Reply to Reviewer report #1's comments

- Line 9: looking at the actual coverage, this is not long term, but multi year. This term is also used appropriately in later parts. Thus, suggest to remove long term.

Answer: We changed 'long term' to 'multiple years of' in the text.

- Line 254: Just wondering, if LLC and LLC-free add up to 1 in the figure (guess it does, but maybe mention it?)

Answer: Yes, LLC and LLC-free occurrence fractions do add up to 1.

- Line 284: "SBL boundary layer" change to "SBL"

Answer: We changed 'SBL boundary layer' to 'SBL' in the text.

- Line 526: Shouldn't this point to figure 2, not 1?

Answer: We thank the reviewer for pointing out the typo. We changed 'Figure 1b ' to 'Figure 2b' in the text.

- Figure 5b: what is this "Case Amount" actually? I thought I understood it, and then was trying to find out what n is of Figure 6. Judging by the numbers (log scale, so bit difficult to read accurately), the n and the Case Amount are the same thing? So Figure 5b could be remove, as the info in in Figure 6? BUT....

- Figure 7: was about to suggest the same as for Figure 5/6 above. But then tried to confirm again, that n and Case Amount are the same thing. However, looking at TWP in 7, it seems slightly above 100. But in Figure 8, it is below 100. So I suggest to define case amount, and if this is n, then remove it from figure 5,7,9 (which seems to agree again with my understanding of n and Case Amount) and only use the n.

Answer: We thank the reviewer for the suggestion. We removed the Case Amount plot in figure 5,7, and 9 as suggested. The Case Amount is basically the same as *n* in Figure 6, except that for Figure 5, it is the mean sample numbers for different comparison groups. Each PBLHT-SONDE method has its own QC flag. Samples with no clear QC flag are removed. Therefore, the effective samples for each method are slightly different. That's the reason why the Case Amount in Figure 5 is slightly different with the *n* in Figure 6.

Reply to Reviewer report #2's comments

My general impression is that the new version of the manuscript reads better. However, it still needs some improvements before it can be accepted for publication. Although I previously suggested some changes, it seems that the authors need a stronger and more explicit supervision. I found a couple major problems in this version that I already pointed out in my previous review and the authors didn't seem to address them satisfactorily.

We are highly appreciated for the reviewer's constructive comments and suggestions, which help to greatly improve the manuscript. We carefully revised the manuscript and tried our best to address the reviewer's comments. The reviewer has a major concern about how Ri is calculated in this study as is presented in the specific comment #12. We want to emphasize that we calculate the 'bulk Richardson number' not Ri. The exact same calculation of 'bulk Richardson number' has been widely used in previous studies for estimating PBLHT. We also modified the definition sentence of Ri_b to emphasize that it is calculated 'between the surface and a given altitude'.

Specific comments:

1. In my second review, I recommended that the authors read the paper carefully again and check every single sentence for consistency. The authors' response was:
   "We carefully revised the manuscript according to the reviewer's comments."
   When I started reading the new version of the manuscript, the first thing that struck me was the abstract, in which you say:"... at three ARM fixed-location atmospheric observatories and from three ARM mobile observatories deployed around the world for various field campaigns, which cover from Tropics to Polar regions and over both ocean and land surfaces." and that information is then repeated in the Introduction.

   According to the above, you have six sources of data in total. Just to remind you one of the points from my previous review:
   L487: "at six ARM observatories located around the world" – This is confusing. Once you say nine (Tab. 1, Fig. 1), then you say six. Again, I strongly encourage the authors to read their paper very carefully and check every single sentence for consistency.

   Response: We thank the reviewer for pointing out the typo. The reviewer is correct that it should be nine.

   Clearly, it was an error not a typo and that error was then repeated in the abstract and the Introduction and is still there. It is difficult for me to understand what the authors mean by "we carefully revised the manuscript" since I find the same simple errors again.

   Answer: we apologize for causing the confusion. In line 118, we pointed that 'ARM operates three mobile facilities (AMFs) which can be requested by scientists through a proposal process for various field campaigns that deploy ARM instruments anywhere in the world for roughly a year'. Each ARM mobile facility could be deployed at a

location for a field campaign and then moved and deployed at another location for a difference field campaign. The three ARM mobile facilities were deployed at various locations around the world for field campaigns during the past decade.

To avoid the confusion, we delete the two 'three's in the abstract.

"which cover from Tropics to Polar regions" – cover what? Use lowercase.

Answer: we changed 'which' to 'These observatories' and used lowercase for 'the tropics' and 'the polar regions' in the text.

I suggest to remove the number of stations from the abstract and describe them in a more general way, for instance as a data set from different climate zones probing a variety of PBL regimes.

Answer: we removed the number of stations from the abstract as suggested. Detailed descriptions of ARM observations are presented in Section 2.

2. Title: should it be "with" or "and"? The current form suggests that radiosonde data is good and ceilometer data needs to be evaluated.

Answer: we agree with the reviewer that PBLHT from radiosonde data is good, so we use 'with' in the title.

3. L30: Not true. PBL height does not characterize the structure of the lowest few kilometers of the atmosphere by any means. It only indicates where the top of that PBL structure is located.

Answer: we deleted this sentence to avoid the confusion.

4. L36-37: "by an inversion layer of potential temperature" – this sentence is unclear. Is it a temperature inversion? Why "inversion layer"? For convective PBLs potential temperature actually does not have an inversion at the top as its stratification changes from neutral to stable, right? Did you mean real temperature, for which temperature inversion can make more sense?

Answer: we thank the reviewer for pointing it out. It should be 'temperature'. We deleted 'potential' in the sentence.

5. There are at least 9 different methods of estimating PBL height. Von Engeln and Teixeira (2013) mention many of them. In large-eddy simulations, it is common to apply gradient methods (for temperature or moisture) or indeed look at turbulence properties. Some examples worth citing:

   Bopape, M.-J.M.; Plant, R.S.; Coceal, O. Resolution Dependence of Turbulent Structures in Convective Boundary Layer Simulations. *Atmosphere* **2020**, *11*, 986. https://doi.org/10.3390/atmos11090986

J. Kurowski, M., P. Malinowski, S. and W. Grabowski, W. (2009), A numerical investigation of entrainment and transport within a stratocumulus-topped boundary layer. Q.J.R. Meteorol. Soc., 135: 77-92. https://doi.org/10.1002/qj.354

Answer: we thank the reviewer for pointing out these previous and recent works. We added references to them.

6. L84: ", the robustness..." – this should either be a new sentence or you should add "and" in between

   Answer: we started a new sentence as suggested.

7. L87: "long term" – this is too vague and means different things for different people; be more specific: multi-year?

   Answer: we replaced 'long term' with 'multiple years of' in the text.

8. L97: "low-level cloud-free" – what are low level cloud free conditions here? What does that low level refer to? Does it mean there are clouds at some upper levels?

   Answer: we defined the 'low-level cloud-free' condition in the line 273 at the 'Results and Discussions' section as: "comparisons are also separated for conditions with and without the presence of low-level clouds below 4 km AGL (referred to as LLC and LLC-free, correspondently), as detected by the ceilometer at the time of the radiosonde launch."

9. L131-132: "three commonly used methods developed by Heffter (1980; referred to as the Heffter method hereafter)" – does it mean you name the three methods as the Hefter method?

   Answer: we rewrote the sentence as:

   'The ARM PBLHT-SONDE VAP implements three commonly used methods including the Heffter method (Heffter 1980), the Liu and Liang method (Liu and Liang 2010), and the bulk Richardson number method (Seibert et al., 2000) to estimate PBLHT from radiosonde data (Sivaraman et al., 2013).'

10. That description of 3 methods is still slightly chaotic.
    Please summarize the text between L130 and L155 with an additional table: Method name, algorithm (e.g., Richardson number, potential temperature gradient), reference, etc. which should help understand your methodology.

    Answer: as suggested, we added a table to summarize the algorithms that are used to estimate PBLHT.

11. L145: when you say between fifth and second levels of sounding data it means nothing to the reader. Explain what levels 5 and 2 mean. Is it at a fixed height which is the same at all stations

under all conditions?
In principle, if you use a threshold for potential temperature difference, it means you assume that there must exist a minimum gradient. So that method also relies on temperature gradient.

Answer: Following Liu and Liang (2010), the potential temperature (θ) difference between the fifth and second level of sounding data (θ_5 - θ_2) is used to represent the near-surface thermal gradient. We added this illustration in the text. (θ_5 - θ_2) might correspond to slightly different heights and pressures for different locations due to different site altitudes and local atmospheric environments. In line 179, we pointed out that 'the radiosonde data is subsampled at a 5 mb resolution, corresponding to vertical height resolutions of 30 to 60 m depending on the atmospheric environment'. It is true that the threshold might be affected by multiple factors. We pointed out in line 198 that 'threshold values of δ_s, δ_u, and $\dot{\theta}_r$ are dependent on the surface type and are empirically determined in Liu and Liang (2010).' The method is not perfect for determining PBL regimes, but it provides a practical way to estimate PBL regimes using radiosonde data.

12. L160: We need to discuss this definition of Ri one more time. According to your description: "Ri at a given altitude can be calculated from sounding data"

$$Ri_b = \left(\frac{gz}{\theta_{v0}}\right)\left(\frac{\theta_{vz}-\theta_{v0}}{u_z^2+v_z^2}\right)$$

As I explained in my previous reviews, this particular definition refers to the entire layer from 0 (surface) up to some height z. This is not Ri at a given altitude and you can only do it for a layer of finite height. If you want to look at a layer of thickness dz, then it should rather be:

$$R_B = \frac{(g/T_v)\Delta\theta_v\Delta z}{(\Delta U)^2+(\Delta V)^2}$$

Because your definition replaces all the gradients with the values at the height z, it implicitly assumes that those values diminish at z=0 (which is true for u and v, but not for theta that uses a difference between level 0 and z in your equation. How exactly is Ri calculated in your data?

If you still claim that Ri in your equation (1) is given at the height z, then we have a problem with understanding basic equations used in your study. I tried to explain it in my previous reviews, but was so far unsuccessful.
It is critically important that you understand and explain the way you use Ri in your study correctly as it is the foundation of your analysis.

Answer: We thank the reviewer for the suggestion. We used the exact same definition of 'bulk Richardson number (Ri_b)' as was widely used in previous studies such as the Equation (7) in Sørensen et al., (1998), the Equation (9) in Seibert et al., (2000), the

Equation (1) in Zhang et al., (2014), and the Equation (1) in Bakas et al., (2020), etc. We calculated and used 'bulk Richardson number (Ri_b)' not Ri. To avoid such confusion, we changed the definition to 'Ri_b between the surface and a given altitude *z* can be calculated from …' in the line 213 in the text. The equation is also consistent with the R_B equation the author suggested given that the PBL is the lowest layer of the atmosphere. Although a critical value (Ri_{bc}) is not well defined for the bulk Richardson number (Ri_b), Seibert et al. (2000) suggests an optimal Ri_{bc} value of 0.25 when applied to radiosonde data. All variables on the right side in the equation (1) can be obtained from radiosonde data and therefore it is straight forward to calculated 'the bulk Richardson number' from radiosonde data.

References:

Bakas, N.A.; Fotiadi, A.; Kariofillidi, S. Climatology of the Boundary Layer Height and of the Wind Field over Greece. *Atmosphere* 2020, *11*, 910. https://doi.org/10.3390/atmos11090910

Seibert, P., Beyrich, F., Gryning, S. E., Joffre, S., Rasmussen, A., and Tercier, P.: Review and intercomparison of operational methods for the determination of the mixing height, *Atmos. Environ.,* 34, 1001–1027, https://doi.org/10.1016/S1352-2310(99)00349-0, 2000.

Sørensen, J.H., Rasmussen, A., Ellermann, T. and Lyck, E.: Mesoscale Influence on Long-range Transport – Evidence From ETEX Modeling and Observations, *Atmospheric Environment,* 32(24): 4207–4217, https://doi.org/10.1007/978-1-4615-4153-0_27, 1998.

Zhang, Y., Gao, Z., Li, D., Li, Y., Zhang, N., Zhao, X., and Chen, J.: On the computation of planetary boundary-layer height using the bulk Richardson number method, Geosci. Model Dev., 7, 2599–2611, https://doi.org/10.5194/gmd-7-2599-2014, 2014.

13. L265: "because Liu-Liang method uses different methods" – confusing

Answer: We changed it to 'because the Liu-Liang method uses different algorithms …as discussed in section 2' to make it clearer.

14. Fig. 11: Make axes labels and ticks larger.

Answer: We thank the reviewer for the suggestion. We use larger axes labels and ticks in the Figure.