Response to Anonymous Referee 2 Comments

We would like to sincerely thank Anonymous Referee 2 for taking the time to review our manuscript and for their helpful comments, which have improved the manuscript. Each referee comment is given below in ***bold italics*** followed by our response to the comment. The line numbers provided in our responses refer to line numbers in the revised manuscript, unless otherwise stated.

This manuscript describes a statistical analysis for a year-round period of Arctic boundary layer observations based largely on radiosondes during the MOSAiC project accompanied by data observed by the DataHawk2 unmanned vehicle. As a central tool, a "self-organizing map" approach was applied to the data. There is no doubt that such a statistical analysis is extremely helpful in addition to all the case studies that have been and are being evaluated. Therefore, the enormous amount of work is greatly appreciated, and after major improvements to the manuscript, I also support the publication of this analysis. However, the manuscript needs a thorough revision that goes far beyond classical "major revisions". I will try to justify this in detail below.

My main concerns are:

The manuscript is extremely difficult to read and understand; this is largely due to the very intensive use of abbreviations. Especially when referring to the different stability regimes of the boundary layer, the abbreviations are not really intuitive and hardly anyone will be able to remember them while reading (see at line 428/29 for example when a sentence almost completely is based on abbreviations). In particular – see Tab 2- the capital "S" is sometimes used for "shallow" and sometimes for "stable" – I have no good suggestion at the moment to improve this, but please consider of a better and simpler way to categorize the different regimes and avoiding abbreviations.

We understand and appreciate your concern. To make the manuscript easier to follow, we have made the following adjustments throughout the paper.

- Reorganized the flow of the paper such that there are separate "Results" and "Discussion and conclusions" sections.
- Cleaned up the SOM figures and provided detailed keys
- Removed some panels from the box and whisker plots that did not add much to the discussion
- Removed all figures and text related to temperature inversions and the UAS data to provide a more concise manuscript

Additionally, where appropriate, we have done our best to rather refer to the general near-surface and aloft stability categories as they relate to the results, instead of listing off the abbreviations.

However in large part, we believe that use of abbreviations is still the best option, and hope that the other changes we have made throughout the manuscript make it easier to read and understand. There are a few reasons for this decision: 1) We believe if a reader takes a few extra moments to review Table 2, they will be able to follow the abbreviations thoughtful the paper. In the end, there are only 5 near-surface regimes and only 3 aloft regimes (which are the same as the near-surface regimes, but only adding an A to indicate "aloft") to remember; 2) A complementary paper containing the same stability regimes as included in the current paper is also under review for publication, and no reviewers have expressed concerns with being able to keep track of them; and 3) If we were to entirely avoid abbreviations, we would instead need to write out the regimes when they are mentioned, but this would perhaps make the paper even more long and confusing, as several of the regime names are quite lengthy.

Very often it is concluded in the paper that the analysis yields unsurprising results or that the results are logically and physically explainable - well, I expect that with the correlations but do you want to evaluate the tool or deliver new scientific findings?

To address this concern, we have made a few adjustments. First, we have now more clearly communicated one of the goals of the paper in the introduction – to use the MOSAiC dataset to see if "textbook" ABL meteorology holds true in the central Arctic, while providing new insight into the quantitative values for the characteristics we show as they relate to stability.

"The results of such a study are firstly valuable to reveal whether current observations agree with past observations and well-known ABL meteorological processes. Additionally, through the use of new methods (i.e., the SOM analysis and detailed stability regime classification), the results also provide further constraints on the vertical structure and features of the Arctic lower atmosphere that may be helpful to improve parameterizations of the central Arctic in weather and climate models." (line 132)

Thus, when we later claim that aspects of the analysis yield unsurprising results, the reader will better understand that this is still an important finding as it relates to the goals of the paper. Next, we have adjusted the results and discussion to focus more heavily on new results which were discovered through this analysis, and have removed the panels from the box and whisker plots that show results which are largely a function of how we define stability.

It is a bit tiring for the reader to have each figure described in such detail (and the figures contain a large amount of detail...), and you should try to find a slightly better and more compact way of presenting and introducing the figures. I know this comment is quite generic but maybe you find a good way to describe your figures in a more comact way.

We have worked to make the description of the figures more concise throughout the paper. One way that we do this is by making the figures themselves easier to read (through larger font and removing unnecessary information) and understand as stand-alone entities (through the addition of detailed keys in the SOM figures which explain what we see in each subplot), and thus less description is necessary in the text. Additionally, the separation of the content of the paper into distinct "Results" and "Discussion and conclusions" sections, allows us to spend less time describing each figure when it comes up in the paper.

When I first read the manuscript and got to Figure 2, I was completely overwhelmed. Why do you need 30 schemes to describe the ABL? With many patterns you only see marginal differences when you look very closely. I think that the manuscript could be made much simpler and more readable if the analysis was limited to a handful of characteristic patterns.

We completely understand being overwhelmed with Fig. 2 in the original manuscript. Thus, as this figure is not instrumental for understanding the analysis (and is rather a demonstration of the methods), we have simplified the figure and moved it to the supplement (see Supplementary Fig. S1 of the revised manuscript). This way, if a reader is very interested to know more about what the 30 SOM patterns look like, and the spread in the observations around the SOM pattern in a given node, they can refer to it, but it is not required to understand the rest of the paper.

Additionally, we understand your questioning as to why we need 30 patterns to describe the ABL, and realize that we need to better describe the purpose of a SOM in the Methods section in order to convince you (and other readers) that 30 patterns are necessary.

The purpose of a SOM is not to find the smallest number of patterns which represent all the possible structures in the data. Rather, a SOM is meant to continuously depict the range of structures present in the training data from one SOM-identified profile to the next - in this case, we show the continuum of ABL vertical structure in the Arctic. Here, the SOM is being used as a way to visualize a large dataset (1377 soundings) in a manageable way while also allowing subtle details in the vertical structure (such as varying heights of stable layers, differing strength of stable layers) to be identified. For example, at first glance, patterns 27 and 28 in Fig. 3 of the original manuscript (Fig. 2 in the revised manuscript) appear to be very similar. However, with a deeper look, we see that the ABL depth between the two patterns differs discernably (leading one to be classified as WS, and the other as VSM), and the strength of the elevated inversion is a bit stronger for pattern 27 (even though they both qualify as -SSA). Then when looking at differences in wind speed and LLJ characteristics for these two patterns in Fig. 5 of the original manuscript (Fig. 4 in the revised manuscript) indeed the characteristics differ notable between the two patterns (e.g., pattern 27 has faster wind speeds, more frequent LLJs, and a lower mean LLJ altitude) which at least partly explain the slight difference between the potential temperature structures of the two patterns.

While we understand that 30 nodes seems like a lot, it is a manageable number of patterns compared to the total sample size of 1377 radiosondes, such that we can actually visualize and understand the range of ABL vertical structures present in the data in one figure (albeit a complex figure). We tested many options for number of nodes (from 20 to 35), and found that certain ABL structures were missing when fewer than 30 nodes were used (e.g., cases with very strong stability either near the surface or aloft were merged with cases with weaker stability, and cases with varying elevated inversion height were merged into the same node).

Additionally, many other SOM studies have used 30 or more nodes (Sheridan and Lee, 2011; Cassano et al., 2016; Nigro et al., 2017), so we do not stray from common practice in doing the same.

Cassano, J. J., Nigro, M., and Lazzara, M.: Characteristics of the near surface atmosphere over the Ross ice shelf, Antarctica, J. Geophys. Res.-Atmos., 121, 3339-3362, https://doi.org/10.1002/2015JD024383, 2016.

Nigro, M. A., Cassano, J. J., Willi, J., Bromwich, D. H., and Lazzara, M. A.: A Self-Organizing-Map-Based Evaluation of the Antarctic Mesoscale Prediction System Using Observations from a 30-m Instrumented Tower on the Ross Ice Shelf, Antarctica, Weather Forecast, 32, 223-242, https://doi.org/10.1175/WAF-D-16-0084.1, 2017.

Sheridan, S. C. and Lee, C, C.: The self-organizing map in synoptic climatology research, Prog. Phys. Geog., 35, 109-119, https://doi.org/10.1177/0309133310397582, 2011.

In order to address all of these points, we have:

1) Added more description in the ‘self-organizing map analysis’ section explaining how a SOM works, including how it is trained, what the goals of such an analysis are, what the resulting product is:

“The SOM analysis uses an unsupervised neural network algorithm to objectively identify a user-specified number of patterns in a training data set (Cassano et al., 2015; Kohonen, 2001). In doing so, this analysis projects high-dimensional input data onto a low-dimensional space as a grid of SOM-identified patterns (Liu and Weisburg, 2011) and provides a compact way to visualize the range of conditions present in the training data. The grid of SOM-identified patterns is referred to as a SOM, or simply a map. Atmospheric applications of SOMs have previously been used to determine ranges of synoptic patterns (Nygård et al., 2021; Cassano et al., 2015; Sheridan and Lee, 2011; Skific et al., 2009; Cassano et al., 2006; Hewitson and Crane, 2002), identify large scale circulation anomalies associated with extreme weather events (Cavazos, 2000), and classify cloud (Ambriose et al., 2000), climate zone (Malmgren and Winter, 1999), precipitation (Crane and Hewitson, 2003), and ice core data (Reusch et al., 2005), to name a few. Most similar to the current study, SOMs have previously been used to identify the range of ABL structures in Antarctica from both tower (Nigro et al., 2017; Cassano et al., 2016) and radiosonde (Dice and Cassano, 2022) data. Here, the SOM analysis is applied to radiosonde profiles of θ_v to identify vertical structure and stability in the lowest 1 km of the atmosphere over the Arctic ice pack during MOSAiC.

A SOM is created by randomly initializing patterns from the input data space and comparing the training data to these patterns. Each sample in the input data is presented to the SOM and compared to all patterns in the initial map. The pattern to which the input data sample is most similar is known as the “winning” pattern, and this pattern, and adjacent neighboring patterns, are modified to reduce the squared difference between it and the input data sample. This process continues for all samples in the training data (Liu and Weisburg, 2011; Cassano et al., 2006) and is repeated thousands of times for the entire training data set until the squared differences between the SOM identified patterns and the training data have been minimized. Further details of how a SOM is trained are given in the papers cited above. Here we use the SOM-PAK software (http://www.cis.hut.fi/research/som-research; Kohonen et al. 1996) to train the SOM presented below.

A critical decision when using SOMs is the number of patterns to be identified by the SOM training, and this depends on the intended application and size of the training data set (Cassano et al., 2006). A greater number of patterns will produce a broader range of structures with more subtle differences between them, and fewer patterns will result in larger variability between and within the patterns. Regardless of the number of patterns identified in the SOM, the SOM provides a smoothly varying, continuous depiction of the range of conditions present in the training data. The output from the SOM training is a two-dimensional array of patterns which are representative of the range of conditions present in the training data (Cassano et al., 2006). The SOM is organized such that the patterns being most similar are located adjacently, and conversely the most different patterns are on opposite sides of the SOM (Dice and Cassano, 2022; Cassano et al., 2016; Liu and Weisburg, 2011). Each sample in the training data is mapped to the resulting SOM pattern with which it has the smallest squared difference resulting in a list of samples for each SOM-identified pattern. This list of data samples can then be used to calculate the frequency of each SOM pattern and for additional analyses. (Dice and Cassano, 2022).” (line 242-275)

2) Explained in more detail how the SOM technique was applied in the current study, and its utility for answering our research questions:

“In this study, a 30 pattern SOM was used to describe the range of lower atmospheric stability profiles, defined by θ_v gradient ($d\theta_v/dz$), present in 1377 MOSAiC radiosonde profiles. Before settling on the 6x5 (30 pattern) SOM, we tested SOMs with size and orientation of 5x4 (20 patterns) to 7x5 (35 patterns). When using 20 patterns, the range in strength of near-surface stability and the varying depths of a weakly stable or near-neutral layer were not fully evident. To fully understand the range of vertical structures in the Arctic, highlighting these differences is important, so the inclusion of additional SOM patterns was necessary. However, with 35 patterns, we found that no additional details were introduced beyond what was shown with 30 patterns. Thus, we determined that 30 patterns is the smallest number to sufficiently describe the range lower atmospheric stability during MOSAiC, retaining fundamental features of vertical structure (e.g., varying height and strength of the θ_v inversion). We also tested the SOM trained with the θ_v profiles rather than the gradient (in the form of the θ_v anomaly compared to 1 km, to remove seasonal temperature dependence), but found that the range in height and strength of the θ_v inversion, as well as the differentiation between a weakly stable or near-neutral layer below a θ_v inversion, were not distinguished.

The profiles of $d\theta_v/dz$ used to train the SOM were derived from radiosonde observations that were first interpolated to a consistent vertical grid of 5 m spacing between 35 m and 1 km (temperature and relative humidity were linearly interpolated and pressure was interpolated with the hypsometric equation). The maximum altitude of 1 km was chosen because it includes the full depth of the ABL in every case and also allows for diagnosing stability immediately above the ABL. Then, θ_v was calculated at 5 m intervals using the interpolated measurements. Finally, profiles of $d\theta_v/dz$ in K $(100\ m)^{-1}$ were calculated as the change in θ_v between adjacent datapoints, resulting in $d\theta_v/dz$ values at 37.5 m, 42.5 m, 47.5 m, and so on, with the last value being at 997.5 m. Training the SOM with $d\theta_v/dz$ profiles resulted in an array of patterns differentiated by the strength and height of the θ_v inversion. As such, observations with similar strength θ_v inversions which occurred at different heights, and observations with similar heights of the θ_v inversion but different strengths, were separated into different SOM-identified patterns.” (line 276-297)

Lastly, while we use the 30 pattern SOM as a starting point for understanding the range of stability profiles present in the MOSAiC soundings, the remainder of the paper does in fact use a reduced number of patterns (stability regimes) to perform further analysis. We view the SOM and the stability regimes as complementary, with the SOM showing details of how the vertical profile of stability varies, while the stability regimes distill these details to the most critical factors of near surface and aloft stability. We have added some text to more explicitly state these things:

“… a more tangible demonstration of the range of vertical structures present during MOSAiC is shown in Fig. 2 (Sect. 3.1) with the mean profiles of $d\theta_v/dz$ and θ_v anomaly for all radiosondes mapped to a given pattern. Results from the SOM analysis will focus on the frequency of occurrence of each pattern and the variability in the vertical structure depending on time of year (e.g., which SOM patterns largely occur in certain seasons). Seasonal analysis in this paper is carried out by grouping observations during September, October, and November as fall; December, January, and February as winter; March, April, and May as spring; and June, July, and August as summer. Additionally, profiles of wind speed (produced by interpolating the zonal and meridional components to the 5 m grid and then calculating total wind speed profiles) and LLJ characteristics in the context of the SOM patterns will be analyzed. Lastly, once the full range of vertical structures was revealed by the SOM, this information was used to develop a set of criteria for classifying stability of any given observation that distills the detail of the SOM to the most critical factors of stability within and above the ABL.” (line 299-309)

A general note on writing style: please try to avoid repetition to strengthen the manuscript.

Furthermore, the sentences are often so complicated and convoluted that a fluent reading - at least for me - was very difficult or impossible. I myself am not a native speaker, but there are enough competent co-authors who can edit the manuscript thoroughly.

We have worked throughout the manuscript to avoid repetition and simplify the sentences. Again we believe that following your suggestion to separate the results and discussion into two separate sections helps with this.

General comment about most of the figures (although here I refer explicitly to Fig. 3): The figure is based on 30 subplots which are by definition quite small but if you try to include even more information in terms of several additional numbers and vertical or horizontal lines, the plots will get really crowded. Even worth in Fig 4 where I am not able to read at all the numbers you included into the subplots – they are simply too small and too many.

We have made several adjustments to make the complex SOM figures easier to read and understand:

Fig. 3 in the original manuscript (Fig. 2 in the revised manuscript):

- Removed the SOM profile – this is shown in what is now supplementary Fig. S1, and it is not necessary for the discussion of this figure
- Made the font of the numbers and letters in each subplot bigger
- Made the axes fonts bigger
- Added a "subplot key" which demonstrates clearly what is shown in each subplot, so that a reader can refer to it, rather than digging through the figure caption to figure out what they are looking at.
- We have opted to retain the horizonal and vertical lines because we think that they importantly demonstrate how each SOM pattern fits into the various stability regime classifications. However, if the referee still feels that these lines should be removed, we are willing to consider removing the lines.

Fig. 4 in the original manuscript (Fig. 3 in the revised manuscript):

- Removed the number in the upper center of each box which indicated the number of radiosonde profiles mapped to that SOM pattern. This information is included in Fig. 3 in the original manuscript (Fig. 2 in the revised manuscript), so is not needed here.
- Removed the number in the center of each box which indicated the number of cases in that pattern/season. We only retain the percentage, as this is the more valuable number for understanding the results
- Made the font of the numbers and letters in each box bigger.
- Added an opaque white box behind the percentage written in each box, so it can be better seen regardless of the shading in each box
- Made the seasonal subplots bigger
- We also include now the annual frequency subplot in this same figure, in response to your next comment (we discuss this figure in the text so you suggest this not be in the supplement).

Fig. 5 in the original manuscript (Fig. 4 in the revised manuscript):

- Made the font of the numbers and letters in each subplot bigger
- Made the axes fonts bigger
- Removed the number in the upper center of each box which indicated the number of radiosonde profiles mapped to that SOM pattern. This information is included in Fig. 3 in the original manuscript (Fig. 2 in the revised manuscript), so is not needed here.
- Added a "subplot key" which demonstrates clearly what is shown in each subplot, so that a reader can refer to it, rather than digging through the figure description to figure out what they are looking at.
- Due to the above bullet point, we have removed "LLJ:" and "2m" from each subplot, as the "subplot key" now indicates what these numbers mean.

I am not convinced that discussing material in the manuscript that has been moved to a supplement is the correct or formal way to do it. If you have to many figures, you should solve this problem differently.

We have reorganized the content of the paper to include all figures which are heavily discussed in the main text of the manuscript, rather than in the supplement. Specifically:

- The annual SOM pattern frequencies (Supplementary Fig. S1 in the original manuscript) has been moved to the main text, and combined into the figure with the seasonal frequencies (Fig. 3 in the revised manuscript)
- The DH2 figure (Supplementary Fig. S2 in the original manuscript) has been removed entirely, as we have chosen to remove the UAS analysis from the current paper (in response to concerns from the other reviewer)
- LLJ frequency (Supplementary Fig. S4 in the original manuscript) has been moved to the main text, and combined into the figure with the LLJ characteristics (Fig. 6 in the revised manuscript)
- TI frequency (Supplementary Fig. S6 in the original manuscript) has been removed entirely, as we have chosen to remove presentation of TI characteristics from the current paper. See response to your next comment for reasoning.

The only figures which remain in the supplement are the figure showing SOM pattern profiles and percentiles of observations (this is adapted from what was Fig. 2 in the original manuscript, and is now Supplementary Fig. S1 in the revised version) and the grid plots showing statistical significance. These figures are all minimally discussed, and not instrumental to understanding the results presented in the paper but may be of interest to some readers seeking additional details not contained in the main text and figures.

About the analysis of temperature inversions: I am a little bit skeptical about this analysis and I wonder of how much of these results are based on self-correlation because the definition of the stability regimes is also based on temperature gradients. You mentioned this issue briefly but this needs to be discussed in more detail.

We agree with the reviewer's critique of this part of the manuscript. Since the temperature inversion results are largely a function of how we classify stability regime (as you note), and do not provide very interesting results, we have removed the temperature inversion results and discussion from the revised manuscript.

Maybe one solution for an improved structure of the entire manuscript would be a stricter separation between explaining the results in one section and discussion and interpretation in another one.

In the revised manuscript, we now separate the content such that the results and corresponding figures are presented in one section ("Results"), and the discussion and interpretation is presented in another section ("Discussion and conclusions"). We believe this helps create better flow of the paper, such that it is easy to understand, and repetition is avoided.

I am not convinced about the meaningful interpretation of parameters averaged over the entire MOSAiC cruise. For example, what can I learn from a statement such as "The average ABL height during MOSAiC was 150 m, and ABL height increases with decreasing stability. " (line 703)? You average ABL height over completely different ABL regimes and the second part of the sentence is somewhat trivial and expected – right?

This is a good point. In the revised manuscript, we only share parameters averaged over the entire MOSAiC cruise when that parameter is largely unchanged between stability regimes (e.g., mean LLJ height is fairly consistent regardless of stability regime). Otherwise, we instead focus on quantities for individual stability regimes (or mean values for groupings of regimes with the same near-surface regime, but different aloft regimes, when appropriate), and how these quantities vary between regimes.

More specific comments:

Introduction (line 35): what do you exactly mean with high temporal and spatial resolution – please specify.

We have largely rewritten the introduction, and so this comment is no longer applicable. We now introduce MOSAiC by saying:

"Thus, there is much to be gained by analysis of more recent data, such as that from the Multidisciplinary drifting Observatory for the Study of Arctic Climate (MOSAiC; Shupe et al. 2020), which observed the central Arctic following one ice floe for a full year from September 2019 to October 2020." (line 110)

Line 39ff: quite generic comment; please provide references - I think it is quite obvious that the Arctic ABL is not necessarily stably stratified in summer - right?

As mentioned above, we have largely rewritten the introduction. Now, we more correctly state that the "Arctic atmosphere over sea ice is typically either stable or near-neutral" (line 39), and then go on to describe in more detail the different mechanisms which contribute to a stable or near-neutral ABL in winter vs. summer:

"Stable conditions are common in Arctic winter (Tjernström and Graversen, 2009) due to persistent longwave cooling in the absence of solar radiation (Brooks et al., 2017) and extended periods of clear skies or thin high clouds (Tjernström and Graversen, 2009), attributable to the lack of open water evaporation. However, intermittent instances of low stratocumulus clouds in winter can force a shallow well-mixed ABL (Morrison et al., 2012; Tjernström and Graversen, 2009; Persson et al., 2002). Such clouds are common during stormy conditions (Brooks et al., 2017; Persson et al., 2002).

Near-neutral or weakly stable conditions are common in Arctic summer (Brooks et al., 2017; Tjernström and Graversen, 2009), often capped by persistent stratiform clouds (Intrieri et al., 2002a; Tjernstrom, 2007; Curry and Ebert, 1992; Liu and Key, 2016; Shupe et al., 2011; Tjernström, 2005, Tjernström et al., 2012; Wang and Key, 2004; Zygmuntowska et al., 2012), which form as ample moisture is advected north either into the Arctic or from the broader ice-free areas across the pan-Arctic region, during the melt season (Sotiropoulou et al., 2016; Tjernström et al., 2019). The ABL is typically decoupled from the cloud layer by a shallow stable layer, such that turbulence is not exchanged between the cloud and the surface (Curry, 1986; Sedlar and Shupe, 2014; Sedlar et al., 2012; Shupe et al., 2013; Sotiropoulou et al., 2014). However, the common advection of warm moist air into the central Arctic can also result in the formation of a shallow, stable ABL (Tjernström et al., 2019; Tjernström, 2005; Cheng-Ying et al., 2011), especially towards the beginning of an advection event, or close to the ice edge (Sotiropoulou et al., 2016; Tjernström et al., 2019). Ice and snow melt in summer may also contribute to the formation of a stable ABL (Kahl, 1990; Gilson et al., 2018)." (line 50-67)

Line 49ff: I assume that - depending on the temperature stratification - the turbulence maybe also increased above the LLJ core because the shear could be similar - right? Furthermore, why should a LLJ weaken stability? The main preconditions for the existence of a LLJ is an almost vanishing turbulent transfer coefficient - typical for stable regimes.

Shear above the LLJ core is typically much less than that below the LLJ core since wind speed below the jet core goes to zero at the surface while that is almost never the case above the LLJ core. The criteria for an LLJ is that the core speed is just 2 m/s greater than the minimum in the wind speed above the jet core. The mean jet core speed during MOSAiC was 11.5 m/s, which is much greater shear when compared to 0 m/s wind speeds at the surface, versus the diminishment of wind speeds above the jet core. However, the wind shear that does exist above the jet core may partly explain diminishment of the capping inversion, and we note that now:

"For both LLJs forced by baroclinicity and inertial oscillations, enhanced wind shear above the jet core may also contribute to turbulent mixing above the LLJ." (line 91)

There are two primary situations in which an LLJ will occur in the Arctic. One such situation is the one you describe: the LLJ forms due to inertial oscillations and is decoupled from the (typically shallow and stable) ABL, such that wind shear from the LLJ does not weaken stability. The other situation is when an LLJ forms due to a baroclinic environment, where wind speeds decrease with altitude according to the thermal wind relationship, and surface friction reduces wind speeds to 0 at the surface, this resulting in an LLJ some distance above the surface. In this case the LLJ is not decoupled from the ABL, and can contribute to weakening of stability in the boundary layer through the enhanced shear from the LLJ.

We have added more discussion in the introduction to clarify these two situations under which an LLJ may form, and the subsequent relationships with ABL stability.

"There are two primary forcing mechanisms for LLJs in the Arctic: baroclinicity and inertial oscillations. Baroclinicity in the Arctic most often occurs near the ice edge (Brümmer & Thiemann, 2002) or due to the passing of a transient cyclone (Jakobson et al., 2013) which creates regions of enhanced temperature contrasts (Koyama et al., 2017). Depending on the wind direction, the horizontal temperature gradient causes the geostrophic wind speed to decrease with height according to the thermal wind relationship (Stull, 1988). This, paired with diminishment of wind speeds at the surface due to friction (Stull, 1988), contributes to the formation of an LLJ at some distance above the surface, typically just above the ABL (Brümmer & Thiemann, 2002). Thus, an LLJ forced by baroclinicity is typically coupled to the surface, and can cause weakening of stability within the ABL due to enhanced shear below the jet core (Banta, 2008; Egerer et al., 2023).

Inertial oscillations in the Arctic can be induced after well-mixed conditions are replaced by increased near-surface stability, for example, after the passing of a storm (Andreas et al., 2010a; Jakobson et al., 2013). In such cases, air aloft becomes decoupled from the surface, ceasing frictional drag, which, along with the impact of the Coriolis force, allows the winds aloft to accelerate to supergeostrophic speeds (Blackadar, 1957; Stull, 1988; Jakobson et al., 2013)." (line 78-90)

Line 101: although you provide an explanation later on, I think a short introduction of what a "self-organizing map" is should be included - simply because you mention it and I think a few words about this technique is essential already at this point and not all readers are knowing about this?!

We have added a short statement in parentheses when the SOM is first introduced, to give a brief summary of what it is:

"A self-organizing map (SOM) analysis (which objectively identifies a user-selected number of patterns present in a training data set) was conducted with the radiosonde profiles to reveal the range of vertical structures observed during MOSAiC…" (line 126)

Methods:

Line 116 ff: is this information about RV Polarstern movement is of interest for your work here – why do you mention this?

You bring up a good point that this statement about when the *Polarstern* travelled under its own power is not relevant to the current work. Thus, the statement has been removed.

About Tab 1:

• What are the sources for the uncertainties? The Vaisala manual? Please provide a reference.

We have added a reference for the Vaisala RS41-SGP uncertainties.

"Instrument specifications and uncertainties for the radiosonde variables are available at: https://www.vaisala.com/sites/default/files/documents/WEA-MET-RS41SGP-Datasheet-B211444EN.pdf (Vaisala Radiosonde RS41-SGP, 2017), and are summarized in Table 1." (line 155)

We have added the citation for the ceilometer uncertainty.

"CBH derivation and uncertainty is discussed in Morris (2016)." (line 178)

Uncertainties for the meteorological tower variables were provided by one of the co-authors, but are also included in Cox et al., submitted, which we hope will be published soon. We already provide a source for the uncertainty in the microwave radiometer data.

• Furthermore, I have serious concerns about the given uncertainty of wind observations: I know that this is the value given in the specifications by Vaisala but there is a lot of discussion about errors in determining the wind velocity - in particular at high latitudes where GPS comes to its end.

We have added a statement to note that the true uncertainty in the winds is likely higher than that given in the datasheet, but this is probably not due to the GPS issues. It is our understanding that GPS issues at high latitudes are largely in vertical accuracy, as all the satellites are down at the horizon. However, because there is visibility on many satellites at any given time even in the Arctic, the horizontal position accuracy (which is used to determine the wind speeds) should be pretty good. Given this, and some other reasons (see text below), we find the winds to be sufficiently reliable for this study.

"It is recognized that the true uncertainty in the winds is likely to be greater than that provided in the data sheet, however after determining that our results changed minimally when additional vertical averaging was applied to the winds (beyond the filtering already applied by Vaisala during their data processing), we find the original winds provided in Maturilli et al. (2021) to be sufficiently reliable for the current study." (line 157)

• Why do you mention uncertainties above 16 km here?

We had mentioned uncertainties above 16 km (for the temperature) and at < 100 hPa (for pressure) simply because they are also included in the data sheet for the Vaisala RS41-SGP. However, as you brought to our attention, uncertainties at those altitudes are not necessary to include in the current manuscript, since the study only uses data at lower levels. Thus, we have removed these high altitude uncertainties from the table.

• I don't understand why a sonic is not enough to estimate the friction velocity?

This is because we use bulk friction velocity, rather than the standard eddy-covariance (EC) value. We have added some text which states more explicitly the difference between the bulk friction velocity and the EC value, as well as justification for our choice.

"Bulk u_* was chosen, as opposed to the standard eddy-covariance value, as the bulk parameterization considers both wind fluctuations and latent heat fluxes (developed using guidance from eddy-covariance data collected during SHEBA; Andreas et al., 2010b) which is more comparable to u_* used in models (e.g., Fairall et al., 2003)." (line 172)

Line 165ff: How can you expect a slope for higher altitudes just based on the lowermost 10 m? Why not simply compare the highest measurement point of the tower with the lowermost observation level of the radiosonde - I assume 12 m (helideck) and 10 m (top of the mast) should compare quite well? If not, you have a problem with the radiosonde – right? Or did I completely misunderstand your approach? – Simply double check the wording.

Since the lowest measurement in the radiosonde is indeed at ~12 m, and the highest tower measurement is at 10 m, we would expect them to line up well, with a similar slope, if there is no issue with the radiosonde measurements (as you state). However, there IS often an issue with the radiosonde measurements in these lowest levels due to the local "heat island" resulting from the presence of the Polarstern (which we already describe in the original manuscript). Thus, we use the tower measurements to identify where the radiosonde measurements are likely incorrect and remove this data. We have adjusted the wording to hopefully make this method more clear:

"Thus, if this "convective layer" was present, then the lowest radiosonde measurements were visually compared to measurements from the met tower to confirm whether the radiosonde measurements were indeed incorrect (e.g., if the lowest few radiosonde measurements were notably warmer then the tower measurement at 10 m). The first credible value of the radiosonde measurements was then taken to be the point at which the tower measurements extrapolated upward would line up with the observed radiosonde measurement, or in the case of a temperature offset between the tower and radiosonde, would have approximately the same slope. All data at the altitudes below this first credible value were removed." (line 194)

Line 168: interpolated => extrapolated? Please check.

Extrapolated is probably the better word. Thank you for this suggestion. The change has been made:

"The first credible value of the radiosonde measurements was then taken to be the point at which the tower measurements extrapolated upward would line up with the observed radiosonde measurement…" (line 196)

Line 172: I thought a low-pass filter removed this pendulum motion? Please comment on this.

It is probably true that the effect of the pendulum motion was already removed during Vaisala's filtering of the data, as we don't really see evidence of this in the published processed data. Thus, we have removed this statement from the revised manuscript.

Line 185: I am somewhat surprised by the high critical Ri value which is two-times higher compared to the "classical" value. Most values published in literature are below 0.25 – do you have an explanation for this?

An explanation for this is described in detail in the Jozef et al. (2022) paper (the subject of which is testing the efficacy of various ABL height detection methods), but in summary this higher critical value is necessary with higher resolution data than is typically used when the "classical" value of 0.25 applies well, and may also be attributed to the methods of calculating Ri_b over a running bin of 30 m throughout the profile (rather than always calculating Ri_b with respect to the surface). Jozef et al. (2022) found that a critical value of 0.25 always identifies an ABL height that is much too shallow for the MOSAiC radiosonde data. We have added some brief text to clarify that this information can be found in the Jozef et al. (2022) paper, as we believe it would take too much space to sufficiently justify in the current manuscript:

"The methodology for calculating the Ri_b profile used to identify ABL height, as well as justification for the use of 0.5 as a critical value (rather than the more traditional value of 0.25) is described in Jozef et al. (2022)." (line 218)

Line 186: I think you can shorten this part a little bit by citing your paper only one times

We have removed the second citation.

Line 189: just to understand it correctly: the gradient is a mean gradient from 35 m to ABL top - right?

The gradient is the overall gradient between 35 m and ABL top: (value at ABL top – value at lowest measurement)/(ABL height – height of lowest measurement). However, we have removed the inclusion of $d\theta v/dz$ over ABL depth from the revised manuscript, and decided it was not necessary to mention the calculation of dV/dz at this point in the manuscript, and thus the sentence has been removed.

Line 225: What do you exactly mean with "theta anomaly profile"? And in which way is one approach "better" than the other one? - please specify.

We have clarified this discussion and the text now reads:

"We also tested the SOM trained with the θ_v profiles rather than the gradient (in the form of the θ_v anomaly compared to 1 km, to remove seasonal temperature dependence), but found that the range in height and strength of the θ_v

inversion, as well as the differentiation between a weakly stable or near-neutral layer below a θ_v inversion, were not distinguished." (line 284)

Line 230 ff: What details are "better" when using 30 patterns instead of 20 or 35? What did I learn from this detail?

We have added some discussion to convince the reader of the choice for 30 patterns versus 20 or 35:

"Before settling on the 6x5 (30 pattern) SOM, we tested SOMs with size and orientation of 5x4 (20 patterns) to 7x5 (35 patterns). When using 20 patterns, the range in strength of near-surface stability and the varying depths of a weakly stable or near-neutral layer were not fully evident. To fully understand the range of vertical structures in the Arctic, highlighting these differences is important, so the inclusion of additional SOM patterns was necessary. However, with 35 patterns, we found that no additional details were introduced beyond what was shown with 30 patterns. Thus, we determined that 30 patterns is the smallest number to sufficiently describe the range lower atmospheric stability during MOSAiC, retaining fundamental features of vertical structure (e.g., varying height and strength of the θ_v inversion)." (line 277)

Line 228ff: I understand that you want to explain details about SOM in a specific part of the paper but you mentioned SOM several time before your explanation - maybe you should at least mention at the beginning what SOM stands for and refer to this point here. I feel that many readers have never heard about SOM before and at least a brief introduction at the beginning could help – or did I have overseen this?!

We now mention very briefly in the introduction that a SOM is:

"A self-organizing map (SOM) analysis (which objectively identifies a user-selected number of patterns present in a training data set) was conducted with the radiosonde profiles to reveal the range of vertical structures observed during MOSAiC…" (line 126)

Additionally, we have added much more detail about how a SOM works, before going on to explaining how the SOM technique was applied in the current study (line 242-275).

Line 244ff: Maybe at this point a comment about the low-pas filtering of GRUAN data is useful and how it effects your data and evaluation?! Or why using 5 m as a grid spacing when the low-pass filtering is at 75 m or so? (see also the comment by Günther Heinemann)

Wind speed was interpolated to a 5 m grid spacing to match the resolution of the interpolated θ_v profiles, such that an average wind speed per SOM pattern could be calculated and visualized in conjunction with the θ_v profiles (see Fig 4 of the revised manuscript). It would not make sense to reduce the resolution of the wind speeds, simply because of the low-pass filtering resolution. However, we have chosen to mention the wind speed interpolation a little later on in the section, as this is not actually relevant for training the SOM, and only confuses the description at this point:

"Additionally, profiles of wind speed (produced by interpolating the zonal and meridional components to the 5 m grid and then calculating total wind speed profiles) and LLJ characteristics in the context of the SOM patterns will be analyzed." (line 305)

As a side note, we use the level 2 radiosonde data, which is not GRUAN processed, but is rather Vaisala processed, because the level 2 data are more reliable at low altitudes. We have added a note earlier in the paper which states that we find the winds provided in the level 2 dataset (processed by Vaisala) to be reliable without additional filtering:

"…after determining that our results changed minimally when additional vertical averaging was applied to the winds (beyond the filtering already applied by Vaisala during their data processing), we find the original winds provided in Maturilli et al. (2021) to be sufficiently reliable for the current study." (line 158)

About Fig 2.: Maybe I missed it but why do I need 30 patterns to describe typical ABL stratifications? For example, what is the difference between pattern 27 and 28? By eye there is no difference. A technical comment on Fig 2: the pattern number and the number of observations is in the same font and partly not well visible - maybe you could provide a color background for the two set of numbers?

We have moved Fig. 2 in the original manuscript to the supplementary figures of the revised manuscript, as it is a rather technical figure, and does not intuitively demonstrate the nuanced differences between SOM patterns, which can all be better seen in Fig. 3 of the original manuscript (Fig. 2 of the revised manuscript) which highlights the differences in stability regime between seemingly similar patterns (e.g., pattern 27 is classified as WS-SSA, and pattern 28 is VSM-SSA due to varying ABL depths). We also hope that our added discussion on the utility of a SOM (line 242-275), and how this this applies to the current study (line 276-309) clarifies why 30 patterns are necessary.

For clarity of the figure (now Supplementary Fig. S1), we have:

- Removed the lines for all cases mapping to each pattern (this information is summarized well with the percentile profiles which we have retained)
- Removed the mean and median profiles
- Changed the colors to be more appealing and visible
- Put an opaque background behind the pattern numbers so they are more visible

Line 259ff: when reading this part, I immediately ask myself if the DH2 observations have a chance to cover all the different patterns because it didn't fly in the Polar night so it should miss the real stable conditions- right?

The fact that the DH2 did capture strongly stable conditions despite not flying in polar night shows that strongly stable conditions can happen outside of polar night (this is also evident by the seasonal SOM pattern and stability regime frequencies observed by the radiosonde, Fig. 3 and Fig. 5 of the revised manuscript). However, due to suggestions from the other reviewer, we have removed the DH2 analysis from the current paper, as it does not add much to the overall takeaways that can already be learned from the radiosonde data.

Line 267ff: Why is a SOM based on anomalies more visual? If you anticipate the result here, I immediately wonder why you used the gradient first and did not start with the analysis of the anomaly right away – I am confused here…

We actually did first try training the SOM with the anomaly profiles, but found that the SOM trained with the anomaly was largely failing to produce any patterns with a distinct near-neutral layer, or to highlight the distinct elevated inversion which is often present. So then we tried training the SOM with $d\theta_v/dz$ profiles, and found a much better representation of the range in height and strength of the θ_v inversion, as well as the differentiation between a weakly stable or near-neutral layer below a θ_v inversion. We have added some text to explain this, and have also adjusted the text to state that the anomaly profile is simply another way to visualize the data, but don't claim that it is the most intuitive way to visualize the data (as what is most intuitive may differ depending on the reader).

"We also tested the SOM trained with the θ_v profiles rather than the gradient (in the form of the θ_v anomaly compared to 1 km, to remove seasonal temperature dependence), but found that the range in height and strength of the θ_v inversion, as well as the differentiation between a weakly stable or near-neutral layer below a θ_v inversion, were not distinguished." (line 284)

"… a more tangible demonstration of the range of vertical structures present during MOSAiC is shown in Fig. 2 (Sect. 3.1) with the mean profiles of $d\theta_v/dz$ and θ_v anomaly for all radiosondes mapped to a given pattern." (line 299)

Line 279ff: I partly understand the motivation to define so many different stability regimes, but I fear that the usefulness for most readers is very limited. These 12 regimes are linked in the manuscript with 12 abbreviations that I definitely cannot remember and when these are mentioned and discussed in the text, I as a reader jump

back and forth to remember the abbreviations. This disrupts the flow of reading, at least for now, whether I can do much with the information or not.

We feel strongly that the use of all 12 regimes is important. In the Results section (e.g., with the histograms and box/whisker plots), we show discernable differences in frequencies and atmospheric characteristics both between the five different near-surface regimes (SS, MS, VSM, WS, and NN), as well as between the different aloft regimes within a certain near-surface regime (e.g., VSM-SSA, VSM-MSA, VSM-WSA). Thus, we feel is it justified to retain all of these regimes, in order to reveal important nuances in the results.

Line 285: A possible solution for a better reading flow could be to distinguish even more clearly between methods and observations - this is only one possibility but in some places these two aspects blur a bit.

We believe we have already well separated the description of the observations (Sect. 2.1-2.2) and the description of the methods (2.3-2.4). Whenever methods of deriving a quantity from the observations is included outside of Sect. 2.2 (e.g., the example referenced here, deriving a specific $d\theta_v/dz$ profile for stability regime identification) this has been done intentionally, because we think it would be confusing for a reader to follow what we did and for what purpose, if it were placed elsewhere. Thus, we choose not to reorganize the separation of observations and methods. However, if we have misunderstood your comment, please let us know.

line 286ff: Why Antarctica? I think MOSAiC should really be sufficient and citing a nonpublished paper from the other side of the world does not really help here…

The point is to demonstrate that the methods are robust across multiple polar locations (i.e., in Antarctica as well as the Arctic), to support the stability regime criteria which are new to this study. We clarify this by now saying:

"The stability regime definitions were developed alongside a similar SOM-based analysis of ABL profiles in Antarctica (Dice et al., submitted), which supports the robustness of these methods for classifying stability in polar regions." (line 317)

We hope by the time of publication of the current manuscript, the Antarctica paper will be in pre-print, but if it is not, we understand that we will need to remove the citation.

line 290ff: Why is the gradient in 42.5 m representative for the AGL? I understand that this value might be representative for the surface layer (at least in summer) but the entire AGL - or do I misunderstand? Please clarify.

Sometimes stability in the lowest ~10 m can differ from that in the rest of the ABL (e.g., there might be a very shallow well-mixed layer below 10 m due to enhanced mixing from the interaction between wind and surface roughness, but above 10 m the atmosphere is stable) but we find that the stability at 42.5 m is high enough to be representative of stability throughout the majority of the ABL. We have added clarifying text in a few places:

"Twelve stability regimes have been defined based on stability within the ABL (hereafter referred to as "near-surface" stability)…" (line 311)

"Since the stability criteria in part depend on stability within the ABL and some observations have an ABL height as low as 50 m, we first include a measurement of $d\theta_v/dz$ at 42.5 m (this determines the near-surface stability), calculated across a 15 m interval between 35 m (lowest point of the profile) and 50 m." (line 320)

Line 325ff: Why are you defining possible regimes that were never observed in the data from MOSAiC? Maybe you have some good reasons but just reading this sentence confuses me.

We have clarified the reasoning:

"VSM-WSA and WS are not represented by a SOM pattern, but do occur rarely in individual profiles, and thus are still defined in Table 2 (see Sect. 3.2 onward). While NN was never observed in an individual MOSAiC profile, we include its definition in Table 2 to support the use of these criteria for observations from other campaigns." (line 361)

Fig 3.: This figure (Fig. 3) contains a lot of information, and I suspect that most readers will have difficulty understanding all the lines and what they mean. Perhaps there is a way to make the diagrams a little clearer.

As described in a repones to a previous comment, Fig. 3 in the original manuscript (Fig. 2 in the revised manuscript) has been revised to improve clarity through the following steps:

- Removed the SOM profile – this is shown in what is now Supplementary Fig. S1, and it is not necessary for the discussion of this figure
- Made the font of the numbers and letters in each subplot bigger
- Made the axes fonts bigger
- Added a "subplot key" which demonstrates clearly what is shown in each subplot, so that a reader can refer to it, rather than digging through the figure caption to figure out what they are looking at.

The colored frame lines describing the regimes should be somewhat thicker to better distinguish the different regimes

We have made the colored frame lines thicker (see Fig. 2 in the revised manuscript)

Also, at this point I wonder how the ABL height is defined! For example, in pattern 8 it is quite difficult to estimate an AGL height even by eye. I assume that the inversion and the entrainment layer are not part of the AGL height according to your definition, right? Is then the term "mixed layer" more appropriate compared to AGL height? You should at least define the phrases carefully at a prominent place.

In addition to a description of how ABL height is calculated (as was already included in the original manuscript), at the same point in the text we have also added a description of what that means physically, in terms of what is included in the ABL per our method:

"ABL height from each radiosonde profile was determined using a bulk Richardson number (Ri_b) based approach in which the top of the ABL was identified as the first altitude in which Ri_b exceeds a critical value of 0.5 and remains above the critical value for at least 20 consecutive meters (Jozef et al., 2022). These criteria typically identify the ABL height as the bottom of the elevated virtual potential temperature (θ_v) inversion (or the bottom of the layer of enhanced θ_v inversion strength) for moderately stable to near-neutral conditions, and at the top of the most stable layer for conditions with a strong surface-based θ_v inversion. The methodology for calculating the Ri_b profile used to identify ABL height, as well as justification for the use of 0.5 as a critical value (rather than the more traditional value of 0.25) is described in Jozef et al. (2022)." (line 213)

Line 358ff: why "perhaps" - you should have the data to evaluate this "unique processes"!

We have added some solid evidence for our hypothesis:

"Pattern 4 is particularly interesting, as there is strong near-surface stability and an elevated region of enhanced stability around 600 m AGL, which may be explained by unique processes occurring primarily in summer. Reported visibility and ceilometer observations suggest a possible low fog layer and additional elevated cloud layer." (line 407).

Line 367: I am not convinced that the Arctic ABL is "always" stably stratified" - in particular in summer this is definitively not the case (see Tjernström et al.) So, I think to sell this as a "new finding" is going too far.

We agree that the Arctic ABL is not "always stably stratified," and were attempting to argue the opposite point – that near-neutral is also common. However, we recognize through further literature review that this is already a notion demonstrated in prior work. Thus, we no longer present this as a new finding, and also have moved this conclusion from the results section to a separate discussion/conclusions section, per another one of your recommendations.

"The SOM patterns (Fig. 2), frequency distribution of stability (Fig. 5a), and ABL height variability (Fig. 5b) highlight that near-surface stability during MOSAiC spanned from strongly stable with a shallow ABL to near-neutral with a deep ABL, with stable and near-neutral conditions occurring with similar frequencies. Stability aloft ranged from strongly to weakly stable. These findings are consistent with Persson et al. (2002), Tjernström and Graversen (2009), and Brooks et al. (2017)." (line 568).

Fig 4: most of the numbers are more or less invisible, at least the numbers in the upper line. Furthermore, black labels on a dark background are quite challenging. I think you should find a much better way to illustrate your point here

We have made many adjustments to improve Fig. 4 in the original manuscript (Fig. 3 in the revised manuscript):

- Removed the number in the upper center of each box which indicated the number of radiosonde profiles mapped to that SOM pattern. This information is included in Fig. 3 in the original manuscript (Fig. 2 in the revised manuscript), so is not needed here.
- Removed the number in the center of each box which indicated the number of cases in that pattern/season. We only retain the percentage, as this is the more valuable number for understanding the results
- Made the font of the numbers and letters in each box bigger.
- Added an opaque white box behind the percentage written in each box, so it can be better seen regardless of the shading in each box
- Made the seasonal subplots bigger
- We also include now the annual frequency subplot in this same figure, in response to your next comment (we discuss this figure in the text so you suggest this not be in the supplement).

Line 377: Again, I am not in favor of discussing material that is not in the manuscript but has been moved to a supplement. Regardless, I don't quite understand the following statement; DH2 couldn't fly in clouds - right? How then can DH2 observations cover all patterns when clouds affect stratification so much - or have I misunderstood something?

The DH2 can fly when there are clouds, but it is not supposed to fly into the actual cloud (it can go all the way to cloud base). This allowed the DH2 to observe the cloud-forced stratification scenarios. However, due to suggestions from the other reviewer, we have removed the DH2 analysis from the current paper.

Line 383ff: Do I understand correctly that LLJs are to be analyzed based on mean values within one individual pattern? As you have already noticed, this must - at least partially - average out the LLJ's with having their wind maximum at different heights - right? Then I have serious doubts whether this is the right method and whether one can make statements about LLJs in this way. However, it might be a misunderstanding on my side but then please clarify and consider rephrasing. In addition, there are from my point of view quite valid remarks of Günther Heinemann about the quality and (in)accuracy of wind measurements in the low altitude and the low-pass filtering of the wind data. Please take these comments seriously.

There may perhaps be a misunderstanding here. While we show average wind speed profiles, the LLJ characteristics (height and speed) are taken as the average of these LLJ characteristics identified from each individual profile in a given SOM pattern. Thus, while the LLJ is not necessarily visually evident in the average wind speed profiles due to slight differences in the heights of the LLJs in individual cases, it is still reasonable to analyze the LLJ characteristics as a mean of those from the individual cases. Additionally, we include the interquartile ranges (IQRs) to show the spread in the LLJ characteristics throughout all of the cases in each SOM pattern. We have revised the discussion to more clearly explain our methods, while more heavily noting what we learn when considering the IQRs of the LLJ characteristics. For example, on the right half of the SOM, the IQRs for LLJ height largely don't overlap with the IQRs for ABL height, whereas on the left half of the SOM, they do, which demonstrates our conclusion that for patterns on the right half of the SOM, the LLJ is largely decoupled from the ABL and thus does not weaken the near-surface stability (and vice versa for the left half of the SOM).

"To understand the influence of mechanical mixing on the stability structures presented by the SOM, we visualize average wind speed profiles for each SOM pattern; additionally, we analyze the LLJ characteristics for each pattern, as the average across all individual cases in each pattern (Fig. 4). As LLJ core height and speed varies across the cases in each pattern, the LLJ is often smoothed out in the average wind speed profile. Interestingly, the average LLJ height was found to be similar across all SOM patterns (roughly 400 m AGL). The higher ABL heights of the weaker stability patterns (WS and NN; on the left side of the SOM) place the LLJ closer to the ABL top than for the stronger stability patterns with lower ABL heights (SS, MS and VSM; on the right side of the SOM). Additionally, the interquartile ranges (IQR) of ABL height and LLJ height overlap for all patterns on the left half of the SOM, and for many patterns, the IQR of LLJ height extends below the average ABL height. Conversely, on the right half of the SOM, the IQR of ABL height and LLJ height only overlap for pattern 23. The LLJ speeds, 2 m wind speeds, and overall wind speed profiles have greater values for the patterns on the left half of the SOM (mean LLJ speed of 12.3 m s^{-1} and mean 2 m wind speed of 5.3 m s^{-1}), compared to the right half (mean LLJ speed of 9.7 m s^{-1} and mean 2 m wind speed of 3.3 m s^{-1}). The LLJ frequency for all SOM patterns is similar, showing that an LLJ was present for 67% – 84% of all observations mapped to any given pattern, with a median LLJ frequency of 76%." (line 418-431)

"Results regarding LLJ height, specifically its relationship to ABL height, support the notion that both baroclinicity and inertial oscillations contribute to LLJ formation in the Arctic. For the SS, MS, and the VSM regimes (represented by patterns on the right half of the SOM), the LLJ core was situated a greater distance above the ABL than for the WS and NN regimes (represented by patterns on the left half of the SOM). This greater distance suggests decoupling between the relatively stable ABL and the LLJ, which is consistent with inertial oscillations as an LLJ formation mechanism. The smaller distance between the ABL and LLJ core for the weaker stability regimes suggests greater coupling between the well-mixed ABL and the LLJ, so inertial oscillations are unlikely to be the formation mechanism, and rather baroclinicity is the more probable cause. The results show that such LLJs have faster speeds, in agreement with Jakobson et al. (2013). The similarity in LLJ core height despite varying stability occurs because of these two different formation mechanisms. Thus, an LLJ can be both a cause and an effect of stability. For a well-mixed or weakly stable ABL, LLJs contribute to the creation of the mechanical turbulence that mixes the ABL. For more strongly stable ABLs, an LLJ can be an effect of the strong stability if the above atmosphere becomes decoupled from the surface." (line 616-628)

Additionally, we have thoroughly considered Prof. Heinemann's remarks, and have shown through various testing metrics that the wind speeds we use are reliable. Specifically relating to your concern, we applied a few different vertical averaging ranges to the wind speed data (we tried a 30 m and 60 m running average), and reproduced the LLJ frequency and box/whisker plots, to see if the results differed from when the original wind profiles were used. We found the results to be largely unchanged, and thus conclude that we are sufficiently confident in the original wind speeds given in the level 2 soundings (used in this study), which had already undergone quite some vertical filtering/smoothing when Vaisala processed them, prior to their publication (see our responses to Prof. Heinemann's comments for more details).

Line 388: Physically, I don't understand this point: a LLJ is by definition linked to a more stable ABL because one precondition for a LLJ is the almost vanishing turbulent exchange coefficient - right? So how can you say that "Thus the LLJ is more closely coupled to the ABL in the weak stability cases…" Maybe I don't understand your point here but then you should clarify it.

Our response to a previous comment of yours is again relevant for this comment:

There are two primary situations in which an LLJ will occur in the Arctic. One such situation is the one you describe: the LLJ forms due to inertial oscillations and is decoupled from the (typically shallow and stable) ABL, such that wind shear from the LLJ does not weaken stability. The other situation is when an LLJ forms due to a baroclinic environment, where wind speeds decrease with altitude according to the thermal wind relationship, and surface friction reduces wind speeds to 0 at the surface, this resulting in an LLJ some distance above the surface. In this case the LLJ is not decoupled from the ABL, and can contribute to weakening of stability.

We have added more discussion in the introduction to clarify these two situations under which an LLJ may form, and the subsequent relationships with ABL stability.

"There are two primary forcing mechanisms for LLJs in the Arctic: baroclinicity and inertial oscillations. Baroclinicity in the Arctic most often occurs near the ice edge (Brümmer & Thiemann, 2002) or due to the passing of a transient cyclone (Jakobson et al., 2013) which creates regions of enhanced temperature contrasts (Koyama et al., 2017). Depending on the wind direction, the horizontal temperature gradient causes the geostrophic wind speed to decrease with height according to the thermal wind relationship (Stull, 1988). This, paired with diminishment of wind speeds at the surface due to friction (Stull, 1988), contributes to the formation of an LLJ at some distance above the surface, typically just above the ABL (Brümmer & Thiemann, 2002). Thus, an LLJ forced by baroclinicity is typically coupled to the surface, and can cause weakening of stability within the ABL due to enhanced shear below the jet core (Banta, 2008; Egerer et al., 2023).

Inertial oscillations in the Arctic can be induced after well-mixed conditions are replaced by increased near-surface stability, for example, after the passing of a storm (Andreas et al., 2010a; Jakobson et al., 2013). In such cases, air aloft becomes decoupled from the surface, ceasing frictional drag, which, along with the impact of the Coriolis force, allows the winds aloft to accelerate to supergeostrophic speeds (Blackadar, 1957; Stull, 1988; Jakobson et al., 2013). For both LLJs forced by baroclinicity and inertial oscillations, enhanced wind shear above the jet core may also contribute to turbulent mixing above the LLJ. A previous study conducted on LLJs in the central Arctic between 25 April to 31 August of 2007 found an LLJ frequency of 46%, a mean LLJ core speed of 7.1 m s^{-1}, and LLJ core altitude typically between 100 and 500 m, with faster LLJs having the jet core located inside the ABL (Jakobson et al., 2013). Additional observational studies in the central Arctic have reported an LLJ frequency of 60-80%, with a higher frequency of LLJs over the pack ice (72%) versus in the marginal ice zone (66%) (Tian et al., 2020; ReVelle and Nilsson, 2008). A similar study to that described in the current paper found LLJs to be present more than 40% of the time in the central Arctic, with typical height below 400 m and speed between 6 and 14 m s^{-1} (Lopez-Garcia et al., 2022). Model studies of central Arctic LLJs have documented a lower frequency, of 20-25% (Tuononen et al., 2015)." (line 78-99)

We have also clarified how these two LLJ formation situations play into the LLJ results that we see, during the discussion of the LLJ results.

"Results regarding LLJ height, specifically its relationship to ABL height, support the notion that both baroclinicity and inertial oscillations contribute to LLJ formation in the Arctic. For the SS, MS, and the VSM regimes (represented by patterns on the right half of the SOM), the LLJ core was situated a greater distance above the ABL than for the WS and NN regimes (represented by patterns on the left half of the SOM). This greater distance suggests decoupling between the relatively stable ABL and the LLJ, which is consistent with inertial oscillations as an LLJ formation mechanism. The smaller distance between the ABL and LLJ core for the weaker stability regimes suggests greater coupling between the well-mixed ABL and the LLJ, so inertial oscillations are unlikely to be the formation mechanism, and rather baroclinicity is the more probable cause. The results show that such LLJs have faster speeds, in agreement with Jakobson et al. (2013). The similarity in LLJ core height despite varying stability occurs because of these two different formation mechanisms. Thus, an LLJ can be both a cause and an effect of stability. For a well-mixed or weakly stable ABL, LLJs contribute to the creation of the mechanical turbulence that mixes the ABL. For more strongly stable ABLs, an LLJ can be an effect of the strong stability if the above atmosphere becomes decoupled from the surface." (line 616-628)

Fig. 5: Again, too much information and details in the plots.

We have revised the Fig. 5 in the original manuscript (Fig. 4 in the revised manuscript) to be make it more clear to understand:

- Made the font of the numbers and letters in each subplot bigger
- Made the axes fonts bigger

- Removed the number in the upper center of each box which indicated the number of radiosonde profiles mapped to that SOM pattern. This information is included in Fig. 3 in the original manuscript (Fig. 2 in the revised manuscript), so is not needed here.
- Added a “subplot key” which demonstrates clearly what is shown in each subplot, so that a reader can refer to it, rather than digging through the figure description to figure out what they are looking at.
- Due to the above bullet point, we have removed “LLJ:” and “2m” from each subplot, as the “subplot key” now indicates what these numbers mean.

Line 415 ff: The is probably one of the most prominent places to say: I am lost in all the details and even more I lost track due to the huge number of abbreviations….

We have restructured how we share the stability regime frequency, focusing on the individual near-surface regimes first, and then talking about aloft regimes more broadly. We hope this is a more intuitive and interesting way for a reader to digest these results:

“The most frequent near-surface regime observed was NN (37% of profiles), followed by VSM (27% of profiles), MS (14% of profiles), and SS (13% of profiles) in decreasing order. WS was observed least frequently (9% of profiles). The total frequency of a stable ABL (combining SS, MS, and WS frequencies) was 36%, just slightly less than the frequency of a near-neutral ABL. The most frequent regime observed aloft was -SSA (66% of VSM cases, 54% of WS cases, and 60% of NN cases had strong stability aloft) followed by -MSA (31% of VSM cases, 39% of WS cases, and 35% of NN cases had moderate stability aloft). Weak stability aloft was infrequently observed (3% of VSM cases, 7% of WS cases, and 5% of NN cases had weak stability aloft). The overall most common regime was NN-SSA, followed by VSM-SSA.” (line 457)

line 454: If a parameter x changes over the depth of the ABL it is not the same as the gradient dx/dz – right?

We have adjusted the wording to more clearly state what we mean. Also, we now only include dV/dz in the revised manuscript, as the $d\theta_v/dz$ result was largely a function of how we defined the stability regimes.

“…change in horizontal wind speed between the surface and top of the ABL (dV/dz)…” (line 471)

Line 455ff: If you cite a figure within the main text, it should be included in the main text - why has it been shifted to a supplement?

It is relatively common practice to include some figures in a supplement which are mentioned in the main text but are not crucial to the understanding of the paper, and rather supply additional details to an interested reader. We have moved most of the figures originally included in the supplement in the original submission to the main text of the revised manuscript, per your suggestion, but we choose to keep the statistical significance figures in the supplement. As the current paper is already quite long and complex, we do not think that adding the statistical significance figures to the main text will help with improve the readability of the paper. The main takeaways of the study can largely be discerned without them, though we do still provide a brief discussion of them in the main text.

line 463: "...the fact that we see this drastic increase also supports the choice of this threshold...." can you please explain this in a little bit more detail?

We have modified the sentence to better state our point:

“The jump in ABL height between the VSM and WS regimes is in part a product of how we define the VSM regime (which requires an ABL height of 125 m or less). However, the magnitude of the increase in ABL height between the VSM regime (mean of 85 m) and WS regimes (mean of 221 m) demonstrates that this threshold was meaningful.” (line 478)

line 465ff: I cannot follow this sentence at all - please double check and consider rephrasing.

We have modified the sentence to clarify our point while not including unnecessary information:

"Additionally, we find that ABL height increases as stability aloft decreases (e.g., the mean ABL height for WS-MSA is greater than the mean ABL height for WS-SSA)." (line 481)

line 468ff: It is quite unusual to start a sentence with a equation - I would avoid it. Furthermore: "shear" => "wind shear".

We have made sure to not start a sentence with an equation, and have changed "shear" to "wind shear":

"SS and MS had the greatest (largely above average) wind shear (dV/dz) within the ABL (Fig. 5c)." (line 483)

So why do you mention Ri with index "b" (for bulk?!) here? If you mention the (local?) gradients then you have the basics for the classical local Ri definition - right?

This is a good point. We have, however, removed the discussion of Richardson number as part of our adjustment of the content to only include the most interesting and new results.

line 475ff: again, a reference outside the paper is not helpful and I suggest to avoiding this. Furthermore, I do not understand your conclusion about the physically meaningful definition of the regimes - please specify what you mean here.

As mentioned in a response to a previous comment, we choose to keep the statistical significance figures in the Supplement, as the statement of what can be gleaned from these figures is sufficiently meaningful, while an interested reader may find additional details in the supplementary figure if they would like.

Next, we have tried to clarify what we mean about the physically meaningful definitions of the regimes, which we have merged with the similar discussion with regards to u_*:

"Significant differences in dV/dz and u_* between most pairs of stability regimes (Fig. S2b) highlights that turbulence properties are distinct for each regime. While perhaps an intuitive statement, it is important to confirm that physically meaningful differences in stability regimes classified largely based on thermal gradient are found for mechanical processes, as well as for turbulence measured by the met tower (a separate platform than the radiosondes used to classify stability regime). This confirmation supports the validity of the stability regime criteria defined in Sect. 2.4." (line 487)

line 477: what is exactly meant by "dV/dz result" - from my point of view this makes no sense

We have clarified this sentence:

"For the weaker stability regimes (WS and NN), winds vary less with height due to greater mixing, which is a common behavior of winds within a weakly stable or near-neutral ABL (Wallace and Hobbs, 2006)." (line 483)

line 479ff: It is surprising that you start the discussion with the Richardson number and now you move on to the friction velocity - so why? Furthermore, you correctly mentioned that Ri describes more the tendency of turbulence development but it is not a measure of the degree of turbulence or the intensity but you concluded already that the near-surface atmosphere is always turbulent. From my point of view this is going too far. This part also needs some careful reconsiderations and not only rephrasing.

As mentioned previously, we have removed discussion of the Richardson number from the manuscript, and thus do not discuss the implication of the Richardson number for turbulence. We instead focus the discussion on u_*.

line 487: Well, turbulence itself might describe a flow but probably not the Arctic - this makes no sense. Maybe I missed it but I suggest to define u_star at the place of first occurrence (around line 145 or so). Also, I suggest to use the word „increased" turbulence.

We have removed the statement that the Arctic environment is characterized by turbulence. Also, we have added a more in-depth description of bulk u_* used in this study when it was first introduced:

"Atmospheric observations of … bulk friction velocity (a theoretical wind speed that expresses the magnitude of stress exerted by wind flowing over the Earth's surface, indicating the magnitude of turbulence; u_*), come from a 10 m meteorological tower… and provide information about near-surface turbulence at the time of each radiosonde launch. Bulk u_* was chosen, as opposed to the standard eddy-covariance value, as the bulk parameterization considers both wind fluctuations and latent heat fluxes (developed using guidance from eddy-covariance data collected during SHEBA; Andreas et al., 2010b) which is more comparable to u_* used in models (e.g., Fairall et al., 2003)." (line 167-175)

Lastly, we have removed the statement that increased u_* indicates increased turbulence here, as we have included a more thorough description of u_* when it is first introduced.

488ff: this not really surprising and I think this statement does not need a supplementary figure - right?

We have revised the discussion to highlight the importance of sharing this unsurprising result:

"Significant differences in dV/dz and u_* between most pairs of stability regimes (Fig. S2b) highlights that turbulence properties are distinct for each regime. While perhaps an intuitive statement, it is important to confirm that physically meaningful differences in stability regimes classified largely based on thermal gradient are found for mechanical processes, as well as for turbulence measured by the met tower (a separate platform than the radiosondes used to classify stability regime). This confirmation supports the validity of the stability regime criteria defined in Sect. 2.4." (line 487)

492ff: I am not surprised that u_star and and wind shear do not show a clear dependency because it is a Richardson number problem - but is this a conclusion as you mentioned?

We have removed the discussion on the dependence (or lack thereof) of u_* and dV/dz from this section, as the revised manuscript instead includes a separate discussion section. With the separation of the "Results" and the "Discussion and conclusions" into different sections, and the removal of Richardson number results from the paper, we hope the discussion on these points is more interesting and the primary takeaways are better highlighted:

"Despite slower wind speeds and lesser u_* for stronger near-surface stability, wind shear (dV/dz) over the depth of the ABL increases with increasing stability, revealing that in strong stability cases, static stability suppresses mechanically generated turbulence, promoting continued ABL stability despite high amounts of wind shear." (line 607)

line 495ff: I cannot follow your argumentation, in particular the last part "...when stability aloft is greater." So, is there a connection between stability at higher altitudes and surface layer mixing? Please explain what you mean here.

This conclusion has been moved to the new, separate "Discussion and conclusions" section, as several results throughout the paper highlight the same conclusion. Additionally, we had revised the description of this conclusion to be more clear:

"While LLJ speed and u_* increase with decreasing near-surface stability, the opposite relationship is seen for stability aloft: LLJ speed and u_* values are greatest when stability aloft is greatest. These results suggest that when the atmosphere is inclined to be strongly stable (e.g., in the absence of clouds during winter), more mechanically generated turbulence is required to fully mix out the near-surface layer than if the atmosphere is inclined to be weakly stable (e.g., in the presence of clouds).” (line 611)

Fig 7 (and maybe other figures as well): you put the labels of the y-axis on the x-axis which is formally not correct and took me a time to understand the plot - please change this labeling.

This change has been made.

line 511ff: Maybe I am wrong and I don't have the details of Banta et al in mind but how can a LLJ exists (or develop) in a well-mixed ABL? Maybe during the transition to a classical Ekman-layer like ABL this makes sense but from a theoretical point of view a LLJ and a well-mixed layer are exclusive; the turbulent exchange coefficient K_m has tend to zero to decouple the LLJ region from the surface - right? In fact, that is the background for your comment in line 513ff but I think this is a precondition for the development of a LLJ....

We have added some text in the introduction to explain how an LLJ can exist when there is a well-mixed ABL. Thus, we believe now that the proper context to support the results presented in the LLJ section are provided.

“Another common feature of the Arctic lower atmosphere is a low-level jet (LLJ), which is a local maximum in the wind speed profile below 1.5 km (Tuononen et al., 2015) that is at least 2 m s^{-1} greater than wind speed minima above and below (Stull, 1988). There are two primary forcing mechanisms for LLJs in the Arctic: baroclinicity and inertial oscillations. Baroclinicity in the Arctic most often occurs near the ice edge (Brümmer & Thiemann, 2002) or due to the passing of a transient cyclone (Jakobson et al., 2013) which creates regions of enhanced temperature contrasts (Koyama et al., 2017). Depending on the wind direction, the horizontal temperature gradient causes the geostrophic wind speed to decrease with height according to the thermal wind relationship (Stull, 1988). This, paired with diminishment of wind speeds at the surface due to friction (Stull, 1988), contributes to the formation of an LLJ at some distance above the surface, typically just above the ABL (Brümmer & Thiemann, 2002). Thus, an LLJ forced by baroclinicity is typically coupled to the surface, and can cause weakening of stability within the ABL due to enhanced shear below the jet core (Banta, 2008; Egerer et al., 2023).” (line 76-86)

line 515ff: maybe you mentioned it earlier but is the "LLJ speed" defined as the maximum speed in the LLJ core or the difference to the surrounding?

The LLJ speed is the maximum speed in the LLJ core. This information was stated on line 194 – 195 of the original manuscript. It can be found in the revised manuscript on line 226-228:

“If an LLJ was found, we identified the LLJ core altitude as the altitude of the maximum in the wind speed, and the LLJ speed as the wind speed at that altitude (Jakobson et al., 2013).”

line 548ff: I think this is physically not really meaningful, maybe it should read as: "...it needs more wind shear in a more stable environment to create mixing ..." or similar. Furthermore, I think the phrase "hypotheses" is going a little bit too far because you never formulated a hypothesis which can now be verified or falsified...

We revised this description following your suggestion to be more physically meaningful, and now include it in the separate “Discussion and conclusions” section. The “hypothesis” we were referring to was previously stated in association with the dV/dz and u_* results, where more turbulence was necessary to mix out an environment with greater stability aloft. By separating the “Results” and “Discussion and conclusions”, we can now make this conclusion one time, with support from the various results:

“While LLJ speed and u_* increase with decreasing near-surface stability, the opposite relationship is seen for stability aloft: LLJ speed and u_* values are greatest when stability aloft is greatest. These results suggest that when the atmosphere is inclined to be strongly stable (e.g., in the absence of clouds during winter), more mechanically generated turbulence is required to fully mix out the near-surface layer than if the atmosphere is inclined to be weakly stable (e.g., in the presence of clouds).” (line 611)

line 555: what do you exactly mean with "excess turbulence"?

The intent was to say that, even in ubiquitously high wind speed environments, a LLJ would contribute to more turbulence production than if there was not an LLJ, as the LLJ speed exceeds that of the rest of the winds throughout the column. We have clarified this wording:

“However, such LLJs can still be important because even if the wind speeds are fast throughout the entire profile up to 1.5 km (for example, during a storm), the slightly greater speed of the LLJ beyond that of the ubiquitously high winds throughout the column supports the production of increased turbulence in the ABL compared to without an LLJ.” (line 634)

Fig 8.: See my comment on Fig 7 about the axis labels

This issue has been fixed for all figures.

Line 567ff: About your supplement: it just includes many further figures which are partly mentioned within the main manuscript but not explained and deeply discussed. I think this is not the right way to use a supplement because the main manuscript should be readable by itself without reading the supplement. If you have a distinct and interesting topic which might be useful for some readers but distract from the red line of the manuscript a supplement might be the right choice but if you have simply too many figures which you want to mention in the manuscript you cannot just move them into the supplement and refer to them (different to an appendix).

The reviewer makes a good point. To address this comment, we have reconsidered which figures to include in the paper, and retain only those which reveal the most interesting and/or new results, and have remove the other figures/panels.

line 577ff: Do you consider TIs as a cause or effect of ABL development?

As noted in previous responses, we recognize that the TI results are largely a function of how stability regime is defined, and thus the results are not the most pertinent. Thus, we have removed all figures and text related to temperature inversions to provide a more concise manuscript.

Line 578ff: About the TI analysis: I am a little bit skeptical about this analysis because I wonder how much of this analysis is based at least partly on self-correlation because the stability regimes are based on the temperature gradients - right? Maybe this should be discussed at least a little bit before interpreting the results.

See our response to your above comment.

Line 581 and general: Maybe better distinguish between explaining figures and results, than interpreting them and finally compare with other studies - the structure is often a bit confusing and you jump back and force

We have restructured the paper to address this and other comments. We now purely share results in the “Results” section, and include a separate “Discussion and conclusions” section where we discuss interpretations of the results, and comparisons to other studies.

line 591ff: I cannot follow your argumentation about the potential for exchange of momentum when TI is well above ABL height - what do you mean? I have the feeling that here is a lot of speculation on play but a careful physical interpretation is missing.

See our response to your previous comments regarding TIs.

line 603: As mentioned earlier, I have the feeling that this part is based on a big portion of self-correlation which has to be ruled out before the interpretation.

See our response to your previous comments regarding TIs.

line 621: why using a decimal value for the second cloud base height? Why 6.1 km and not 6 km??

The original threshold of 6.1 km was chosen, as this threshold was given by an online source. However, the thresholds for low, mid-level, and high clouds vary between sources, so in the end, choosing a threshold is a bit arbitrary. Thus, for simplicity, we have changed the mid-level cloud base height threshold to 6 km. The results are essentially unchanged.

line 633ff: What do you mean with the statement that "… a regime is driven by the radiative signature of clouds..."? and further on in line 634 what are those other mechanisms? This explanation is hardly to follow and needs some careful rephrasing.

We intended to mean that the VSM and NN regimes are driven by the enhanced downwelling radiation produced by clouds versus clear skies. The other mechanisms we refer to are primarily longwave cooling and wind speeds. We now clarify all these things when the interpretation of the cloud-related results are discussed in the "Discussion and conclusions" section.

"Low clouds, correlated with greater LWP, were observed with greater frequency for cases with weaker stability both within the ABL and aloft, highlighting the ability of low clouds and enhanced moisture content to support turbulent mixing both near the surface through enhanced downwelling longwave radiation, and below cloud base though cloud top radiative cooling. In such cases, a well-mixed ABL can be coupled to the cloud layer and extend through the depth of the cloud to cloud top, though a shallow stable layer may decouple a well-mixed ABL from a low cloud. Conversely, mid-level and high clouds were observed with greater frequency for cases with stronger stability, highlighting that in such cases, the cloud is likely to be decoupled from the surface, allowing the strong stability to persist." (line 644)

The summary will certainly need to be completely revised when the previous analyses have been appropriately re-sorted and revised - so I have refrained from detailed comments on this chapter now.

We have indeed revised the summary greatly based on the reorganization of the paper, and our original "Summary and Conclusions" section is now a "Discussion and Conclusions" section. We are open to hearing your feedback on this new section during the next round of reviews.