Reviewer #2 reply:

We thank the reviewer for the supportive review of the manuscript. We appreciate the detailed comments about the usage of ERA5 and its potential performance issues in the Arctic. We revised the manuscript including a more thorough literature research regarding ERA5 performance and adding minor observation-reanalysis comparisons. Below, we repeat the reviewer's comments in black and write our response in blue. The line numbers in the responses are valid for the revised manuscript.

This is a generally well-written presentation on the meteorology experienced during the HALO-AC3. I enjoyed reading it and am confident it will be a useful contribution to the larger research community. I like many of the figures. My main comment is that I rather wished for more description of how well the ERA5 reanalysis can be trusted for this part of the world. The analysis relies heavily on ERA5, including for cloud and precipitation phase. Has there been any comparison of the ERA5 products to the in-situ data yet? It may still be early stages for this, but some of the drop sonde quantities provided in the manuscript would be very easy to compare against ERA5. A literature review of other assessments is mostly absent, other than some in section 2.4.4. My own cursory web search revealed at least these two: Seethala et al., 2021; Loeb et al., 2022, but I would expect there to be more. I was hoping to see a more systematic assessment of the ERA5 quantities.

- We performed a more dedicated literature research regarding the performance of ERA5 in the Arctic (line 122-126). For the HALO-(AC)³ campaign, ERA5 data can probably be more trusted than in other years because our dropsondes have been assimilated. We also added this information to the manuscript: "During HALO-(AC)³, dropsonde measurements launched from HALO were assimilated into ERA5." (line 126-127). We now include a comparison of IVT of the Atmospheric Rivers from ERA5 to dropsondes (line 377-379 and Appendix B), and IWV from Ny-Alesund radiosondes to values from ERA5 (line 316-317). For brevity and to focus on the analysis of the weather events and their climatological context, we keep the comparison of ERA5 with observations to a minimum. More detailed comparisons are expected to be part of upcoming publications.
  The precipitation phase is a known problem even with observations due to deficits in model microphysics. We observed liquid precipitation over the sea ice with HALO's cloud radar during the Atmospheric River events. We added this information to the precipitation paragraph of the warm air intrusions and Atmospheric Rivers chapter (4.1) to ensure the reader that ERA5 can be trusted in this case: "Liquid precipitation over sea ice was also observed by the cloud radar onboard HALO during research flights. " (line 389-390)

Was the data assimilation consistent for the entire timespan of the ERA5 climatology, so that statements about 'maximum records' (e.g. line 281) are fair to make?

- ERA5 assimilates various sources of information and the used number of observational data has increased steadily throughout time. Therefore, the quality of reanalysis in general has increased over time especially as more satellite data are assimilated. With the percentile information in the manuscript, we assess if a value is close or far from a new maximum or minimum.

The other main comment is that in several places there are references to place names whose geography the reader may not be aware of, such as Franz Josef land. Fig. 1 might be a place to add some helpful geographic annotations, also for the ocean basins (Fram Strait, Greenland Sea, Barents Sea), as is section 2.1

- We agree that the reader can much better follow our analysis when we point out the major land marks in Figure 1. Therefore, we added the ocean basin labels. The location of Franz Josef Land relative to Svalbard is now described in the manuscript ("...over Scandinavia and around Franz Josef Land (northeast of Svalbard, Fig. 6b).", line 232-233) but did not fit in Figure 1 without overloading it.

Specific comments:

Abstract: The acronym ERA5 is spelled out, but that for HALO-AC3 is not. My guess is that more readers will know what ERA5 is, than HALO-AC3. I'd suggest spelling out HALO-AC3 and seeing if the journal will accept ERA5 as is.

- Agreed. We now briefly explain the acronyms HALO and (AC)3 to clarify the origin of the campaign name: "Centered around the High Altitude and Long Range (HALO) research aircraft and the collaborative research project on Arctic Amplification (AC)3, the airborne field campaign HALO-(AC)3 took place from 07 March to 12 April 2022. " (line 2-3)

Abstract, line 10: include years of the ERA5 climatology.

- The years of the ERA5 climatology is embedded in the abstract as follows: "Compared to the ERA5 climatology (1979-2022), record breaking vertically integrated poleward heat and moisture fluxes ..." (line 8-9)

Abstract, line 11: not a good idea to expect the reader to know what a 'shapiro-keyser' cyclone is, you can leave out the name reference

- We agree that the detailed characteristics of this cyclone are not needed here and thus replaced it by "strong cyclone" (line 11).

Abstract, line 16: 'untypically' => 'atypically'

- Done.

Intro, line 32: my recollection is that the Francis and Vavrus, 2015, was highly debated after it was published, leading to a US CLIVAR report, and spurring other work by e.g. E. Barnes at CSU. A bit more detailed literature review here would make this portion more impactful.

- We extended the literature review regarding the impact of climate change on the jet stream: "A more meandering jet would result in an increasing number of poleward moist and warm air intrusions (MWAIs) and southward cold air outbreaks (CAOs). However, the tropical upper troposphere warms while the Arctic lower stratosphere cools, reducing meridional temperature gradients at higher altitudes (Lee et al., 2019, Stendel et al., 2021). The frequency of meridional transport through the North Atlantic has increased during the last decades while it stayed constant or even decreased in other regions (Mewes and Jacobi, 2019). You et al. (2022) found a positive trend in the frequency and duration of atmospheric blocking over the Barents Sea especially in winter, supporting the statement of an enhanced North Atlantic pathway for meridional transport. " (line 31-37)

Line 67: space between performance of

- Well spotted. Space has been added.

Line 120: Nimbus -> Nimbus

- Thank you for spotting also this typo, but we removed the description of satellite sensors for brevity.

Lines 161-164: this is slightly confusing as written. Do Guan and Wailer use a IVT threshold of 100 kg/m/s and you use 50? Maybe combine those two phrases into one sentence if so.

- Indeed, for the Arctic, we use the threshold 50 kg m-1 s-1 while it is 100 kg m-1 s-1 in the original Guan and Waliser revised Atmospheric River detection algorithm. We rephrased the second sentence to: "In this study, ARs were identified with a global algorithm by Guan and Waliser (2015) in its revised version (Guan et al., 2018), adapted to the lower moisture content of the Arctic (Lauer et al., 2023)." (line 137-139)

Line 212: how is the polar low's center determined.

- Thank you for pointing out that we missed giving this information. We added "pressure minimum" in paranthesis to clarify this: "We analyse the environment for Polar Low formation with a set of conditions (C1-C6) suggested by Radovan et al. (2019) and Terpstra et al. (2016) in a 200 km radius around the Polar Low's centre (pressure minimum):" (line 157-158)

Line 212: using the max 10 m wind gust as opposed to the mean wind assumes ERA5 underestimates polar low wind gusts…do you know this for sure?

- We did not intend this impression and realized that this could be formulated more clearly. We did not use mean wind because this quantity is expected to yield lower wind speeds than actually present at some places in the Polar Low due to ERA5's coarse resolution. Using either a higher resolution reanalysis (like CARRA) or using wind gusts instead capture the mean wind speed in a rather small scale feature like a Polar Low better. We added the information that "Wahl et al. (2017) [1] found that scales of multiples of the grid cell spacing are required to realistically represent the energy spectrum of a wind field. We decided to use the maximum 10 m wind gust instead of mean wind to get a better estimate of the near-surface wind field of this small-scale phenomenon that might be hidden due to the coarse resolution of ERA5." (line 163-166).
  [1]: Wahl, S., Bollmeyer, C., Crewell, S., Figura, C., Friederichs, P., Hense, A., Keller, J. D., and Ohlwein, C. (2017): A novel convective-scale regional reanalysis COSMO-REA2: Improving the representation of precipitation. *Meteorologische Zeitschrift* 26 (4), 345-361, doi: 10.1127/metz/2017/0824.

Line 218: the drop sonde vorticity calculation: at what time? What was the center of the drop sonde circle? The vorticity calculation should be easy to compare to that from ERA5, how does ERA5 do?

- The dropsondes have been launched between 06:55 and 07:53 UTC on 08 April. This information has been added to chapter 4.3, where also other time information regarding the Polar Low is given: "Dropsonde measurements between 06:55 and 07:53 UTC show high values of relative vorticity in the lowest 2 km and above 6 km, indicating cyclonic rotation." (line 455-456). We added a brief comparison of the ERA5 vorticity to the dropsonde vorticity: "When averaging ERA5 vorticity over the grid points closest to the dropsonde positions, we find a disagreement to the dropsonde measurements below and good agreement above 4 km height. (...)The disagreement between ERA5 and the dropsondes could be due to a misrepresentation of the Polar Low's wind field in the reanalysis or due to spatio--temporal mismatches of its position." (line 456-457, 459-460). The centre of the dropsonde circle was the centre of the circle flown by HALO and thus slightly off the pressure minimum seen in ERA5. This can be seen in Figure 13b.

Line 248: southerly winds not obvious for the central region in fig. 3d…would suggest removing ‘and central’.

- We agree that this was not as clearly visible in the central region for 07-09 March. Therefore, when we shortened this section, we summarized the typical wind pattern of the warm period as: "This pressure constellation resulted in a consistent southerly and southwesterly flow with only a few short--lived interruptions in the three measurement regions. The interruptions can be seen as near--surface temperature drops and wind direction change (i.e., 11 March, Fig. 3c, d)." (line 201-203)

Lines 248-255: it's hard to visualize what you are saying just from fig. 3, would suggest adding in some spatial circulation figures like what you have within fig. 4.

- As this is only preconditioning the HALO-(AC)³ period, we did not include a spatial circulation figure. This part has been dropped to reduce the length of this section as suggested by reviewer 1 and to decrease the focus on the preconditioning.

Line 278: the stated maximum IVT_north of 388 kg/s/m doesn't seem consistent w Fig. 2. Is that because the maximum is an hourly-mean?

- In Figure 2, regional (or area) averages have been computed averaging over both latitudes and longitudes. To make it more clear that latitude averages (not area averages) are meant, we rephrased this part to: "Simultaneously, the latitude--averaged IHT_north and IVT_north exceeded the previous maxima from 1996 ($9.44*10^{10}$ W m-1 vs. $9.32*10^{10}$ W m-1, and 388 kg m-1 s-1 vs. 384 kg m-1 s-1, Fig. 5)." (line 216-217)

Line 280-282: what's the difference between 'latitude-averaged' and 'area-averaged' IVT?

- Latitude-averaged is only a 1-dimensional averaging while area-averaged is 2-dimensional and respects the increasing data point density of a regular lat-lon grid with increasing latitudes. To stress the latitude averaging, we rephrased this part to: "In Fig. 5, we show latitude--averages of IVT_north and the vertically integrated meridional heat flux IHT_north over the central region to .... " (line 189-190)

Line 283: how are you defining MWAI intensity? Winds?

- We distinguish between weak and strong MWAIs through IVT_north thresholds (strong if exceeding central-region-average of 100 kg m-1 s-1). We rephrased the distinction between weak and strong MWAIs more clearly: "An MWAI is considered weak (strong) when IVT_north is below (equal or above) 100 kg m-1 s-1. " (line 136)

Line 294: I don't follow "The moisture flux decreased faster than the heat flux". Is this from the atmosphere to the ocean? Or the turbulent fluxes coming off of the ocean?

- To clarify that we mean the atmospheric heat and moisture fluxes, we rephrased this part to: "After the AR, much drier but still relatively warm air followed, leading to a strong reduction in IVT_north and a slight reduction in IHT_north (Fig. 5). " (line 222-223)

Line 302: "frontal structure representative of a Shapiro-Keyser cyclone". Better to just describe the frontal structure, as many readers, including myself, will not know what you are talking about.

- We understand that not every reader might be familiar with the term but during the campaign, this event was always called "Shapiro-Keyser" cyclone. In upcoming studies, this term might also be used as it represents a turning point in the campaign

and would be lost in other cyclones if this classification was removed. We dropped the brief description of the Shapiro-Keyser cyclone characteristics for brevity and only refer to literature.

Line 385: it could be interesting to discuss how the subsidence is evolving as well, as that would also influence the static stability.

- We analyzed ERA5 based vertical velocity anomalies at 850, 700 and 500 hPa over the cold period compared to the 1979-2022 climatology. We found a slight positive subsidence anomaly at 850, 700 and 500 hPa in the central Arctic, coinciding with the enhanced static stability in this region (cold anomaly at 2m, warm anomaly at 850 hPa). This information has been added in line 281-282: "This area also shows slightly positive subsidence anomalies at 850 hPa (not shown). "

Line 415: not fully following how surface conditions explain a high tropopause height. I think you can just say 'vertical advection lifts the tropopause to 12.9 km' and be done with it.

- Agreed, we rephrased this part to: "Vertical advection of heat and moisture lifted the tropopause to 12.9 km (measured by the 12 UTC radiosonde on 12 March, Fig. 7a)." (line 307-308)

Line 425: these radiosonde profiles are also an opportunity to assess the corresponding ERA5 profiles.

- We understand that a comparison between ERA5 and observations is beneficial for scientists using ERA5, but we would like to keep the focus of the manuscript on the weather (and sea ice condition) analysis and climatological context. Additionally, we currently do not know which sondes were assimilated and which not, which makes a fair comparison difficult. Additionally, the orography around Ny-Alesund is very complex, adding to the difficulties for the ERA5-radiosonde comparison. Nevertheless, we added the ERA5 based IWV estimate close to Ny-Alesund for 24 March 2022: "Northerly winds corresponding to the MCAO period led to extremely dry conditions with IWV down to 1.1 kg m-2 (closest grid point in ERA5 with land--fraction < 0.25: 1.5 kg m-2) on 24 March at 06 UTC (Fig. 7b)." (line 316-318)

Line 450: '2023) that' => '2023), '

- Well spotted. We corrected it

Line 527: it should be relatively straightforward to figure out if the latent heat fluxes are increases because q_sat-q_air is increasing or because the wind speeds are increasing. How much is the q_sat increasing? I would think the SST would not be changing all that much?

- We investigated the regional distribution of q_sat - q_air and found that the southern region indeed featured higher differences between q_sat and q_air (especially later on 01 April and on 02 April 2022). As the sea surface temperature over the southern

region is higher (mostly by 3-4 K) than over the central region, q_sat is increased in the south. This shows that q_sat – q_air is responsible for enhanced latent heat fluxes. We added this information in section 4.2: "We found that larger differences between the specific humidity of the air and specific humidity at saturation in the southern compared to the central region were responsible for the increased latent heat fluxes (not shown)." (line 411-412)

Line 565: 'lied' => 'lay'

- Has been corrected.

Line 573: please include a figure comparing the drop sonde vorticity and wind speed profiles to that from ERA5. The drop sonde circle should also give you a divergence profile and updraft speed that can be compared to that from ERA5 in a figure. Please do so.

- We added ERA5 vorticity to the plot (see response to your comment on line 218 of the unrevised manuscript). Also with respect to the other reviewer, we do not show divergence profiles (and comparisons to ERA5 in this respect) to avoid adding more content to the manuscript. Furthermore, a detailed publication on the divergence measurements from the research flights is underway.

Line 605: 'a' => 'an' (in front of easterly)

- Has been corrected.

Line 653: insert 'the' before 'absence'

- Has been corrected.

Figures/Tables

Table 1: the northern part of the northern region is hard to understand initially from this table, however, figure 1 shows the study area very well. I had to look down to figure 1 to understand the table. It might be best to either have the figure before the table, or to add a 4th column to the table for the northern part of the northern region

- We agree that a fourth column helps to separate the extension from the main part of the northern region. The table has also been moved to the new Appendix A, which contains detailed methods, because the exact coordinates would only be necessary for full reproducibility.

Fig. 1: spell out what NYA means in the caption.

- We added the description of "NYA" in the caption: "The orange label NYA in the zoomed domain marks the location of Ny-Alesund."

Fig. 3: the graphic on the left looks a bit odd, as there is no real need for us to know the day of the week I don't think. I would suggest adding a color bar to the top of the right-hand panels that has the identification information.

- We agree that the calendar like graphic has some redundant information like the weekday. However, events can be more directly attributed to a certain date while it might be more difficult to read the exact date in a colorbar-like time series (i.e., as illustrated below in the same colours used in Fig. 3a). In case we did not understand your idea, we would kindly ask for an elaborated comment.

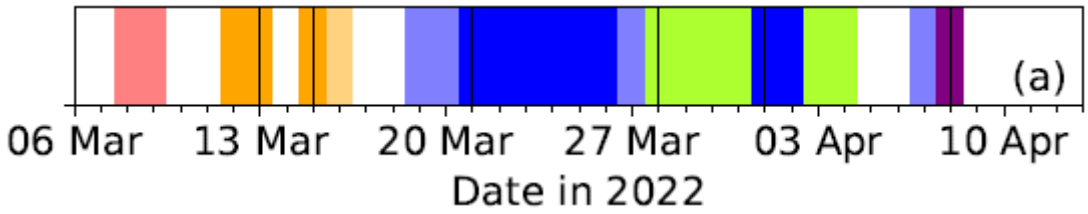

Fig. 4: include dates in caption.

- Done.

Fig. 5: are these hourly values? Would be worth putting in caption.

- Yes, theses are hourly values. We added it to the caption: "Hovmöller diagram of hourly vertically integrated meridional fluxes of (a) heat (IHT_north), and (b) moisture (IVT_north) during HALO-(AC)^3, averaged over the central region latitudes."

Fig. 8: some strange overlapping of lat/lon labels in a-c, would suggest just removing a few.

- We removed the overlapping lat and lot labels.

Fig. 9, caption: I don't understand the last sentence, and both the top and far right histograms need a basic description. Also, the values shown in here for

- We think that you referred to the horizontal lines that should mark the measurement region boundaries? We changed the colours to the respective colours used in Fig. 3.

Fig. 10: fewer lat/lon labels and bigger plots would be nice. You could just leave most of the lat labels out.

- We reduced the number of lat and lon labels and increased the subplot size.

Fig. 13: including the ERA5 vorticity values on panel a would be nice as would be an additional plot showing the wind speeds/divergences. Panel b needs SLP labels.

- We added the ERA5 vorticity to the plot but do not show wind speeds or divergence due to the length of the manuscript. Panel b now has mean sea level pressure labels. We also corrected an error regarding the dropsonde vorticity uncertainty computation.