I have reviewed the revised version of this manuscript, and especially the author response to reviewer comments. First, I would like to thank the reviewers for a thorough review and several very relevant comments. Second, I find that the authors have responded appropriately to these comments and that this have substantially improved the quality of the paper. Third I would like to apologize for lengthening and already long process; it was unusually difficult to find reviewers for this paper.

I have, however, one remaining issue that I would like to have resolved before I accept this paper for publication. This deals with an – as I believe – insufficient discussion on the key feature that this paper deals with; the atmospheric boundary layer and its depth.

Observing the boundary layer depth – or the height of the boundary-layer top – from space is a very timely issue; having a global climatology of the from space would open up a new chapter in boundary layer meteorology. This is also pursued in Nasa's Decadal Survey Incubation program and the NASA PBL Study Team (see DOI: 10.1175/BAMS-D-23-0228.1). In the light of this it would be important to discuss the fundamental problem: What is a boundary layer and how can its characteristics be estimated from space?

The text does an excellent job of describing the technical challenges with different metrics but it never clarifies these issues, which I believe makes the interpretation difficult. For example, I believe that the attempts by the authors to explain the differences between model and observations for the case studies by heterogeneity and sonde balloon drift are less than convincing, misguided and maybe even misleading. I think that the reasons instead lie in the fact that the authors are comparing apples and pears.

The atmospheric boundary layer by definition is the layer of the lowest of the atmosphere closest to and in direct contact with the Earth's surface, where mixing is maintained by turbulence. This cannot be directly simulated by models and hence not by ERA5; instead it is parameterized. Therefore, in ERA5 this layer is diagnosed from boundary-layer theory using a version of the critical Richardson number, Ric. None of this can be observed, neither from space, nor from surface based lidar and not from radiosondes. Instead different proxies are used; most commonly some kind of mixing concept often involving thermal structure, e.g. identifying inversions in temperature or moisture; sometimes also using aerosols.

In this context it is necessary to realize that just because the thermodynamic profiles suggests mixing has happened doesn't mean it is still ongoing. Both in the context of the residual layer and for decoupled cloud layers, a layer with seemingly well mixed potential temperature may be much deeper than the actual boundary layer as defined using a critical Ric. In such cases the inversion in potential temperature may not be the top of the boundary layer (cf. e.g. DOI:10.1002/2017JD027234) and the definition of it becomes a matter of choice. If the vertical gradient of the wind speed goes to zero at a lower height, Ri > Ric which will indicate a shallower boundary layer than the (main) inversion. Also, aerosols may remain unchanged in a residual layer, whereas in the actual boundary-layer it is affected by deposition, chemistry or clouds.

What I'm looking for here is not a solution to this problem, because there may not be one. I'm asking for an insightful paragraph or maybe just a few lines discussing this, acknowledging that differences between different methods and different instruments and methods may not indicate that one or the other is correct and the other wrong; it may just be that they measure different thing, none of which may be the actual boundary layer.