Response to Anonymous Referee #1

We appreciate the time and effort that you have dedicated to provide valuable feedback and helpful comments, which significantly improved our manuscript.

During the review process, we noticed an inaccuracy in the calibration of the hot-wire sensors that warranted a recalibration of the data. The changes are minimal and do not affect the conclusions of this manuscript, but all analysis were performed again with the re-calibrated data. The new data set is currently being in the publication process of PANGAEA and accepted as a data descriptor publication in Nature Scientific Data.

Here is the point-by-point response to your comments. We included the original comment in *blue and italics* with the lines referring to the first manuscript. Our response follows in black, and the line numbers refer to the revised version.

This study evaluated various observing methods used to identify the surface mixed layer (SML) height during the Multidisciplinary drifting Observatory for the Study of Arctic Climate (MOSAiC) campaign, using in-situ turbulence measurements from a tethered balloon platform as ground truth for the SML height. The study finds a critical bulk Richardson number (~0.12) that can be extrapolated for use with radiosonde observations. This appears to be one of the main conclusions of the study and is worthy of documentation based on careful analysis by the investigators. However, I found the scientific exploration of cloudless versus cloudy boundary layers to be lacking and feel that the basis for using other methods (particularly Monin-Obukhov similarity theory) was not properly motivated. What can the community use from these conclusions going forward? I think emphasizing the derivation of a critical bulk Richardson number is indeed significant and important. However, I found little value in defining the differences in the SML height for cloudless versus cloudy boundary layers without any type of physical interpretation. Moreover, the manuscript was a little difficult to follow in terms of structure and most importantly in terms of grammatical fluency. Some of the figures (particularly Fig. 11) also seemed to display results/statistics that weren't entirely relevant and were hard to follow. While I think the scientific merit of the study is there and I commend the investigators for this, there needs to be some significant restructuring and improvements for it be acceptable for publication. I am therefore recommending reconsideration following major revisions.

In response to your general comment, we have revised our introduction and elaborated on the motivation to distinguish between the two types of the Arctic atmospheric boundary layer (ABL): cloudless and cloudy conditions.
We agree also that using the Monin-Obukhov similarity approach was not explicitly explained and motivated, and we have now extended the description (see Sect. 2.2.3) to include this. We consider MO as a useful complement to estimate the surface mixing layer (SML) height for the time between airborne measurements. However, this is only possible if the boundary conditions for this method are met.
We agree that parts of the manuscript were difficult to read and hopefully have now improved the reading flow. Furthermore, we have followed your advice and improved Fig. 11, as this figure was quite difficult to interpret.

We hope that we have sufficiently addressed your minor and major concerns with our changes to the manuscript described below.

General comments

- **Section 2.2:** *It's not entirely clear what you plan to do here. After reading through the entire section it seems you are describing the different methods available to calculate the SML height, and I think it's great to discuss the advantages and disadvantages of each method. But when introducing the section, I think it would help the reader to more explicitly state that you are going to use a variety of calculations and that you consider one of them to be a superior method given the instrument availability. Moreover, I think emphasizing the use of in-situ measurements to derive a critical bulk Richardson number for use with radiosonde measurements is prudent.*
  Response: Thank you for this valuable comment. We have followed your suggestion and summarized our approaches at the beginning of Section 2.2:

  "In this chapter, we present several methods for determining SML heights based on different available data sets. First, we discuss a method to derive SML heights that benefits directly from local turbulence measurements and will serve as a reference for other approaches in the following (Sect. 2.2.1). A second method is based on the mean temperature and wind speed profiles in the context of the bulk Richardson number criterion (Sect. 2.2.2). Finally, we compare the previous results with SML heights based on the application of the Monin-Obukhov similarity theory and thus on the determination of surface fluxes (Sect. 2.2.3)." Lines 149-153

  We have also expanded the introduction and already included a brief description of the different approaches there.

  "By observing turbulence by in situ measurements, the SML height can be derived directly. One can either define the SML height as the height where ε drops significantly with height, or when the flow is considered non-turbulent based on a "turbulence threshold" (Shupe et al., 2013; Brooks et al., 2017). [...] We use the turbulence-based SML heights as reference to derive a critical bulk Richardson number for the winter and spring. The in situ turbulence perspective not only allows to derive a critical value but also to evaluate the bulk Richardson number approach." Lines 71-83

- **Section 3.3:** *Can you provide any speculation as to why the critical bulk Richardson number estimates are so much smaller than the theoretical value?*

  Response: For the bulk Richardson number method, the mean temperature and wind gradient of a layer is approximated by the difference from the surface to the layer height. It is assumed that the wind speed at the surface tends to zero, but the surface skin temperature is measured individually, and there are already different approaches in the application. By definition, the skin temperature should be used. However, this is rarely measured and the temperature at 2m height is used instead. This value, especially for radiosondes, is still subject to some measurement errors and, in particular, does not correspond to the skin temperature. The determination of the critical Richardson number is therefore directly related to the method used and is not universal. To show the influence of the method for calculating Ri_b, we have added the analysis of Ri_b with the skin temperature as a surface reference. Although this does not explain all the differences between our critical value and the "theoretical value", it does show how much the critical Richardson number depends on the type of calculation and the surface reference used.

In Sect. 2.2.2, we mentioned the problem of the surface reference temperature by adding the following paragraph:

“According to the basic concept of the surface bulk Richardson number approach, the lower reference level is the surface where the mean horizontal wind velocity equals zero U ($z = z_0$) and θ_0 is the skin temperature. Since other studies are often based on radiosonde observations only, the temperature at 2 m is then typically used for θ_0, even though this value may differ from the skin temperature, especially for stably stratified conditions with strong surface temperature gradients. To be consistent with other studies, we first use temperature observations at 2 m from the *Met City* tower for θ_0. Then, for comparison, the same analysis is performed using the skin temperature measured during MOSAiC.” Lines 195-200

In Sect. 3.3, we have included the following paragraph to compare the influence of the different surface temperatures on the critical bulk Ri number:

“Applying the same analysis with the skin temperature as θ_0, we derive $Ri_{bc} = 0.16$ (Fig. 6). This difference shows clearly how strongly the derived value for Ri_{bc} depends on the selected surface reference. Furthermore, we can assume that the temperature measured at a height of 2 m on a mast is also significantly more accurate than a corresponding measurement with a radiosonde, which has comparatively large inaccuracies in the lower ranges.” Lines 283-286

– **Section 3.5:** *I found the use of Fig. 11 alone to be rather confusing in the context of this discussion. The intent of the discussion appears to be describing the applicability of MO theory for reasonably locating the SML height. While I think it’s interesting to point out where MO theory fails and why, this could be better captured by tweaking Fig. 11 to (1) show bounded regions, for example via shading, that are discussed in the text and (2) by perhaps using a twin axis on the top of each panel that corresponds to profile number. You often discuss the profile number, but one has to refer all the way back to Fig. 2 to know the profile numbers to which you are referring. In general, could your conclusions from Section 3.5 be summarized in a better way, for example something like repeating Fig. 8 (which made the comparison very clear)?*

Response: We fully agree that Fig. 11 was confusing and have edited the figure to improve clarity according to your suggestions. We no longer show the temporal progress but have decided to show the profiles stacked as in Fig. 2. The x-axis now shows the profile numbers. The periods discussed in the text are highlighted with bounding and numbered in Roman. In addition, we have added a twin axis on top of the figure to give the month of the year. The shaded area in the plot shows the corresponding daylight condition as in the first version. To further improve the readability of the figure, we have decided to remove the Monin-Obukhov length L in a separate panel. In the text, we describe that L is used to derive h_{MO} and that no h_{MO} can be derived if L is negative.
While we agree that a scatter plot can be very helpful for interpretation, this is unfortunately not straightforward. The MO method complements the balloon perspective under certain conditions and fails when there is a second source of turbulence aloft (as in the case of clouds). These constraints complicate the interpretation of a scatter plot and, therefore, it is not possible to draw a general conclusion about whether the MO method over- or underestimates the SML heights. Thus, we aim to show that the MO method can bridge gaps between balloon measurements but is limited with respect to clouds and the occurrence of LLJs. We have added an explanation in Sect. 2.2.3.

"With the Monin-Obukhov scaling method, referred to as the MO method hereafter, the SML height can be derived continuously and complements balloon-borne SML height estimates. However, Eq. (5) is only valid for stable stratification ($B_s < 0$ and $L > 0$) and the method fails if further sources of turbulence are present at higher levels (such as clouds or Low-Level Jets (LLJ)). Again, the in situ turbulence method is helpful to assess the applicability of the MO method (Sect. 3.5)." Lines 225-229

Thank you for your valuable feedback. We hope that our changes have addressed your concerns.

- **Summary and conclusion:** *The Summary and Conclusion section leaves a lot to be desired. I think you could elaborate quite a bit on what you intended to do with this study. It seems you are comparing various methods of identifying SML height using a turbulence-based method as ground truth. This is a unique aspect of this study that should be emphasized more. For example, when you discuss "confusion in the literature" on Line 344, this is the first time this supposed confusion is mentioned. I think you should be more explicit in stating that for this study and future MOSAiC studies, you determined that a critical bulk Richardson number of 0.12 was found and can be reliably used going forward (perhaps even give this exact value in the abstract). This was emphasized appropriately in the abstract but not enough in the final section, in my opinion. Moreover, the last paragraph is very short, but it seems like you are starting to speculate about remaining science questions that need to be addressed and how improved datasets can be used to address those questions. I think this is worthy of more discussion here. Discussion of future needs for estimating the SML height would also be useful. If the in-situ turbulence-based method is the best, what do future campaigns need in order to make this a reality?*

  Response: We strongly agree with your comment and have completely rewritten the "Summary and conclusion" section. Based on your comments, we have extended the discussion of the intention of our study and further developed its unique aspect with in situ turbulence-based SML height estimates. We followed your suggestion to include recommendations for further studies by discussing the advantages and disadvantages of the different methods and their limitations.

  We finish the section now with our main conclusion:

  "Finally we come to the conclusion that turbulence profile measurements are the most reliable method for SML height determination but at least profile measurements with radiosondes are necessary to either determine the SML height using the Richardson approach or to check the boundary conditions in combination with MO theory." Lines 422ff

- **General:** *In general, the separation of cloudless versus cloudy boundary layers lacks any significant conclusions. I think readers would appreciate some extrapolation of these results with some physical interpretation. Why is determining the SML height for these two environments important? Are there modeling implications? Does it mean anything regarding cloud formation processes? As it stands, you seem to have arbitrarily separated the two, but the reader is left wondering why you did this in the first place.*

  Response: Thanks for your comment. We have addressed the separation between cloud and cloudy conditions in the Introduction and refer to the study of Solomon et al. (2023).

  "We have distinguished between those two states of the atmosphere, as many models have difficulties to reproduce the bimodal distribution of the terrestrial radiation (Solomon et al., 2023)." Lines 39-40

As the height of the surface mixing layer is a parameter directly influenced by the presence of clouds, this separation was made to see whether a critical value to estimate the SML height using the bulk Richardson number approach needs to consider cloud conditions. We agree that this was not well explained in our first version of the manuscript and hope the reader is now adequately guided.

- **General:** *I commend the investigators for being brief and concise, but there is considerable room for further explanation, particularly regarding your methods and conclusions. I would take the time and space to elaborate quite a bit about your approach, the importance of it, and what the important conclusions are.*
  Response: We agree and refer to our comment about the summary and conclusion which has been completely revised.

<u>Specific comments</u>

- **Line 23:** *Perhaps be a little more explicit regarding the vertical exchange under stable stratification. Vertical exchange of what? Aerosols? Heat? Momentum?*
  Response: We have complemented this sentence to "the vertical exchange of energy" as this is the main driver of amplified warming in Arctic winter. Please see Lines 25-26.

- **Line 31:** *"Sea ice surface" itself is not a condition; perhaps rephrase to state what condition the sea ice surface produces that makes it unique*
  Response: Thanks for your remark. We have changed this sentence:

  "The Arctic ABL is formed under unique conditions, such as strong cooling of the sea ice surface and the lack of solar radiation during winter, which favors the evolution of stable atmospheric layering (Persson et al., 2002; Tjernström and Graversen, 2009; Morrison et al., 2012; Brooks et al., 2017)." Lines 34-36

- **Line 48:** *Become even more important for what? Some of these statements are very vague and could use more specificity.*
  Response: In the Arctic, long-lasting low-level clouds are especially important for the surface radiative energy budget as they alter the radiative fluxes at the surface. To name this directly, we have changed this sentence.
  "While cloudless conditions are rather scarce in the Arctic (Intrieri et al., 2002b), frequently occurring low-level mixed-phase clouds are of major importance for the surface radiative energy budget." Lines 53-55

- **Line 65:** *I think you need to introduce what exactly the Richardson number is (ratio of buoyancy to shear) rather than just saying what it is used for. This is a big part of your study, but a brief description of how to interpret the Richardson number would be useful.*
  Response: We have now extended the introduction to the bulk Richardson number and its use to derive the SML height based on a critical value. We hope that this makes the interpretation of the bulk Richardson number clear.
  "Another approach to define the SML height applies the bulk Richardson number Ri_b (Andreas et al., 2000; Zilitinkevich and Baklanov, 2002; Dai et al., 2014; Zhang et al., 2014; Jozef et al., 2022; Peng et

al., 2023). Ri_b is derived from the ratio between shear and buoyancy and is a measure of whether turbulence tends to increase or decrease. The SML height is defined by the height where turbulence can not be sustained because the bulk Richardson number exceeds a critical value of Ri_b (Zilitinkevich and Baklanov, 2002; Andreas et al., 2000). This critical value, however, is under discussion and varies, for example, among sites (Vickers and Mahrt, 2004)." Lines 75-80

- **Lines 70-71:** *Please include a reference for Monin-Obukhov similarity theory, and perhaps a brief description of what exactly it is.*
  Response: We have added a reference to the Monin-Obukhov similarity theory:

  "Further, continuous, surface-based energy flux measurements can be used to estimate the SML height using Monin-Obukhov similarity theory (Zilitinkevich, 1972; Vickers and Mahrt, 2004). The Monin-Obukhov similarity theory describes the near-surface turbulent exchange processes based on surface measurements. This approach can complement the balloon-borne SML height estimates between balloon launches." Lines 83-87

- **Line 99:** *You haven't yet defined the turbulence probe here, though I assume it is the hot wire anemometer package introduced on line 90. Perhaps specifically call this the turbulence probe on line 90.*
  Response: You are right, thank you. We have revised this and introduced the term 'turbulence probe', which we use to refer to the hot-wire anemometer package.

  "A hot-wire anemometer package specifically designed for turbulence observations (Egerer et al., 2019), hereafter referred to as 'turbulence probe', was used to measure the data of this study." Lines 107-108

- **Lines 105-106:** *I don't think you've described the irradiance measurements/instrumentation yet.*
  Response: We recognize that we missed a proper description of the measurements and the instrumentation and added this in Section 2.1.

  "Continuous, near-surface observations of meteorological and turbulence parameters were performed at a location on the ice floe called *Met City* (Shupe et al., 2022). Measurements with an ultrasonic anemometer/thermometer were taken at a meteorological tower at 2 m, 6 m, and 10 m heights serving as surface reference for our balloon observations. Furthermore, the upward (↑) and downward (↓), broadband terrestrial irradiances *F* (in units of W m−2) were collected at the Atmospheric Surface Flux Station with a Precision Infrared Radiometer (Shupe et al., 2022). The net irradiances $F_{net} = F^{\downarrow} - F^{\uparrow}$ are taken as a proxy for the radiative energy budget at the surface. Furthermore, the measurements at *Met City* include the surface radiometric skin temperature." Lines 124-130

- **Line 116:** *You cite Illingworth et al. (2007), but provide no mention of what remote sensing instrumentation (lidar? radar?) or product from Cloudnet is used. In addition, you say you use Cloudnet but in the following sentence say that it poses challenges for low cloud layers. Is the cumulative surface radiative flux the only complementary measurement used? Regarding this, further down on line 120, you again mention radiation measurements but give no information on the instrument. I think some more information needs to be provided here.*
  Response: We agree that the information was not sufficient. Low-level clouds, as occurring frequently in the Arctic, are a known challenge for remote sensing instrumentation and therefore special

algorithms are applied. However, as we are interested in the effect of clouds on the surface radiative energy budget, we use the surface radiative flux measurements as a main cloud proxy. Additionally, we compare this to 360° all-sky photographs. Eventually, we decided to not use remote sensing instrumentation to characterize the cloud situation and, therefore, removed the Cloudnet information from the paper. Regarding the radiation measurements, please see the comment above.

"The cloudy Arctic ABL can be very shallow, especially in winter, and hence poses challenges on remote sensing approaches to detect the very low cloud layers (Griesche et al., 2020). For the question of the influence of clouds on the ABL dynamics, however, the net irradiances are of main importance, which are directly influenced by the clouds. Therefore, the terrestrial net radiation F_{net} measurements at Met City, where F_{net} is the cumulative surface irradiance, are used as an independent "cloud indicator". Cloudless conditions prevail when the net radiation is below -25 W m^{-2}, while higher values are associated with clouds (Wendisch et al., 2023). To avoid ambiguous allocations, data with net radiation in the range between -28 W m^{-2} and -22 W m^{-2} are not considered. Additionally, the cloud condition was manually compared with 360° photographs and all sky total imager observations (as far as possible regarding daylight conditions)." Lines 140-147

- **Line 142:** *You have not defined what "r" is in the equation for Taylor's hypothesis. Please do so and also provide a reference for Taylor's hypothesis, with perhaps a brief mention of what the hypothesis includes and why it is valid.*
  Response: We now explained "r" as the 'spatial increment' and added information about Taylor's hypothesis, which allows to transfer time to length scales.

  "The averaging in Eq. (1) denoted by the angle brackets is performed over all *t*, so the structure function $S^{(2)}$ is a function of time lag τ which has been calculated by applying Taylor's frozen turbulence hypothesis (transferring time lag to spatial increment $r = U \cdot \tau$) with the mean flow velocity *U* (see Stull (1988) for more details)." Lines 162-165

- **Line 176:** *You have not defined what the distinct layers of "δz" are in this context. Please do so.*
  Response: We have added "δz = 30 m".

  "Compared to other bulk *Ri*-definitions, where mean gradients are estimated for distinct layers of δz = 30 m (Jozef et al., 2022), for example, the Ri_b-approach applied here fails if multiple turbulent layers are present." Line 202f

- **Figure 3 caption and panel c:** *In the last sentence when you say that the SML heights derived by in situ turbulence are indicated by triangles, are you referring to the two left-pointing triangles on the right of the plot? Perhaps describe this a bit more clearly.*
  Response: Yes, that was the intention, and we hope the explanation we have given is now sufficient.

  "[...] The SML heights for cloudless (orange) and cloudy conditions (blue) derived by in situ turbulence records are indicated by left pointing triangles on the right in (c)."

- **Line 214:** *The first sentence here seems out of place, as you start discussing the bulk Richardson number for the cloudless profile and then move immediately toward discussing the cloudy profile. Suggest moving the first sentence to the discussion in the previous paragraph and starting this*

paragraph with the discussion of only the cloud profile.
Response: Thank you for your suggestion, we have implemented it directly. Lines 243f

– **Fig. 9:** *You have an inlaid panel, I assume to emphasize structure during a small part of the observing period, but this panel is not discussed anywhere in the text nor in the caption of the figure. I suggest either removing the inlay or specifically addressing it in the text.*
Response: That is true and we recognize that we did not provide any explanation of the panel in the text. Originally, we have shown the cloud case of Fig. 3 in the inlaid panel but now decided to remove the inlay and updated Fig. 9.

– **Line 304:** *It's rather hard to eyeball the specific 29 December event to which you are referring here. In addition, you reference profile #s 13 to 16, but this doesn't really mean anything since the x-axis of Fig. 11 is date and not profile #. Perhaps use a shaded bounding region in Fig. 11 to indicate this period that you explicitly discuss in the text.*
Response: We fully agree with you and have revised the Fig. 11 as discussed above.

– **Line 308:** *"…if not vertical profiles are available." Vertical profiles of what, exactly? Cloudy profiles that may otherwise be detected by remote sensing instrumentation?*
Response: Indeed, we have missed an explanation. Thank you for your remark, we have added information on the measured parameters.

"Therefore, with rapidly changing cloud cover, SML height determinations using MO theory are only partially successful. F_{net} observations can at least indicate these possible problems if no vertical profiles of thermodynamic or turbulent parameters are available." Lines 346-348

– **Line 309:** *Again, referring to this specific time period is kind of hard for the reader. Suggest using a shaded bounding region to indicate this period to make the text discussion easier.*
Response: We agree and followed your suggestions as described above.

<u>Technical comments</u>

We highly appreciate all your technical comments and have incorporated all suggestions you made here. In the following, we refer to the Lines in the revised manuscript, where this

– **Line 23:** *"extend" should be "extent"*
Response: Line 25
– **Line 43:** *there should not be a comma between "clouds" and "impact"*
Response: Line 49
– **Line 69:** *"turbulence based" should be "turbulence-based"*
Response: Line 81
– **Line 70:** *Not sure what you mean by "evaluate the bulk Richardson number approach"; is the turbulence-based SML height used to evaluate the bulk Richardson number approach?*
Response: Yes, this is exactly what we wanted to say here. We have changed this sentence and hope that this becomes clear now. Please see Lines 81 f.

"We use the turbulence-based SML heights as reference to derive a critical bulk Richardson number for the winter and spring. The in situ turbulence perspective not only allows to derive a critical value but also to evaluate the bulk Richardson number approach."

- **Line 70:** *"surface based" should be "surface-based"*
  Response: Line 84
- **Line 73:** *comma splice–use conjunction after comma*
  Response: We have rephrased the paragraph and removed the comma splice. Lines 88-92
- **Line 82:** *comma splice–use conjunction after comma*
  Response: We added a conjunction after comma. Line 98
- **Line 86:** *"The setup enables continuously vertical profiling" should be "The setup enables continuous vertical profiling"*
  Response: Line 102
- **Line 94:** *I'm not sure what is meant by "an one component"; do you mean "a single component"?*
  Response: Yes, we have meant a single component but have removed this information to prevent misunderstanding. Line 113
- **Line 100:** *"a few hundreds meters" should be "a few hundred meters" or "a few hundreds of meters"*
  Response: Line 119
- **Lines 104-105:** *"taken at a meteorological tower in 2 m…height" should be something like "taken at a meteorological tower at 2m…heights"*
  Response: Line 125f
- **Line 128:** *I don't think there needs to be a comma after "then"*
  Response: We have rephrased this paragraph. Please see Lines 149-153.
- **Line 135:** *"…as turbulence measure" should be "…as the turbulence measure"*
  Response: We have rephrased this sentence:
  "For our application, we chose the energy dissipation rate ε to quantify turbulence […]" Lines 156f
- **Line 141:** *Do you mean angle brackets?*
  Response: Yes, we do, thank you. Line 162f
- **Line 161:** *Comma splice; please include a conjunction after the comma*
  Response: We have added a conjunction. Line 185
- **Line 179:** *"…majority of the SML heights lies…" should be "…majority of the SML heights lie…"*
  Response: Line 205f
- **Line 180:** *"multi" should be "multiple"*
  Response: Line 206
- **Line 202:** *Probably not a need for the word "two" before "cloudless"*
  Response: We have removed it. Line 233
- **Lines 258-259:** *Here and in many other places in the manuscript, you are using "respectively" incorrectly, which requires one to list two features and two names attached to those features. For example, the correct use here would be something like the following: "Figure 7b shows the distributions of Ribc for cloudless and cloudy conditions in orange and blue lines, respectively."*
  Response: Thanks for pointing this out, we have revised the manuscript and removed the wrongly used "respectively".
- **Line 266:** *No need for a comma in this sentence.*
  Response: Line 300
- **Line 300:** *Comma splice. Use a conjunction.*
  Response: We have rephrased the sentence. Line 335

Response to Anonymous Reviewer #2

We thank your for the valuable comments and appreciate the time and effort you dedicated to your feedback. These comments contributed to a significant improvement of our manuscript.

During the review process, we noticed an inaccuracy in the calibration of the hot-wire sensors that warranted a recalibration of the data. The changes are minimal and do not affect the conclusions of this manuscript, but all analysis were performed again with the re-calibrated data. The new data set is currently being in the publication process of PANGAEA and accepted as a data descriptor publication in Nature Scientific Data.

Here is a point-by-point response to the reviewers' comments. We included the original comment in *blue and italics* with the lines referring to the first manuscript. Our response follows in black, and the line numbers refer to the revised version.

General comments

This study examines the characteristics of the Arctic atmospheric boundary layer (ABL), more specifically the surface mixed layer (SML), during the Multidisciplinary drifting Observatory for the Study of Arctic Climate (MOSAiC) campaign, using measurements from a tethered balloon platform including in-situ turbulence measurements. The measured profiles of the turbulence dissipation rate are used to diagnose the depth of the SML. This diagnosis serves as a reference for the evaluation of the bulk Richardson method and the surface flux-based method. I think the most significant result of the study is the determination of a critical bulk Richardson number for the diagnosis of the height of the SML. Furthermore, two typical states of the Arctic ABL were observed and characterized: cloudless situations with a stable and shallow ABL, and cloudy conditions with a mixed ABL. The paper is nicely organized and mostly well written. The presentation is mostly clear. While I think the paper may be acceptable with only minor revisions following the comments below, I think it could be significantly strengthened following the suggestions of the other reviewer.

We thank you for reviewing our paper and appreciate your feedback and comments that improved the quality of our study. We have revised the manuscript considering the comments of both Anonymous referees with the hope that we have addressed all minor and major concerns. Please find the responses to your comments below.

Specific comments

- **L23-L24:** *"... plays an important role as stable stratification hampers the vertical exchange and leads to a near surface warming contribution to Arctic amplification...". This is not clear to me. Wouldn't more stable conditions and reduce vertical exchange lead to surface cooling?*

  Response: This is somewhat misleading, the lapse rate effect describes the additional warming due to the fact that additional heat added to the system (by global warming) is distributed in a smaller volume (due to shallow SMLs) and therefore contributes to surface warming. This mechanism is described in detail by the cited work of Bintanja et al., 2011.

– **L172-L173:** *"For surface temperature, we use observations at the 2 m height". Would this not lead to a significant underestimation of the true surface inversion strength? Please comment on the differences.*
Response: This is a very good point and indeed there is a strengthening of the inversion when the radiative skin temperature is considered. During MOSAiC the air temperature at 2 m height and the skin temperature were measured. We have added a second bulk Richardson number calculated with the skin temperature as surface reference to account for the inversion down to the surface.

"According to the basic concept of the surface bulk Richardson number approach, the lower reference level is the surface where the mean horizontal wind velocity equals zero U ($z = z_0$) and θ_0 is the skin temperature. Since other studies are often based on radiosonde observations only, the temperature at 2 m is then typically used for θ_0, even though this value may differ from the skin temperature, especially for stably stratified conditions with strong surface temperature gradients.
To be consistent with other studies, we first use temperature observations at 2 m from the *Met City* tower for θ_0. Then, for comparison, the same analysis is performed using the skin temperature measured during MOSAiC." Lines 196-201

We derived a critical value for both bulk Richardson numbers. The critical values are 0.12 for the reference temperature at 2 m and 0.16 when we use the skin temperature. Either temperature can be used depending on the purpose of the analysis. As skin temperature is not always measured, we concentrated on the standard meteorological parameters that are most likely to be measured during field campaigns. This approach should ensure applicability to other campaigns.

"Applying the same analysis with the skin temperature as θ_0, we derive $Ri_{bc} = 0.16$ (Fig. 6). This difference shows clearly how strongly the derived value for Ri_{bc} depends on the selected surface reference. Furthermore, we can assume that the temperature measured at a height of 2 m on a mast is also significantly more accurate than a corresponding measurement with a radiosonde, which has comparatively large inaccuracies in the lower ranges." Lines 284-287

– **L205:** *"with a less well mixed, neutrally stratified layer below the inversion" What do you mean? To me the sub-cloud layer looks like a rather well-mixed turbulent layer..*
Response: We fully agree and have change the sentence accordingly.

"The first case represents a cloudless ABL with a pronounced surface inversion, and the second case describes a cloudy ABL and the resulting elevated inversion at the cloud top with a well-mixed layer below the inversion." Line 234-236

– **L208:** *"a slightly stably stratified layer". Seems still rather stable to me. Do you mean a "less stably stratified layer"?*
Response: Yes, thank you, we have edited the sentence.

"Cloudless conditions prevailed during a profile observed on 5 March 2020 with a strong surface-based temperature inversion ($\Delta\theta \approx 7$ K within about 40 m) up to 50 m followed by a less stably stratified layer above (Fig. 3a)." Lines 238-239

– **L209-L210:** *"Near the surface, the wind speed increases with height, peaking at about 50 m again and continuing almost constantly until the maximum height of the profile." I find this sentence unclear, please reformulate..*

Response: We agree and rephrased this sentence.
"The wind velocity *U* increases with height from the surface up to a height of about 50 m, and then remains almost constant up to the maximum height of the profile (Fig. 3b)." Lines 239-241

- **L261:** *"the differences for the two mean case depending…" What do you mean? Please clarify.*
  Response: We referred to the two cases of cloudy and cloudless conditions but we recognize that this was not clear in that sentence. We have rephrased it.
  "While we have derived Ri_{bc} for cloudless and cloudy conditions, the differences for the two typical ABL types are negligible." Lines 293-294

- **Fig. 3:** *The near surface conditions (with strong gradients) are difficult to see. You might want to increase the size of the panels.*
  Response: Yes, that was rather hard to see in the first figure. We have enlarged the size slightly. We hope that this improves readability.

Technical comments

Thank you for your feedback on the language. We have incorporated all the suggested changes and provide the Line numbers below.

- **L23:** *vertical extend --> vertical extent*
  Response: Line 25
- **L203:** *alternates between two cloudless and cloudy --> alternatives between cloudless and cloudy*
  Response: Line 233
- **L214:** *almost linearly increases --> increases almost linearly*
  Response: Line 244
- **L214:** *The new paragraph should start with "The example for vertical stratification under cloudy conditions…"*
  Response: Yes, we agree and have changed this accordingly. Please see Lines 243f.
- **L221:** *decrease almost abruptly --> decrease quite abruptly*
  Response: Line 251
- **L231:** *As it is often --> As is often*
  Response: Line 261
- **L232:** *already begins clearly inside --> already begins inside*
  Response: Line 262
- **L237:** *near surface --> near the surface*
  Response: Line 267
- **L238:** *low values --> lower values*
  Response: Line 268
- **L241:** *turbulence continuously reaches --> turbulence reaches*
  Response: Line 271
- **L280:** *SML height smaller than 150 m --> SML height less than 150 m*
  Response: Line 314
- **L298:** *at the low humidities --> at low humidity values*
  Response: Line 333
- **Fig. 11:** *at surface --> at the surface*
  Response: We have changed this, please see caption of Fig. 11.