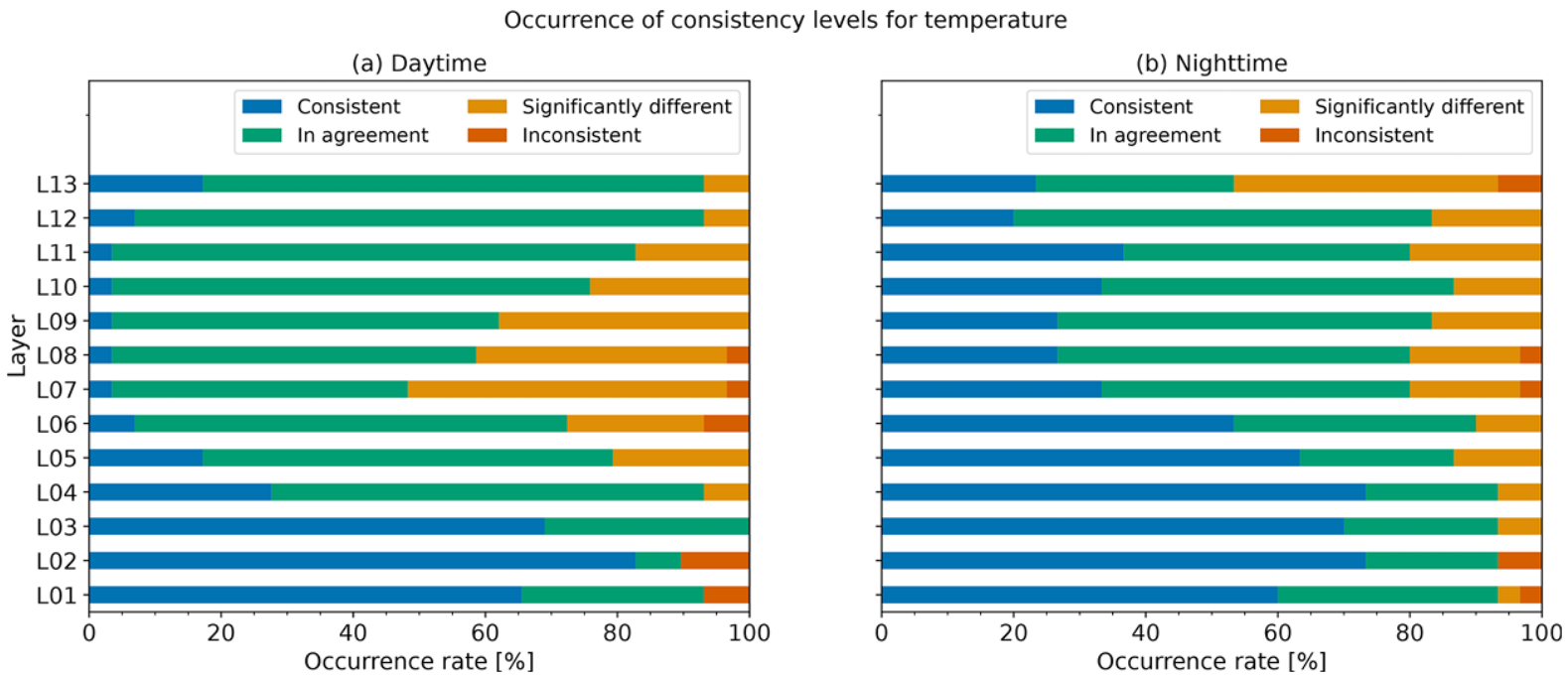
Received: 8 November 2021 – Discussion started: 23 February 2022
Revised: 19 September 2022 – Accepted: 19 September 2022 – Published: 19 October 2022



Comparison of iMS-100 GDP and RS92 GDP†

Type B/A - (individual/ensamble)

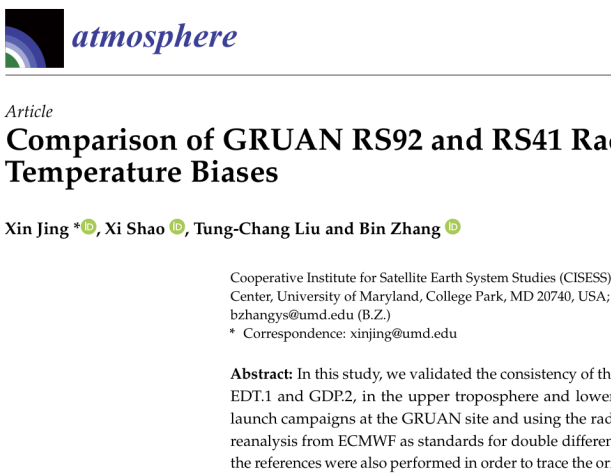
$\lvert m_1 - m_2 \rvert < k\sqrt{u_1^2 + u_2^2}$	True	False	Significance level
$k = 1$	Consistent	Suspicious	32 %
$k = 2$	In agreement	Significantly different	4.5 %
$k = 3$	–	Inconsistent	0.27 %



†Thanks to Hoshino et al, 2022

Comparison of RS41 GDP and RS92 GDP

Jing et al (2023) compared RS41 and RS92 GDP.2s, using $\sim$ 1000 profiles at 10 GRUAN stations.



atmosphere

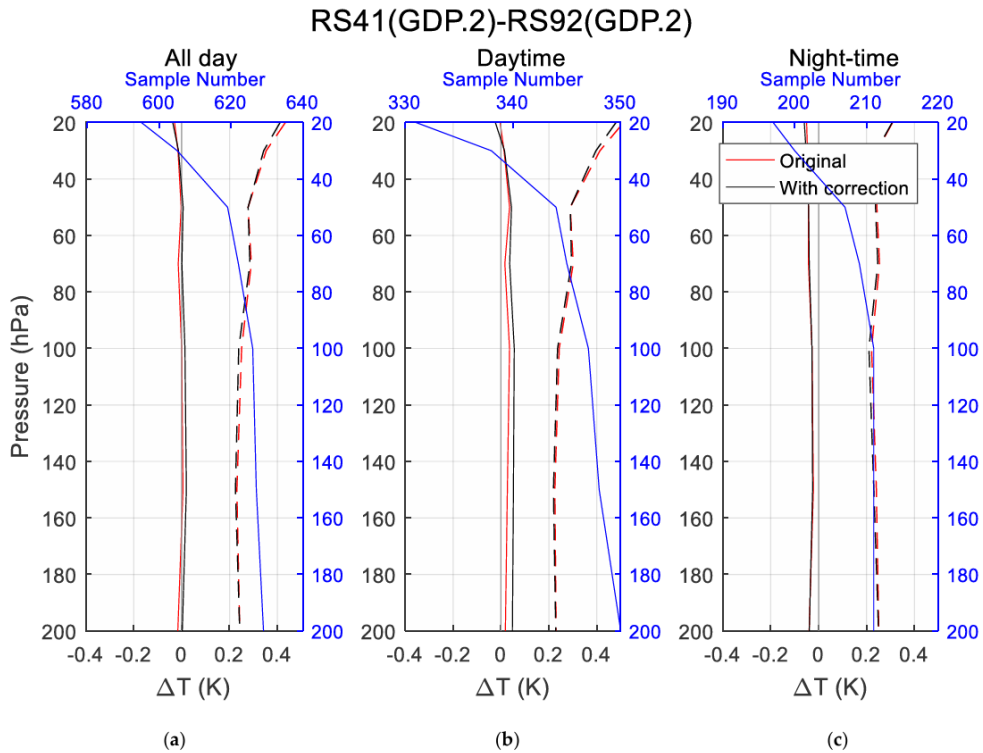
Article

Comparison of GRUAN RS92 and RS41 Ra Temperature Biases

Xin Jing *, Xi Shao, Tung-Chang Liu and Bin Zhang

Cooperative Institute for Satellite Earth System Studies (CISESS) Center, University of Maryland, College Park, MD 20740, USA; bzhangys@umd.edu (B.Z.)
* Correspondence: xinjing@umd.edu

Abstract: In this study, we validated the consistency of th EDT.1 and GDP.2, in the upper troposphere and lowe launch campaigns at the GRUAN site and using the rad reanalysis from ECMWF as standards for double differer the references were also performed in order to trace the or



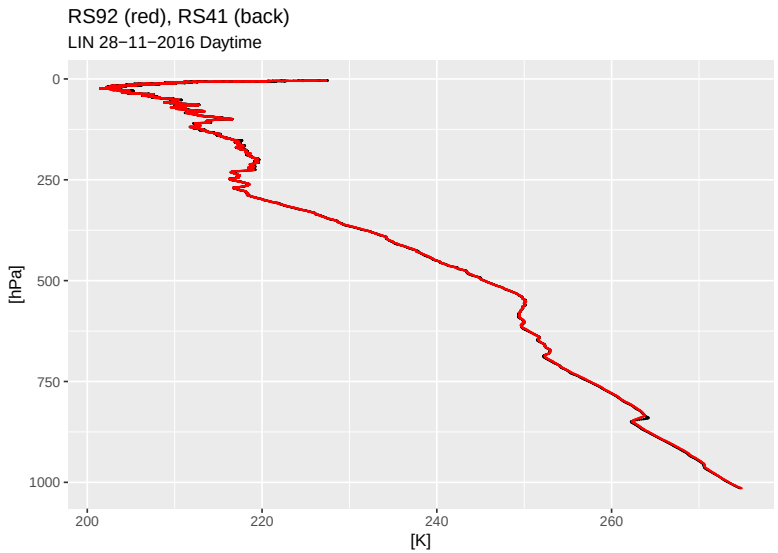
Further developments

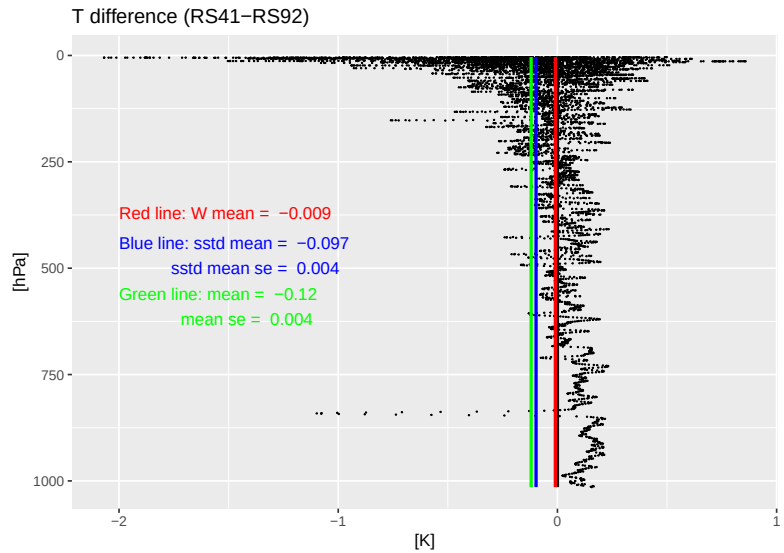
In the analysis of GDP differences $\Delta_{i,j}$ we may consider some additional issues:

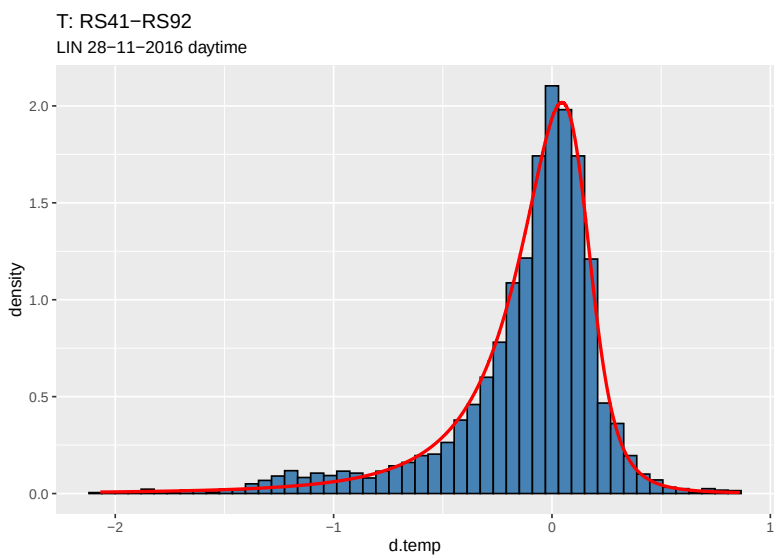
- Is interpolation an issue?
- Should we use uncertainties?
- Are we limited to the Gaussian distribution?
- What about vertical and spatial autocorrelation?
- What about cross-correlation between different variables?

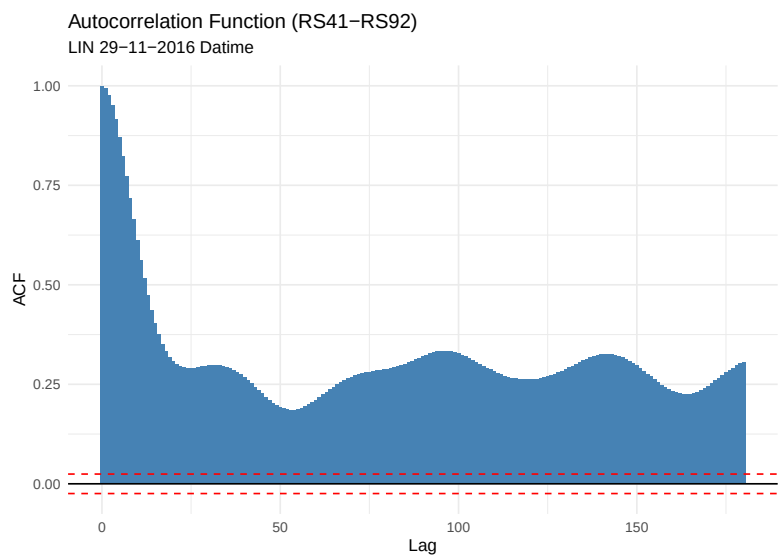
Toy dataset: RS41-RS92 GDP comparison

LIN 28-11-2016 Daytime









Interpolation

Interpolation "may" have two opposite effects:

- smoothing the observed signal
- increasing measurement uncertainty

I will consider

- small gaps (high resolution profiles)
- large gaps (atmospheric layers, mandatory levels, ...)

Interpolation generalities

Atmos. Meas. Tech., 13, 6445–6458, 2020
https://doi.org/10.5194/amt-13-6445-2020
© Author(s) 2020. This work is distributed under
the Creative Commons Attribution 4.0 License.

Atmospheric Measurement Techniques
Open Access
EGU

$$m_t = (1-\alpha(t))y_{t^-} + \alpha(t)y_{t^+}$$

$$t^- < t < t^+$$

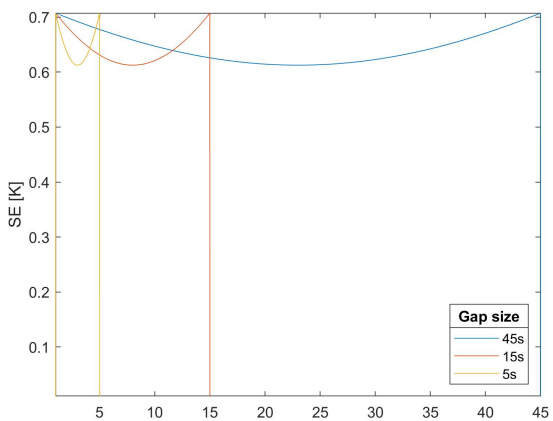
Interpolation uncertainty of atmospheric temperature profiles

Alessandro Fassò[1], Michael Sommer[2], and Christoph von Rohden[2]

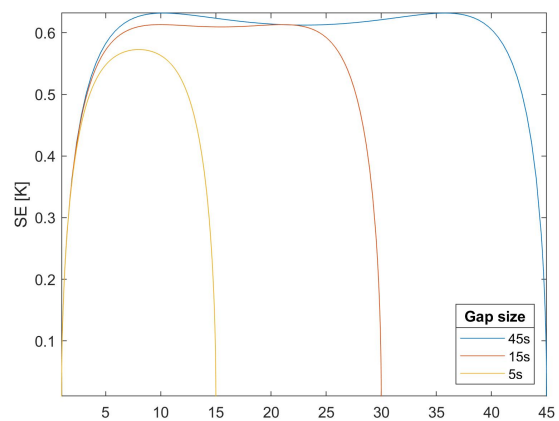
[1]Department of Management Engineering, Information and Production, University of Bergamo, Dalmine, Italy
[2]GRUAN Lead Centre, Deutscher Wetterdienst, Lindenberg, Germany

$$\begin{aligned} u_t^2 &= E[(m_t - z(t))^2] \\ &= \sigma_\epsilon^2 + 2\sigma_y^2\{1-\alpha+\alpha^2\} \\ &+ 2\{\alpha(1-\alpha)\gamma(t^+ - t^-) - \alpha\gamma(t^+ - t) - (1-\alpha)\gamma(t - t^-)\} \end{aligned}$$

White Noise



Autocorrelated



Small gaps

As expected: small gaps ⇒ small interpolation uncertainty.

In the RS41-R92 example, I compared

- time shifting
- linear interpolation to
  - linear alt grid
  - lin press grid
  - log press grid

d	sd(d.temp)	rmse(d.temp)	mean(alt)
-2	0.45	0.46	16960
0	0.41	0.42	16960
2	0.37	0.39	16960
4	0.34	0.37	16960
6	0.33	0.36	16960
8	0.33	0.37	16960
interp.lin.alt	0.46	0.46	17930
interp.lin.press	0.16	0.17	6923
interp.log.press	0.36	0.38	18464

RS41$g.Measurement.StartTime = "2016-11-28T10:45:34.096Z"
RS92$g.Ascent.StartTime = "2016-11-28T10:45:36"

Large Gaps and/or co-location issues

The Immler inequality must be intergated by a third component

$$|m_1 - m_2| < k\sqrt{u_1^2 + u_2^2 + u_{coll}^2}$$

For example Fassò et al (2014, AMT) used functional regression for comparing Beltsville and Sterling RSs.

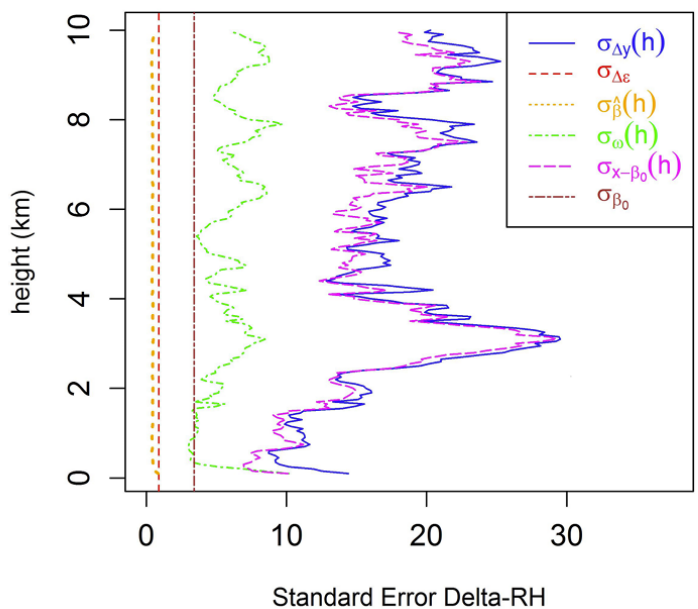


Table 1. Budget of total collocation uncertainty for relative humidity (Δrh) in Beltsville–Sterling, using the HFR model of Eqs. (14) and (15).

Source of uncertainty		$\bar{\sigma}^2$	$\bar{\sigma}^2$%	$\bar{\sigma}$
Total uncertainty	Δ_y	343.8	100 %	18.54
Collocation drift	Δ_μ	343.0	–	18.52
Bias (adjustable)	β_0	11.6	3.4 %	3.40
Environ. error (reducible)	$x \sim \beta_0$	293.2	85.5 %	17.12
Environ. error (irreducible)	ω	37.1	10.8 %	6.09
Sampling error	$\hat{\beta}$	0.21	0.1 %	0.46
Measurement error	Δ_ε	0.81	0.2 %	0.90

to the collocation difference of water vapour content in the dry air mass, as shown by Eq. (14) and Fig. 5. Note that this collocation difference of water vapour content stems from a combination of the short-term variability of water vapour

Considering uncertainty

When comparing GDPs we have uncertainties of all single measurements.
Assuming

$$\Delta_{i,j} \sim NID(\beta, u^2_{A,i,j} + u^2_{B,i,j})$$

the MLE estimation of the bias β is the weighted average:

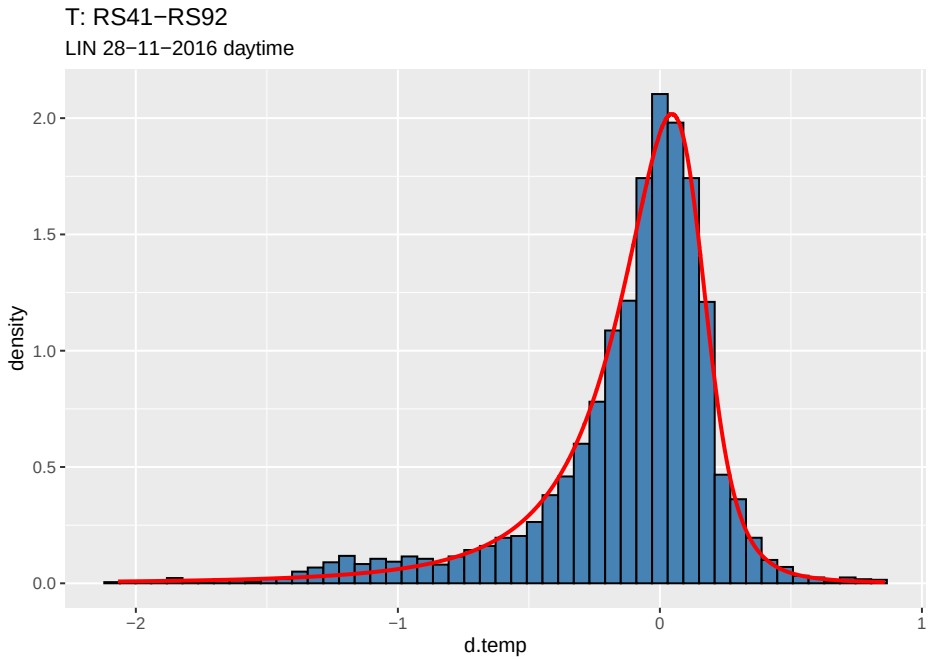
$$\hat{\Delta}(S,P) = \frac{1}{W_{S,P}} \sum_{i,j \in S,P} \Delta_{i,j} w_{i,j}$$

$$w_{i,j} = \frac{1}{u^2_{A,i,j} + u^2_{B,i,j}}$$

Non Gaussian uncertainties

The Skew-t distribution

	est	se
xi	0.325	0.005
omega	0.397	0.010
alpha	-3.585	0.171
nu	2.862	0.141



As a consequence the Immler's inequality is to be substituted by the corresponding percentiles.
For example, for k=2, we have

$$\tilde{y}_{2.5\%} < m_1 - m_2 < \tilde{y}_{97.5\%}$$

which, clearly, is not centred on the mean.

Thanks for your questions!