

Authors' Response to Reviews of

Assessing the sea ice microwave emissivity up to submillimeter waves from airborne and satellite observations

Nils Risse, Mario Mech, Catherine Prigent, Gunnar Spreen, and Susanne Crewell
The Cryosphere, <https://doi.org/10.5194/egusphere-2024-179>

RC: *Reviewers' Comment*, AR: Authors' Response, Manuscript Text

1. RC1, Dr. Tim Hewison

1.1. General comment

RC: *The manuscript presents a valuable analysis of sea ice emissivity measurements, including novel observations at submillimeter wavelengths, which will be of considerable interest to the remote sensing community, given the forthcoming AWS, ICI and Sterna satellite missions. Importantly it relates the variability on the different scales observed by airborne and satellite sensors. The authors could consider further analysis to include a quantification of the scene variability on other scales - e.g. through the use of variograms/structure functions. This could extend the application of the results to other applications. It is generally well-written and the conclusions, in particular, are clear. However, there are several cases where key details of the methodology are missing, which would make it very difficult to reproduce the results. Examples are given below. Furthermore, more attention needs to be paid to uncertainties - especially those introduced by various assumptions (see below).*

AR: The authors would like to thank Dr. Tim Hewison for providing highly valuable and constructive feedback on this manuscript. We have carefully considered all the comments and provided responses below. The revised version provides further details on the methodology to facilitate the reproducibility of the results. Referee comments are given in bold text.

AR: Variograms (or semi-variograms; e.g., Mälicke et al., 2020) or structure functions (e.g., Kitchen, 1989) describe the spatial dependence as a function of separating distance. Variograms are thus conducted with data of similar spatial resolution, e.g., MiRAC or satellites. They provide the basis for spatial interpolation, which we did not aim for in this work. However, we agree that the parameters of a variogram (nugget, sill, effective range) and their dependence on environmental conditions (e.g. for soil moisture measurement networks; Lakhankar et al., 2010) or frequency in our case could provide new information, but such an analysis was out of the scope of this study. Instead, we focused on the emissivity variability as a function of footprint size. This allows us to bridge from small-scale airborne observations to large-scale satellite observations. Hence, it also provides a representation of the underlying spatial covariance described by a variogram. A key difference is that this approach does not use a distance but a radius, and we do not quantify the deviations from the central location but average them and compute the variability over all samples. Hence, we change spatial resolution with increasing radius that can be compared to satellite observations, e.g., 5 or 16 km. However, to allow further analysis of the data using other geostatistical methods, such as variograms, we provide the derived emissivities on PANGAEA.

Kitchen, M. (1989), Representativeness errors for radiosonde observations. *Q.J.R. Meteorol. Soc.*, 115: 673-700. <https://doi.org/10.1002/qj.49711548713>

Lakhankar T, Jones AS, Combs CL, Sengupta M, Vonder Haar TH, Khanbilvardi R. Analysis of large scale spatial variability of soil moisture using a geostatistical method. *Sensors (Basel)*. 2010;10(1):913-32. doi: 10.3390/s100100913. Epub 2010 Jan 25. PMID: 22315576; PMCID: PMC3270877.

Mälicke, M., Hassler, S. K., Blume, T., Weiler, M., and Zehe, E.: Soil moisture: variable in space but redundant in time, *Hydrol. Earth Syst. Sci.*, 24, 2633–2653, <https://doi.org/10.5194/hess-24-2633-2020>, 2020.

1.2. Line 106

RC: *Please provide a reference to a document describing the calibration and bias correction procedures.*

AR: We added two references to other work that describe the calibration and bias correction procedures. The reference Mech et al. (2019) describes the calibration procedure, including internal calibration and absolute calibration. The bias correction procedure follows the approach that Konow et al. (2019) applied to airborne observations of the HAMP radiometer onboard the HALO research aircraft, which we transfer to the Polar 5 aircraft.

Konow, H., Jacob, M., Ament, F., Crewell, S., Ewald, F., Hagen, M., Hirsch, L., Jansen, F., Mech, M., and Stevens, B.: A unified data set of airborne cloud remote sensing using the HALO Microwave Package (HAMP), *Earth Syst. Sci. Data*, 11, 921–934, <https://doi.org/10.5194/essd-11-921-2019>, 2019.

The instrument receivers were calibrated with a two-point calibration using liquid nitrogen and an internal target at the beginning of each campaign. In addition, MiRAC-A performed gain calibrations every 15 min, and MiRAC-P every 20 min during flights using an internal target ([Mech et al., 2019](#)). After the campaign, we applied a bias correction of the 89 GHz TBs [following Konow et al. \(2019\)](#) based on Passive and Active Microwave Radiative Transfer (PAMTRA; Mech et al., 2020) forward simulations by using dropsonde profiles under clear-sky conditions over the open ocean extended by ERA5 reanalysis (Hersbach et al., 2020) to the top of the atmosphere and a sea surface temperature analysis (UK Met Office, 2012) as input.

1.3. Line 116

RC: *Does this uncertainty combine systematic and random effects? It is important that they can be treated separately in evaluating the uncertainty of the emissivity estimates on different spatial scales.*

AR: Yes, the uncertainty of the measured TB combines systematic and random effects. The random noise used in our work is an upper bound. We estimate it from 20 consecutive 30 s measurement intervals over open ocean under clear sky during ALOUD RF10. The TB standard deviation averaged over all 20 intervals ranges from 0.2 to 0.3 K depending on the channel. The bias correction is performed with a limited number of clear-sky dropsonde observations that have a random and systematic uncertainty (relative humidity, temperature, surface temperature, surface wind speed, forward simulation, ...). Therefore, the reference itself is very likely biased, which we account for by increasing the 89 GHz TB error by 2 K. In our study, we treat both components of the TB error as systematic and do not separate them when averaging multiple observations. Generally, it is correct that random TB noise would cancel out when averaging multiple observations. Our approach aims at reducing the complexity of the emissivity error estimation while still providing a robust quantitative estimate. A separation of random and systematic errors would have a negligible effect on the emissivity error because of the much smaller contribution of random TB errors compared to systematic error sources that affect the emissivity, i.e., surface temperature, atmospheric temperature, relative humidity, and TB bias correction (89 GHz channel). Our final uncertainty represents a confidence interval dominated by systematic

uncertainty. We will add this information and a short discussion to the manuscript.

The TB noise is about 0.5 K for MiRAC-A (Küchler et al., 2017) and -P (Mech et al., 2019), [which is an upper bound of the observed TB noise of 0.2 to 0.3 K depending on the channel from a homogeneous time series during ACLOUD RF10. This random noise cancels out when averaging, but we do not consider this here as systematic effects dominate the overall emissivity uncertainty \(see Sect. 3.2\).](#) Hence, we assume the overall TB uncertainty from bias correction and noise to be 2.5 K at 89 GHz and 0.5 K at all other frequencies.

1.4. Line 153

RC: *What uncertainty is added by this assumption? Could it introduce significant biases? E.g. during strong surface inversions - in this case, would it be better to assume a linear change from the flight level to the surface?*

AR: We now added information on the lowest flight altitude to the paragraph to clarify that it is just about 100 m and not 3 km above ground. The quantification of emissivity uncertainty due to atmospheric temperature and humidity profiles described in the manuscript also includes the near-surface atmosphere. Therefore, it also includes extrapolation uncertainties close to the ground.

Tjernström, M. and Graversen, R.G. (2009), The vertical structure of the lower Arctic troposphere analysed from observations and the ERA-40 reanalysis. *Q.J.R. Meteorol. Soc.*, 135: 431-443. doi: 10.1002/qj.380

We assume constant temperature and humidity from the lowest flight altitude [of about 100 m](#) down to the surface if no dropsonde information is available over sea ice. [The air temperature measured at these heights differs less than 5 K from the mean surface temperature, which indicates that the profiles capture typical Arctic surface temperature inversions \(e.g., Tjernström and Graversen, 2009\).](#)

1.5. Line 158

RC: *What infrared emissivity is assumed to open water? Is this a function of sea state?*

AR: We assume an infrared emissivity of 0.995 for both sea ice and water, close to the constant value of 0.996 in Hoyer et al. (2017) and Thielke et al. (2022) for mixed sea ice and ocean surfaces. The emissivity of 0.995 represents the observed infrared emissivity of various ice types and sea water within the KT-19 band pass from 9.6 to 11.5 μm (Hori et al., 2006). We do not vary the IR emissivity of sea water along the flight track due to discontinuous information on the presence of water within the radiometer footprint and a lack of accurate sea water state description. This approach aligns with previous studies on sea ice emissivity that observed mixtures of open water and sea ice (e.g., Hewison and English, 1999). We clarified the approach in the revised manuscript.

Høyer, J. L., Lang, A. M., Tonboe, R., Eastwood, S., Wimmer, W., and Dybkjær, G. Report from Field Inter-Comparison Experiment (FICE) for ice surface temperature. Danish Meteorological Institute (2017).

Thielke, L., Huntemann, M., Hendricks, S., Jutila, A., Ricker, R., and Spreen, G. (2022). Sea ice surface temperatures from helicopter-borne thermal infrared imaging during the MOSAiC expedition. *Sci. Data*, 9, 364. doi:10.1038/s41597-022-01461-9

The infrared TB is converted to surface skin temperatures with an infrared emissivity of 0.995 [similar to Hoyer et al. \(2017\) and Thielke et al. \(2022\)](#), which approximates the infrared emissivity of typical sea ice types [and ocean](#) with an accuracy of 0.01 to 0.02 (Hori et al., 2006).

1.6. Line 163

RC: *Is this bias compensated for? How is it accounted for in the uncertainty analysis?*

AR: The bias between KT-19 and the thermal infrared-based satellite reanalysis by Nielsen-Englyst et al. (2023) is not compensated for or directly considered in the uncertainty analysis. We did not apply an offset to NE23 to match KT-19. This results in a potential overestimation of the 89 GHz emissivities from satellites compared to MiRAC during ACLOUD, where NE23 temperatures are 4 to 6 K lower than KT-19. However, this offset would not lead to a relevant change of the spectral variability that we observe. The emissivity change can be estimated directly from our uncertainty analysis. The surface temperature error of ± 3 K for ACLOUD results in 2 % emissivity uncertainty. If we multiply this by a factor of 2, we find that the offset between MiRAC and satellites is about 4 % or 0.03 in emissivity. This is equivalent to the 89 GHz uncertainty during AFLUX for MiRAC. Therefore, we do not shift the NE23 estimates to match KT-19. We included the bias between NE23 and KT-19 in the detailed discussion of the spectral emissivity variations in Sect. 5.3

Low spectral differences occur during ACLOUD near nadir from 89 to 243 GHz (MHS and MiRAC) and at vertical polarisation from 91 to 150 GHz (SSMIS; Fig. 8a). [The higher satellite emissivity can be explained by the underestimation of the NE23 skin temperature compared to KT-19.](#)

1.7. Line 165

RC: *(1) It would be helpful to include a short summary of how the NE23 analysis is derived. (2) It should also be clarified whether the NE23 analysis is used instead of the KT-19 measurements for emissivity calculations airborne as well as satellite measurements, and, if so, why, and (3) how this contributes to the overall uncertainty - especially in the context of surface temperature gradients through sea ice and any overlying snow layers.*

AR: Response to (1): The NE23 product obtains daily sea and sea ice surface temperatures from clear-sky thermal infrared satellite observations derived by optimal interpolation. The reference contains further details on the included sensors. During ACLOUD and AFLUX, the product uses the operational OSISAF IST product (OSI-205; Dybkjaer et al., 2018) based on 11 and 12 μm AVHRR observations and an AVHRR-based cloud mask. Sea ice concentration fields from OSISAF are used to separate sea and sea ice temperatures. We summarized this in the revised manuscript.

Dybkjaer, G., Eastwood, S., Borg, A.L., Høyer, J.L., Tonboe, R., 2018. Algorithm theoretical basis document (ATBD) for the OSI SAF Sea and sea ice surface temperature L2 processing chain. OSI205a and OSI205b.

AR: Response to (2): The NE23 is only used for satellite, not airborne observations. We clarified this in Sect. 2.4. This is a natural choice as KT-19 measures simultaneously to MiRAC at a comparable spatial resolution. On the other hand, KT-19 is not used for satellites because the NE23 product observes the sea ice at a similar spatial resolution (about 5 km) as the satellites and provides better spatial coverage than KT-19. We will mention this in the revised manuscript.

AR: Response to (3): Using two different data sets comes at a cost of biases between both propagating into the emissivity estimates. Lower temperature biases occur during AFLUX (similar to 1.5 K found for KT-19

observations during IceBridge flights described in the reference publication), and higher temperature biases occur during ACLOUD. These higher biases during ACLOUD affect the comparison of satellite and airborne emissivities shown in Fig. 8. The satellite emissivities might be higher than the airborne emissivities due to the lower NE23 surface temperature (see the reply on line 163 and update of the manuscript). Generally, we expect low systematic differences between both estimates because they rely on the same measurement principle. They are not affected by surface temperature gradients through sea ice and any overlying snow layers because they measure only the upper few millimeters of the snow or ice (Warren, 1982).

Warren, S. G. (1982), Optical properties of snow, Rev. Geophys., 20(1), 67–89, doi: 10.1029/RG020i001p00067.

We use KT-19 as input to the sea ice emissivity calculation for MiRAC. We also require an accurate description of the surface temperature at the satellite footprint scale with higher spatial coverage than KT-19. Therefore, we use the daily Level 4 Arctic sea and ice surface temperature reanalysis with a resolution of $0.05 \times 0.05^\circ$ (Nielsen-Englyst et al., 2023), which matches the AMSR2 satellite footprint size, hereafter referred to as NE23. The product derives daily gap-free sea and ice surface temperatures from clear-sky thermal infrared satellite observations sensitive to the upper few millimeters of the snow or ice (Warren, 1982) and passive microwave-based sea ice concentration.

We use the nearest NE23 ice surface temperature pixel to the satellite footprint as input to the sea ice emissivity calculation [for satellites](#).

1.8. Line 169

RC: *Is there any potential here for confirmation bias? I.e. if the multiyear ice concentration maps are derived based on satellite observations, and assumed emissivity spectra?*

AR: Yes, there might be some potential for confirmation bias when using multiyear ice concentration from passive and active microwave observations to explain signatures from airborne passive microwave observations. The product from the University of Bremen incorporates additional information to minimize the risk of false interpretation of similar microwave signatures of both major ice types. These are temperature-dependent corrections and ice drift corrections (Melsheimer and Spreen, 2023). Therefore, this product, in combination with visual airborne imagery, provides a robust approach to classify the observed ice floes in the marginal sea ice zone. Furthermore, different microwave frequencies are used to create the product compared to our airborne observations, i.e., 5.3 GHz for ASCAT (VV polarization, normalized to 40°) and AMSR2 (19 and 37 GHz, V and H polarization, 55° ; Melsheimer and Spreen, 2022). We included this explanation in the revised manuscript.

Although the multiyear ice concentration product incorporates microwave observations of AMSR2 and ASCAT that might correspond to those at MiRAC frequencies, the implemented temperature and drift corrections increase independence between multiyear ice concentration and MiRAC TB.

1.9. Line 177

RC: *What variability is typically observed within the $\pm 2h$ window?*

AR: The temporal variability is hard to quantify for the swath data, where footprints are always at slightly different locations. Therefore, we opt for a qualitative discussion of this by showing the time series of the 89 GHz TB from MHS collocated with MiRAC's 89 GHz channel as an example (Fig. R1). The same analysis

was performed for the other satellite instruments, and similar results were obtained. We generally find low temporal variability within a 4-hourly time window, especially for AFLUX, where flights cover a small domain. Higher variability occurs during ACLOUD RF23 for MHS. This flight covers a large area, which causes an artificial temporal drift. We added the outcome of this analysis to the revised manuscript. The paragraph also includes an analysis of the sea ice drift based on RC2 on line 178.

We ensure simultaneous observations by filtering collocations within a ± 2 h window, which maximizes the number of satellite overpasses and minimizes the effects of sea ice drift. [The sea ice drifts less than 2.5 km within 2 h in the study area during flight days based on National Snow and Ice Data Center \(NSIDC\) sea ice drift data \(Tschudi et al., 2020\). Moreover, the relatively long time window is not problematic as the analysis of satellite data reveals much higher spatial than temporal variability \(not shown\).](#)

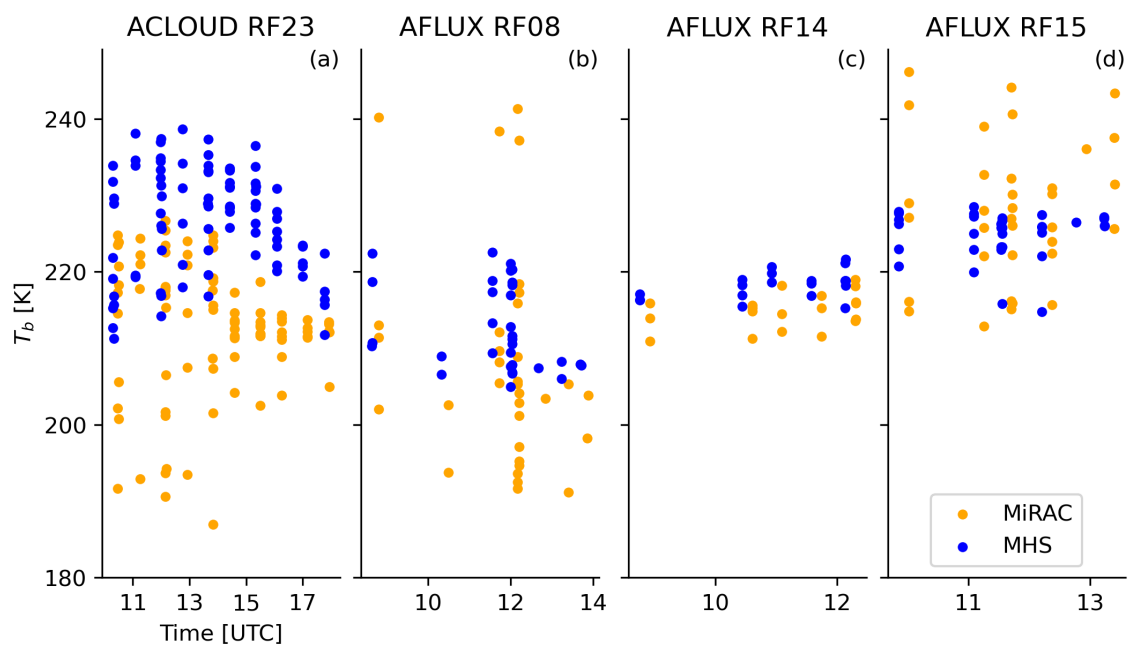


Figure R1: Temporal variability of the TB from MHS and MiRAC resampled to the MHS footprints at 89 GHz for all flights where MiRAC provided 89 GHz TBs. MiRAC and satellites are separated by 10 min for better visibility.

1.10. Line 205

RC: *It should be clarified whether this transmissivity refers to the layer between the aircraft and the surface. And what about atmospheric emission?*

AR: We included the information that transmissivity refers to the layer between the aircraft and the surface. The atmospheric emission is contained in the final term of Eq. (3), i.e., the upwelling atmospheric radiation at the observation height T_b^\uparrow .

The TB observed at aircraft or satellite height, denoted as T_b , is given by

$$T_b = T_s \cdot e \cdot t + T_b^\downarrow \cdot t \cdot (1 - e) + T_b^\uparrow, \quad (1)$$

with the surface emissivity e , surface temperature T_s , atmospheric transmissivity in viewing direction [between the surface and the aircraft or satellite height](#) t , downwelling atmospheric radiation at the surface T_b^\downarrow , and upwelling atmospheric radiation at the observation height T_b^\uparrow .

1.11. Line 221

RC: *It would still be interesting to compare the results for all 183GHz channels - albeit with increased uncertainties.*

AR: We agree that comparing emissivities at all 183 GHz channels is interesting. Therefore, we showed all 183 GHz channels that satisfy the surface sensitivity criterion in Fig. 3 of the manuscript. The figure shows that these channels provide similar emissivities, with slightly lower emissivities at the 183 ± 2.5 GHz channel compared to 183 ± 7.5 GHz channel. However, this difference lies within the uncertainty range and might be caused by the higher water vapor absorption at the inner channel and measurement uncertainties of atmospheric water vapor and temperature. These uncertainties directly propagate into the emissivity. The lack of direct measurements of the downwelling atmospheric radiation and related uncertainties in the atmospheric profile causes higher uncertainties at 183 GHz than during previous studies (e.g., observations with the airborne MARSS radiometer in Hewison et al., 1999, or Wang et al., 2017b). We provide all emissivities and estimated uncertainties in the published emissivity data to allow for different surface sensitivity thresholds for other applications. We justified our decision in the revised manuscript.

Only 183 and 340 GHz observations during ACLOUD lie below the surface sensitivity threshold and are ~~therefore~~ excluded [to avoid highly uncertain emissivity estimates](#).

1.12. Line 225

RC: *This is an underestimate in the case of strong surface inversions, which are common over sea ice.*

AR: See our response to the comment on line 153.

1.13. Line 232

RC: *How could the uncertainties be estimated for satellite observations?*

AR: The same approach can be applied to satellite observations. However, we do not perform this uncertainty estimation as reasoned in the revised manuscript.

The uncertainty estimation is performed only on aircraft and not on satellite observations [because the MiRAC channels already include most satellite channels. A notably higher emissivity uncertainty occurs for satellites near 183 GHz compared to MiRAC due to the higher atmospheric contribution.](#)

1.14. Line 246

RC: *It is interesting to note that the results from the specular assumption appear more Gaussian. Why could that be?*

AR: We visualized the data of the histogram as a function of observed TB and research flight to better answer this question (Fig. R2a and b). The histograms shown in the manuscript are drawn below (Fig. R2c and d). The histogram shapes are partly determined by differences between the research flights and the dependence of the bias on the observed TB (for specular only). The difference between the flights causes the bimodality of the histogram for Lambertian reflection. The reason for the differences between the flights is likely a bias in the atmospheric profile. While the bias for RF08 is almost perfectly centered around 0 K for Lambertian reflection, the other flights show a slight positive bias. The positive bias could be related to an overestimation of the relative humidity or a misrepresentation of its vertical profile.

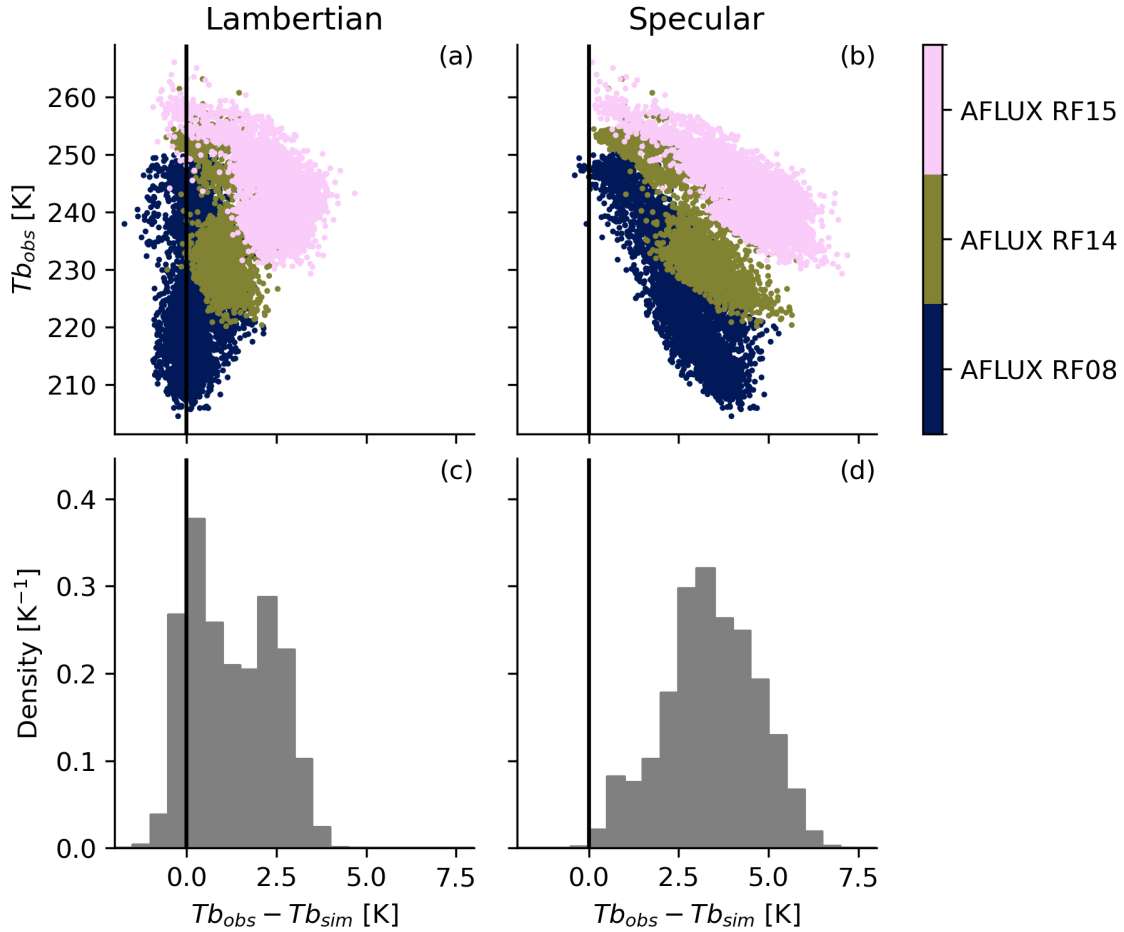


Figure R2: Difference between observed ($T_{b,obs}$) and simulated 183.31 ± 5 GHz TB ($T_{b,sim}$) using 183.31 ± 7.5 GHz emissivities under (a, c) Lambertian and (b, d) specular surface reflection during AFLUX. (a, b) Scatter plot between the difference and the observed TB. (c, d) Histogram of the difference with a TB bin width of 0.5 K.

1.15. Figure 3

RC: *I found this figure confusing - it took several readings before I understood it. It could also be expanded to a full page width.*

AR: This figure highlights the observations for an example flight transect. We will expand it to a full page in the revised manuscript. Also, we made an adjustment based on the second reviewer's comment to improve the visibility in panel (j) and added the new version to the RC2 response document (see comment on Figure 3 therein).

1.16. Table 3

RC: *How exactly is the mean relative uncertainty calculated?*

AR: The mean relative uncertainty is calculated by dividing the uncertainty by the emissivity and averaging this value over all samples. We clarified the table caption.

Values denote the sample count (Cnt.), median (Mdn.), interquartile range (IQR), and ~~mean~~-relative uncertainty averaged over all samples (Unc.).

1.17. Figure 5

RC: *I would also be interested to see the emissivity plotted as spectra for each cluster.*

AR: We agree that a representation as spectra is interesting (see Fig. R3). However, the 89 GHz channel is measuring at a different polarization and angle, so we decided not to include this representation in the manuscript.

1.18. Line 377

RC: *What are the implications of this assumption? Noting the results of Wang et al., 2017b differ from Harlow (2007): and Haggerty and Curry (2001), who found an increase in emissivity with frequency for sea ice between 150 and 220 GHz.*

AR: In the context of our comparison between airborne and satellite observations, we aimed to use a satellite channel less affected by atmospheric water vapor emission and as close to the MiRAC 183 GHz channel as possible. The goal of the sentence is to indicate that spectral emissivity gradients might be present. We have written this more clearly in the revised manuscript.

The near-nadir (0 to 30°) 157 GHz MHS and 165.5 GHz ATMS channels are comparable to MiRAC's nadir 183 GHz channel. We compare these satellite channels instead of the 190.31 and 183.31±7 GHz channels due to their higher surface sensitivity and lower uncertainty, although spectral emissivity gradients might occur (e.g., Hewison et al., 2002). Other channel or instrument combinations differ in incidence angle or polarization, making footprint-level comparisons less meaningful. ~~For the comparison, we assume low emissivity gradients over the 157 to 183 GHz frequency range on satellite scale (Wang et al., 2017b)~~.

1.19. Figure 6b

RC: *This is a very useful result. But would it be better to divide by reflectivity (1-emissivity)? This might normalise the distribution.*

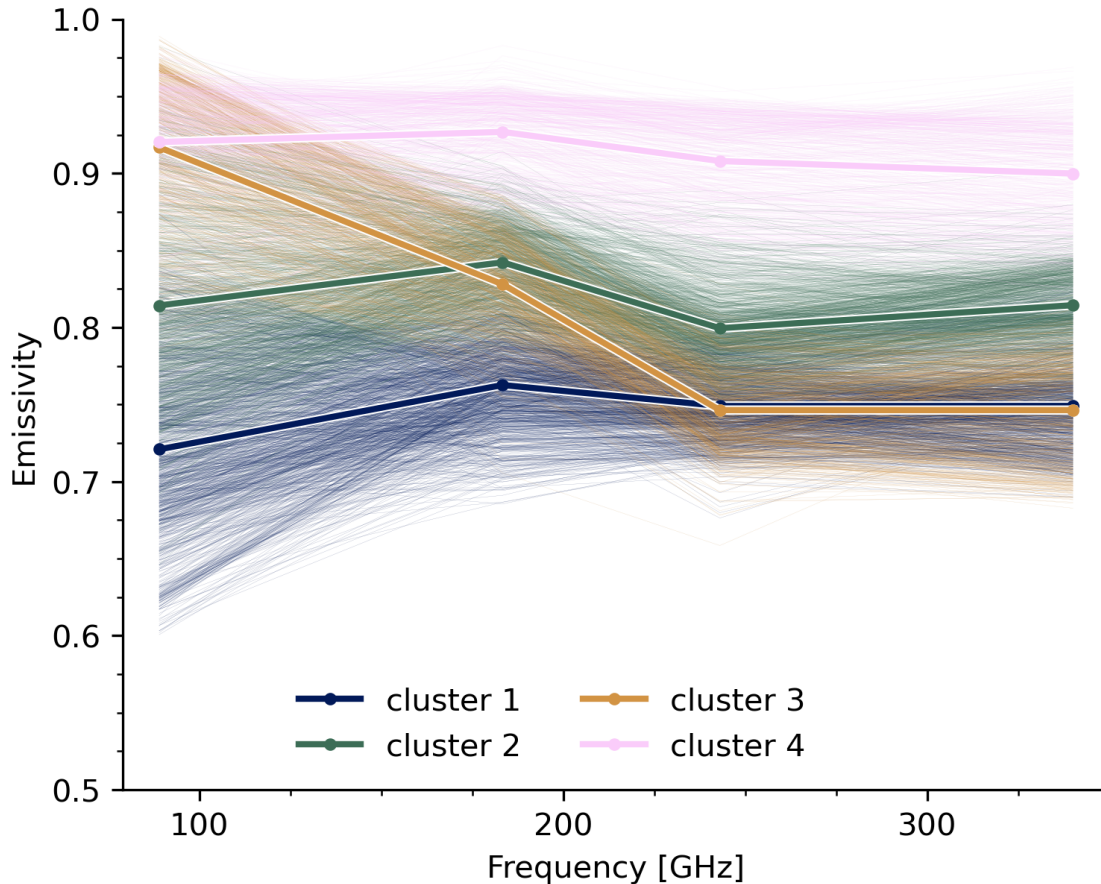


Figure R3: Sea ice emissivity spectra as a function of frequency. Thin lines show all samples and thick lines the cluster mean.

AR: Figure 6b highlights the dependence of the emissivity variability on the footprint size. The interquartile range acts as measure for variability. Dividing this measure by the mean or median reflectivity is more difficult to interpret, i.e., lower values would occur for lower emissivity compared to a higher emissivity with the same underlying interquartile range. This makes it difficult to compare different channels. Therefore, we show the emissivity interquartile range without applying a normalization by reflectivity.

1.20. Line 396

RC: *How exactly are the observations averaged to ensure equal spatial sampling?*

AR: The temporal and spatial collocation criteria are described in Sect. 2.5. Temporal alignment is ensured through the ± 2 h time window and spatial alignment by the sensor-dependent distance threshold and a certain number of MiRAC observations within that distance. We simply average all MiRAC footprints that are collocated with one satellite footprint. We do not perform a weighted average because MiRAC does not cover the entire footprint area, and thus, little or no improvement is expected from a weighted mean. We modified

the sentence by adding a reference to the collocation section and replacing "equal" with "comparable."

The MiRAC observations are averaged to the [collocated](#) footprints of each satellite instrument to ensure ~~equal~~ [comparable](#) spatial sampling ([see Sect. 2.5](#)).

1.21. Figure 7

RC: *How exactly is the MiRAC emissivity resampled to satellite footprints?*

AR: See our response to the comment on line 396.

1.22. Figure 7

RC: *It would also be useful to plot the IQR for the emissivity derived from MiRAC and satellites independently. This figure could also be expanded to full page width.*

AR: The interquartile range is only available for MiRAC. Here, it is used as a measure of hectometer-scale emissivity variability within a single satellite footprint. This is different from Fig. 6, where the IQR was calculated over all samples. We also performed the same calculation here and provided the results in Tab. 4 and 5 for MiRAC (after averaging to satellite footprint), MHS, and ATMS.

1.23. Line 401

RC: *It is not clear how the Lambertian assumption introduces a bias less than 2% for MiRAC.*

AR: The statement is based on the comparison of 89 GHz emissivities under specular and Lambertian reflection (Fig. R4). The figure shows the percentage of specular emissivities exceeding Lambertian emissivities for the different frequencies for ACLOUD and AFLUX. It shows that the difference increases with decreasing emissivity and increasing atmospheric opacity. Similar results are described by Matzler (2005) and Karbou and Prigent (2005). For the reader, we extracted the relevant information from this figure, i.e., the percentage by which a fully specular emissivity would exceed the fully Lambertian emissivity provided in all our figures. This is only done for 89 GHz because we do not have polarization information from space that hints at specular contributions to the reflection and found that Lambertian emissivity is more consistent. However, we made a mistake and extracted only the values for AFLUX, not for ACLOUD. We corrected this in the revised version. For ACLOUD, the uncertainty due to surface reflection could exceed the emissivity uncertainty that we estimated. This is due to the higher optical depth compared to AFLUX. At the same time, it is unlikely that sea ice is purely specular at 89 GHz (e.g., a specular parameter of 0.5 was found over Antarctica in summer at 50 GHz in Guedj et al., 2010, which also shows good results over sea ice in Bormann, 2022) and the true uncertainty is lower. An exact quantification is impossible at this point. We modified line 255 in Sect. 3.3 and line 401 in Sect. 5.3.

Change of line 255:

For MiRAC observations [at 89 GHz](#), fully specular emissivities ~~are about 2% (0.6%) higher than Lambertian emissivities of 0.7 (0.9) at 89 GHz under 25° incidence angle~~ [exceed fully Lambertian emissivities by about 6 to 2% during ACLOUD and 3 to 1% during AFLUX for Lambertian emissivities from 0.6 to 0.8](#). This emissivity uncertainty is in the order of or lower than the uncertainty due to the surface temperature assumption [since sea ice is not fully specular at 89 GHz \(Bormann, 2022\)](#).

Change of line 401:

As expected, the 89 GHz emissivity shows a polarization signal of about 0.1. ~~Hence, the purely Lambertian assumption in our emissivity calculation introduces an emissivity bias less than 2% for MiRAC~~ This indicates a specular contribution to the surface reflection and an underestimation of the emissivity under purely Lambertian reflection at 89 GHz for MiRAC (see Sect. 3.3).

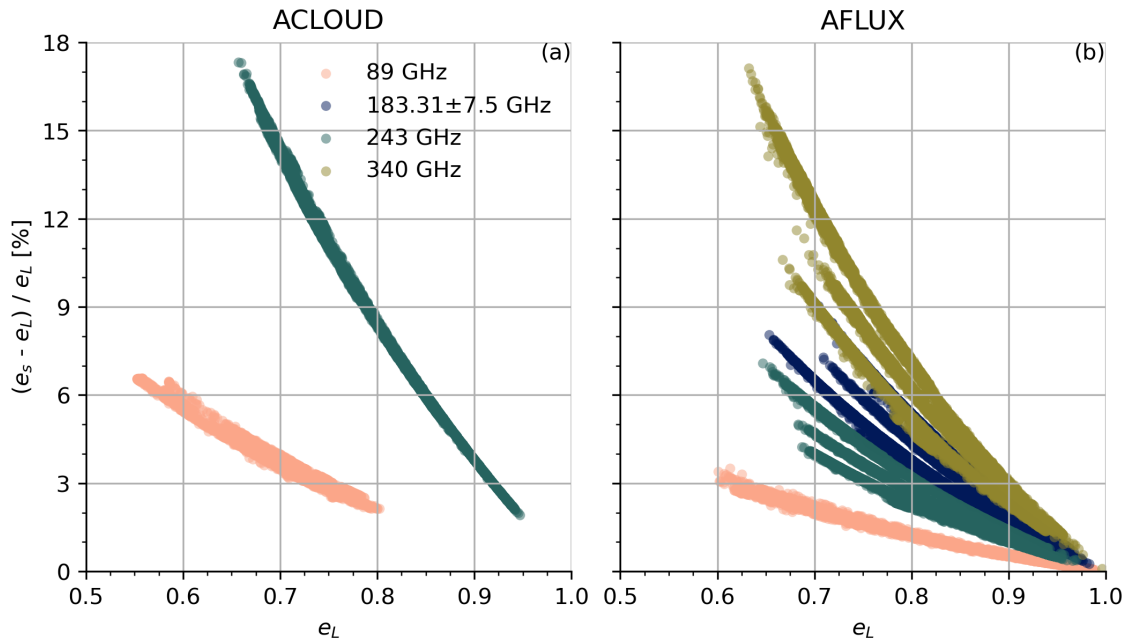


Figure R4: Relative difference between specular and Lambertian emissivity as a function of Lambertian emissivity for (a) ACloud and (b) AFLUX and the four channels.

1.24. Table 4

RC: *Is the relative bias here calculated from the mean difference? It does not seem to match the difference of the median values.*

AR: Yes, the relative bias is calculated from the mean difference normalized element-wise by the MiRAC emissivity. It mostly matches the tendency of the median value, but not always as the emissivity distribution is slightly skewed (see Fig. 8b for MHS and ATMS).

1.25. Figure 8

RC: *This is potentially a very useful figure, but is confusing to interpret - especially the labelling on the x-axis should be improved. It may also help to more clearly distinguish V & H polarisation. It would be better supplemented by a table of values.*

1. Why are multiple values shown for 243GHz (2) 340GHz (3)?
2. Any idea why the 89V results for AMSR2 are out-of-family?

3. *AMSR2 and SSMIS are conical scanners, with V and H polarisations, not QV as shown.*

4. *Why are no ATMS results are given for ACLOUD?*

AR: We updated Fig. 8 of the manuscript. We rotated the x-axis labels and removed the transparency of the count bars. We did not indicate the V/H polarization with, e.g., a hatch pattern because they are already clearly separated at 89 GHz, and it requires too many different patterns to indicate QV, QH, and nadir. Therefore, the polarizations are indicated only on the x-axis. We do not think that values in a table provide additional value to this figure, as variability is a key element that we want to highlight here. The following replies refer to the four specific questions.

1. Multiple boxes are shown for 243 and 340 GHz due to the collocation method that matches MiRAC observations to each satellite observation. One satellite observation consists of a specific footprint and channel. The separation by channel was performed because MiRAC's 89 GHz is not available for low flight altitudes (see Sect. 2.5), and we want to ensure comparable spatial sampling, at least for the same spectral bands, for better comparison. We consider the 243 and 340 GHz channels as part of the >100 GHz band and, therefore, show the same collocation as for the 183 GHz MiRAC channel. Therefore, we also do not show AMSR2 for MiRAC channels above 100 GHz. The distributions for the different sensors are similar, which is an encouraging result. Only for AMSR2, we find differences at 89 GHz due to the low count of AMSR2 footprints during ACLOUD (see question 2).
2. The AMSR2 89V results differ because the sensor only covers the western part of ACLOUD RF23 after applying the collocation method. Only 23 footprints collocate with MiRAC's 89 GHz channels, as indicated by the low count in the bar plot of Fig. 8a. The two overpasses are at 09:46 and 11:25 UTC, respectively. The first MiRAC sample is at 11:35 UTC. SSMIS covers also the eastern part, which has a lower emissivity at vertical polarization.
3. We calculated QV from H and V for AMSR2 and SSMIS at the given incidence angle for comparison with MHS and ATMS, although the incidence angle differs here with comparable footprint sizes. This is mentioned also in line 401.
4. No ATMS results can be shown for ACLOUD because no ATMS footprint passes the collocation criteria. This has been mentioned earlier in line 183. We added a note to the caption of Fig. 8.

The 88 to 92 GHz satellite footprint count might be lower than above 150 GHz because satellite footprints are excluded if the nearest MiRAC channel contains no emissivity. [Note that no ATMS overpass occurred during ACLOUD.](#)

1.26. Line 465

RC: *How much difference is expected from nadir to 25°? I would not expect much, following Hewison and English (1999): Airborne Retrievals of Snow and Ice Surface Emissivity at Millimetre Wavelengths. IEEE Trans. Geosci.Remote Sensing, Vol.37, No.4, 1999, pp.1871-1879, doi:10.1109/36.774700*

AR: The QH emissivity at 89 GHz of Fig. 5 in Hewison and English (1999) shows angular variation for all sea ice types. Between nadir and 25°, this is about 0.02 for nilas and similar for other ice types. We expect similar differences for our airborne observations between nadir and horizontal polarization at 25° because the polarization difference observed by SSMIS and AMSR2 at 89 GHz is about 0.1 near 53° (Fig. 8). Therefore, we consider the different polarization and viewing angle of MiRAC-A and MiRAC-P observations as one of the limitations of this study.

1.27. Line 263

RC: *Typo: Word -> World*

AR: Done.

Notably, surface structural variations from 3 to 4 km suggest the presence of young ice, which defines ice in the transition stage between nilas and first-year ice (~~Word-World~~ Meteorological Organization, 2014), possibly formed within leads among the thicker multiyear ice.

1.28. Line 473

RC: *I suggest “resolve” instead of “capture”, noting it will still affect the mean emissivity.*

AR: Done.

Spatial resolution: MiRAC’s hectometer scale may not ~~capture~~ resolve smaller sea ice features such as ridges or melt ponds, which could influence emissivity.