Author's response to review #1 of research article AMT-2022-314

We would like to thank anonymous referee #1 for their constructive review, which significantly improved the quality of the manuscript.

Answers to specific reviewer comments

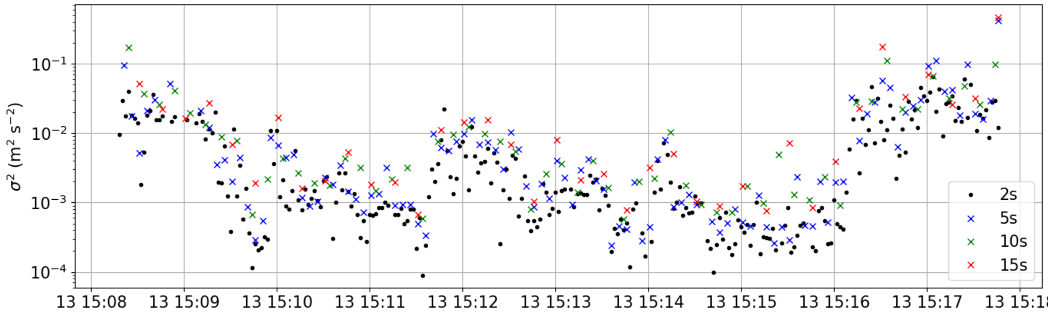
Major comments

1. **Figure 2 shows a 2 s spectrum for pitot and hotwire fluctuations. Is 2 seconds interval spectrum sufficient for such a turbulence structure analysis? For example, a typical ECOR system uses at least 10 mins data.**

   We are aware that with this short averaging time, we exclude larger scales of the energy spectrum and only capture contributions from smaller eddies when calculating variances. The averaging time is always a compromise that is impacted by various practical reasons. In this paper, we were originally using consistent 2s interval lengths for HW calibration, ε calculation, and σ^2 calculation. The consistent interval ensures that we can use ε and σ^2 in the same equation (Eq. 8 in the manuscript) with a common resolution.

   Further reasons for using a relatively short interval are: 1) Shorter intervals provide the highest spatial resolution. The main purpose of our vertical profile measurements is to resolve thin layers. When using longer averaging times, the vertical profiles are blurred; 2) At larger intervals, airspeed variations due to the DH spirals start to be visible; 3) For HW calibration, the calibration coefficient varies a lot and the averaging interval should be as short as possible; 4) For ε, the interval is of minor importance because its value can be estimated from any spectral data in the inertial subrange; 5) For σ^2, the averaging interval has the greatest impact, but as the resulting variance is found to be rather insensitive to the interval (see Fig. 5), the short interval of the other parameters is used for σ^2 as well (for the DH2).

   The main difference to an ECOR system is that we are not aiming to capture low-frequencies in the spectrum, but instead, we aim to resolve a vertical profile and the turbulent exchange between shallow layers in the ABL. The plot copied below shows a time series of estimated variances for different DH2 averaging intervals, showing an expected decrease in resolution with increasing averaging time, yet the magnitude of the variance increases only slightly relative to the range in variance over time (altitude). After careful consideration of all influencing factors, we decided to increase the averaging time to 5s as the best compromise for this work. The impacted parameters throughout the manuscript are changed correspondingly.



We included this discussion in the manuscript in Sect. 2.3.2, line 306ff.

2. **Line 200: How does the author select the value of C? What is the appropriate range of the C values? Similar situation for the appropriate range for the Kolmogorov constant (in line 222). Will you please provide a guideline for those parameters' determination? For example, are those constants unique to the Arctic environment or general cloudy conditions?**

   Thank you for this comment. The constant C has to be determined from observations and depends on stability. The value of C=0.35 of earlier studies was confirmed by the observations in Hanna 1968, which encompass a large variety of (stability) conditions. We revisited the literature and decided to use the value of C=0.41 instead, which seems to be more suitable for stable conditions (Lee 1996). We've revised the calculations and updated the text passage (line 206ff).

   The Kolmogorov constant is universal and does not depend on the Arctic environment or cloudy conditions. Experimental data support that for the three-dimensional energy spectrum α=1.5. For one-dimensional measurements, the corresponding constant depends on whether the component measured is longitudinal or lateral to the flow (relative wind): α_{long}=0.5, α_{lat} = 4/3*α_{long} = 24/55*α ≈ 0.65 (Pope 2000, p.232). As we use the spectral method for the DataHawk2 horizontal measurements, we use α_{long}=0.5 in this study.

3. **How many flights do you compare the DH2 measurements with BELUGA? Do you have some statistics to confirm the DH2 performance?**

   The flights illustrated in Figure 1 of the manuscript were used to compare DH2 measurements and BELUGA (in total four days each with one or two profiles per platform). Please also see the comment below for a comparison between the two platforms for these flights individually. These were the only flights during which the DH2 and the BELUGA sonic anemometer flew concurrently. Therefore, we do not have more robust statistics from MOSAiC to compare DH2 and BELUGA measurements.

4. **Figures 9, 10, and 11 show that the temperature difference between DH2 and BELUGA (for the potential temperature profiles) is 2-4 C. That is relatively large. Do you have any explanations for the data quality and the meaningfulness of using the comparison? Do you have other ground comparisons to determine the temperature measurement's uncertainty range? Similar concerns with the dissipation rate, wind speed variance and the gradient Richardson number. They were plotted on a log scale, and it is hard to understand how accurately the new approach derived parameters compared with BELUGA.**

   The temperature difference in Fig. 9 is in the range of 2-4 ^{o}C only below 800m altitude. Above, the temperature profiles agree very well. Therefore, we assume that this difference is not a systematical measurement offset, but is caused by spatial and temporal heterogeneity of the ABL. Also, Fig. 10 shows an almost perfect agreement between DH2 and BELUGA temperature measurements. The plots also include a comparison to meteorological mast measurements, which as well do not show a systematic offset. Figure 11 does not have BELUGA measurements. A detailed comparison of dissipation rates and variances derived from BELUGA and DH2 is shown in Figs. 4 through 6, pointing out that the DH2 can resolve smaller turbulent eddies than BELUGA, but within the resolved scales, the measurements agree.

Further, DH2 measurements of temperature, wind, and humidity were compared to those from the radiosondes as an established platform by Jozef et al. (2022), and they found that DH2 and radiosonde profiles of the aforementioned variables were similar to each other, such that features including ABL height, low-level jets, and inversions were in agreement between DH2 and radiosonde measurements taken at approximately the same time. For example, when comparing ABL height from DH2 and radiosonde observations within ~3 hours of each other, no significant difference at the 5% significance level was found. Additionally, Hamilton et al. (2022) provide detailed statistics on the performance of the DH2 when compared to radiosonde observations within 1 hour of the DH2 launch during MOSAiC, showing reasonable agreement of temperature and wind. The Jozef et al. (2022) paper provides in the supplementary figures the profiles of bulk Richardson number from all DH2 flights and the corresponding radiosonde (closest radiosonde to DH2 launch, within ~3hrs). These plots show that Ri_b from DH2 and radiosonde are generally in good agreement. The Ri_b profiles from the radiosondes were less noisy than that from the DH2, but this can be attributed to smoother profiles due to the lower vertical resolution of measurements from the radiosonde versus the DH2.

We have added comparisons to radiosoundings to the revised manuscript in Sect. 2.1.2, line 111ff.

Minor comments

- **Equation 9 used equation 10 in Siebert et al. (2006) for the u component. How do you derive C2 =2.6 for vertical velocity components?**

  For the longitudinal spectrum: $C_{2,long}$=2 (as also noted in Siebert 2006). For the lateral spectrum: $C_{2,lat}$ = 2* 4/3 ≈ 2.66 (Pope 2000). In isotropic conditions, the Kolmogorov theory predicts a 4/3 ratio between the spectra of lateral and longitudinal wind velocity components in the inertial subrange (Kaimal et al. 1972).

- **In section 2.3.1, the structure of this section is confusing. Before Line 232, the author introduced the method used by Siebert et al. (2006), then starting in line 332, "a different established method is applied to derive dissipation rates." Please list the equations for the other method. What are the connections between the two methods? Do you plan to compare them? Or do they complement each other? Which method is more suitable for the Arctic environment? What are the pros and cons of choosing each method in Fig 3?**

  The formulation was misleading, we do not use a different method other than the ones introduced: for DH2 the spectral method, and for BELUGA the second-order structure function. We have clarified this in the revised text in section 2.3.1, line 273ff. We also added in line 224: “Both techniques estimate dissipation rates at inertial subrange scales and are independent on the larger scales.” Fig.3 compares the different methods applied to the DH hotwire data for one day where the hotwire data quality allows applying both methods.