Review of “Synergistic retrieval of atmospheric boundary layer height from multi-source observations over the Arctic Ocean during MOSAiC” by Yuliang Liu, Shulei Li, Lei Liu, Shuai Hu, Xuyang Li, Rongying Shao

Overview

This paper uses a large subset of the MOSAiC radiosounding data to identify the top of the atmospheric boundary layer, and train a machine learning model using profile data from multiple instruments (microwave radiometers, infra-red spectral radiometer, ceilometer backscatter, and radiosonde wind profiles) to retrieve the ABL depth. The model is tested against a different subset of the MOSAiC data, and an independent data set from the north slope of Alaska, and with both all input data sources, and with individual data types excluded. Testing is evaluated overall and for a variety of conditions (clear skies, clouds, and different stabilities).

Overall the machine learning model does a good job of identifying ABL depth, with accuracy being condition dependent — most effective under clear skies, with larger errors in cloudy conditions.

One conclusion is that the boundary layer depth during MOSAiC (and presumably over Arctic sea ice in general) is controlled mainly by mechanically generated mixing at the surface. This is an important result, and one that I don’t think I have seen clearly demonstrated before. Nice work!

The paper is generally well written, but with quite a few (relatively minor) issues relating to missing technical details, or clarity of information, and to the logical flow of the narrative (see detailed comments). These should be easy to fix.

Overall I believe the paper to be acceptable for publication with only minor revision.

Detailed comments

One overarching point regarding the presentation of the performance of the ABL depth retrievals is that the various measures (RMSE etc) are given as absolute values in metres. These look good overall (eg, for the MOSAiC ‘all’ evaluation, RMSE = 42.4m, bias = -10m), but the significance of the uncertainty/bias depends very much on the absolute depth of the boundary layer. A 42m uncertainty is a minor issue for a 1000m deep BL, but much more significant for a BL of 100m or less. It would be useful to also give uncertainties as fractions of BL depth, and to note any functional dependence on depth if present.

Line 7 – the construction of this sentence is a little odd, “…remain scarce, and ABL structure…” – the two statements either side of “ , and “ are both true, but completely unrelated, while as written the grammar suggests the second statement is in some way dependent on the first. Better to rephrase something like “The Arctic boundary layer (ABL) structure is governed by stability, clouds, and wind shear, but is poorly constrained because of the scarcity of in situ observations”

Line 14 – “The ABLH annual cycle is found to peak in…” – this is ambiguous, what peaks? – the depth of the ABL, or the ‘cycle’ (of some property of ABLH)? -> “The depth of the ABL peaks in May, and has minima in August and December…”

Line 46–47 – the statement: “persistent surface-based temperature inversion and longwave radiative cooling are the primary mechanisms maintaining the boundary layer in a near-neutral state” is a little misleading. A surface-based inversion is, by definition, stable, and longwave radiative cooling at the surface will promote stability not near-neutral conditions. Turbulent mixing generated mechanically by wind shear at the surface will tend to oppose this, and maintain a near-neutral condition. Note that a few lines down, at line 53, you cite the same Tjernström paper for the statement “longwave radiative cooling and the absence of solar radiation promote the widespread formation of stable boundary layers” – same process (longwave cooling) but different result.

Line 51 – “…warm and moist air masses transport northward…” -> ‘transport’ should be ‘transported’ or better ‘advected’

Line 55 – “…atmospheric stratification stability…” -> “…atmospheric stability…”

Line 66–68 – this statement start “Several observational <u>techniques</u>…” but then lists <u>instruments</u>. The instruments on their own, or even the measurements from them, don’t constitute <u>techniques</u> – the measurements must be processed in some way to derive BL depth; and multiple processing/analysis techniques may exist for any given measurement type. Rephrase something like “ABL depth can be diagnosed from a variety of measurements from instruments such as…”

Line 74 – “…a few tens of metres to less than 100 m” – 100m here is meant as an upper limit, so “…a few tens of metres to of the order of 100 m”. Current phrasing is unclear since a few tens of m is definitely less than 100 m, so both lower and upper limits of range are sort of the same, and upper limit vague.

Line 76–77 – the sensitivity of infra-red radiometers to the structure above clouds is generally none rather than just ‘substantially reduced” – once liquid water path is above the black-body limit for IR (40-50 g m^{-2}) the cloud is opaque.

Line 114 – a reference for the ARM INTERPSONDE product would be useful here.

FIGURE 1 – the ship track shown on figure 1 is a little misleading since it makes no distinction between the ice drift and the transits (where there are very limited measurements). See Shupe et al. (2022)(fig 2) for a better approach, distinguishing between ice drift and transits. It would also be useful to show the ice extent.

Line 127 – ‘…unified time and height grid’ – the time resolution (1 minute) is given, but there is no information on the vertical resolution given, please provide this.

Line 130 – ‘…the Richardson number method.’ – on its own this doesn’t provide the reader with enough information to know exactly what was done, there are multiple possible approaches. Is this a gradient Richardson number? A bulk Richardson number? What is vertical resolution for former and lower/upper limits for latter. If following a previous study then give a reference.

Table 1. This might be better moved to an appendix of supplementary information. For most readers the precise details of the channels used will be of little relevance.
Also, in all the column headings ‘wavenumber’ has a typo and misspelled ‘wavenymber’

Line 171 – again for the INTERPOSONDE product we’re told that the winds are provided at ‘100 vertical levels’ but not what the vertical resolution is, or what vertical range those 100 levels span – is that all the levels, or just a low-atmosphere subset.

Line 181–184 – the list of features used to enhance the model’s ability is rather general, e.g. ‘characteristic height information…from backscattered profiles’, but insufficient for a reader to repeat the analysis. While a detailed explanation of all of these features might be too long to include in the main text, it is relevant. I would encourage its provision as supplementary information.

Line 197 – ‘zs’ – the s should be subscript.

Figure 2. – the figure shows several example profiles of temperatures, wind and humidity from radiosonde profiles with the height of the ABL top, derived from the Richardson number, indicated. It would be really useful to the reader to also show the profile of Richardson number from which the ABL top is diagnosed.

Line 268 – it is not clear from the text whether the RMSE’s given here refer to the test data set of the validation data set. It can be inferred they apply to the test data only after reading further, but I found it easy to confuse the two when looking back for values referred to in later discussion. Make this explicit.

Line 287–290 – the logic of the statement “*In particular, under foggy and cloudy conditions, the RMSE increases by 30.0 m and 21.9 m, respectively, indicating that the mechanical mixing and vertical-shear information provided by wind profiles are key constraints for identifying the boundary layer top under stable boundary-layer and low-visibility conditions*” seems flawed – it implies the importance of the information from the wind profiles is demonstrated by a reduction in the accuracy of the retrieval. I agree the information from the wind profiles is important, but it should be unaffected by low visibility. I think what is meant here is that the importance of this information is demonstrated by the fact that while the retrieval performance is degraded, it is still reasonable, even where important information on thermodynamic structure (e.g. IR spectrometer profiles in low visibility conditions) is lost, and the algorithm relies much more on the winds.
Consider rephrasing.

Figure 7 – it would be interesting here to also indicate the height of the ABL top derived directly from the Richardson number profiles, maybe just at the times of radiosondes.
Y-axis label for panel (d) ‘PS’ should be ‘P_s’

Line 367 – “In spring and summer, higher near-surface temperature generally corresponds to higher ABLH…” – should make a distinction here between a higher (seasonal) mean temperature, and a higher temperature within the diurnal cycle (which is what is really referred to here). As stated it is ambiguous as to which is meant.

Line 369–371 – “In autumn, the diurnal pattern becomes less regular, suggesting that the boundary layer gradually shifts from being thermally dominated to being jointly controlled by thermodynamic, dynamical, and radiative processes.” – this is not obvious to me. Why is The ABLH diurnal cycle here considered ‘less regular’ than that in spring? Both show significant variation over the day. Or do you mean a less obvious relationship to temperature?
I find the diurnal temperature record here a little odd – for all other seasons the cycle looks more or less continuous across midnight (the jump from 2300 to 0000), but here there is a large temperature jump across that 1-hour boundary…is that real?

Figure 4 – How are the seasons defined here? The simple, and widely used 3-month groupings June/July/August = Summer, Sept/October/Nov = Autumn, don’t map neatly to the changes in surface energy budget in the Arctic. Shupe et al. (2022) define seasonal periods (and transitions) via a consideration of the energy budget and meteorological conditions; this has seasons of unequal lengths which might then impact statistical robustness (OK, but needs acknowledging if similar here).

And a sanity check since it is not stated. These diurnal cycles are based on local times not the UTC time given in all the archived data files…right?

Line 381 – you divide the results into 3 subsets, ‘coupled, CBH-above-ABLH, and ‘ABLH-above-CBH’. It is not clear what the definition of ‘coupled’ is here.

Figure 11 – this is an interesting and informative analysis, but I wonder what the rationale is for choosing latent heat flux as one of the independent variables that may influence ABLH rather than (or as well as) sensible heat flux. Latent heat flux is almost always small in the Arctic, and has a small influence on convective forcing than the sensible heat flux. Where convection is a potential influence, perhaps the buoyancy flux (from the virtual potential temperature flux) is more relevant, including the effects of both sensible and latent heat.

Line 417 – “suggesting that a relatively warmer surface favors boundary-layer development” – this statement relates to the relationship with T_s-T_2, so is misleading as given, implying a relationship with absolute surface temperature rather than near-surface temperature gradient.

Line 418 – “Latent heat flux shows a moderate correlation with ABLH (r = 0.38), which may reflect the promoting effect of enhanced surface-atmosphere coupling under moist or cloudy conditions on boundary-layer deepening” – this seems unlikely. Under moist conditions (and in the presence of low cloud the near-surface will be relatively moist) we would expect a <u>smaller</u> latent heat flux. A higher flux would be expected in low humidity conditions.

Line 423-428 – the discussion of the relationship between low level jets and ABLH claims the LLJ’s promote deeper boundary layers. The arguments are physically reasonable, and I think this probably is the case…but. The discussion lacks any recognition of potential confounding – or at least complicating – factors. Figure 12a, and the associated discussion, link jet core speed with ABLH. There is certainly a strong correlation, but no recognition that jet core speed is dependent on the background wind speed, and a strong background wind drives a higher wind stress, stronger surface mixing, and a deeper ABL. Figure 12a is more convincing, but there remains the possibility that jets might be more common for deeper boundary layers, biasing the statistics.

Overall, I think LLJs probably do contribute to maintaining deeper boundary layers, but the discussion needs more nuance.

Line 437 – the discussion here refers to the no-microwave-band model, but it’s not stated why this is used…is it simply that no microwave radiometer is available for this data set?

Line 444–445 – “As shown in figure 13…” – you mean figure 4, the values quoted are for the MOSAiC validation set in figure 4.