Article

COAT Project: Intercomparison of Thermometer Radiation Shields in the Arctic

Carmen García Izquierdo [1,*], Sonia Hernandez [1], Marina Parrondo [1], Alberto Casas [1], Angelo Viola [2,†], Mauro Mazzola [2], Andrea Merlone [2,3] and Yves-Alain Roulet [4]

1 Centro Español de Metrología, 28760 Tres Cantos, Madrid, Spain
2 National Research Council, Institute of Polar Sciences, 00185 Rome, Italy
3 Istituto Nazionale di Ricerca Metrologica, 10135 Torino, Italy
4 Federal Office of Meteorology and Climatology MeteoSwiss, 3003 Bern, Switzerland
* Correspondence: mcgarciaiz@cem.es
† Deceased author.

Abstract: A metrological field intercomparison of thermometer radiation shields in the Arctic was conducted with the aim of obtaining information to increase the worldwide comparability of air temperature measurements. Air temperature measurements are performed by different combinations of thermometers and shields. The response of each system (thermometer + shield) to local meteorological conditions depends on the system itself, limiting the comparability of air temperature measurements. Ten different models of radiation shields were included in the intercomparison, involving two campaigns: (1) the laboratory campaign, where all the instrumentation was calibrated just before and just after the field campaign, and (2) the field campaign that lasted 14 months where 41 thermometers were sampled every 2 min. All the delivered data were subjected to quality control to assure the robustness of the conclusions. A reference shield was defined, and the other shields were compared to the reference one for the conditions where maximum divergences were expected, solar irradiance being the highest impact factor. A maximum divergence value of 1.29 °C was derived for one of the shields and, for all the shields, the difference from the reference one decreases with wind speed. Finally, the uncertainties associated with the shields intercomparison were calculated.

Keywords: air temperature measurements; thermometer radiation shields; arctic; metrology for climate; uncertainties in air temperature measurements; intercomparison of instruments in the field

1. Introduction

The Standing Committee on Measurements, Instrumentation and Traceability (SC-MINT) in the Infrastructure Commission (INFCOM) of World Meteorological Organization (WMO) includes in its terms of reference the duty of coordinating and conducting intercomparisons in cooperation with relevant networks as appropriate [1]. WMO recognizes field-based instrument intercomparisons as a fundamental tool to increase the comparability of measurements taken by different systems at different places and at different times. The role of metrology in these field intercomparisons is important, as it was stated by the WMO-CIMO (Commission for Instruments and methods of Observation), in its 17th session (October 2018) [2], and currently it is included in the terms of reference of WMO/INFCOM/SC-MINT [1]. Additionally, the WMO, aware of the importance of air temperature measurements in the evaluation of climate change and the strong influence of external factors on air temperature measurements, recommended the organization of comparisons of radiation shields in extreme climates [2]. Comparisons of radiation shields have been performed in different climate conditions [3–8], but an intercomparison of radiation shields in the polar climate was still required. Air temperature measurements for environmental applications are performed by thermometers protected from the solar radiation and rain by a shield. Combinations of different models of thermometers and

radiation shields are used around the world, increasing the difficulty in the generation of air temperature measurements with robust comparability. In addition, the response of each system (thermometer + radiation shield) to fast temperature changes, solar radiation, wind, humidity, etc., depends on the system itself.

These facts motivated the organization of a comparison of radiation shields in the Arctic climate, which was carried out under the umbrella of the European EMPIR project 19SIP06 Coat [9]. This project is linked to and increasing the impact of the MeteoMet projects [10,11] and supported by the WMO (World Meteorological Organization). The general objective is to increase the comparability of air temperature measurements in Arctic conditions using three specific objectives:

-    To test and evaluate as many radiation shields, commonly used by national weather services, as possible, at the same location and time under polar conditions.
-    To update the knowledge on the performance characteristics and operational factors of radiation shields in Arctic conditions.
-    To estimate the relative different impacts of solar radiation, wind speed and precipitation on air temperature measurements inside different models of radiation shields.

In addition, this intercomparison intends to provide reliable information to a better understanding of some measurement uncertainty components included in the WMO measurement quality classification scheme for surface observations [12]. Quantification of the coupling factor between different shields to same thermometer model is derived in this paper as well as the influence of some environmental factors on air temperature measurements under polar conditions.

Accomplishing these objectives implies laboratory campaigns and a field campaign in the Arctic where air temperature and ancillary quantities are continuously monitored over around one year of field measurements at the research station in Ny-Ålesund, Svalbard (Norway).

This document describes the comparison site in terms of its climate and its design. Then, all the instrumentation involved in the comparison and all the problems found in obtaining the data are outlined. In addition, the quality control applied to all the data, the analysis of the data and the corresponding conclusions are reported in this document.

2. Intercomparison Field, Data and Methods
2.1. Intercomparison Field

The organization of a field intercomparison requires the selection of the optimal place, and although the local climate is the driver component, other aspects such as available and experienced staff at the local comparison field, accessibility, power supply, remote communications and other practical issues need to be considered.

The comparison field is located in the Arctic research station, at NY-Ålesund, Norway (78°55′00″ N 11°56′00″ E). In general, Ny-Ålesund has a polar climate [13–15], despite winter temperatures being warmer than in other Arctic areas due to the North Atlantic drift. The monthly mean, maximum and minimum temperatures during the intercomparison period (16 July 2022 to 1 September 2023) are displayed in Figure 1a [15].

The wind at the comparison site at 10 m height [14], and during the comparison period, flowed mainly to the southeast direction, with wind speeds values indicated in Figure 1b:

-    Lower than 5 m/s for the 34.45% of the events;
-    In the interval (5, 10) m/s for the 42.18% of the events;
-    In the interval (10, 15) m/s for the 18.08% of the events;
-    In the interval (15, 20) m/s in the 4.19% of the events;
-    Higher than 20 m/s for the 1.1% of the events.

Ny-Ålesund was covered by snow from the 15 October to the 1 June, as it is shown in Figure 1c [13] where 0 means no snow, 1 = mostly free of snow ground, 2 = equal parts of snow covered and snow free ground, 3 = mostly snow-covered ground and 4 = completely snow-covered ground.

Figure 1. Climate at Ny-Ålesund at the dates of this shields intercomparison. (**a**) Monthly mean, maximum and minimum temperatures [15]. (**b**) Wind at the comparison field at 10 m height [14]. (**c**) Snow cover and snow depth in Ny-Ålesund [15]. The snow cover is identified by a number 0–4: 0 means no snow, 1 = mostly free of snow ground, 2 = equal parts of snow covered and snow free ground, 3 = mostly snow-covered ground and 4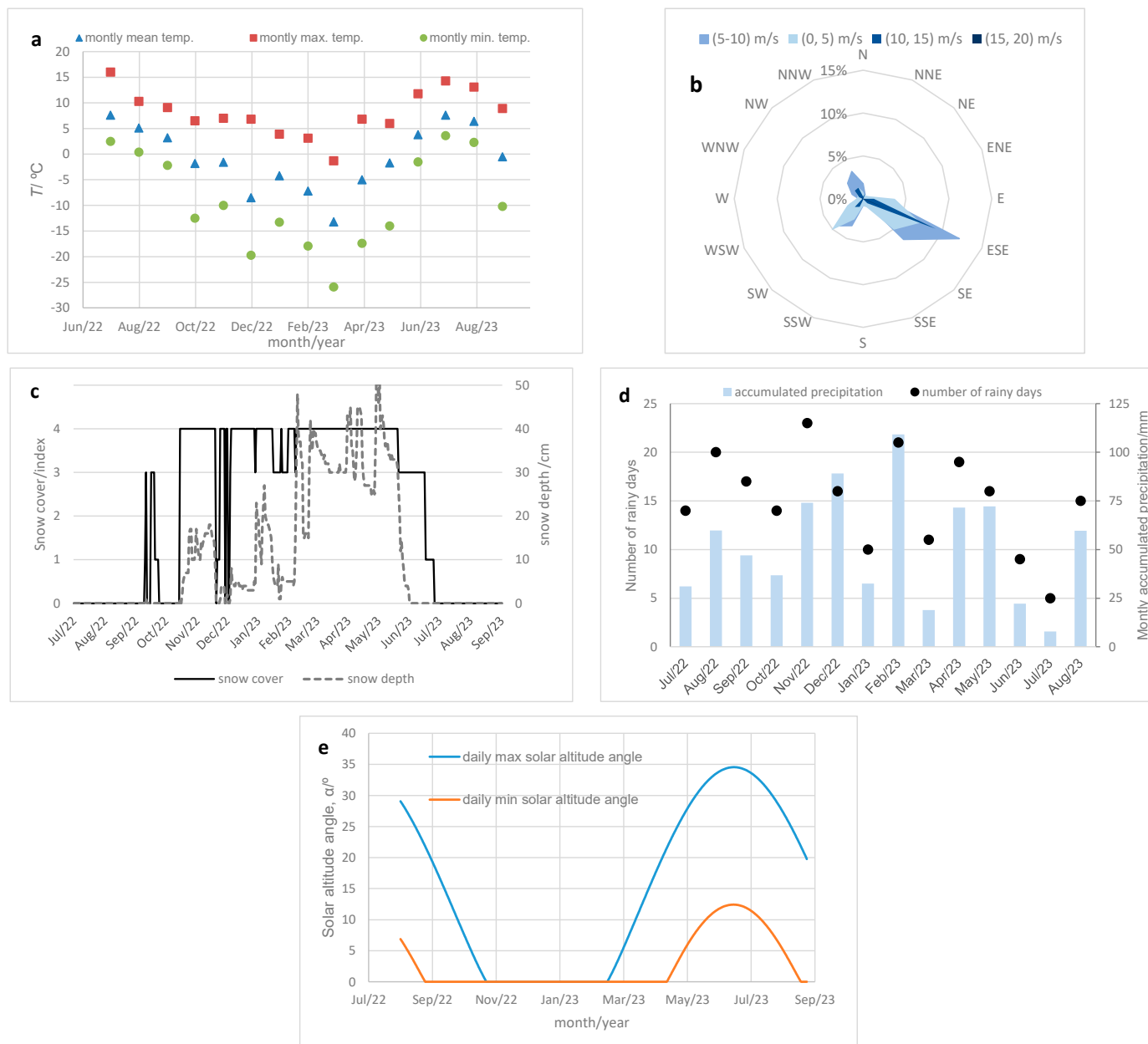 = completely snow-covered ground. (**d**) Monthly number of rainy days and monthly accumulated precipitation [15]. (**e**) Daily maximum and minimum solar altitude angle [16,17].

Rainfall events in Ny-Ålesund during the intercomparison period are presented in Figure 1d, which also includes the monthly accumulated precipitation [15].

In addition, the theoretical maximum and minimum solar altitude angle at Ny-Ålesund, calculated following [16,17], is shown in Figure 1e. This information is important due the solar radiation being the main impact factor on the radiation shields' behavior, and this impact is depending on the angle of solar radiation incidence on the shield. The altitude angle is also important in evaluating the impact of the reflected solar radiation by the ground on air temperature measurements.

The comparisons site is in a flat terrain covered by homogenous natural ground with grass lower than 10 cm height along all the field intercomparison period. It is designed as a rectangular grid with 20 poles, each of them 4 m separated from the adjacent ones (Figure 2. The distance between the poles was decided as a compromise to make interactions between the screens insignificant while keeping all of them on a small area to guarantee the same meteorological conditions. Except for the Stevenson screen that is installed on a specific platform, each pole holds only one radiation shield, and all the shields and thermometers are set so that the temperature measurements are at 2 m height above the ground level during the summer when the ground is not covered by snow. This configuration, in combination with the fact that there are no major obstacles around the comparison field [18–20], guarantees that all the shields are freely exposed to temperature, solar radiation, wind and precipitation.
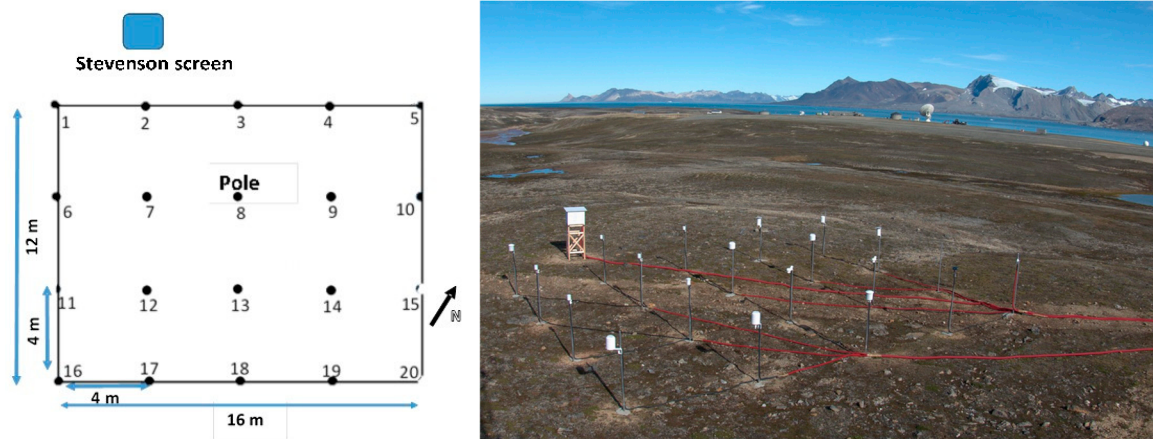

Figure 2. Design of the field experiment. A flat area where the adjacent poles are separated 4 m. The comparison field is free of obstacles.

2.2. Instrumentation

The instrumentation involved in this intercomparison is summarized in Table 1 with a short description about the different shields, their main characteristics, the used thermometers and the period of exposition to the extreme Arctic conditions. Ten different models of radiation shields were included. In some cases, shields of the same model were used under different configurations in terms of the thermometer and/or the data logger, and in other cases, shields of the same model (and under the same configuration in terms of thermometers and datalogger) worked as natural ventilated and also as active ventilated shields. In addition, redundant shields were included to provide indications of the thermal homogeneity of the comparison field. That is the case of shields in pole 1 and in pole 18 and the shields in poles 5 and 16. Furthermore, 2 ultrasonic anemometers were deployed, each of them at different poles: pole 19 and pole 20. These ultrasonic anemometers provided wind measurements at 2 m (same height as thermometers), and one of them also provided readings of sonic temperature. But the analysis of the sonic temperature is out of the scope of this paper.

Table 1. Details of the radiation shields and the thermometers involved in the field comparison.

Pole	Type of Shield	Ventilation	Pt-100	Pt-100 Black	Other Thermometer	Datalogger	Comparison Period	Comments
1	round multiplate	natural	P1/01, P1/02	P1/03b	0	fluke 1586A (n° 2)	16 July 2022–31 August 2023	shield same model as 18
2	round multiplate	natural	P2/01	0	0	fluke 1586A (n° 2)	16 July 2022–31 August 2023	shield same model as 3
3	round multiplate	natural	0	P3/01b	0	fluke 1586A (n° 3)	16 July 2022–31 January 2023	shield same model as 2
4	round multiplate	natural	P4/01	P4/02b	P4/own: Pt-100	fluke 1586A (n° 2)	16 July 2022–31 August 2023	shield same model as 10
5	round multiplate	natural	P5/01, P5/02	P5/03b	0	fluke 1586A (n° 2)	16 July 2022–31 August 2023	shield same model as 16 and 17
6	round multiplate	natural	P6/01	0	0	fluke 1586A (n° 2)	16 July 2022–31 August 2023	shield same model as 7 and 11
7	round multiplate	natural	0	P7/01b	0	fluke 1586A (n° 2)	16 July 2022–31 August 2023	shield same model as 6 and 12
8	round multiplate	natural	P8/01, P8/02	P8/03b	0	fluke 1586A (n° 2)	16 July 2022–6 February 2023	
9	round multiplate	natural	P9/01, P9/02	P9/03b	0	fluke 1586A (n° 2)	16 July 2022–31 August 2023	
10	round multiplate	active (12 V)	P10/01, P10/02	P10/03b	0	fluke 1586A (n° 3)	16 July 2022–31 August 2023	shield same model as 4
11	round multiplate	natural	0	0	P11/own: T&HR sensor (20 V)	own datalogger A	16 July 2022–1 April 2023, 20 May 23–31 August 2023	shield same model as 6 and 7
12	round multiplate	active (12 V)	P12/01	P12/02b	0	fluke 1586A (n° 3)	16 July 2022–31 August 2023	
13	Specific design	active (20 V)	0	0	P13/own: T&HR sensor (20 V)	fluke 1586A (n° 3)	16 July 2022–31 August 2023	shield same model as 14
14	Specific design	active (20 V)	P14/01	0	0	fluke 1586A (n° 3)	16 July 2022–31 August 2023	shield same model as 13
15	round multiplate	natural	P15/01, P15/02	P15/03b	0	fluke 1586A (n° 1)	1 February 2023–31 August 2023	
16	round multiplate	natural	P16/01, P16/02	P16/03b	0	fluke 1586A (n° 3)	16 July 2022–31 August 2023	shield same model as 8 and 17
17	round multiplate	natural	P17/01	0	P17/own: Pt-100 and T&HR sensor	P17/01: fluke 1586A (n° 3). P17/own: own datalogger B	16 July 2022–31 August 2023	shield same model as 8 and 16
18	round multiplate	natural	P18/01, P18/02	P18/03b	0	fluke 1586A (n° 3)	16 July 2022–31 August 2023	shield same model as 1
Stevenson screen	wood	natural	SS/01, SS/02	0	0	fluke 1586A (n° 1)	16 July 2022–31 August 2023	

The thermometers included in Table 1 and used in this shield intercomparison are identical 4-wire platinum resistance thermometers, Pt-100 (IEC-751, Class 1/10 DIN) [21], from the same manufacturer. They are encapsulated in a stainless-steel sheath, and their dimensions are 6 mm diameter and 60 mm long. For some of these thermometers, the emissivity of the sheath was increased by painting them with Nextel Velvet coating; these thermometers are referred in this document as Pt-100 black. The aim of using these black thermometers is to study the impact of potential residual solar radiation inside the shields.

All the thermometers were calibrated at Centro Español de Metrología (CEM) in stirred liquid baths and in a climatic chamber [22] with traceability to the International Temperature Scale of 1990 (ITS-90) [23]. The calibrations of all the thermometers were performed in the same configuration as used in the field in terms of datalogger and datalogger channel. This fact reduces the possible impact of Pt-100 self-heating [24] in the measurements. The expanded uncertainty ($k = 2$) [25,26] associated with the calibration of the temperature reading system (thermometer + datalogger) was 0.02 °C. The drifts [25] of the temperature reading systems (thermometers + dataloggers) were determined by their

calibration just before and just after the complete external exposure in the Arctic. The drift was lower than 0.02 °C for all the temperature systems with the exception of one related to pole 12. The thermometer at pole 12 was connected to a data logger with a resolution of 0.1 °C, and its drift was lower than 0.2 °C. All the instrumentation at pole 12 (shield + thermometer + datalogger) is from the same manufacturer. The drift evaluation of the thermometers at pole 8 was not possible due to these thermometers and the corresponding shield disappearing under a big storm. The drifts are low enough to not impact on the air temperature measurements taken along the complete field experiment and hence on the conclusions of this document. Nevertheless, the source uncertainty due to the drift of the thermometers is included in the total uncertainty budget.

Depending on the internal volume of the shield, one, two or three thermometers were included inside the shield. In almost all the shields with three Pt-100, one of them was of high emissivity sheath. The other two Pt-100 (both with low emissivity sheath) increase the reliability and the assurance of the experiment. This redundancy in thermometers guarantees the existence of useful measurements even in the cases where one of the thermometers fails or drifts (due to multiple reasons such as transportation, the deployment and/or during the exposure to extreme conditions).

This intercomparison implies the measurement of 41 thermometers and supplementary quantities: wind speed and direction (at 2 m and at 10 m), solar irradiance, humidity and pressure. Wind at 2 m is measured by two ultrasonic anemometers: Vaisala WMT700 (pole 20) and Lambretch ultrasonic sensor (pole 19). Wind at 10 m and solar irradiance are measured by the instrumentation included in the Amundsen-Nobile Climate Change Tower (CCT) [14], which is closer than 10 m to the comparison field. Wind measurements at 10 m are performed by the Young Marine Wind Monitor 05106 anemometer, and the solar irradiance is measured by a Kipp and Zonen CNR1 net radiometer. Information about snow cover and precipitation at the comparison site was extracted from the Norway climate database [15]. The sun elevation was determined theoretically for the Ny-Ålesund coordinates [16,17].

The deployment of all the instrumentation was performed following the manufacturers' recommendations with special attention to the correct applied DC voltage (when needed), the connections, and the datalogger communication to the computer in the case of active ventilated shields, thermometers with digital output and dataloggers.

2.3. Data Acquisition

All the comparison instruments were sampled every 2 min. The resistance values of the Pt-100 were acquired by 3 dataloggers Fluke 1586A, each of them with two modules: 1586–2586 High-Capacity Input Module. A PC located at the comparison site monitored the fluke dataloggers by an Ethernet connection and using the Fluke DAQ Application Software (version 6.0). Daily, this software creates a file in the computer with all the channels readings. For the thermometers with digital output, a datalogger of the same thermometer's manufacturer was used (dataloggers A and B). Datalogger A is connected to the PC by Ethernet, the measurements are stored in datalogger A and they can be downloaded to the PC by our own software. Datalogger B is connected to the PC by an RS485-USB adaptor, and the measurements are accessible by HyperTerminal. In addition, the power supplied to the active ventilated radiation shields, datalogger A and datalogger B was also monitored during all the field campaigns. Daily, all the data from the five dataloggers are downloaded to a computer located at CEM for the checking and store. In this way, the raw data were kept updated in two computers, one of them at the comparison site (Arctic) and the other one at CEM. The reason for performing daily control of the data is to detect problems at the very early stage and to solve them in short time. With the daily control, the following problems were detected:

- The ventilation in the pole 4 shield stopped working since almost the beginning of the comparison (around the 20 July). As the pole 10 shield is the same model as the one

at pole 4, this comparison provides information of the same model shield under two different configurations: natural ventilated and artificial ventilated.
- On the 26 August, the power supply applied to datalogger A failed, and it was changed the 2 September.
- On 1 February 2023, a new radiation shield was included in the comparison (pole 15), which was installed by National Research Council, Institute of Polar Sciences (CNR-IPS) local staff.
- On 6 February 2023, after a big storm, the shield and the three corresponding thermometers (Pt-100) in pole 8 disappeared. The drift of these thermometers could not be evaluated.
- On 1 April 2023, datalogger B, related to poles 11 and 19, stopped working. The manufacturer sent a new datalogger that was deployed the 20 May. Unfortunately, the configuration of the new datalogger sampled the data every 10 min instead of every 2 min. This fact could not be checked until the final field campaign due to the communication between the local computer, and the new datalogger B could not be arranged remotely. All the data were kept at the internal memory of the datalogger, and these data were downloaded after the field campaign.
- From 12 April to 19 April, the local PC lost its power supply, and the recovery took some time due to bad meteorological conditions.
- From 14 May 2023 to 28 May 2023, the internal memory of the dataloggers was almost full, which created some irregularities in the sampling time interval, and data from some channels were lost. The usual dataloggers working was reached after the internal memory was cleared out by CNR-IPS local staff.
- In September 2023, when the experiment was dismantled, the thermometer of the shield in pole 3 was found on the ground. By its measurements (provided along all the comparison period), the date the thermometer fell to the ground was estimated: the 4 October 2022. The shield at pole 3 (Pt-100 black) is the same model as the shield at pole 2 (usual pt-100). Although this paper cannot provide a complete analysis of this shield in terms of residual inner solar radiation, important information about its behavior is still included in this report.

All these problems in losing data are quantified in Table 2. In the case of the fluke dataloggers and datalogger B, there was a total failure rate of 5.5%, mainly focused in April and May; meanwhile, for datalogger A, the failure rate was higher due to the time needed to replace the initial datalogger A with a new one (April and May).

Table 2. Percentage of lost data during the field campaign in the Arctic.

	Percentage of Lost Data: Fluke	Percentage of Lost Data: Datalogger A	Percentage of Lost Data: Datalogger B
July 2022	9.3	9.7	11.3
August 2022	1.0	17.9	4.3
September 2022	0.0	5.3	0.1
October 2022	0.0	0.0	0.0
November 2022	0.0	0.0	0.0
December 2022	0.0	0.0	0.0
January 2023	0.0	0.0	0.0
February 2023	0.3	0.0	0.0
March 2023	0.1	0.0	0.1
April 2023	25.3	96.7	25.3
May 2023	39.1	58.8	39.1
June 2023	2.0	0.0	0.0
July 2023	2.3	0.4	0.0
August 2023	0.0	0.0	0.0
total	5.5	33.8	5.5
total without July 2022	5.3	10.1	5.2

2.4. Data Quality Control

CEM developed a specific software to convert the Pt-100 resistance values into temperature values in degrees Celsius, applying the corresponding calibration curve of each sensor. This software also performs a quality control of all the resistance values to avoid any corruption of this comparison's conclusions by the possible wrong data.

The quality control consisted in removing the data in the following instances:

1. Null data. For Pt-100, null data are considered when the resistance value is equal to zero or when the resistance value is void. In the case of thermometers with digital output, null data are considered when there are no data.
2. Erroneous data. Erroneous data are only applicable to Pt-100 and includes the data having exactly the same figure as the previous one. The fluke dataloggers record resistance values with a resolution of 0.1 mΩ (~0.25 mK). The probability of having two consecutive resistance values with the same figure is considered negligible, and this fact was contemplated as a wrong value.
3. Inconsistency data: Resistance values out of the temperature range $(-40, 40)\,°C$.
4. High variability of the data: Data variation higher than 5 °C in relation to the previous value.

Detailed information about the percentage of the removed data for each thermometer and for each month of the comparison due to the quality data control is included in Table S1 and Figure S1 in Supplementary Materials to this document. The data quality control showed that a high percentage of July 2022 data did not pass the quality control, so July 2022 was not considered in the analysis here presented.

Overall, 94% of thermometer measurements passed the data quality control, and for some thermometers, the percentage of valid data was 98.7%. This is the case for the thermometers at pole 1, pole 5, pole 8, pole 9, pole 15 and the Stevenson screen. Then, these shields are the main candidates to be selected as reference ones.

3. Experimental Results and Discussion

3.1. Reference Screen

There is not a globally recognized reference shield to be used in air temperature measurements, and the most optimal shield depends on the local climate. As it is shown in several publications [3–8], solar radiation is the highest impact factor on shields' behavior, mainly at low wind speeds. ISO 17714:2007 [27] states that "*screens that are cooler during the day and warmer during the night are likely to be giving measurements that are closest to the truth*". Following this sentence, defining the most optimal reference radiation shield should be based on the analysis of daily maximum and minimum temperatures, for days with a significant amount of downward solar radiation and when the screens are subjected to a more pronounced diurnal–nocturnal solar cycle.

In this analysis, the establishment of the reference radiation shield is based on the measurements when direct sunlight reaches Ny-Ålesund, that is in the period between the 8 March and the 8 October, with the polar day between 18 April 2023 and 24 August 2023 [15]. Another factor considered by other authors [3–8] is performing the analysis under clear-sky conditions, but clear-sky events are rare at Ny-Ålesund [13,15]. For this reason, this work considers the measurements under all sky conditions, and the analysis is driven by the days with an average solar irradiance higher than a defined threshold values: the daily average of downward shortwave radiation higher than the corresponding values is included in Table 3. The values in Table 3 were defined by considering the climate information provided in [13]. Then, the establishment of the reference shield is based on the measurements of 114 days out of the 245 days with solar radiation reaching Ny-Ålesund. These 114 days were selected after analyzing the downward short-wave radiation (daily average, max and min) reaching the comparison field (Figure 3).

Table 3. Lower limits for average daily downward shortwave radiation. Values based on information provided in [13].

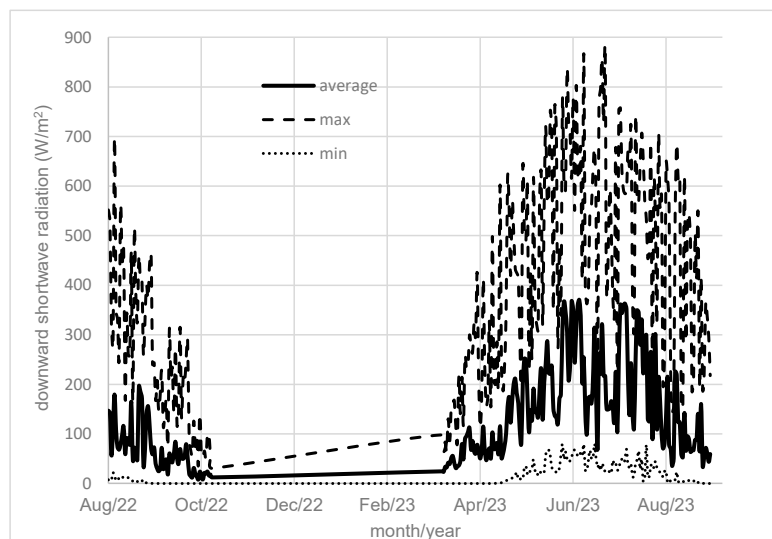| Downward Shortwave Radiation Lower Limit Values. Values Based on [13] (W/m^2)						
March	**April**	**May**	**June**	**July**	**August**	**September**
50	100	175	200	175	100	50

Figure 3. Daily average, maximum and minimum downward shortwave radiation in the comparison field.

The design of the radiation shield is very important in considering it as a reference. The ventilation inside the shields is a key factor in reducing the radiation effects on air temperature measurements. In principle, active ventilated shields are the candidates to work as reference for this comparison. But ISO 17714:2007 [27] recommends not choosing an active ventilated shield as reference unless its performance was fully checked (in terms of its potential psychrometric cooling effects, the effect of the ambient wind on the air flow, if a large aspiration rate heats the sensor, if the air exhausted by the fan may re-circulate and influence the incoming air, and the effect of air flow obstructions on temperature measurements).

In this intercomparison, despite the voltage applied to the fan in the active ventilated shields being monitored along all the field campaign, there is not information about the real aspiration rate inside the shields nor was a complete previous characterization of the active shields performed in the laboratory. In addition, for many shield designs, natural ventilation is effective enough as soon as the wind speed is higher than 1 m/s [27–30]. Figure 4 shows the histogram of wind events (at 2 m) at the comparison field and for the days selected in terms of solar radiation limits (Table 3). Only 17.6% of the events had wind speed lower than 1 m/s, and as we see later in this work, these events happen mainly at night and in moments with low solar irradiance. This justifies the resolution of choosing a natural ventilated shield as the reference one.

Considering the amount of available data after the quality control check for each of the thermometers, the shields at pole 1, pole 5, pole 9 and the Stevenson screen are the optimal candidates to be considered as the reference shield. For each of these shields, the 10-min mean temperature values were used to find the daily extremes air temperatures and the corresponding time of occurrence. The 10 min period was chosen to avoid the effects of possible multiple extreme daily values and to minimize the impact of small-scale turbulences. The extreme daily temperatures in the comparison field were determined by averaging the extreme values for each of the shields. For each day, the difference between

each shield extreme value and the averaged one was calculated. Figure 5a shows these differences for maximum and minimum temperatures for the potential reference shields.
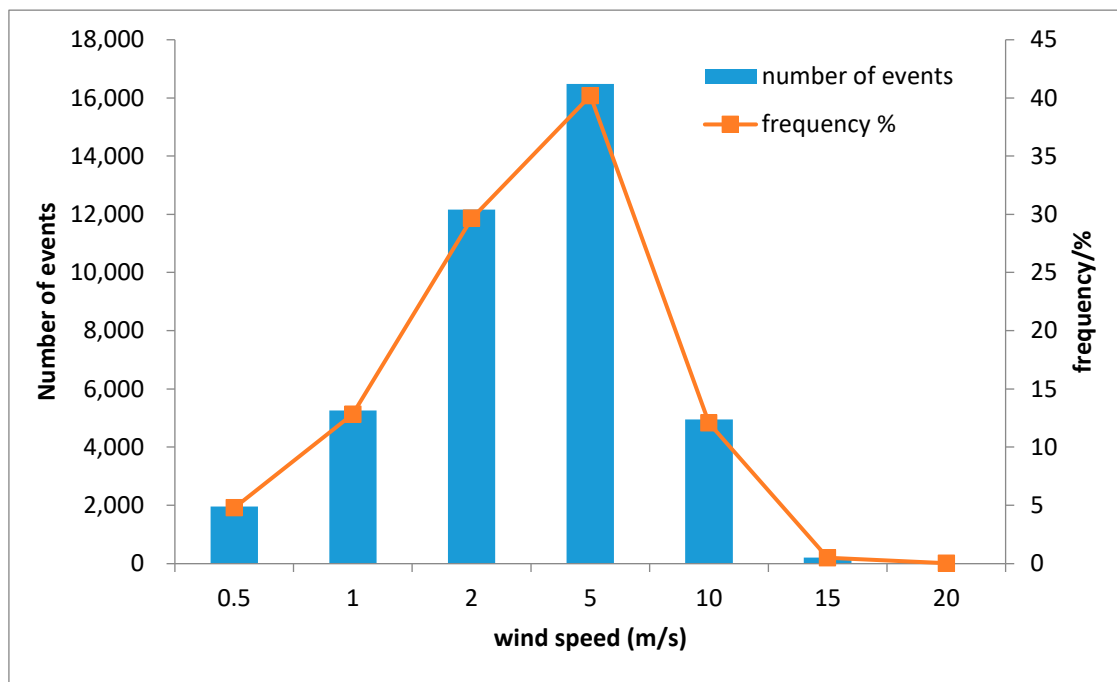

Figure 4. Histogram of the wind events (at 2 m) for the period with solar radiation at NY-Ålesund and with average downward shortwave radiation higher than the values in Table 3.
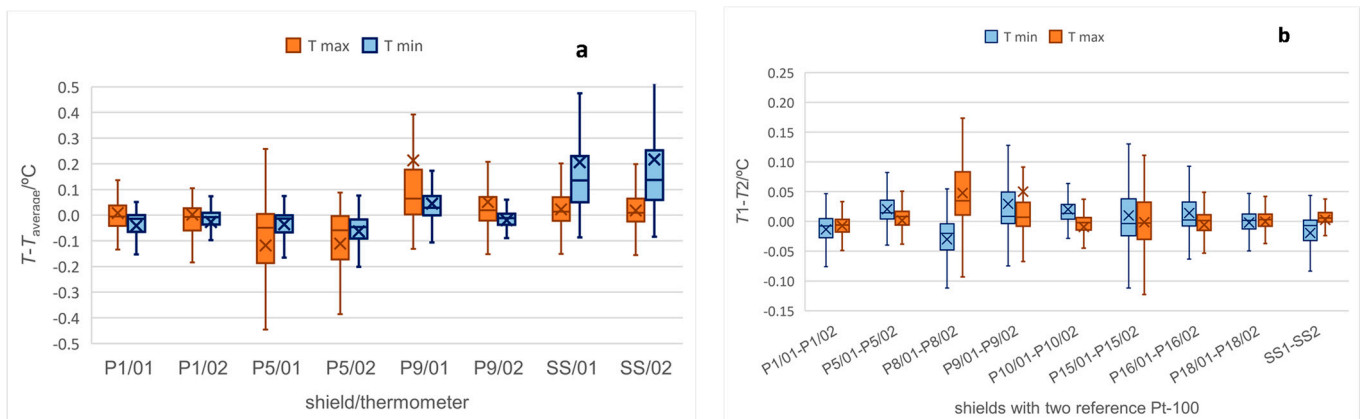

Figure 5. Readings of the thermometers in the potential reference shields for the daily maximum and daily minimum temperatures: (**a**) Differences to the average daily extreme temperatures. (**b**) Differences between the two Pt-100 included the shields for all the comparison period. "x" is the average value, "-" is the median, "□"is the 25–75% interval and "I" is the complete interval.

Focusing on maximum temperatures, the shield at pole 5 could be the reference one, and focusing on minimum temperatures, the Stevenson screen would be the preferred one. But the high variability of the thermometer readings at these temperatures should be also considered in the final decision.

The selection of the reference shield is also based on the reliability of the thermometers' readings inside the shields. As it was previously explained, the reliability is improved by the use of two Pt-100 (Table 1). Having evaluated that the drift of the thermometers is low enough (lower than 0.02 °C), the difference between the two synchronized thermometers' readings was considered as an indicator of the thermal homogeneity inside the shield and

as an indicator of the reliability in the defined temperature inside the shield. For daily extreme temperature conditions, the difference between the thermometer's readings are displayed in Figure 5b for all the shields with two Pt-100.

For the determination of the reference shield, despite shield 1 and shield 5 show very similar behavior, shield 1 was chosen as the reference one due to its lower measurements dispersion in Figure 5a,b. The Stevenson screen also showed an excellent performance but it was not chosen as the reference due its theoretical delay in response time [29].

The shield at pole 1 was chosen as the reference one based of several facts, which were already previously explained:

1. Covering long comparison period (Table 1);
2. Redundant shields deployed in the comparison field. Same model shields at pole 1 and at pole 18 (Table 1);
3. Amount of available data after the data quality control check;
4. Colder daily maximum temperatures and the dispersion of the measurements (Figure 5a);
5. Warmer daily minimum temperatures and the dispersion of the measurements (Figure 5a);
6. Reliability of the measurements provided by the two thermometers included in the shield and thermal homogeneity inside the shield (Figure 5b).

Figure 5b shows that the dispersion of the differences between the readings of the thermometers is a bit high for some shields. This dispersion could be an indicator of internal residual solar radiation (affecting in a different way each thermometer due to their position inside the shield) or noisy events. In the next sections, the temperature of the shield is calculated by the average of the two thermometers' readings for the shields with the complete interval of the difference between the two thermometers' readings in the range $\pm 0.1\ ^\circ$C. In the other shields, the reading of the two thermometers is studied.

3.2. Comparison of the Shields under Daily Maximum and Minimum Air Temperature Values

Once the reference radiation shield is defined, the differences between the shields and the reference one are evaluated in terms of daily maximum and minimum temperatures based on 2-min data. Figure 6 shows the distribution of these differences, where two periods are distinguished in terms of solar radiation: polar night (from 9 October 2022 to 4 March 2023, with maximum solar irradiance lower than 60 W/m^2 and maximum solar elevation lower than 5°) and the rest of the comparison period.
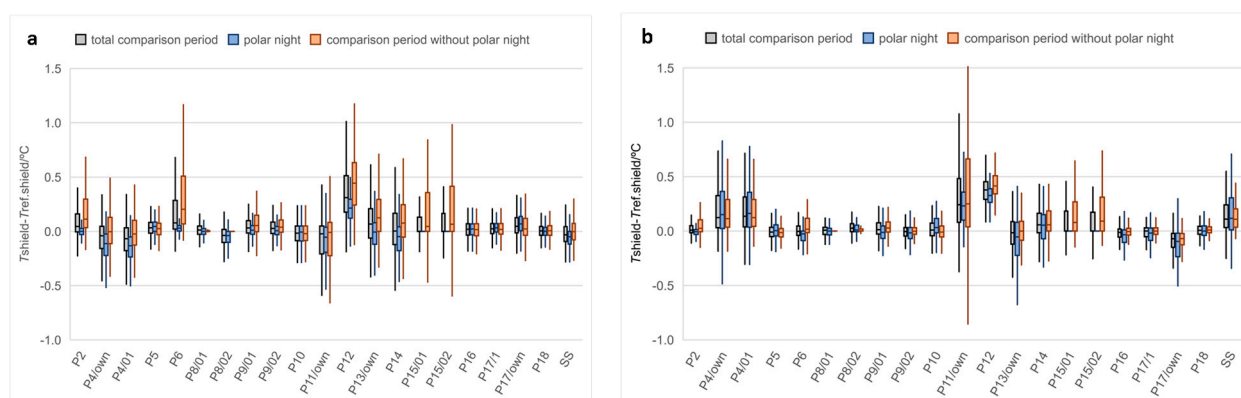

Figure 6. Difference between shields and the reference shield (**a**) for daily maximum temperatures and (**b**) for daily maximum temperatures. "-" is the median, "□"is the 25–75% interval and "I" is the complete interval.

Figure 6 shows that for some of the shields, the difference to the reference one is independent of the solar irradiance (same differences for the polar nights as for the rest of

the comparison period). In other shields, such as pole 6, 11, 12, and 15, a dependence of solar radiation could be derived. This is deeply analyzed later in this paper (Section 3.5).

3.3. Lag-Time Response

The lag-time response of the shields in relation to the reference one was also studied for daily maximum and minimum temperatures. Figure 7 shows the difference between the time of the extreme values happening for the different shields and the time of extreme values for the reference shield. For almost all the shields, the lag-time response seems independent of solar irradiance conditions, having similar values for polar night than for the rest of the comparison period. But it is worth saying that the lag-time response is in coincident with the sampling period, 2 min. This time delay is probably more related to the accuracy in sampling all the datalogger channels than to the real response time of the thermometers inside the corresponding shields. Having this fact in mind, the shield at pole 11 (natural ventilated) with its own thermometer shows systematically maximum and minimum temperatures later than the reference shield, and something similar happens in the shield at pole 13 (active ventilated). In the case of the shield at pole 11, the delay is due to the answer time of the thermometer instead due to the shield, since the shield at pole 6 (same model as the one at pole 11) does not show the same time response delay as in pole 11. Something similar happens with the shield at pole 13 (same model as the one at pole 14, both active ventilated). It is interesting to highlight that despite the usual thought that the active ventilated shields answer faster than the natural ventilated ones, this study does not show the same conclusion, which is probably due the comparison area is quite windy (Figure 4), and wind mitigates the response time delays between the shields. The same reasoning could be the explanation for the lack of lag-time in the Stevenson screen. For the Stevenson screen, a delay response time was observed in some previous works [28–32]; here, this effect is not conclusive. The time resolution of the experiment, 2 min, is not good enough for lag-time response studies, and a less windy place could be more adequate to study this effect.
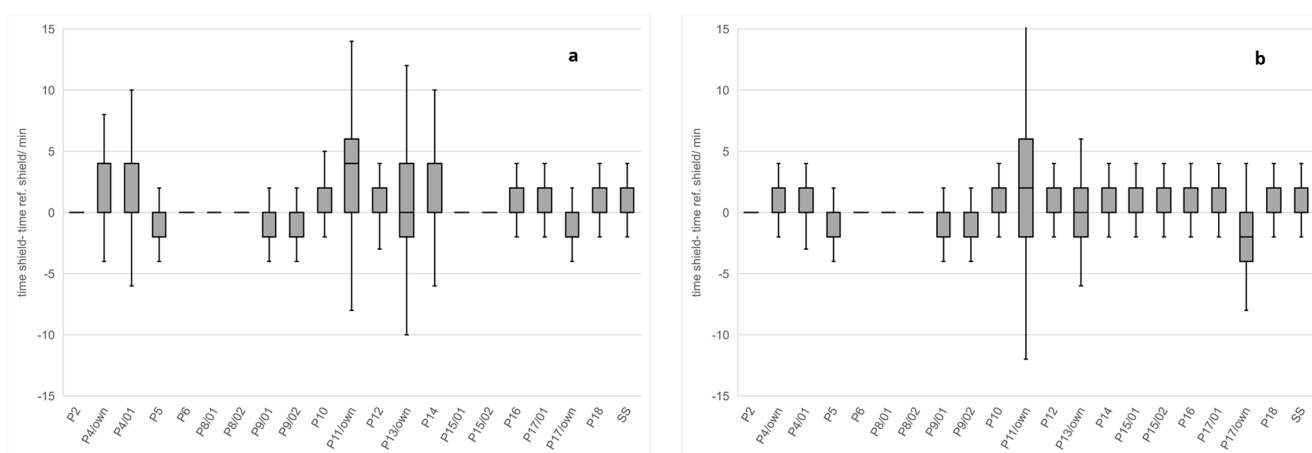

Figure 7. Response time of the different shields in relation to the reference shield for (**a**) daily maximum temperatures and for (**b**) daily minimum temperatures. "-" is the median, "□"is the 25–75% interval and "I" is the complete interval.

3.4. Effect of Ground Covered by Snow on the Behavior of the Shields

The shields' performance in relation to the snow conditions on the ground was also evaluated. For this purpose, the difference to the reference shield for daily extreme temperature values was analyzed in terms of the temporal period with the ground completely covered by the snow and the temporal period without snow on the ground. Figure 8 displays these differences. The establishment of the period of time covered by snow is based on the information provided in Figure 1c, considering only the days when the snow covering index takes the value of 3 or 4. The period of time without snow on the ground

is defined by the days with a snow covering index of 0 and 1. Due to the solar irradiance being the main impact factor, the measurements taken in the period of polar nights were excluded from this study.
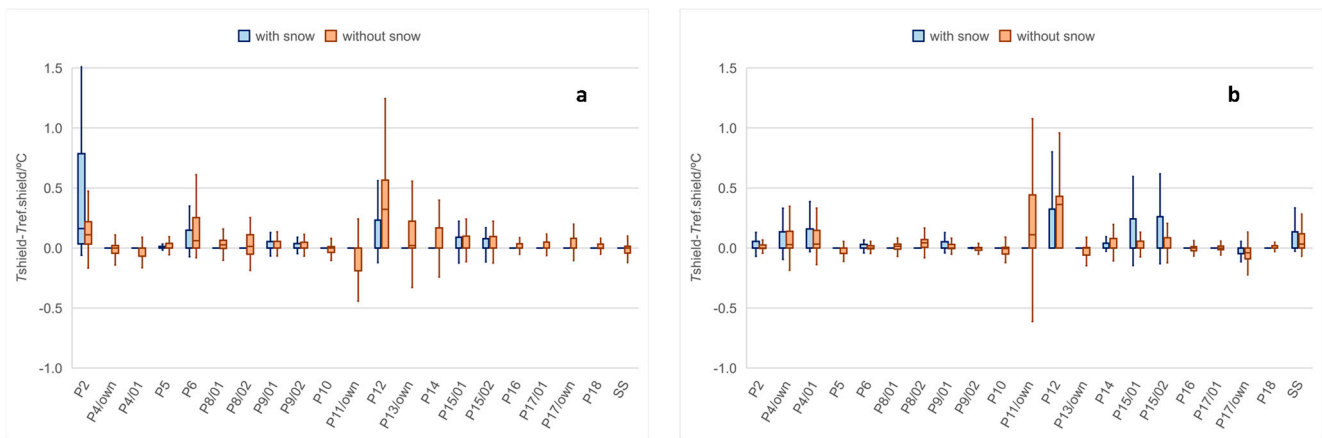

Figure 8. Differences to the reference shield for (**a**) daily maximum temperatures and for (**b**) daily minimum temperatures under different ground conditions during the comparison period excluding polar nights. "-" is the median, "□" is the 25–75% interval and "I" is the complete interval.

The information provided in Figure 8 suggests that the majority of the shields are not sensitive to the reflected solar radiation by the snow covering the ground. An explanation for this could be in terms of the sun elevation at Ny-Ålesund during the snow covering period. Figure 1e shows that the maximum sun elevation for the snow covering period is 34.5°. For this incidence angle, the reflected solar radiation component hardly reaches the inner walls of the shields, and the difficulty in reaching the inner part of the shields increases with the reduction in solar elevation. Despite this fact, the dispersion of the values at pole 15 and mainly at pole 2 seems to be affected by the ground covering.

Regarding the shield at pole 2, it shows higher measurements dispersion when the ground is covered by snow and at daily maximum temperatures that usually correspond to high solar irradiance conditions. Despite this fact, the two medians' difference is lower than 0.05 °C.

Regarding the shield at pole 15, the measurements dispersion is higher when the ground is covered by snow and only for daily minimum temperatures, happening mainly at moments with low values of solar radiation and in coincidence usually with daily lower wind speed events (Figure 9).
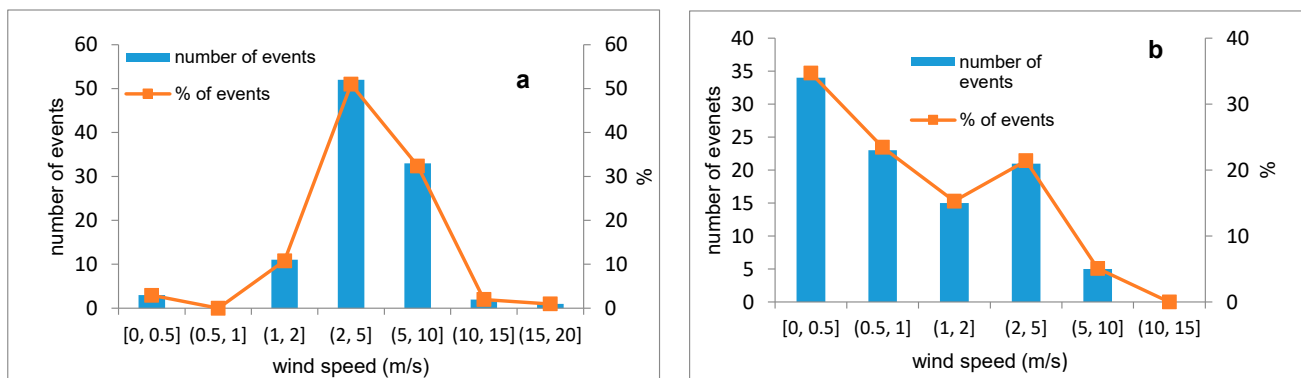

Figure 9. Histogram of the wind events for (**a**) daily maximum temperatures and (**b**) for daily minimum temperatures during the comparison period and for days with mean solar irradiance higher than the averaged climatic values (Table 3).

3.5. Direct Solar Irradiance Impact on the Performance of the Shields

As it was explained for the selection of the reference shield, the clear days at Ny-Ålesund are rare [13,15]. For this reason, the solar irradiance impact on radiation shields performance is evaluated on days with mean solar irradiance higher than the one established in Table 3 and for all sky conditions. As [27] stated, the impact of solar irradiance strongly depends on wind speed values, expecting higher influences for low wind speeds. Figure 9 displays the wind events for days fulfilling the requirements established in Table 3 in terms of solar irradiance and for the moments of daily maximum and minimum temperatures. Figure 9 concludes that usually the daily maximum temperatures happen in coincidence with wind speed higher than 5 m/s; meanwhile, daily minimum temperatures happen when the wind speed is lower than 5 m/s.

Figure 10 shows the performance of the shields in relation to the reference one, at the times of daily maximum and minimum temperatures, for the days achieving the conditions established in Table 3 in terms of solar irradiance values and for different wind speeds.
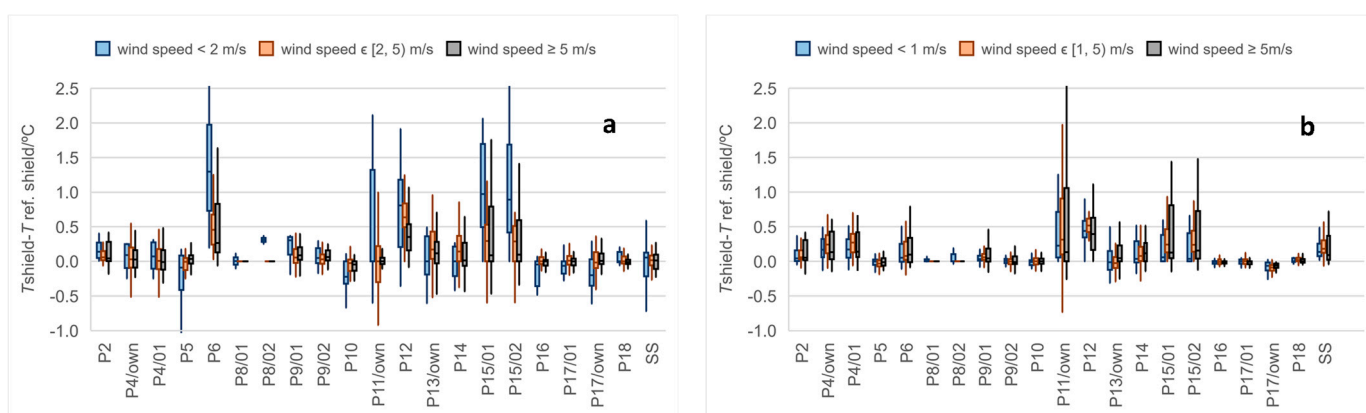

Figure 10. Differences in the reference shield for (**a**) daily maximum temperatures moments and for (**b**) daily minimum temperatures moments, for days reaching the requirements regarding solar irradiance (Table 3) and for several wind speeds. "-" is the median, "□"is the 25–75% interval and "I" is the complete interval.

The information provided in Figure 10 shows that the differences of the shields in relation to the reference one are higher for daily maximum temperatures. In addition, the distributions of the measurements are wider for daily maximum temperatures as well as the impact of the wind speed on the shields' performance. The reason for these points is that the daily maximum temperatures usually happen at moments with high values of solar irradiance and, as it was already said before in the paper, the solar irradiance is the main impact factor on the shields' performance.

In order to detect the maximum differences between the shields and the reference one, the performance of the shields was analyzed for daily maximum temperatures and for the days with mean daily solar irradiance higher than the values included in Table 3. Figure 11 displays the result of this analysis for each of the shields, and it provides information about the following:

- Impact of the solar irradiance on the shield performance in relation to the reference shield.
- Impact of the wind speed on the shield performance.
- The presence of residual solar radiation inside each shield model is derived by the difference between the readings of the black Pt-100 and the usual Pt-100s.
- Shield performance for different thermometers.
- Reliability of the temperature measurements inside the shield with two Pt-100.
- Each of the figures (Figure 11a–j) includes a comparison of the same model shield under different configurations.
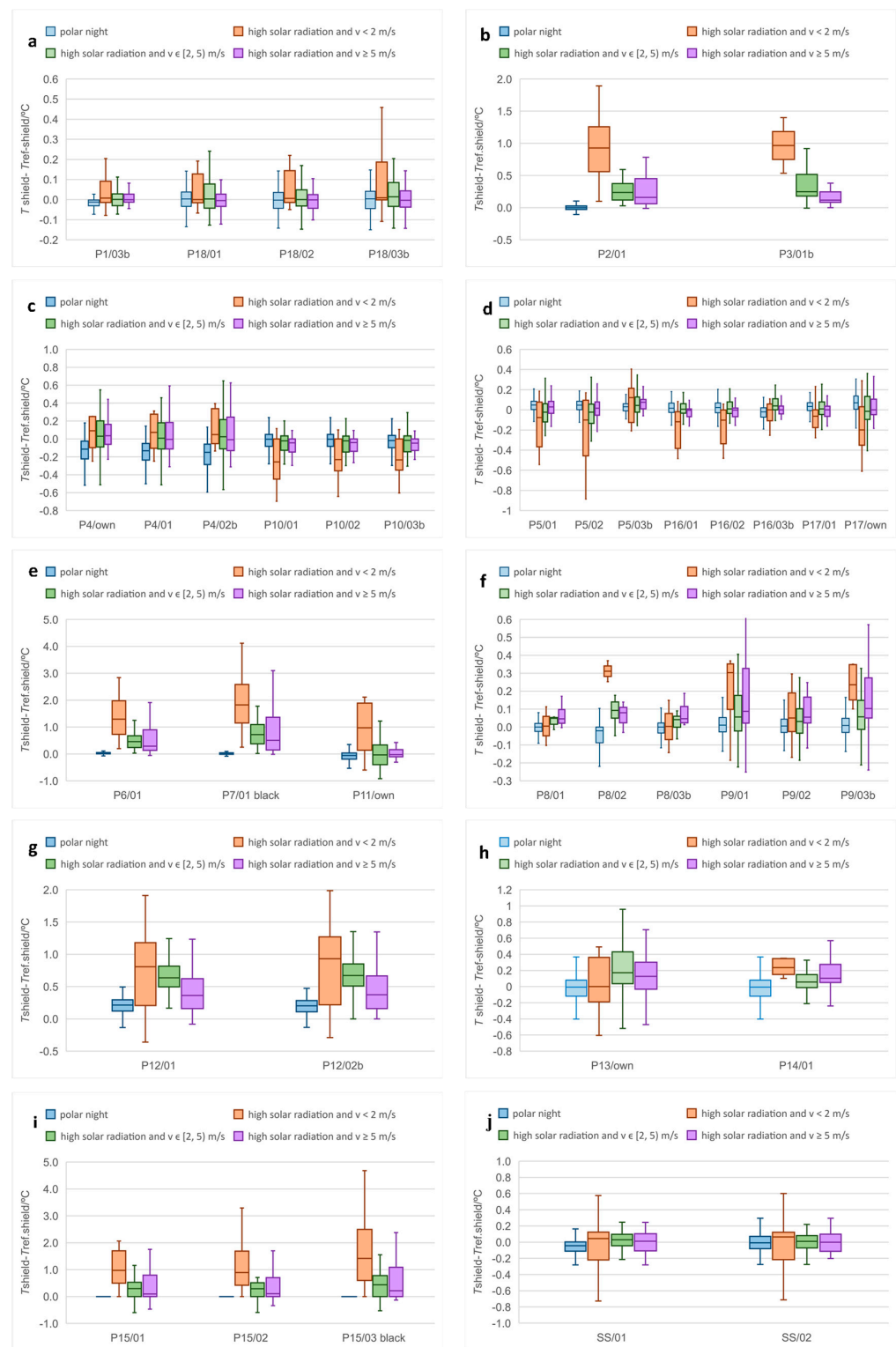
Figure 11. Behavior of the shields in relation to the reference shield at daily maximum temperatures for (**a**) shields at pole 1 and at pole 18, (**b**) shields at pole 2 and at pole 3, (**c**) shields at pole 4 and at pole 10, (**d**) shields at pole 5, at pole 16 and at pole 17, (**e**) shields at pole 6, at pole 7 and at pole 11, (**f**) shields at pole 8 and at pole 9, (**g**) shields at pole 12, (**h**) shields at pole 13 and at pole 14, (**i**) shields at pole 15 and (**j**) Stevenson screen. "-" is the median, "□"is the 25–75% interval and "I" is the complete interval.

The numerical values of the differences between the shields and the reference one are included in Table 4.

Table 4. Median of the temperature differences between the shields and the reference one at daily maximum temperatures under high values of solar irradiance and under different wind conditions.

	Polar Night	v < 2 m/s	v ϵ [2, 5) m/s	v $\geq$ 5 m/s
difference to ref. shield, P2, (°C)	0.00	0.93	0.24	0.16
difference to ref. shield, P4/own, (°C)	−0.11	0.09	0.03	0.04
difference to ref. shield, P4/01, (°C)	−0.13	0.07	0.01	−0.01
difference to ref. shield, P5, P16 = (P5/01 + P5/02 + P16/01 + P16/02)/4, (°C)	0.03	−0.11	0.00	0.01
difference to ref. shield, P6, (°C)	0.03	1.29	0.46	0.29
difference to ref. shield, P8 = P8/01, (°C)	0.00	0.01	0.05	0.05
difference to ref. shield, P9 = P9/02, (°C)	0.01	0.05	0.03	0.06
difference to ref. shield, P10 = (P10/01 + P10/02)/2, (°C)	0.00	−0.29	−0.03	−0.05
difference to ref. shield, P11/own, (°C)	−0.05	0.97	−0.04	−0.02
difference to ref. shield, P12, (°C)	0.21	1.04	0.64	0.45
difference to ref. shield, P13/own, (°C)	−0.01	0.11	0.20	0.16
difference to ref. shield, P14, (°C)	−0.01	0.24	0.06	0.10
difference to ref. shield, P15 = P15/01 + P15/02)/2, (°C)		1.01	0.42	0.19
difference to ref. shield P17/own, (°C)	0.08	−0.20	−0.01	0.00
difference to ref. shield, P18 = (P18/01 + P18/02)/2, (°C)	0.00	0.05	0.00	−0.02
difference to ref. shield, SS = (SS/01 + SS/02)/2, (°C)	−0.03	0.05	0.02	0.01

3.5.1. Comparison of Shields at Pole 1 (Reference Shield) and at Pole 18

The reference shield and the one at pole 18 are both the same model: big enough to hold three Pt-100 and with natural ventilation. Figure 11a highlights the independence of the shield performance from the solar irradiance, since the median values are very similar for polar nights and for moments with high solar irradiance values, even when the wind speed is lower than 2 m/s. In addition, Figure 11a shows that the residual solar radiation inside the shield is negligible; only an increase in the dispersion of the measurements for wind speed lower than 2 m/s is observed for the high-emissivity thermometers.

3.5.2. Comparison of Shields at Pole 2 and at Pole 3 to the Reference One

The shields at pole 2 and at pole 3 are the same model, both with natural ventilation and with a small internal volume. Then, only one Pt-100 is included in each of the shields: a usual one at pole 2 and a black one at pole 3. During the dismantling of all the instrumentation, in September 2023, the black thermometer P3/01b was found on the ground. By analyzing the corresponding measurements, it was estimated that the black thermometer left the shield at pole 3 at the beginning of October 2022. This means there are only a couple of months of measurements for the shield at pole 3 and without data for polar nights. Focusing on Figure 11b, it can be derived that the solar irradiance impacts on this shield performance mainly for wind speed lower than 2 m/s. The ventilation inside the shield is good enough for significantly reducing the impact of solar irradiance, mainly for wind speeds higher than 5 m/s. In addition, Figure 11b shows that the difference between the median values of the black Pt-100 and the usual Pt-100 are negligible, meaning the residual solar radiation inside the shield is very limited.

3.5.3. Comparison of Shields at Pole 4 and at Pole 10 to the Reference One

The shields at pole 4 and at pole 10 are the same model shields, but one of them works with natural ventilation (pole 4) and the other one works with active ventilation (pole 10). Figure 11c displays the performance of these shields, and it is in agreement with the statement established in [25] regarding active ventilated shields providing higher temperature values for events without solar radiation (polar nights) and lower temperature values for high solar radiation events. But this decrease in temperature values is

cushioned by the increase in wind speed. In addition, comparing the readings of the black Pt-100 included in both shields with the other two thermometers readings, it can be concluded that the residual solar radiation inside the shield is negligible independently of the shield configuration (natural-active ventilation). Figure 11c also shows that these two configurations (natural/active ventilation) show a negligible impact on temperature measurements mainly for wind speeds higher than 2 m/s. The main difference between these two configurations is in terms of the dispersion of the measurements, being higher for the natural ventilated mode. On the contrary, the natural ventilated mode shows a smaller dependence on wind speed.

3.5.4. Comparison of Shields at Pole 5, at Pole 16 and at Pole 17 to the Reference One

The same model shields were installed at poles 5, 16 and 17: a big dimension shield with natural ventilation. P5 and P16 were working in the same configuration in terms of the thermometers. Both shields included three Pt-100s, and one of them had a high emissivity sheath (black). The shield at pole 17 included a usual Pt-100 and also the complete manufacturer system (shield + thermometer + datalogger). Figure 11d provides the comparison of these shields in relation to the reference one, and it can be derived that these shields provide lower temperature values than the reference one, mainly at low wind speeds. Despite this fact, the variation in the median values with wind speed is lower than 0.1 °C, and it decreases with wind speed. The median values are independent of solar irradiance conditions (same value for polar nights and for moments of high solar irradiance) mainly for wind speed higher than 2 m/s. By comparing the reading of the black thermometers against the other thermometers, it can be concluded that the residual solar radiation inside the shield impacts mainly when the wind speed is lower than 2 m/s, reaching the highest value of 0.2 °C in the P5 shield. Figure 11d also shows that the thermometer supplied by the manufacturer (P17/own) provides lower temperatures for low wind speed moments, but on the contrary, it is noisier.

3.5.5. Comparison of Shields at Pole 6, at Pole 7 and at Pole 11 to the Reference One

The same model shields were installed at poles 6, 7 and 11. The unique difference between them is that pole 11 includes a complete manufacturer system, with a shield, thermometer and a datalogger supplied by the same manufacturer, and the temperature readings are provided with a resolution of 0.1 °C. Another fact to be considered regarding pole 11 is that the datalogger failed the 1 April, and the new datalogger was installed at the comparison field on the 20 May. The new datalogger provided values every 10 min instead of every 2 min (as the other dataloggers did). The change in datalogger implied losing the data of April and May 2023, which were months that were extremely important in terms of solar radiation conditions. In addition, having data values every 10 min means fewer details about the real extreme temperature values at the pole 11 system during the comparison. Pole 6 includes the same Pt-00 as the other shields, and the shield at pole 7 includes a black Pt-100.

Figure 11e compares the measurements of the same shield model but with different configurations in terms of the thermometers, and it shows that this shield is not immune to solar irradiance, and it is affected by residual solar radiation inside the shield. This last fact is deduced from the difference performance when this shield is combined with a high-emissivity thermometer (black one at pole 7), where the error in temperature measurements can be close to 2 °C for wind speed lower than 2 m/s. The system at pole 11 shows a cushioned impact of solar irradiance, which is probably due to the own thermometer sensor (white T&HR sensor) being more protected against the direct impact of residual solar radiation inside the shield than when a usual Pt-100 is used. In all the configurations, the solar irradiance effect decreases with wind speed.

3.5.6. Comparison of Shields at Pole 8 and at Pole 9 to the Reference One

The radiation shields at poles 8 and 9 are the same model. The unique difference is that the shield at pole 8 has its inner walls painted in white, while the shield at pole 9 has its inner walls painted in black. The shield at pole 8 and the associated thermometers disappeared in February 2022 after a storm, making impossible the evaluation of the thermometers drift after the field campaign. Thus, the available measurements are limited to just few months, and almost all of them are in the polar night period. The available P8 data for the conditions of daily maximum temperatures in days with solar irradiance higher than the values included in Table 3 are very limited. Despite this, the difference between this shield and the reference one under these conditions is displayed in Figure 11f. P8/02 shows a no coherent performance, which was probably due to it suffering some drift after its initial calibration, so the measurements of the thermometer P8/02 are not considered. The comparison of the median values in Figure 11f indicates that the performance of this shield is independent of the solar irradiance conditions since the medians of the thermometer P8/01 are similar for polar nights and for the days with high solar irradiance. Also, the wind speed influence seems minor. In addition, the residual solar radiation inside the shield is negligible in terms of the difference between the median values of thermometer P8/01 and P8/03 b. For the shield at pole 9, Figure 11f shows that the dispersion of the measurements is bigger than that for the shield at pole 8, but it could be due to there being less measurements for pole 8 than for pole 9. In addition, the readings of P9/01 are not coherent with the readings of the black thermometer, P9/03b, mainly for low wind speed, nor are they coherent with thermometer P9/02 in terms of the measurements dispersion. Despite the drift of the three thermometers evaluated reaching values lower than 0.02 °C, it was decided that the shield performance is defined by the readings of the thermometer P9/02. The thermometer P9/01 was probably affected by some external noise that disturbed its measurements. Focusing on the median value of P9/02, a very limited influence of the solar irradiance conditions is derived, since the median value was 0.05 °C for polar nights and for high solar irradiance moments. In addition, the impact of external wind speed is negligible. By the different performance of P9/01 and P9/03b, a residual internal solar radiation can be derived with a maximum effect of around 0.2 °C when the wind speed is lower than 2 m/s.

3.5.7. Comparison of Shields at Pole 12 to the Reference One

The shield at pole 12 is actively ventilated, and it includes two Pt-100, one of which is painted black. As shown in Figure 11g, this shield shows noisy measurements, even at polar nights with a lack of solar irradiance. For the period of polar nights, this shield is warmer than the reference one (natural ventilated), which is in agreement with the statement that the ventilated shields provide higher values of temperature at night [25]. On the contrary, this shield does not follow the rule that the active ventilated shield provides lower maximum temperature values [25] (event with solar radiation). The high difference in relation to the reference shield and the strong dependence of these differences with wind speed indicate this shield is affected by solar irradiance. On the contrary, Figure 11g shows similar measurements' median values for the two Pt-100, independently of the thermometer emissivity. This is probably due to the combination of two facts: the residual solar radiation inside the shield is low and the active ventilation inside the shield reduces the possible effect of this (low) inner residual solar radiation.

3.5.8. Comparison of Shields at Pole 13 and at Pole 14 to the Reference One

The shields at poles 13 and 14 are same model, active ventilated shields. Their internal volume was not big enough to include more than one thermometer. The shield at pole 13 includes a thermometer from the same manufacturer as the shield; meanwhile, the shield at pole 14 includes a Pt-100. Figure 11h displays that by comparing median values, the solar irradiance has a maximum effect of 0.2 °C when measurements are performed by the Pt-100, and this effect is reduced with wind speed. This conclusion is not so clear for the

thermometer P13 due to the dispersion in its measurements being higher, probably as a consequence of its more datalogger limited resolution.

3.5.9. Comparison of Shields at Pole 15 to the Reference One

The shield at pole 15 was installed 1 February 2023, 7 months after the starting of the field campaign. This implies the available data are reduced regarding the other shields, mainly for the period of polar nights when more stable measurements were generated due to the lack of solar radiation. The several figures along this paper suggest this shield is affected by solar radiation mainly at low wind speed conditions and for maximum daily temperatures (Figure 10a). In addition, the difference between the two thermometer readings showed a dispersion higher than 0.1 °C for the total readings interval (Figure 5b). This high dispersion could be due to the lack of immunity to solar irradiance. But despite the dispersion of the thermometers readings being high, the median value for the two thermometers P15/01 and P15/02 is very similar (Figure 11i). The median of the difference in relation to the reference shield for both thermometer P15/01 and P15/02 is around 1 °C for extreme conditions in terms of temperature, solar irradiance and low wind speed. The impact of the residual solar radiation inside the shield is analyzed by the readings of a black Pt-100. The black Pt-100 shows higher median values for all wind conditions and the difference between thermometer decreases with wind speed. Finally, the performance of this shield is very similar to the reference one for wind speed higher than 5 m/s and also for the polar night (lack of solar radiation) conditions.

3.5.10. Comparison of Stevenson Screen to the Reference One

The Stevenson screen is big enough to hold three Pt-100, but despite this fact, a thermometer with a high-emissivity sheath was not included. The two Pt-100 included in the Stevenson screen are the usual ones. Figure 11j highlights the very low dependence of the shield performance on the solar irradiance, since the median values are very similar for polar nights and for moments with high solar irradiance values. The maximum difference is for the events with wind speed is lower than 2 m/s, and this difference is lower than 0.1 °C. These differences decrease with wind speed.

3.6. Uncertainties

Each shield performance (t_i) in comparison to the reference one ($t_{reference}$), when both shields are under the same conditions in terms of solar radiation, temperature and wind speed, is analyzed by Equation (1) and with more detail, by Equation (2), where the input quantities identified by δ are estimated to be zero but not their associated uncertainties. Equations (2) and (3) included components associated with the calibration of the thermometers ($\delta t_{ref,calibration}$, $\delta t_{i,calibration}$), the drift of the thermometers ($\delta t_{ref,drift}$, $\delta t_{i,drift}$), the involved dataloggers ($\delta t_{ref,reading\ system}$, $\delta t_{i,reading\ system}$) and the thermal homogeneity of the comparison field during the field campaign ($\delta t_{thermal\ homogeneity\ of\ the\ comparison\ field}$)

$$\Delta t = t_i - t_{reference} \tag{1}$$

$$\Delta t = \left(t_i + \delta t_{i,calibration} + \delta t_{i,drift} + \delta t_{i,reading\ system}\right) - \left(t_{ref} + \right.$$
$$\left. \delta t_{ref,calibration} + \delta t_{ref,drift} + \delta t_{ref,reading\ system}\right) + \tag{2}$$
$$\delta t_{thermal\ homogeneity\ of\ comparison\ field}$$

To calculate the associated uncertainty of the shield performance in comparison to the reference shield, the law of propagation of uncertainties [24] should be applied to (2), resulting in

$$u^2(\Delta t) = u^2\left(t_i - t_{ref}\right) + u^2(\delta t_{,calibration}) + u^2(\delta t_{i,\ calibration}) + u^2\left(\delta t_{ref,drift}\right) + u^2\left(\delta t_{i,\ drift}\right) +$$
$$u^2\left(\delta t_{ref,\ reading\ system}\right) + u^2\left(\delta t_{i,reading\ system}\right) + u^2\left(\delta t_{thermal\ homogeneity\ of\ comparison\ field}\right) \tag{3}$$

The description and quantification of the different uncertainty sources is performed:

3.6.1. Uncertainty Contribution Due to the Standard Deviation of the Differences between the Shield under Study and the Reference Shield, $u(t_i - t_{reference})$

In this report, the behavior of the shields are analyzed in comparison to a reference one by the median of all the measurements performed under conditions that magnified the difference between shields. The means, measurements for daily maximum temperatures and daily mean solar irradiance higher than the values are included in Table 4.

Choosing the median value implies that values near the bounds are less likely than those near the median. It is then reasonable to choose a symmetric triangular probability distribution, having equal sloping sides and with lower bound a− = 0% of all the measurement intervals and a higher bound a+ = 100% of the complete measurements interval. In a triangular distribution, the standard uncertainty of the median is described by (4), where a is the half-width of the probability density function.

$$u\left(t_i - t_{reference}\right) = \frac{a}{\sqrt{6}} = \frac{((a_+ - a_-)/2)}{\sqrt{6}} \tag{4}$$

3.6.2. Uncertainty Contribution Due to the Calibration of the Thermometers, $u(\delta t_{ref,calibration})$, $u(\delta t_{i,calibration})$

The uncertainty component due to the calibration of the thermometers is the value of the expanded calibration uncertainty value divided by the coverage factor. The calibration of the Pt-100 was performed just before and after the field experiment, using the corresponding datalogger as the reader of the Pt-100. Then, the expanded calibration uncertainty was 0.02 °C (k = 2), including the datalogger calibration.

For the case of the datalogger with a resolution of 0.1 °C, the expanded uncertainty associated with the calibration of the system (thermometer + datalogger) is 0.06 °C (k = 2).

Then, the uncertainty component due to the calibration of the systems thermometer+ datalogger is derived from (5):

$$u(t_{i,calibration}) = \frac{U_{i,calibration}(k = 2)}{k} \tag{5}$$

3.6.3. Uncertainty Contribution Due to the Drift of the Thermometers, $u(\delta t_{ref,drift})$, $u(\delta t_{i,drift})$

This uncertainty component was evaluated by the difference between the calibrations performed just before and just after the field campaign. The analysis of these calibrations gave the drift value of 0.02 °C for all the Pt-100 and also for the thermometer associated with datalogger B. Despite some thermometers showing a drift of 0.01 °C, in this analysis, the drift value of 0.02 °C was chosen to simplify the calculations. In the case of datalogger A, with a resolution of 0.1 °C, a drift lower than 0.02 °C was derived for all the calibration temperatures, with the exception of −25 °C, where a drift of 0.1 °C was observed. In order to simply the calculations, a drift of 0.1 °C is considered for the system thermometer + datalogger A. The corresponding probability distribution was considered as the rectangular one, and the uncertainty component is calculated by (6)

$$u\left(t_{i,drift}\right) = \frac{thermometer\ i\ drift}{\sqrt{3}} \tag{6}$$

3.6.4. Uncertainty Contribution Due to the Readings during the Field Campaign, $u(\delta t_{ref,reading\ system})$, $u(\delta t_{i,reading\ system})$

This uncertainty component was evaluated by the combination of the system (thermometer + datalogger) resolution and the stability of the system (thermometer + datalogger). The resolution of the systems is 0.01 °C, with the exception of the datalogger A, with a resolution of 0.1 °C. The repeatability of the different systems is evaluated by analyzing the different behavior of the thermometers under the same field exposure conditions. For this purpose, the measurements taken during polar nights are considered, mainly, the

dispersion of the all measurements for polar nights. Both components with a rectangular distribution are described in Equation (7):

$$u^2\left(\delta t_{i,reading\ system}\right) = u^2(\delta t_{resolution}) + u^2(\delta t_{repeat.}) = \left(\frac{resolution}{\sqrt{12}}\right)^2 + \left(\frac{repeatability}{\sqrt{12}}\right)^2 \tag{7}$$

3.6.5. Uncertainty Contribution Due to the Thermal Homogeneity of the Comparison Field, $u(\delta t_{thermal\ homogeneity\ of\ the\ compariosn\ field})$

This uncertainty component was evaluated by the differences between the median values corresponding to a shield at pole 1 and the shield at pole 18, which are both of the same model (Table 4). The dispersion of all the differences is also considered in this uncertainty component. Both components with a rectangular distribution are described in Equation (8):

$$u^2\left(\delta t_{thermal\ homogeneity\ of\ comparison\ field}\right) = u^2(\delta(t_{shield\ 1} - t_{shield\ 18})) + u^2\left(\delta t_{dispersion}\right) =$$
$$\left(\frac{t_{shield\ 1} - t_{shield\ 18}}{\sqrt{12}}\right)^2 + \left(\frac{\delta t_{dispersion}}{\sqrt{12}}\right)^2 \tag{8}$$

The final quantification of the expanded uncertainties is included in Table 5. This quantification is based on the data for daily maximum temperatures and for the days with daily mean solar irradiance higher than the values included in Table 4.

Table 5. Final expanded uncertainty ($k = 2$) of the Table 4 values, the median of the differences between each of the shields and the reference shield.

	Polar Night	v < 2 m/s	v ∈ [2, 5) m/s	v ≥ 5 m/s
U(P2-ref. shield), $k = 2$ (°C)	0.22	0.76	0.34	0.36
U(P4/own-ref. shield), k = 2 (°C)	0.66	0.64	0.73	0.65
U(P4/01-ref. shield), $k = 2$ (°C)	0.66	0.64	0.73	0.65
U(P5, P16-ref. shield), $k = 2$ (°C)	0.31	0.41	0.39	0.31
U(P6-ref. shield), $k = 2$ (°C)	0.22	1.1	0.54	0.82
U(P8-ref. shield), $k = 2$ (°C)	0.20	0.20	0.24	0.17
U(P9-ref. shield), $k = 2$ (°C)	0.27	0.31	0.33	0.26
U(P10-ref. shield), $k = 2$ (°C)	0.63	0.68	0.65	0.61
U(P11/own-ref. shield), $k = 2$ (°C)	0.62	1.2	1.0	0.58
U(P12-ref. shield), $k = 2$ (°C)	0.46	1.0	0.60	0.65
U(P13/own-ref. shield), $k = 2$ (°C)	0.58	0.66	0.69	0.68
U(P14-ref. shield), $k = 2$ (°C)	0.58	0.50	0.56	0.58
U(P15-ref. shield), $k = 2$ (°C)		0.87	0.75	0.91
U(P17/own-ref. shield), $k = 2$ (°C)	0.35	0.46	0.44	0.33
U(P18-ref. shield), $k = 2$ (°C)	0.31	0.30	0.35	0.27
U(SS-ref. shield), $k = 2$ (°C)	0.30	0.59	0.33	0.30

4. Conclusions

A field intercomparison of thermometer radiation shields was conducted at NY-Ålesund, Norway (78°55′00″ N 11°56′00″ E) for around 14 months, covering the different meteorological conditions in such an extreme environment. Different designs of radiation shields were included in this intercomparison. The air temperature measurements were performed by the same model of resistance thermometers, Pt-100, all of them connected to fluke dataloggers. In addition, complete systems (radiation shield + thermometer + datalogger) of a unique manufacturer were also part of this intercomparison. All the thermometers were calibrated just before and just after the field campaign.

Overall, 94.5% of the recorded data passed the quality control test defined in this work, assuring the robustness of the conclusions here presented.

A reference shield for this intercomparison was selected. Despite it being recommended that reference shields are actively ventilated [27], this exercise showed a natural ventilated shield with the best performance under the specific climatic conditions in the Arctic.

The behavior of each shield against the reference one was studied for daily maximum and minimum temperatures, and it was concluded that the difference of the shields to the reference increases with solar irradiance and decreases with wind speed. The response time of each of the shields was also analyzed, reaching the conclusion that the sampling time is not low enough to deliver reliable statements. The impact of the ground surface state on the shields' performance was analyzed, reaching the conclusion that the ground covered by snow does not have a high impact on the performance of the shields, probably due to the low solar elevation angle, which is typical in the Arctic. This low elevation angle means that the solar radiation reflected by the snow/ground does not reach easily the inner of the shields. On the contrary, the low solar elevation angle increases the direct solar radiation impact on the shields.

Finally, the difference between the shields and the reference one was evaluated for the conditions when the highest divergence between shields is expected: days with high solar irradiance, at daily maximum temperatures and at different wind speeds. In addition, the impact of the residual solar radiation inside the shields was also estimated by the use of high emissivity Pt-100 (black). The quantification of the difference between shields was performed in terms of the median of the measurements taken under the extreme conditions already explained. The maximum difference (1.29 °C) was found for the shield at pole 6 and for wind speeds lower than 2 m/s. This difference decreases with wind speed, reaching a value of 0.29 °C when the wind speed is higher than 5 m/s. The same dependence between solar irradiance impact and wind speed was observed in other shields. On the contrary, other shields, such as the ones at pole 8, pole 9, and the Stevenson screen, show a poor dependence of solar irradiance and of wind speed.

The uncertainty associated with the quantification of the difference between shields was also evaluated, where the dispersion of the measurements is the highest uncertainty source. The calibration and drift of the thermometers were also considered as well as their reproducibility by the dispersion of the measurements at polar nights (events without solar irradiance). The uncertainties range from 0.27 to 1.2 °C, and as a general rule, the uncertainty decreases with wind speed.

References

1. Standing Committee on Measurements, Instrumentation and Traceability (SC-MINT). Available online: https://community.wmo.int/en/governance/commission-membership/INFCOM/Officers/Management-Group/SC-MINT (accessed on 13 May 2024).
2. Commission for Instruments and Methods of Observations. 17th Session, Amsterdam, 12–16 October 2018. WMO-n° 1227. Available online: https://library.wmo.int/records/item/56292-commission-for-instruments-and-methods-of-observation?offset=4 (accessed on 13 May 2024).
3. Lacombe, M.; Bousri, D.; Leroy, M.; Mezred, M. *Instruments and Observing Methods. Report n° 106*; WMO Field Intercomparison of Thermometer Screens/Shields and Humidity Measuring Instruments: Geneva, Switzerland, 2011.
4. Van der Meulen, J.P.; Brandsma, T. Thermometer screen intercomparison in De Bilt (The Netherlands), Part I: Understanding the weather-dependent temperature differences. *Int. J. Climatol.* **2008**, *28*, 371–387. [CrossRef]
5. Brandsma, T.; Van der Meulen, J.P. Thermometer screen intercomparison in De Bilt (The Netherlands), Part II: Description and modelling of mean temperature differences and extremes. *Int. J. Climatol.* **2008**, *28*, 389–400. [CrossRef]
6. de Haij, M.; Bijma, J.; Proksch, N. (Too) Hot or Not? Field Experiment with the KNMI Thermometer Screen. Available online: https://www.knmi.nl/kennis-en-datacentrum/publicatie/too-hot-or-not-field-experiment-with-the-knmi-thermometer-screen (accessed on 13 May 2024).
7. Kaspar, F.; Hannak, L.; Screiber, K.-J. Climate Reference Stations in Germany: Status, Parallel Measurements and Homogeneity of Temperature Time Series. Available online: https://www.adv-sci-res.net/13/163/2016/asr-13-163-2016.pdf (accessed on 13 May 2024).
8. Hoover, J.; Yao, L. Aspirated and non-aspirated automatic weather station Stevenson screen intercomparison. *Int. J. Climatol.* **2018**, *38*, 2686–2700. [CrossRef]
9. Available online: https://www.euramet.org/research-innovation/search-research-projects/details/?tx_eurametctcp_project[project]=1679 (accessed on 13 May 2024).
10. Merlone, A.; Lopardo, G.; Sanna, F.; Bell, S.; Benyon, R.; Bergerud, R.A.; Bertiglia, F.; Bojkovski, J.; Böse, N.; Brunet, M.; et al. The MeteoMet project–metrology for meteorology: Challenges and results. *Meteorol. Appl.* **2015**, *22*, 820–829. [CrossRef]
11. Merlone, A.; Sanna, F.; Beges, G.; Bell, S.; Beltramino, G.; Bojkovski, J.; Brunet, M.; Del Campo, D.; Castrillo, A.; Chiodo, N.; et al. The MeteoMet2 project–Highlights and results. *Meas. Sci. Technol.* **2018**, *29*, 025802. [CrossRef]
12. Annex 1G of the WMO GIMO Guide to Instruments and Methods of Observation, (WMO-No. 8), 2023 Edition. Available online: https://library.wmo.int/viewer/68695/?offset=1#page=1&viewer=picture&o=bookmark&n=0&q= (accessed on 13 May 2024).
13. Maturilli, M.; Herber, A.; König-Langlo, G. Surface radiation climatology for Ny-Ålesund, Svalbard (78.9° N), basic observations for trend detection. *Theor. Appl. Climatol.* **2015**, *120*, 331–339. [CrossRef]
14. Mazzola, M.; Viola, A.P.; Lanconelli, C.; Vitale, V. Atmospheric observations at the Amundsen-Nobile Climate Change Tower in Ny-Ålesund, Svalbard. *Rend. Fis. Acc. Lincei* **2016**, *27* (Suppl. S1), 7–18. [CrossRef]
15. Norsk Klima Service Center. Available online: https://seklima.met.no/ (accessed on 13 May 2024).
16. Astronomical Calculations: Solar Coordinates. Available online: https://squarewidget.com/solar-coordinates/ (accessed on 13 May 2024).
17. NOAA Solar Calculator. Available online: https://gml.noaa.gov/grad/solcalc/ (accessed on 13 May 2024).
18. WMO-N°8 WMO Guide to Meteorological Instruments and Methods of Observations, Part I, Chapter I, Annex 1D, 2021 Edition. Available online: https://library.wmo.int/idurl/4/68695 (accessed on 25 January 2024).
19. Coppa, G.; Quarello, A.; Steeneveld, G.; Jandrić, N.; Merlone, A. Metrological evaluation of the effect of the presence of a road on near-surface air temperatures. *Int. J. Clim.* **2021**, *41*, 3705–3724. [CrossRef]
20. Garcia Izquierdo, C.; Coppa, G.; Hernández, S.; Merlone, A. Metrological Evaluation of the Building Influence on Air Temperature Measurements. *Atmosphere* **2024**, *15*, 209. [CrossRef]
21. *EN 60751:2008*; Industrial Platinum Resistance Thermometers and Platinum Temperature Sensors. IEC: Geneva, Switzerland, 2007.
22. Nicholas, J.V.; White, D.R. Traceable Temperatures: An introduction to temperature measurement and calibration, 2nd ed. *Meas. Sci. Technol.* **2002**, *13*, 1651. [CrossRef]
23. Guide to the Realization of the ITS-90.BIPM/CCT:2018. Available online: https://www.bipm.org/documents/20126/41773843/Guide_ITS-90_5_SPRT_2021.pdf/c4bbbe56-4118-eef7-47cb-3ea234db40b8 (accessed on 25 January 2024).
24. Izquierdo, C.G.; Hernández, S.; González, A.; Matias, L.; Šindelářová, L.; Strnad, R.; del Campo, D. Evaluation of the self-heating effect in a group of thermometers used in meteorological and climate applications. *Meteorol. Appl.* **2019**, *26*, 117–129. [CrossRef]
25. *JCGM 200:2012*; International Vocabulary of Metrology—Basic and General Concepts and Associated Terms (VIM). 3rd ed. JCGM 200:2008 with Minor Corrections. Joint Committee for Guides in Metrology: Sèvres, France, 2012. Available online: http://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf (accessed on 25 January 2024).
26. *JCGM 100:2008*; Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement. Joint Committee for Guides in Metrology: Sèvres, France. Available online: http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf (accessed on 25 January 2024).
27. *ISO 17714:2007*; Meteorlogy-Air Temperature Measurements—Test Methods for Comparing the Performance of Thermometer Shields/Screens and Defining Important Characteristics. ISO: Geneva, Switzerland, 2007.

28. Buisan, S.T.; Azorin-Molina, C.; Jimenez, Y. Impact of two different sized stevenson screens on air temperature measurements. *Int. J. Climatol.* **2015**, *35*, 4408–4416. [CrossRef]

29. Perry, M.C.; Prior, M.J.; Parker, D.E. An assessment of the suitability of a plastic thermometer screen for climatic data collection. *Int. J. Climatol.* **2007**, *27*, 267–276. [CrossRef]

30. Harrison, R.G. Natural ventilation effects on temperatures within Stevenson screens. *Q. J. R. Meteorol. Soc.* **2010**, *136*, 253–259. [CrossRef]

31. Harrison, R.G. Lag-time effects on a naturally ventilated large thermometer screen. *Q. J. R. Meteorol. Soc.* **2011**, *137*, 402–408. [CrossRef]

32. Harrison, R.G.; Burt, S.D. Quantifying uncertainties in climate data: Measurement limitations of naturally ventilated thermometer screens. *Environ. Res. Commun.* **2021**, *3*, 061005. [CrossRef]