Report from interlaboratory comparison of air thermometer calibration procedures

EURAMET project 1459

Åge Andreas Falnes Olsen, Dubhaltach MacLochlainn, Carmen García Izquierdo, Denis Smorgon, Regina Deschermeier, Carolyn Eckerleben, Florian Bubser, Michal Voldán, Magnus Holmsten, Peter Pavlasek, Milan Ioan Maniur, Seda Oğuz Aytekin, Paul Carroll, Stephanie Bell, Iska Kolaveri, Christina Hofstätter-Mohler, Jan Nielsen, Peter Rothmund, Reidun Anita Bergerud, Miruna Dobre, Debby Van Den Berghe, Jovan Bojkovski, Patrick Raab, Helmut Mitter, Tanja Vukićević, Alexandra Kowal, Justyna Dobosz, Semir Cohodarevic, Richard Högström, Emese Turzó-András, Eric Georgin, Rafał Jarosz, Slavica Simic, Evmorfia Kokkini, Javier de Lucas Veguillas

19.09.2023

Justervesenet

Authors and affiliations

Åge Andreas Falnes Olsen, Reidun Anita Bergerud, Peter Rothmund	Justervesenet (Norway)
Dubhaltach MacLochlainn	NSAI (Ireland)
Carmen García Izquierdo	CEM (Spain)
Denis Smorgon	INRiM (Italy)
Regina Deschermeier, Carolyn Eckerleben, Florian Bubser	PTB (Germany)
Michal Voldán	CMI (Czech Republic)
Magnus Holmsten	RISE (Sweden)
Peter Pavlasek, Milan Ioan Maniur	SMU (Slovakia)
Seda Oğuz Aytekin	TUBITAK (Türkie)
Paul Carroll, Stephanie Bell	NPL (Great Britain)
Iska Kolaveri	DPM (Albania)
Christina Hofstätter-Mohler	BEV (Austria)
Jan Nielsen	DTI (Denmark)
Miruna Dobre, Debby Van Den Berghe	SMD (Belgium)
Jovan Bojkovski	UL (Slovenia)
Patrick Raab, Helmut Mitter	BEV E+E (Austria)
Tanja Vukićević	MBM (Montenegro)
Alexandra Kowal, Justyna Dobosz	INTiBS (Poland)
Semir Cohodarevic	IMBiH (Bosnia and Herzegovina)
Richard Högström	VTT MIKES (Finland)
Emese Turzó-András	BFKH (Hungary)
Eric Georgin	LNE CETIAT (France)
Rafał Jarosz	GUM (Poland)
Slavica Simic	DMDM (Serbia)
Evmorfia Kokkini	EIM (Greece)
Javier de Lucas Veguillas	INTA (Spain)

Contents

1 Introduction

The interlaboratory comparison (ILC) was initiated in 2019 as an important first step in the EURAMET project 1459. The aim of the ILC was first and foremost to gain knowledge later used in developing guidelines for air thermometry, particularly within the realm of calibration and dissemination of the temperature scale. While the ILC protocol asked the participants to report a set of main data that followed a rather strict formular, they were also encouraged to perform different characterisation experiments and to explore different strategies for calibration. The idea was to have a set of common baseline data and analyse them in a traditional ILC framework, but at the same time gather exploratory data that could help advance the knowledge in air temperature metrology. The results were intended to be used as input to guidance documents for the calibration of thermometers for air temperature measurements, and possibly to some initial best practice guides for practical air thermometry in the field. The data used in this report is deposited in Zenodo [1].

Air temperature metrology finds obvious applications in many different fields, such as meteorology and climate studies. It is also important in metrology fields such as length, mass and humidity, as well as in cold chain management for the pharmaceutical and biotech industries. In the automotive industry it is also an important parameter in various quality control and production stages, such as accelerated aging tests and surface coating. Within the humidity community precision air temperature measurements have long been recognised as a crucial, but challenging activity. A previous ILC was carried out between 2009 and 2012 where two different probes were circulated between 20 European laboratories, with the temperature ranging between -40 °C and 100 °C [2]. The deviations from consensus ranged from 2 mK to more than 200 mK, with uncertainties spanning 30 mK to almost 400 mK. The present ILC expands the scope by both having a wider range of thermometer models, a much larger collection of travelling standards, and expands the temperature range down to -80 °C.

Recent work by de Podesta *et al* [3] has pointed out a fundamental challenge with precision metrology of air temperature. It was found that in the presence of a heating source, either as irradiation or internal heat dissipation, the temperature error exhibits a dimensional dependence such that larger thermometers lead to a larger error. Furthermore, the effect is exacerbated for smaller wind speeds.

In the present ILC, probes of diverse geometry and design were circulated in an attempt to elucidate some of the issues in air temperature metrology. The probes were calibrated at -40 °C to +60 °C in steps of 20 °C, with a subset of participants going down to -80 °C. It was organised in three loops, with a single link laboratory (JV). The topology is illustrated in Figure 1.

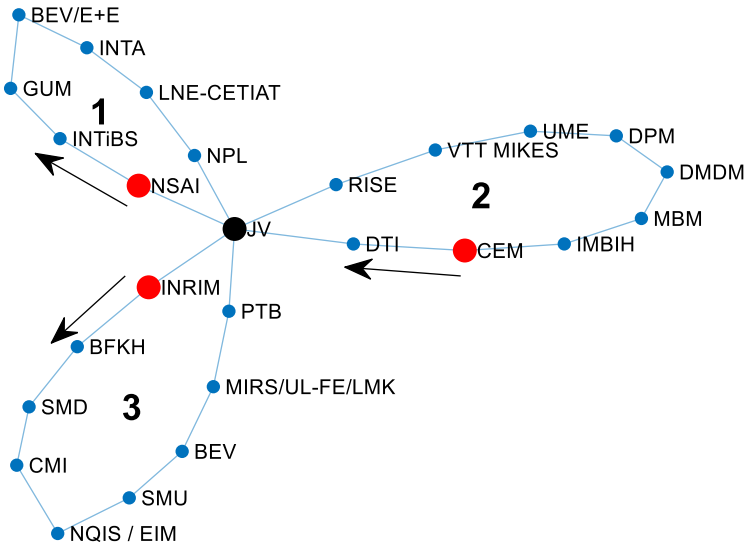


Figure 1 Overview of the ILC topology. JV acted as link laboratory between 3 loops. NSAI, CEM and INRiM were pilots in loops 1, 2 and 3, respectively.

Some participants could not reach the lowest designated temperature in their loop.

Table 1 Overview of participant contributions, with actual temperatures realised, and the type of observations reported.

Name	Role	Temperatures (°C)	Quantities
NSAI	PILOT	-80, -60, -40, -20, 0, 20, 40, 60	Air, liquid, selfheating
INTiBS	PARTICIPANT	-80, -60, -40, -20, 0, 20, 40, 60	Air
GUM	PARTICIPANT	-80, -60, -40, -20, 0, 20, 40, 60	Air, selfheating
BEV/E+E	PARTICIPANT	-80, -70, -60, -40, -20, 0, 20, 40, 60	Air
INTA	PARTICIPANT	-60, -40, -20, 0, 20, 40, 60	Air
LNE-CETIAT	PARTICIPANT	-80, -70, -60, -40, -20, 0, 20, 40, 60	Air
NPL	PARTICIPANT	-80, -60, -40, -20, 0, 20, 40, 60	Air
JV	LINK	-40, -20, 0, 20, 40, 60	Air, selfheating
CEM	PILOT	-40, -20, 0, 20, 40, 60	Air, selfheating, liquid, hysteresis
DTI	PARTICIPANT	-40, -20, 0, 20, 40, 60	Air
RISE	PARTICIPANT	-40, -20, 0, 20, 40, 60	Air
VTT MIKES	PARTICIPANT	-40, -20, 0, 20, 40, 60	Air
UME	PARTICIPANT	-40, -20, 0, 20, 40, 60	Air
DPM	PARTICIPANT	-20, 0, 20, 40, 60	Air
DMDM	PARTICIPANT	-20, 0, 20, 40, 60	Air
MBM	PARTICIPANT	-40, -20, 0, 20, 40, 60	Air
IMBIH	PARTICIPANT	-40, -20, 0, 20, 40, 60	Air
INRIM	PILOT	-40, -20, 0, 20, 40, 60	Air, liquid, selfheating
BFKH	PARTICIPANT	-40, -20, 0, 20, 40, 60	Air
SMD	PARTICIPANT	-40, -20, 0, 20, 40, 60	Air
CMI	PARTICIPANT	-40, -20, 0, 20, 40, 60	Air
NQIS / EIM	PARTICIPANT	-40, -20, 0, 20, 40, 60	Air
SMU	PARTICIPANT	-40, -20, 0, 20, 40, 60	Air
BEV	PARTICIPANT	-40, -20, 0, 20, 40, 60	Air
MIRS/UL-FE/LMK	PARTICIPANT	-40, -20, 0, 20, 40, 60	Air
PTB	PARTICIPANT	-40, -20, 0, 20, 40, 60	Air

1.1 Probes

The circulating probes comprised 8 different models from 6 different manufacturers. Separate items were circulated in the loops, bringing the total number of circulating probes to 23 (the Physicus probe specimen in loop 3 was unstable during pre-circulation tests at the pilot, and the pilots and coordinator decided to exclude it). The table shows an overview of the probes used.

Table 2 Overview of probes used in the circulation. The diameter listed is at the presumed location of the sensing element. The Calpower and Physicus probes have a thicker shaft closer to the leads. The Wika TR60 probe was designed with ventilation fins, probably to maximise the surface area.

Model	Serial number	Dimensions	Loop
BEV E+E probe	B-1	Ø6 mm/L230 mm	1
	B-3		2
	B-5		3
BEV E+E probe high reflectivity	I-4		1
	I-5		2
	I-6		3
Calpower NS	NS02	Ø3 mm/L80 mm	1
	NS04		2
	NS08		3
MBW probe	1066	Ø3 mm / L40mm	1
	1064		2
	1065		3
PHYSICUS PT100/10	702/18	Ø5 mm/L117 mm	1
	703/18		2
Vaisala TMP1	P5150501	Ø6 mm/L130 mm	1
	P5150502		2
	P5150503		3
Wika CTP5000-170B	W3450254/CNZF-10-1	Ø6 mm/L350 mm	1
	W3450254/CNZF-10-2		2
	W3450254/CNZF-10-3		3
Wika TR60 Special	WK1	Ø7.76 mm/L44 mm Ø19.7 mm/L62 mm	1
	WK2		2
	WK3		3

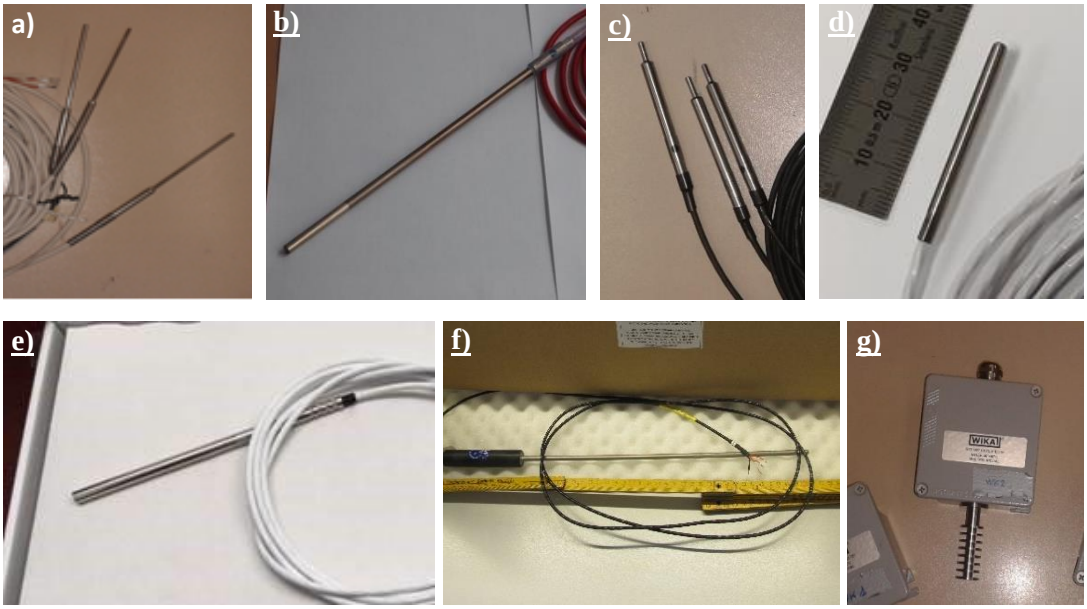


Figure 2 The circulating probes. a) Calpower, b) BEV E+E probes, c) Physicus, d) MBW, e) Vaisala, f) Wika CTP5000, g) Wika TR60. For the TR60 the box was removed prior to circulation.

Photos of all probe models used are shown in Figure 2.

1.2 Schedule and execution

The original schedule anticipated that measurements should be conducted between July 2019 and July 2020. Each participant was allowed 3 weeks of measurement and 1 week for shipping, and a substantial effort was invested in designing the order of measurements such that all participants received the probes at the most convenient time for each.

The loops were organised so as to deliberately collect all participants with anticipated customs clearance requirements in loop 2, with the exception of JV/Norway which acted as link laboratory and therefore had to measure in all loops. Customs clearance was performed with temporary imports. With one exception this worked well, but a minor reorganisation was necessary in loop 3 due to issues between the Czech Republic and Norway.

The Wika TR60 probe used in loop 2 had fragile lead contacts initially, and at RISE the final ice point check saw a reduction in resistance corresponding to around 25 mK. After consultation with the pilot and coordinator, RISE resoldered the leads and added new electrical insulation. Prior to forwarding the probe to VTT/MIKES, another batch of ice point data were recorded, which now were in line with the value obtained upon reception at RISE.

In March 2020 the Covid pandemic presented a major challenge. After the initial shock, transportation across borders proceeded without any major delays, however the impact on laboratory work varied substantially between the countries. In some cases NMIs saw an increase in calibration activity in the initial lockdown phase. In other countries access to the laboratories was restricted, and the participants therefore could not carry out all measurements as planned. In a few cases staff got infected with Covid 19 and for obvious reasons this caused further complications and delays. Invariably, calibrations and essential activities were prioritised, resulting in a reduced availability of time for ILC activity. The delays were exacerbated when the original tight schedule could not be adhered to, since some laboratories had to wait longer than planned to find a new available time for the ILC measurements. The final data were acquired in Q3 2021, and the final reports were received in Q4 2021.

1.3 Participant setups

Participants used a variety of setups. Most consisted of different climatic test chamber models, but in three cases the measuring volume was within an enclosure submerged in a liquid bath. 11 out of 25 reports used a subchamber inside a larger enclosure.

Participant	Current (mA)	Chamber	Thermostat	Chamber volume (L)	Subchamber volume (L)
BEV/E+E	0,5	Lauda RP 4090 CW	Bath	-	0.64
		SIMTECH ST 70240 (radiation shield used)	chamber	249.5	-
		SIMTECH ST 70240	chamber	249.5	1.90
GUM	1	CTS	chamber	351.0	37.70
INTA	1	Vötsch VT 7034	chamber	326.3	1.88
INTiBS	1	No device info	chamber	-	5.88
JV	1	Weiss SB22/160/40, sn222/19811	chamber	159.2	26.15
LNE-CETIAT	1	CTS.	chamber	3.2	-
NPL	1	Temperature Applied Sciences	chamber	421.9	1.52
NSAI	1	Vötsch VT 7034	chamber	281.3	-
CEM	1	Kambic KK-340CHULT	chamber	-	-
DMDM	1	Vötsch (-20C) Thunder Scientific (>=0 °C)	chamber	- -	- -
DPM	1	Angelantoni DY110	chamber	112.7	-
DTI	1	Heraeus Vötsch HC 7057 Thunder Scientific 2500ST (>0 °C)	chamber	563.6	- -
IMBIH	1	No device info.	chamber	197.9	17.52
MBM	1	Kambic KK-190 CHULT	chamber	177.9	-
RISE	1	Etanol/water baths	Bath	-	0.17
UME	1	WEISS TECHNIK WKL 100/40	chamber	96.8	-
VTT MIKES	1	Heraeus-Vötsch HC4020	chamber	204.0	-
BEV	1	Weiss Technik WK3 340/40	chamber	332.8	0.91(*)
BFKH	1	Vötsch VC4100	chamber	992.8	-
CMI	1	No device info.	Bath	-	-
INRIM	1	Vötsch VT7011	chamber	123.5	4.42
MIRS/UL-FE/LMK	1	ThunderScientific 2500	chamber	43.3	-
		Vötsch 7110	chamber	1000.0	27.00
NQIS / EIM	1	HERAEUS HC4033	chamber	326.6	-
PTB	1	CTS GmbH. C-40/350	chamber	351.0	-
SMD	1	Vötsch HC 4033	chamber	326.9	-
SMU	1	Vötsch VC 4018.	chamber	195.8	-

(*) BEV used a wind tunnel subchamber where two pipe sections at different diameters are joined by a gradually narrowing neck. The volume quoted is the volume of the narrow section where the probes are located during the calibration.

Two participants reported results independently from more than one setup. For BEV E+E the setups are labelled *setup 1* (using the Lauda RP 4090 CW), *setup 2* (using the SIMTECH ST 70240 with a radiation shield) and *setup 3* (using the SIMTECH with a subchamber). For MIRS/UL-FE/LMK the setups are labelled *setup 1* (using the Thunder Scientific 2500) and *setup 2* (using the Vötsch 7110). Two participants used different climatic test chambers above and below 0 °C. Apart from BEV E+E, all participants used the same probe current (1 mA).

2 Probe characteristics

The probes were subjected to a suite of characterisation measurements at the pilots. In addition, some participants carried out their own set of characterisations. This section summarises the findings.

The pilots recorded (R, T) data before and after the circulation, and also performed measurements in liquid baths. Self heating in liquid baths were carried out in all loops, and in addition the loop 2 pilot acquired data on self heating, humidity levels and hysteresis in air.

The pilot data acquired before and after circulation provided important drift and robustness information.

2.1 Self heating

Self heating was evaluated by the pilots, and some of the participants on a voluntary basis. Data is available from air measurements, liquid bath measurements and in a few cases, in TPW cells. Since the self heating effect is small, in the order of a few 10's of mK or less, the temperature of the isothermal enclosure needs to be maintained stable enough for the appropriate evaluation , or that the true reference temperature is recorded simultaneously. There is noticeable scatter in the results reported, suggesting that the bath or air temperature may not have been known or recorded correctly in some cases.

The data has been reported as (R, T) pairs at the probe currents of 1 mA, 1.41 mA and 2.82 mA. The latter value was only used by one of the pilots. The change in reported resistance between different probe currents was computed and converted to an equivalent temperature by

$$\Delta_R = \frac{R(T, 1.41) - R(T, 1)}{\partial R/\partial T\big|_T} \qquad 2\text{-}1$$

The equation computes the difference in resistance at two probe currents (1.41 mA and 1 mA) at a specific temperature T, and divides by the slope of the $R(T)$ curve at the same temperature T. When a participant repeated the measurement several times the average is computed.

The measured temperature *T* is used to compute the difference in temperature at the two probe currents: $\delta = T(1.41) - T(1)$. The difference is small, and does not imply any deliberate change in the temperature. However, the temperature in the bath or climate chamber will invariably fluctuate and it is necessary to take this into account when we compute the net selfheating.

The net self heating Δ_{SH}, in temperature units, is simply computed by

$$\Delta_{SH} = \Delta_R - \delta \qquad 2\text{-}2$$

The self heating is computed for every reported case, i.e. at all nominal temperatures, for all 23 probes in circulation, and at multiple participants. However, it is natural to expect that the self heating should be very similar for probes of the same manufacturer and model, and we will therefore group all data for a specific model together (i.e. data measured on 2 or 3 individual probes) when compiling statistics.

The results scatter substantially, with some cases of a negative self heating. However, by excluding obvious outliers (such as the negative points), and then using the median as the best estimate, we obtain the results summarised in Table 3.

Table 3 Median self heating observed for each probe model. The BEV E+E probes are practically the same, and the surface finish is not expected to substantially change the self heating properties.

Probe	In air (°C)	In liquid (°C)
Wika TR60 Special	0.023	0.017
MBW probe	0.019	0.006

Vaisala TMP1	0.018	0.011
Calpower NS	0.013	0.004
Wika CTP5000-170B	0.008	0.004
PHYSICUS PT100/10	0.007	0.002
BEV E+E probe high reflectivity	0.006	0.002
BEV E+E probe	0.006	0.002

As expected, the self heating is larger in air than in liquid baths. The Wika TR60 probe was harder to handle in liquid baths due to its shape. The Wika TR60 probe exhibits the highest self heating in both liquid and air. Its design includes a set of thin, radial protrusions on the cylindrical metal sheath whose intention presumably is to enhance the thermal contact between the sensing element and the medium. On the other hand, the probes are quite large, which would tend to weaken the thermal contact with the free air and the sensing element, and also exacerbate the dimensional dependence of the temperature difference between the thermometer wall and the surrounding air [3].

The difference between the probes is substantial, with almost a factor of 5 from the lowest to the highest self heating in air.

The two BEV E+E probes are as expected the same. Their design is identical except for the coating, which for one of them is highly reflective.

The self heating does not appear to depend strongly on temperature, see Figure 3 to Figure 9 The self heating as a function of temperature for the probes. The data is compiled from reports by JV, NSAI, CEM and INRIM. The error bars represent scatter in observations, not uncertainty.. There is perhaps a small increasing tendency with temperature, which may be explained by the fact that the dissipated electrical power in the resistor is proportional to resistance, which increases with temperature.

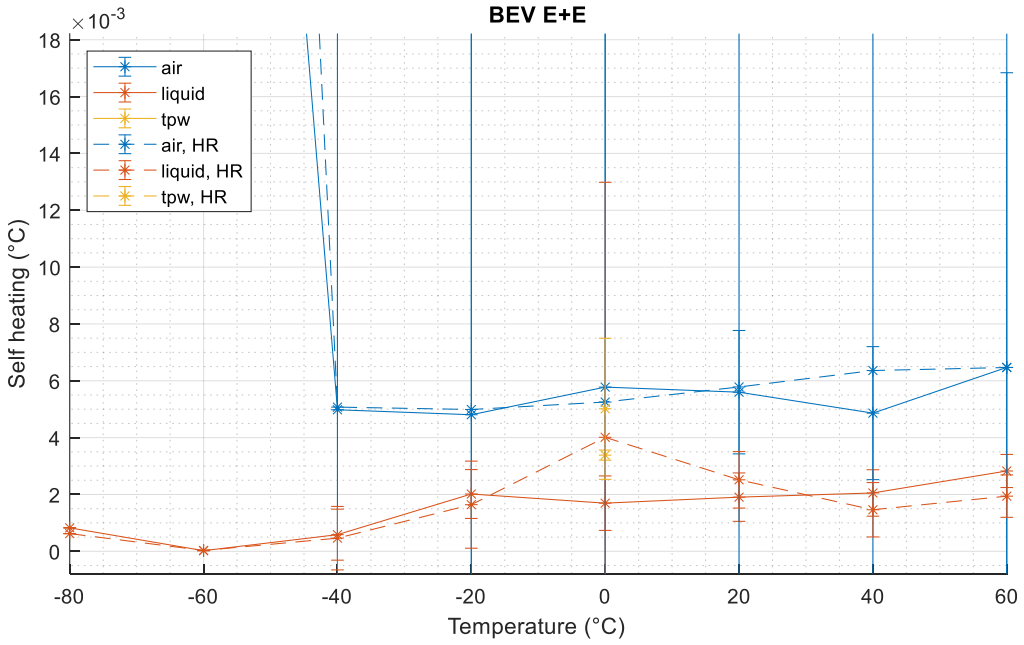


Figure 3

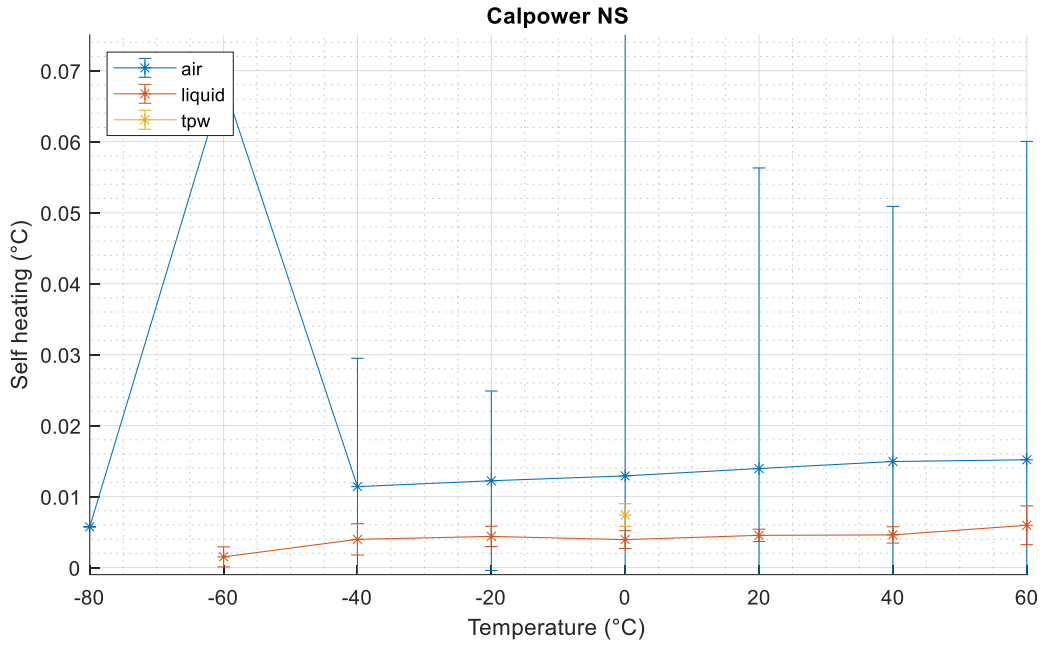


Figure 4

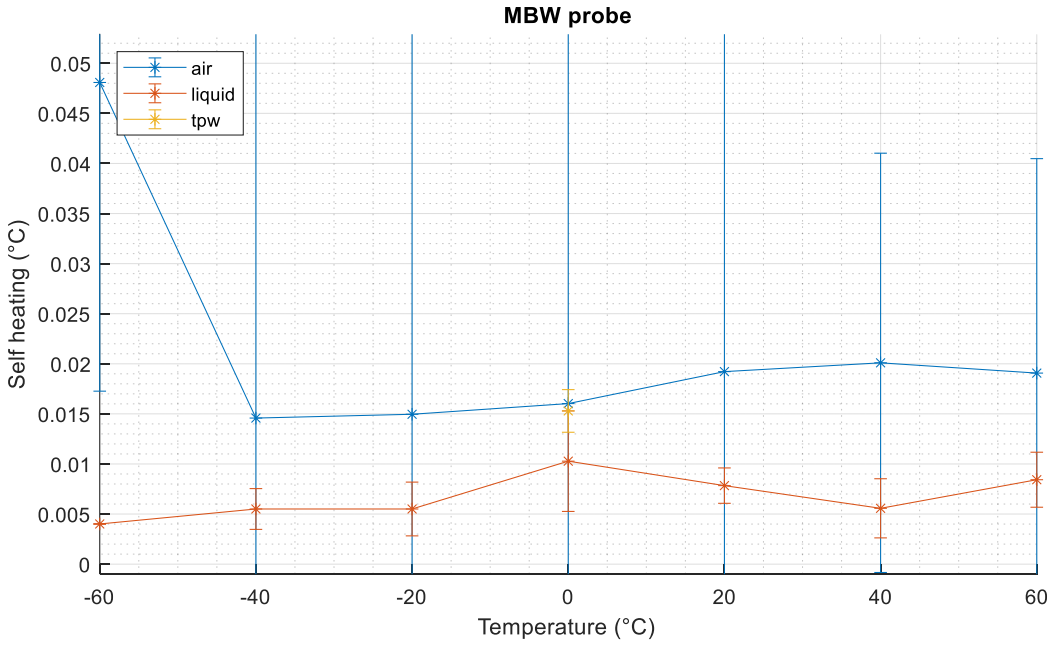


Figure 5

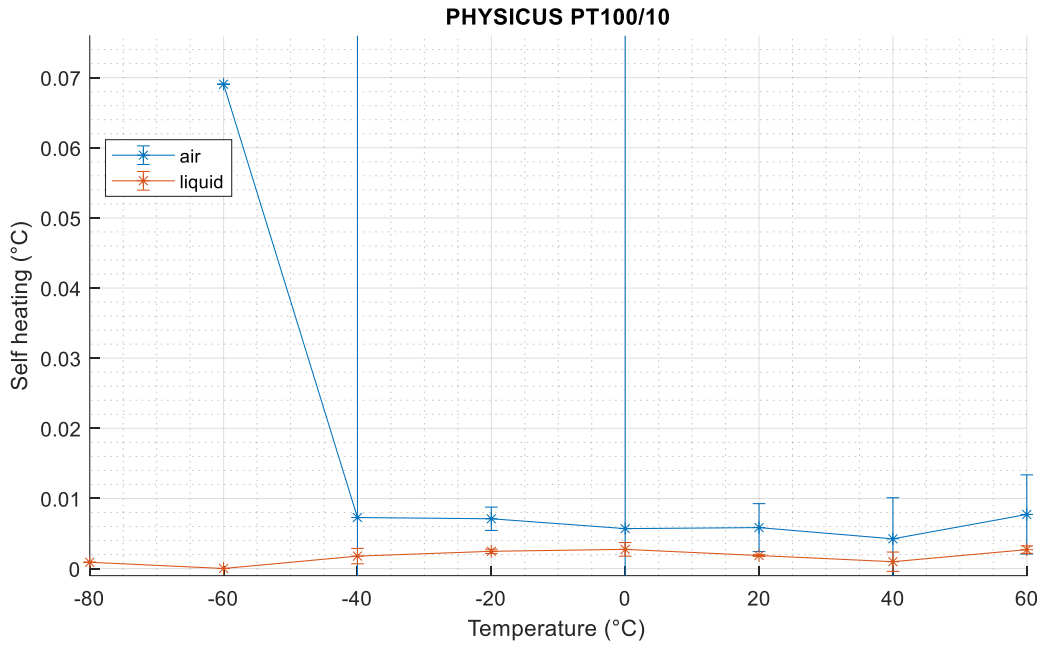


Figure 6

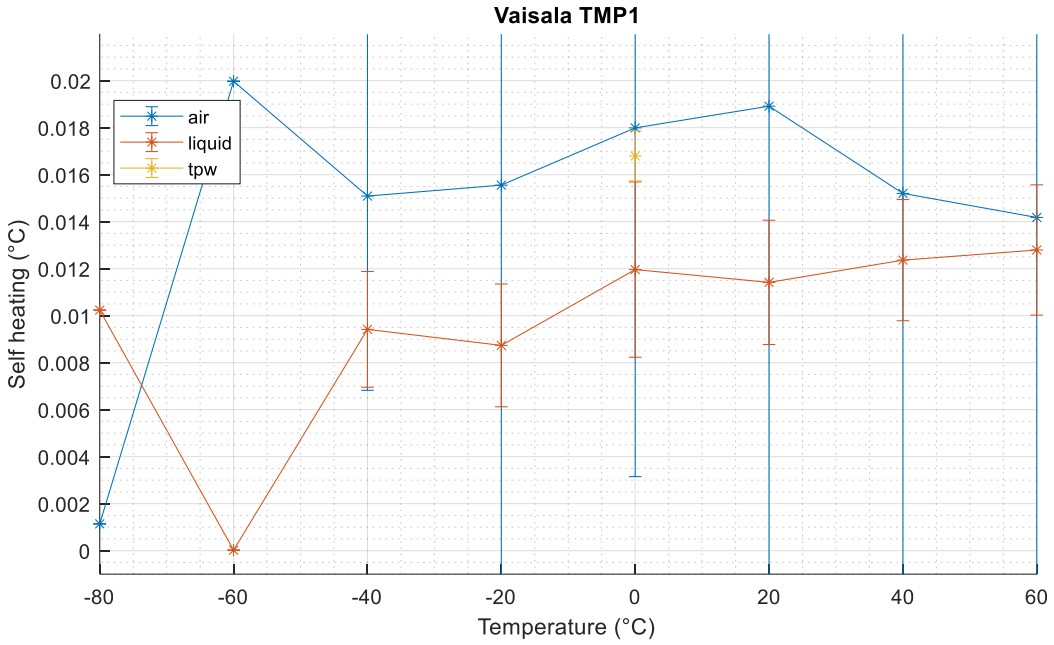


Figure 7

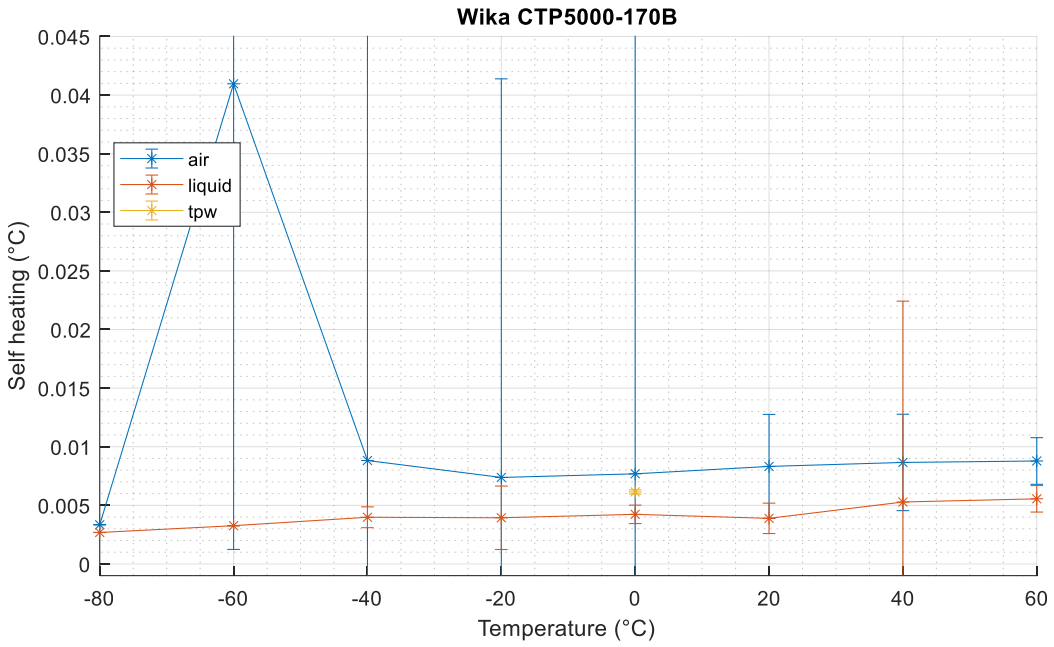


Figure 8

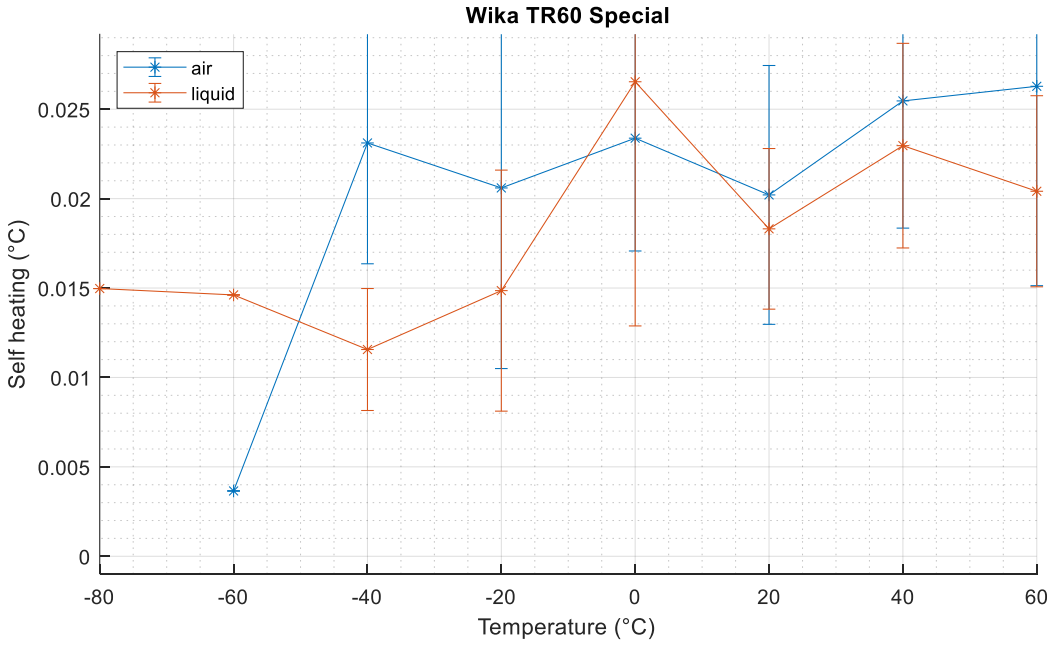


Figure 9 The self heating as a function of temperature for the probes. The data is compiled from reports by JV, NSAI, CEM and INRIM. The error bars represent scatter in observations, not uncertainty. The temperatures on the x-axis are nominal temperatures, and the TPW values are grouped with the ice point values.

The figures illustrate the large scatter in self heating evaluation results. For the air data this could be caused by difficulty in achieving stable conditions in the climate chambers, and reported changes in probe resistance may also be affected by an actual temperature change in the air.

2.2 Liquid bath data

Pilots measured the resistance at all measurement points in liquid baths. They repeated the measurements when the probes were returned after circulation in the loop. We expect that the liquid bath data represent the best achievable calibrations of the probes.

The drift in liquid baths were computed at each nominal temperature for each probe. First we compute the difference in resistance, ρ_λ, after and before the circulation: $\rho_\lambda = \rho_{\lambda 1} - \rho_{\lambda 0}$, and convert to temperature by dividing with the sensitivity of the ITS-90 reference function. Then we subtract the corresponding reference temperature difference $\tau_\lambda = \tau_{\lambda 1} - \tau_{\lambda 0}$ to obtain the net drift as

$$\delta_\lambda = (\rho_{\lambda 1} - \rho_{\lambda 0})\frac{\partial T}{\partial R} - (\tau_{\lambda 1} - \tau_{\lambda 0})$$

The pilots reported uncertainty as well as the bath data. The uncertainty in the drift value can be computed from

$$u_{\delta\lambda} = \sqrt{\left(u_{\rho\lambda}\frac{\partial T}{\partial R}\right)^2 + {u_{\tau\lambda}}^2}$$

where $u_{\rho\lambda}$ and $u_{\tau\lambda}$ are the largest of the initial/final uncertainties. This assumes that the uncertainties at the pilots are perfectly correlated before and after circulation. This is an optimistic assumption and $u_{\delta\lambda}$ is probably a slightly small value, but not substantially. Typically, the uncertainty in individual δ_λ-points is below 10 mK (at 95 % coverage).

Table 4 Liquid bath drift data, comparing probe characteristics before and after circulation. All values are expressed in units of °C. The highlighted cells show large drift (red, greater than 0,1 °C) and moderate drift (yellow, between 0,02 °C and 0,1 °C), just as a guide to the eye. Larger drift is observed in loop 1 compared with the other loops.

Model	Loop	-80 °C	-60 °C	-40 °C	-20 °C	0 °C	20 °C	40 °C	60 °C
BEV E+E	1	0.920	0.006	0.006	0.138	0.131	0.002	-0.086	-0.151
BEV E+E	2			0.002	0.000	0.001	0.001	0.006	0.007
BEV E+E	3			-0.005	-0.003	-0.004	-0.006	-0.010	-0.012
BEV E+E HR	1	1.500	-0.004	-0.005	0.312	0.090	-0.013	-0.125	-0.195
BEV E+E HR	2			0.000	0.001	0.023	0.001	0.002	-0.001
BEV E+E HR	3			-0.002	0.001	-0.002	-0.004	-0.006	-0.009
Calpower NS	1	0.229	0.094	0.070	0.031	0.005	0.006	0.008	0.021
Calpower NS	2			-0.003	-0.004	0.000	0.000	0.003	0.004
Calpower NS	3			0.000	-0.002	0.000	0.007	0.023	0.035
MBW probe	1	0.028	0.016	0.012	0.008	0.004	0.006	0.006	0.004
MBW probe	2			-0.003	-0.002	0.000	0.001	0.001	-0.003
MBW probe	3			-0.006	-0.007	0.020	-0.013	0.002	-0.003
PHYSICUS	1	0.615	0.349	0.223	0.160	0.111	0.114	0.094	0.082
PHYSICUS	2			0.015	0.027	0.002	0.014	0.019	0.015
Vaisala TMP1	1	0.245	0.134	0.157	0.165	0.065	0.042	0.038	0.037
Vaisala TMP1	2			-0.003	-0.007	-0.004	-0.003	-0.001	-0.002
Vaisala TMP1	3			-0.055	-0.055	-0.057	-0.060	-0.060	-0.058
Wika CTP5000	1	0.015	0.009	0.010	0.008	0.007	0.008	0.008	0.009
Wika CTP5000	2			0.007	0.007	0.007	0.008	0.010	0.012
Wika CTP5000	3			0.006	0.013	0.000	0.012	0.015	0.015
Wika TR60	1	0.047	0.036	-0.004	-0.015	-0.027	-0.045	-0.046	-0.048
Wika TR60	2			0.052	0.048	-0.018	-0.008	-0.011	-0.018
Wika TR60	3			-0.006	-0.019	0.006	-0.012	-0.010	-0.005

Table 4 summarises the results. All values are in units of °C. The cell formatting is to aid the eye. Red cells contain drift values larger than ±0,1 °C (in total 19 out of 154), yellow cells contain values larger than ±0,02 °C (in total 32 out of 154), and the rest (103) are values within the interval ±0,02 °C. The table permits a couple of observations.

An intriguing observation is that the probes in loop 1 seem to suffer larger drift than in the other loops. The large drift was observed quickly by the pilot, and it was decided to remeasure some of them at low temperature. The repeat measurements were performed several months after reception of the probes, during which they had been stored idly at ambient conditions. It is the remeasured data which is shown in Table 4.

Because loop 1 was the only loop in which the probes were subjected to the coldest temperatures a hypothesis was proposed that the lowest temperatures would affect some of the probe models, perhaps by some hysteretic humidity exchange with the surroundings. The probes in loops 2 and 3 were then tested by subjecting them to -80 °C for a prolonged period. They were placed inside a container, immersed in a liquid bath, and maintained at low temperature for several hours. The resistance was continuously logged, along with the bath temperature (for surveillance). The table below summarises the results.

Table 5 Drift results from loops 2 and 3 at -80 °C. The probes were the same make and model as those that were found to drift substantially in loop 1. The tests were performed by a long exposure to low temperature, while their output and the reference temperature were monitored continuously.

	CEM	**INRIM**
Vaisala	35 mK	10 mK
BEV	5 mK	10 mK
BEV high reflectivity	25 mK	25 mK
Physicus	580 mK	-
Duration	24 hours	55 hours
Bath drift		<0.1 mK

The Physicus probe do drift with a similar magnitude as that observed in loop 1, but both BEV probes are nearly unaffected by this long duration exposure to low temperature. The reason for the larger drift in loop 1 remains unknown.

A second observation is that an observed drift value at one temperature does not necessarily predict the drift at another temperature. On average the drift decreases with increasing temperature in the limited temperature range used here. However, for individual probes the pattern is less clear.

2.3 Ice point data

Ice point/TPW data were reported by all participants, based on measurements at reception and just before the probes were shipped to the next participant. A summary plot is shown in Figure 10, where both values (upon reception and just before dispatch) are plotted. The graph suggests that the probes do drift during the circulation, but not in a systematic manner.

There are a few examples of very large jumps at the same participant, of the order 0,1 °C. This may be attributed to some sort of error, in the transcription of results or in the measurements. Some of the probes were difficult to immerse properly in liquid and this may have affected the ice point results. However, it is difficult to test this hypothesis.

The Vaisala probe used in loop 3 appears to change ice point resistance by almost 0,3 °C on the first leg of the circulation. Since the probe was measured twice by the pilot both before and after the circulation one could conclude that the initial step change is due to a physical change in the probe. But as seen in Table 4 the drift deduced from the liquid bath measurements is virtually temperature independent, and around 60 mK. Since the remaining ice point checks seem to be stable and corroborate the liquid bath observation it is reasonable to conclude that the initial ice point data is an outlier.

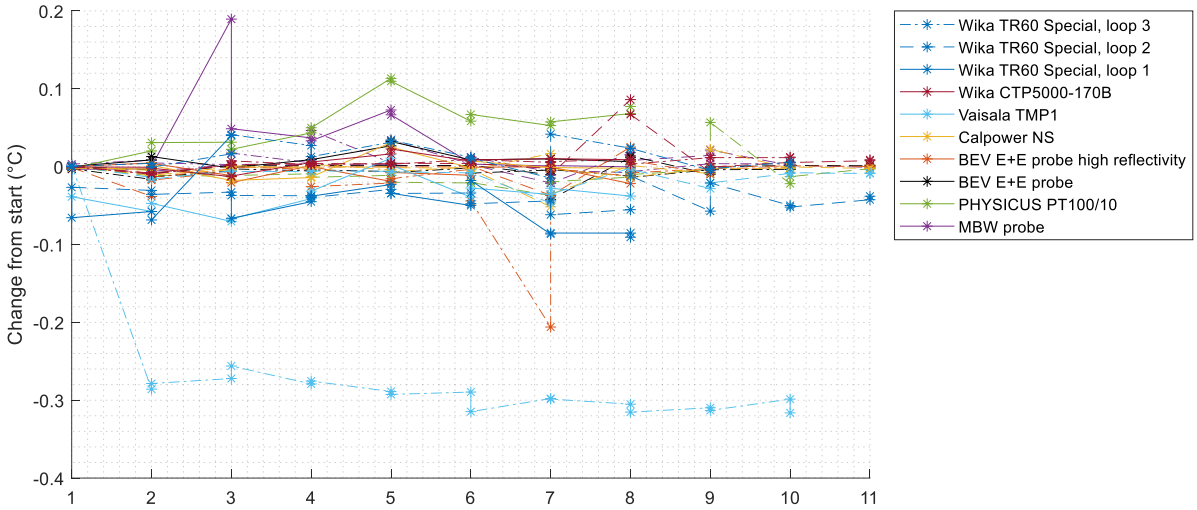


Figure 10 Change in the reported ice/tpw values. The colors distinguish the probes, while the line style distinguishes loops: solid lines are loop 1, dashed lines are loop 2 and dashed-dot lines are loop 3. The legend shows the pattern for the Wika TR60 probe. The initial value reported by pilots is used as the origin for all probes, and change in resistance has been converted to a change in temperature.

2.4 Probe drift – summary

Table 6 shows a summary of the drift data, represented by three different possible parameters: (i) liquid bath drift at -40 °C, which is the data from the appropriate column in Table 4; (ii) the span of the TPW or ice point values reported by the participants during circulation; and (iii) the uncertainty in the consensus value at -40 °C, if we compute it without linking the different loops (the computation is explained in Section 3.4).

The table rows have been sorted by liquid bath drift. A closer inspection of the columns for the ice point checks and the consensus value reveals that a large value for one of the drift metrics does not necessarily imply a large value at a different temperature for the same probe. In particular, the ice point values from the pilots are small for some of the probes that turn out to have a very strong drift as measured by the liquid bath data – and vice versa.

During planning of the ILC the intention was to use ice point/TPW data as an indication for drift during circulation. However, participants measured these values in a number of different ways, in some cases even in liquid bath thermostats set to 0 °C. There are cases of quite substantial changes from one participant to the next, and even from reception to dispatch at the same participant. The ice point data is probably strongly heterogeneous.

The liquid bath data represent a more thorough check for drift, but was only carried out at the pilots. There is no way to tell whether the drift has been gradual during the circulation.

The uncertainty of the consensus value (see Section 3.4) is not a direct measure of the probe drift, but can point to individual probes that seem to cause more difficulty for the participants. A large

uncertainty suggests that the data scatter is strong, which could be caused by drift, but also inherent issues with the probe stability.

Table 6 Summary table of three different drift metrics. The data was sorted according to the last column. All values are in °C. The three largest values in the two other columns have been highlighted. The ice point measurements do not appear to be a good predictor of the drift at -40 °C, as measured in liquid baths.

Model	Loop	Ice point span	Consensus uncertainty at -40 °C	Absolute value of drift, in bath, -40 °C
PHYSICUS	1	0.046	0.019	0.223
Vaisala TMP1	1	0.032	0.011	0.157
Calpower NS	1	0.019	0.010	0.070
Vaisala TMP1	3	0.127	0.092	0.055
Wika TR60	2	0.025	0.148	0.052
PHYSICUS	2	0.037	0.043	0.015
MBW probe	1	0.077	0.021	0.012
Wika CTP5000	1	0.014	0.012	0.010
Wika CTP5000	2	0.039	0.021	0.007
Wika CTP5000	3	0.009	0.105	0.006
MBW probe	3	0.022	0.020	0.006
Wika TR60	3	0.018	0.322	0.006
BEV E+E	1	0.014	0.029	0.006
BEV E+E HR	1	0.011	0.025	0.005
BEV E+E	3	0.007	0.037	0.005
Wika TR60	1	0.052	0.017	0.004
Vaisala TMP1	2	0.019	0.038	0.003
Calpower NS	2	0.030	0.065	0.003
MBW probe	2	0.017	0.059	0.003
BEV E+E HR	3	0.018	0.023	0.002
BEV E+E	2	0.022	0.059	0.002
Calpower NS	3	0.007	0.164	0.000
BEV E+E HR	2	0.025	0.092	0.000

3 Data processing

The data reported by the participants consists of pairs of resistance and temperature values (R, T). The temperatures reported were the actual air temperature as prepared by the participant, and should be measured in the best way possible by each of the participants. The resistance values were the DUT readings. Associated uncertainties (U_R, U_T) were requested for both values independently.

The uncertainties requested from the participants did not include probe characteristics. In particular, it is customary to assess the self-heating of the probes, and some laboratories also characterise hysteresis by cycling humidity and/or temperature.

In this section we present the data analysis, including the pre-processing, linkage and consensus value computation. To avoid clutter in notation, we do not explicitly add subscript symbols to distinguish between data points (probes, loops and temperatures). Instead we adopt the following conventions:

- Roman letter R and T (possibly with subscripts) indicate reported values prior to preprocessing.
- Greek letters ρ and τ indicate preprocessed values, and thus represents comparable invariants. Essentially ρ is a corrected resistance to correspond with a common temperature τ.

- We bundle data from the same probe model using the observations at JV.
- In equations we do not explicitly distinguish between the probe models. But all calculations are repeated on each model independently.

The data comprises more than 1600 unique combinations of (participant, probe, temperature)-triplets. For participants measuring all 8 probes at temperatures from -40 °C to 60 °C we have available 48 unique datapoints. In practice the actual number of reported points ranged from 40 (due to a limited temperature range) to 138 (at JV, which measured 23 probes).

3.1 Data cleaning

In a few cases it was necessary to rectify misunderstandings or obvious typing mistakes. When an error was suspected in the main data point reported by a participant, they were asked to revise their reporting.

Some participants only reported uncertainty for one of the quantities, resistance or temperature. In those cases we assumed that the reported uncertainty was a combined uncertainty, which we trusted the participant to compute according to standard procedure, so the missing values were taken to be 0 in all calculations.

Some participants added rows for repeated realisations in the main reporting tables. In those cases the average value was used as that participant's contribution, while the uncertainty was the largest reported uncertainty at that point.

3.2 Preprocessing

Participants realised slightly different temperatures near the nominal values. The observations were aligned to a common reference temperature by extrapolating the reported resistance values from the corresponding temperature to the nominal temperature. For small temperature deviation a linearisation is sufficient, hence a resistance correction is computed from

$$\Delta_R = \rho - R_{DUT} = (\tau - T_{ref}) \left.\frac{\partial R}{\partial T}\right|_{T=\tau} \qquad 3\text{-}1$$

The nominal temperature is designated τ and the corresponding resistance ρ. Each probe has its individual $R(T)$ curve (obtained from the pilot laboratories' measurements in liquid baths), from which we can compute the sensitivity coefficient $\partial R/\partial T$. The linearisation requires that $\tau - T_{ref}$ is reasonably small, but for the current investigation the linearisation error is less than 1 mK and negligible in all cases, with a possible exception at -80 °C.

Equation 3-1 is applied for each probe, at each participant, and at each nominal temperature. The ρ from the same probe at the same nominal temperature are the invariants used to compare the results.

The uncertainty u_ρ follows from Equation 3-1, noting that u_τ= 0:

$$u_\rho = \sqrt{{u_{DUT}}^2 + \left(u_{ref} \cdot \left.\frac{\partial R}{\partial T}\right|_{T=\tau}\right)^2} \qquad 3\text{-}2$$

The lowest temperature at -80 °C proved to be challenging for most of the participants. In practice, several data points were acquired at -70 °C instead. In those cases we have computed Δ_R from Equation 3-1 using τ = -70 °C, but adding the data to the pool at -80 °C. If the corrections at participants is temperature dependent in a systematic way this procedure will lead to an increased

and unknown error at those points because the Δ_R at -70 °C differs from that at -80 °C. Unfortunately, the results do suggest systematic trends for many participants, but there is no simple way to resolve this problem.

3.3 Loop links

The JV data were used to link the loops. The linkage is computed as a model specific correction. For a given probe model, JV recorded data for three units (except the Physicus probe, which was not circulated in loop 3). For each model, a model average is computed from the three datasets available at JV, which is used to compute a correction for each specific probe to its model average:

$$L_i = \langle \rho_{JV} \rangle - \rho_{i,JV}, \qquad i \in \{1,2,3\} \tag{3-3}$$

where $\langle \rho_{JV} \rangle$ is the average observed resistance, at JV, for all probes of the same model: $\langle \rho_{JV} \rangle = \sum \rho_{i,JV}/3$. The linkage uncertainty follows from the standard GUM equation

$$u_L^2 = {u_{JV}}^2 + {u_{i,JV}}^2 \tag{3-4}$$

The linkage parameter is a correction added to all results. After this correction we assume that data for the same probe type is comparable, and they are pooled for consensus building.

3.4 Consensus values

The results from section 2 suggest that the probes drift noticeably during circulation, and that there is no unequivocal way to compensate for that drift. There may also be other differences between the laboratories, such as irradiation levels, wind speed and placement in the air chamber, and this may further add to unknown differences in the outcomes. We therefore choose to model the results with the random effects model, which assumes that there is an unknown additional random error at each participant, but with a common mean (0) and variance. The observed resistance r_i at each participant can then be written as a sum

$$r_i = \rho + u_i + \varepsilon_i \tag{3-5}$$

Here ρ is the true resistance, u_i the uncertainty at each laboratory, which is estimated as the reported uncertainty, and ε_i is an additional, unknown random error that is different at each laboratory, but drawn from the same distribution with variance σ^2. This error is not an uncertainty contribution but a correction that was unknown to the laboratories when they recorded the data. The error is attributed to changes in the characteristics of the traveling probes in this report, perhaps caused by changes inflicted during transportation (for instance changing humidity conditions or mechanical shocks) although the source of the error is strictly speaking unknown. Since the ice point data do not show any systematic trend for most of the probes, this error term is best modelled as a random variable whose variance is estimated from the data. The best estimate for ρ is the weighted mean:

$$\bar{\rho} = \frac{1}{\sum w_i} \sum r_i w_i$$
$$w_i = \frac{1}{\left(u_i^2 + \sigma^2\right)} \tag{3-6}$$

The unknown variance can be estimated in different ways. Here we use the DerSimonian-Laird procedure (explained in [4] [5] [6]). The unknown variance is estimated from the observations in a 2-

step procedure. First, a zero-σ estimate $\hat{\rho}$ is computed, along with the appropriate zero-σ weights $\hat{w}_i$. These parameters are used to compute the Cochran Q-statistic from observations, which is a measure of the sum of relative errors:

$$Q = \sum_{i=1}^{n} \hat{w}_i (r_i - \hat{\rho})^2 \qquad 3\text{-}7$$

The sum is taken over all n observations for one particular temperature and probe. The estimate of σ^2 is

$$\sigma^2 = \max\left\{0, \frac{Q - n + 1}{\sum \hat{w}_i + \sum \hat{w}_i^2 / \sum \hat{w}_i}\right\} \qquad 3\text{-}8$$

The last equation is then fed back into Equation 3-6 to compute the final $\bar{\rho}$ and w_i. The standard uncertainty of $\bar{\rho}$ is

$$\bar{u} = \frac{1}{\sqrt{\sum w_i}} \qquad 3\text{-}9$$

The estimate of σ^2 from Equation 3-8 is prone to underestimation of the true value [4] [5]. Koekpe *et al* proposed a modified procedure where the scatter of σ^2 is also estimated via a Monte Carlo estimation. The method leverages an analytical estimate of the scatter from Biggerstaff and Tweedie [6], which is used to construct a probability density for σ^2, which is typically larger than the value from Equation 3-8. We have adopted this method here, and the uncertainty in the consensus values are expanded somewhat as a result of this.

The calculations are carried out separately at each nominal temperature and for each probe model, to provide unique consensus values for at each temperature and probe. This is necessary to ensure that we only compare identical thermometer states – the thermometer resistance changes with temperature, and differs between probes at any given temperature. However, this does not in itself imply any fundamental difference from a statistical viewpoint. There is no a priori reason to believe that one probe is more difficult than the others, or that one temperature presents any particular challenge compared with the others; hence, deviations from the consensus should be comparable across temperature and probes. However, we will return to this in section 4.

3.5 Degree of equivalence

The final results are the deviations from the consensus value along with its uncertainty. This is computed as

$$d = \bar{\rho} - \rho \qquad 3\text{-}10$$

and the uncertainty

$$u_d = \sqrt{\bar{u}^2 + {u_\rho}^2} \qquad 3\text{-}11$$

The normalised error can be computed from

$$E_n = \frac{d}{2u_d} \qquad 3\text{-}12$$

3.6 Correlations

Correlations between datapoints are ignored. In almost all cases the largest reported uncertainty contributions are laboratory specific, such as chamber uniformity and stability. The traceability of reference equipment is also from internal references at the institutes in most cases, and does not in any case dominate the uncertainty. An overview of the uncertainty contributions that were reported is found in Section 5.1.

4 Main results

This section presents the unilateral degree of equivalence for each participant, using consensus values computed for each probe and temperature. Some data points have been excluded from the consensus computation: they are shown in the next subsection.

4.1 Outliers and suspicious datasets

There are three categories of irregularities in the data. The first is connected with the drift observed in liquid baths at the pilots, which revealed substantial drift in loop 1 for some of the probes. The second is a participant which seems to have a temperature dependent deviation from consensus. The third is a participant which reports very large uncertainty.

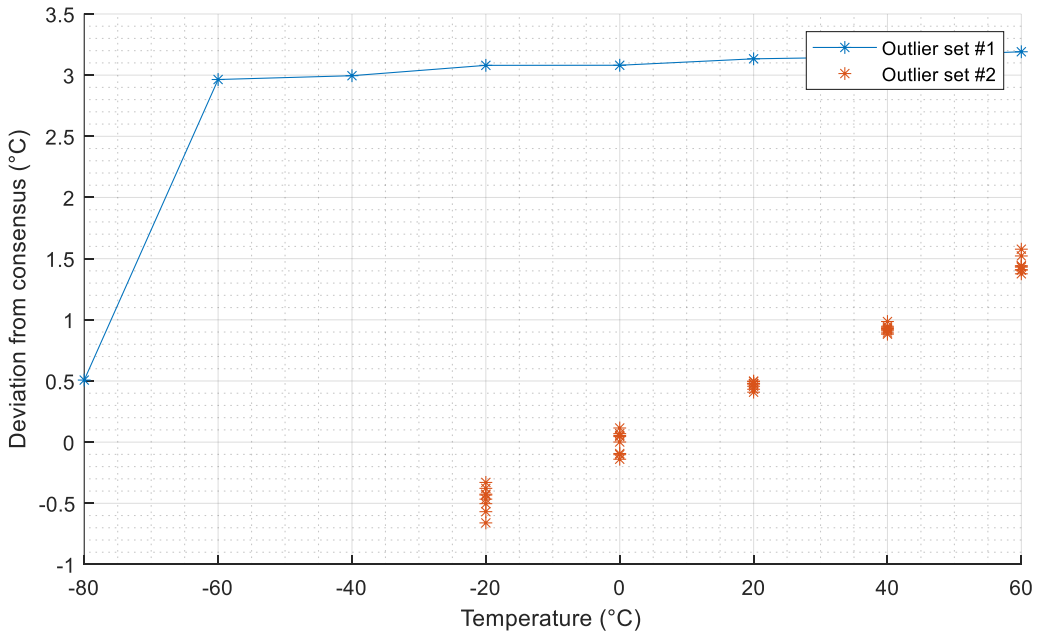


Figure 11 Outlier data sets. The excessive deviation is so large that it affects statistical descriptors such as standard deviation and mean. It was recorded for one of the high drift probes. The other dataset is mostly acceptable if we consider each temperature in isolation, but the trend arises suspicion. Such a linear and consistent behaviour (for all circulating probes) suggests that the laboratory has a systematic error in either the reference equipment or in the way the probes are interrogated.

4.1.1 Excessive offset for one probe

Figure 11 shows the distance from consensus for the B-1 probe at one of the participants (in blue). The distance is excessive, but without an obvious explanation. However, the B-1 probe is one of the probes that exhibited large drift in the liquid bath data.

4.1.2 Linear trend

Figure 11 shows the deviation from consensus for a particular participant (in red). The points exhibit an acceptable scatter, but a distinct linear trend in temperature.

4.1.3 Excessive reported uncertainty

Finally, one participant reported standard uncertainties (k=1) in the range 0.12 °C to 0.4 °C. Among the reported uncertainties in this ILC the typical values are much smaller, on the order of 0.02 °C. The participant data was used as normal in computation of the consensus value, but because the consensus is computed as a weighted mean the contribution from this participant was small compared to the other participants.

4.2 Aggregate results

4.2.1 Consensus values

The consensus uncertainties are summarised in Table 7, and graphically in Figure 12. The table also shows the contribution from the unknown random effects term (what the uncertainty would have been if the reported uncertainties were 0). Two participants reported measurements for more than one setup, but only one of their datasets were used in the consensus computation.

Table 7 Consensus standard uncertainties at all temperatures and for all probes. The values are converted to temperature units (°C) from the SPRT reference function. The numbers in parenthesis are the random effects contribution, expressed as a standard uncertainty. In many cases this is a substantial contribution.

°C	BEV E+E	BEV E+E HR	Calpower	MBW	Physicus	Vaisala	CTP5000-170B	TR60
-80	0.198 (0.155)	0.214 (0.115)	0.129 (0.066)	0.158 (0.081)	0.869 (0.452)	0.175 (0.090)	0.102 (0.052)	0.135 (0.099)
-60	0.098 (0.049)	0.084 (0.039)	0.035 (0.016)	0.075 (0.034)	0.040 (0.018)	0.032 (0.014)	0.037 (0.017)	0.055 (0.025)
-40	0.027 (0.015)	0.032 (0.018)	0.058 (0.031)	0.022 (0.011)	0.021 (0.012)	0.035 (0.019)	0.039 (0.023)	0.130 (0.074)
-20	0.020 (0.011)	0.028 (0.016)	0.024 (0.013)	0.018 (0.010)	0.014 (0.009)	0.016 (0.009)	0.008 (0.004)	0.060 (0.035)
0	0.006 (0.002)	0.005 (0.002)	0.013 (0.006)	0.011 (0.005)	0.021 (0.012)	0.011 (0.005)	0.005 (0.002)	0.018 (0.009)
20	0.004 (0.001)	0.007 (0.003)	0.007 (0.003)	0.004 (0.001)	0.007 (0.003)	0.006 (0.002)	0.004 (0.001)	0.003 (0.000)
40	0.008 (0.004)	0.011 (0.006)	0.013 (0.007)	0.005 (0.000)	0.011 (0.007)	0.008 (0.004)	0.004 (0.000)	0.004 (0.000)
60	0.010 (0.005)	0.013 (0.007)	0.018 (0.010)	0.008 (0.004)	0.026 (0.016)	0.011 (0.005)	0.005 (0.002)	0.009 (0.004)

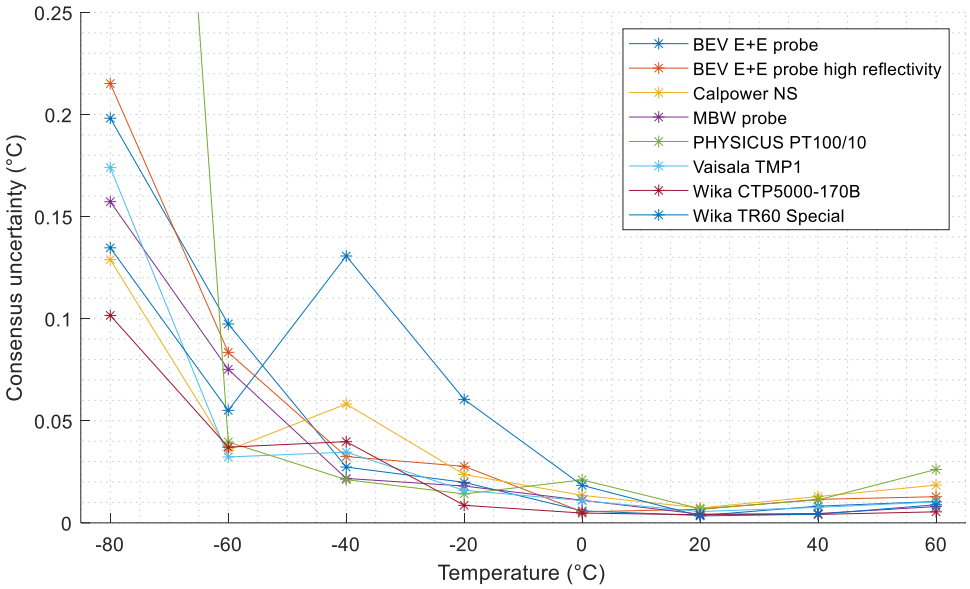


Figure 12 Uncertainty in the consensus values versus temperature. The points are the taken from Table 7. There is a consistent temperature dependence of the uncertainty, which to a large extent is determined by the uncertainty attributed to the unknown, random effects. Below -40 °C we have fewer datapoints, which may partly explain the larger scatter.

4.2.2 Temperature dependent deviation

The uncertainty in the consensus values are plotted in Figure 12. The uncertainty partly represents the reported uncertainty from the participants, and partly represents the between-lab variability. The uncertainty tends to grow with decreasing temperature from a minimum around 0 °C to 20 °C. This pattern may also be observed from the scatter of the consensus deviations from all participants. It seems 20 °C somehow makes it easier to obtain consistent conditions across the various laboratories. At present there is no explanation for this observation. A feasible, but at present speculative, reason could be that it is easier to obtain a uniform background temperature in the chambers and hence the effect of irradiation is reduced.

4.2.3 Aggregates at each participant

Since the participants measured (R, T) points at 5-8 nominal temperatures for 7-8 probes, there are many observations for each participant: up to 138 (JV which measured all 23 probes) and as low as 40 (in a few cases participants could not realise all the temperatures). This enables a second independent assessment of laboratory performance by extracting statistics from all the data points available.

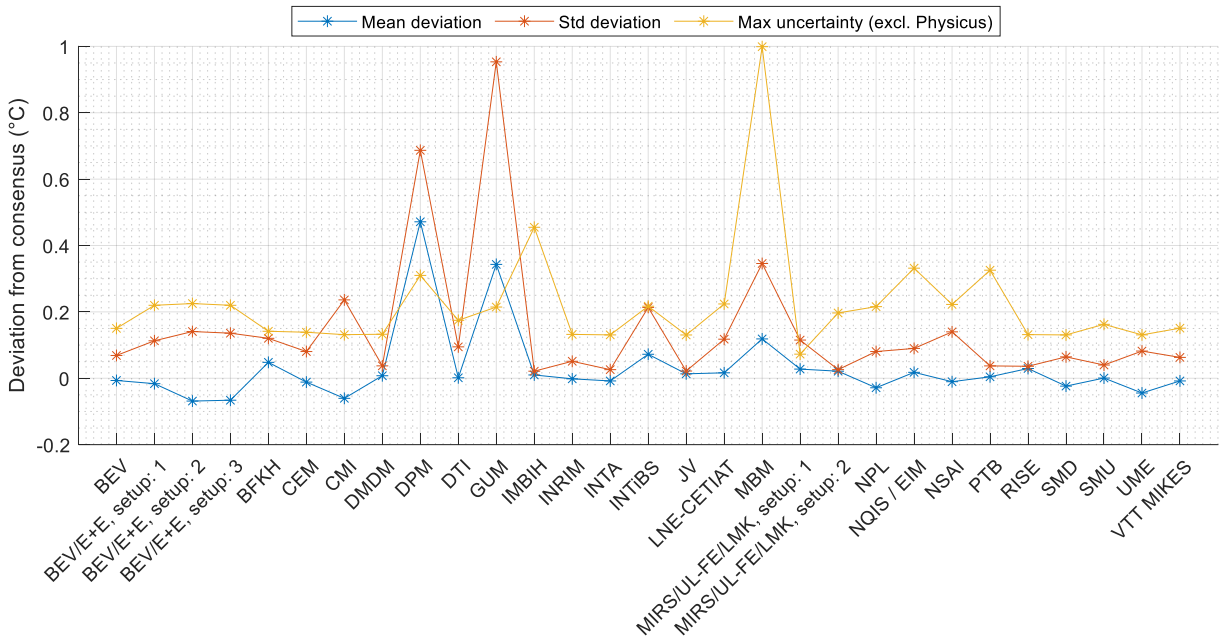


Figure 13: Plot of three different metrics for the deviation at each participant. The blue line shows the average distance from the consensus value for all the points reported by the participant. The red and yellow lines are two different ways to view the scatter. The red line shows the standard deviation of the observed distances from consensus: there is no strict uncertainty involved, but it does represent the independent scatter in the measurements. A low value suggests the participant is able to consistently reproduce its performance across the temperature range and for the different probes.

Figure 13 shows three features computed for each participant. Two of them are directly computed from the distance from consensus: its average and its standard deviation. The average value (blue line) should be close to zero if there is no systematic bias at the participant. A small standard deviation (red line) suggests a high degree of the repeatability at the participant. Both these metrics are direct observables from the data and are related to the scatter of the results at each participant. Finally, the yellow line shows the maximum uncertainty among all the unilateral degree of equivalence-uncertainties for each participant. It is composed of both the laboratory reported uncertainty and the uncertainty of the consensus value and can be regarded as a worst case uncertainty for each laboratory. If the laboratory is in agreement with the consensus value (E_n<1) for all data points the red line should be below the yellow line. This should be true even if the laboratory fails for a small number of cases.

In a few cases the observed standard deviation of points (red line) is larger than the maximum uncertainty. When this occurs the reported uncertainty from the participant is probably much smaller than the deviations from the consensus. A closer inspection of the results (see Section 4.3) reveal that this is the case. For GUM the observation is explained by a few extreme outliers. For DPM we have observed a systematic, temperature dependent which leads to large offsets while the reported uncertainty is quite small. For CMI the reported uncertainty is probably too small, perhaps with some contributions not taken into account.

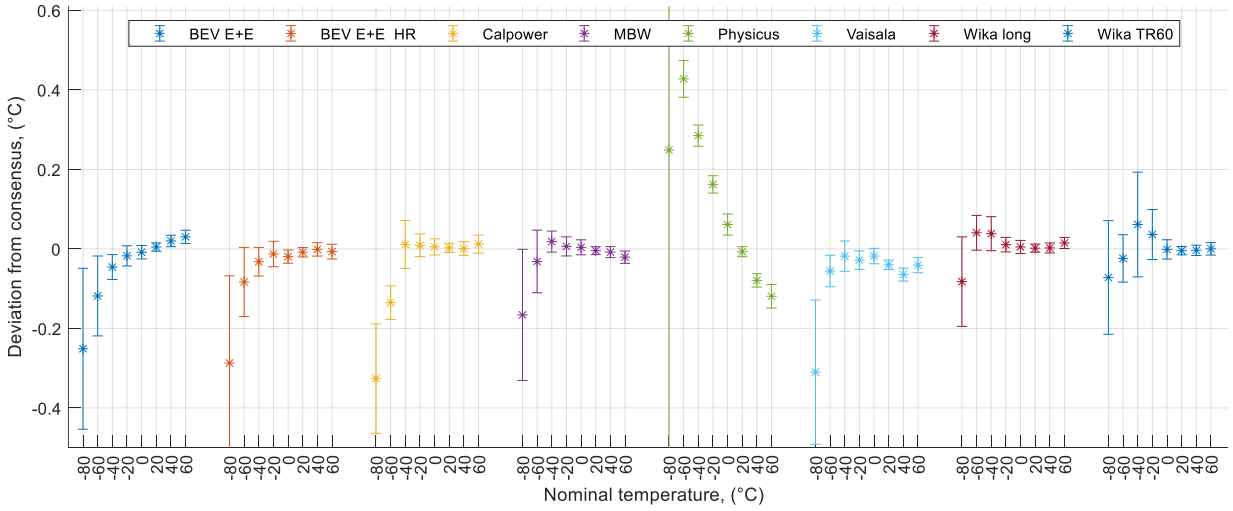
4.3 Individual degree of equivalence

This section contains the main results for individual participants. The tables show the results as a deviation from consensus in °C, and a standard uncertainty (k=1). The data is also presented graphically, in some cases zoomed in to highlight small details.

4.3.1 BEV E+E, setup 1

Setup 1 used an enclosure immersed in a liquid bath.

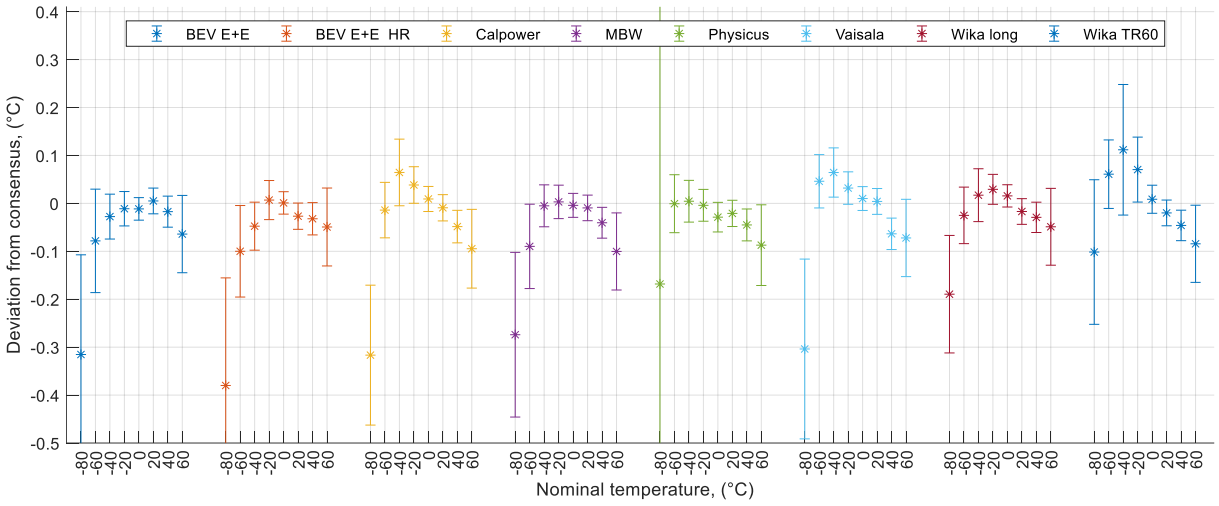
T (°C)	BEV E+E	BEV E+E HR	CTP5000-170B	Calpower	MBW	Physicus	TR60	Vaisala
-80	-0.252 ± 0.204	-0.286 ± 0.220	-0.082 ± 0.113	-0.327 ± 0.138	-0.166 ± 0.165	0.244 ± 0.870	-0.073 ± 0.143	-0.311 ± 0.181
-60	-0.118 ± 0.101	-0.083 ± 0.087	0.040 ± 0.044	-0.136 ± 0.042	-0.032 ± 0.079	0.428 ± 0.046	-0.024 ± 0.060	-0.055 ± 0.040
-40	-0.047 ± 0.031	-0.029 ± 0.036	0.037 ± 0.043	0.011 ± 0.060	0.017 ± 0.026	0.285 ± 0.027	0.056 ± 0.131	-0.022 ± 0.038
-20	-0.020 ± 0.025	-0.009 ± 0.032	0.011 ± 0.018	0.011 ± 0.028	0.005 ± 0.024	0.162 ± 0.022	0.037 ± 0.063	-0.030 ± 0.023
0	-0.008 ± 0.017	-0.020 ± 0.017	0.005 ± 0.016	0.005 ± 0.020	0.004 ± 0.019	0.061 ± 0.027	-0.001 ± 0.024	-0.017 ± 0.019
20	0.005 ± 0.010	-0.009 ± 0.012	0.002 ± 0.010	0.002 ± 0.012	-0.004 ± 0.010	-0.007 ± 0.013	-0.004 ± 0.010	-0.040 ± 0.012
40	0.020 ± 0.014	-0.003 ± 0.017	0.002 ± 0.013	-0.000 ± 0.018	-0.008 ± 0.014	-0.079 ± 0.017	-0.003 ± 0.013	-0.064 ± 0.016
60	0.031 ± 0.017	-0.008 ± 0.018	0.015 ± 0.014	0.011 ± 0.023	-0.020 ± 0.016	-0.119 ± 0.030	0.000 ± 0.016	-0.039 ± 0.019



4.3.2 BEV E+E setup 2

Setup 2 used a climate chamber with a radiation shield protecting the sensors.

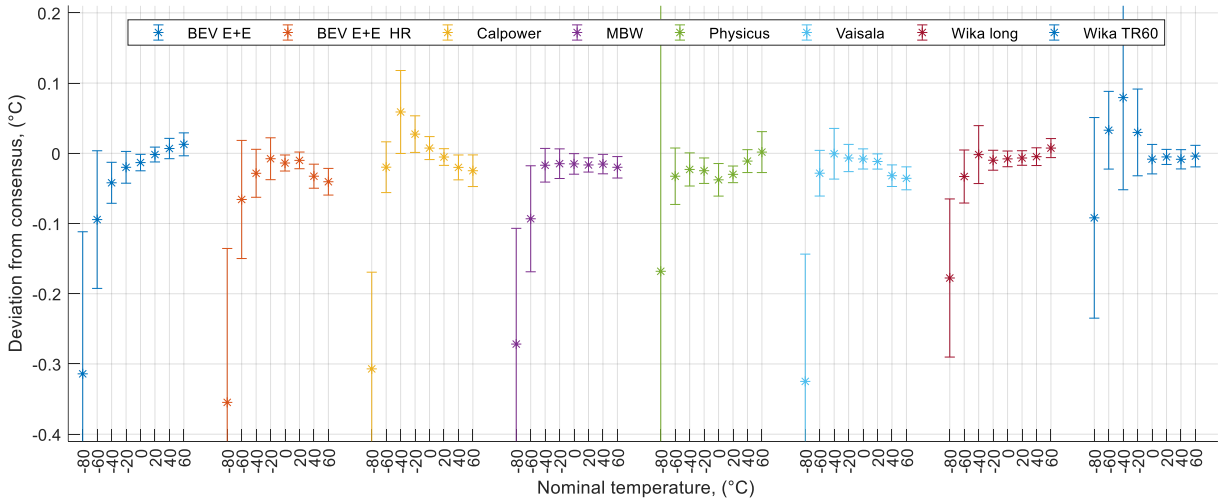
T (°C)	BEV E+E	BEV E+E HR	CTP5000-170B	Calpower	MBW	Physicus	TR60	Vaisala
-80	-0.316 ± 0.210	-0.379 ± 0.225	-0.189 ± 0.123	-0.317 ± 0.146	-0.274 ± 0.172	-0.173 ± 0.871	-0.102 ± 0.151	-0.305 ± 0.188
-60	-0.078 ± 0.108	-0.100 ± 0.095	-0.025 ± 0.059	-0.014 ± 0.058	-0.090 ± 0.088	-0.001 ± 0.061	0.061 ± 0.071	0.046 ± 0.056
-40	-0.029 ± 0.047	-0.044 ± 0.050	0.016 ± 0.055	0.065 ± 0.069	-0.006 ± 0.044	0.004 ± 0.044	0.107 ± 0.135	0.061 ± 0.051
-20	-0.014 ± 0.036	0.011 ± 0.041	0.029 ± 0.031	0.040 ± 0.038	0.002 ± 0.035	-0.004 ± 0.033	0.071 ± 0.068	0.030 ± 0.034
0	-0.012 ± 0.023	0.001 ± 0.023	0.016 ± 0.023	0.010 ± 0.026	-0.004 ± 0.025	-0.029 ± 0.031	0.009 ± 0.029	0.011 ± 0.025
20	0.005 ± 0.027	-0.027 ± 0.027	-0.017 ± 0.027	-0.009 ± 0.027	-0.009 ± 0.027	-0.021 ± 0.027	-0.020 ± 0.027	0.004 ± 0.027
40	-0.017 ± 0.032	-0.033 ± 0.034	-0.029 ± 0.032	-0.049 ± 0.034	-0.040 ± 0.032	-0.045 ± 0.033	-0.046 ± 0.032	-0.063 ± 0.033
60	-0.064 ± 0.081	-0.051 ± 0.081	-0.049 ± 0.080	-0.096 ± 0.082	-0.100 ± 0.080	-0.087 ± 0.084	-0.084 ± 0.080	-0.071 ± 0.081



4.3.3 BEV E+E setup 3

Setup 3 used a subchamber inside a climate chamber.

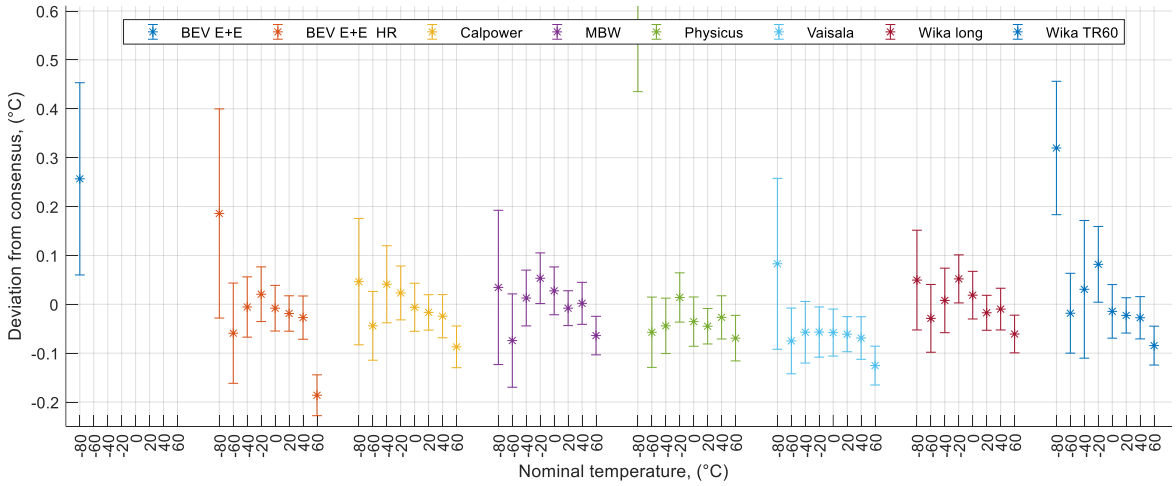
T (°C)	BEV E+E	BEV E+E HR	CTP5000-170B	Calpower	MBW	Physicus	TR60	Vaisala
-80	-0.315 ± 0.204	-0.354 ± 0.220	-0.177 ± 0.113	-0.308 ± 0.138	-0.272 ± 0.165	-0.173 ± 0.870	-0.093 ± 0.143	-0.326 ± 0.181
-60	-0.094 ± 0.099	-0.066 ± 0.084	-0.033 ± 0.038	-0.020 ± 0.036	-0.093 ± 0.075	-0.033 ± 0.040	0.033 ± 0.055	-0.028 ± 0.033
-40	-0.043 ± 0.029	-0.025 ± 0.034	-0.003 ± 0.041	0.059 ± 0.059	-0.018 ± 0.024	-0.023 ± 0.024	0.074 ± 0.130	-0.004 ± 0.036
-20	-0.023 ± 0.023	-0.004 ± 0.030	-0.010 ± 0.014	0.029 ± 0.026	-0.016 ± 0.021	-0.025 ± 0.018	0.030 ± 0.061	-0.009 ± 0.019
0	-0.013 ± 0.012	-0.014 ± 0.011	-0.008 ± 0.011	0.008 ± 0.016	-0.015 ± 0.014	-0.038 ± 0.023	-0.008 ± 0.021	-0.007 ± 0.014
20	-0.002 ± 0.011	-0.010 ± 0.012	-0.007 ± 0.010	-0.006 ± 0.012	-0.017 ± 0.010	-0.030 ± 0.012	-0.005 ± 0.011	-0.011 ± 0.011
40	0.007 ± 0.014	-0.034 ± 0.017	-0.005 ± 0.013	-0.021 ± 0.018	-0.015 ± 0.014	-0.011 ± 0.016	-0.008 ± 0.014	-0.031 ± 0.016
60	0.013 ± 0.016	-0.042 ± 0.019	0.007 ± 0.014	-0.026 ± 0.022	-0.019 ± 0.015	0.002 ± 0.029	-0.004 ± 0.015	-0.034 ± 0.017



4.3.4 GUM

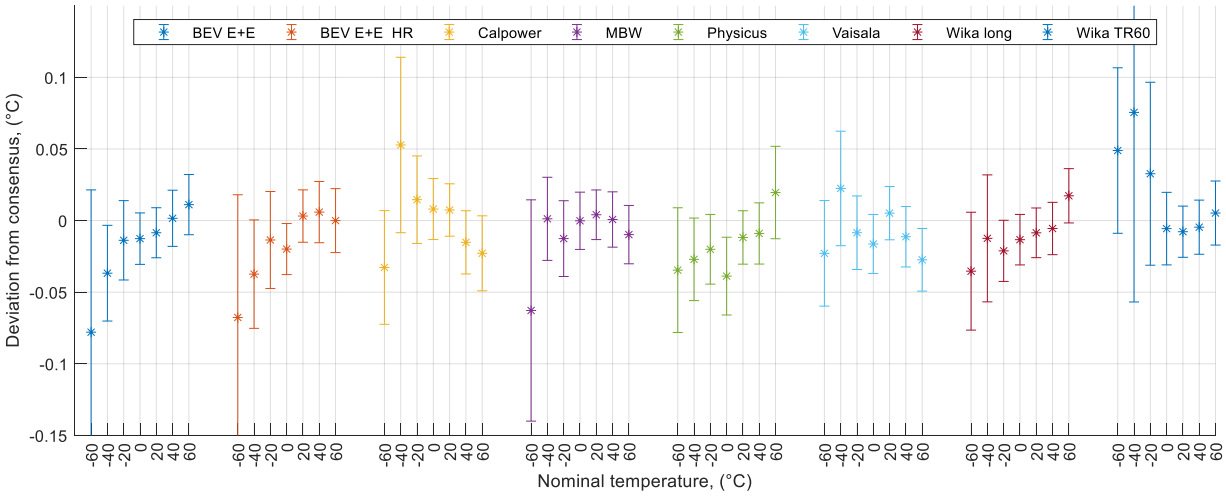
The plot has been zoomed to show details for the majority of points, which excludes 7 datapoints for the BEV E+E probe. See the table.

T (°C)	BEV E+E	BEV E+E HR	CTP5000-170B	Calpower	MBW	Physicus	TR60	Vaisala
-80	0.256 ± 0.199	0.187 ± 0.215	0.050 ± 0.103	0.046 ± 0.130	0.034 ± 0.158	1.301 ± 0.869	0.319 ± 0.137	0.082 ± 0.175
-60	2.870 ± 0.115	-0.059 ± 0.102	-0.029 ± 0.069	-0.044 ± 0.070	-0.074 ± 0.095	-0.057 ± 0.072	-0.018 ± 0.082	-0.075 ± 0.067
-40	2.978 ± 0.059	-0.002 ± 0.062	0.007 ± 0.066	0.041 ± 0.079	0.011 ± 0.057	-0.044 ± 0.057	0.026 ± 0.140	-0.061 ± 0.063
-20	3.065 ± 0.053	0.024 ± 0.056	0.052 ± 0.049	0.025 ± 0.055	0.052 ± 0.052	0.014 ± 0.050	0.082 ± 0.077	-0.059 ± 0.051
0	3.082 ± 0.048	-0.008 ± 0.047	0.019 ± 0.049	-0.006 ± 0.049	0.027 ± 0.049	-0.035 ± 0.050	-0.014 ± 0.055	-0.057 ± 0.048
20	3.136 ± 0.036	-0.019 ± 0.036	-0.017 ± 0.036	-0.017 ± 0.036	-0.008 ± 0.036	-0.045 ± 0.036	-0.023 ± 0.036	-0.061 ± 0.036
40	3.192 ± 0.044	-0.029 ± 0.044	-0.010 ± 0.043	-0.025 ± 0.044	0.002 ± 0.043	-0.027 ± 0.044	-0.027 ± 0.043	-0.068 ± 0.044
60	3.198 ± 0.040	-0.187 ± 0.042	-0.061 ± 0.039	-0.088 ± 0.043	-0.063 ± 0.039	-0.069 ± 0.047	-0.084 ± 0.040	-0.124 ± 0.040



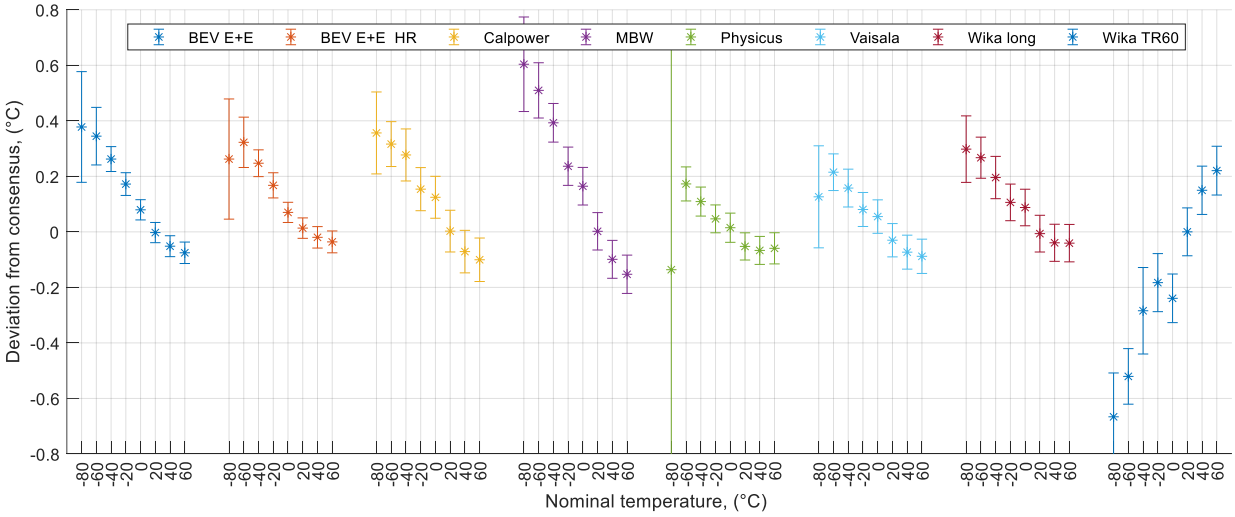
4.3.5 INTA

T (°C)	BEV E+E	BEV E+E HR	CTP5000-170B	Calpower	MBW	Physicus	TR60	Vaisala
-60	-0.078 ± 0.100	-0.068 ± 0.085	-0.036 ± 0.041	-0.033 ± 0.040	-0.063 ± 0.077	-0.035 ± 0.044	0.049 ± 0.058	-0.023 ± 0.037
-40	-0.038 ± 0.033	-0.034 ± 0.038	-0.013 ± 0.044	0.053 ± 0.061	-0.000 ± 0.029	-0.027 ± 0.029	0.071 ± 0.132	0.019 ± 0.040
-20	-0.016 ± 0.028	-0.010 ± 0.034	-0.021 ± 0.021	0.016 ± 0.030	-0.014 ± 0.026	-0.020 ± 0.024	0.033 ± 0.064	-0.010 ± 0.026
0	-0.013 ± 0.018	-0.020 ± 0.018	-0.013 ± 0.018	0.008 ± 0.021	-0.000 ± 0.020	-0.039 ± 0.027	-0.005 ± 0.025	-0.016 ± 0.021
20	-0.008 ± 0.017	0.003 ± 0.018	-0.009 ± 0.017	0.007 ± 0.018	0.004 ± 0.017	-0.012 ± 0.019	-0.008 ± 0.018	0.005 ± 0.019
40	0.002 ± 0.020	0.005 ± 0.021	-0.006 ± 0.018	-0.016 ± 0.022	0.001 ± 0.019	-0.009 ± 0.021	-0.004 ± 0.019	-0.010 ± 0.021
60	0.012 ± 0.021	-0.001 ± 0.022	0.017 ± 0.019	-0.024 ± 0.026	-0.009 ± 0.020	0.020 ± 0.032	0.005 ± 0.022	-0.026 ± 0.022



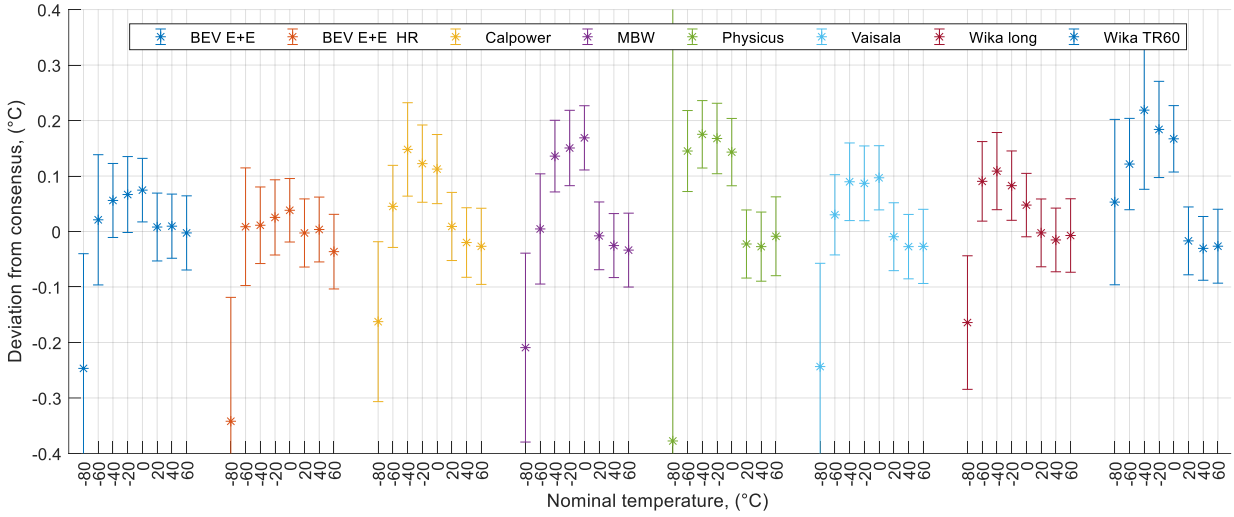
4.3.6 INTiBS

T (°C)	BEV E+E	BEV E+E HR	CTP5000-170B	Calpower	MBW	Physicus	TR60	Vaisala
-80	0.377 ± 0.201	0.263 ± 0.217	0.298 ± 0.120	0.356 ± 0.148	0.604 ± 0.170	-0.141 ± 0.870	-0.667 ± 0.158	0.125 ± 0.184
-60	0.345 ± 0.104	0.322 ± 0.090	0.267 ± 0.074	0.316 ± 0.081	0.510 ± 0.099	0.173 ± 0.061	-0.521 ± 0.100	0.215 ± 0.066
-40	0.261 ± 0.045	0.251 ± 0.048	0.195 ± 0.076	0.277 ± 0.094	0.391 ± 0.069	0.109 ± 0.052	-0.289 ± 0.155	0.154 ± 0.068
-20	0.170 ± 0.041	0.171 ± 0.045	0.106 ± 0.066	0.155 ± 0.078	0.235 ± 0.069	0.047 ± 0.050	-0.182 ± 0.104	0.079 ± 0.061
0	0.079 ± 0.036	0.070 ± 0.036	0.088 ± 0.066	0.125 ± 0.075	0.164 ± 0.068	0.015 ± 0.053	-0.239 ± 0.088	0.056 ± 0.060
20	-0.002 ± 0.036	0.013 ± 0.037	-0.006 ± 0.066	0.002 ± 0.075	0.002 ± 0.067	-0.052 ± 0.049	0.000 ± 0.086	-0.030 ± 0.060
40	-0.051 ± 0.038	-0.021 ± 0.039	-0.039 ± 0.067	-0.072 ± 0.077	-0.099 ± 0.068	-0.067 ± 0.050	0.150 ± 0.087	-0.072 ± 0.061
60	-0.075 ± 0.039	-0.037 ± 0.039	-0.041 ± 0.067	-0.102 ± 0.078	-0.152 ± 0.069	-0.059 ± 0.056	0.221 ± 0.088	-0.087 ± 0.062



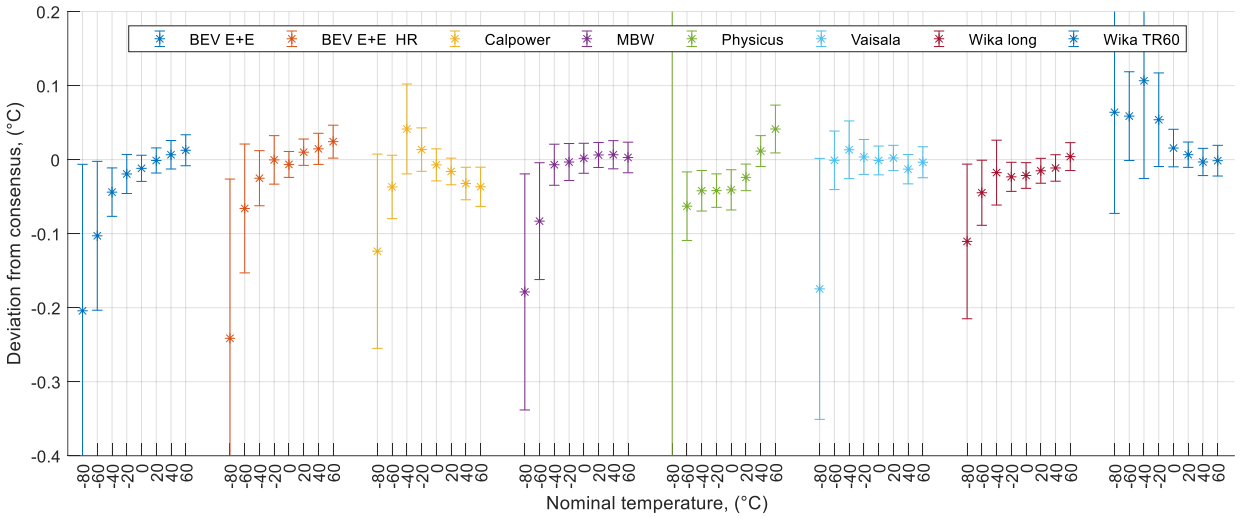
4.3.7 LNE-CETIAT

T (°C)	BEV E+E	BEV E+E HR	CTP5000-170B	Calpower	MBW	Physicus	TR60	Vaisala
-80	-0.248 ± 0.209	-0.341 ± 0.224	-0.164 ± 0.121	-0.163 ± 0.145	-0.209 ± 0.170	-0.382 ± 0.871	0.052 ± 0.149	-0.244 ± 0.186
-60	0.021 ± 0.118	0.008 ± 0.106	0.090 ± 0.072	0.045 ± 0.074	0.005 ± 0.099	0.145 ± 0.073	0.121 ± 0.082	0.030 ± 0.072
-40	0.055 ± 0.067	0.015 ± 0.069	0.108 ± 0.069	0.148 ± 0.084	0.135 ± 0.064	0.175 ± 0.061	0.214 ± 0.142	0.086 ± 0.070
-20	0.064 ± 0.068	0.029 ± 0.068	0.083 ± 0.062	0.124 ± 0.070	0.149 ± 0.068	0.168 ± 0.063	0.184 ± 0.087	0.085 ± 0.067
0	0.075 ± 0.057	0.038 ± 0.057	0.048 ± 0.057	0.113 ± 0.062	0.169 ± 0.058	0.143 ± 0.061	0.168 ± 0.060	0.097 ± 0.058
20	0.008 ± 0.061	-0.003 ± 0.061	-0.002 ± 0.061	0.009 ± 0.061	-0.008 ± 0.061	-0.022 ± 0.061	-0.017 ± 0.061	-0.009 ± 0.061
40	0.010 ± 0.058	0.002 ± 0.058	-0.015 ± 0.057	-0.021 ± 0.063	-0.025 ± 0.058	-0.027 ± 0.062	-0.030 ± 0.057	-0.026 ± 0.058
60	-0.002 ± 0.067	-0.038 ± 0.067	-0.007 ± 0.066	-0.028 ± 0.069	-0.033 ± 0.067	-0.009 ± 0.071	-0.026 ± 0.067	-0.025 ± 0.067



4.3.8 NPL

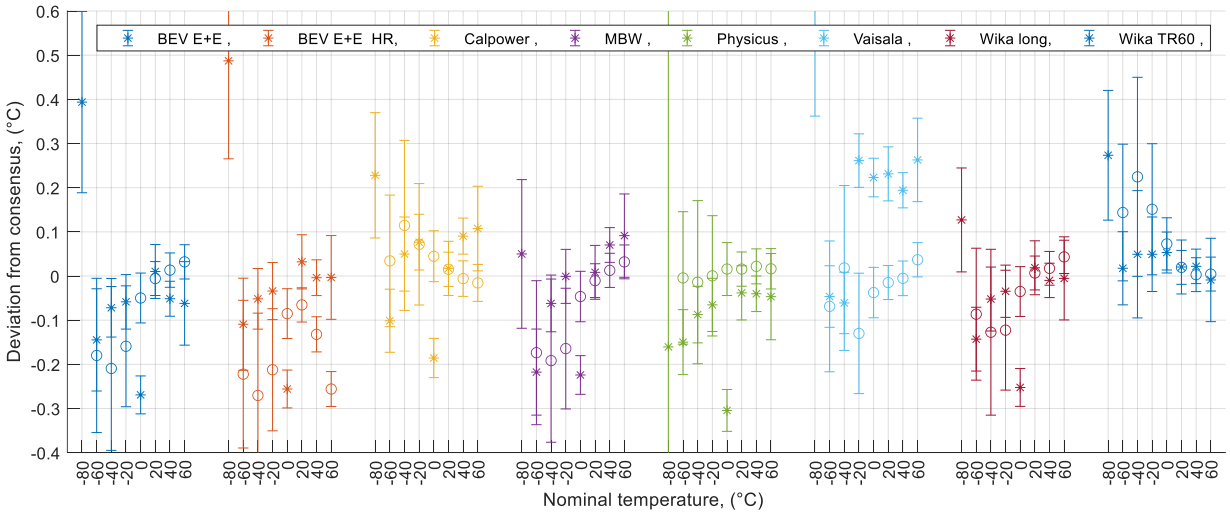
T (°C)	BEV E+E	BEV E+E HR	CTP5000-170B	Calpower	MBW	Physicus	TR60	Vaisala
-80	-0.205 ± 0.200	-0.241 ± 0.216	-0.110 ± 0.105	-0.124 ± 0.132	-0.179 ± 0.160	-0.451 ± 0.869	0.063 ± 0.137	-0.175 ± 0.176
-60	-0.103 ± 0.101	-0.066 ± 0.087	-0.045 ± 0.044	-0.037 ± 0.043	-0.083 ± 0.079	-0.063 ± 0.047	0.059 ± 0.060	-0.001 ± 0.040
-40	-0.045 ± 0.033	-0.022 ± 0.037	-0.018 ± 0.044	0.041 ± 0.060	-0.008 ± 0.028	-0.042 ± 0.027	0.102 ± 0.131	0.010 ± 0.039
-20	-0.022 ± 0.026	0.003 ± 0.033	-0.023 ± 0.020	0.015 ± 0.029	-0.004 ± 0.025	-0.042 ± 0.023	0.054 ± 0.063	0.002 ± 0.024
0	-0.012 ± 0.018	-0.007 ± 0.017	-0.021 ± 0.017	-0.007 ± 0.022	0.002 ± 0.020	-0.041 ± 0.027	0.016 ± 0.025	-0.000 ± 0.019
20	-0.001 ± 0.017	0.010 ± 0.018	-0.015 ± 0.017	-0.016 ± 0.018	0.006 ± 0.017	-0.024 ± 0.018	0.007 ± 0.017	0.003 ± 0.017
40	0.007 ± 0.019	0.013 ± 0.021	-0.011 ± 0.018	-0.033 ± 0.022	0.007 ± 0.019	0.012 ± 0.021	-0.003 ± 0.018	-0.012 ± 0.020
60	0.013 ± 0.021	0.023 ± 0.022	0.004 ± 0.019	-0.038 ± 0.026	0.004 ± 0.021	0.041 ± 0.032	-0.001 ± 0.021	-0.002 ± 0.021



4.3.9 NSAI

NSAI piloted loop 1. The data in the plot indicate the difference before and after the circulation. The numbers in the table are the average values, with the worst case as uncertainty.

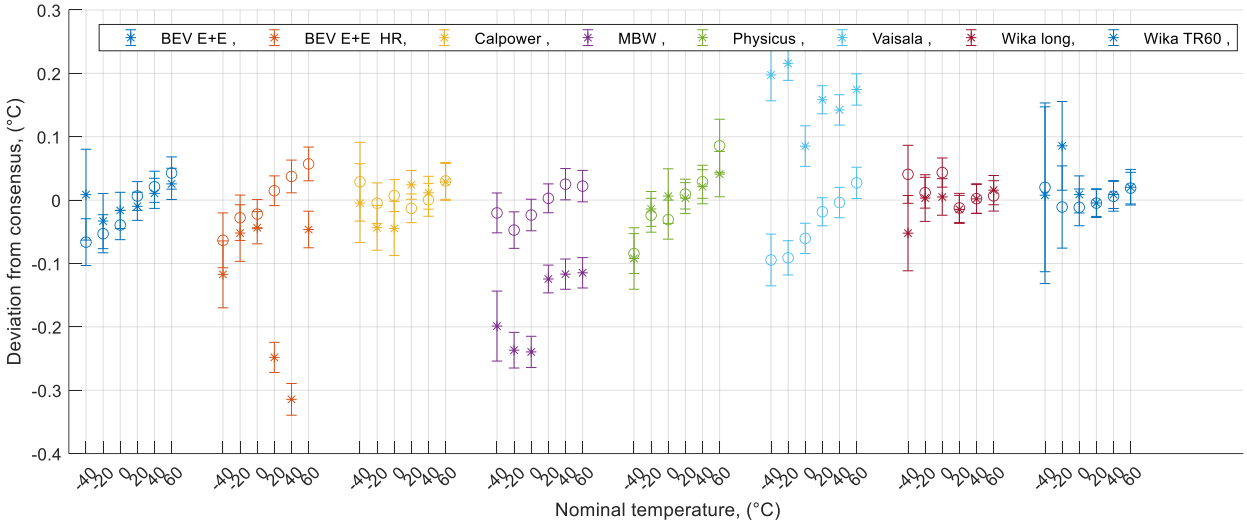
T (°C)	BEV E+E	BEV E+E HR	CTP5000-170B	Calpower	MBW	Physicus	TR60	Vaisala
-80	0.393 ± 0.207	0.488 ± 0.222	0.127 ± 0.118	0.227 ± 0.142	0.050 ± 0.168	-0.165 ± 0.871	0.273 ± 0.147	0.546 ± 0.184
-60	-0.162 ± 0.175	-0.166 ± 0.167	-0.115 ± 0.149	-0.034 ± 0.149	-0.196 ± 0.163	-0.077 ± 0.150	0.081 ± 0.155	-0.057 ± 0.148
-40	-0.142 ± 0.185	-0.157 ± 0.186	-0.091 ± 0.188	0.082 ± 0.192	-0.128 ± 0.185	-0.051 ± 0.185	0.132 ± 0.225	-0.025 ± 0.187
-20	-0.111 ± 0.137	-0.119 ± 0.138	-0.079 ± 0.136	0.076 ± 0.137	-0.084 ± 0.137	-0.033 ± 0.136	0.101 ± 0.148	0.064 ± 0.136
0	-0.159 ± 0.056	-0.171 ± 0.056	-0.144 ± 0.056	-0.070 ± 0.058	-0.135 ± 0.057	-0.144 ± 0.060	0.064 ± 0.059	0.093 ± 0.057
20	0.002 ± 0.061	-0.017 ± 0.061	0.013 ± 0.061	0.016 ± 0.061	-0.001 ± 0.061	-0.011 ± 0.061	0.020 ± 0.061	0.109 ± 0.061
40	-0.019 ± 0.040	-0.069 ± 0.040	0.004 ± 0.039	0.042 ± 0.041	0.042 ± 0.039	-0.009 ± 0.040	0.013 ± 0.039	0.095 ± 0.040
60	-0.015 ± 0.095	-0.131 ± 0.095	0.019 ± 0.094	0.045 ± 0.096	0.062 ± 0.094	-0.015 ± 0.098	-0.002 ± 0.094	0.151 ± 0.095



4.3.10 CEM

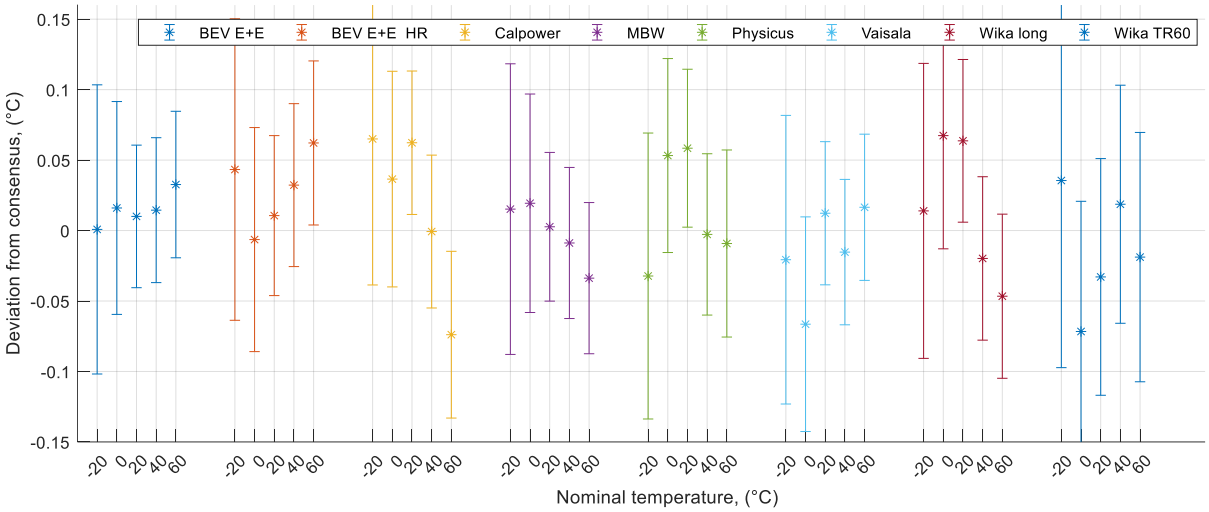
CEM piloted loop 2. The data in the plot indicate the difference before and after the circulation. The numbers in the table are the average values, with the worst case as uncertainty.

T (°C)	BEV E+E	BEV E+E HR	CTP5000-170B	Calpower	MBW	Physicus	TR60	Vaisala
-40	-0.030 ± 0.072	-0.087 ± 0.053	-0.007 ± 0.059	0.012 ± 0.062	-0.111 ± 0.055	-0.088 ± 0.049	0.009 ± 0.139	0.048 ± 0.041
-20	-0.046 ± 0.044	-0.036 ± 0.045	0.007 ± 0.037	-0.023 ± 0.035	-0.143 ± 0.029	-0.019 ± 0.027	0.038 ± 0.070	0.061 ± 0.027
0	-0.028 ± 0.029	-0.033 ± 0.025	0.024 ± 0.029	-0.018 ± 0.043	-0.132 ± 0.025	-0.012 ± 0.043	-0.000 ± 0.029	0.013 ± 0.032
20	-0.002 ± 0.023	-0.117 ± 0.024	-0.013 ± 0.023	0.005 ± 0.023	-0.061 ± 0.023	0.006 ± 0.024	-0.005 ± 0.023	0.070 ± 0.022
40	0.016 ± 0.025	-0.140 ± 0.026	0.002 ± 0.023	0.005 ± 0.026	-0.046 ± 0.025	0.025 ± 0.027	0.008 ± 0.023	0.070 ± 0.024
60	0.035 ± 0.025	0.004 ± 0.029	0.011 ± 0.024	0.028 ± 0.029	-0.046 ± 0.025	0.063 ± 0.042	0.020 ± 0.028	0.102 ± 0.025



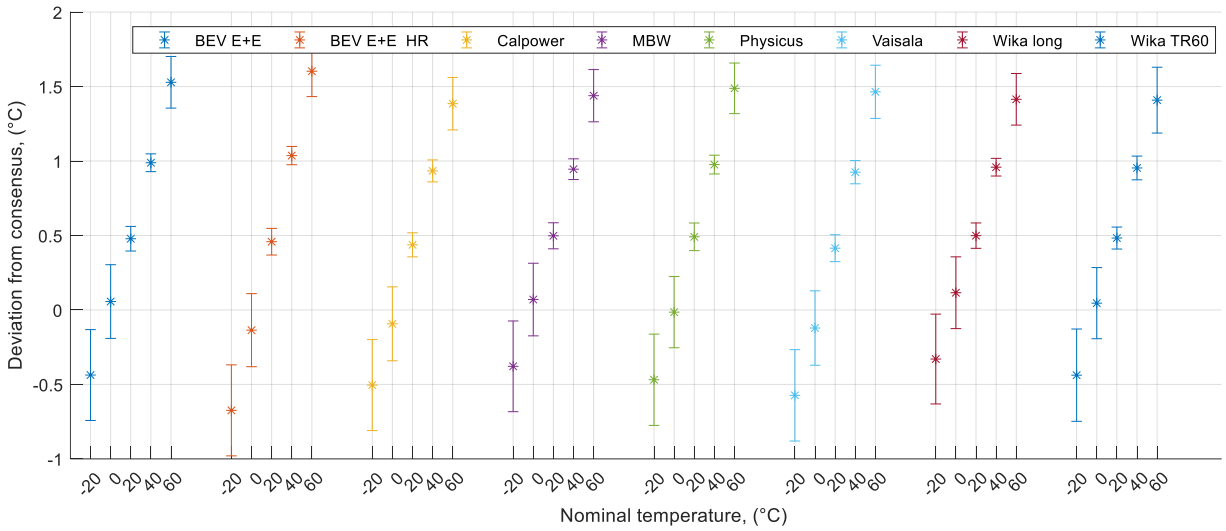
4.3.11 DMDM

T (°C)	BEV E+E	BEV E+E HR	CTP5000-170B	Calpower	MBW	Physicus	TR60	Vaisala
-20	-0.002 ± 0.103	0.047 ± 0.107	0.014 ± 0.105	0.067 ± 0.104	0.014 ± 0.103	-0.032 ± 0.101	0.036 ± 0.133	-0.022 ± 0.102
0	0.016 ± 0.076	-0.007 ± 0.079	0.067 ± 0.080	0.037 ± 0.076	0.019 ± 0.077	0.053 ± 0.069	-0.071 ± 0.092	-0.066 ± 0.076
20	0.010 ± 0.051	0.010 ± 0.057	0.064 ± 0.058	0.062 ± 0.051	0.003 ± 0.053	0.058 ± 0.056	-0.033 ± 0.084	0.013 ± 0.051
40	0.015 ± 0.051	0.031 ± 0.058	-0.020 ± 0.058	-0.001 ± 0.054	-0.009 ± 0.054	-0.003 ± 0.057	0.019 ± 0.084	-0.014 ± 0.052
60	0.033 ± 0.052	0.061 ± 0.058	-0.047 ± 0.058	-0.075 ± 0.059	-0.033 ± 0.054	-0.009 ± 0.066	-0.019 ± 0.088	0.018 ± 0.052



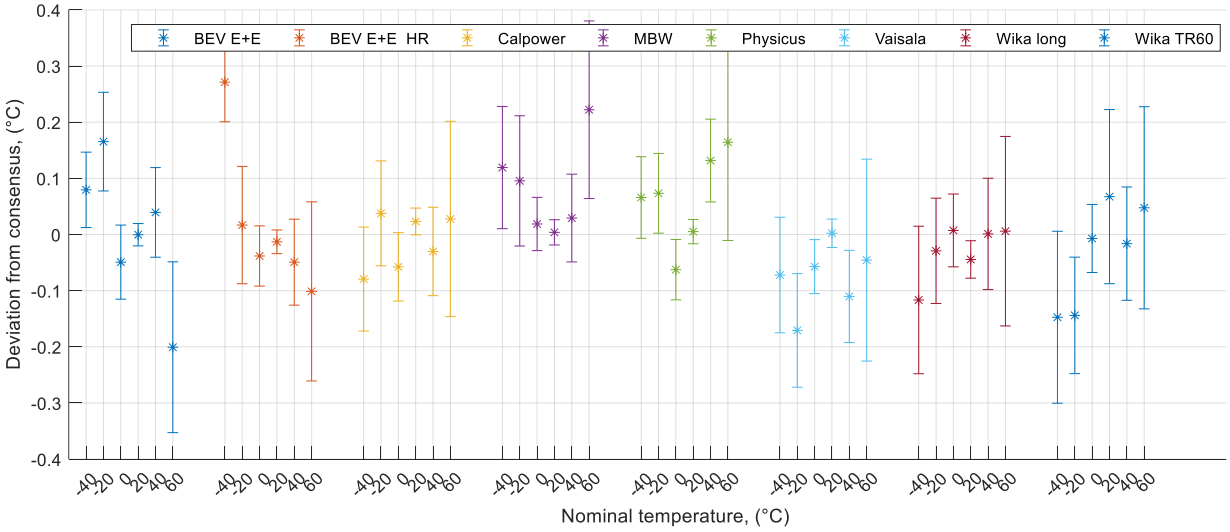
4.3.12 DPM

T (°C)	BEV E+E	BEV E+E HR	CTP5000-170B	Calpower	MBW	Physicus	TR60	Vaisala
-20	-0.440 ± 0.305	-0.670 ± 0.305	-0.330 ± 0.302	-0.503 ± 0.305	-0.380 ± 0.305	-0.469 ± 0.306	-0.438 ± 0.310	-0.575 ± 0.307
0	0.056 ± 0.247	-0.136 ± 0.245	0.116 ± 0.241	-0.093 ± 0.248	0.070 ± 0.244	-0.015 ± 0.239	0.046 ± 0.239	-0.121 ± 0.250
20	0.478 ± 0.083	0.458 ± 0.090	0.499 ± 0.085	0.437 ± 0.081	0.498 ± 0.087	0.491 ± 0.093	0.483 ± 0.074	0.415 ± 0.090
40	0.989 ± 0.060	1.035 ± 0.061	0.958 ± 0.059	0.933 ± 0.074	0.945 ± 0.070	0.976 ± 0.063	0.954 ± 0.080	0.926 ± 0.078
60	1.529 ± 0.173	1.602 ± 0.170	1.414 ± 0.173	1.384 ± 0.176	1.439 ± 0.175	1.488 ± 0.170	1.409 ± 0.221	1.466 ± 0.179



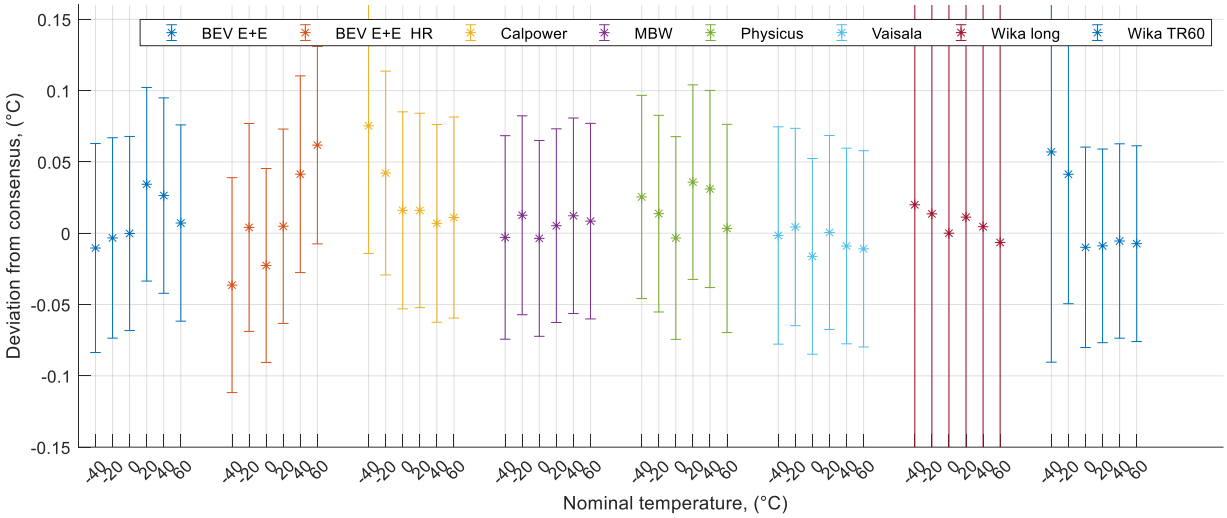
4.3.13 DTI

T (°C)	BEV E+E	BEV E+E HR	CTP5000-170B	Calpower	MBW	Physicus	TR60	Vaisala
-40	0.078 ± 0.067	0.275 ± 0.071	-0.117 ± 0.131	-0.079 ± 0.092	0.118 ± 0.109	0.066 ± 0.073	-0.152 ± 0.152	-0.076 ± 0.103
-20	0.163 ± 0.088	0.021 ± 0.104	-0.029 ± 0.094	0.039 ± 0.093	0.094 ± 0.116	0.073 ± 0.071	-0.144 ± 0.104	-0.173 ± 0.101
0	-0.049 ± 0.066	-0.038 ± 0.054	0.007 ± 0.065	-0.057 ± 0.061	0.019 ± 0.047	-0.063 ± 0.054	-0.006 ± 0.061	-0.057 ± 0.048
20	-0.000 ± 0.020	-0.013 ± 0.021	-0.044 ± 0.033	0.023 ± 0.024	0.004 ± 0.022	0.005 ± 0.022	0.068 ± 0.155	0.003 ± 0.025
40	0.040 ± 0.080	-0.051 ± 0.077	0.001 ± 0.099	-0.031 ± 0.079	0.030 ± 0.078	0.132 ± 0.074	-0.016 ± 0.101	-0.109 ± 0.082
60	-0.200 ± 0.152	-0.103 ± 0.159	0.006 ± 0.169	0.027 ± 0.174	0.223 ± 0.158	0.164 ± 0.175	0.048 ± 0.180	-0.044 ± 0.180



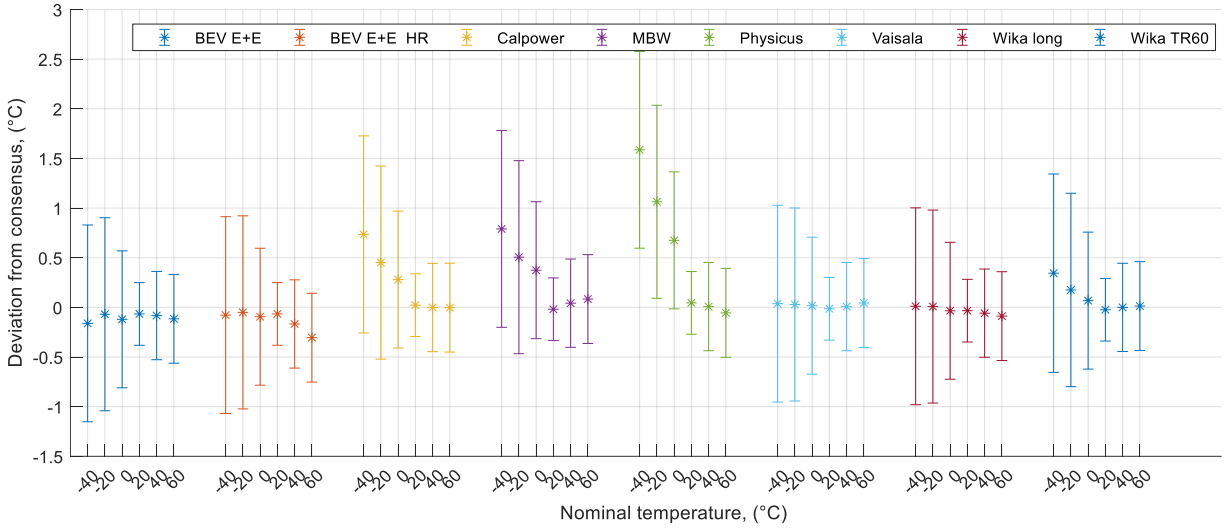
4.3.14 IMBIH

T (°C)	BEV E+E	BEV E+E HR	CTP5000-170B	Calpower	MBW	Physicus	TR60	Vaisala
-40	-0.012 ± 0.073	-0.033 ± 0.075	0.019 ± 0.455	0.075 ± 0.089	-0.004 ± 0.071	0.025 ± 0.071	0.052 ± 0.147	-0.005 ± 0.076
-20	-0.006 ± 0.070	0.008 ± 0.073	0.013 ± 0.453	0.044 ± 0.071	0.011 ± 0.070	0.014 ± 0.069	0.042 ± 0.091	0.003 ± 0.069
0	-0.000 ± 0.068	-0.023 ± 0.068	0.000 ± 0.453	0.016 ± 0.069	-0.004 ± 0.069	-0.003 ± 0.071	-0.009 ± 0.070	-0.016 ± 0.069
20	0.034 ± 0.068	0.004 ± 0.068	0.011 ± 0.453	0.016 ± 0.068	0.005 ± 0.068	0.036 ± 0.068	-0.009 ± 0.068	0.001 ± 0.068
40	0.027 ± 0.069	0.040 ± 0.069	0.005 ± 0.453	0.006 ± 0.069	0.012 ± 0.069	0.031 ± 0.069	-0.005 ± 0.068	-0.008 ± 0.069
60	0.008 ± 0.069	0.060 ± 0.069	-0.006 ± 0.453	0.010 ± 0.070	0.009 ± 0.069	0.003 ± 0.073	-0.007 ± 0.069	-0.009 ± 0.069



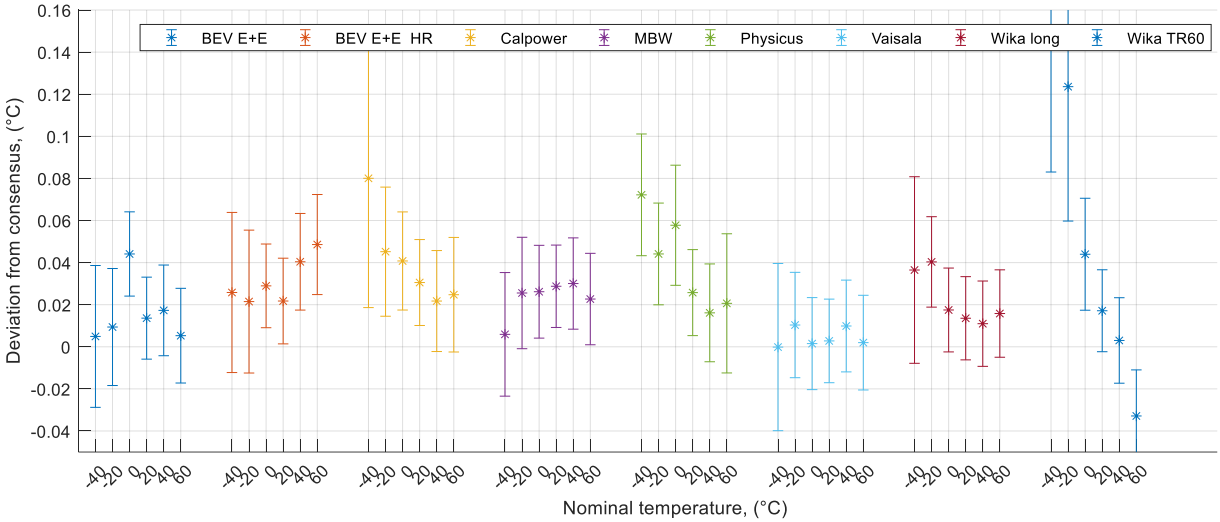
4.3.15 MBM

T (°C)	BEV E+E	BEV E+E HR	CTP5000-170B	Calpower	MBW	Physicus	TR60	Vaisala
-40	-0.163 ± 0.991	-0.074 ± 0.991	0.010 ± 0.991	0.734 ± 0.992	0.788 ± 0.991	1.587 ± 0.991	0.339 ± 0.999	0.034 ± 0.991
-20	-0.072 ± 0.972	-0.047 ± 0.972	0.008 ± 0.972	0.453 ± 0.972	0.504 ± 0.972	1.063 ± 0.972	0.176 ± 0.974	0.027 ± 0.972
0	-0.121 ± 0.690	-0.095 ± 0.690	-0.034 ± 0.690	0.280 ± 0.690	0.374 ± 0.690	0.675 ± 0.690	0.069 ± 0.690	0.017 ± 0.690
20	-0.066 ± 0.316	-0.066 ± 0.316	-0.033 ± 0.316	0.022 ± 0.316	-0.019 ± 0.316	0.045 ± 0.316	-0.024 ± 0.316	-0.014 ± 0.316
40	-0.082 ± 0.444	-0.169 ± 0.444	-0.059 ± 0.444	-0.002 ± 0.445	0.042 ± 0.444	0.008 ± 0.445	-0.000 ± 0.444	0.008 ± 0.444
60	-0.115 ± 0.447	-0.307 ± 0.447	-0.088 ± 0.447	-0.004 ± 0.447	0.084 ± 0.447	-0.055 ± 0.448	0.013 ± 0.447	0.045 ± 0.447



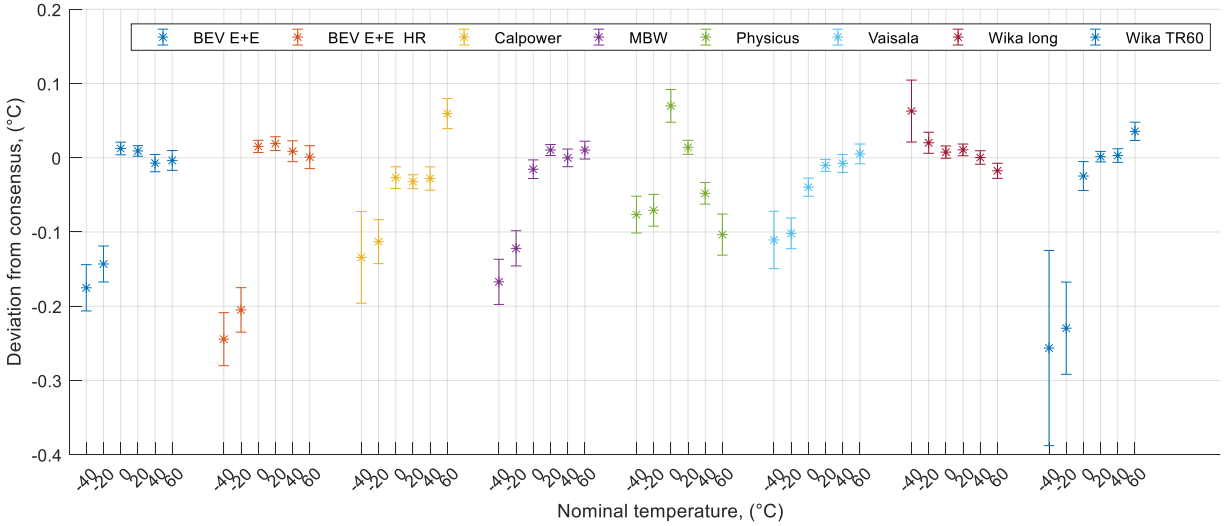
4.3.16 RISE

T (°C)	BEV E+E	BEV E+E HR	CTP5000-170B	Calpower	MBW	Physicus	TR60	Vaisala
-40	0.004 ± 0.034	0.029 ± 0.038	0.036 ± 0.044	0.080 ± 0.061	0.005 ± 0.029	0.072 ± 0.029	0.210 ± 0.132	-0.004 ± 0.040
-20	0.007 ± 0.028	0.025 ± 0.034	0.040 ± 0.021	0.047 ± 0.031	0.024 ± 0.026	0.044 ± 0.024	0.124 ± 0.064	0.009 ± 0.025
0	0.044 ± 0.020	0.029 ± 0.020	0.018 ± 0.020	0.041 ± 0.023	0.026 ± 0.022	0.058 ± 0.029	0.045 ± 0.027	0.002 ± 0.022
20	0.014 ± 0.019	0.021 ± 0.020	0.014 ± 0.020	0.030 ± 0.020	0.029 ± 0.020	0.026 ± 0.020	0.017 ± 0.019	0.003 ± 0.020
40	0.018 ± 0.022	0.039 ± 0.023	0.011 ± 0.020	0.021 ± 0.024	0.030 ± 0.022	0.016 ± 0.023	0.003 ± 0.020	0.011 ± 0.022
60	0.006 ± 0.022	0.047 ± 0.024	0.016 ± 0.021	0.024 ± 0.027	0.023 ± 0.022	0.021 ± 0.033	-0.033 ± 0.022	0.003 ± 0.023



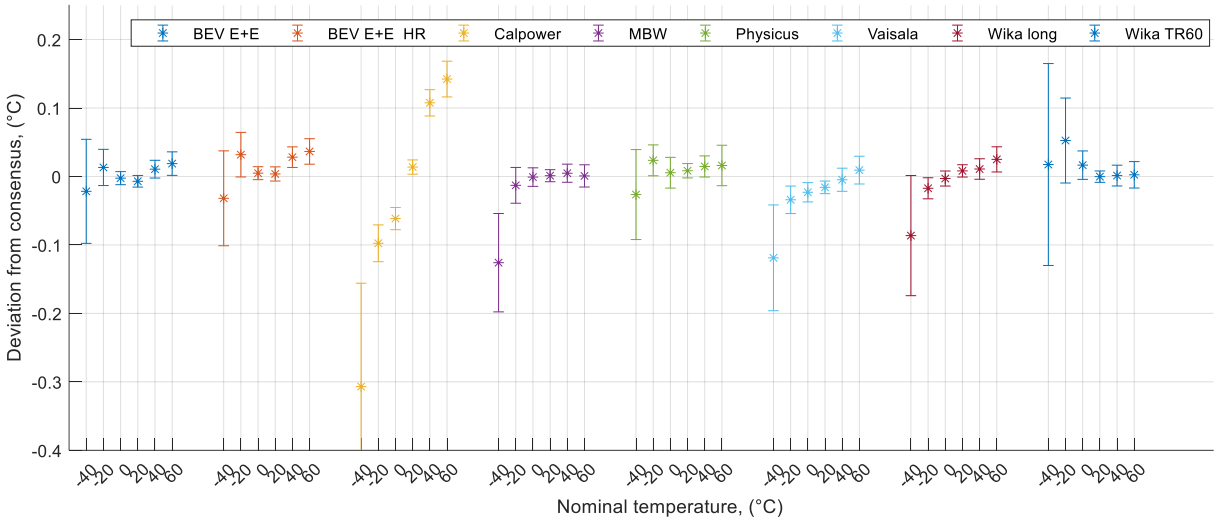
4.3.17 UME

T (°C)	BEV E+E	BEV E+E HR	CTP5000-170B	Calpower	MBW	Physicus	TR60	Vaisala
-40	-0.176 ± 0.031	-0.241 ± 0.035	0.062 ± 0.041	-0.134 ± 0.061	-0.169 ± 0.030	-0.077 ± 0.025	-0.261 ± 0.131	-0.114 ± 0.039
-20	-0.146 ± 0.024	-0.201 ± 0.030	0.020 ± 0.014	-0.111 ± 0.029	-0.123 ± 0.024	-0.071 ± 0.021	-0.229 ± 0.062	-0.104 ± 0.021
0	0.013 ± 0.008	0.015 ± 0.008	0.008 ± 0.008	-0.026 ± 0.015	-0.016 ± 0.012	0.070 ± 0.022	-0.024 ± 0.019	-0.039 ± 0.012
20	0.009 ± 0.007	0.019 ± 0.009	0.011 ± 0.008	-0.033 ± 0.009	0.011 ± 0.007	0.014 ± 0.009	0.002 ± 0.007	-0.010 ± 0.008
40	-0.007 ± 0.012	0.007 ± 0.014	0.000 ± 0.009	-0.029 ± 0.016	0.000 ± 0.012	-0.048 ± 0.015	0.003 ± 0.009	-0.007 ± 0.012
60	-0.003 ± 0.013	-0.000 ± 0.015	-0.017 ± 0.010	0.058 ± 0.020	0.011 ± 0.012	-0.104 ± 0.028	0.036 ± 0.012	0.007 ± 0.014



4.3.18 VTT MIKES

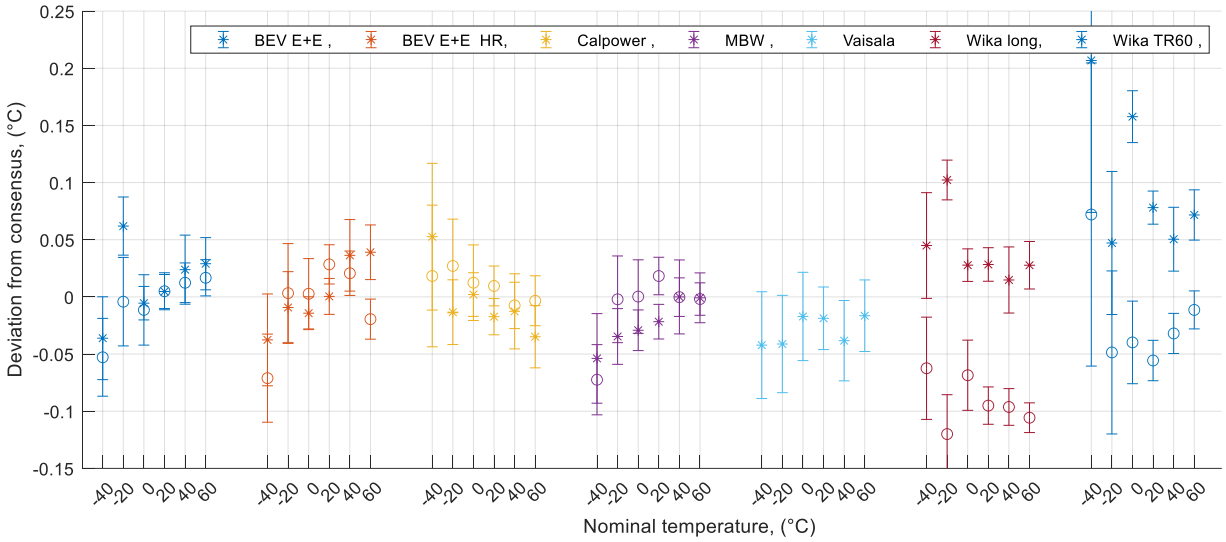
T (°C)	BEV E+E	BEV E+E HR	CTP5000-170B	Calpower	MBW	Physicus	TR60	Vaisala
-40	-0.023 ± 0.076	-0.028 ± 0.069	-0.087 ± 0.088	-0.307 ± 0.151	-0.127 ± 0.072	-0.026 ± 0.066	0.013 ± 0.147	-0.122 ± 0.077
-20	0.011 ± 0.026	0.036 ± 0.032	-0.017 ± 0.015	-0.096 ± 0.027	-0.014 ± 0.026	0.024 ± 0.023	0.053 ± 0.062	-0.036 ± 0.020
0	-0.002 ± 0.009	0.005 ± 0.009	-0.003 ± 0.011	-0.061 ± 0.016	-0.001 ± 0.013	0.006 ± 0.022	0.017 ± 0.021	-0.022 ± 0.014
20	-0.007 ± 0.008	0.003 ± 0.010	0.008 ± 0.009	0.013 ± 0.010	0.001 ± 0.009	0.008 ± 0.010	-0.000 ± 0.008	-0.016 ± 0.009
40	0.011 ± 0.013	0.027 ± 0.015	0.011 ± 0.015	0.107 ± 0.019	0.005 ± 0.013	0.015 ± 0.015	0.002 ± 0.015	-0.004 ± 0.017
60	0.019 ± 0.017	0.035 ± 0.019	0.025 ± 0.018	0.141 ± 0.026	0.002 ± 0.016	0.016 ± 0.029	0.003 ± 0.019	0.011 ± 0.020



4.3.19 INRIM

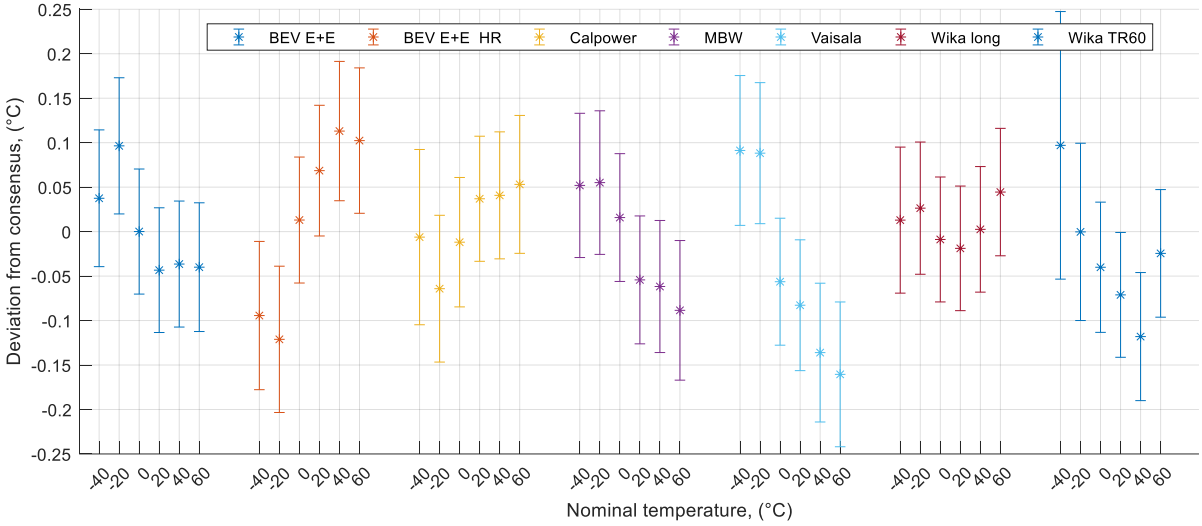
Inrim piloted loop 3. The data in the plot indicate the difference before and after the circulation. The numbers in the table are the average values, with the worst case as uncertainty.

T (°C)	BEV E+E	BEV E+E HR	CTP5000-170B	Calpower	MBW	TR60	Vaisala
-40	-0.046 ± 0.036	-0.051 ± 0.040	-0.010 ± 0.046	0.035 ± 0.064	-0.065 ± 0.039	0.135 ± 0.132	-0.046 ± 0.047
-20	0.026 ± 0.039	0.001 ± 0.043	-0.009 ± 0.034	0.008 ± 0.041	-0.019 ± 0.038	-0.000 ± 0.071	-0.043 ± 0.043
0	-0.008 ± 0.031	-0.006 ± 0.031	-0.020 ± 0.031	0.008 ± 0.033	-0.015 ± 0.032	0.060 ± 0.036	-0.016 ± 0.039
20	0.005 ± 0.016	0.014 ± 0.017	-0.033 ± 0.016	-0.004 ± 0.017	-0.002 ± 0.016	0.011 ± 0.018	-0.018 ± 0.027
40	0.018 ± 0.030	0.027 ± 0.031	-0.041 ± 0.029	-0.011 ± 0.033	0.000 ± 0.032	0.010 ± 0.028	-0.037 ± 0.035
60	0.023 ± 0.023	0.008 ± 0.024	-0.039 ± 0.021	-0.020 ± 0.027	-0.001 ± 0.022	0.030 ± 0.022	-0.015 ± 0.031



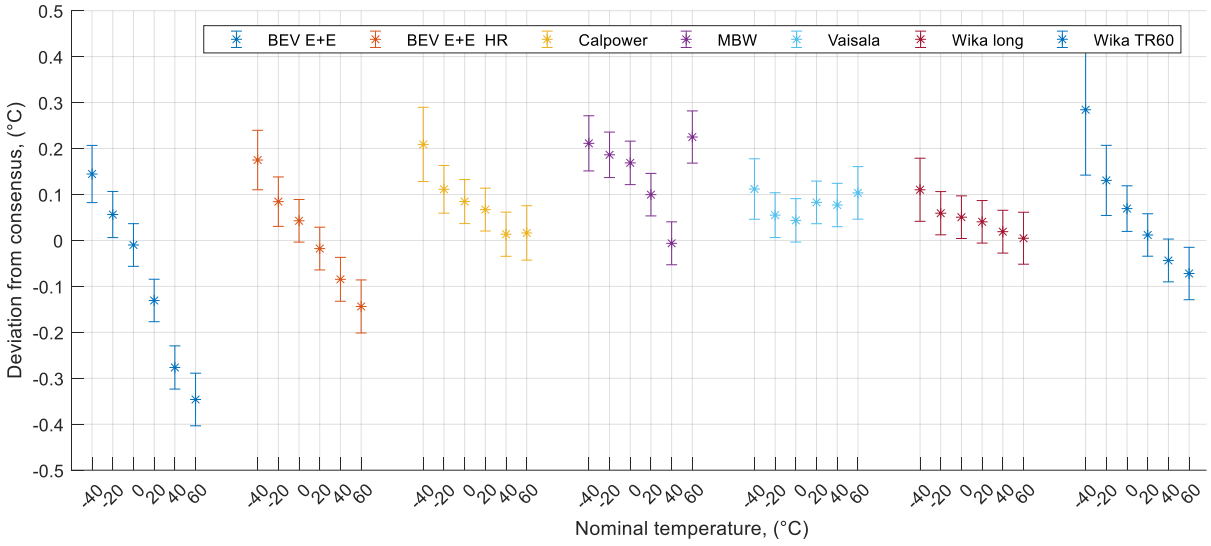
4.3.20 BEV

T (°C)	BEV E+E	BEV E+E HR	CTP5000-170B	Calpower	MBW	TR60	Vaisala
-40	0.037 ± 0.077	-0.094 ± 0.083	0.013 ± 0.082	-0.006 ± 0.099	0.052 ± 0.081	0.096 ± 0.150	0.091 ± 0.084
-20	0.096 ± 0.077	-0.121 ± 0.082	0.026 ± 0.074	-0.064 ± 0.083	0.055 ± 0.081	-0.000 ± 0.100	0.088 ± 0.079
0	0.000 ± 0.070	0.013 ± 0.071	-0.009 ± 0.070	-0.012 ± 0.073	0.016 ± 0.072	-0.040 ± 0.073	-0.056 ± 0.071
20	-0.043 ± 0.070	0.069 ± 0.073	-0.019 ± 0.070	0.037 ± 0.070	-0.054 ± 0.072	-0.071 ± 0.070	-0.083 ± 0.074
40	-0.036 ± 0.071	0.113 ± 0.078	0.003 ± 0.071	0.041 ± 0.071	-0.062 ± 0.074	-0.118 ± 0.072	-0.136 ± 0.078
60	-0.040 ± 0.072	0.102 ± 0.082	0.044 ± 0.072	0.053 ± 0.078	-0.089 ± 0.078	-0.024 ± 0.072	-0.161 ± 0.081



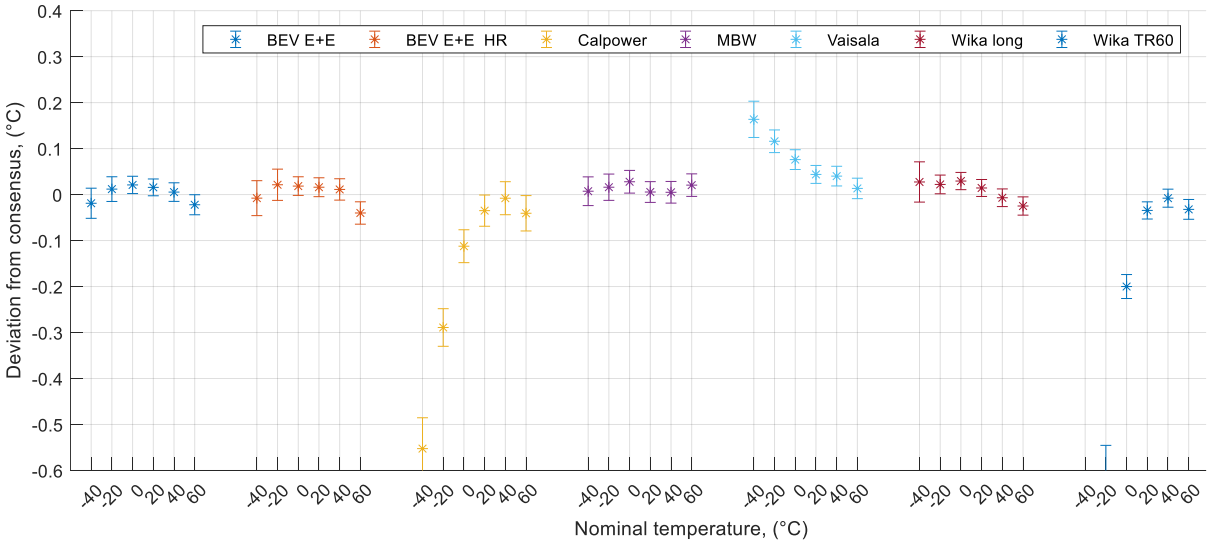
4.3.21 BFKH

T (°C)	BEV E+E	BEV E+E HR	CTP5000-170B	Calpower	MBW	TR60	Vaisala
-40	0.143 ± 0.062	0.179 ± 0.065	0.109 ± 0.069	0.209 ± 0.080	0.210 ± 0.060	0.279 ± 0.142	0.108 ± 0.066
-20	0.054 ± 0.050	0.088 ± 0.054	0.059 ± 0.047	0.113 ± 0.052	0.185 ± 0.049	0.131 ± 0.076	0.053 ± 0.049
0	-0.010 ± 0.046	0.043 ± 0.046	0.051 ± 0.046	0.085 ± 0.048	0.169 ± 0.047	0.070 ± 0.050	0.044 ± 0.047
20	-0.131 ± 0.046	-0.018 ± 0.047	0.041 ± 0.046	0.067 ± 0.047	0.100 ± 0.046	0.012 ± 0.046	0.083 ± 0.046
40	-0.276 ± 0.047	-0.086 ± 0.048	0.019 ± 0.047	0.013 ± 0.048	-0.006 ± 0.047	-0.043 ± 0.047	0.078 ± 0.047
60	-0.346 ± 0.057	-0.145 ± 0.058	0.005 ± 0.057	0.015 ± 0.059	0.226 ± 0.057	-0.072 ± 0.057	0.105 ± 0.057



4.3.22 CMI

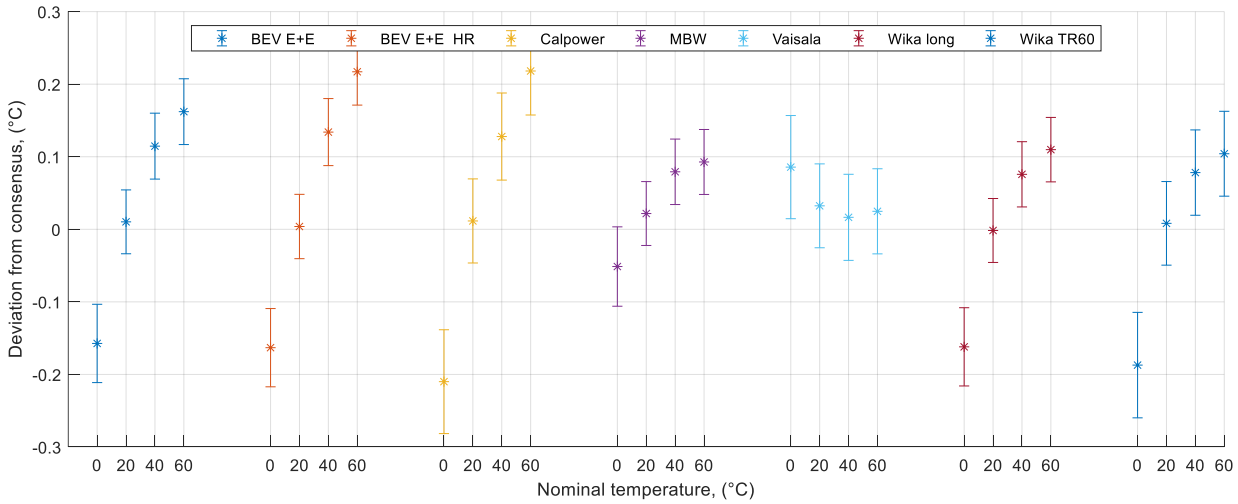
T (°C)	BEV E+E	BEV E+E HR	CTP5000-170B	Calpower	MBW	TR60	Vaisala
-40	-0.020 ± 0.033	-0.004 ± 0.038	0.027 ± 0.044	-0.553 ± 0.067	0.006 ± 0.031	-1.255 ± 0.131	0.160 ± 0.040
-20	0.009 ± 0.027	0.025 ± 0.034	0.022 ± 0.020	-0.287 ± 0.041	0.015 ± 0.029	-0.609 ± 0.063	0.114 ± 0.025
0	0.021 ± 0.019	0.018 ± 0.020	0.029 ± 0.019	-0.112 ± 0.036	0.028 ± 0.025	-0.199 ± 0.026	0.077 ± 0.021
20	0.016 ± 0.018	0.016 ± 0.021	0.014 ± 0.018	-0.035 ± 0.034	0.006 ± 0.023	-0.034 ± 0.019	0.044 ± 0.019
40	0.006 ± 0.020	0.010 ± 0.023	-0.007 ± 0.019	-0.009 ± 0.036	0.005 ± 0.024	-0.007 ± 0.020	0.041 ± 0.022
60	-0.022 ± 0.022	-0.041 ± 0.024	-0.025 ± 0.020	-0.042 ± 0.038	0.021 ± 0.025	-0.032 ± 0.021	0.015 ± 0.022



4.3.23 MIRS/UL-FE/LMK setup 1

The data here was acquired in the Thunder Scientific 2500 climate chamber.

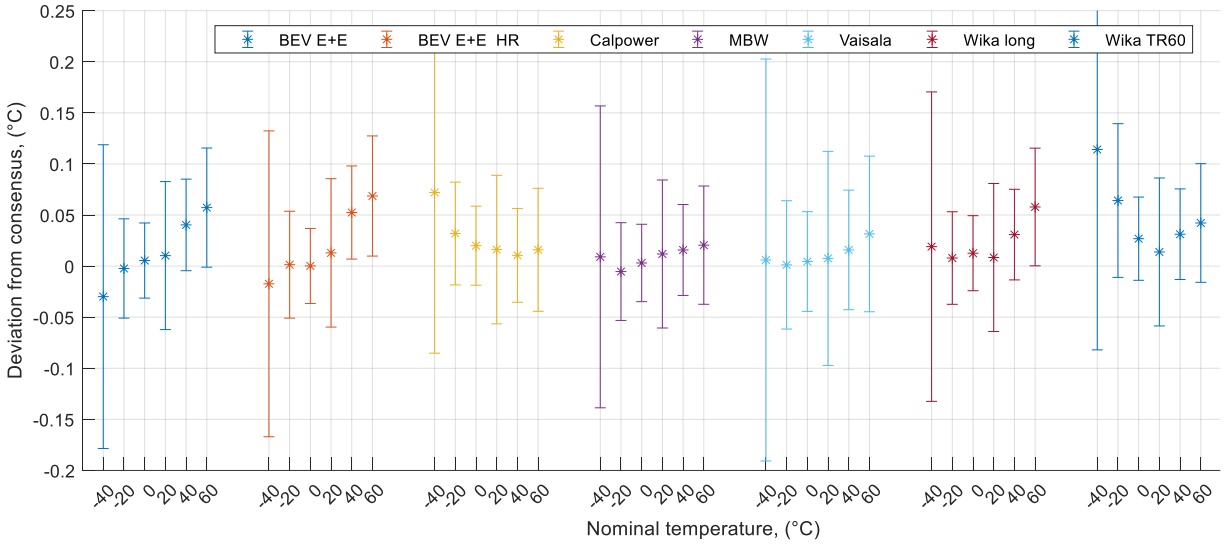
T (°C)	BEV E+E	BEV E+E HR	CTP5000-170B	Calpower	MBW	TR60	Vaisala
0	-0.157 ± 0.054	-0.163 ± 0.054	-0.162 ± 0.054	-0.210 ± 0.072	-0.051 ± 0.055	-0.186 ± 0.073	0.086 ± 0.071
20	0.010 ± 0.044	0.003 ± 0.044	-0.002 ± 0.044	0.011 ± 0.058	0.022 ± 0.044	0.008 ± 0.058	0.033 ± 0.058
40	0.115 ± 0.045	0.133 ± 0.046	0.076 ± 0.045	0.127 ± 0.060	0.079 ± 0.045	0.079 ± 0.059	0.017 ± 0.059
60	0.163 ± 0.045	0.216 ± 0.046	0.110 ± 0.044	0.217 ± 0.061	0.093 ± 0.045	0.104 ± 0.058	0.026 ± 0.059



4.3.24 MIRS/UL-FE/LMK setup 2

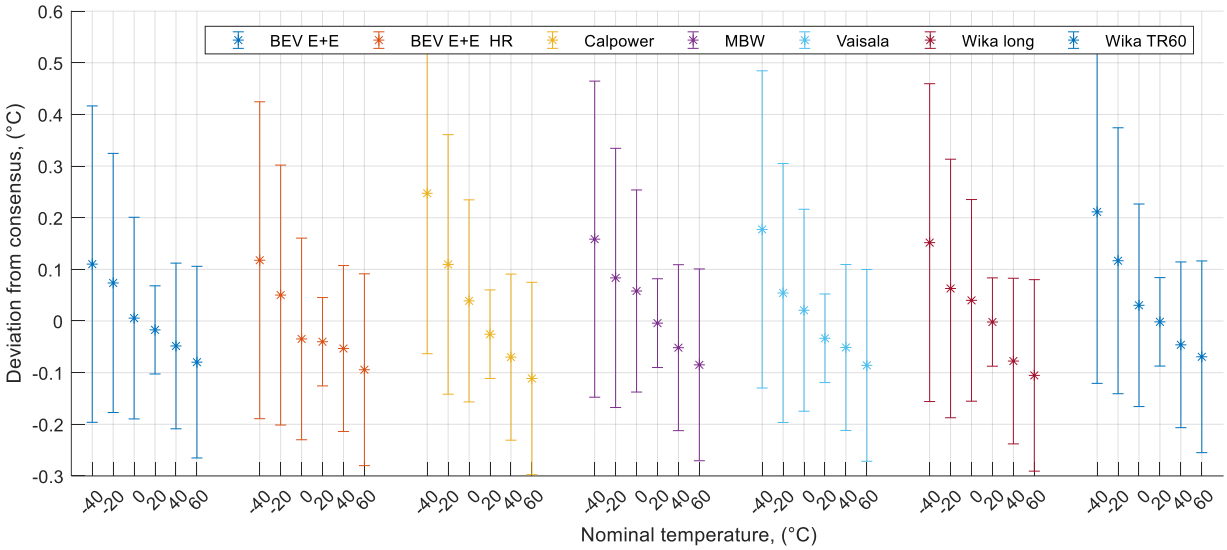
The data here was acquired in a Vötsch 7110 climate chamber.

T (°C)	BEV E+E	BEV E+E HR	CTP5000-170B	Calpower	MBW	TR60	Vaisala
-40	-0.031 ± 0.149	-0.014 ± 0.150	0.018 ± 0.151	0.072 ± 0.157	0.008 ± 0.148	0.109 ± 0.196	0.003 ± 0.197
-20	-0.005 ± 0.049	0.005 ± 0.052	0.008 ± 0.045	0.033 ± 0.050	-0.006 ± 0.048	0.065 ± 0.075	-0.001 ± 0.063
0	0.005 ± 0.037	-0.000 ± 0.037	0.013 ± 0.037	0.020 ± 0.039	0.003 ± 0.038	0.028 ± 0.041	0.005 ± 0.049
20	0.010 ± 0.072	0.013 ± 0.073	0.008 ± 0.072	0.016 ± 0.073	0.012 ± 0.072	0.014 ± 0.072	0.008 ± 0.105
40	0.041 ± 0.045	0.051 ± 0.046	0.031 ± 0.044	0.010 ± 0.046	0.016 ± 0.044	0.032 ± 0.044	0.017 ± 0.059
60	0.058 ± 0.058	0.067 ± 0.059	0.058 ± 0.058	0.015 ± 0.060	0.021 ± 0.058	0.042 ± 0.058	0.033 ± 0.076



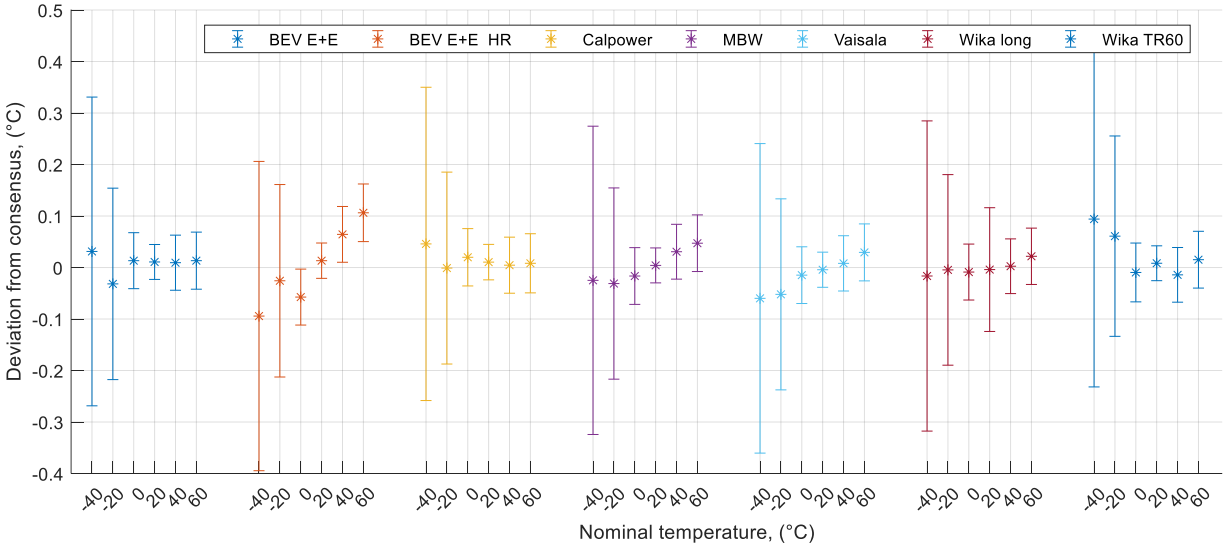
4.3.25 NQIS/EIM

T (°C)	BEV E+E	BEV E+E HR	CTP5000-170B	Calpower	MBW	TR60	Vaisala
-40	0.109 ± 0.306	0.121 ± 0.307	0.151 ± 0.308	0.247 ± 0.311	0.157 ± 0.306	0.206 ± 0.332	0.174 ± 0.307
-20	0.071 ± 0.251	0.054 ± 0.252	0.063 ± 0.250	0.111 ± 0.251	0.082 ± 0.251	0.117 ± 0.257	0.052 ± 0.251
0	0.006 ± 0.195	-0.035 ± 0.195	0.040 ± 0.195	0.039 ± 0.196	0.058 ± 0.196	0.031 ± 0.196	0.021 ± 0.196
20	-0.017 ± 0.085	-0.041 ± 0.086	-0.002 ± 0.085	-0.026 ± 0.086	-0.004 ± 0.086	-0.001 ± 0.086	-0.033 ± 0.086
40	-0.048 ± 0.160	-0.055 ± 0.161	-0.078 ± 0.160	-0.071 ± 0.161	-0.051 ± 0.161	-0.046 ± 0.160	-0.050 ± 0.161
60	-0.079 ± 0.186	-0.096 ± 0.186	-0.105 ± 0.185	-0.112 ± 0.186	-0.084 ± 0.186	-0.069 ± 0.186	-0.084 ± 0.186



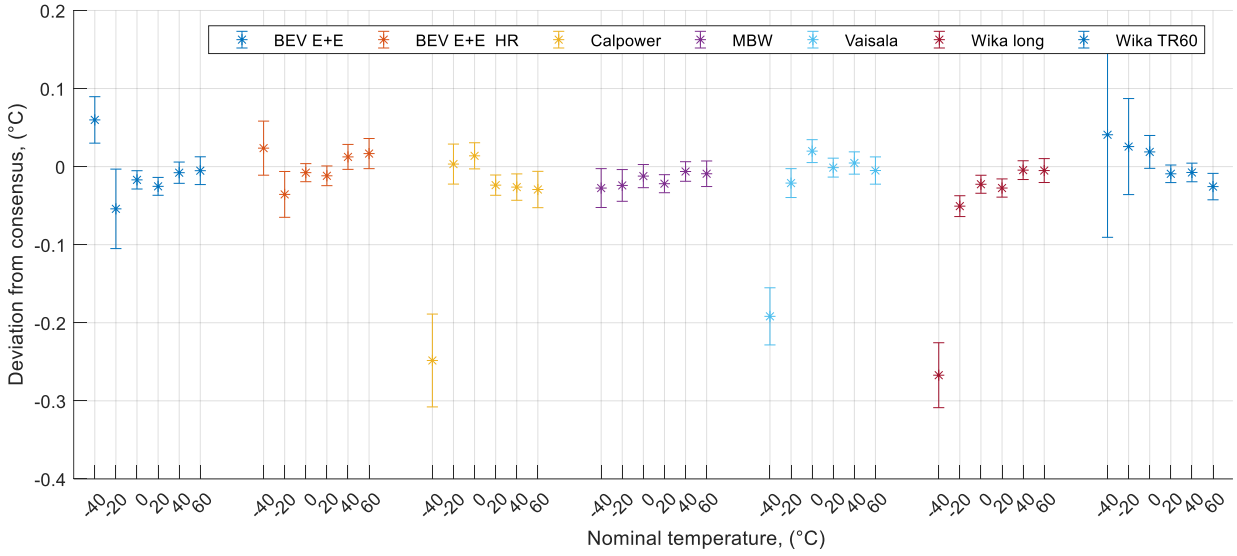
4.3.26 PTB

T (°C)	BEV E+E	BEV E+E HR	CTP5000-170B	Calpower	MBW	TR60	Vaisala
-40	0.030 ± 0.300	-0.091 ± 0.300	-0.017 ± 0.301	0.046 ± 0.304	-0.026 ± 0.299	0.089 ± 0.326	-0.063 ± 0.301
-20	-0.034 ± 0.186	-0.022 ± 0.187	-0.005 ± 0.185	0.001 ± 0.186	-0.032 ± 0.186	0.061 ± 0.194	-0.054 ± 0.185
0	0.013 ± 0.054	-0.058 ± 0.054	-0.009 ± 0.054	0.020 ± 0.056	-0.017 ± 0.055	-0.009 ± 0.057	-0.014 ± 0.055
20	0.011 ± 0.034	0.013 ± 0.034	-0.004 ± 0.120	0.010 ± 0.034	0.004 ± 0.034	0.008 ± 0.034	-0.004 ± 0.034
40	0.010 ± 0.053	0.063 ± 0.054	0.003 ± 0.053	0.004 ± 0.054	0.031 ± 0.053	-0.014 ± 0.053	0.009 ± 0.054
60	0.014 ± 0.055	0.105 ± 0.056	0.022 ± 0.055	0.007 ± 0.057	0.048 ± 0.055	0.015 ± 0.055	0.031 ± 0.055



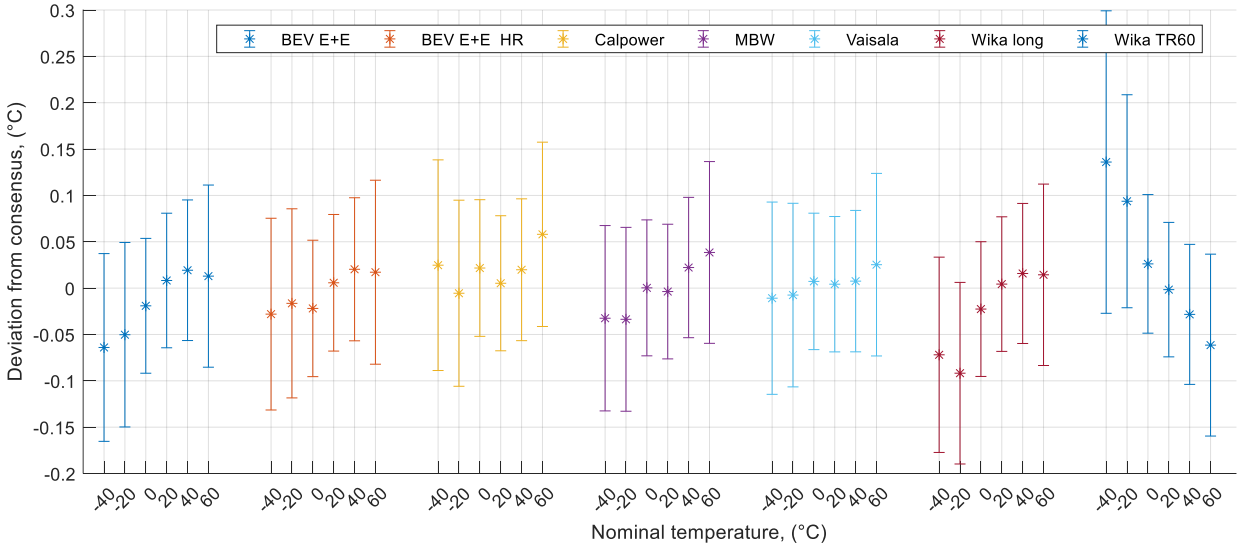
4.3.27 SMD

T (°C)	BEV E+E	BEV E+E HR	CTP5000-170B	Calpower	MBW	TR60	Vaisala
-40	0.059 ± 0.030	0.027 ± 0.034	-0.268 ± 0.041	-0.248 ± 0.059	-0.029 ± 0.025	0.036 ± 0.131	-0.195 ± 0.037
-20	-0.057 ± 0.051	-0.032 ± 0.029	-0.051 ± 0.013	0.005 ± 0.026	-0.025 ± 0.020	0.026 ± 0.061	-0.023 ± 0.018
0	-0.017 ± 0.012	-0.008 ± 0.011	-0.023 ± 0.011	0.014 ± 0.017	-0.012 ± 0.015	0.020 ± 0.021	0.020 ± 0.015
20	-0.025 ± 0.011	-0.012 ± 0.013	-0.027 ± 0.012	-0.024 ± 0.013	-0.022 ± 0.012	-0.009 ± 0.011	-0.001 ± 0.012
40	-0.007 ± 0.014	0.011 ± 0.016	-0.005 ± 0.012	-0.027 ± 0.017	-0.006 ± 0.013	-0.007 ± 0.012	0.006 ± 0.014
60	-0.005 ± 0.018	0.015 ± 0.019	-0.005 ± 0.015	-0.030 ± 0.023	-0.008 ± 0.016	-0.025 ± 0.017	-0.003 ± 0.018



4.3.28 SMU

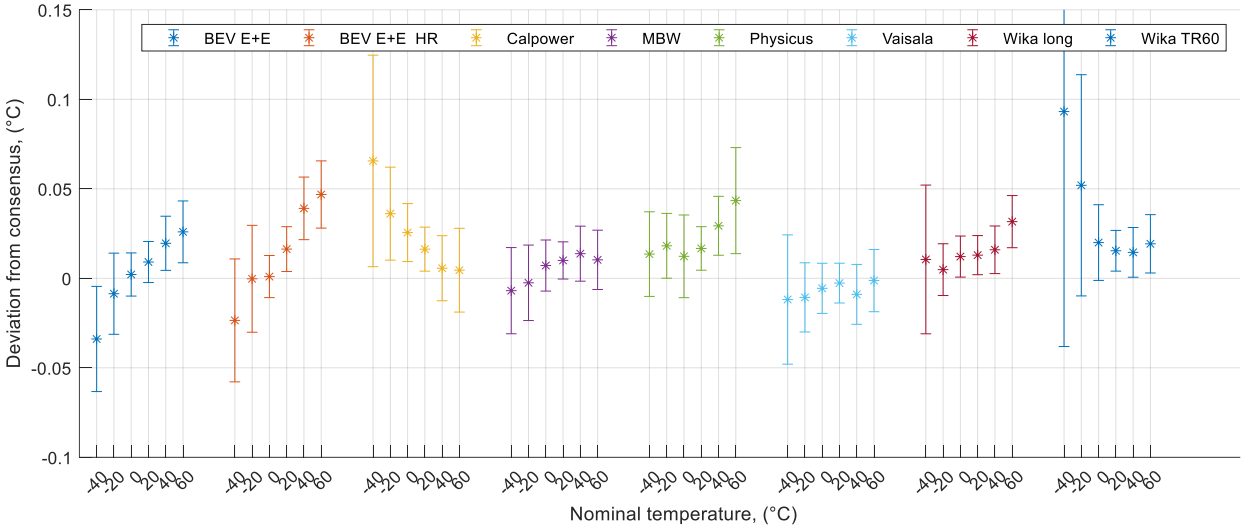
T (°C)	BEV E+E	BEV E+E HR	CTP5000-170B	Calpower	MBW	TR60	Vaisala
-40	-0.065 ± 0.101	-0.025 ± 0.103	-0.073 ± 0.105	0.025 ± 0.113	-0.034 ± 0.100	0.131 ± 0.163	-0.014 ± 0.104
-20	-0.053 ± 0.100	-0.013 ± 0.102	-0.092 ± 0.098	-0.004 ± 0.100	-0.035 ± 0.099	0.094 ± 0.115	-0.009 ± 0.099
0	-0.019 ± 0.073	-0.022 ± 0.074	-0.023 ± 0.073	0.022 ± 0.074	0.000 ± 0.073	0.027 ± 0.075	0.008 ± 0.074
20	0.008 ± 0.073	0.005 ± 0.074	0.004 ± 0.073	0.005 ± 0.073	-0.004 ± 0.073	-0.001 ± 0.073	0.005 ± 0.073
40	0.020 ± 0.076	0.019 ± 0.077	0.016 ± 0.076	0.019 ± 0.076	0.022 ± 0.076	-0.028 ± 0.076	0.008 ± 0.076
60	0.013 ± 0.098	0.016 ± 0.099	0.014 ± 0.098	0.057 ± 0.099	0.039 ± 0.098	-0.061 ± 0.098	0.027 ± 0.099



4.3.29 JV

JV measured on all probes, which means that each model probe was measured three times (twice for the Physicus probe). The deviation values are the model averages, while the uncertainties are the worst case for each (probe model, temperature) case.

T (°C)	BEV E+E	BEV E+E HR	CTP5000-170B	Calpower	MBW	Physicus	TR60
-40	-0.035 ± 0.029	-0.020 ± 0.034	0.010 ± 0.041	0.066 ± 0.059	-0.008 ± 0.024	0.014 ± 0.024	0.088 ± 0.130
-20	-0.011 ± 0.023	0.004 ± 0.030	0.005 ± 0.014	0.038 ± 0.026	-0.004 ± 0.021	0.018 ± 0.018	0.052 ± 0.062
0	0.002 ± 0.012	0.001 ± 0.012	0.012 ± 0.011	0.026 ± 0.016	0.007 ± 0.014	0.012 ± 0.023	0.021 ± 0.021
20	0.009 ± 0.011	0.016 ± 0.012	0.013 ± 0.011	0.016 ± 0.012	0.010 ± 0.010	0.017 ± 0.012	0.016 ± 0.011
40	0.020 ± 0.015	0.038 ± 0.017	0.016 ± 0.013	0.005 ± 0.018	0.014 ± 0.015	0.029 ± 0.016	0.015 ± 0.014
60	0.026 ± 0.017	0.046 ± 0.019	0.032 ± 0.015	0.003 ± 0.023	0.011 ± 0.017	0.043 ± 0.030	0.020 ± 0.016



5 Auxiliary information

5.1 Uncertainty budget

An uncertainty budget for the reference temperature was supplied in the reporting template. Some participants included detailed calculations for all temperatures, while other participants only reported details for an example temperature.

Table 8 Overview of the uncertainty contributions used by the participants.

	Calibration	Drift	Observations	Self heating	Radiative heating	Air flow cooling	Stability of chamber	Heat transfer supports	Dynamic temperature	Heat transfer wires	Uniformity of chamber	Humidity	Hysteresis	Reproducibility	Non-linearity	Bridge SPRT	Bridge SPRT drift	Bridge SPRT resolution
NSAI	X	X	X	X	X	X	X	X	-	X	X	-	-	-	-	-	-	-
INTiBS	X	X	X	-	-	X	X	-	-	-	X	-	-	-	-	-	-	-
GUM	X	X	X	X	X	-	X	-	-	-	X	-	-	-	-	-	-	-
BEV/E+E	X	X	X	X	-	-	X	X	-	-	X	-	-	-	-	-	-	-
INTA	X	X	X	X	X	X	X	X	X	X	X	-	-	-	-	-	-	-
LNE-CETIAT	X	X	X	X	X	X	-	X	X	X	X	-	-	-	-	-	-	-
NPL	X	X	X	X	X	X	X	X	X	X	X	-	X	X	X	-	-	-
JV	X	X	X	X	X	X	X	-	X	-	X	-	-	-	-	-	-	-
CEM	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DTI	X	X	X	X	X	-	-	-	-	-	X	-	-	-	-	-	-	-
RISE	X	X	X	X	X	X	X	X	-	X	X	-	-	-	-	-	-	-
VTT MIKES	X	X	X	X	X	-	-	-	-	X	X	-	-	-	-	-	-	-
UME	X	X	X	X	X	X	X	X	X	X	X	-	-	-	-	-	-	-
DPM	X	X	X	X	X	-	X	-	X	X	X	-	-	-	-	-	-	-
DMDM	X	X	X	X	X	X	X	X	X	X	X	-	-	-	-	-	-	-
MBM	X	X	X	X	X	-	X	-	-	-	X	-	-	-	-	-	-	-
IMBIH	X	X	X	-	X	X	X	-	-	-	X	X	-	-	-	-	-	-
INRIM	X	-	X	-	X	-	X	-	-	-	X	-	-	-	-	-	-	-
BFKH	X	X	X	X	-	-	X	-	-	-	X	-	-	-	-	-	-	-
SMD	X	X	X	X	-	-	X	-	-	-	X	-	-	-	-	-	-	-
CMI	X	X	X	X	X	X	X	X	X	X	X	-	-	-	-	-	-	-
NQIS / EIM	X	X	X	X	X	X	X	X	X	X	X	-	-	-	-	-	-	-
SMU	X	X	X	X	-	-	X	-	-	-	X	-	-	-	-	X	X	X
BEV	X	X	X	X	X	-	X	-	-	-	X	-	-	-	-	-	-	-
MIRS/UL-FE/LMK	X	X	X	X	X	-	X	-	-	-	X	-	-	-	-	-	-	-
PTB	X	X	X	X	X	-	X	-	-	X	X	-	-	-	-	-	-	-

Table 9 provides an overview of the contributions, along with the range of attributed values. As can be seen the magnitude of the contributions varies substantially. For the chamber uniformity the variation is two orders of magnitude. The variation is not in itself a problem, since uncertainty budgets should be tailored to the experiments. Indeed, some of the participants used subchambers specifically to reduce the issue of chamber uniformity, and some participants even immersed the chamber in liquid baths. We cannot rule out, however, that some participants have been either excessively optimistic or pessimistic.

Table 9 Range of the numeric values for contributions to uncertainty. The last column is a measure of the importance of each contribution according to the reported values from participants. The numeric value is a count of the number of participants for which the contribution is the largest contributor to combined uncertainty, in at least half the measurement temperatures.

Contribution	Smallest reported value, °C	Largest reported value, °C	Largest at N participants
Uniformity of chamber	0.001	0.185	13
Stability of chamber	0.000	0.058	3
Air flow cooling	0.000	0.020	1
Calibration	0.001	0.025	1
Observations	0.000	0.043	1
Radiative heating	0.000	0.030	1
Self heating	0.000	0.031	1
Bridge SPRT	0.003	0.003	0
Bridge SPRT drift	0.003	0.003	0
Bridge SPRT resolution	0.000	0.000	0
Drift	0.000	0.040	0
Dynamic temperature	0.000	0.040	0
Heat transfer supports	0.001	0.010	0
Heat transfer wires	0.001	0.024	0
Humidity	0.004	0.004	0
Hysteresis	0.003	0.003	0
Non-linearity	0.003	0.003	0
Reproducibility	0.003	0.003	0

6 Discussion, conclusions and summary

The protocol of the ILC did not prescribe specific measurement procedures to follow. In contrast, participants were encouraged to use their own methods, and if time permitted, to collect auxiliary characterisation data. The purpose was to investigate different calibration methods for air sensors, and to search for patterns that could help establish a recommended best practice for air thermometer calibration. With this in mind it is not surprising that the results are inconsistent when judged by standard evaluation methods of ILC results.

However, some other observations probably cannot be explained simply by a diversity in methods. The circulating probes were found to change during circulation. The protocol prescribed an ice point or TPW check both immediately after receiving and just before sending the probes, and the intention was to use these measurements to assess the drift. However, the measurements in liquid baths performed at the pilots revealed that some of the probes changed substantially, and we have been unable to identify any patterns that could have enabled a systematic correction due to drift. The observed changes are erratic in temperature, and in some cases the same probe model appears to behave differently depending on the loop. One possible, but unverified, explanation is that some of the probes had been subject to liquid ingress at one or more participants during the circulation, and that this has affected their electrical resistance.

Because of the changes (or drift) it is not obvious that the participants can be directly compared to each other. However, by analysing the data with the random effects model it was hoped that the estimated between-lab variance could be realistically modelled from the data. It was found that the

contribution of the between-lab variance to the consensus uncertainty was substantial. It was also found that the consensus uncertainty was strongly temperature dependent: quite small at around 20 mK between 0 °C and 20 °C, increasing to 200 mK at -80 °C.

The data allowed an independent assessment of laboratory performance by aggregating the unilateral degrees of equivalence (DoE) from each probe and temperature. The participants reported 40 to 60 data points depending on the temperature range and the number of probes (in the case of loop 3 only 7 probes). In each case the average of the deviation from consensus was computed, along with the associated standard deviation. We posit that this is a measure of the consistency of the participant, with the attractive feature that it is independent of the reported uncertainty. The caveat is that not all participants contribute equally to the consensus calculation – the smallest reported uncertainties will have higher weights – and hence that those participants who contribute more to the consensus are also more likely to be close to it.

Low temperatures seem to be challenging. This is seen in the unilateral degrees of equivalence, where the deviations are typically large for the low temperatures. Below -40 °C there are some additional technical hurdles, and some of the participants in loop 1 could not reach below -70 °C.

Self heating results are surprisingly scattered, but typically the scatter is smaller for liquid measurements than air measurements. A plausible mechanism is that the sensors have long stabilisation times in air, and this may affect the self heating results if the participants have exposed the sensors to a specific temperature for different durations. In air the participants could also have measured under different conditions such as windspeed, humidity or pressure. The stability of the thermostat used (typically a climate chamber) might also have an influence. However, it seems that a closer inspection of the self heating evaluation is warranted, and that perhaps the community should agree on a best practice.

The protocol offered an uncertainty budget for the reference air temperature measurement. The most common main contributor to uncertainty was the chamber uniformity, but its input value varied substantially between participants. This is not unreasonable given the wide range of methods employed. Around half of the participants used a subchamber inside a climate chamber, with the purpose of improving uniformity and shielding temporal variations in the air temperature.

An important aim for the ILC was to identify performance differences in calibration methods. However, a fundamental limitation with the current dataset is the lacking stability of the circulating probes. This makes it more difficult to ensure that differences in performance between participants can be attributed to their method, and not to the circulating probes themselves. Nevertheless, a few lessons for the future can be gleaned from the results:

- Self heating characterisations should be scrutinized. In some cases they will make a noticeable contribution to uncertainty, but the large scatter in the amount of self heating attributed in this ILC to the same probes makes it hard to pinpoint exactly what the contribution should be.
- Chamber uniformity is often a large contribution to uncertainty. Measures to improve this should be taken.
- In air it is also important to expand the array of recommended probe characteristics to measure. The list should at least include the sensitivity to windspeed. A challenge is that many effects are governed by more basic physical characteristics, such as shape, internal design and heat transfer, and it may not be possible to decouple directly measurable

quantities. For instance, the self heating will also depend on the windspeed, not just properties of the probe itself.

- A future ILC with much more stable artefacts should be considered.

Currently it has not been explored yet whether the data may be used to identify certain methods or setups that are better, in the sense of more homogeneous, than others. Again the probe stability is an issue, but it might be possible to identify smaller subsets of data (restricted to some of the more stable probes, or to a subset of participants) that could shed more light this question. The current report has not explored this avenue, but future work will proceed in this direction.

7 Bibliography

[1] Å. A. F. Olsen, Data from interlaboratory comparison of air thermometer calibrations, ATM ILC, 2019-2021, Zenodo, 2023, DOI: 10.5281/zenodo.8409783.

[2] M. Heinonen, M. Anagnostou, J. Bartolo, S. Bell, R. Benyon, R. A. Bergerud, J. Bojkovski, N. Böse, C. Dinu, D. Smorgon, K. Flakiewicz, M. J. Martin, S. Nedialkov, M. B. Nielsen, S. Oğuz Aytekin, J. Otych, M. Pedersen, M. Rujan, N. Testa, E. Turzó-András, M. Vilbaste og M. White, «Comparison of Air Temperature Calibrations,» *International Journal of Thermophysics,* vol. 35, p. 1251–1272, 2014.

[3] M. de Podesta, S. Bell og R. Underwood, «Air temperature sensors: dependence of radiative errors on sensor diameter in precision metrology and meteorology,» *Metrologia,* vol. 55, p. 229–244, February 2018.

[4] A. Koepke, T. Lafarge, A. Possolo og B. Toman, «Consensus building for interlaboratory studies, key comparisons, and meta-analysis,» *Metrologia,* vol. 54, pp. S34-S62, 2017.

[5] A. A. Veroniki, D. Jackson, W. Viechtbauer, R. Bender, J. Bowden, G. Knapp, O. Kuss, J. P. T. Higgins, D. Langan og G. Salanti, «Methods to estimate the between-study variance and its uncertainty in meta-analysis,» *Res. Syn. Meth.,* vol. 7, p. 55–79, March 2016.

[6] B. J. BIGGERSTAFF og R. L. TWEEDIE, «INCORPORATING VARIABILITY IN ESTIMATES OF HETEROGENEITY IN THE RANDOM EFFECTS MODEL IN META-ANALYSIS,» *Statistics in Medicine,* vol. 16, pp. 753-768, 1997.

[7] P. McBrien, «A theoretical semi-empirical model for precise air temperature metrology,» 2021.

[8] C. García Izquierdo, S. Hernández, A. González, L. Matias, L. Šindelářová, R. Strnad og D. del Campo, «Evaluation of the self-heating effect in a group of thermometers used in meteorological and climate applications,» *Meteorol Appl,* vol. 26, p. 117–129, January 2019.