Response to Reviewer 1

First of all, we would like to thank the reviewer for the valuable, in-depth comments to our manuscript. We appreciate the reviewer took the time to provide exceptionally detailed feedback which we believe has helped to further enhance the quality of this review article. In the following we address all concerns point-by-point.

1. **The atmospheric boundary-layer height (ABLH) definition issue**

- In the first sentence of the abstract, the authors start with a definition which does not appear as general as it could be. This goes back behind Seibert et al. (2000) who, after intense discussion, replaced "pollutants" by "trace substances" (which must not necessarily be pollutants) and allowed heat, moisture or trace substances to be released not only at the Earth's surface but also inside the ABL or even from above (e.g., by entrainment).

Thank you for pointing out these inaccuracies. We simplified the start of the abstract:

"The atmospheric boundary layer (ABL) defines the volume of air for the dilution of heat, moisture and trace substances. Quantitative knowledge on the temporal and spatial variations of the heights of the ABL and its sublayers is still scarce, despite their importance for a series of applications (including, e.g., air quality, numerical weather prediction, greenhouse gas assessment and renewable energy production)."

- In the abstract, the authors implicitly identify the ABLH with the mixing height, although later in the manuscript they interpret it differently (e.g., when they identify the ABLH as the top of the SBL height + RL height).

We removed the word "rapidly" from the first sentence of the abstract to reduce the impression that we consider ABLH to be synonymous to the MLH. Also, we are now explicitly referring to the "heights of the ABL and its sub-layers".

- Common ABLH definitions often introduce a time scale of interaction between the earths surface and the atmosphere. The word "rapidly" (L2) is not very specific in this context. And in section 1 (L103) it appears, that the authors do not want to decide whether to focus on processes with a duration of about one hour or, alternatively, one day – which is a significant difference. It becomes obvious later, that they seem to prefer the latter one.

As mentioned, we removed the word "rapidly" from first sentence of the abstract. We mention the "indirect" response at daily time scales because we consider the residual layer to be part of the ABL. To clarify our understanding of the time scales, we have re-worded the sentence (now L56) to:

"It responds directly to surface forcing at time scales of less than one hour (Garratt, 1994) while indirect effects (e.g. in the residual layer) can extend to daily time scales."

- In the introductory chapter (L62ff) the authors start with introducing the ABL with a poetic description of some of their relevant characteristics. Again here, I would prefer to have a clear definition.

The paragraph (line 55) was re-structured and now reads:

"The ABL is the lowest part of the troposphere where direct interactions with the Earth's surface (land and sea) take place (Seibert et al., 2000). It responds directly to surface forcing at time scales of less than one hour (Garratt, 1994) while indirect effects (e.g. in the residual layer) can extend to daily time scales. Exchange mechanisms include the transfer of momentum, radiation, heat, moisture, particles and gases. The ABL defines the volume in which heat, moisture and trace substances are primarily dispersed following either the release at the surface or some altitude within the ABL or the entrainment from the free troposphere (FT) above. Exchanges with the FT takes place via entrainment and ejection processes (Stull, 1988). Horizontal variations in ABL dynamics stem from a combination of synoptic atmospheric conditions (e.g., atmospheric stability, wind shear, cloud dynamics) and surface forcings (driven by contrasts in e.g., surface cover, roughness, topography) (Garratt, 1994; Seibert et al., 2000)."

2. Section 2 on Measurement principles.

- This section gives a comprehensive overview on the different profiling techniques. It is, however, not well-balanced. While microwave radiometers, Doppler lidars, and ceilometers are described extensively, most of the other techniques are treated quite short. This appears understandable considering the fields of scientific expertise of the authors and also the fact that these three techniques are obviously seen as the most promising candidates for more extensive networks. But then the authors should explain and motivate this from the beginning. Also, one gets the impression that the different subsections were contributed by different co-authors but not adequately harmonized in the final editing process.

Section 2.2 was significantly re-structured to account for your comments. We extracted the general introduction to the lidar technology which is now discussed at the start of the section (lines 150-165) to reduce repetition while keeping the relevant information in the individual subsections.

We also now provide better context to the objectives of this section (lines 183-188):

"In the following sections the emphasis is placed on in-situ platforms and ground-based remote sensing instruments that are to-date commonly used to observe the ABL and can be considered the most promising candidates for extensive measurement 185 network operations (Sect. 2.3). These are radiosoundings for in-situ profiling (Sect. 2.2.1; note that significant advances are expected for network operations of uncrewed areal systems), passive radiometers for temperature profiling (Sect. 2.2.2), Doppler wind lidars for profiling of wind and turbulence (Sect. 2.2.3), and finally automatic lidars and ceilometers for aerosol profiling in the ABL (Sect. 2.2.4)."

- In general, the instruments description could probably be a bit shortened by referencing to the instrumental literature – the recently published "Handbook on Atmospheric Measurements" (Ed. T. Foken, Springer) would be an ideal source of references for this.

Section 2.2 was significantly re-structured to account for your suggestions. Details on instrument descriptions were shortened, especially for the DWL and the aerosol lidars. References to the "Springer Handbook of Atmospheric Measurements" are now included as this is obviously a key textbook in this context.

- I wonder about the so-called "automatic lidars". Does one really need this term? And is it still appropriate today? A Doppler lidar or Vaisala's DIAL, once properly configured, may run automatic as well over weeks, months … (as the authors also state – L531)

The term *automatic lidars and ceilometers* (ALC) describes aerosol lidars that provide continuous observations of attenuated backscatter and cloud-base altitude estimates. We consider "ALC" as an established term that discriminates those automatic systems from research-grade aerosol lidars that use high laser power and can not be operated continuously. We have added the following explanation (lines 445-451) to provide the necessary context:

"Ground-based lidar systems available for the profiling of aerosols differ greatly in laser power and wavelengths utilised (Foken, 2021). It can be generally differentiated between high-power lidar systems and the comparatively low-power automatic lidars and ceilometers (ALC). The latter is a collective term that refers to both ceilometers which traditionally focused on cloud base height estimation and those backscatter lidars primarily designed to continuously provide aerosol profile information (such as micro pulse lidars; MPL). Capabilities and limitations of high-power aerosol lidars have been outlined for the Raman lidar (see description in Sect 2.2.2), a research-grade lidar which is able to sample water vapour and at times temperature, in addition to aerosol properties (Table 2).)"

- The affiliation of the single systems does not appear straightforward in any case. DIAL and Raman lidar are primarily thought to provide humidity and temperature profiles, but they are not discussed under "thermodynamic profiling".

Section 2.2 was significantly re-structured to account for your suggestions. DIAL and Raman lidar are now introduced in the thermodynamic profiling subsection and then referenced in Section 2.2.4. The introduction of the DIAL was slightly extended (lines 270-275):

"A differential absorption lidar (DIAL) transmits laser beams at two wavelengths exploiting the differential attenuation (Lammert and Bösenberg, 2006) to derive vertical profiles of water vapour (Behrendt et al., 2007) or trace gases such as CO2 (Gibert et al., 2008), CH4 (Robinson et al., 2015), ozone (Banta et al., 1998; Ravetta and Ancellet, 1998), or NO2 (Piters et al., 2012). Thanks to recent developments, compact DIAL systems are becoming increasingly available that allow for continuous water vapour profiling of the ABL, using a significantly lower pulse energy compared to the Raman lidar (Newsom et al., 2020)."

- The sequence of presentation does not always appear logical. Section 2.2.1 starts with towers and tethered balloons, but instead of continuing with radiosondes (called the classical standard in-situ profiling technique in the introduction) the authors then switch to aircraft before coming back to the radiosonde. Similarly with the lidar systems: After a quite specific description of the principle of Doppler lidars, section 2.2.4 starts with a very general definition of a lidar system.

We changed the order of presentation. Radiosondes are now discussed following tethered balloons. A general description of lidar systems is now presented earlier (lines 150-165) to reduce repetition in the sub-sections.

- Concerning the radiosondes time resolution one might remark that the synoptic practice at weather services (to launch sondes typically at 00 UT and 12 UT) implies that the representation of typical ABL regimes depends on longitude. While we usually have a "good" SBL / CBL sounding in Central Europe, the standard ascents just represent the transition periods in certain regions of Asia and America. But to brake a lance for the sondes: They provide synchronous profiles of temperature, humidity and wind from just one sensor system, while with remote sensing one needs two or three (expensive) systems for that … The tracking error (mentioned in L261) is not that a big issue anymore since former radar tracking has been replaced by GPS in most networks today.

According to your valuable remarks, we have improved the wording at multiple lines in the manuscript

Line 213: "The main advantages of radiosonde data are: (i) observations of temperature, pressure, humidity, wind speed and direction are collected simultaneously using the same measurement system;…"

Line 220: "While these coordinated launches at synoptic times are required to take the extremely valuable global snapshot of the atmosphere, they generally limit the representation of the ABL diurnal evolution at a given place. Where the launch times occur e.g. during morning growth and/or evening decay of the mixing layer, diurnal minima or maxima may not be captured."

Lines 228-236: "Other specific problems that can result in systematic errors in derived ABL characteristics include humidity sensor uncertainties in cold and dry or cloudy conditions (Seidel et al., 2010;Wang andWang, 2014) and significant horizontal displacement of the balloon during the ascent (Schween et al., 2014). This drift means observations are affected by spatial variations in ABL dynamics which can be challenging for data analysis and interpretation. Some stations operate automatic launch systems that can introduce temperature and humidity uncertainties in the lowest altitudes (< 200 m) as sondes are located in climate-controlled chambers before being released into ambient air (Madonna et al., 2020). Site-dependent radar tracking uncertainties (Seibert et al., 2000) that have caused errors in the wind profiles at low altitudes are no longer a concern as GPS tracking is now used instead. Careful removal of discontinuities induced by changes to the operating system helps to harmonise long-term records (Madonna et al., 2022)."

3. The section on ABL height retrievals contains, to my opinion, a few errors and debatable points:

- The statement about the Ri method in L664f is wrong: Why should the Ri method not be able to provide a layer height estimate in stable conditions? Ri exceeds its critical value somewhere in the stable range (Ri>0) and marks the level where negative buoyancy finally suppresses mixing by shear. In fact, it is seen as an advantage that the Ri method can be applied consistently to convective and stable boundary layers as it is done in many schemes applied to NWP output fields or radiosonde data. Also, the discussion later in this section (L736ff.) is irritating: How can the parcel method "provide good results under a wide range of meteorological conditions" if its application is limited to the CBL? And how can it be superior to the Ri method if the latter one is a generalization of the first one?

Thank you for pointing out this mistake. The statement in L664f of the earlier manuscript version was clearly wrong, as the Ri method provides a SBLH estimation as used in a lot of publications based on radiosonde, ground-based profiling observations and satellite measurements. We have adapted the mention of the bulk-Richardson number method in this section accordingly. Some studies found the parcel method to be more robust for application to long-term data, which is explained by the fact that fewer input data are required. We have amended a paragraph in Section 3.1.2 (lines 711-718) to clarify this point.

"From an analysis using MWR data, Collaud Coen et al. (2014) found the parcel method to be more robust and hence better suited for automatic real-time detection of the CBLH compared to the bulk-Richardson method because the latter requires more input data. In addition to the temperature (and pressure) data, wind profile observations are needed which introduces additional measurement uncertainties and missing values from a second system. Such issues are slightly reduced when the methods are applied to e.g. radiosonde data, as here both wind and temperature are gathered by the same measurement system. Due to the simplicity of the parcel method it is more likely to capture shallow layer heights. Seidel et al. (2010) conclude that diurnal and seasonal variations based on this method generally tend to have a greater amplitude and can be considered more consistent than those derived from other approaches. This is in agreement with the analysis by Collaud Coen et al. (2014)."

- In general, I have the feeling that Ri number based ABLH estimation appears a bit underrepresented in the manuscript, as do other turbulence-based methods. In the conclusions section it does not occur at all. This is probably due to the fact that current schemes applied to single ground-based remote sensing systems do not allow application of this approach. However, in the outlook one could discuss that a synergy of different instruments might allow to use this approach in the future.

In response to some of your other comments, the synergy aspect of the bulk-Richardson method is now presented more clearly in Section 3 and mentioned in the conclusions (line 1369).

- When mentioning the height of the SBI it should be noted that this might be accepted as a stability-based criterion, however when asking for the processes behind the derived values it might suffer from ambiguity since the upper part of a surface inversion is often governed from radiative cooling alone while turbulent mixing is present in the lower part only.

To clarify the physical interpretation of the SBIH, we added the following sentence to Section 1.1 (Line 83):

"While vertical mixing mainly occurs in the lower levels of the temperature inversion, a combination of other processes such as radiative cooling, subsidence or horizontal advection shapes the depth and the magnitude of the *SBI*."

- “The vanishing gradient in the air temperature profile marks the base of an elevated inversion” – I do not understand this: At the base of an elevated inversion the temperature gradient should become positive. And mixing might already be suppressed in a zone of constant or even slightly decreasing-with-height air temperature, if the potential temperature gradient becomes positive.

For clarification, we changed the first item in the list of MBLH detection methods (line 668ff) to

“To derive the MBLH from temperature profiles, both the base of an elevated temperature inversion (Seidel et al., 2010) and the height of the maximum positive θ gradient (Stull, 1988; Seidel et al., 2010) have been applied. As vertical mixing can already be reduced for a certain layer below a positive vertical gradient in air temperature, the criterion based on potential temperature can be more suitable to assess the limitations of vertical mixing. However, to calculate \theta, profiles data of atmospheric pressure are required which may not always be available.”

- It is not the refractive index, but the refractive index structure parameter which is the primarily observed variable by sodar and RWP (L774/775). Later (L810ff) follows a redundant paragraph just replacing Cn2 by SNR (which is somehow equivalent), these two pieces of text might be put together.

Thank you for pointing this out. The paragraph (now line 747ff) has been changed to

“Turbulence can be diagnosed from the refractive index structure parameter (Sect. 2.1) observed by sodar and RWP (Sect. 2.2.3). The peak in the vertical profile of the refractive index structure parameter caused by small-scale buoyancy fluctuations across the entrainment zone has been found to coincide with the MBLH (White, 1993; Angevine et al., 1994; Wilczak et al., 1997). Given these fluctuations are associated with relatively high SNR in sodar and RWP observations, some methods assign CBLH to a local peak in RWP SNR (Liu et al., 2019a; Collaud Coen et al., 2014)..”

- Section 3.3 starts with a very positive assumption – can it be really taken as given that the distribution of aerosol and moisture is in parts a result of previous mixing processes? I would more generally say that the “distribution of … can be a result of either mixing, transport / advection, accumulation or formation processes or any combination of these”

To clarify that the distribution of atmospheric constituents results from a combination of many processes, the sentence has been rephrased (now line 815ff) taking into account your suggestions.

“The distribution of aerosols and moisture usually results from a complex combination of processes, including emission, formation, accumulation, deposition, transport (advection), but also mixing. Profiles of attenuated backscatter (Sect. 2.1) hence trace some aspects of the recent history of ABL dynamics.”

- To get a quick overview on the capabilities of the different systems with respect to range (lower and upper limits) and height resolution, another Table would be very helpful (or additional columns in Table 2)

Given the diversity in sensor models, it can be tricky to put general numbers to these instrument capabilities. For example, the respective near-range capability and signal strength in greater altitudes of DWL and ALC depends significantly on the laser power and instrument specifics. Also for RWP, a range of systems are available with different range capabilities. Obviously, for research-grade lidars, the strength lies more in the greater altitudes while the opposite is true for sodar systems. And also in terms of range resolution, the answer to this question is not easy to put in general terms because recent manufacturers start to increase the range resolution of the observed products by applying oversampling techniques in the firmware. We decided not to include this information in Table 2 because for most systems it would indicate that the min and max range covered and also vertical resolution depend on the specifics of the instrument model. Still, we now highlight in more general terms that the near-range capabilities and maximum range covered are usually connected for any lidar system (lines 158ff).

"In general, there is a relation between the near-range and far-range capabilities of lidar systems. While high-power systems have a monitoring range of several kilometers (some reaching the stratosphere), they require an increasingly large telescope which again increases the blind-zone near the sensor. Also the vertical resolution of the recorded profile used to increase with laser power and vertical range extent, however, manufacturers increasingly apply oversampling procedures to the data products whereby increasing the number of range gates."

We also highlight the aspect of range capabilities in the conclusion section.

4. The sensor synergy section provides a nice summary and synthesis, a few possible additions might be considered

- When discussion the potential of sensor synergies (L1036f) it might be mentioned, in addition, that the availability of observations from several sensor systems could be used to assess and quantify the uncertainty of any ABLH retrieval and to add a quality flag to the derived values (as has been done for the different retrievals from radiosonde profiles by Beyrich and Leps, 2012).

Thank you for this valuable suggestion to mention uncertainty assessment which is obviously an important objective of the ABL ground-based remote sensing community. We added the following sentence (line 1023ff) and also mention this synergy application in the conclusions.

"In addition, combining observations from several sensor systems provides crucial information for the assessment and quantification of layer detection uncertainties (e.g., Cohen et al., 2015) and could be exploited to assign quality flags derived layer heights (as has been done for the different retrievals from radiosonde profiles by Beyrich and Leps, 2012).

- In Section 4.2 on the morning growth I miss the sodar which has been shown to be an ideal instrument to observe the CBL growth (at least in the early stages, up to a few hundreds of meters, depending on the instruments characteristics).

To emphasize the importance of sodar observations for shallow layer conditions and the CBL growth, we added the following sentence (line 1097):

"Given their ability to capture turbulence even in the near range (Sect.2.2), sodar systems are particularly valuable for the monitoring of the growth (onset) of the CBLH (Beyrich, 1995)."

- The statement "radiosonde ascents are rare between sunrise and solar noon" (L1098) should not be made that general, first this depends much on the geographical longitude, and second the timing of the morning growth phase varies with the season. The more severe problem is the fact that radiosoundings just provide a snapshot-like picture of ABL structure which is a serious limitation during phases of strong non-stationarity, as the morning growth.

Thank you for pointing out this inaccurate statement. We rephrased the beginning of this paragraph to (line 1090ff):

"The time of CBLH morning growth is characterised by substantial temporal variations (Halios and Barlow, 2017), especially where the surface energy balance exhibits a pronounced diurnal cycle. Compared to the limited temporal coverage of radiosonde ascends (Sect. 2.2.1), the continuous monitoring enabled by remote sensing profilers is a clear advantage for the assessment of this period characterised by significant seasonal but also day-to-day variability."

- Another too strong statement (L1107): “Analysis of RWP data allows for the detection of CBL morning growth” – this at least depends on the type of RWP. UHF profilers at around 500 MHz definitely miss the early stages of CBL growth due to a lowest measurement height at around 500 m.

To clarify this point, the sentence (now line 1104ff) was changed to

“The detection of CBLH from RWP data during morning growth relies on the careful differentiation between the turbulent signature generated by the entrainment of RL air into the CBL and variations near the RLH (Bianco et al., 2022).”

5. Section 5 on ABL climatology goes, to my opinion, a bit beyond the scope of the manuscript. It illustrates what could be achieved once we will have available data sets from ground-based remote sensing instruments on ABLH over climatological time scales (30 years +). However, current discussion basically relies on either radiosonde and model data or on measurements over shorter time scales. The text in this section is mainly descriptive, touching many aspects without going into depth, and leaving questions open. Given that the manuscript has already a considerable length I would recommend to consider omitting this chapter and just discussing the climatology perspectives (with mentioning the different aspects covered by the subsections) in the conclusions section.

We understand that section 5 can be regarded to go beyond the scope of the current publication. Also to shorten the manuscript and make it more accessible to the reader, we are following your suggestion to remove the chapter and all its mentions in the abstract and other sections.

The Figures and Tables

- Figure 2 appears too simple and partially confusing:
  - What is the relevance of the different colors – this should be explained.
  - The unspecified “exchange” symbols only occur at the top of the growing CBL in the morning – so no exchange between the Earth’s surface and the atmosphere, no exchange between the fully developed CBL and the FT?
  - Solar irradiance only occurs in the morning, but not later in the day, and radiative cooling only occurs around sunset, but not later during the night?
  - Why does wind-shear turbulence occur in the RL only, but not in the SBL (where it is usually the only mechanism to maintain some turbulence) or at the transition between the ABL and the FT?
  - The EZ symbol should be repeated at the top of the fully developed CBL.
  - I consider it for unusual that the development of the CBL starts before sunrise.

Thank you for your valuable and detailed comments. Figure 2 was updated to address the concerns listed.

- Also, Figure 3 is not in all aspects a successful illustration of typical ABL variables profiles
  - The vertical dashed lines are not explained, I assume that they should mark the geostrophic wind speed. If so, one may question why it is constant with height in the upper and lower row, but decreases with height in the middle row, this is an unnecessary complication.
  - The upper panel represents an early-morning CBL. Figure 2 suggests a thick EZ at this stage (which shall probably be illustrated by the three different colors along the z-axis representing the CBL <but this is not explained>). However, this does not go together with the well-mixed ⍰v profile across these three sub-regions. If ⍰v appears well-mixed, the sharp drop in tracer concentration at half the CBL height is

not straight-forward to explain, nor is the large zone of supergeostrophic winds across most of the CBL+RL. And the ⍰w profile close to the surface appears to be non-typical for a (even shallow) CBL.
- In the SBL (lower panel) it is suggested that the SIH and the height of the LLJ axis coincide, but this is not necessarily the case … And since w is zero at the surface, sigmaw cannot have a maximum there.
- The horizontal yellow dashed line and shading is said to mark the ABLH. I feel this to be a bit irritating – it suggests that the uncertainty (variability?) in ABLH (illustrated by the shading?) is almost independent of the state of the ABL. Moreover, the behavior of the different variables at this mean ABLH appears a bit inconsistent: If ⍰v shows a very sharp increase just at the mean (?) ABLH, why should c then be much more smoothly distributed? And why should c show a smooth transition on the top of the ABLH, but a sharp transition at the top of the CBLH when entrainment is expected to mix the profiles to some extent?

Thank you for your valuable and detailed comments. Figure 3 was updated to address the concerns listed and merged with Figure 2 to make the link more obvious.

- Figure 6 does not really show results within the scope of the manuscript (ABLH from ground-based remote sensing), panels a) to c) are not marked in the Figure, even if one can easily deduce what is meant here.

As Section 5 has been removed from the manuscript, we also removed Figure 6.

- Table 1 could be optimized a bit, to my opinion:
- I suggest to differentiate optically between atmospheric and so-called measurement variables either by sorting these types in two different columns or by printing the measurement variables in, e.g., Italics.
- Variables representing basically the same characteristic might be grouped together, e.g., the different temperature and humidity variables, wind speed and the wind vector components etc.
- I suggest to replace "dynamic processes" by "dynamic and turbulent processes"
- What about turbulent fluxes or combined indices, like Richardson numbers, which may be used to detect the ABL height as well.

Table 1 was adapted according to your suggestions. Of course many additional variables can be calculated form these variables. We have added the following sentence to the table caption to clarify this point this:

"Based on the variables listed here, other higher-order variables or parameters can be calculated (such as turbulent fluxes or Richardson numbers) that are valuable for characterising the ABL, especially where observations from multiple systems are available for synergy applications."

- Table 2 suffers from some omissions
- DIAL is not primarily used to measure temperature (also not mentioned in the text)
- Raman lidar may also used to measure temperature, and – of course – all the aerosol parameters listed for an aerosol lidar
- RWP basic measurement variables are missing: Cn2, vr, also TKE could be derived
- Sodar can also measure vr, u, v, w, sigmaw,
- The list of network operations might call for a list of acronyms.

We have updated Table 2 according to your comments.

Minor Issues

- L74: The poor time resolution of (operational) radiosondes is not the only disadvantage of this technique.

Here in the introduction (line 24ff) we added the following sentence. Further details on the limitations of radiosonde profiling are listed in Section 2.2.

“Sampling the ABL vertical profile has historically been mostly achieved using radiosondes. While these balloon ascends provide indispensable information, their temporal resolution is usually insufficient to capture the full diurnal evolution of the ABL dynamics and the significant horizontal drift of the balloon during the ascent means observations are affected by spatial variations in ABL dynamics which can be challenging for data analysis and interpretation.”

- L78: It would be good to give some rough numbers on the “entire ABL vertical extent” already here for orientation.

Changed to (line 29ff)

“Significant advances in ground-based remote sensing measurement technology and algorithm development now allow for continuous profiling of the entire ABL vertical extent (ranging from a few tens of metres to >3 km, or even higher, depending on geographic settings and synoptic conditions) at high temporal and vertical resolution (Illingworth et al., 2019; Cimini et al., 2020) and automatic detection of ABL sublayer heights from different atmospheric quantities (Collaud Coen et al., 2014; Duncan et al., 2021).”

- L123: It is not only the height of a surface inversion which is meaningful for vertical dilution processes but the internal structure of that inversion as well.

In response to this and your other comments on the topic of SBI, this now reads (line 81ff):

“The vertical profile of air temperature in the SBL often shows a characteristic surface-based temperature inversion (SBI), whose height (SBIH) can be very meaningful in restricting vertical dilution. While vertical mixing mainly occurs in the lower levels of the temperature inversion, a combination of potential other processes such as radiative cooling, subsidence or horizontal advection shapes the depth and the magnitude of the SBI. “

- L133-134: What about the role of wind shear for the exchange between the ABL and FT?

Beginning of the paragraph (line 93ff) changed to

“Exchange between the CBL and the FT (or the RL) occurs via the penetration of the CBL thermals into the air aloft and the entrainment of warm and (in the absence of clouds) dry air into the CBL. As horizontal wind speeds are usually lower in the CBL compared to the FT or RL (Figure 2), wind shear at the CBLH further generates mechanical turbulence that contributes to the entrainment.”

- L138: Shouldn’t that be “RL or the CBL” instead of “RL or the SBL”?

For clarification, this now reads (line 99ff):

“The ABL transition to the F T is marked by a strongly positive temperature lapse rate, the capping inversion (CI). EZD is greater when the temperature difference between ABL and F T is weak (AMS, 2017). The CI often coincides with a sharp vertical decrease in specific humidity and significant vertical wind shear (Figure 2).”

- L142: “may be located” – this sounds rather vague. If Cu clouds form in the ABL, isn’t the ABLH than normally above the CBH?

Changed to “… is located …”

- L191: The link to the Rn measurements either calls for some additional explanation or it should be omitted here.

Sorry, we can not locate the sentence to which this comment is referring to.

- L238: Wind speed and wind direction from radiosondes haven't been measured with sensors lifted by the balloon for decades, they are derived from the replacement of the balloon tracked by radar or GPS.

The sentence (now line 203ff) has been rephrased to:

"In-situ measurements of air temperature and humidity are taken by sensors that are being lifted up by an helium-inflated aerostatic balloon while atmospheric pressure, wind speed and direction are derived along the flight path via satellite tracking (e.g. GPS)."

- L266ff: This description of a RASS is a bit too simplified and not fully correct. First there are two types of RASS – a radar wind profiler + an acoustic source and, alternatively, a sodar + an electromagnetic source. Second, with the RWP as wind measurement system, the acoustic component is normally not operated as a sodar but just provides the sound signal detected by the electromagnetic component. Furthermore, with RWP, the sound source is normally not operated in pulse mode. Also the given range limitation for RASS temperature measurements (L273) are not true in a general sense, but depend on the system characteristics: While this limit applies to 1-GHz RWP with RASS capability and to sodar-based RASS, a RASS using an RWP at around 500 MHz allows temperature measurements well above 1 km

Thank you for pointing out these inaccuracies. We have adapted the RASS description accordingly (lines 276ff).

"RASS systems either combine a radar wind profiler with a source of acoustic signals (e.g., sodar) or a sodar system with a source of electromagnetic signals (Emeis, 2010; Foken, 2021). From the Doppler shift of the respective returned signal the speed of sound is measured as a function of altitude, from which the profile of virtual temperature can be deduced. The uncertainty in temperature can be < 0.5 K, provided a number of careful corrections are applied (Görsdorf and Lehmann, 2000). Temporal resolution depends on the application with 10 minutes averaging being typical. The vertical resolution of the profile depends on the length of the pulse transmitted, with RASS systems usually configured to have a resolution of 30-60 m. As for many ground-based remote sensing instrument types, the capabilities to capture information in the near-range or greater altitude, respectively, depends on the specific RASS system characteristics. While sodar-based RASS or 1-GHz radar wind profilers with RASS capability reach their maximum range at about 500 m, measurements well above 1 km can be obtained with RASS systems using a radar wind profiler at about 500 MHz."

- L270: Görsdorf and Lehmann (2000) achieve this accuracy by applying a number of careful corrections, the authors might wish to mention this.

Thank you for pointing out this important detail. The sentence (line 278) was changed to

"The uncertainty in temperature can be < 0.5 K, provided a number of careful corrections are applied (Görsdorf and Lehmann, 2000)."

- L288f (and others): I would recommend to be cautious with the term "root mean square error" (RMSE) which implies that the truth is known – a difficult issue for measurements in the real atmosphere with different instruments relying on different measurement principles, it could be better to use the term "root mean square difference" (RMSD) instead.

We agree with your assessment and replaced the term "RMSE" by "RMSD".

- L340ff: The discussion of rmsd and bias values between different instruments is just half of the truth since the determination of the ABLH height in many situations is based on gradients rather than on absolute values.

We added the following statement (line 189ff) to highlight this important aspect:

"During the discussion of respective sensor capabilities, it is obviously of interest to assess the agreement of observations obtained from different sensors in terms of absolute values. However, it should be kept in mind that layer height retrievals methods (Sect. 3) tend to exploit relative changes (such as vertical gradients) which means aspects such as sensor response time of in-situ measurements, or vertical resolution are generally also critical to consider."

- L359: For ABLH determination from sodar measurements see also Beyrich (1997).

Reference included.

- L357ff (sodars): The possibility to measure wind and turbulence (at least ⍰w) with sodars is completely missing in the discussion here (see also Table 2). It might be confusing for the reader to read here that a sodar primarily measures CT2, while in Table 2 Cn2 is listed. Concerning the limitations of sodars, besides the emission of sound, at least equally relevant is the sensitivity to environmental noise.

Paragraph adapted accordingly (lines 350ff).

- L451-455: This reads if DWL would have no operating limitations ("can operate under all weather conditions"). First a certain aerosol load is necessary to create backscatter. Second, rain typically distorts the measurements and makes vr derivation difficult or impossible. Third, DWL does not provide profile information much above the base of low clouds – a severe limitation in winter in many regions.

The paragraph has been reworded to more accurately present the limitations of the DWL observations (lines 431ff).

- L521: Isn't that a bit low? – ECMWF data coverage survey typically gives 500-700 temp observations at 00 UT.

This was a mistake. Now corrected to ~800 stations that do launches at least once but mostly twice per day.

- L546: None of the 128 RWPs in China is visible in Figure 4c.

This is correct. We have asked several agencies to provide information on measurement locations of networks that are not shown on this Figure. In the case of the China networks, we received the response that the station locations are confidential. To clarify that the shown illustration is only a subset of the true number of global measurement sites, we added the following sentence to the caption of Figure 3 (formally Figure 4):

"Note that this is by no means a comprehensive representation of all profiling instruments being operated. Additional networks do exist but meta data, such as station locations are not always be easily accessible."

- L618ff: I would not say, that "radiosondes may observe spatial variations in the ABL dynamics" this would require measurements at the same height at different places (as with a scanning lidar). I would instead say that "Vertical profiles derived from radiosondes may be influenced by spatial variations in the ABL dynamics and do not necessarily represent the ABL structure just above the launch site.". In addition to surface heterogeneities and the synoptic flow, also orography may play a role here.

We rephrased the sentence according to your suggestion (line 575).

- L658: Can Ri number methods really be grouped as temperature-based approaches to determine the ABL height? The wind profile significantly affects the Ri based retrievals.

In fact, we agree it is not easy to find the appropriate category for the discussion of the Richardson method as it is obviously using the synergy from multiple variables. We decided to include it into the temperature-based methods given its close connection to the parcel method. We have amended the section in several places to highlight that it is a synergy approach and that wind shear is of high relevance for this layer height retrieval.

Line 545ff: “Certain approaches (such as the bulk-Richardson method here described in Sect. 3.1) are in fact not only associated with only one of those categories as they exploit a combination of atmospheric variables. While some measurement systems capture several variables simultaneously (e.g. radiosondes), the synergy between measurements from different ground-based remote sensing profilers (e.g. combining the temperature profile from MWR and wind profile from DWL) is a promising approach as it allows for these multi-variable parameters to be calculated.”

Line 625ff:” While the parcel method is only applicable under unstable atmospheric conditions, the bulk-Richardson method takes into account the implications of wind shear contribution to turbulence generation. It is applicable under all stability regimes. The bulk-Richardson number Rib represents the ratio of turbulence induced by thermal buoyancy and wind shear, respectively, and profiles of both temperature and horizontal wind are required to calculate Rib. It is essentially a synergy approach that combines thermodynamic and dynamic effects and could as well be grouped into dynamic retrieval methods (Sect. 3.2).”

Line 739ff: “Using mean wind profiles (Figure 2), the most commonly used layer detection approach is the bulk-Richardson method. Looking at the relation between thermally-induced buoyancy and shear-induced turbulence, this synergy method (see discussion in Sect. 3.1) is applicable under all stability conditions (i.e. for detection of both SBLH and CBLH).”

- L732: One may add that these retrievals are typically most uncertain during the evening transition period.

We added the following sentence (line 699):

“Uncertainties are usually greatest during the evening decay (Sect. 4.4) of the CBL as this is a time of significant change.”

- Section 3.1.2: The uncertainty analysis and quality flags provided by the scheme suggested in Beyrich and Leps (2012) might be mentioned in this subsection as a possibility to characterize the limitations of these (thermodynamic) methods.

Thank you for pointing out this valuable approach. We have included the following paragraphs in Section 3 and the conclusions:

Line 601: “Beyrich and Leps (2012) developed a scheme that utilises the agreement between different methods (in their case thermodynamic and wind-based detection applied to radiosonde profiles) to quantify the uncertainty of the layer heights at a given moment and assign quality flags accordingly. This is a promising approach that could be further extended where data from multiple systems are available simultaneously which allow for a range of detection methods based on different atmospheric variables to be applied in synergy.”

- L785: What are “low atmospheric waves”? Do they occur at low altitudes or do they have low frequencies?

Changed to “gravity waves”.

- L816: For the sodar – RWP
- sensor synergy the paper by Beyrich and Görsdorf (Boundary-Layer Meteorol. 76, 1995, 387-394) would be a nice reference.

Indeed very suitable. Reference included.

- L836: It could be made a bit clearer that the effect of rain is twofold here: First the measurements of w or radial winds can be corrupted by contributions from the fall speed of rain drops. On the other side, rain on the antenna (e.g., of a sodar) or telescope (e.g., of a DL) might generally prevent a useful retrieval.

Implications of precipitation on the measurement uncertainty in sodar and DWL observations are now highlighted more clearly in Section 2.2.

- L995: "… passage of synoptic fronts." – One may remark that the ABL is generally poorly defined under these conditions (the "gust weather" or "dissolving" ABL type according to Schneider-Carius, 1953) which are dominated by larger scale processes such that the surface – atmosphere interaction is of minor relevance.

Thank you for pointing this out. We have rephrased the sentence accordingly (line 976):

"Most methods naturally struggle with reliable layer detection and attribution during precipitation or the passage of synoptic fronts (de Bruine et al., 2017; Yang et al., 2017) as the ABL is generally poorly defined during those times when surface-atmosphere interaction splay a minor role compared to larger-scale processes."

- L1127: "… agree within a few hundred metres …" add "… or even less" – A few hundred metres would still be a considerable uncertainty, but values of 50 m – 150 m are possible for a well-developed CBL with a clear capping inversion (see also line 1514f).

Changed according to suggestion.

- L1172f: Following the definition of the EZ, the determination of its thickness from temporal or spatial variations (spatial variations underly the original water tank studies analysed be Deardorff in the 1970ies) of CBLH would be the preferable way – gradients of mean quantities are a secondary approximation.

We rephrased the sentence to (lines 1170ff):

"The EZD can be estimated either exploiting temporal (or spatial) variations in CBLH (Sect. 3), providing a direct measure of the entrainment process. Alternatively, it can be approximated based on gradients of mean observed quantities between the CBL and FT (RL)."

- L1262-1263: I am not sure about this generalization – First: What is "high latitudes", second the cited paper refers to the marine ABL concerning the large seasonality over high latitudes: "**Marine PBL depth** has a somewhat similar seasonal march to that over continents but is weaker in amplitude and is shifted poleward. The **maximum seasonal amplitude over oceans** occurs over the Arctic."

This comment is no longer relevant as section 5 has been removed from the manuscript.

- L1290f: Cloud dynamics certainly affect CBL growth rates, however, other variables are of relevance as well, such as moisture content of the ABL, lapse rate in the FT above the ABL, (downward) synoptic-scale vertical motions …

This comment is no longer relevant as section 5 has been removed from the manuscript.

- L1348ff: many of these studies cannot really called "long-term climatological observations".

This comment is no longer relevant as section 5 has been removed from the manuscript.

- L1463f: Lower stability would imply an increasing vertical temperature gradient, wouldn't it?

This comment is no longer relevant as section 5 has been removed from the manuscript.

- Concerning the cited literature I may additionally direct the authors attention to a paper by Zhang et al. (2020): Research progress on estimation of the atmospheric boundary layer height. J. Meteorol. Res. 34, 482-498, doi: 10.1007/s13351-020-9910-3

Thank you for pointing this out. We included the reference in the introduction (line 36):

"In their recent review Zhang et al. (2020) stress that interpretation of ABL height data should always take into account the specifics of both the retrieval algorithm (e.g. which atmospheric variable is analysed?) and the input data (e.g. characteristics of the sensor used for data acquisition)."

Some language and technical issues

- L138: Probably better replace “gradient” by “difference” here.

Changed according to suggestion.

- L168: insights on - insights into

Changed according to suggestion.

- L170: “tracer variables may portray” (they do not necessarily …)

Changed according to suggestion.

- L190: record- recorded

Changed according to suggestion.

- L191: eventually replace “atmosphere” by “atmospheric volume”

Changed according to suggestion.

- L251: add “worldwide” to the “177 sites”.

Changed according to suggestion.

- L267: “The frequency … is varied”

Changed according to suggestion.

- L491: “wide range of models”

Changed according to suggestion.

- L502: “ABL layers” is doubling the “layers”

Changed to “ABL sub-layers”.

- L507: delete “of” after operators

Changed according to suggestion.

- L536-37: The official name of DWD in English is: German Meteorological Service.

Changed accordingly.

- L547: Close the brackets after JMA

Changed accordingly.

- L547: covering - utililizing (or: working at)

Changed according to suggestion.

- L555: “and they are the major …”

Changed according to suggestion.

- L593: “at higher altitudes”

Changed according to suggestion.

- L594: “… first range gates”

Changed accordingly.

- L658: Vogelezang and Holtslag were definitely not the first that introduced the Ri methods, thus write “e. g., “ or introduce primary reference.

Included the following additional references: “… bulk-Richardson method (Hanna, 1969; Vogelezang and Holtslag, 1996; Zilitinkevich and Baklanov, 2002).”

- L662f: better write “… parcel methods that additionally considers the shear contribution to turbulence generation.”

Changed accordingly.

- L695: Who is “L.”?

Formatting error was corrected.

- L774: “… can be diagnosed from the refractive index structure parameter”

Changed accordingly.

- L796: “scanning-strategy”

Changed accordingly.

- L822: What is “random forest machine learning”?

To simplify the sentence, we deleted “random forest” as it seems not relevant to explain the specifics of the ML approach.

- L877: Add a comma after “backscatter” to make this better readable.

Separated into two sentences.

- L918: Tennekes is only once the author of that paper.

Formatting error corrected.

- L947: Cohn and Angevine were definitely not the first to estimate EZ thickness, thus write “e. g., “ or introduce primary reference.

Added “e.g.”

- L1169: add “… (accurately) by ALC”

Changed according to suggestion.

- L1177: thickness

Changed accordingly.

- L1283: add “h” after the times given (or write 1200 LT etc. instead)

This sentence has been removed.

- L1445: daily -- daytime?

This sentence has been removed.

- L1500: turbulence-based

Changed accordingly.

- L1536: add “to” after “expected”

Changed accordingly.

Response to Reviewer 2

We thank the reviewer for the positive response to our manuscript and the valuable feedback.

Below, we respond to all comments point-by-point.

The paper "Atmospheric boundary layer height from ground-based remote sensing: a review of capabilities and limitations" by Kotthaus et al., provides a summary of state-of-art boundary layer height estimates using ground based remote sensing systems. Such reviews are important as there has been significant development in novel techniques and instruments, and many papers go un-noticed unless highlighted by such a review. The authors have done a really nice job going into details of PBL definitions (some clarifications needed), type of instruments, new algorithmic developments and current deficiencies.

The reviewer does have some additional clarifications, and comments which would be helpful if addressed in the revisions.

Major Comments:

1. Definition of ABLH is mostly provided from a lens of onshore applications or land-based ABLH, how do these definitions hold up in an offshore marine boundary layer? There needs to be some discussion on that front in the initial sections (Section 1.1).

Given the climatology section 5 has been removed from the manuscript, the introduction was also shorted to remove most aspects related geographic and land-cover specific ABL dynamics.

2. In Figure 3, it is indicated that the vertical velocity variance estimates can be used to estimate the stable boundary layer depth, the reviewer is not aware of many papers discussing/showing that methodology (except for Pichiguina and Banta 2010). So not a widely accepted methodology, as the vertical velocity variance is very low during nighttime conditions. Please provide appropriate citation for other references or reconsider this statement.

As you mention correctly, the vertical profile of the vertical velocity variance shown in Figure 3c was incorrect. This was also pointed out by Reviewer 1. We have adapted the figure accordingly.

3. The latest COSMIC-2 satellites have better spatial and temporal resolution and can provide better boundary layer height estimates compared to legacy COSMIC/other RO satellite data. Please mention something about COSMIC-2 to the reader, I don't see that in the article.

We added the following sentences (line 172ff):

"Following the success of COSMIC, the promising COSMIC-2 mission was launched in 2019 to provide radio occultation data at even higher through deeper tropospheric penetration (50% within 200 m of Earth's surface). These observation enable improved detection of the ABLH and superrefraction at the top of the ABL (Ho et al.,2020; Schreiner et al., 2020)."

4. Table 2, Network operation, the Atmospheric Radiation Measurement (ARM) network is missing in the list.

Table 2 has been updated.

5. What is the importance of knowing ABLH during daytime or nighttime transition periods? That needs to be discussed, as models tend to deviate significantly during those time periods. Do we expect a given instrument to perform well during those transition periods?

Both morning growth and evening decay are discussed in Section 4 regarding the monitoring of the diurnal cycle of the ABL. We have added the following sentences to the conclusions (line 1318):

"The morning growth and evening decay of the CBL also poses challenges to numerical simulations for a range of applications, including air quality, greenhouse gas assessment and numerical weather prediction. When using observations for model evaluation or comparisons, it is crucial to carefully consider the specific uncertainties of the respective measurement used. Also, it is important to understand which atmospheric variable is used for layer detection, as it can introduce systematic biases if e.g. turbulence-derived layer heights are compared to results exploiting aerosol profiles."

6. For DWL, another issue is the power of a given scanning Doppler lidar to reach the boundary layer. Some low powered scanning DLs fall short of reaching the ABLH during convective conditions, due to attenuation of the signal, increased Cn2 effects, instrument noise, etc.

A more general description of lidar systems and their capabilities is now presented in lines 151-166, which highlights this point more clearly. It is also repeated in the DWL section and in the conclusions.

7. The NY Mesonet network in the US is missing: http://www.nysmesonet.org/data/profiler#stid=prof_alba. They have DLW and MWR profilers.

Thank you for pointing this out, The NYS Mesonet MWR are now shown on Figure 3 and mentioned in Table 2.

8. Offshore ABLH should be given a separate section here, as there are challenges in measuring them due to trapped aerosol layers, internal boundary layers, coastal effects etc.

Section 5 has been deleted from the manuscript.

9. ABL climatology sections seems out of place here. Once you define these advantages and disadvantages, the climatology will be digested with a "grain of salt" by the reader. Unless you put some uncertainty plots etc. I would encourage the authors to reconsider this section. Perhaps another climatology article would be most appropriate. This would also reduce the length of the article.

We understand that section 5 may go beyond the scope of the current publication. Also to shorten the manuscript and make it more accessible to the reader, we are following your suggestion to remove the chapter and all its mentions in the abstract and other sections.

Minor Comment

1. I am not a big fan of Table of contents for a journal paper but will leave it to the editor to decide. Figure 1 encapsulates this nicely.

We agree with your comment and removed the Table of contents for the final publication.