



Systematic errors in borehole temperature monitoring records in permafrost – Part 1: Stability of calibrated temperature sensor offsets.

Soňa Tomaškovičová¹, Fredrik Johansson¹, Michele Citterio², and Thomas Ingeman-Nielsen¹

¹Department of Environmental & Resource Engineering, Technical University of Denmark, Kongens Lyngby, Denmark

²Department of Glaciology & Climate, Geological Survey of Denmark and Greenland, Copenhagen, Denmark

Correspondence: Soňa Tomaškovičová (soto@dtu.dk)

Abstract. This study investigated some of the systematic errors that contribute to the uncertainty budget associated with short- and long-term temperature measurements in permafrost. We quantified errors associated with four types of sensors' behaviors that are not addressed by the most commonly carried out single-point 0 °C bias correction: i) long-term drift of the 0 °C-offset using sensor recalibration data after six and nine years, ii) non-linear sensor offsets dependence on temperature (multiple-step calibration), iii) drift of the 0 °C-offsets following thermal stress, and iv) hysteresis of temperature values measured by the sensor. We measured sensor drift rates up to 0.0157 °C per year, with 95.3% of evaluated sensors drifting towards warmer temperatures. Extrapolated over 30 years, the observed positive drift of 0.471 °C would be comparable to the rate of permafrost warming reported from some permafrost regions, assuming that at the deployment temperatures, the drift rate remained nearly linear. We also confirmed non-linear sensor offsets across the –20 °C to 20 °C range, with larger errors at negative temperatures. These results support the argument for multi-point calibration, particularly at and below 0 °C, to improve the representation of phase-change processes, especially in fine-grained and saline soils. Additionally, 43 thermal cycles between –20 °C and 20 °C in the lab produced measurable offset drift, while the sensor readings exhibited hysteresis, indicating dependence on prior thermal state. Based on the physical mechanisms underlying these errors, we argue that similar drift-inducing and drift-enhancing processes are likely present in other sensor technologies commonly used in permafrost monitoring. We therefore recommend extending current calibration best practices to include periodic recalibration of individual sensors, comprehensive reporting of sensor metadata and calibration histories, and calibration guidelines tailored separately to long-term monitoring and short-term measurement campaigns. Such measures would improve the quality, traceability, and comparability of permafrost temperature records that we are acquiring as a permafrost community.

1 Introduction

Long-term timeseries of ground temperatures are essential data for assessing ground thermal state and trends, and for calibration and validation of the ground thermal model component of Earth Systems models. The reliability of conclusions drawn from these datasets and models depends, among other factors, on accuracy and long-term stability of the measurement systems (combination of sensors and loggers) used to acquire the temperature data.



The accuracy of a measurement system can be estimated by a calibration procedure, whereby the sensor offsets from a known, traceable standard are quantified at various temperatures in the expected temperature range. The sensor's records may then be adjusted by applying a correction offset to align with the said standard. Long-term stability can be evaluated by repeating such calibrations after a period of deployment in the field. The information from the (re)calibration procedures can be used to quantify and correct systematic errors in temperature observations, which could otherwise go undetected and produce biases in the datasets.

Temperature sensors for field deployment can be ordered mounted on strings connected to a datalogger, and are typically supplied without explicit calibration data for each sensor on the string; the manufacturer-specified accuracy is typically in the range from 0.01 °C to 2 °C. Consequently, the best practices for permafrost monitoring (Noetzli et al., 2021) recommend 0 °C bias correction to determine the individual sensor's actual offset from true 0 °C, a recommendation mostly based on experiences from mountain permafrost. As sensors may have nonlinear error dependence on temperature, the accuracy and offsets determined at 0 °C may not be valid across the entire temperature range where phase change occurs. Hence for permafrost soils, multiple-point temperature calibrations have been recommended (Osterkamp, 1984). The most recent World Meteorological Organization (WMO) guide to measurement of cryospheric variables (WMO, 2024) recommends at least three temperature points for calibration around 0 °C and preferably at the extremes of the expected temperature range.

Sensor's accuracy may be further affected by drift-inducing or drift-enhancing factors, such as exposure to temperatures outside of the sensor's specified operating ranges, fast temperature changes (thermal shocks), strain within the sensing element due to mechanical stresses from cycles of thermal expansion and contraction, from pressures from deformation of the borehole casing, or from freezing of any water that may have leaked inside the casing. Factors like these may deteriorate the manufacturer-specified sensor accuracy, stability and drift. Ideally, sensor re-calibration is performed at regular time intervals (recommendations by Noetzli et al. (2021) suggests every 5-10 years) to determine whether drift is occurring. In practice, this is often difficult because of the costs associated with the logistics around retrieving and replacing the sensors. In some cases, the sensor retrieval is not possible, e.g. when borehole walls deform and trap the sensor strings inside the casing, or when water leaks into the casing and freezes the strings in ice. These challenges are common to boreholes both in sedimentary and in rocky, blocky permafrost terrain.

While measurement accuracy is recognized to be important and guidelines for sensor calibration and recalibration exist, researchers rarely report calibration procedures, individual sensor errors, or how those errors change over time when publishing permafrost temperature data or research based on them. The template for delivering ground temperature data to the Global Terrestrial Network for Permafrost (GTN-P) database (GTN-P, 2026) does not require nor mention reporting information on submitted data uncertainty, on calibration practices nor on individual sensor type and characteristics. Some datasets in the database contain information on 'accuracy', but this is reported as a single value which presumably refers to the manufacturer-specified accuracy for the used sensor model in general, not individually determined offset of each sensor used to debias the dataset. Some datasets are reported to be 'quality checked', though it is unclear what this entails, i.e. visual assessment only, filtering and elimination of erroneous measurements, elimination of outliers. It is typically also unclear whether the data have



been corrected for frost heave effects and surface elevation changes due to long-term permafrost degradation, and whether they were debiased, by applying the 0 °C-offsets or offsets at multiple calibration steps.

60 We hypothesize that combination of the multitude of factors affecting sensor performance under a long-term monitoring project (target design lifetime of 2-3 decades, Noetzli et al. 2021) ends up constituting a non-negligible contribution to the uncertainty budget of the ground temperature measurements. Moreover, because of the objective difficulties associated with sensor recalibration, it is desirable to estimate the magnitude of the drift that can be expected in otherwise undisturbed borehole installations.

65 In this study, we report on the stability of calibrated offsets of DS18S20 temperature sensors mounted on Geoprecision sensor strings, and we evaluate whether their performance is adequate for permafrost monitoring applications. We also extensively discuss the physical processes liable for introducing drift errors to this and to other sensor technologies, in support of the argument for systematic reporting of calibration practices and offsets, to document more realistic uncertainty ranges for long-term permafrost temperature records. The permafrost group at the Technical University of Denmark (DTU) (permafrost.dtu.dk)
 70 has instrumented and has been operating 65 (as per 2026) permafrost temperature monitoring installations in Greenland using temperature sensor strings and dataloggers from Geoprecision. This article therefore also serves as a quality documentation and reference to help interpret uncertainties in these datasets.

The distinguishing features of our study are i) the use of high performance temperature calibration equipment (as opposed to ice bath calibration or immersion in snow), ii) evaluation of long-term 0 °C offset drift (after up to 9 years) in both sensors
 75 previously deployed in field monitoring and in sensors stored at stable conditions, and iii) the design of our experiments inspired by real, field scenarios and temperature ranges as opposed to metrology protocols difficult to reliably implement in the field. Although we do quantitatively test only one type of temperature sensor (DS18S20), we argue (in the Sect. 4) that the physical causes of the quantified errors are inherent to other, including newer sensor types and technologies commonly used in permafrost monitoring.

80 The effects we quantified were the following:

1. *Accuracy over a temperature range from –20 °C to +20 °C.* Evaluation of sensor offsets from the true temperature at seven steps of a temperature range realistic for the field installations in coastal areas of Greenland.
2. *Long-term stability of the 0 °C offsets.* Quantification of the drift of the 0 °C sensor offsets first determined in 2016 (29 sensors) and 2019 (14 sensors), and re-calibrated in 2025, for three sensor strings.
- 85 3. *Response to cyclic thermal stress.* Evaluation of the drift of the 0 °C offset based on a stress test consisting of 43 temperature cycles with a center at 0 °C and peak-to-peak amplitude of 40 °C.
4. *Hysteresis of the 0 °C offset.* Investigation of whether the accuracy of the sensor measurements depended on its previous thermal state; potential hysteresis in the sensor response would affect the accuracy of representation of the freeze-thaw processes.



90 This study is the first in a series of three articles focusing on quantification of the uncertainty budget of borehole temperature records in permafrost. Part 2 of the series will evaluate the contribution of systematic errors due to the borehole installation techniques and materials; this includes consistency and inter-comparability of temperature records from boreholes instrumented with the same casing materials, the effect of use of different casing materials (plastic vs. metal) on the rate of heat transfer and the estimated thermal characteristics of the soil. Part 3 will evaluate the combined effect of the previously quantified measurement errors on outcomes of ground thermal models.

2 Methods

We performed the calibration experiments on temperature sensors of the type DS18S20 mounted on strings from the manufacturer Geoprecision. The sensor-logger system uses a two bandgap-generated voltage source based thermometer (a digital temperature sensor with 12-bit resolution) and M-Log5W dataloggers. Different sensor strings were subjected to different experiments, subject to their availability for the tests in the lab as opposed to their deployment in the field.

Table 1. Manufacturer-supplied specifications of the investigated sensor DS18S20

Sensor type	DS18S20
Accuracy	$\pm 0.5^{\circ}\text{C}$ @ -10 to $+85^{\circ}\text{C}$ $\pm 2^{\circ}\text{C}$ @ -55 to $+125^{\circ}\text{C}$
Resolution	0.065°C
Drift	$\pm 0.2^{\circ}\text{C}$
Datalogger	M-Log5W-DALLAS

* Based on a 1000-hour stress test at 125°C with supply voltage = 5.5V.

We conducted comparative calibrations of the field sensors in a Fluke 7320 calibration bath (hereafter *the calibration bath* or *the bath*) filled with the Julabo Thermal G water-glycol based bath fluid. The bath has an operating range of temperatures from -20°C to $+150^{\circ}\text{C}$, and a manufacturer specified stability of 0.005°C and a uniformity of 0.005°C at -20°C and $+25^{\circ}\text{C}$. Hence it is possible to measure multiple sensors in the bath simultaneously, given that the bath liquid is allowed to flow freely around and between the sensors.

A calibrated secondary reference PRT sensor of the type Fluke 5628, S/N 4524, (*the reference sensor*) was installed at the center of the bath, and connected to a Fluke 1586A Super-DAQ Precision Temperature Scanner (*the temperature scanner*) which controlled the data acquisition of the reference sensor.

Calibration of the field sensors in the bath was conducted in a temperature-controlled room at a temperature set to 2°C , to minimize the heat transfer from the surroundings to the reference and field sensors through the wiring connecting them to the respective data loggers. The temperature scanner was located in an adjoining room at a stable room temperature (approx. 21°C). To ensure proper alignment of temperatures measured by the field sensors and the reference sensor in time, the clocks



Table 2. Overview of devices under various tests.

String ID	String alias	No. sensors	History	Tests
A521E0	E0	15	deployed for ca. 6 years	0 °C offset, Long-term drift, Multi-step calibration
A521E4	E4	14	deployed for ca. 6 years	0 °C offset, Long-term drift, Multi-step calibration
A539CC	CC	14	never deployed	0 °C offset, Long-term drift
A5395B	5B	18	never deployed	0 °C offset, Multi-step calibration
A53CEE	EE	11	never deployed	Drift under thermal stress, Hysteresis
A521DE	DE	4	deployed for 5.5 years	Drift under thermal stress, Hysteresis
A521E3	E3	5	deployed for 5 years	Drift under thermal stress, Hysteresis

of the dataloggers measuring the field sensors and the clock of the temperature scanner measuring the reference sensor were synchronized.

115 2.1 Reference sensor temperature validation using the triple-point-of-water

The reference sensor was supplied with a NIST (National Institute of Standards and Technology) traceable calibration, and as part of our quality control procedures, the sensor is traceably re-calibrated at intervals. Between formal re-calibrations, we perform regular reference sensor validations using a triple-point-of-water cell of the type Fluke 5901B-G (*the TPW cell*) with a Fluke 9210 maintenance apparatus (*the TPW maintenance apparatus*).

120 The triple-point-of-water (TPW) is a fundamental thermodynamic reference point where pure, oxygen-free water reaches an equilibrium state between three phases – solid, liquid, and vapor – at 0.01 °C, thus providing a highly reproducible calibration gold standard. Preparation of the TPW cell consists of filling the inner chamber with ethanol and supercooling the water in the cell to −5 °C using the TPW maintenance apparatus. Freezing of the cell is initiated by shaking. The supercooling and freezing is followed by a controlled thawing phase that ensures a stable temperature in the TPW cell at 0.01 °C over a period
 125 of several hours. Validation of the reference sensor temperature is performed by inserting it in the prepared TPW cell and measure the temperature during 30 minutes at 10 second intervals. During the validation experiment, the reference sensor, TPW cell and maintenance apparatus are placed in a temperature controlled room at 2 °C, to minimize the heat transfer from the surroundings to the reference sensor under test. The offset of the reference sensor ΔT^{ref} is found as the mean difference between the reference temperature $T_{TPW\ cell}$ and the sensor's recorded temperatures T_i^{ref} :

$$130 \quad \Delta T^{ref} = \frac{1}{N} \sum_{i=1}^N (T_{TPW\ cell} - T_i^{ref}) \quad (1)$$

where $T_{TPW\ cell} = 0.010$ °C inside the TPW cell.

Using this procedure, we have documented a total system drift of the reference sensor and temperature scanner of <0.001 °C/year at 0.01 °C.



2.2 Multiple-step offset calibration

135 The field sensors' accuracy was evaluated by conducting multiple-step traceable measurements at temperatures within a range commonly encountered in the field: -20°C , -10°C , -5°C , 0°C , $+5^{\circ}\text{C}$, $+10^{\circ}\text{C}$, $+20^{\circ}\text{C}$. The string-sensor system was placed in the calibration bath, with the reference sensor immersed in the centre of the bath and the string sensors placed around it. After allowing the bath temperature to stabilize, the acquisition period lasted 30 min at each temperature step, during which the reference sensor was recording every 10 s, to capture any potential instabilities in the bath temperature, and the string was
140 recording every 30 s. After checking the reference sensor records, these were down-sampled to 30 s records, to match the string record frequency.

When comparing the bath temperature measured using the sensor string system with the temperature measured by the reference sensor, the calibration offset ΔT for each sensor on the string at each temperature step were found as:

$$\Delta T = \frac{1}{N} \sum_{i=1}^N (T_i^{ref} + \Delta T^{ref} - T_i) \quad (2)$$

145 where T_i^{ref} is the temperature measured by the reference sensor, and T_i is the temperature measured by the sensor under calibration. The calibration offset ΔT is the bias correction constant for the sensor under calibration. This correction may be applied to the field data as to obtain a corrected temperature reading:

$$T^{corr} = T^{field} + \Delta T \quad (3)$$

where T^{field} is the temperature measured by the deployed sensor, and T^{corr} is the corrected temperature reading of the
150 measured system.

2.3 Long-term stability of the sensor offsets

The long term stability of the field sensors was evaluated by taking new traceable measurements (in spring 2025) and comparing them with the historical measurements.

For the strings E0 and E4, the 0°C -offsets were first measured in spring 2016, when the strings were calibrated prior to field
155 deployment in Greenland (Ingeman-Nielsen, 2016). The strings were installed in boreholes in sedimentary permafrost, north of the town Ilulissat (69.2198°N , 51.0986°W), in summer 2016 as part of site investigations for construction of a new airport. The strings remained in the boreholes (the inner casings of $\varnothing 32\text{ mm}$ PE tubing, not filled with liquid, the outer protective casing of metal) for approximately 6 years, experiencing seasonal temperature variations between -24°C to 18°C (uppermost sensor). As the strings could be removed easily from the boreholes, we concluded that they had not been submerged or exposed to ice
160 pressures during the deployment. The monitoring ended in 2022 when the strings were retrieved from the boreholes and were since stored at a stable room temperature of around 21°C until spring 2025 when they were re-calibrated.

In 2016, the 0°C -offsets were determined using equipment different from the 2025 calibration setup. In 2016, we used a Julabo F32-ME cooling bath (stability $\pm 0.01^{\circ}\text{C}$) and a Fluke 1524 reference thermometer with 5627A RTD sensor, manufacturer-



calibrated to an accuracy better than $\pm 0.046^\circ\text{C}$ at 0°C . This system had a resolution of 0.001°C . To assess the reproducibility
165 of the calibrations, two to three independent repeated calibrations of at least two sensors on each string were performed. The
absolute difference between repeated calibrations for the same sensor was on average $0.02 \pm 0.02^\circ\text{C}$ (1σ , assuming a semi-
normal distribution), with a maximum observed difference of 0.07°C .

For the string CC, the 0°C calibration offsets were first measured in spring 2019 using the same calibration setup as for the
2025 calibrations. The string was never deployed in field monitoring and has been stored in our laboratory, at around 21°C
170 room temperature, for 6 years.

The 0°C offsets from 2016 and from 2019 were compared to the offsets determined in 2025, nine and six years later,
respectively. We calculated the rate of change of the 0°C -offset per year, per each sensor.

2.4 Drift due to thermal stress

Fast temperature changes can induce drift in sensor accuracy in excess of the drift stated in the specifications (Table 1). We
subjected 20 sensors of the strings EE, DE and E3 to a stress test consisting of 43 pre-programmed temperature cycles with
175 the following repeating pattern: $-20^\circ\text{C} \rightarrow 0^\circ\text{C} \rightarrow +20^\circ\text{C} \rightarrow 0^\circ\text{C}$. Each temperature setpoint was kept for 2 hours, which
allowed the bath temperature to stabilize. The temperatures of the strings were logged every 60 seconds, and the reference
sensor was logged every 10 seconds. The potential drift of the offset was evaluated by comparing the temperature readings of
the strings at 0°C to the reference sensor readings at 0°C . For the reference sensor and for each sensor on a string, a fixed 25
180 minutes segment of stable measurements at 0°C was used to calculate a mean temperature reading. The data were separated
into heating cycles (following -20°C) and cooling cycles (following 20°C) to account for potential hysteresis effects. The
error (ε) between a string sensor and the reference sensor was calculated as the difference in means of each segment of the
cooling and heating cycles respectively:

$$\varepsilon_i = \frac{1}{N_1} \sum_{j=1}^{N_1} T^{ref} - \frac{1}{N_2} \sum_{k=1}^{N_2} T^{sensor} \quad (4)$$

185 where i is the index of the analyzed segment, N_1 is the number of measurements in the analyzed segment for the reference
sensor, N_2 is the number of measurements in the analyzed segment for the sensor in test. We performed Mann-Kendall trend
tests ($\alpha = 0.05$) on the sensor measurement errors (ε) to detect monotonic trends that indicate systematic sensor drift. The test
provides a judgement of whether a trend is present, and the magnitude of the slope over the timeseries.

2.5 Hysteresis of sensor response

190 Hysteresis was evaluated based on data collected in the test of response to cyclic thermal stress described in the previous Sect.
2.4, by comparing the errors between the field sensors and the reference sensor reading at 0°C , from the cooling and heating
segments respectively. A one-sample t-test was performed for each sensor to test the null hypothesis that the mean change in
error equals zero, which would mean that there is no systematic hysteresis effect. The test was conducted across all temperature
cycles, using a significance level of $\alpha = 0.05$.



195 3 Results

3.1 Repeatability of the calibration setup and stability of the calibration bath

We evaluated the repeatability of our TPW setup for 0 °C validation of the reference sensor, based on a four year measurement record. We conclude that our 0 °C-offset calibration setup is very reliable, with good short-term stability (3 identical measurements within one week) and inter-operator consistency (4 different operators across the four years determined consistent 0 °C-offsets of the same reference sensor). Over the four years of use, the reference sensor showed a very small drift corresponding to an average of 1 mK per 2 years.

The stability of the calibration bath was confirmed to be well within manufacturer's specifications, with standard deviation of the measurements taken by the reference sensor during the calibration steps typically below 0.004 °C.

3.2 Long-term drift of the 0 °C offsets

205 Out of the 43 sensors tested across the three strings (E0, E4, CC), all exhibited drift of the 0 °C-offset (ΔT) (Fig. 1). 41 (95.3%) sensors showed positive drift, thus introducing warm bias to measurements, if left undetected and uncorrected. 2 sensors can be evaluated as stable, with a negative drift so small (-0.0001 °C (string E0) to -0.0006 °C (string E4) that it was near the accuracy level of our calibration setup. The maximum rate of drift of the 0 °C-offset was 0.0157 °C/year of positive drift (string E4); the minimum positive drift was 0.0014 °C/year (string E4); the average positive drift across the three strings was 0.0064 °C/year. These rates of sensor drift would cause a warm bias in the data in the range of 0.045 – 0.471 °C (min - max) in 30 years, assuming the drift was linear in that time frame.

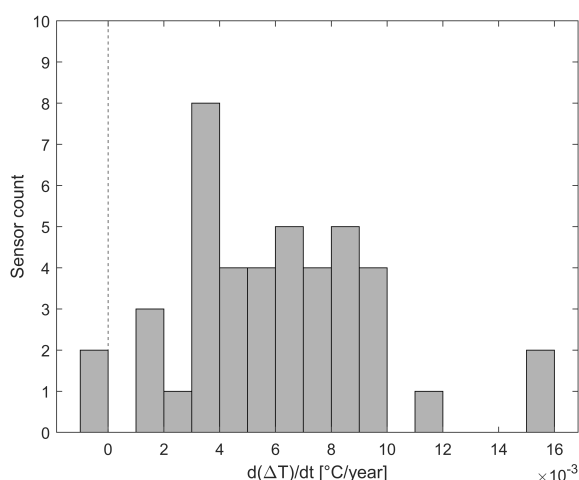


Figure 1. Long-term drift of the 0 °C-offsets (ΔT) of the individual sensors of the sensor strings E0, E4 (both measured in 2016 and in 2025) and CC (measured in 2019 and in 2025). The drift is expressed as the rate of change of the ΔT in °C per year. Vertical dashed line at 0 °C separates negatively and positively drifting offsets.



3.3 Offsets in the temperature range -20°C to $+20^{\circ}\text{C}$

We found that 4 out of the 47 sensors subjected to the multi-step calibrations did not live up to the manufacturer's specifications on accuracy (Table 1). The absolute offsets of the four sensors at -10°C were larger than the stated 0.5°C (Fig. 3.3).

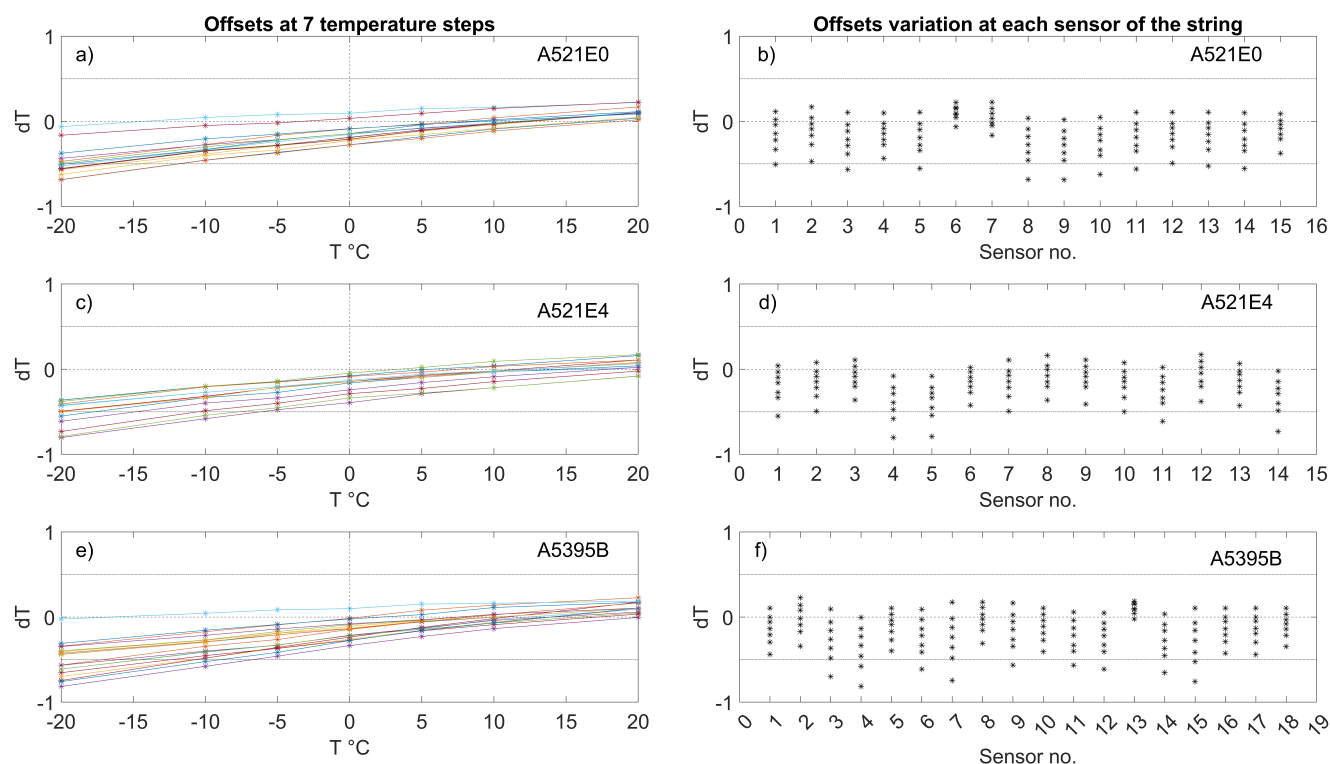


Figure 2. Multiple-step calibration for the sensor strings A521E0, A521E4 and A5395B.

215 The absolute offsets were larger at temperatures below 0°C than at temperatures above (Fig. 3.3). This suggests that the multi-step calibration is particularly relevant for determining the sensor offsets at temperatures below 0°C , which have the most implications for determination of soil freezing characteristics, especially of fine-grained soils and saline soils.

3.4 Drift under thermal stress

For the three strings EE, DE, and E3 subjected to 43 cycles of thermal stress, the Mann-Kendall test determined that 16 out of 20 sensors showed a statistically significant trend ($p < 0.05$) in the temperature readings compared to the reference sensor in the heating cycle, and 16 (different) sensors out of 20 sensors showed a trend in the cooling cycle. The drifts found were both negative and positive. The largest magnitude of the slope were $-0.000375^{\circ}\text{C}$ per cycle. Assuming that the thermal cycles simulated sensor wear throughout yearly freeze-thaw cycles, 30 such cycles (30 years of summer and winter extreme temperatures) could contribute up to -0.011°C drift in either cooling or warming direction.



225 3.5 Hysteresis

We performed a one-sample t-test on the thermal cycling data to evaluate whether the temperature sensors readings were affected by hysteresis. The test showed a hysteresis in some of the sensors. Across the three strings (EE, DE, and E3), 7 out of 20 sensors showed a significant change ($\alpha = 0.05$) in the error due to hysteresis. The difference in error between the cold and warm plateaus was maximum 0.009°C .

230 The confirmed good stability and repeatability of the bath fluid temperature (Sect. 3.1) confirms that the observed apparent hysteresis is not related to bath performance.

4 Discussion

4.1 Drift-inducing and drift-enhancing factors in different thermal sensor technologies

The DS18S20 sensors in the Geoprecision probes used in this study derive their temperature reading from the difference
 235 between two on-chip voltage sources with different temperature coefficients, based on the well-understood silicon bandgap principle. A comprehensive modern treatment is available in Pertijs and Huijsing (2006), and the same principle was already used by Hilbiber (1964) to obtain a very stable voltage reference with very low temperature sensitivity. The exact analytical expression derived by Widlar (1967) showed that the base-emitter voltage is nearly – but not exactly – linear with absolute temperature, which makes linearization techniques necessary. Verster (1968) demonstrated how to implement a very linear
 240 thermometer, and Timko (1976) described the design of the commercially available AD590, which achieved $\pm 0.5^{\circ}\text{C}$ at 25°C calibrated accuracy, $\pm 0.3^{\circ}\text{C}$ linearity from -55 to $+150^{\circ}\text{C}$, and is still in production. In the DS18S20 sensor, the base-emitter voltage difference between two bipolar transistors operated at different current densities is digitized by an on-chip analog-to-digital converter (ADC), mapped to a temperature value in $^{\circ}\text{C}$, and made available to be read out in digital form. This architecture eliminates several error sources present in designs using discrete analog temperature sensors and analog-to-digital
 245 converters (discussed below). However, both static errors and time-dependent drift still occur.

While the details of the DS18S20 linearization and factory calibration procedure are not publicly available, single-temperature set point calibration is standard practice for this class of digital temperature sensors to correct the per-die offset of the linearized temperature output, keeping the slope fixed to a nominal value dependent on the semiconductor process (Pertijs and Huijsing, 2006). Any die-to-die variation of the slope term remains uncorrected and results in a temperature-dependent error that
 250 grows with distance from the factory calibration temperature. This is consistent with the per-sensor variation in offset-versus-temperature slopes observed in our multi-step calibrations, and with the systematically larger offsets at negative temperatures, which lay furthest from the likely factory trim temperature. Additionally, any static linearization error adds another systematic effect. Multi-step characterization of each individual device allows to quantify the combined error and, through curve fitting, it may allow to correct for it (Maxim Application Note 208). Left uncorrected, these systematic errors are also consistent with
 255 the pattern of broadening temperature uncertainty towards very low and very high temperatures commonly observed in the sensors accuracy specifications.



Thermo-mechanical stress on the silicon die has been identified as a major mechanism introducing long-term drift in bandgap-based temperature sensors (Fruett et al., 2003), and drift patterns observed in this study are consistent with this mechanism. Internal stresses from differential thermal expansion between the silicon die, the die-attach adhesive, and the epoxy molding compound shift the base-emitter voltage through the piezjunction effect. This produces both an initial offset at the time of manufacturing of the sensor, and a progressive drift as the materials undergoes mechanical relaxation and aging (Fruett et al., 2003). This mechanism operates even at constant room temperature, which may explain the significant drift we observed in the sensor strings 5B and CC, which have been stored in the lab and never deployed. In the field, such drift would superimpose a systematic long-term trend to the actual climatic signal, and could only be detected and prevented (or corrected for) through periodic sensors re-calibrations.

Mechanical loads can also be transmitted to the die through solder joints and the printed circuit board on which the sensor is mounted. Borehole casing deformation, ice pressures on a temperature probe, differential thermal contraction of the probe assembly, as well as gradual relaxation of the stress in the printed circuit board from differential thermal expansion across soldering, storage and operating temperatures can all contribute to this external stresses. Thermal cycling may further induce cumulative changes in the die stress, potentially producing both cycling-induced drift and hysteresis. The DS18S20 sensor datasheet specifies a 0.2°C typical drift after a 1000-hour stress test at 125°C . At room or permafrost-site temperatures, an equivalent degree of stress relaxation may take orders-of-magnitude longer time to achieve.

Other sources of error, such as sensitivity to supply self-heating, may be present; these depend on the operating parameters configured by the user. The ca. 5 mW dissipated by the DS18S20 sensor during a single conversion period of 750 ms are unlikely to produce a warm bias exceeding the manufacturer specified accuracy at the temperature sampling rates typical of permafrost monitoring. However, using higher temperature sampling rates, e.g. during calibration in the bath, may introduce a non-negligible bias, especially if the aim of the calibration is to characterize each individual sensor to a tighter uncertainty than the manufacturer's 0.5°C specification. The exact materials and construction details of both the DS18S20 sensor and the assembled Geoprecision probes are not known in sufficient detail to reliably quantify self-heating under all temperature sampling rates. In the case of continuous sampling, the constant ca. 5 mW dissipation could introduce a worst-case warm bias reaching steady-state magnitude of up to 0.6°C after some minutes, depending on which DS18S20 package version is used by Geoprecision. This worst-case estimate is based on assumption that the Geoprecision probe contains a DS18S20 sensor in the TO-92 package option (from size considerations), with junction-to-ambient thermal resistance of 180°C/W (from typical values for similar packages), and still air surrounding the sensor. In practice, this thermal resistance and consequently the temperature rise may be significantly reduced by the pins soldered to the printed circuit board, any potting compound filling the probe, or by the direct contact between the DS18S20 and the metal can of the probe due to the power dissipation in the probe. Estimating the worst case self-heating due to a single reading of the sensor – a scenario closer to a typical field operation – is even more uncertain because temperature transient is controlled not only by the unspecified thermal resistance of the package, but also by the thermal time constants of the die, package, board, potting and probe, none of which are known. Because the ADC in the DS18S20 sensor is of sigma-delta type, the effective self-heating error is probably close to the average of the temperature ramp, so roughly half of the self-heating temperature peak. This temperature peak will be smaller than the



steady state estimated above, but it is difficult to make a more exact calculation. Empirical evidence suggests that at sampling frequency of 1 Hz, a DS18B20 configured for 12 bit resolution (matching the DS18S20 operation) may result in ca. 0.06 °C self-heating (Petrich et al., 2020).

295 The DS18S20 is one out of several iterations in the Dallas/Maxim/Analog Devices 1-Wire temperature sensor family. The original DS1820 used a dual-oscillator (slope-counter) architecture for temperature-to-digital conversion, providing a resolution of 9 bits output which could be extended to 12 bits by reading out some additional data and carrying out a simple calculation; the DS18S20 reimplemented this with a 12-bit sigma-delta ADC and truncating the result to 9-bit format, to ensure drop-in backward compatibility with the DS1820, and internally synthesises the additional bytes needed for emulating a
 300 DS1820 for extended-resolution calculation; the DS18B20 uses the same design of DS18S20 and lets the user configure the bit size of the output (Analog Devices Application Note 4377). The typical specification charts of temperature error vs. temperature in the datasheets of these three parts show that the temperature non-linearity is the same between DS18S20 and DS18B20, but radically different between the former and the earlier DS1820 for which it was designed as a drop-in replacement. For some higher-accuracy devices like the Texas Instruments TMP117, the datasheet shows that the residual non-linearity of the
 305 temperature error over temperature follows a different curve for each of the two package options offered, even though both remain within the same stated accuracy limits. This illustrates how the different behavior of very similar devices with the same datasheet specifications may become detectable when calibrated to tighter limits than their original specifications, requiring individual multi-point characterization of each device.

Finally, for some sensors like the DS18B20, the prevalence of counterfeit devices may be very high outside of the distribution
 310 channels authorized by the manufacturers; testing of over 1000 units from more than 70 vendors revealed at least seven distinct clone families with variable accuracy and undocumented drift characteristics (Petrich et al., 2020).

Our study focused on a single sensor type whose semiconductor-based technology is inherently susceptible to drift over time, even when not deployed. One of the strengths of our study is that we quantified drift in both previously deployed sensors and sensors stored under stable environmental conditions, thereby isolating drift from other deployment-related factors. This
 315 type of drift can only be corrected through repeated calibration. Because the digital sensor design allows greater confidence in identifying the source of drift, our findings provide a clear baseline for assessing sensor stability.

Other commonly used temperature sensor technologies are used in permafrost monitoring. These typically use discrete-component designs in which a negative temperature coefficient (NTC) thermistor, platinum resistance temperature detector (RTD), or thermocouple is read by an external ADC through an excitation circuit which typically include precision voltage,
 320 current or resistance references. Each component in this chain contributes to the total measurement uncertainty, and each may drift and respond to thermal and mechanical stress. NTC thermistors are ceramic semiconductors widely used because of their high sensitivity and moderate cost. Their strongly nonlinear resistance-temperature characteristics require correction through an empirical fit whose coefficients may need recalibration as the sensor ages and drifts (Steinhart and Hart, 1968). While the exact physical processes leading to long-term drift are still being investigated, the trend is towards increasing resistance over
 325 time, which would introduce a cooling bias into long-term temperature timeseries, if left uncorrected (Feteira, 2009). Platinum RTDs have very good long-term stability under controlled conditions, but they are expensive, have low sensitivity, and under



field conditions the wire-wound elements tends to be susceptible to strain, mechanical shock and contamination; meanwhile the thin-film RTDs face tradeoffs in terms of long-term stability, response time and vibration resistance. Thermocouples, while cheap, have limited accuracy and sensitivity. This highlights the need to repeat similar investigations across these sensor types.

330 Discussing extensive details of each of these temperature sensor types is beyond the scope of this paper. Our key message is that when characterizing temperature sensors and complete probes, it is important to keep in mind the effects of the numerous physical factors influencing their accuracy and long-term stability. Over time, semiconductor production lines undergo periodic die revisions and changes in packaging materials, any of which can subtly shift the population statistics of sensor accuracy, linearity, drift and sensitivity to mechanical stress, often without being reflected in the datasheet specifications or in
 335 an updated part number. Similarly, discrete-component probe designs are subject to bill-of-materials changes, parts substitution, differences between batches, firmware and calibration procedures, that can subtly alter the system's accuracy and drift characteristics between production lots without requiring any change in the product's model designation. These facts support the argument for repeated calibration of individual sensors designated for long-term monitoring projects.

4.2 Implications for calibration practices in permafrost monitoring

340 The recommendation to calibrate permafrost temperature sensors at 0 °C (Noetzli et al., 2021) is largely based on experience from mountain environments, where permafrost occurs in bedrock and coarse, blocky materials. In these settings, most pore water does indeed freeze at 0 °C because of low ion concentrations and relatively small fraction of water adsorbed onto mineral surfaces. Sensor offsets at 0 °C can theoretically be estimated from spring and autumn zero-curtain periods. While this approach may be viable for sensors installed in the active layer, below the active layer, distinguishing early sensor drift
 345 from naturally occurring warming or cooling trend becomes difficult without recalibration or duplicate measurements (Luethi and Phillips, 2016). Moreover, even in rock, freezing-point depression can occur due to solutes, pressure, pore size, and pore-surface effects, resulting in mobile supercooled unfrozen water present at temperatures as low as −10 °C (Mellor, 1970; Lock, 1990). We therefore support the recommendation to apply at least the 0 °C-offset correction to every sensor, including those used in long-term rock permafrost monitoring.

350 The limitations of single-point calibration become more pronounced in soils, particularly fine-grained soils. Here, the freezing temperature of pore water is commonly depressed below 0 °C because of dissolved ions and the larger proportion of adsorbed water. As a result, even small temperature errors can significantly affect estimates of latent heat exchange during freeze–thaw processes. Furthermore, phase transitions in fine-grained soils occur gradually over a broader temperature range than in rock and coarse-grained material (e.g., Lovell Jr, 1957; Romanovsky et al., 2000; Tomašková and Ingeman-Nielsen,
 355 2024). The nonlinear temperature dependence of sensor errors observed in our multi-step calibration demonstrates that an offset determined at 0 °C cannot be assumed valid across the full phase-change range. Consequently, the multi-point calibration approaches advocated by Osterkamp (1984) and the WMO (2024) are particularly relevant for monitoring of sedimentary permafrost.

Studies have argued that analyses of permafrost temperature trends are less sensitive to calibration errors than absolute
 360 temperature measurements (Noetzli et al., 2021; Biskaborn et al., 2019). However, without initial calibration and periodic re-



calibrations, sensor drift cannot be quantified. Our results show that, over the expected service life of a long-term monitoring installations (20 – 30 years, Noetzli et al., 2021), sensor drift can reach 0.314 °C to 0.471 °C (in 20 and in 30 years, respectively) for the particular sensors and circumstances we tested. This is assuming that the sensor drift is linear over the period of deployment. The drift behavior, while presumably slowing over time, may still be expected to be in the high rate part of the drift curve at the deployment temperatures, which are much lower than the stress-test temperature given by the manufacturer (Table 1), according to the temperature dependence of reaction rates (Arrhenius equation, Atkins et al. 2023). Such rates of drift would be large enough to account for the permafrost warming reported in recent studies (Etzelmüller et al., 2020; Biskaborn et al., 2019), and that is without accounting for other error sources contributing to sensor drift and the total measurement uncertainty.

A further challenge for geophysical applications is that records of the exact sensor type used in deployments, and whether re-calibrations were performed, are often incomplete or unavailable in permafrost temperature databases. This makes it less straightforward to apply realistic uncertainty ranges to the reported datasets. Our findings highlight the importance of reporting detailed sensor metadata (sensor type, calibration procedure incl. used calibration equipment, re-calibration, individual sensor offset and drift characteristics) alongside permafrost temperature datasets.

The numerous additional contributors to the uncertainty budget of permafrost temperature measurements were beyond the scope of this study. These include, for example, the influence of moisture on electronic components (highlighted e.g. by Luethi and Phillips, 2016), variations in logger enclosure temperature, fluctuations in battery voltage, and other instrumental and environmental factors. We speculate that some of the largest sources of uncertainty in ground thermal measurements arise from the physical characteristics of borehole installations themselves. These include drilling methods, borehole diameter, backfill material, casing material, and the accuracy with which sensors are positioned at their intended depths. In addition, geomorphic and cryogenic processes, such as seasonal frost heave and thaw settlement, and long-term permafrost degradation accompanied by surface elevation changes, may alter sensor positions relative to the ground surface. Such changes could introduce apparent temperature trends that might otherwise be interpreted as evidence of permafrost warming or cooling. The effects of these uncertainties will be reported as Part 2 and 3 of this article series.

5 Conclusions

We reviewed and quantified some of the systematic errors that affect field measurement systems for permafrost temperature monitoring, and discussed their importance for short- and long-term monitoring applications. In the process, we also presented a suite of (re-)calibration procedures suitable for thermal sensors intended for long-term ground temperature monitoring in periglacial environments. Our key findings can be summarized as follows:

1. We confirmed and quantified long-term drift in the 0 °C offsets of the tested digital sensors DS18S20. Drift rates reached 0.0157 °C per year, with 95.3% of sensors drifting toward warmer temperatures. Extrapolated over 30 years, the observed drift (0.471 °C) would be comparable to the magnitude of permafrost warming reported in recent studies, and could therefore introduce substantial bias into permafrost thermal state evaluations, if left uncorrected. Other sensor tech-



nologies contain additional components and mechanisms contributing to the drift, making it difficult to apply realistic uncertainty ranges to the existing datasets. We therefore argue for the importance of reporting detailed sensor metadata including at least the sensor type, and ideally the individual sensor offsets, drift characteristics, (re-)calibration procedure and equipment, re-calibration,) alongside long-term permafrost temperature datasets.

2. We confirmed that sensor offsets from true temperatures are non-linear across the temperature range typical of periglacial environments (-20°C to 20°C). Although all tested sensors met the manufacturer-specified accuracy between -5°C and 20°C , 8.5% sensors failed to meet this specification at -10°C . Offset magnitudes were generally larger at negative than at positive temperatures, indicating that sub-zero measurements are more susceptible to systematic error when sensor- and temperature-specific corrections are not applied. Such errors affect not only the accuracy of short- and long-term monitoring records but also calculations of phase-change energy balance and soil freezing characteristics, particularly in fine-grained and saline soils. We therefore recommend extending the current practice of 0°C bias correction to include calibration and correction at multiple sub-zero temperatures, which would benefit both short- and long-term applications.
3. We quantified sensor drift after 43 cycles of thermal stress between -20°C and 20°C in a majority of the sensors. The magnitude of the drift seen in our experiments motivates a recommendation to perform regular re-calibration of field sensors that experience thermal cycling.
4. We quantified hysteresis of the sensors. The recommendation is that it is taken into account in the design of the experiment if temperature cycles are present in the application.
5. Based on the evidence presented in this study, we argue that the individual calibration of each temperature sensor or probe, both before and after their field deployment, should always be attempted as an essential task of the monitoring program. This is due to the proven existence of known physical processes liable for introducing long-term instrumental drift errors in sensors of various technologies. The magnitude of these errors alone is comparable to the slow and small geophysical signals associated with permafrost warming and thawing. We therefore strongly recommend incorporating periodic sensor recalibration into the best practices for long-term permafrost monitoring. The realities of field deployments do on occasion make it difficult or impossible to retrieve and recalibrate sensor strings from the deformed or otherwise blocked boreholes. In settings where sensor retrieval is impossible due to borehole deformation or casing failure, maintaining data quality may require installing newly calibrated sensor strings in replacement boreholes approximately every five years. While such measures may be unnecessary for short-term studies focused primarily on process understanding, they are essential for generating long-term temperature records with quantifiable and realistic uncertainty ranges. Where neither re-calibration nor replacement boreholes are feasible, the most appropriate course of action would be to document broader uncertainty ranges for the produced temperature timeseries than the sensors and probes manufacturer specifications may suggest. That is at least until extensive experience is gathered on the exact type of sensor that confirms long-term stability under actual field conditions.



Data availability. Data used in this study are available through the following link: <https://figshare.com/s/412a58427e877835ae1e>. This link will be updated to a DOI during the review process.

Author contributions. Soňa Tomašková: Conceptualization, Methodology, Formal Analysis, Investigation, Visualization, Writing - Original Draft, Writing - Review & Editing, Funding Acquisition.
Fredrik Johansson: Conceptualization, Methodology, Formal Analysis, Investigation, Writing - Original Draft, Writing - Review & Editing.
Michele Citterio: Conceptualization, Methodology, Writing - Original Draft, Writing - Review & Editing.
Thomas Ingeman-Nielsen: Conceptualization, Methodology, Validation, Data Curation, Resources, Writing - Original draft, Writing - Review & Editing, Funding acquisition.

Competing interests. The authors declare no competing interests.

Acknowledgements. This study was part of project TARAJULIK, funded by the Greenland Research Council, grant agreement 80.30. The study was also supported by the ILLUQ project, funded by the European Union Horizon Europe research and innovation programme (project no. 101133587), the Greenland Integrated Observing System (GIOS) funded by the Danish Ministry of Higher Education and Science (agreement no. 5229-00005b), and the COLDHUB project, funded by NordForsk (project no. 97338).



References

- Analog Devices Application Note 4377: Comparison of the DS18B20 and DS18S20 1-Wire Digital Thermometers, Application Note AN4377, March 2009. <https://www.analog.com/en/resources/design-notes/1wire-protocol-pdf-of-ds18s20-vs-ds18b20-digital-thermometers.html>. Retrieved on the 15. June 2026.
- 445 Atkins, P. W., De Paula, J., and Keeler, J.: Atkins' physical chemistry, Oxford university press, 2023.
- Biskaborn, B. K., Smith, S. L., Noetzli, J., Matthes, H., Vieira, G., Streletskiy, D. A., Schoeneich, P., Romanovsky, V. E., Lewkowicz, A. G., Abramov, A., et al.: Permafrost is warming at a global scale, *Nature communications*, 10, 264, 2019.
- Etzelmüller, B., Guglielmin, M., Hauck, C., Hilbich, C., Hoelzle, M., Isaksen, K., Noetzli, J., Oliva, M., and Ramos, M.: Twenty years of European mountain permafrost dynamics—the PACE legacy, *Environmental Research Letters*, 15, 104 070, 2020.
- 450 Feteira, A.: Negative temperature coefficient resistance (NTCR) ceramic thermistors: an industrial perspective, *Journal of the American Ceramic Society*, 92, 967–983, 2009.
- Fruett, F., Meijer, G. C., and Bakker, A.: Minimization of the mechanical-stress-induced inaccuracy in bandgap voltage references, *IEEE Journal of Solid-State Circuits*, 38, 1288–1291, 2003.
- GTN-P, 2026: Global Terrestrial Network for Permafrost - Tutorials. <https://gtnp.arcticportal.org/index.php/component/content/article/15-database/6-tutorials>. Retrieved on the 12. June 2026.
- 455 Hilbiber, D.: A new semiconductor voltage standard, in: 1964 IEEE international solid-state circuits conference. Digest of technical papers, vol. 7, pp. 32–33, IEEE, 1964.
- Ingeman-Nielsen, T.: Kalibrering af termistorstrenge til boringer i Ilulissat, Technical University of Denmark, Department of Civil Engineering, 2016.
- 460 Lock, G. S. H.: The growth and decay of ice, Cambridge University Press, 1990.
- Lovell Jr, C.: Temperature effects on phase composition and strength of partially-frozen soil, *Highway Research Board Bulletin*, 1957.
- Luethi, R. and Phillips, M.: Challenges and solutions for long-term permafrost borehole temperature monitoring and data interpretation, *Geographica Helvetica*, 71, 121–131, 2016.
- Maxim Application Note 208: Maxim Application Note 208 - Curve Fitting the Error of a Bandgap-Based Digital Temperature Sensor. <https://www.mrsoft.fi/dl/Curve>
- 465 Mellor, M.: Phase composition of pore water in cold rocks, 292, Corps of Engineers, US Army, Cold Regions Research and Engineering Laboratory, 1970.
- Noetzli, J., Arenson, L. U., Bast, A., Beutel, J., Delaloye, R., Farinotti, D., Gruber, S., Gubler, H., Haeberli, W., Hasler, A., et al.: Best practice for measuring permafrost temperature in boreholes based on the experience in the Swiss Alps, *Frontiers in Earth Science*, 9, 607 875, 2021.
- 470 Osterkamp, T.: Temperature measurements in permafrost, Tech. Rep. FHWA-AK-RD-85-11, State of Alaska Department of Transportation and Public Facilities, Division of Planning and Programming, Research Section, 1984.
- Pertijs, M. A. and Huijsing, J.: Precision temperature sensors in CMOS technology, Springer Science & Business Media, 2006.
- Petrich, C., Sæther, I., Dang, N. P., Kleven, , and O'Sadnick, M.: A Note on Remote Temperature Measurements with Ds18B20 Digital Sensor, 25th IAHR International Symposium on Ice, Trondheim, 23 -25 November 2020, 2020.
- 475 Romanovsky, V., Osterkamp, T., et al.: Effects of unfrozen water on heat and mass transport processes in the active layer and permafrost, *Permafrost and Periglacial Processes*, 11, 219–239, 2000.



- Steinhart, J. S. and Hart, S. R.: Calibration curves for thermistors, in: Deep sea research and oceanographic abstracts, vol. 15, pp. 497–503, Elsevier, 1968.
- 480 Timko, M. P.: A two-terminal IC temperature transducer, IEEE Journal of Solid-State Circuits, 11, 784–788, 1976.
- Tomaškovičová, S. and Ingeman-Nielsen, T.: Quantification of freeze–thaw hysteresis of unfrozen water content and electrical resistivity from time lapse measurements in the active layer and permafrost, Permafrost and Periglacial Processes, 35, 79–97, 2024.
- Verster, T.: p–n junction as an ultralinear calculable thermometer, Electronics letters, 4, 175–176, 1968.
- Widlar, R. J.: An exact expression for the thermal variation of the emitter base voltage of bi-polar transistors, Proceedings of the IEEE, 55, 485 96–97, 1967.
- WMO: Guide to meteorological instruments and methods of observation, Volume II - Measurement of Cryospheric Variables, Tech. rep., World Meteorological Organisation, 2024.