Response to Anonymous Referee #1

We greatly thank the reviewer for his/her comments. Below are our responses in blue

Reviewer comments

General: An interesting contribution on the synergetic use of different satellite measurements to monitor some PBL characteristics. I have a few major concerns, and various minor questions comments that I'd like to see addressed in an update. I rated the manuscript as Major Revision, mostly to leave the authors more time if any of my comments require larger work.

Major Issues:

- I am missing some more information on the systematically different sonde data vs. the RO data. Figure 2 and 3 seem to show different conditions, thus potentially sampling different areas, with RO data having CWV data up to 3cm, while the sonde data extends only up to about 1.5cm. Is the RO sharpness primarily identifying regions that are more humid while the sonde data is in less humid areas?

As the reviewer points out, the difference is mainly sampling different areas, that is, when the sharpness parameter is greater than 2.5 (top panel in figure 4) the comparison encompasses regions where a lot of cumulus clouds might be present (see Figure 3 for reference), hence, the BL-CWV is higher. However, when the comparison is made for sharpness parameters greater than 3 (bottom panel in figure 4) the comparison is mostly over stratus regions and the BL-CWV values resemble more the sonde values. The sondes on the other hand are always restricted to stratus regions since we only used robust inversions, that is, when the three different methods (explained in the text) agree within 200m.

The following sentences will be modified,
(P4 line 7 of the original manuscript): That is, those inversions where the boundary layer inversion height estimates of the tree methods agree within 200m, **which mostly occur in stratus regions**.

(P5 line 8 of the original manuscript): This improvement arises because when using a larger sharpness parameter, we are ensuring that most pairings are in the stratus regions **(see Figure 3 for reference)** where the AMSR-MODIS technique should work better. **A larger sharpness parameter also reduces the range of the BL-CWV comparison by excluding the high values found under cumulus regimes. This makes the comparison ranges (that is, in Figure 4-bottom) similar to the ones found in the sonde comparison where the sondes used are restricted to stratus regions by using the robust inversion criteria. That is, when the three different methods to find the inversion (explained in section 3.1) agree within 200m.**

Is there any chance to also compare RO to sonde to improve the understanding? After careful consideration, we feel that such comparison belongs to a GPS-RO validation paper and not here. However, we hope that our previous explanation will suffice for the reviewer.

Or to include more data?
We included as much radiosonde and GPS-RO data as we could find.

- Page 2, Line 31 (P2/L31): points to systematic issues in the December months. Just "removing" a month because it does not seem to fit, and use a another data version that fits, would need a much more substantiated justification. Thus I'd like to see more info on what might cause this issue and why the version 6.1 was used now. And then, why 6.1 is not used throughout.

The issue was a one-off coding error on the MODIS processing algorithm (per the personal communication with Richard Frey). We discover the issue after the whole dataset was produced with version 6.0 and a reprocess of the AMSR-MODIS using version 6.1 through-out the entire time period is currently outside our possibilities due to the large time involved in downloading the terabytes of MODIS data. We will modify the following sentence: Instead, version 6.1 was used for all December months **as recommended by the MODIS team. A full reprocessing of the AMSR-MODIS dataset using MODIS version 6.1 (or the latest MODIS version) is left for a future AMSR-MODIS version.**

Minor / Editorial Issues:

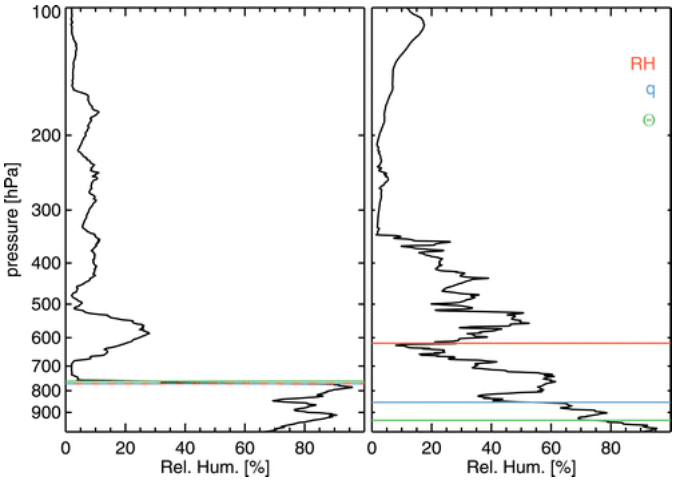
- Section 2: would it be worthwhile to point out that MODIS measures around 1:30pm? The following sentence will be modified to: All these instruments orbit in tandem measuring the same volume of air within minutes of each other, that is, by design, these measurements are collocated; their equatorial crossing time is ~1:30pm.

- P2/L27/L33: the estimated errors are mentioned here, could you include also whether this is a systematic or random uncertainty / error? Both of these estimates are random errors. This will be reflected in the text.

- P3/L11: "as the number of the number": correct? The sentence will be changed to: It also shows its associated standard deviation as well as the number of single observations (MODIS pixels) …

- P4/L7: "data below 200m or above 4km": I could of course look at the cited article, but maybe it can be better explained here. Does this mean if the inversion layer is below/above, the sonde is not used? Or is this data below 200m not used in the integral (which would be more of an issue)?
We only exclude this data in the BL height determination analysis, the sentence will be modified to: As in Millán et al. (2016), **during the inversion height determination**, we exclude all the data below 200 m or above 4 km **to avoid artifacts caused by temperature inversions near the surface as well as to avoid free-tropospheric features**. **Further**, we use only robust inversions, that is, …

- P4/L8: How much data is removed in these screening steps? And did you try to loosen the threshold of the three methods to be within 200m to increase the sample size?
We did not quantify how much data was screen by those steps, because such a stringent threshold is needed to ensure that the radiosondes used represent a well behave boundary layer profile. As shown in the following figure, relaxing such threshold could result in using sonde profiles where the boundary layer is not well defined.



Examples of relative humidity (RH) sonde measurements: (left) a robust inversion and (right) an unrobust inversion. Color lines display the boundary layer height determined using the location of the minimum vertical gradient of RH (red), the location of the minimum vertical gradient of specific humidity *q* (blue), and the location of the maximum vertical gradient of potential temperature Θ (green). Figure 3 from Millán et al (2016) doi: 10.1175/JAMC-D-15-0143.1

- P4/L24: The RO processing articles seem rather outdated, I assume that the processing uses some more recent algorithms, e.g. 1DVar (I might be wrong though). And where does the data come from, UCAR? The version 2.6 though is not a UCAR identifier as far as I can see. It would also be interesting to get more info on the humidity background if a 1DVar was used. Otherwise, RO will not provide very accurate info on humidity in the mid troposphere. The data comes from the JPL retrieval algorithm, for which those are the most current references. The following sentence will be modified to: In particular we use version 2.6 **of the JPL processing algorithm.**

- P6/26: When using ERA-I data, is exactly the same data at the same time used, or is this a larger data set? The following sentence will be added at the end of that paragraph: Note that for each region, ERA-Interim data from the nearest synoptic time (0,6,12,18 UT) to the measurement local time was used.

- Figure 1: Please use different color range for BL-CWV and its std dev, to allow more visualization of the std dev values. After careful consideration, we decided that it was best to leave the colorbar as it is. This way, the reader gets an immediate sense of the size of the standard deviation, as opposed to have to look at the ranges of two colorbars.

- Table 2: What is the last column? It seems not the number of obs. It is the correlation between BL-CWV and the number of observations. The caption of the table will be change to: Climatological (top) and interannual (bottom) correlation coefficients between BL-CWV and **SST, LTS BL-CTH and the number of observations**.

- Figure 4: There appear to be also negative CWV in this plot, is that found in certain regions/for low CWV values? Yes and no, that is, they do tend to occur more at low CWV values but not only at those places. Any reason why this is not excluded from the data set? Seems to point to the MODIS overestimation you mentioned. We believe that excluding these points could lead to a high bias of the BL-CWC. The following sentence will be added after the MODIS overestimation discussion: The overestimation of the MODIS CWV above the clouds could lead to negative values in the AMSR-MODIS dataset (as can be seen in Figure 4). However, we do not recommend that these negatives values are excluded of any analysis of the AMSR-MODIS dataset because some negative values will be due to the noisy nature of the MODIS measurements over cloudy pixels, and excluding those will lead to biasing high.

- Figure 5: These rectangular/boxes are from a publication in 1993. Some question on that: Why is the Australian not included? The Australian region is a relatively weaker SC region and hence is harder to interpret. Why not use all identified regions in that article(or at least the mid-lat marine stratus ones)? Again, because we focus in the robust SC regions. And last, is there a good reason to update the boxes with the latest data available? We have a better, higher resolution picture of our planet compared to 1993. Several studies have used these boxes and we wanted to compare against them without adding an extra level of complexity by changing the study regions, but the reviewer is correct, and a paper studying the impact of choosing different regions in the SC vicinities and its impact upon stratus amount and BL-CWV would be interesting.

- Figure 6, 7, 8, 9: my color print out shows almost exactly the same colors for Peruvian and Namibian. I cannot distinguish them at least. On screen it is okay. We do have high end color printers, thus I assume this might also happen for others printing it out. When zooming into Figure 5, I noted that the boxes have the same color, but I think that is not really necessary and limits your color options. The color for the Namibian region will be changed for a darker purple.

- Figure 6: Maybe I missed it somewhere in the discussion, but this has no NH/SH shift, but that is found in Figure 7 and others. In 6, it seems they all peak around July/August. The difference is not as visible in Figure 6 as it is in Figure 7, however, it is there. As stated in the text (P6 line 19 of the first version): The annual cycle us notably stronger in the Peruvian and Namibian regions with maxima during August *(as opposed to July)* and the peak lasting from June to November.

- Figure 6: Is the result better visible if it is normalized to the total in the area?
Below is the figure as requested by the reviewer, as shown is really similar to the Figure shown in the original paper. The reason is because the area of the regions is really similar (all regions shown are 10 by 10 degrees) which is roughly 1e6 KM2 (the number we used originally to scale them. In responding to this question, we realize that the y-label was wrongly shown as counts*1e6 when it should have been counts / 1e6. This will be updated in the new manuscript.

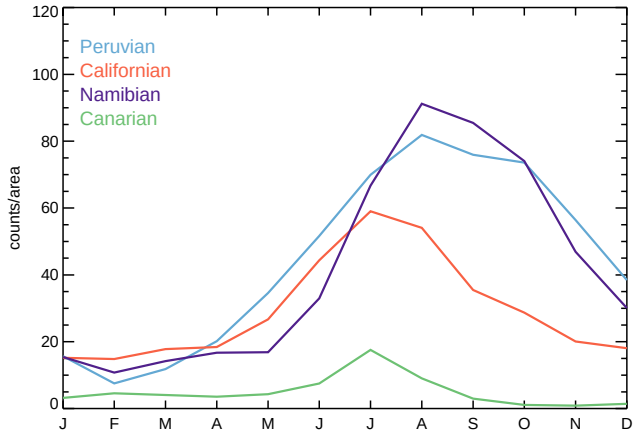


Figure: Annual cycle of the total AMSR-MODIS number observations (scaled by the total area of each region) for the regions delimited in figure 5 by the rectangular boxes

Also, these regions are different in size and cover different subsidence, does that have any impact?
As mention above, they are really similar in size, and while they do have a somehow different subsidence (see figure below) , the counts are not consistently correlated with subsidence (in this case using omega500 from ERA-Interim), as LTS is (as shown in Figure 7 of the manuscript).

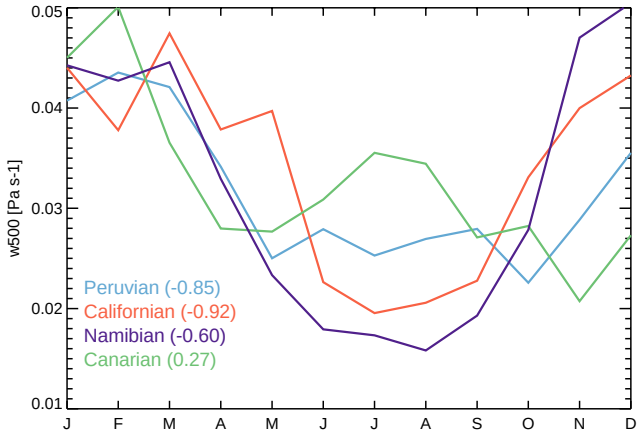


Figure: Annual cycle of w500 (from ERA-Interim) for the regions delimited in figure 5. The number of brackets are the correlation coefficient between the annual cycle of the number of observations and these w500 cycles.

We will add in the manuscript: Other parameters (MODIS CTP, AMSR SST, ERA-Interim w500 and ERA-Interim Surface Pressure) were analyzed in a similar manner but none of them were strongly correlated with the number of observations across the four regions used here.

- Generally, is there any impact visible of the different AMSR instruments used here? We didn't find any, that is, we analyze time series of the AMSR total CWV and SST globally as well as in the regions study and did not find any visible impact / discrepancies. The following sentence will be added in the paragraph describing the AMSR instruments: Note that, no discrepancies nor visible impacts were found in time series from these two instruments.