The paper by Cimini et al. presents an interesting approach to derive continuous Mixing Layer Heights (MLH) from multi-frequency, multi-angle microwave brightness temperatures that are calibrated to MLH from lidar. This technique avoids the loss of information involved in first retrieving temperature and humidity profiles and then applying radiosonde diagnostics. In summary, the manuscript is well written and presents a novel concept for possible application to several observations. A weakness of the concept is its pure statistical calibration and the paper lacks information on the physical processes behind. I this respect I have three main issues which need to be taken into account before final publication:

1) I am strongly missing some **physical explanation** what information from the set of microwave brightness temperatures (Tb) contributes most strongly to the MLH estimate. Such understanding is not only important because of scientific curiosity but also to further improve the method. Because a simple regression algorithm is used to retrieve MLH from Tb with lidar MLH as “truth” this should be quite simple to identify by systematically reducing the observation vector y. For example it is rather interesting to know if the K-band (humidity information) improves the MLH estimation. Other questions are: What is the contribution of the low elevation angles – particular of the transparent ones where inhomogeneity can cause problems? What is the effect of cloud sensitive channels? What information causes the extrapolation of MLH to lower altitudes (cf. page 4982, line 21)?

In particular I have my doubts about the statistical significance of training the regression for each month and worry that overfitting takes place. Waht is the physical reason behind selecting months (not seasons)? For an operational application anyway an algorithm based on prior data needs to be available. Therefore a robust algorithm is necessary and requires an understanding which information contributes to the retrieval result (see above). Based on understanding a further separation by phenomena (like day/night or cloud/no cloud) could make sense but a separation in respect to months does not make sense to me. The author need to give information on how strongly do the results change when all months are used at the same time?

2) The manuscript does not mention the effect of **clouds** – in particular boundary layer clouds. A large fraction of all data presented is certainly affected by that. For the STRAT2-D algorithm I assume that MLH height is coupled to cloud base? What is the effect for the microwave radiometer (see above)? What happens in broken cloud situations that are characteristic for well developed boundary layers? What effect has the averaging interval on that?

3) The authors should comment on the value of **MLH retrievals during night**. Some publications argue that “..these methods are most applicable to the daytime or convective boundary layer and not the night-time or stable boundary layer. At night, surface diagnostic methods are a good proxy for the depth of the stable boundary layer.” (Schmid and Niyogi, 2012). Being provocative one could also say that based on the presented results (Fig. 7) the nocturnal MLH is just 200 m – an estimate which fits into the uncertainty range of all the different estimates. Is there at least one convincing example that MLH retrieved from Tb provide important information for applications like air quality which can not be seen by lidar due to the overlap effect?

Additional points:

Abstract: For people who only read the abstract the location of the study (Sirta or “a typical mid-latitude site”) needs to be mentioned.

„The proposed method provides results that are more consistent with radiosonde estimates than MLH estimates from MWR retrieved profiles. – where is it shown?

2.1 Radiosonde data:
- The authors provide the number of radiosondes available but should also give here – in the beginning – the time interval used in this study.
- The threshold for the bulk Richardson number is taken from the “grey literature” with different values for day and night. The literature is rich in different numbers and I wonder why the authors didn’t choose a more common one like 0.25 used also with ERA-Interim (van Engelen and Texeira, Journal of Climate, 2013). The authors should give at least an estimate on the sensitivity of the threshold. An problem for radiosonde determined MLH is the vertical resolution as several older studies use only low resolution profiles. In fact Schmid & Niyogi (2012) exploit variability of high resolution profiles for a new method to derive planetary boundary layer.

2.2 MWR data:
- The authors should specify Tb uncertainty – I know it is tricky – but at least they need to provide the numbers, which resemble the measurement error (Eq. 1).
- spelling Lönhert = Löhnert

2.3 Lidar data
This section should mention temporal and vertical resolution and the overlap height
- First the authors provide the comparison of the “original” STRAT2-D algorithm with 53 radio soundings (Tab. 1) while later an improved STRAT2-D algorithm (after Pal et al., 2013) is used. It would be good to know if the improved algorithm improves the rather poor skill during night in Tab. 1. Or is this a feature of the difficult estimation of MLH from soundings?
- I am surprised that the morning transition is just taken at 11 UTC. In my opinion one strength of lidar observations is that the raise in MLH from sunrise onward can be observed rather well. From the statistics presented in the paper I can not infer how well this is done by MWR. As Fig. 2 illustrates the transition zone (several hours) is most demanding and I wonder how results change if this times are excluded in the analysis.

3. Methodology
- p4979 line 23: “accepts non-unique solutions” sounds strange. Maybe “leads to” ?
- p4980 line 7: “consists of .. state vector includes MLH..”
- Here the aouthors mention 10 min bins while later only hourly values are given.
- The algorithm is trained separately for day and night. Why don’t you show the results also separately in Tab. 2. - especially when considering the large differences in Tab. 1. Alternatively, you could just color the dots for nighttime in blue in Fig. 5 and give the separate results in the lower right corner. They hopefully all lie in the lower left corner.

4. Results
p4983,line 25: MLH itself does not depend on climatology.
p4984,line 12: The authors might want to mention that RS estimates are only representative for one “slanted” profile that might be by pure chance related to an especially strong eddy while the lidar estimates are for temporal averages.

5. Summary
p4985, l23: consistently
- I find it a bit dangerous to talk about seasonal statistics as the considered time interval covers just a few months which might strongly be affected by certain weather types.
Fig. 5: Please enlarge the letters.