

*First of all, we would really like to thank the Editor for the dedicated time and for the chance to further improve the scientific quality of the manuscript. In the following text, we have included the response to all comments, together with a description of the corresponding changes made in the revised manuscript. For clarity, the editor's comments are presented with black and responses with **blue**. All line numbers mentioned refer to the uploaded track-changes version of the manuscript.*

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?

We have modified the introduction and included the following text (Lines 31-44), to support the discussion on the boundary layer and the fundamental problem of how its characteristics can be estimated from space:

"Over the open Atlantic, the Marine Atmospheric Boundary Layer (MABL) is typically shallow and influenced by the relatively constant sea surface temperature, while boundary layers in coastal and island regions experience terrestrial-marine interactions that increase their variability (Garratt, 1994; Wood, 2012). A limited number of studies over years have addressed the detection and analysis of MABL using lidar data, primarily due to practical and observational challenges over the ocean(e.g. Atlas et al. 1986; Flamant et al. 1997; Pena et al. 2015). Given these constraints, satellite observations can provide an important means of obtaining information in remote regions lacking in-situ and ground-based remote sensing data, while also enabling the development of global climatologies (Teixeira et al., 2025).

Although the BL is a near-surface phenomenon, several satellite measurements can indirectly infer its properties, particularly its depth and spatial or temporal variability. The Cloud-Aerosol

Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) mission, has been widely used to derive global BL height climatologies over ocean and land and is therefore essential for studying lower troposphere characteristics (e.g. Liu et al. 2024). Nevertheless, when interpreting satellite-derived BL characteristics, it is crucial to decode the measurements appropriately, as the definition and identification of the BL can vary depending on the chosen approach and physical parameter. The Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) of the CALIPSO satellite, can measure, among others, backscattered light from aerosols and clouds. Hence, in this case, the top of the lowest aerosol layer often coincides with the BL top, since aerosols are typically well mixed within the BL and drop sharply above it (Li et al., 2017)."

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.

We agree that the apparent discrepancies between model and observational estimates largely arise because the different approaches represent distinct physical aspects of the boundary layer. While we had noted this point in the manuscript (see below: lines 432–434 and 455–456), we have now expanded the discussion of the first case and the conclusions. We hope that this addition explicitly acknowledges that the different methods may not measure the same quantity, which explains the observed differences without implying that one approach is necessarily more accurate than another.

lines (432-434) "The differences observed across the instruments largely reflect the distinct definitions and retrieval methods used to estimate the BL top, emphasizing that no single dataset provides a complete picture on its own."

lines (455-456): "Hence, the height detected for the BL top needs careful treatment and the interpretation is highly dependent on the definition and methods used. "

We have changed the discussion at the end of the section 3.3.1 (Lines 379-383):

"Upon assessing all the BL results together, we find that the two lidars are in good agreement, consistently capturing the well-mixed aerosol layer. In contrast, the radiosonde indicates the strongest inversion at around 1 km, which is relatively high for a BL in this region and differs significantly from the lidar results. As discussed by Brooks et al. (2017), an apparently well-mixed potential temperature profile may extend into a residual layer where turbulent mixing is no longer active, leading to an overestimation of the actual BL height. Moreover, while lidar detects the top of the aerosol mixing layer, radiosondes diagnose stability changes that may reflect remnants of

earlier mixing. Therefore, these differences between lidar- and radiosonde- derived boundary layer heights can be expected, particularly under conditions of weak turbulence or decoupled layers."

We have also enriched the explanation of how BL top is derived from ECMWF dataset (Lines 141-146).

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, Ri_c . 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 Ri_c . 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, $Ri > Ri_c$ 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.

We have modified the conclusions as following (Lines 444-453):

"It is important to note that differences between the BL heights derived from different instruments/model and methods do not necessarily imply that one is correct and the other is wrong. Rather, they often reflect the fact that each technique responds to a different physical aspect of the boundary layer. For instance, in ERA5, the BL top is not explicitly resolved but diagnosed from boundary-layer theory using a critical Ri_c , representing the depth of active turbulent mixing. Radiosonde-derived heights are typically based on thermodynamic structure, identifying the strongest temperature or humidity inversion, which may correspond to a residual

layer rather than the actively mixed layer. Lidar measurements, in contrast, detect gradients in aerosol backscatter, which trace the extent of aerosol mixing but may remain unchanged even after turbulence ceases. An apparently well-mixed potential temperature profile may thus extend well above the dynamically defined boundary layer (Brooks et al., 2017). Therefore, the discrepancies observed between model, lidars, and radiosonde estimates likely arise because these approaches describe related but not identical layers within the lower atmosphere.”