Reply to Referee #1

We deeply thank Referee #1 for her/his constructive review which has been very fruitful to improve the manuscript.

We have significantly revised the manuscript in response.

Below is an answer point by point to the review.

Main comments :

- The introduction should emphasize more clearly why previous methods from the literature are insufficient and how this new method is addressing the shortcomings. This is done for one specific approach Angevine et al. (1994) but is this really the only one available so for application to RWP data?
  Indeed, we realized that the introduction was too poor in documenting previous works made on Zi retrieval based on RWPs. We have largely enriched it.

- In section 2.1 it should be stated more explicitly which auxiliary data are being used and how. Try to make it clear to a reader who may be interested in repeating this work elsewhere, what type of observations are required in addition to the RWP.

  We have revised the initial Table 1 in order to make the use of the dataset clearer.
  In particular, we changed the directing point of view, and first considered the context (P2OA multi-instrumented site, BLLAST and LIAISE), then the RWP profilers. From there, we considered the auxiliary instrumentation, that was used in the different steps of this study. We specified this use in the table (see Table 1 page 6).

  We also have drawn another table later in the manuscript (Table 3 in the revised manuscript, page 8), with the list of variables needed at the different step. Some of them being optional. The text then gives more details and definitions on those, but at least, the table clearly summaries this information.

- In section 2, separate the description of the different steps: start by introducing data acquisition and pre-processing (e.g. filtering, quality control, averaging), then introduce the calculation of new parameters.
  We have made those aspects clearer in the revised version.
  A first part introduces the measured variables, then we address the data process, and later on, we deal with the data averaging and the calculation of new parameters.

- It is rather unusual to introduce result figures in the methods section. Maybe consider referencing figures that are discussed later but focus on the description of the data in section 2

Yes, that is true. We initially thought that it would help understanding the algorithm, but it actually brought some confusion, some complexity in the paper organisation, and also some redundancy in the manuscript.
We therefore moved those figures to later in the manuscript, in a section dealing with the 3 illustrative examples, all gathered together only after the algorithm was described.
We also decided to separate the initial « Results » section in two new sections : one with those three illustrative examples, and the other with the statistic evaluation of the algorithm with the radiosoundings. (sections 4 and 5)

- Try to organise section 3, maybe using a table (?) that gives an overview on the input variables to the algorithm with a short comment of the ABL feature they respond to (e.g. mixing at CBL top or rather RL height, etc) and then also a table to the parameters and thresholds, including the definition of tinit with auxiliary observations.
We have made this clearer with the new Table 3 mentioned before (page 8).

- It should be clearly visible how other data such as surface humidity or sensible heat flux are being used and if they are not available at a different sites, what would be the implication of working without such information?
- When the 2 m relative humidity measurements are missing : If there is not such measurement, the algorithm does not detect fog, and may gives Zi estimates during those events. In this case, the estimate is likely to be unaccurate, with attribution of Zi at the top of the fog cloud.
- When the surface flux measurements are missing : tinit is determined from only Cn2 at the third gate or from Ziepsi. Considering surface flux measurements slightly increases the number of Zi estimates (3%), with a few earlier estimates which might not be appropriate, but also a gain in the detection of Zi in certain specific days. The change remains very small, as quantified in the text. (page 13, lines 292-293)
We have specified this better in the final discussion of the revised version (page 25, lines 522 to 526)

- Also when writing, try to reference future sections when appropriate, i.e. at times threshold values are introduced and then explained at a later stage but the reader might not know that further information will be provided later.
We have checked this aspect, and the new organisation actually improved a lot this issue.

- Regarding the flag system in section 3.3.3: this is a very promising approach. You discuss its application based on case studies. However, please also comment on the performance of this automatic characterisation based on a diverse and longer dataset. Could this tool be used as a reliable interpretation of the CBLH results without looking at individual days carefully? If so, it would be interesting, how often the different classes are being detected at your site.
We have discussed the use of the flags and Zi various estimates in a specific point further in the present response.
This is indeed an important point about the potential of CALOTRITON in further documenting the complex structure of the low troposphere.

We made some preliminary statistics on the flags, in order to document the occurrence of textbook situations versus more complex cases. And we also give some statistics on the differences between the different types of Zi estimates. The results are indeed interesting, with for example :

(1) Over the 8-year time series at P2OA :

- 17 % of the days with more than 75 % of the data at flag 1. That means 17 % of days quite close to « textbook cases ».
- 3 % of the days show a significant difference between Zi_NP3_std and Zi_NP3_sub for more than 25 % of the data. This would mean that ITBL are not very frequent at P2OA.
- 72 % of the days show a significant difference between Zi_NP3_std and Zi_NP3_sup for more than 25 % of the data. Those are notably related to the large occurrence of cloud layers above the CBL top which generates an inversion.

(2) During the LIAISE campaign, at Els Plan, the flags show the complexity of the low troposphere :

- 0 % of the days with more than 75 % of the data at flag 1. → There is no textbook case during LIAISE.
- 75 % of the days show a significant difference between Zi_NP3_std and Zi_NP3_sub for more than 25 % of the data. This means that ITBL occurs very frequently during the LIAISE campaign at Els Plan.
- 92% of the days show a significant difference between Zi_NP3_std and Zi_NP3_sup for more than 25 % of the data. → There are established upper inversion in the LIAISE area.

Even if those results need to be taken with caution, they interestingly reveal the higher complexity of LIAISE study area, with consistent result. Case studies will help better qualifying the potential of this flag system.

In the concluding discussion of the revised manuscript, we give those intersting examples of statistics to illustrate the potential of using the flags and Zi estimates for the interpretation of the low troposphere vertical structure, and occurrence of elevated inversions and ITBL (page 26-27, lines 552 to 568), but also point on the need to better qualify this potentiality before generalization.

- When discussing the radiosonde comparison in Section 4.2, please reference literature on how you interpret the diversity in results from different radiosonde methods. Maybe it would be more conclusive to work with the subjective method only for the evaluation of the new algorithm and move the other scatterplots to the supplement material? It is not clear what we learn from looking at all the different results.

Since we still find interesting to show the issues raised by the different automatic estimates from in situ thermodynamic profiles, we decided to find a compromise between your suggestion, and the original proposition. Thus, we kept the parcel method and only one of the gradient method, and then showed the subjective method and the selection (« converging ») set (Fig 11 page 23). This enables us to enlarge the figure, and avoid redundancy without loosing the messages. We also made those messages clearer, about what we learn from those comparisons (page 23).

- Figures 2, 4, 7: Please consider a different way to present the data. The figures are very noisy and it is difficult to find the relevant information. Maybe try a different colour scale for the shading in the background? And reduce the amount of layer heights that are being shown? Choose symbols that are always clearly visible, even when several layer heights agree, i.e. they should not overlay each other so that they are not visible. Is it essential to always show all 6 panels? Maybe some could be moved to a figure in the supplement material?

  It is true that those figures were complex. We used some different colors in the back for the height-time section, with some transparency. We also changed the line colors and the symbols, to have them all more visible. Finally, we here decided to remove the windspeed and wind direction panels. The wind is useful to understand the structure and estimated Zi (shear for the first example, strong wind for the second, circulations for the third), but it actually is visible on the isolated radiosounding profiles shown in other fugures. We thought those may be sufficient.

  Former figures 2, 4, 7 have accordingly been changed (now Figures 4, 6, 8 in revised manuscript).

- Figure 6: please add mean and median or other statistics in the figure to facilitate comparison of the barcharts.

  We have added this information on former Figure 6 (now Fig. 3).

<u>Minor comments :</u>

P2, l26: The motivation paragraph of the introduction is very short. Might be useful to add a few more aspects that highlight why the height of the CBL is a variable that should be better characterised based on observations. Maybe move lines P3, L67-71 to beginning of the introduction, i.e. why is CBLH important for studies in complex terrain atmosphere dynamics?

We have re-orientated the start of the introduction toward the importance of estimating the CBL depth, and enriched it (page 2 lines 23 to 29). However, we kept the lines P3, L67-71 of the former manuscript there, because they motivate the monitoring of the CBL depth on this specific P2OA-CRA site, where we're making the long term series.

P2, Introduction: While it is important to highlight the diversity of techniques that are being used to measure the CBL height, it could be useful for this paper to highlight especially the shortcomings of previous methods applied to wind and turbulence measurements. i.e. demonstrate why a new algorithm is needed. And maybe provide some insights on the strengths of UHF input data compared to other measurements.

We have discussed those points, see page 4-5 lines 90 to 127. We notably citet and discussed the work of Heo 2003, Compton 2013, Collaud Coen 2014, Molod 2015, Bianco 2002, 2008, Liu 2019.

P3, Introduction: Please have a look at the recent review on ABLH detection measurements conducted by the PROBE COST action: https://doi.org/10.5194/amt-16-433-2023

We were not aware of this very interesting and well made review when we wrote the article, due to time concomitance. This is a quite timely review. As an exhaustive review, we took it as a starting reference in the revised manuscript.

P4, L85: “CBL height”

Typo corrected.

P4, L85: Is the time series 22 years long? On Page 3, line 67 it is stated that the UHF observations started in 2010. And according to table 1 observations started in 2011. Please give consistent information.

It is true this is not easy to follow. 22 years is the total length of this UHF RWP time series at P2OA-CRA (with some breaks when it is moved to external field campaigns). But only the 2015-2022 time period was used for the algorithm developpement. 2018 is a year taken for the configuration optimization (common, year with the CT25k ceilometer). 2011 is the year of the BLLAST field campaign.

We made all this clearer in the manuscript.

P5, l101: Please comment on the maximum range with a good confidence level.

We added this comment (page 6, lines 156 to 162), and added this specification in table 2.

The maximum height for this mode is usually around 3 km a.~g.~l., but may be only 500 m or 1000 m in winter, when dry anticyclonic conditions occur. It can reach 7 or 9 km within deeper clouds and rain. For the detection of Zi, we limited the search within the first 3000 m, which we made clearer in the revised version (page 8 line 194).

P5, l104: Is there a name for the retrieval of the 3d wind from the radial velocity? Or at least a reference? Is this done with an internal algorithm by the radar or does it requires post-processing? Are the data filtered for noise? Or any other quality control applied?

The 3D wind retrieval is the typical Velocity Volume Processing method. We specified it more precisely. It is used here with an internal algorithm (improved relatively to the manufacturer software), as well as the preceding step of selecting the meteorological spectral peak. Unfortunately, this procedure is not specifically published, but used in many studies and associated publications. We refered to one of them (page 7, lines 164 to 168).

P5, l110: Please provide reference for aperture correction. Even if it is the manufacturer’s user manual.

Thank you for pointing this. It is not a manufacturer source. We used the same technique as Jacoby-Koaly 2002, and so refered to them in the revised version (page 7, lines 173-174).

P5, l111: What is the distance between the radar location and the radiosonde launch site?

It is only a few tens of meters each time. We made this clearer in the revised version (page 22, line 457) . We did it by making clearer the position of the RWP and colocated radiosoundings, for each field experiment.

P11, l154: no plural for “fog”

Typo corrected.

P11, L161: averaging across how many gates?

We specified this (page 10, line 215). The averaged is made up to 3000 m maximum.

P11, L116-172: Maybe better structure the method description into data preparation (such as averaging and cleaning) and then start to introduce the new dimensionless variable and the detection procedure.

Effectively, that makes things clearer. We have made it clearer the data process upstream of the new parameter calculation. We actually added a sub-section 3.2.2 for this (page 10).

P11, L177: Please reference literature that describes the use of TKE or dissipation rate for the detection of the CBL height. Has your approach been used before? Same threshold? Also using UHF profiler data as input?

There is no dedicated publication for this Zi retrieval technique. But we did use this technique before and described it in the corresponding publications. We commented on this point more precisely and citet those references (page 3, lines 57-58).

P13, L236: Why is it important to give more weight to Cn2 rather than sigma_w?

Cn2 remains « the lead » for the Zi retrieval, that is the sommital inversion must have a significant weight.

The subsequent search for the optimum order ‘x=3’ confirms this.

P17, L323: Any studies that can be referenced regarding the complexity of ABL dynamics in the study area in Spain?

LIAISE is a pretty recent field experiment, and the complexity is still more revealed through the analysis of its dataset, but one recent publication, associated to LIAISE, can already testify to this complexity (Jimenez 2023, Mangan 2023). We cite it in the introduction of the third illustrative example (page 19, line 408).

P19, L354: Comment on how reliable the flag system is expected to be over a longer time period? Do you consider it useful for interpretation of cases studies or is it in fact a useful tool to automatically characterise the complexity across several years of measurements? How can this be assessed?

We discussed this point in the concluding discussion (pages 26-27, lines 550 to 568).

- We first consider flag as index of complexity, and reliability of the « best » Zi estimate. Making statistics on the flags is possible, to assess the occurrence of textbook cases versus complex cases.

- But we also consider the flags and the various estimates of Zi definitely useful for case studies. We are starting to use them in the context of the LIAISE experiment, where many complementary observations can help us understand the complex 3D and multi-processes situation.

- What is not easy to say and assess, is the possibility of further interpretating the flags and Zi estimates (with definitions of multi-layers like ITBL and RL) from a statistical point of view.

The case by case studies mentioned before is a very useful step to estimate how the interpretation may be generalized, for a statistical use of the flags and « Zi » various estimates.

P20, L364: How did you choose the appropriate method and why? Please reference literature and comment on the uncertainty that this could have on the comparison results.

We did not intend to be exhaustive on those methods, but wished to consider some methods that were relevant enough for the comparison with the RWP measurement point of view.

We actually chose to try several typical methods, knowing that :

- the parcel method had proved satisfactory in previous experiences, and has common aspects with a search of turbulence layer

- the gradient methods may very well correspond to some local maximum of Cn2

- the subjective method is always very useful for the precise understanding of what is happening.

We did omit the relative humidity maximum which though proved to be useful in some past experiences, but was not more relevant than the previously mentioned here.

We also omitted the bulk Richardson method, in spite of its large use. The reason is that we had experienced that it was not always appropriate. However, this method could appear relevant, due to its combination of temperature gradient and wind gradient. We did estimate it in response to Referee #2, and verified whether it was more appropriate. But we found that using this technique was nit better than using the parcel method, and did not change our message on the automatic detection methods from in situ profilings, nor on the evaluation of CALOTRITON.

We probably missed several of them, less commonly used.
We added some discussion about this aspect (page 23, lines 474 to 479 and page 24, lines 484 to 487).

P20, L378: What is the physical meaning of your CBLH is a stable layer is advected near ground level below?

It is true that as soon as an internal stable (or even simply different) layer is advected within the current CBL, then this CBLH looses its meaning and « status », turning from CBLH to residual layer top…

We made this point clearer in the revised manuscript (page 23, lines 473-476), by simplifying the formulation.

It is one reason why the use of Zi_NP3_sub is interesting. One can keep monitor the previous CBL, and change to RL, and have an idea of the entrance of an IBL within, and of its depth (whatever its internal stability characteristics).

P24, L465: Yes, you can analyse a long time series but with the current assessment restricted to convective conditions omitting shallow boundary layer development in winter.

Right, we corrected this in the conclusion. The latter has been significantly modified, following Referee #2 comments. But this point is made at the start of the conclusion (page 25, lines 519 to 521..