Reply to Reviewer #1's comments

- *The author aimed to compare the ceilometer- and radiosonde-estimated PBLHTs under stale, unstable and RL, cloudy and cloud-free conditions. But what is the stability parameter used and how is the RL defined in this study? How the cloudy and cloud-free condition is defined? It should be explained.*

We pointed out in line 145 that 'The Liu-Liang method classifies the boundary layer regime as CBL, SBL, or NRL by comparing $\theta_5 - \theta_2$ with a stability threshold δ_s'. We added another sentence at line 146 'For CBL, $\theta_5 - \theta_2 < -\delta_s$; for SBL, $\theta_5 - \theta_2 > +\delta_s$; and for NRL, $-\delta_s < \theta_5 - \theta_2 < +\delta_s$' to provide more details about how the stability regime is determined.

We pointed out in line 179 that 'The Vaisala CL31 ceilometer detects up to three cloud layers simultaneously and measures vertical visibility'. To make it more clear, we added a sentence in line 180 'Ceilometer cloud detections are used to distinguish cloudy and cloud-free conditions'.

- *The observation data used in this paper include both over land and ocean. However, what is the difference between the accuracy of PBLH estimation over land and ocean? It is suggested to be explained in the manuscript.*

We agree with the reviewer that when evaluating the retrieved variables, it is important to examine the retrieval bias, accuracy/differences between the retrievals and the ground truth. However, a great challenging for PBLHT estimations is that there is no ground truth to evaluate with. We pointed this out in the line 221. It is difficult to obtain the overall accuracy of the two ways of estimating PBLHTs. We believe that good comparisons between ceilometer- and radiosonde-estimated PBLHTs generally indicate more reliable PBLHT estimations.

- *In Figure 2, what is the reason for the great difference in PBLH retrieved by different methods at 18:00 LT? According to the attenuated backscatter coefficient, it is well mixed within the PBL, generally, the uncertainty of PBLH retrieving should be relatively small under this condition?*

We agree with the reviewer that for well mixed PBL, the uncertainty of PBLHT retrieving should be relatively small. The PBL structure at 18:00 LT is more complicated. In lines 227-229 we pointed out that 'At 17:30 LT on February 10, there is a weak stable layer developed near the surface, where the low altitude atmosphere is still well-mixed. This is a typical structure of a residual layer overlaying a weak stable layer. PBLHT CEIL and PBLHT Heffter captured the top of the residual layer, while PBLHT Liu-Liang is underestimated. PBLHT bulk Richardson is quite low, because it takes the top of the weak stable layer as the PBLHT'.

- *In Figure 3, The profiles of backscatter and Richardson number is incomplete, which will lead us to doubt the rationality of the data. In addition, what are the reasons for the difference of PBLH retrieved by different methods? Because the defect of the method or the structures of the PBL? should be explained.*

The Vaisala CL31 ceilometer has a field-of-view of 0.83 *mrad* and receives considerable background signals when pointing vertically. Therefore, subtracting background signals during the post-processing procedure leads to noisy ceilometer backscatter profiles above PBL when the atmosphere is free of clouds or aerosol layers. Early studies show that CL31 ceilometer is capable of detecting aerosol layers and can be used to estimate PBLHT (Münkel et al., 2007).

The Bulk Richardson number increases dramatically above PBL and is out of the x-axis range in Figure 3. We pointed out in line 154 that 'The bulk Richardson number *Ri* represents the ratio of thermally produced turbulence to that generated by vertical wind shear. Since wind shear produced turbulence is greatly reduced above the top of atmospheric boundary layer, *Ri* increases dramatically at the top of SBL.'

As for the reasons of the difference of PBLHT retrieved by different methods, we believe that it is because of the limitation of the measurements, the defect of the methods, and the complicated structures of the PBL. (1) temperature, humidity, and aerosol intensity measurements only indirectly reflect PBL structures. Direct measurements of PBL turbulence structures with high temporal and vertical resolutions are not available; (2) the retrieval methods are often based on empirical relations, which might not be applicable to certain complicated PBL structures; (3) it is still challenging to obtain reliable PBLHT estimations under stable boundary layer conditions. In this study, we show that 'under unstable boundary layer conditions, ARM low- and mid-latitude land observatories have higher correlation coefficients and good comparisons between PBLHT CEIL and PBLHT SONDE. ARM observatories at the ocean surface and under stable boundary layer conditions have weak correlation coefficients between PBLHT CEIL and PBLHT SONDE.'

Reference:

Münkel, C., Roininen, R.: Automatic Monitoring of Boundary Layer Structures with Ceilometer. vol. 184 Vaisala News., 2010.

Reply to Reviewer #2's comments

Main comment:
This is an interesting paper that has an element of novelty: it presents a comparison between ceilometer and radiosonde data for a range of PBL sites and regimes. I generally think that the paper is worth publishing although it needs major improvements.

We thank the reviewer for these constructive suggestions and comments. We carefully revised the manuscript according to the reviewer's comments.

Specific comments:

You show correlation coefficients, but I don't see any calculations of and the discussion on the biases between the tested methods. Can you please add it (if not to the paper, then to the Supplement)? This is an important part that should complete the whole picture.

What is the significance of your correlation coefficients? You calculate those values but don't seem to comment on their significance. To me it looks like only convective PBL can be relatively well probed with the two instruments but I couldn't find anything about the overall accuracy/differences between the two ways of calculating it.

We thank the reviewer for these two very constructive comments. We agree with the reviewer that when evaluating the retrieved variables, it is important to examine the retrieval bias, accuracy/differences between the retrievals and the ground truth. However, a great challenging for PBLHT estimations is that there is no ground truth to evaluate with. We pointed this out in the line 221. It is difficult to obtain the overall accuracy of the two ways of estimating PBLHTs. We believe that good comparisons between ceilometer- and radiosonde-estimated PBLHTs generally indicate more reliable PBLHT estimations.

You could shorten your PBLHT acronym to PBLH and still be well understood.

We thank the reviewer for the suggestion. In literature both PBLHT and PBLH have been used (Huang et al., 2011; Nelson et al., 2021). We prefer PBLHT in the manuscript to keep consistent with the ARM PBLHT-VAP.

References:

Huang, M., et al.: Multi-scale modeling study of the source contributions to near-surface ozone and sulfur oxides levels over California during the ARCTAS-CARB period, Atmos. Chem. Phys., 11, 3173–3194, https://doi.org/10.5194/acp-11-3173-2011, 2011.

Nelson, K. J., Xie, F., Ao, C. O., & Oyola-Merced, M. I. (2021). Diurnal Variation of the Planetary Boundary Layer Height Observed from GNSS Radio Occultation and Radiosonde Soundings over the Southern Great Plains, *Journal of Atmospheric and Oceanic Technology* (published online ahead of print 2021).

Abstract:
L7: How can a parameter influence atmospheric processes? Processes are controlled by the physics, not by PBLHT. Maybe you meant that their representation in climate models depends on PBLHT?

We agree with the reviewer that processes are controlled by the physics. We deleted this sentence.

You mention PBL types (stable vs unstable, cloud free and cloudy conditions. More information is needed on latitudes (mid-lat vs tropics?) and types of the surface (maritime vs continental).

As suggested, we added a sentence of "which cover from Tropics to Polar regions and over both ocean and land surfaces" in the line 13.

"Under unstable boundary conditions" – you can simply say "For convective boundary layers"

We thank the reviewer for the suggestion. As shown in Figure 4 and discussed in lines between 403-407, "The CBL regime generally has a small fraction for all the observatories and is negligible for the TWP, MAO, and SGP observatories. Since the CBL and NRL have similar state variables and pollutant profiles, we combine the CBL and NRL regimes and refer to it as the unstable boundary layer condition, in contrast to the SBL regime, which stands for the stable boundary layer condition." Therefore, we used "unstable boundary layer conditions" in stead of calling it "convective boundary layers".

Introduction:
L24: and moisture

We added this in the sentence.

L27-30: Unclear: is the depth a parameter that determines the structure of the lowest few km? The depth is the result rather than a cause of the PBL processes.

We agree with the reviewer and changed the word "determines" to "characterizes".

L33-35: I think a different classification is more common: stable, unstable, neutral PBLs, depending on their mean stratification. This is however not what Liu-Liang method is based on. Generally, that classification is somewhat simplified and may not be appropriate for transient cases, which should be mentioned.

We agree with the reviewer that the classification using Liu-Liang method is simplified. We added a sentence in lines between 58-61: "It should be noted that atmospheric boundary layer stability ranges from very stable to strongly unstable. Classification of atmospheric boundary layer stability into these three major regimes is simplified and may not be appropriate for transient atmospheric conditions (Mahrt 1999)."

It may be helpful to clarify that Liu-Liang method is simply a temperature gradient method.

As suggested, we added a sentence "For example, both Heffter (1980) and Liu and Liang (2010) use potential temperature gradient as a key parameter to determine the PBLHT for CBL and NRL regimes." in lines between 69 and 71.

L36: for CBL it is both convection and turbulence that cause strong mixing across the PBL

We thank the reviewer for the comment. We added "convection" in the sentence.

L37: for shallow CBLs there should be some temperature inversion at the top, but for deeper CBLs (e.g. in the tropics) it is more difficult to determine the top of the PBL

We thank the reviewer for pointing out this situation. We added a sentence "For deep CBLs such as in the Tropics, however, it might be difficult to determine the top of the boundary layer using the potential temperature inversion (Kepert et al., 2016)" and a reference in lines between 52 and 54.

Mention briefly about different definitions of PBLHT used in atmospheric models and observational studies: maximum Richardson number, temperature inversion, moisture gradient, minimum refractivity gradient.

We thank the reviewer for the great suggestion. We added the following sentence to talk about PBLHT determinations in atmospheric models in line 63-65.

The PBLHT in atmospheric models is usually calculated by using either diagnostic equations that take surface fluxes and the initial temperature profile as inputs or by using the Richardson number profile to find the first level where the Richardson number exceeds a critical value (Seibert et al., 2000).

L127: what does theta with two dots above mean?

It should be one dot above theta, which represents the potential temperature gradient. We corrected this typo.

What is the vertical resolution of the soundings? What is the accuracy of wind and temperature measurements and thus the overall accuracy of the methods used?

We mentioned in line 109-110 that "SONDE … with a 1-second resolution". To make it clearer to readers, we now added in line 165-166: "corresponding to vertical height resolutions of several meters to more than 10 meters depending on the atmospheric dynamic environment."

We also added the accuracies of wind and temperature measurements in line 165: "The accuracies of radiosonde measured temperature and wind speed are 0.2 °C and 0.2 m/2, respectively".

As we pointed out in line 221 that there is not ground truth for PBLHT determinations, it is difficult to have an overall accuracy of the methods used in PBLHT-SONDE VAP.

L136: Explain why Ri dramatically increase at the top of SBL.

We added a sentence in the line 201: "Since wind shear produced turbulence is greatly reduced above the top of atmospheric boundary layer,".

Eq. 1: That equation is different from a typical one for bulk Richardson number including temperature and velocity gradients. I think this is because you calculate mean properties in the entire layer between z=0 and PBLH, which should be clarified in the text. I am skeptical about using such a bulk method for determining PBL height. The thicker the layer, the more risk that you omit important turbulence activity between 0 and z. That is why maximum Richardson number method can be more beneficial: instead of looking at one thick layer we can look at a number of thin layers for which Ri is calculated.
Does your Richardson number method really use Eq. 1 or looks into different layers within PBL?

We thank the reviewer for pointing out an alternative way to use the maximum Richardson number method for improving PBLHT estimations with radiosonde data. We use PBLHTs estimated with the bulk Richardson method directly from the ARM PBLHT-SONDE VAP (Sivaraman et al., 2013), which implements the equation (7) in Sørensen et al. (1998). The bulk Richardson number is calculated at each height, so it is not the mean property in the entire layer between z=0 and PBLHT. As we mentioned, the goal of the study focuses on investigating the robustness of ceilometer-estimated PBLHTs. Improving PBLHT estimations from radiosonde data using the maximum Richardson number method is out of the scope of this study.

References:

Sivaraman, C., McFarlane, S., Chapman, E., Jensen, M., Toto, T., Liu, S., and Fischer, M.: Planetary boundary layer (PBL) height value added product (VAP): Radiosonde retrievals, U.S. Department of Energy Rep. DOE/SC-ARM-TR-132, 36 pp., https://www.arm.gov/publications/tech_reports/doe-sc-arm-tr-132.pdf, 2013.

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.

3.1 LLC-free unstable boundary layer conditions – I suggest to make those section titles more self-descriptive, for instance:
LLC-free unstable boundary layer conditions -> Cloud-free unstable PBL (or similar)
LLC-free stable boundary layer condition -> Cloud-free stable PBL (or similar)

As suggested, we changed those section titles to be more self-descriptive.

L315: typo: Richardons -> Richardson

We corrected the typo.

L361: You claim that "machine learning techniques have the potential to greatly improve PBLHT estimates" – but you didn't prove it in this paper, so this claim has no foundation and should be removed. The same sentence is present in the abstract, and it is totally unjustified as far as I can see.

We deleted discussions that are related to machine learning techniques in both the abstract and summary sections.

Figures;
Fig 3: What are the units of backscatter?

The units of ceilometer backscatter coefficient is $sr^{-1}m^{-1}$. To reduce the data storage size and be easier shown in the figure, the units of ceilometer backscatter coefficient was converted to (1/(sr*km*10000)). To be consistent with the standard units, we now converted it back to $sr^{-1}m^{-1}$ in both Figure 2 and Figure 3.

Fig 4: Is it a climatology? Or just a selected time period? I don't understand it.
Explain what fraction means here (fractional occurrence?).

Figure 4 shows the occurrence fractions of boundary layer regimes and low-level cloud-free/cloudy conditions during the selected time period. Fraction is the occurrence fraction. We added the information in the figure and the caption.

Fig 11: explain the range of the boxes and whiskers in the caption.

We added a sentence "Horizontal bars, boxes and whiskers represent the median, interquartile range and range of the data" to explain the range of the boxes and whiskers in the caption.

You could cite this paper:

1. Fritz, A. M., Lapo, K., Freundorfer, A., Linhardt, T., & Thomas, C. K. (2021). Revealing the morning transition in the mountain boundary layer using fiber-optic distributed temperature sensing. Geophysical Research Letters, 48, e2020GL092238. https://doi.org/10.1029/2020GL092238

(they show how important it is to measure PBL transitions with high spatio-temporal resolutions and suggest using a temperature gradient method)

We appreciate the reviewer's suggestion. We agree that this paper highlights the need for high spatial-temporal PBLHT estimations. We added a sentence "Observing atmospheric boundary layer transitions with high temporal-spatial resolutions is required to investigate atmospheric thermodynamic processes (Fritz et al., 2021)." and the reference to this paper in lines between 73-74.