**Response to reviewers for the paper *A multi-instrument fuzzy-logic boundary-layer top detection algorithm* by Elizabeth N Smith and Jacob T Carlin -** *Anonymous Reviewer # 1*

We thank the reviewer for their time and comments on our paper. To guide the review process we have copied the reviewer comments in black, italicized text. Our responses are in regular blue font. We have responded to all the referee comments and, when appropriate to do so, shared the resulting alterations to our paper in blue, italicized text. Any line number references are to those in the originally submitted manuscript.

*This paper presents the results of using a fuzzy logic algorithm with a combination of doppler lidar, radiance [interferometer] and microwave radiometer, with validation done by comparisons with radiosonde data. I feel the paper needs a major revision as the presentation is confusing and the results are not very clear. I realize that there are lots of different cases to present, and many options in the radiosonde comparisons, but in the end, it's not obvious as to how these retrievals might ultimately be used. But I think a rewrite could make it much easier to extract this information. My specific comments are:*

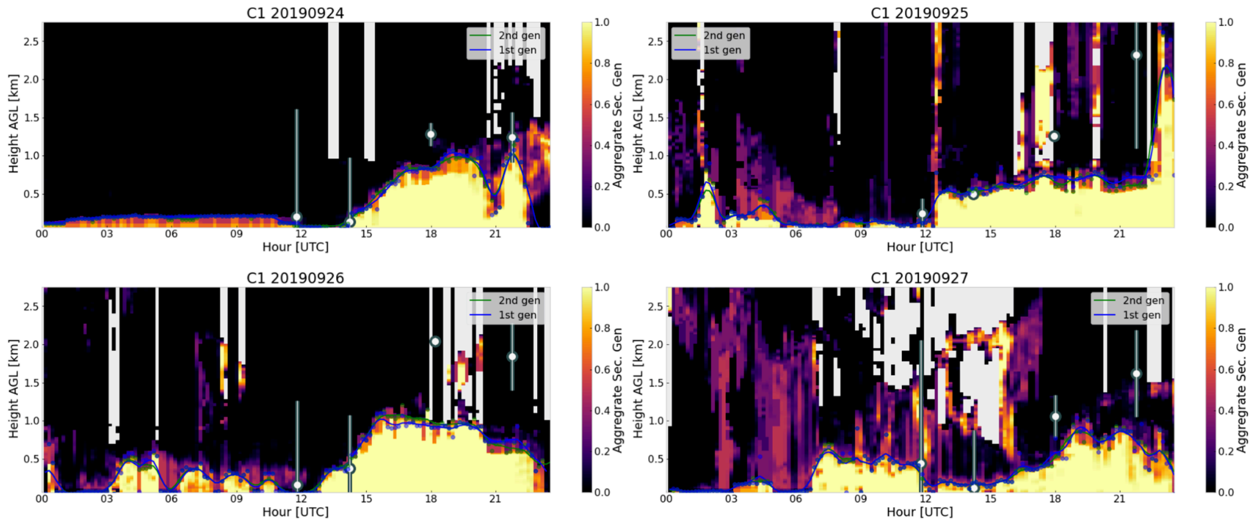
1. *Both the and abstract should state more clearly what observations are being used for the retrieval and validation (see the first sentence above).*
   We modified language in both the abstract and the Summary and Outlook sections to be more explicit about the three instruments used to develop the algorithm, while remaining clear that the algorithm could be adapted to use other similar observation types.
   <u>Abstract</u>: In the second paragraph the first sentence and last sentences were modified:
   *This study introduces a fuzzy-logic algorithm that leverages the synergy of multiple remote-sensing boundary-layer profiling instruments: Doppler lidar, infrared spectrometer, and microwave radiometer.*
   *…*
   *While developed with the three instruments mentioned above, the fuzzy-logic boundary-layer top detection algorithm, called BLISS-FL, could be adapted for other wind and thermodynamic profilers. BLISS-FL is released publicly fostering collaboration and advancement within the research community.*
   <u>Summary and Outlook</u>: In the opening paragraph of this section we added a parenthetical after platforms
   *While this algorithm was developed for use with the CLAMPS platforms (i.e., Doppler lidar, infrared spectrometer, and microwave radiometer), it could be applied to or adapted for similarly instrumented facilities such as but not limited to the Department of Energy's Atmospheric Radiation Measurement profiling facilities.*

2. *Line 110: do you mean the first generation step in B18, or is this changed in the current algorithm?*
   We mean there is an addition to the algorithm at a certain stage, and we will highlight it when we get there. In addition to the existing language "with any deviations from or expansions upon B18 highlighted", we modified the text to more clearly state "*At this stage the algorithm design deviates from the design of B18*" at the beginning of the paragraph which started on line 125 in the original manuscript. This language should make it clearer that the differences are now intentional and by design, not only differences in available variables. We hope that adding this language at the location when the addition to the algorithm is introduced makes it clear that prior to this the algorithm design is generally consistent.
3. *Line 126: Why are buoyant processes important during the night? This seems counter intuitive as buoyancy should grow during the daytime.*
   Buoyancy does grow during the day. Buoyant processes are not limited to the growth of buoyancy. Specifically in the scope of the algorithm we are discussing, the first generation step relies heavily on measures of mixing, or in other words observations showing that mixing is ongoing. In this case, that means mostly mechanical mixing observable by Doppler lidar. Of course, the buoyant processes related to daytime heating can lead to that mixing, but those buoyant processes are not directly observed by the Doppler lidar. Buoyant processes are important during the night when the daytime boundary layer decays and the mechanically driven processes may become less dominant, which in the scope of this algorithm can make the Doppler lidar observations of mixing less informative. As such information about the presence of the daytime capping inversion, thermodynamic structure in the residual layer, and the evolution of any surface inversion is critical at night. In non-canonical cases, the classical "Stull" evolution is not representative of boundary-layer evolution including mechanical or buoyant processes. Including both, specifically through the night, is useful during transitions and during non-canonical nocturnal boundary-layer cases.
4. *Line 131: So you mean that the B18 algorithm was not successful in estimating overnight BL height?*
   The reviewer brings up a fair point that the B18 algorithm is applied throughout the full diurnal cycle. In their paper, they do not offer any assessment or verification at night. In any case, the current phrasing implies that they only applied their method to daytime hours which is inaccurate. We will rephrase this for accuracy.
   *In order to capitalize on the availability of thermodynamic profiles and to improve the capability and robustness of our algorithm during nocturnal hours when the mechanical mixing that backscatter-based instruments (e.g., Doppler lidar) observe can be absent or limited to the residual layer (Schween et al. 2014), the first-generation step also includes temperature inversion height as an input variable (Fig 2c).*
   *Schween, J. H., A. Hirsikko, U. Löhnert, and S. Crewell, 2014: Mixing-layer height*

*retrieval with ceilometer and Doppler lidar: From case studies to long-term assessment. Atmos. Meas. Tech., 7, 3685–3704, https://doi.org/10.5194/amt-7-3685-2014.*

5. *Section 3: Do radiosonde observations measure thermodynamic BLH rather than the ML height? If so, is it really a good comparison?*
   In the scope of this algorithm development, we view mixing and thermodynamic contributions to boundary-layer height as coupled. The algorithm includes measures of mixing as it occurs and indicators mixing has occurred. Thermodynamic BLH parameters can be related to indicators (i.e., mixing has occurred) while MLH would be related to measures (i.e., mixing is occurring). Also, as mentioned in the paper, radiosonde observations are a commonly available and known source of data, which allows the comparison to be made against a dataset that is already integral and 'standard' in many atmospheric data communities.
6. *Line 402: It is fine not to recommend a preferred radiosonde PBLH algorithm, but it would be really helpful to see a plot like*
   In the absence of any additional context, we are treating this comment as if it is related to comment 7 as it seems incomplete on its own. We apologize if our assumption is incorrect.
7. *Figure 3 is really helpful to see the evolution of the PBLH estimate during the course of a day. It would be even better if you could plot the radiosonde estimates of PBLH on top of this. There is lots of discussion of how the algorithm does at different times of the day, but much of this could be clarified with a comparison like this.*
   The strength of an algorithm like the one we present here is we are able to capture or at least make estimates about BL evolution that is not able to be captured in typical radiosonde observations. In our case, we are fortunate to have some additional temporal resolution in the CHEESEHEAD radiosonde dataset (which featured up to 4 radiosonde launches/day) that makes this sort of comparison partially possible. However, the days and regimes in which comparison is available are not fully representative of the comparison we do with the full radiosonde dataset. CHEESEHEAD data are collected in a forest canopy, which perhaps has implications for BL height evolution and BL height detection which are not accounted for in this work. In any case, it is the dataset we have available, so we can include the comparison as requested. We now include a series of four days from the CHEESEHEAD project while the 4-times daily radiosonde releases were occurring. This comparison shows a spread of 'performance,' but generally shows what is expected based on the scatter plot comparisons. Differences are smaller during the early part of the daytime BL. The biggest differences typically occur latest in the afternoon.

   To include this figure in this text, we had to reorganize some of the text in Section 4. Previously the text read as introducing the median soundings scatter plot, describing what it shows, then comparing to the scatter plot of all the sounding methods to describe why the median is important. Now the text reads as introducing the median scatter plot, moving straight to comparing to scatter plots of all methods to describe why all methods are important, then returning to the median scatter plot to describe what it shows. Then we follow that description with complementary information now available from showing the aggregates from CHEESEHEAD with soundings overlaid with respect to time.



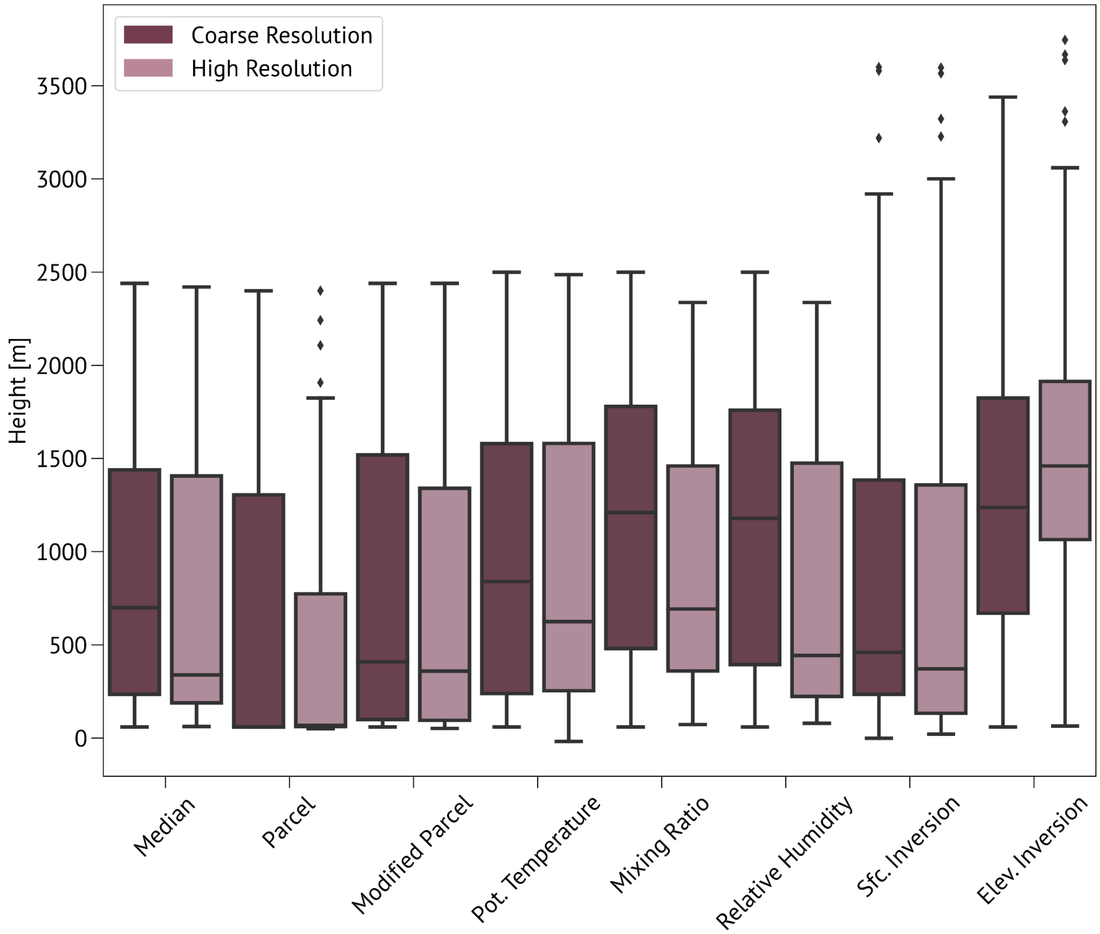

8. *The summary section (5) doesn't really have any concrete conclusions. As you say, it's hard to make a fair comparison since a 100 m difference means something very different depending on the time of day. But plotting the height estimates in comparison to radiosondes as a function of time would really help with this.*
   The radiosonde time series plot is now included in the manuscript in response to another comment. In a more direct response to this comment, the Summary and Outlook section, particularly the 4th paragraph, was revised to be more explicit about conclusions and outcomes. We discuss that while there is agreement between radiosondes and the algorithm, there is a pattern of underestimation. We speculate about the reasons behind this including the limitations of remote sensors. We also now more explicitly discuss that exploring errors through comparison with respect to time would require more data, specifically more data around the diurnal cycle to create discrete samples and or develop normalization techniques. The changes appear starting at approximately line 422 in the original manuscript.

   *This comparison suggests fairly strong agreement between the techniques, but still with a maintained pattern of the fuzzy logic algorithm applied to CLAMPS observations underestimating BL height compared to the median of radiosonde-based methods. Given the instrumentation types included in the algorithm development and evaluation (Doppler lidar, infrared spectrometer, microwave radiometer), we speculate that even in the best-case comparisons, remote-sensing-based observations are simply more inclined to detect the lowest possible indication of BL height (especially in cases where the BL top itself may be a layer with depth). These types of instruments are also more likely to lose measurement capability and resolution with distance from the surface. These reasons* *could combine to result in systematically lower estimates of BL height than those provided by in-situ platforms like radiosondes.*

*Unlike some similar algorithms, this approach has the capability to utilize thermodynamic observation information and kinematic observation information (when available) to provide BL height estimates throughout the diurnal cycle. Specific analysis is focused on the early morning and late overnight periods. While some statistics included suggest perhaps even minor improvement compared to the daytime group, this may be a misleading result. For example, a bias of 100 meters means something different when the BL height is O(100 m), which is common in the overnight and morning hours, compared to O(1000 m) BL height values in fully developed daytime and afternoon BL. More data is needed in this comparison to understand the role time of day plays in how the fuzzy logic algorithm behaves. More data could provide the opportunity to build time-dependent samples or to develop temporal normalization approaches to control for this important sensitivity.*

9. *Figure 7: It's not clear what you mean by dark and light colors.*
   In the original manuscript, “dark” and “light” colors referred to the shading/brightness of boxplots for each set of colors (e.g., for the modified parcel method, dark green referred to the results using coarse radiosonde data and light green referred to the results using high-resolution data). We revisited this figure to determine how to present the information more clearly. We came to the conclusion that using different colors for each method and relying on the light/dark shading to show the differences between the box pairs for each method was too confusing. We updated the figure to use a single color scheme for all methods, with the labels along the x-axis describing each method. We also are more specific about using the language of ‘pairs’ to point out that two boxes go with each method. For all of the pairs the darker shade depicts the results for the given method (or median) using coarse resolution data and the lighter shade shows the results for the same method using high resolution data. We added a legend to the figure to reiterate what

the light and dark shades refer to. The updated figure is shown below.




10. *Table 3 appears to be the primary result of this work. Would it help to include a percent difference along with the absolute difference?*
We have added percent differences at the suggestion of the reviewer. The values generally support the same outcomes. There is a roughly -33% bias overall, decreasing with the first two exclusions. When broken down by time of day the morning bias (~ -42.5%) is roughly twice the afternoon bias (~ -20%).