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

April 2022

This study presents a very valuable comparison of several BLH detection methods as a function of atmospheric stability computed by several methods. This analysis and the climatology are the most interesting parts of the study and the discussion has been largely improved in the reviewed version of the manuscript.

I still have a main comment about the choice of the vocabulary used for some conclusions. The authors choose to strictly apply the BL height detection methods provided at ARM (e.g. even if several publications mentioned that a better threshold should be applied for the H80 method) and, as mentioned several time in their answers, they did not compute the values by themselves. A major consequence of this choice is that BLH estimated by Ri is taken as the true boundary layer height. The fact that the sole expression "boundary layer" is used regardless of method and time is characteristic of a simplified vision of a basic boundary layer with ONE characteristic height. It is however obvious that the authors have a large knowledge on the topic and that the words "fail" (lines 652, 661, 673, 907, 908) and "true" (lines 622, 842, 903) just tend to emphasize the results by dichotomous considerations. Surprisingly, the word "truth" is used three times to warn the readers "none of the BL height considered here are clearly the truth" (line 444). The Ri method allows a good BL characterization in most cases but it remains limited for some atmospheric conditions. I ask therefore the authors to really discuss the complexity of the BL height determination in the comparison between MWR and ERA5 results by nuancing the notions of "true" and "false".

We thank the reviewer for taking the time to provide these comments. We reply to the minor points below in blue. As a general response, we did not mean to give the impression that the BL height estimated by Ri is true boundary layer height. We are of the option that all methods applied to the radiosondes provide valid estimates of the broadest definition of the BL - the part of the atmosphere influenced by the surface. However, we do acknowledge that some methods are designed to identify only certain types of BLs that could be called a mixed layer or mixing layer.

Regarding the MWR diagnosed BL heights, we have carefully revised the text in this section to stress we are comparing ERA5 and the MWR and it is unclear which is correct. We have revised the manuscript to remove words such as fail and true. The exception is the text describing when the MWR retrieval really does "fail" which is occasionally the case due to for technical reasons.

Minor comments

1. Line 3: the use of the two "and" leads to confusion about the role of " the level of agreement". This sentence has been split into two sentences to make it clearer.

2. Line 230-231: The problem is not of "no truth", but of the complexity of BL, the definition of BL height and the fact that BL height is more a layer than a precise altitude. In that sense the use of the word "subjective" is completely inappropriate since your study relies to measured/modeled data. These sentences have now been revised. We delete "subjective" and now refer to the complexity of the BL.

3. Lines 288-289: What are the bias/problems bounded to the altitude differences used to define stability classes? Which method is the most trustful one? The stability determined from the eddy covariance system reflects the stability of the surface layer whereas from the sounding, the stability reflects more that of the lower boundary layer. It is hard / impossible to know which is most trustworthy. However, we hypothesis that the stability from the sounding may be more representative of the large-scale environment as it is not directly above the forest canopy. However, on the other hand, the radiosonde derived stability estimate is from one instantaneous measurement, whereas the stability from the eddy covariance system is a 30-minute average of very high resolution observations. Some of the discrepancies in Figure S1 may be due to the differences in altitude used / the two diagnostics measuring slightly different things but as we discuss in the conclusions, we believe that the stability threshold used by the LL10 scheme is not appropriate for this location.

4. Lines 338-339: please give the "surface height"used for the sounding instead of "much lower". It would be also nice to have a comparison with the "surface height" of the MWR. This sentence has been revised to now read *However, in ERA5 this is typically around 10 m a.g.l. whereas in the radiosonde soundings this is much lower - typically 2 m a.g.l.*. The potential temperature at the "surface" that the MWR uses is taken from the automatic weather station which incorporated with the MWR. This means that the "surface" potential temperature is at a height of 1.5 m above the surface. We have added this to the manuscript in section 4.2.

5. Fig. S2: the x labels of plot e and f are missing. Thank you for pointing this out. This has now been corrected.

6. Lines 473-479: please add a sentence about the results of Fig. S2: the Heffter method overestimates BLH mostly for unstable EC conditions (similarly to Fig. 4). LL10 overestimate RI mostly for stable conditions (different to Fig. 4, the answer being probably given by Fig. S1). The apparent difference for how the LL10 BL heights compared to the $Ri_{0.25}$ BL heights for different stabilities is related to the different definition and bins of stability. In Figure 4b, the red points include neutral and unstable points identified from the LL10 scheme. The points where the LL10 method has BL heights between 200 - 800 m, which generally are larger than the BL heights, are red meaning either unstable or neutral stratification. In Figure S2, the points in the this area are mainly light blue (near neutral stable), middle blue (weakly unstable) and orange (near neutral unstable). Figure S1 tells us all of these cases are most likley to be diagnosed as neutral by the LL10 scheme and thus appear as red points in Figure 4. Therefore, we see not discrepancies in the results in Figures 4 and S2 that need to be explained.

7. Lines 661-673: In case of weak stability and very low wind speed, the Ri method applied by ERA5 leads to very high, constant BLH. The description is exact, but I do not think that you can qualify the MWR BLH estimate as a failure and ERA5 BLH estimate as the real value. The ERA5 BLH follows the aerosol top layer (Fig S4 c)), which is not always considered as BL top (e.g. the authors do not consider RL top as BLH in previous sections of the manuscript dealing with MWR BL detection during night). Moreover, as mentioned in your paper, MWR is an instrument and ERA5 a model and weak unstability is detected (Fig. S4 d) when MWR presents a CBL development. The only use of the expression "boundary layer" for the various methods and times is an easy shortcut of this manuscript. However it induces a too easy appreciation of the "true BLH", but BL, especially during night, is often complex. Then, to which boundary layer or sublayer correspond ERA5 BLH of the 28th of October 2018? We agree with the reviewer here in that a "true" BL height is not a good term as it implies an exact height whereas in reality, and depending on the application and the method used, a range of values / heights is potentially "correct" or at least appropriate. We have tried to stress this more in parts of the manuscript. In the introduction we added text to stress that our definition of the BL is very broad and that we acknowledge there are many different types of the BLs and that the BL structure is complex. We have also modified the text in lines 661 - 673 to remove terms such as "fail" or "true".

8. Line 673: see previous comment on success and failure. This sentence has now been removed.

9. Lines 795-796: As already commented in the first review, this conclusion is right only if Ri is the right method to resolve BL height. The authors answered that Ri applied by ERA5 leads to

similar results to Ri applied to RS. This proves that ERA5 temperature and wind profiles are similar to sounding profiles, but it does not demonstrate that Ri is the right method. We agree that we cannot conclude that the Richardson number is the "right" method to diagnose the BL, although previous studies have suggested that this is often a good method to use with model / reanalysis data. We have now revised this sentence to be clear that the results show that the temperature and wind profiles in ERA5 and the radiosondes are similar and that if the same method is applied to both to identify the BL height, good agreement can be expected.