Response to Reviewers – "Boundary-layer height and surface stability at SMEAR II, Hyytiälä, Finland in ERA5 and observations"

Victoria Sinclair, Jenna Ritvanen, Gabin Urbancic, Irina Statnaia, Yurii Batrak, Dmitri Moisseev and Mona Kurppa

February 2022

We thank the reviewer for their constructive comments on our submitted manuscript. We have copied the comments of the reviewers in black here and include our response to each individual comment in blue.

Reviewer 1

Main comments:

1. The parcel method is a commonly and widely used method in the CBL determination. It is applied by the automatic BL detection from the MWR but not on RS. A comparison of this method with the already applied on RS would have been worth. Similarly, the Heffler and part of LL10 methods applied on RS could also be used for MWR.
First, we would like to note that the manuscript does not aim to compare BL heights obtained from the MWR to those obtained from radiosondes. Second, we would like to stress that our study does not aim to develop new methods to identify the BL height nor improve pre-existing methods. Third, we want to point out that our choice of methods to identify the BL height from the radiosondes was based solely on the options provided within the ARM Value-Added-Product (VAP), "Planetary Boundary Layer Height": https://www.arm.gov/capabilities/vaps/pblht. As we state in section 2.2, the radiosondes were made during an intense campaign when one of the ARM mobile facilities was in Hyytiälä. ARM routinely computes and makes available this VAP for all permanent sites and all campaigns. For example, this VAP is available from ARM's permanent sites since 2001 (Oklahoma), 2002 (Barrow), 2013 (Grasiosa) and from 2001 - 2014 (Tropical west Pacific). Therefore, to remain consistent with these other stations, and other ARM campaigns, we only used the methods already in this VAP. This is the main reason we do not apply the parcel method to the radiosonde data. We have added text to the beginning of section 3 to stress this more clearly.

Similarly, the Heffler and part of LL10 methods applied on RS could also be used for MWR.
The aim of this manuscript was to take pre-existing methods, that are used by others already, to determine the BL height. This is why we only use the manufacturer provided algorithm to

1

diagnose the BL height from the MWR. In theory, we could apply the Heffter method to the MWR data but, in our opinion, it does not make sense to apply a modified LL10 method to the MWR. This is because we would be introducing yet another method to identify the BL height which is not the aim of this paper. Furthermore, given that the radiosonde observations and the microwave radiometer observations do not overlap in time and that we cannot compare them, there is limited scientific value of attempting to apply the same methods to the MWR as are provided in the ARM VAP.

2. The MWR has data since June 2018 leading nowadays to a 3 years time series. A one-year time series is presented in Fig. 8. A climatology with these measurements (at least the year presented) would also be very valuable and allow a comparison with the ERA5 long-term time series. We have created figures of the monthly and diurnal climatology of the BL height diagnosed from the MWR based on the data period considered in the manuscript - September 2018 to August 2019 (Figure 1 in this response). This figure includes all of the MWR data. When the mean diurnal cycle is considered in the warm season, two peaks in BL height are evident; one around 11 UTC and a second one in the middle of the night - around 00 UTC. This strongly suggested to us that the MWR does not diagnose the BL correctly at night in the warm season. Based on the climatology of stability class from the eddy covariance measurements, we know that this is when very stable or weakly stable conditions are common. Hence we concluded that in some cases when the BL is stable the MWR does not identify a physically meaningful BL height - it over-estimates the BL height considerably potentially identifying a residual layer. This is confirmed by Figure 2 in this response which only includes data when the MWR has determined the BL to be convective. When the stable points are removed, the diurnal evolution of the BL height from the MWR is much more reasonable. Given this issue with the stable BLs, a fair comparison with the ERA5 climatology figures is not straightforward and therefore, we do not add these new figures to the manuscript.

In addition, we also investigated if our results would have changed if we included additional data by also creating the climatology for June 2018 - Dec 2021 (Figure 3). By comparing this figure to Figure 1, we see that there are only small differences. Therefore, we do not include any additional MWR data in the revised manuscript as this would require considerable work and would not change the main conclusions.

3. The radiosondes data are only available from February to mid-September. What would be the impact of the October-January (i.e. most of the fall and winter periods) on the comparison between the methods? On the validation of ERA5? If ERA5 cannot be considered as validated for Fall and Winter, should the comparison between ERA5 and MWR and the climatology be discussed differently ?

Firstly, we do compare ERA5 to the radiosondes BL height estimates for February which is the coldest month of the year in southern Finland and certainly is a winter month. Thus, we are of the opinion that it is somewhat unjustified to say that ERA5 is not validated for winter. Secondly, our results show that we have validated ERA5 BL heights for all classes of surface stability - it is not the case that we are completely missing certain types of the BLs in our 8 month validation period. This is shown in Figure 4 in this response which, for the 23 years of surface stability data from the eddy covariance measurements, shows that, climatologically,
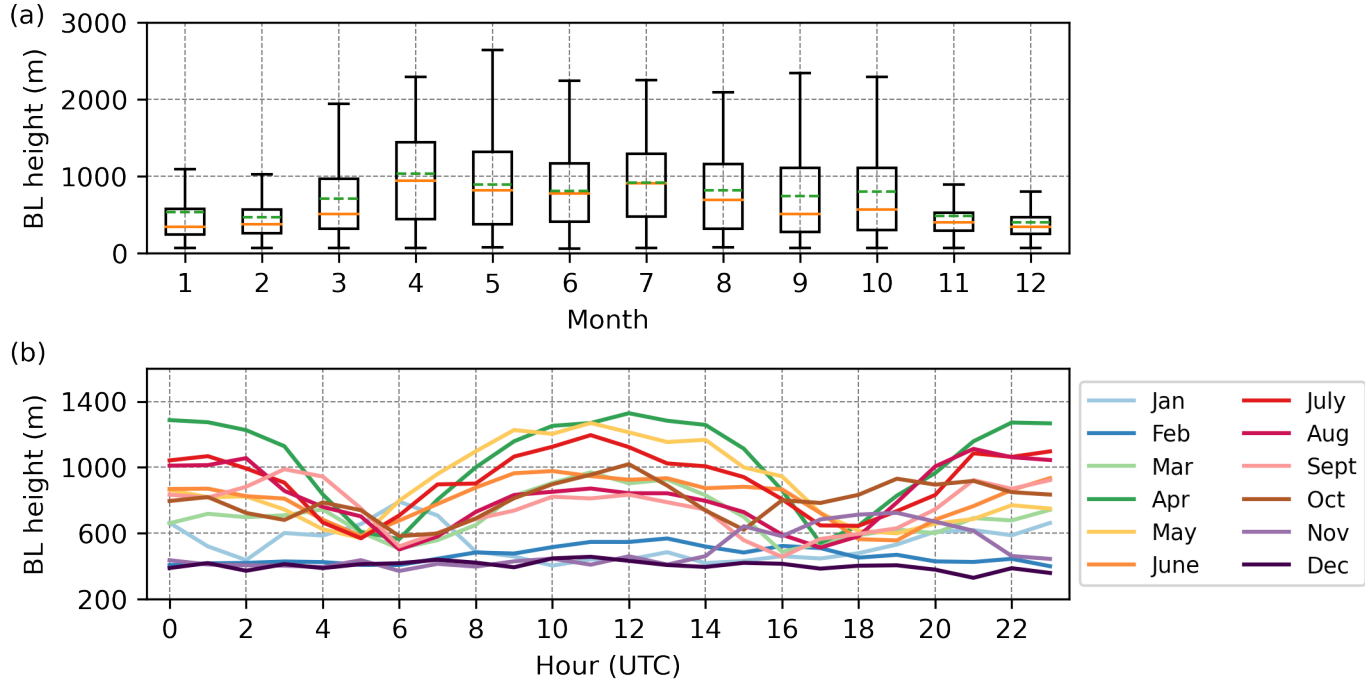
2

Figure 1: The (a) annual and (b) diurnal cycle of the boundary layer height from microwave radiometer measurements for the Sep 2018 - Aug 2019 period included in the manuscript. The figures contain both the convective and stable BLH MWR measurements.

it is only the weakly stable and near-neutral stable BLs that are more common in October to January (the time period we did not perform a comparison for) than in February to September. Furthermore, Figure 4 shows that even in the time period we do consider, these types of BLs are sampled enough i.e. the orange bars are not tiny. However, we have added some text to section 5.2 and to the discussion to reminder a reader that the comparison with ERA5 does not cover a full year.

4. Discussion of Fig. 5, 6 and 7:(§ 5.1, 5.2 and 5.3): the comparison between the methods as a function of time and of stability is very interesting. The description is fine and could be published as it. It could however be largely improved if the structure of these § would rely on the differences between methods. E.g. (1) The largest differences are found between Heffler and the other methods in very unstable cases due to the required temperature difference of 2K. Is it possible to improve this Heffler detection by lowering the required T difference in case of unstable situation? (2) ERA5 has the large discrepancy in case of very stable situations - is it due to the lower vertical resolution, due bias in T profile or in wind profiles? (3) ERA5 agree better with $Ri_{0.5}$ for very unstable cases - due to the lower vertical resolution (L457-463). (4) at 00 UTC or in case of stability, LL10 leads to higher MLH - inherent to the used method since LL10 estimate the top of the stable layer or the LLJ and the Ri a very shallow height that does not really correspond to a physical layer. (5) ERA5 << MWR in case of stability - MWR measures
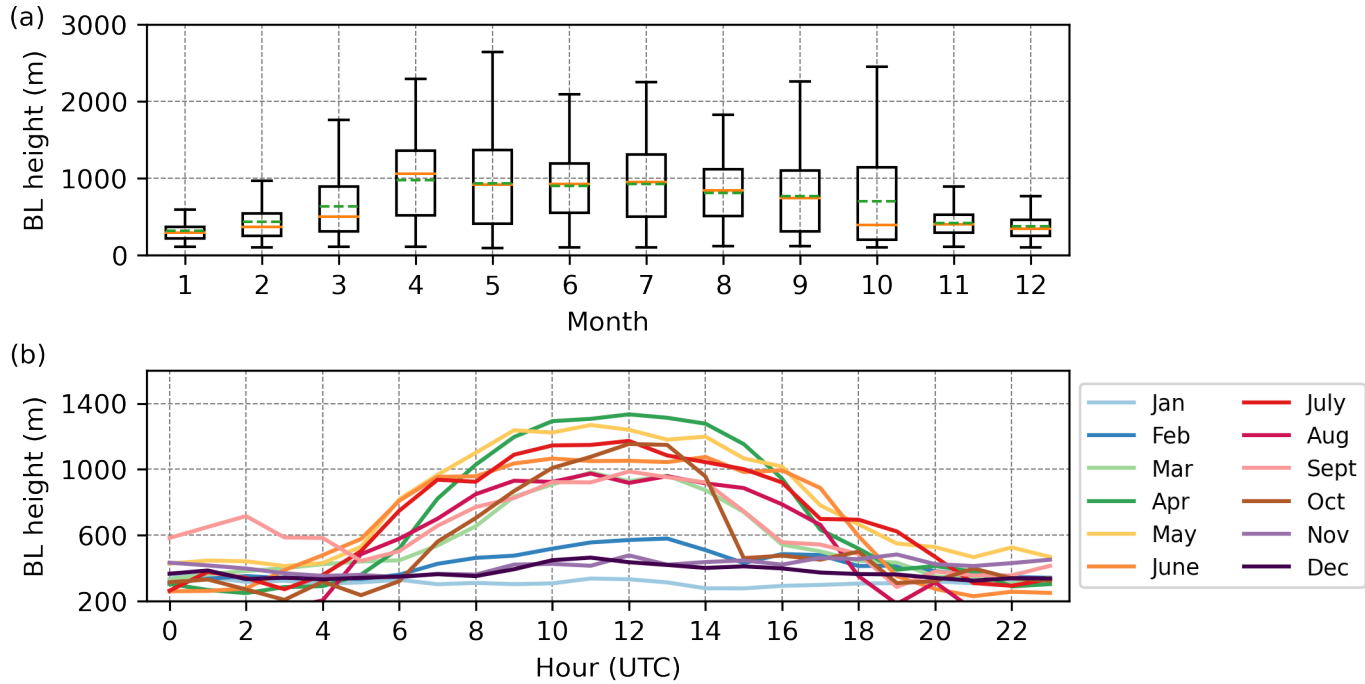
3

Figure 2: The (a) annual and (b) diurnal cycle of the convective boundary layer height from microwave radiometer measurements for the Sep 2018 - Aug 2019 period included in the manuscript.

the top of the stable layer and Ri another usually shallower layer. (6) ERA5 << MWR in case of stability - MWR measures the top of the stable layer and Ri another usually shallower layer. I am aware that a reorganization of the section represents a lot of work but I think that it would improve the manuscript. An alternative would be to enhance these points in the discussion.

Overall, these comments have been very helpful. Although we have not hugely re-structured section 5 we have made some major changes to parts of the manuscript. Firstly, we have added details to section 3 to discuss and highlight the differences between the methods we use to diagnose the BL height. Here we now also stress that the different methods do not all quantify exactly the same physical thing and thus even a priori we should not expect the different methods to agree. We also agree that it is important to be aware of the sensitivity of the diagnosed BL height to the various thresholds applied in the different methods. As such we now include more critical analysis of this in sections 3, 5 and 8 and more references to previous studies which have considered these sensitivities. Based on the specific comments here, we have also included more more physical interpretation to the detected differences in section 5.

Some specific comments relating to the specific points raised: (1) As we want to evaluate pre-existing methods included within the ARM VAP, it is not in the scope of this current study to further develop / alter the Heffter method. (2) To determine the cause of the ERA5 discrepancies in stable conditions would require considerable additional analysis which is not in the scope of the current study. It would also very challenging to separate the difference causes of error as
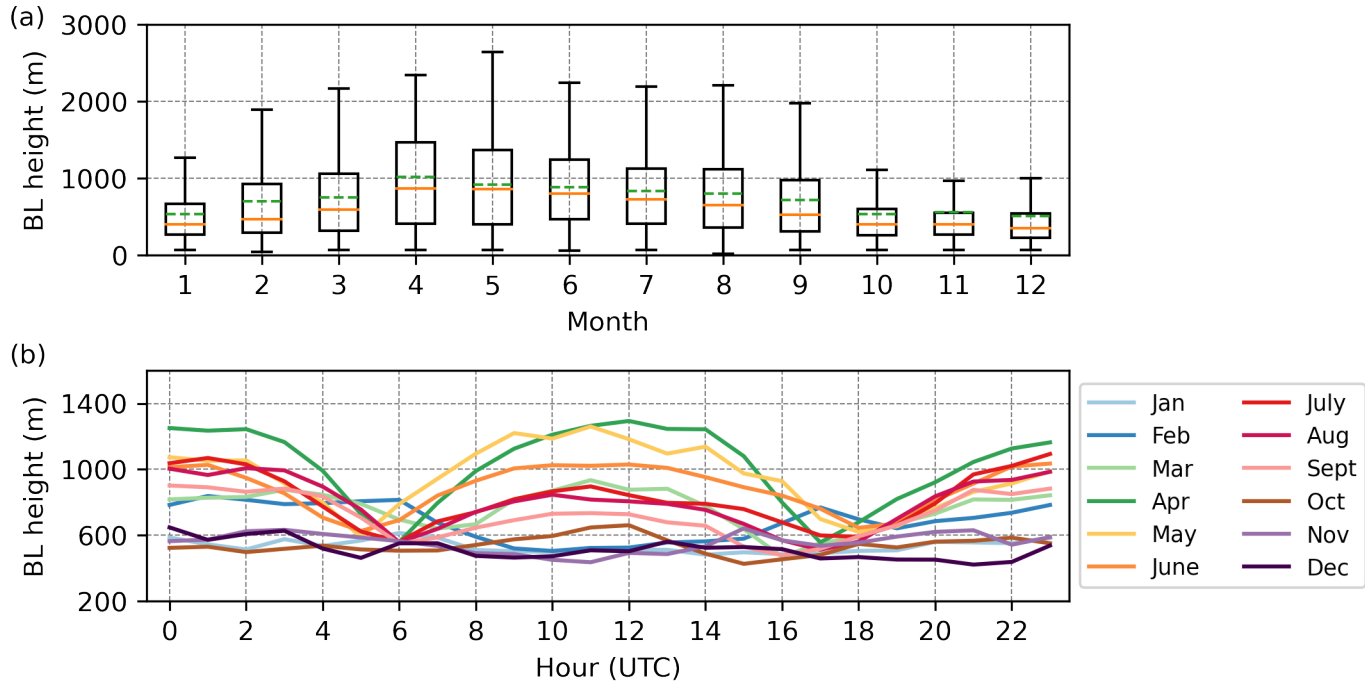
4

Figure 3: The (a) annual and (b) diurnal cycle of the boundary layer height from microwave radiometer measurements for June 2018 - Dec 2021 period. The figures contain both the convective and stable BLH MWR measurements.

the limited resolution in ERA5 may limit the accuracy of the temperature and wind but also any inadequacies in the boundary-layer parameterization scheme could also cause errors. (3) ERA5 may agree better with $Ri_{0.5}$ because of the resolution but also a warm surface temperature bias in ERA5 could cause this (We now discuss the potential impact of the difference in resolution between the radiosondes and ERA5 in more depth in the revised manuscript). (4, 5) We now attempt to highlight the differences between the methods, and the likely impacts of these differences on the BL height, more in section 3 and section 5. (6) For the unfiltered data, yes, ERA5 has much shallower BL heights than the MWR for very stable conditions. As stated in our response to major point 2 above, this discrepancy may be due to the MWR incorrectly identifying the height of some very stable BLs.

Minor comments

1. L 17: please mention that the climatology rely only on ERA5 time series. Revised to now state that the climatology is based on ERA5.

2. L 220: what is the initial vertical grid of the radio-sounding? What are the reasons for sub-sampling RS vertical profiles and the potential consequences on each PBLH detection method? The radiosondes measure data every two seconds. Therefore the vertical grid in terms of pressure is not uniform. On average, below 700 hPa, there are measurements every 1.2 hPa but this can
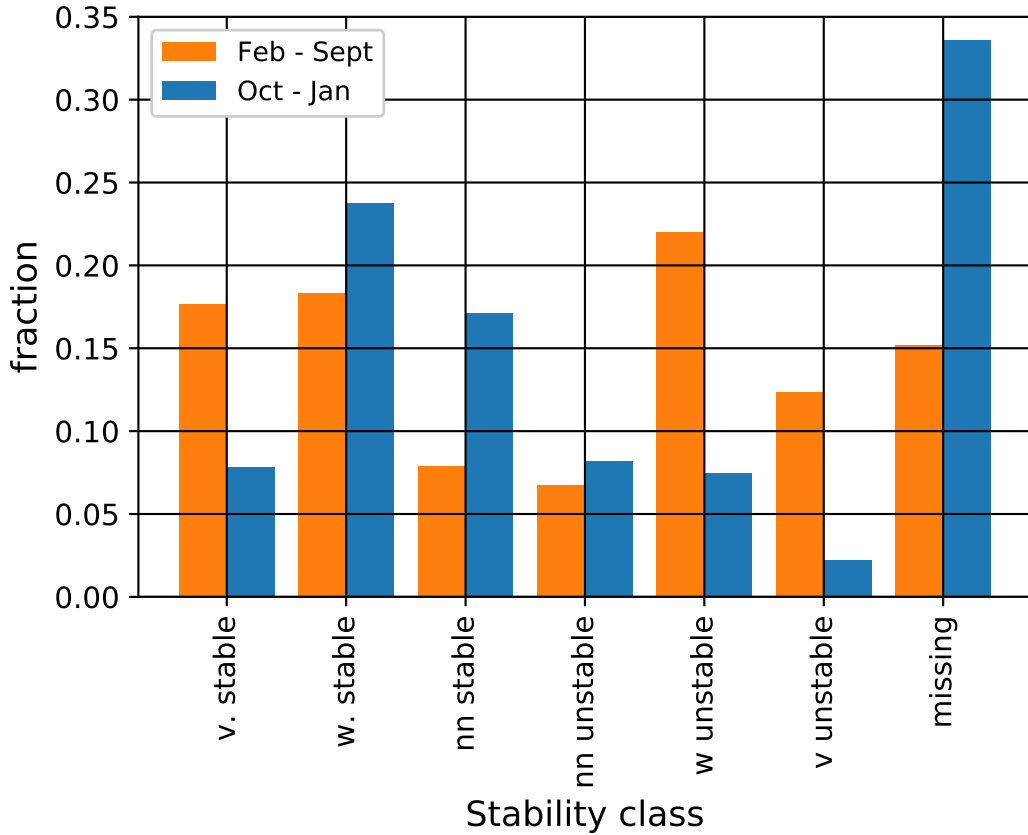
Figure 4: The fraction of times that different stability classes are detected for the months with no radiosondes (October to January) and for the months for which ERA5 was validated (February to September).

vary (standard deviation of 0.12 hPa) due to different ascent rates. This information has been added to section 2.2 where we previously stated the resolution as 10 - 15 m. Therefore, one reason for the sub-sampling is to put all of the radiosonde data on a vertical grid with the same resolution. A second reason is to smooth the profiles which makes computing gradients less noisy.

3. L 222-229 H80: Is the PBLH very sensitive to the chosen potential temperature difference of 2K and to the 15 hPa vertical resolution? In general increasing the resolution means the vertical potential temperature gradients are larger and thus the BL height can be diagnosed lower. Also if the 2 K threshold was decreased, lower BL heights would be diagnosed. However, as stated above, it is out of the scope of this study to revise the pre-existing methods that are used to diagnose BL height but it is important to be aware of the limitations, and the sensitivities, of all the methods used to identify the BL height. Therefore, we have attempted to add a more in-depth critical discussion of this in section 3 and in the discussion section. We have added additional reference to previous studies that have considered these sensitivities.

4. L 230-240: For convective BL, LL10 has a large similarity with the parcel method. The differences between both methods are 1) the parcel method take the potential temperature at ground (2 m) and LL10 at 150 m, 2) the parcel method use an instability threshold of 0 (and 0,5 for LL10) and 3) LL10 takes the first level with the potential temperature gradient > 4.0 K km^{-1} higher than the level with the potential temperature difference with the one at 150 m < 0.5. Differences 1) and 3) lead to lower and higher PBLH than the parcel method, respectively. Are the PBLH sensitive to the various thresholds? What is the difference between LL10 and the parcel method in case of unstable situations? This is important since the MWR use the parcel method. We do not estimate the BL height from the parcel method and thus we cannot quantitatively answer how it differs from the LL10 method. The BL heights are likely sensitive to the various thresholds and we now include more discussion about this in the revised manuscript and discuss previous studies which have analysed these sensitivities.

5. L249: -9999 corresponds to a missing value but it's real value is not important and differs as a function of the programming language. Yes, this is true. We have deleted this sentence as how a missing value is specified is not necessary information.

6. Figure 2: The RS profile corresponds to the 5 hPa or 15hPa vertical resolution? a plot of the bRi number could also help to understand the difference between the threshold and explain the failure of this method at 12 UTC. A smaller vertical extend of the y scale for the 00 UTC case could help to see the differences between the methods. The radiosonde data is the raw, high resolution data measured by the radiosonde every 2 seconds / approximately every 1.2 hPa. We have added this information to the caption. In addition we have also changed the scale on the y-axis for the top and bottom rows.

7. L 306: it would help to know what are the potential reasons leading to underestimation by the Ri methods. It has also to be mentioned that ERA5 (also a Richardson number-based method) is ok contrarily to Ri applied to RS. This case study has proved difficult to fully understand as we took the ARM provided values and did not compute them ourselves. We believe that the underestimation by the Ri methods maybe related to the choice of surface values in this case or is affected by the vertical potential temperature profile which has some weakly stable layers present (see Figure 5 in this response). We have added text to the manuscript about this. We have also added more information to section 3 regarding the comparison of ERA5 (Richardson number method with $Ri_c = 0.25$) BL heights to those from the radiosondes using the "same" method. This is not an exact, fair comparison as the vertical resolution and surface values differ. Therefore, it cannot be expected a priori that these two methods should agree perfectly.

8. L324-325: the subject of "is within the range..." is not clearly defined. This has now been revised to read "the BL height from ERA5 is within the range of values estimated from the radiosondes".

9. L348-349 and 353-355: the fact that the nocturnal (or stable) BL are different between MWR and ERA5 rely mostly on the applied method. ERA5 uses Ri that gives a MLH almost at ground in case of atmospheric stability; the layer given by Ri during the night is however physically/thermodynamically not well defined. MWR search for the vanishing potential temperature gradient corresponding to the top of the stable layer. Both methods cannot be directly compared
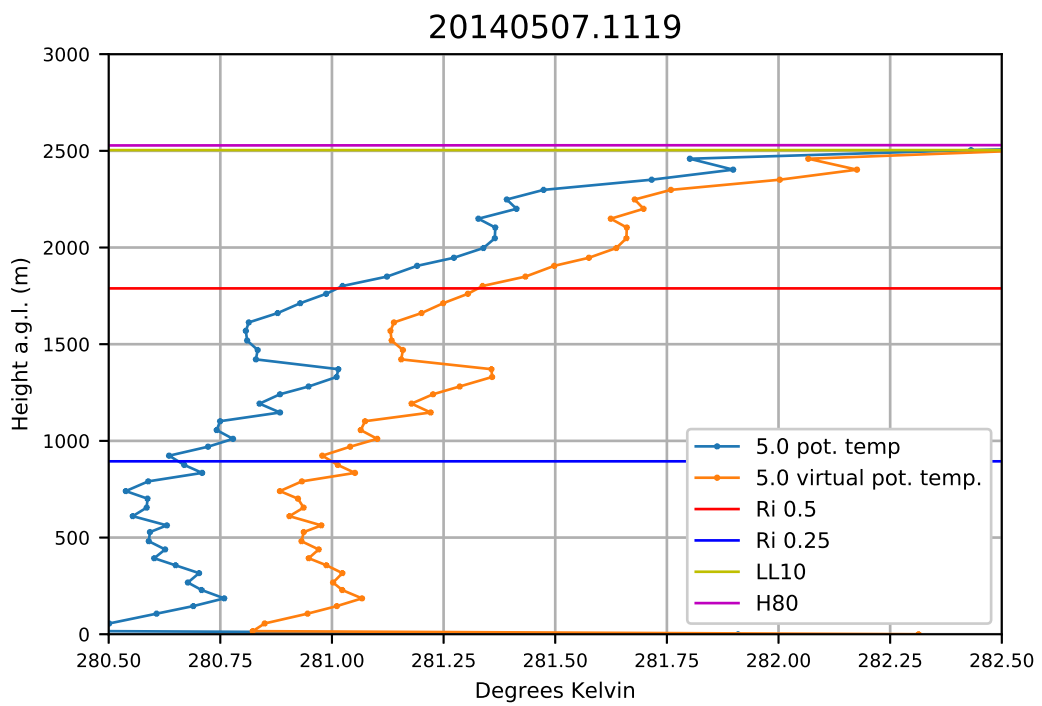
7

Figure 5: Vertical profiles of potential temperature (blue) and virtual potential temperature (orange) on the 5-hPa vertical grid at 12 UTC on 7th May 2014. Horizontal lines should the BL height diagnosed by the four different methods.

in such a case. There are two separate issues here. Firstly, as the reviewer correctly states, the method applied by the MWR to identify the BL height is very different and thus not directly comparable to the method that ERA5 uses. We have revised section 3 and the discussion to highlight the differences between the methods more and to more clearly state that we would not expect the methods to agree exactly. The second issue is similar to that discussed in our response to major point 2 above and highlighted by figures 1b and 3b. The MWR algorithm does appear to identify physically incorrect BL heights at night (when the BL is stable).

10. L377- 378: this is not obvious. The same data are used but the comparison is between BLH. This refers to the correlation coefficients between the different methods applied to the radiosonde data. Although the methods are all different and thus not expected to agree perfectly, we expected the correlations to be at least positive. However, we agree that it is not a given that the correlations would be statistically significant. We have therefore revised this sentence.

11. L391-393: grammatical problems. We have revised this sentence.

12. L396-399: Per definition, the parcel method always detect a lower BLH than the Ri. The cases LL10>Ri and LL10<Ri should then be discussed as a function of the differences (see comment L230-240) between LL10, the parcel method and the Ri one. We do not use the parcel method to identify the BL height from the radiosondes and therefore to discuss this would be speculative. However, as said above, we now include more critical discussion of the differences between the methods.

13. L 420-421: "The LL10 scheme and especially the H80 method produce the deepest BLs when the surface layer is very unstable." It seems then that in case of large unstability, the LL10 "over-shooting parcel" is responsible of the higher BLH than the Ri method. Is my assumption right? This is certainly one possibility. However, the LL10 scheme also searches for the level where the temperature exceeds the temperature at 150 m (T_{150m}) plus 0.5K. Therefore if T_{150m} + 0.5 is greater than the near surface temperature used by the Richardson number method, this would also result in deeper BLs in the LL10 method.

14. L427-428 + Fig 6:" When weakly stable and near-neutral stable conditions are considered, the LL10 scheme has the shallowest BLs and the narrowest distributions": From which stability threshold the LL10 method apply the neutral and the stable detection method? This info can help interpreting Fig. 6. As stated in the caption of Figure 6, the stability classes plotted are from the eddy covariance. In the LL10 scheme, stability is determined based on the potential temperature difference $\theta_5 - \theta_2$ (as described in section 3): if $\theta_5 - \theta_2$ is less than -1K a convective BL is present, if it is greater than +1K a stable BL is present and if it is between -1K and +1K a neutral BL is present. We have also now compared the LL10 stability classes to the EC stability classes in section 3 and have added the related figure to the supporting information.

15. L504-510: For high stability the Ri (ERA5) leads per definition higher MLH than the parcel method (MWR) à this explain the results (see main comment 4) We are not sure we fully understand this comment as these lines refer to where we discuss unstable situations: *"for very unstable BLs the MWR diagnoses deeper BLs than ERA5 but for less unstable BLs the MWR*

9

has shallower BLs than ERA5." We do agree that the Ri method with a critical value greater than zero should give deeper BLs than the parcel method though.

16. L518-519: "These statistical results from the very stable case support our finding from the case study (Section 4.2) that the height of the BL diagnosed by the MWR under stable conditions can be significantly overestimated": I do not agree with this formulation. The method applied to MWR in stable cases is just different and try to measure another sublayer than ERA5. This sublayer corresponds also not to aerosol layers measured by ceilometers. Please see the response to major point 2 above and also Figures 1b and 3b in this response. We do believe that in some stable cases the MWR does diagnose unrealistically deep BLs and thus does over-estimate the BL height and that the differences between the MWR BL heights and the ERA5 values are not solely due to the difference in methods.

17. L542-544: "Note that the stable BL height values from the MWR have been filtered out, so most of the MWR BL heights during summer nights are not seen in Figure 8." Why ? I expect (due to the applied methods) that the removed MWR BL were higher than the ERA5. Is this right ? See the response to major point 2 above.

18. L563-564: do you have a tentative explanation for these kind of cases ? One hypothesis is that these weakly stable BLs are dominated by shear-driven turbulence and not buoyancy-driven turbulence. As the MWR only considers the potential temperature profile, it will not be able to estimate accuratly the BL depth in such conditions. ERA5 which uses a Richardson number method does account for shear driven turbulence and this could explain the deeper BLs in ERA5 in this case.

19. L566-572: it is however very important here to consider (and write a reminder for the reader) that most of the fall and winter periods (15 september-January) were not taken into account in the comparison between RS and ERA5. ERA5 was then considered as a good BLH retrieval method for spring and summer, when the atmospheric stability is very different than during fall and winter. MWR remains a measuring system and ERA5 a reanalysis, so that MWR results cannot be discarded without any clear reason. In fact, you wrote in the discussion (L728-729) "Thus, ERA5 still cannot capture the depth of very stable BLs accurately, which is likely due to deficiencies in the BL parameterization of lack of resolution". It is then very important to take this conclusion into account when describing and discussing the climatology. The comparison covers 1st Feb - 15 September 2014 and as we have written in response to major point 3 above, we do believe that we have verified ERA5 for all stability classes. As these lines refer to the 1-year (September 2018 - August 2019) comparison between ERA5 and the MWR we do not add information here concerning the time period of the radiosonde / ERA5 comparison as we feel it would be confusing for a reader. We do add elsewhere reminders that ERA5 was only compared against the radiosondes for 1st Feb - 13 Sept.

20. L574: Fig. 8 shows a complete year of MWR data. Why not at least provide the seasonal cycle of this first year of measurement? We have now created a new figure (Figure 2 in this response) which shows the average seasonal and diurnal cycle of convective BL heights derived from the MWR. For reasons discussed in our response to major point 2 above we excluded the stable cases as including them leads to clearly unphysical BLs during summertime nights (Figures 1b

and 3b in this response). Since we only include the convective BLs, this figure is not directly comparable with those from ERA5 and this is why we decided not to include it in the revised manuscript.

21. L575 as explained in a previous comment, ERA5 was validated only for spring and summer. In fall and winter (as seen in the climatology) weakly stable and near-neutral stable cases are much more frequent. In these cases, Ri can be largely influenced by the wind component (see results L584-586). This should be discussed anywhere. ERA5 was validated for February which is the coldest month in southern Finland. Figure 4 in this response does show that near neutral stable and weakly stable cases do occur more in the months not considered (October - January) than those we did considered. We have added some text based on previous studies about how Richardson number method behaves in weakly stable cases.

22. L638-645: same comment as for L575: the night seasonal cycle of BLH should also be discussed as a function of the wind compound in Ri method. We have analysed the wind climate using measurements made at 67.2 m and these show, as expected, that stronger winds occur in winter than in summer and that calm conditions are more common in summer than in autumn / early winter. While we appreciate that low-level wind speed is not the same as low-level wind shear, we do add some details concerning the seasonal wind variation to section 6.

23. L660-661: this conclusion is right only if Ri method is the right method to resolve BL height. We believe this conclusion is valid since there is good agreement between ERA5 and the Richardson number methods applied to the radiosondes. This at least means ERA5 represents the vertical structure of potential temperature and winds in a relatively ok manner.

24. L752-754: "A key outcome of our analysis is that the MWR does not reliably estimate the BL height under stable conditions, which at Hyytiälä, occur commonly at night between April and September." Once again, ERA5 and MWR just apply different methods estimating different sublayers in case of stable conditions. While we agree that ERA5 and the MWR use different methods, we still believe that our statement is valid (see response to major point number 2). We have revised the text in section 5.3 to make it clearer what the problem is with stable BLs and why we have removed them from Figure 8.

25. L760: once again, ERA5 was validated mostly in spring and summer. This should also be taken into account in the discussion. We have added text to the discussion to reminder a reader that we do not consider a full year, but we do sample all types of BL stability.

26. Fig 4: would it be possible to color dots of c), e) and f) with EC stability ? (or all plots with EC stability) We have made this figure and include it in the supporting material and as Figure 6 in this response. We did not not include this new figure in the manuscript for two reasons (1) we want to highlight how the stability from the LL10 scheme affects the diagnosed BL heights from the LL10 method and how they compare to BL heights from other methods and (2) some of the information about how the BL height from different methods relate to each other is already presented in Figure 6 of the manuscript, but in our opinion, in an easier to interpret manner. We also feel that including stability from the EC data on some panels and stability from the LL10 method on other panels would be confusing.
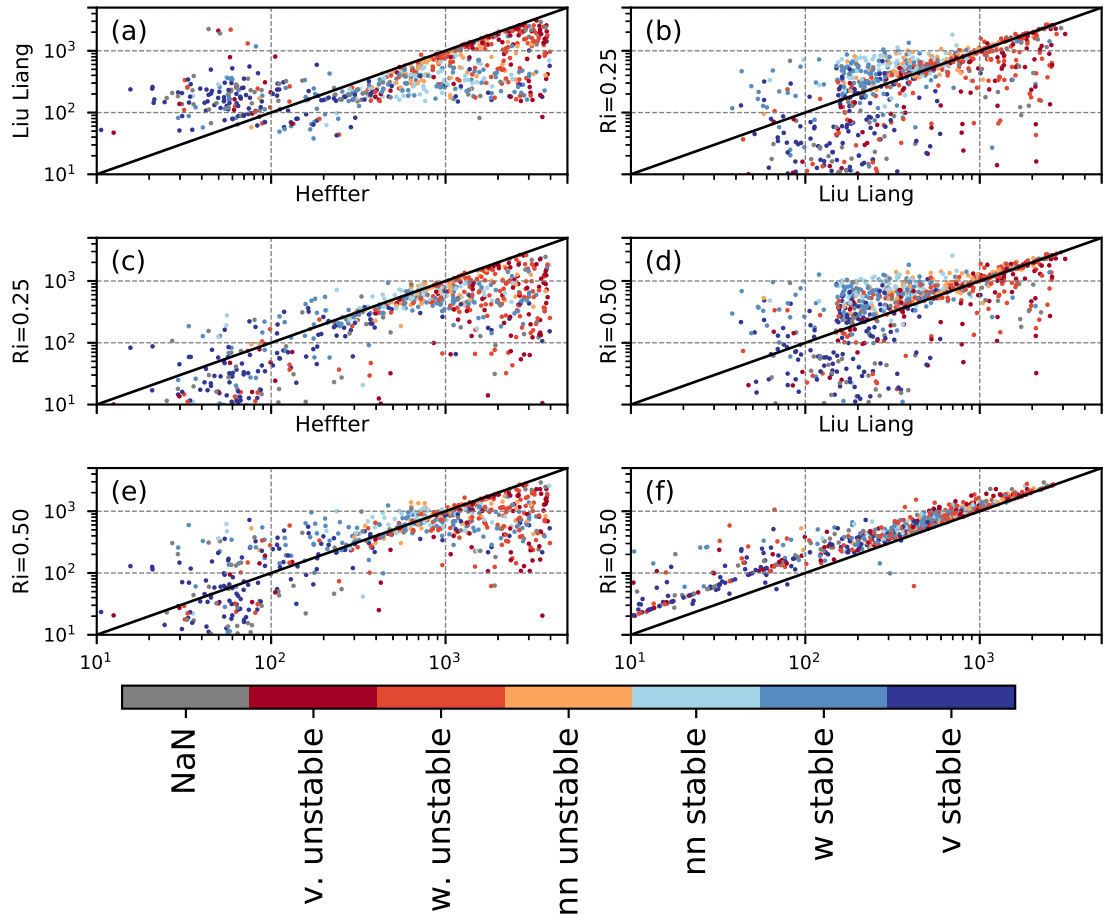
Figure 6: Scatter plots showing the relation between the BL height diagnosed from radiosondes taken during BAECC using 4 different methods. Data is from 1st February 2014 - 13th September 2014. Colours indicate the stability class from the EC measurements. Black solid lines shows the 1-to-1 line. Note the logarithmic scale.

27. Fig 7: since the MWR time series appears thereafter, I would indicate in the legend the period of comparison. The data included in Figure 7 is from the start of September 2018 to the end August 2019. We have added this information to the caption and also to section 2.3.

28. Fig 9: please also indicate the used period of time. We have now added the time periods to the caption.

29. Fig 11: mention that a) and b) correspond to ERA5 BL heights We have revised the caption to state that the BL heights are from ERA5.

30. Fig 12 (and similar figure in the supplement): it's not clear if all the plotted correlations (r>0.75) are statistically significant. Are correlations with r>0.75 sometimes also statistically significant ? All the values that are shown in Figure 12 are statistically significant and we have modified the caption to state this. This is likely due to the very large sample size (40 years). We assume the second question here was meant to be "are correlations with r<0.75 sometimes also statistically significant?". Yes, this does occur. However, in response to reviewer 2, we have modified Figure 12 to quantify how often the BL height at each grid point is within 150 m of the BL height at Hyytiälä.

31. Bigger font size in Figure S1 would be nice.We have increased the font size in this figure.