Atmospheric Measurement Techniques Discussions

Atmos. Meas. Tech. Discuss., referee comment RC2
https://doi.org/10.5194/amt-2021-383-RC2, 2022

Comment on amt-2021-383

Anonymous Referee #2

Referee comment on "Testing the efficacy of atmospheric boundary layer height detection algorithms using uncrewed aircraft system data from MOSAiC" by Gina Jozef et al., Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2021-383-RC2, 2022

The manuscript “Testing the efficacy of atmospheric boundary layer height detection algorithms using uncrewed aircraft system data from MOSAiC”, by Gina Jozef, John Cassano, Sandro Dahlke, and Gijs de Boer, evaluates different published methods for the determination of the ABL height for a unique data set sampled by an uncrewed aircraft system over the sea ice of the Arctic Ocean. These “objective” methods are verified against a “subjective” or visual method to evaluate the height of the ABL, and the robustness of this approach is shown by applying it also to radiosonde profiles sampled in close spatiotemporal proximity. The manuscript is well within the scope of AMT but requires major revisions before it can be accepted for publication.

Here my major comments:

The information given in the introduction on the concept of an atmospheric boundary layer height is very superficial. Here the authors should expand and clarify why this concept is important, for which applications it is used, and also in which way its definition is debatable. In my eyes, it is a diagnostic parameter used to quantify the altitude up to which the direct surface-atmosphere interaction can be considered relevant. It is some sort of simplification, which is helpful for many applications, but it is not a physical property of the atmosphere, e.g. it may not be continuous for example during regime transitions. A more critical reflection on this should be included. It should also be clear that either method for identifying the ABL height is only providing an estimate (one could even claim that there is no true ABL height only different methods to estimate or diagnose it).

This also has implications for your line of argumentation, which seems to be based on the

assumption that your "subjective method" is giving the true ABL height. I agree that a visual evaluation of the ABL height by an expert may be generally better than any objective method, but also the expert can be wrong, e.g. due to misleading observations (DH2 and RS provide only a "quasi-snapshot").

More details on the calculation of the Rib number and the determination of underlying parameters (gradients) are required. In particular, I lack details on the determination of the wind speed from the helical flight patterns. There is a reference to unpublished work, but some important details should be mentioned here. I assume you use the 10Hz 3D wind data? How do potential time lags and/or inaccuracies in IMU or GNSS data influence instantaneous wind measurements? I also noted that the DH2 temperature profiles start at lower altitudes than the corresponding wind speed profiles. Why is this? In which coordinate system is the wind speed measured, relative to sea ice or in earth coordinates? I think this has only implications for the Richardson number at the first level, but this is still important to mention. Furthermore, the response time of the temperature and humidity sensors and their implications for the computation of Ri and theta_v and determination of the ABL height should be discussed. For theta_v two very different response times are combined (T and RH sensor). What effect does bin-averaging, including ascent and descent data have? How may non-stationarity, e.g. a substantial temperature change near the surface within the ~30m flight time affect the results, and are there any observations indicating that this may have been an issue (One could simply make use of surface-based observations to detect non-stationary conditions during the flight period)?

I am missing a general assessment of the different sampling methods (radiosonde vs DH2). There are some important differences like the radiosonde can only sample during ascent, differences in the vertical climb speeds, response time, wind measurements. It' is therefore well possible that one of the data sets is generally more smooth or has higher uncertainties in particular in a specific altitude region or under certain conditions.

The results/discussion section is kept on a rather general level and has little content compared to the methods chapter. Interesting research questions are not addressed (systematically), e.g.:

- How do different sampling and data processing methods affect the differences for radiosonde vs DH2 based ABL height estimates?
- In Sect 3.2 you address the question of how stability or specific features in the ABL structure cause certain methods to perform better or worse than others (you also scratch on the surface of this in Sect. 2.2 ff) but this is done in a rather episodic manner. Table 3 would be a good starting point for expanding on this. In the corresponding section, you mention that one could list problematic features to be used in a pre-screening procedure. I think you should have the data and knowledge to propose such a list. I consider this as very relevant for the research community using similar systems to determine the ABL height.
- How sensitive are the different methods to the choice data processing methods, or sensors used, e.g. vertical averaging procedure? This would also be of interest to the community, e.g., to make adjustments for different measurement systems.

Minor comments/suggestions

Since this is a technical paper small details may need some additional attention, e.g. the difference between level and layer (which has a thickness) and the corresponding indexing. Are meteorological parameters, gradients at level k averaged over k to k+1, and how exactly are averages and gradients determined?

I also regard the term "subjective" as misleading. In Section 2.3.1-3 you describe criteria that could also be automated (you may have done that). I would therefore claim it is a semi-objective method, where the final decision is made through a visual interpretation by an expert. "Visual" and "automatic" may be the better terms to distinguish the two types of methods. When reading the abstract my first impression was that it is strange to evaluate an "objective" method with a "subjective" one. Generally speaking, one may want to trust an objective method more than a subjective one. This confusion could be avoided by sticking to the terms "visual" and "automatic".

In general, I also see the potential for condensing the content of in particular the introduction and methods sections.

Readability could be increased by avoiding some passive phrases and using “we”, as done elsewhere.

Specific comments, technical corrections, and suggestions for improvement:

L15: “fixed-wing uncrewed aircraft system”

L16: consider introducing an abbreviation for the ABL height, e.g. simply "Z" or "H" or with subscript "ABL"

L16: “the ABL structure”.

L18: “the ABL height”. In general, I have the feeling that some articles are missing, in particular when abbreviations are used. Please add wherever this could increase the readability. Note that if you would simply use “H” instead of “ABL height” you would not need the article in this case.

L39: “pack ice” might be better.

L49-60: This paragraph could be rewritten, giving a general description of the ABL and its typical structures of the ABL and above. This would then naturally lead to the concept of the ABL height.

L51ff: “...the ABL is mostly impacted by interactions between the atmosphere and sea ice surface features, including the generation of turbulence through surface energy fluxes emitted from open water regions such as leads…”. This is hard to understand. I think you want to mention both, mechanical and buoyant production of turbulence, but it reads like buoyancy is part of the mechanical production (by interaction with surface features). It should also be made clear that buoyancy is mostly negative over sea ice.

L54: I consider mixing as a turbulent process so “radiative mixing” could be misleading. How about “turbulent mixing forced/triggered by radiation–cloud interaction”...

L55: Consider mentioning the effect of ice edges e.g., at leads. The roughness is increased due to the freeboard.

L59: include "(LLJ)" here.

L76: only one “)”

L76: Here one could expand a bit on where the ABL height could be identified be when there is a capping LLJ? At the core, somewhere below or above? Are there different opinions about this?

L114: Is the SHT-85 really measuring at 100Hz. The response time of this sensor is rather slow, but of course, it is allowed to oversample.

L117: Here more details on the 3D wind estimates could be given (see comment above).

L110-L120: Can you also provide similar information (doesn’t have to be as detailed) for the radiosonde sensors (here or elsewhere). Does the radiosonde also contain a Vaisala sensor (I think the corresponding radiosonde sensor would be RS41)? Can the radiosonde sensor also be included in Table 2?

L147ff: Were there any cases when a clear determination of the boundary layer height was not possible, even though the max altitude was sufficient. Potential reasons could be non-stationarity or internal boundary layers?

L152-155: Does this mean data from both, the ascent and descent were used? If so you may cause a kink at the level where you start using the first data after takeoff but have data from the descend before landing.

L174, 175, 177: Should be "an SBL/NBL" but "a CBL". Check the entire manuscript for this type of typo.

L176: You mean that the SBL "can range from …" but what is written refers to theta_v.

L183: It is not clear what "i" refers to.

L208-209: Your data may even suggest that there is a tendency to more stable conditions during the seasons you observed.

L227: "The Bulk Richardson number"

L227-228: Consider rephrasing the first sentence to allow for buoyant suppression. "buoyancy" may be better than "buoyant production"

L240ff: Here you should be more precise. See for example https://glossary.ametsoc.org/wiki/Bulk_richardson_number: "In the limit of layer thickness becoming small, the bulk Richardson number approaches the gradient Richardson number, for which a critical Richardson number is roughly Ric = 0.25 … Unfortunately, a critical value is not well defined for the bulk Richardson number, leading to uncertainty in turbulence likelihood for values near the critical value."

L247: Here more detailed information would be necessary. The raw data is bin averaged using 1-m bins,. then you use 30-m bins for Rib with 5-m resolution. Note that the choice of bin size may have implications for the choice of thresholds, e.g., for Ri_b. This could be discussed further.

L250-255: Assuming that in the lowermost ~5m theta_v will have its strongest gradient this method has its weakness when no data from this level can be used, as the CBL would be found at higher theta_v, thus resulting in an underestimation. If only the ascent data below ~5m is ignored but descent data is used this should be clearly indicated. An alternative would be to use IR surface temperature estimates. Overall this appears like a very objective method, which can be fully automated.

L250. "theta_v" without "the"

L251: Replace "identify" e.g. with "determine" to avoid repetition.

L266: Consider changing to "slope of theta_v". At which heights are these multiple shifts?

L271: The determination is not made entirely based on the humidity. In the last step, theta_v is used again.

L278-280: If DH2 data is only considered from altitudes above a certain threshold this statement is not well supported unless surface temperature estimates from an onboard IR sensor are taken into consideration. Note that IR surface temperature estimates may be subject to uncertainties related to sensor temperature stability, the emissivity of the surface, radiation flux divergence, and sensor tilt.

L284-285: This is often related to an inflection point in the wind profile or at least the layer where wind shear approaches zero. Showing both profiles of theta_v and wind speed would be illustrative, for the interpretation of Ri_b.

L290-L291: This statement implies that the correct SBL height is known. One could also choose to define the SBL height as the level where there is such a clear shift in Ri_b. One may then simply end up with a different height.

L305: See previous comment: Does this imply that the ABL height can be determined with a resolution of 5m or 30m?

L308-309: Do you use equations 1-3 for the determination of the regime? Please include a statement, making this crossreference since this (is as you indicate) slightly different from the Liu-Liang method.

L318: Note that the height of the lowest level is critical. If the levels close to the surface are not sampled this may become an issue. You may want to discuss this later on.

L320: “the atmosphere”

L323: The notations for vertical gradients are not consistent, compare e.g., line 304. Please stick to one notation.

L326, Eq 6: Gradients would have to be determined between two layers so from k to k+1 (or k-1 to k+1). It is hard to follow which of these two levels is chosen as the ABL height. Depending on your resolution this makes a difference for the ABL height. Since this is a reoccurring issue for all methods relying on vertical gradients a general statement at the beginning of section 2.2 would help.

L334-335: The chosen threshold is quite different from the originally proposed one. What is the reason for considering this as inappropriate?

L341-342: This statement appears very general. TKE is just one way to define the SBL height. This goes back to my general comment on a more critical reflection on ABL heights. It should also be moved to the introduction.

L343: Below the ABL there is only the surface, which is only buoyant in an oceanographic sense.

L344: simpler: "the SBL height" or "the height of an SBL".

L356: Consider changing the subscript to account for different ranges of theta^dot_r for different regimes (compare Sections 2.4.1.1-3). BTW, what does the r stand for?

L391: better "starts" - "extends" may be associated with an upward direction.

L396: layer or level?

L402: “the stability regime”

L413: “local maximum”

L416: “local minimum”

L420: Can you provide a brief interpretation of this figure as done for the previous methods/figures?

L429: I suggest using the term "threshold value" to avoid a bit of the discussion on the critical value of Ri (see also comment above). One could interpret the following paragraphs as you were trying to find Ri_bc for the transition between turbulent and laminar flow based on your observations, but in fact, this is not the scope of this paper and you don't make use of any turbulence observations.

L445-447: I get an idea of what is meant here, but I would suggest reformulating this sentence to make it more clear. Can this be broken up into two sentences?

L451-452: Can you summarize the main differences that may play a role?

L453: "...applied to radiosonde data ..." is enough, "to identify ABL height" is implicit.

L454: It would be natural to state the number of the radiosonde profiles you used somewhere here.

L458: no “the” before “theta_v”

L459: Consider changing to e.g., “create profiles of the same parameters as for the DH2 data”

L462: can you give the reason for this inaccuracy?

L469: Stick to one common unit for vertical temperature gradients. I propose K/km.

L474-476: Do you mean: "Similar figures for all available DH2 and radiosonde profiles can be found in …". Is it possible to use a hyperlink to get directed to this online supplementary material?

L481: Consider reformulating, e.g.: "In general, the deviation between ABL heights from DH2 and the radiosonde increases with decreasing time proximity".

L507: This is one example where the subjective method appears as the “truth”, although it is most likely not perfect, either. Consider adding "... compared to the subjective method".

L510-511: Please note that there have been some high-level debate on the use of p-values and the 5% statistical significance, see e.g., https://www.nature.com/articles/d41586-019-00857-9. I am not an expert in statistics but I recommend at least using a somewhat "softer" formulation, like "can be considered statistically significant when the p-value is less than 5% (or 0.05)." If your p-value is 0.05, this means there is still a 5% chance that your result is completely random.

L513 and elsewhere: Consider using superscripts "Ri_b^0.5" and Ri_b^0.75.

L519-520: See comment above

L520: What is more complex, the method or the result from the method?

L532: “R^2” without “number”

L534: “0.5 to 1”

L556-559: Very wordy sentence for saying that you assess the (cumulative) frequency distribution for the difference of the objective methods relative to the subjective one.

L571: Consider changing to “number of cases within each (relative difference) category”

L575: Consider replacing “predicts” since it's a diagnostic method.

L581-584: Here, I would like to see some discussion on such aspects. Are there differences in the sampling or data processing methods that may lead to the fact that different threshold values work best? Such discussions may be very useful for the research community as they may have to adapt threshold values depending on their observational approach.

L588: “the ABL heights”

L595: Consider replacing “it is not consistent enough to be reliable” with “it is not reliable”.

L598-599: Do you mean the LLJ core? This statement could need a reference.

L600-601: “throughout the whole profile”

L625: radiosondes were launched from the deck so "right at the surface" should be changed to "close to the surface"

L628: I suggest changing to "during polar night".

L630: There is at least a debate on whether the free atmosphere above the ABL is really laminar or rather weakly turbulent. The choice of a threshold value for Rib largely depends on the vertical resolution you use to compute Ri_b.

L649: only "several", "different" is implicit

L649: "methods (i.e., Liu-Liang …)"

L652: You could state the threshold values 0.5 and 0.75.

L656: It is occasionally quite good, better to use"largely".

L663: Is this also true for largely ice-free areas in the Arctic, which are likely underrepresented by a sea-ice-based campaign?

L671: This repetition should not be necessary.

L672: It should be safe to use active voice: “would change minimally”

L679: “These similar conclusions” (plural)

L681: "no method" or "no single method"

L681-683: Again: The objective methods may be better than a visual inspection by a non-expert. A combination of both visual + objective may be better. For the semi-automatic approaches, the list of features that may cause certain methods to fail would be very useful. One different approach could be to use an ensemble of automatic methods and visually inspect only the profiles for which the resulting ABL heights diverge.

Figure 2: Caption: “each flight” is misleading when showing only selected flights

Figure 3: The caption is extremely long. The legends could be merged and plotted only once (applies to more figures)

Figure 7: This figure should be redone. Here are some suggestions: Use different color schemes for the figures, e.g., not two different shades of green in the same panel (Panel 5 also has several different shades of green for the horizontal lines to indicate ABL heights from the different methods). What are the dashed lines in Panel 5? Since the ABL heights shown as text in Panel 1 and 2 are also related to Panel 3-5 it would make more sense to put them in a small table (2 lines 6-7 rows), placed under the 5 panels. Use one common legend for all 5 panels. Consider using a logarithmic scale for Ri (only if all values are positive) or a narrower range. Condense the caption and structure it better. These suggestions could partially also improve the other figures.

Figure 8: Consider, using a smaller range for the y-scale in the bottom panel and rather mention that a few outliers are not visible with this scaling. It may also be simpler to use “Relative difference” instead of “Absolute value of the percent difference”.

Figure 10: This may be my personal preference, but it might be better to display this as a CDF plot (four lines) or a histogram (four bars for each bin) using bins with a constant width (e.g., ranging from 10% to 20%). It would also be possible to combine the CDF and histograms in one panel. Consider using a different y-label, e.g., "frequency of occurrence" and only one common x-label. For the "No ABL Height found" class you could simply add NaN or display them differently, e.g., plot them as shaded bars or horizontal dashed lines (sort of downward from 100%). I also note that the bars in the last two columns don't add up to 100%. Do the missing cases indicate a relative difference exceeding 100%?