The authors would like to thank the reviewer for their time in reviewing our work and for their thoughtful comments, questions, and recommendations. Our responses are in blue.

Reviewer 2 comments

The article is motivated by the need to accurately model the boundary layer entrainment flux and changes in the property concentration due to entrainment in the budget of scalar property equations for use in numerical weather prediction models.

The authors aim to robustly estimate the entrainment fluxes from ground based remote sensing instruments. To this effect, the validity of the residual assumption used by Wakefield et al 2023 (W23) is investigated over the course of the evolving BL using LES framework, and a reinterpretation of the mixing diagram method presented in W23 is offered.

The scientific method, the results, and the discussions that follow are compelling. However, the article suffers from a few minor shortcomings that hamper readability. I recommend that the authors carefully consider below listed minor suggestions/revisions in their resubmission to the journal:

- Line 53-55: "Because of the vertical resolution of the profiling instruments used in W23, they computed the mean of the mixed layer properties only from 0.1zi−0.5zi, where zi is the depth 55 of the BL determined from the TROPoe retrievals using a parcel method". Please clarify what is the instrument vertical resolution. Is this superior to the LES resolution? Do the authors wish to suggest that the data resolution used by W23 is insufficient to resolve the BL? I fail to understand the point of this statement.

Thank you for highlighting this section. We apologize for the confusion here and have corrected our statement to properly reflect the work that was done in W23. The sentence now reads “To avoid the surface layer and entrainment zone, W23 computed the mean of the mixed layer properties only from $0.1 z_i - 0.5 z_i$, where z_i is the depth of the BL determined from the TROPoe retrievals using a parcel method.”

- Line 67: MD here is used without defining the abbreviation. The abbreviation is defined in section 2.1. Please move the definition of MD to line 67 instead for clarity.

Thank you for your close reading. We have added the definition to line 67.

- The end of section 1 could very much use a breakdown of the article sections to guide the reader through the document.

Thank you for your suggestion. We have added “Section 2 describes the methods used in this study. Section 3 presents the results of whether the residual assumption for deriving entrainment is valid, a comparison of different definitions of the mixed layer for calculating ENT2, and variability across different dates and boundary layer depth definitions. Section 4 offers a discussion of the results. Section 5 highlights conclusions and presents opportunities for future work on this topic. “ to the end of section 1.

- A nearly periodic drop in the BL depth is observed when minimum potential temperature flux method is considered for single column (left plot of figure 2). See at times 0930, 1030, 1130, 1230, and 1400 hour time stamps. Is there an obvious explanation for such an observation? Such a systematic trend could potentially affect the relative magnitudes of ENT1 and ENT2 for single column data and brings into question the representativeness of the BL depth estimate using the minimum potential temperature flux method. (Overall, section 2 is well designed to crisply convey the investigation methodology).

Thank you for your observation. Since the single column fluxes are being calculated over a 30-minute window size, this could be the result of a rounding error. In Rosenberger et al. 2024, they used this method of calculating higher order moments from single column output and found that the fluxes had more variability than higher order moments such as variance and skewness.

- Are ENT1 and ENT2 estimates presented in Figure 5 computed from single-column data, slab averaged data, or average of the equidistantly placed columns? Please specify this upfront when describing figure 5. (If such a detail was already resented in section 2 earlier, ignore this comment).

Thank you for your comment. These ratios come from the slab averaged data, and to reflect this, we have changed the sentence on line 136 to read “Figure 5 shows the ratio of the slab derived ENT1 to ENT1+ENT2...”

- Further, in figure 5: There appears to be no systematic trend in the ENT1/(ENT1+ENT2) ratios using the three methods even in the afternoon data when compared across the 5 different days. For instance: for days represented by the square and circles, the BL estimated using the maximum humidity variance seems to provide consistently the smallest ratio of ENT1/(ENT1+ENT2) while the BL depth estimated using level of neutral buoyancy provides the largest ratios. Is there a specific reason as to why such a trend is not observed for the other 3 days?

Thank you for your question. We found that the ratio between ENT1 and the total entrainment depends on where the BL depth falls in the entrainment zone, so when one BL depth is higher than another, it will have a smaller contribution from the ENT1 term. The level of neutral buoyancy is not always going to be the deepest definition. We are not arguing that, in using a specific definition, there will always be a larger contribution to the ENT1 term, rather, we see that the contribution is dependent on the relative depths of the BL definitions.

- Lines 179-181 are vague. "Slightly better job" is purely qualitative here. It is hardly distinguishable from visual inspection of figure 6, especially in the afternoon data. Some effort to quantify such a difference would be more defensible and provide further evidence to the use of full zi as a proxy for the mixed layer.

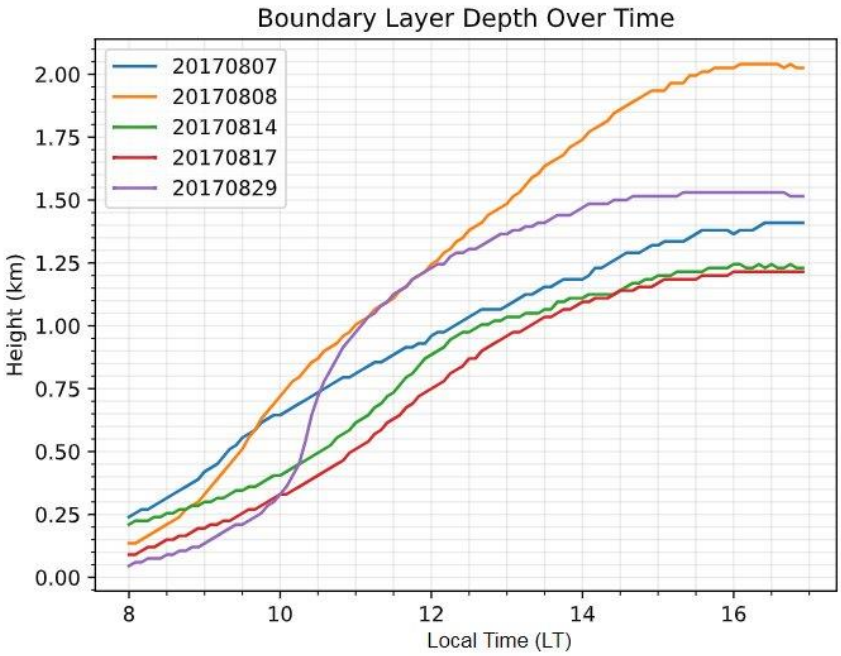
Thank you for your comment. We agree that we need to quantify the distance of the respective clusters in Figure 6. We found this by calculating the pairwise distance across each cluster and included those distances in the text, starting at line 180. “For the morning, the pairwise distance of the respective clusters is calculated, and for the restricted method that distance is 1.28 (kJkg-1), and for the full method, that distance is 1.10 (kJkg-1). In the afternoon, there is much more overlap between the two different methods, though that overlap tends to underestimate both ENT2 contributions. The pairwise distance for the restricted method in the afternoon is 1.58 (kJkg-1) and for the full method, it is 1.45 (kJkg-1).” We changed “tighter” to “as the cluster around the slab value in the morning is closer, according to the pairwise distance of farthest points in the cluster” in line 187.

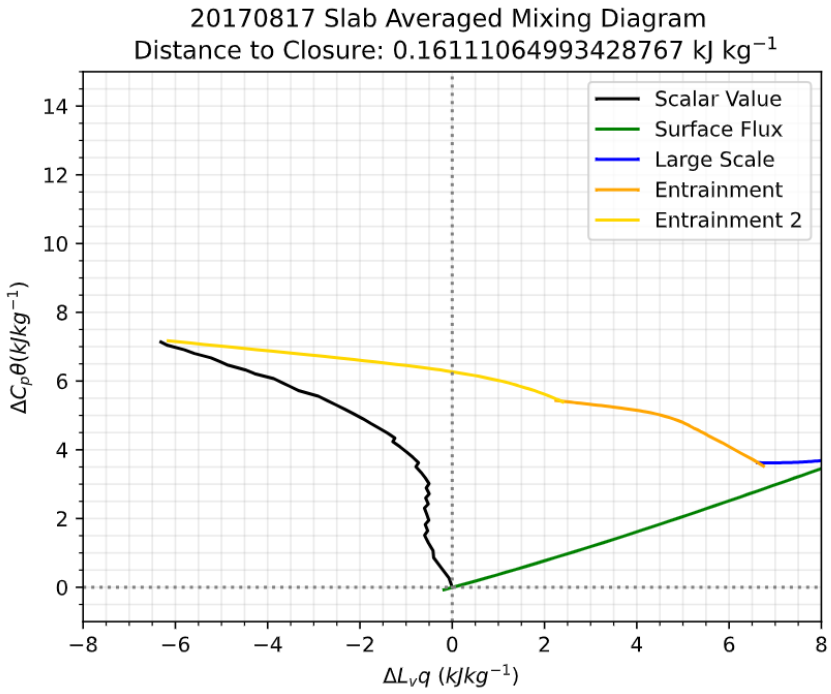
- I fully concur with RC1's comments that lines 193-196 are confusing and need a revision. This is only clear after reading it numerous times, but not readily intuitive from visual inspection. A brief explanation here is warranted.

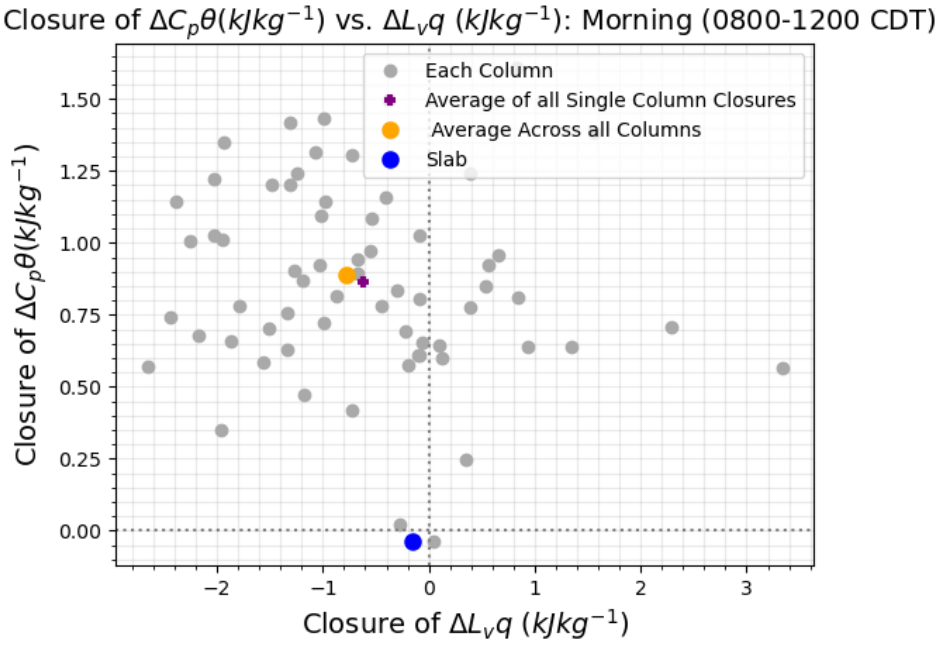
Thank you for seconding the RC1’s comment on this section. We have changed the phrasing here to say, “We see that during the morning (left), the average across the 64 columns and the average of all the columns is very close. This means that the average single column will yield a similar closure value to the full array. In the afternoon, the average of all the columns (purple) is closer to the slab value (blue) than the average across all 64 columns (orange), so for this case, the average single column replicates the slab values better than the result from averaging across the full array.”

- The mean closures presented for August 17th dataset appears as a clear outlier. Is there a specific reason for this?

Thank you for your question. The 17th has a rapidly changing latent heat surface flux while the boundary layer is shallower than that of the other dates. We believe that this is due to the fact that there was rainfall at night on August 16. There was no rainfall on any of the other dates preceding the date that was selected. By adding additional columns into the mean, the degree of closure improves. The reason that it increases at first is likely due to the specific columns selected in the calculation.







- In figure 9 it is not possible to differentiate between the different days of the dataset. It would be nice to differentiate the legends of the closure estimates for average of single columns for clarity.

Thank you for your suggestion. The points in figure 9 have been color coordinated by date.

- Lines 221-223: "ENT2 is larger the higher in the entrainment zone the BL depth definition." This sentence doesn't make sense to me. What do the authors intend to say here?

Thank you for your comment. We see that, when the boundary layer depth definition falls higher in the entrainment zone, the magnitude of the second entrainment term is larger than the first entrainment term – or that the component of total entrainment due to ENT2 increases closer to the free troposphere. To make this point clearer, we have added "The magnitude of ENT2 is larger the higher in the entrainment zone the BL depth definition. This means that the contribution to the total entrainment from ENT2 increases the closer to the free troposphere the BL depth definition."

I recommend that the article be published subject to the authors addressing the above mentioned minor suggestions/revisions.