Comments on the manuscript

Boundary-Layer Height and Surface Stability at SMEAR-II, Hyytiälä, Finland in ERA-5 and Observations

Submitted by

VA Sinclair, J Ritvanen, G Urbancic, I Statnaia, Yu Batrak, D Moisseev, and M Kurppa

for publication in **Atmos. Meas. Tech. (AMT)**

General Assessment

The paper deals with the estimation of the atmospheric boundary layer height (ABLH) and surface layer stability for the Station for Measuring Ecosystem-Atmosphere Relations (SMEAR) in Hyytiälä, Southern Finland. This station has been in operation for about 25 years with the aim to study land surface atmosphere interaction processes including trace gas and aerosol measurements. The station is part of the European Research Infrastructure Networks ICOS and ACTRIS, its comprehensive characterization with respect to atmospheric mixing potential is thus of high practical relevance. Before 2018, profile measurements of atmospheric variables have been performed at Hyytiälä during short-term field campaigns only using radiosondes, a microwave radiometer has been installed at the site in 2018. To create a long-term ABLH climatology for the site, the authors therefore use ERA-5 data which are validated against radiosonde and microwave radiometer (MWR) observations for two shorter time periods, unfortunately the reference radiosonde and MWR measurements do not cover the same time period.

The paper is well written and well structured. The approach chosen follows a sound methodology, and the results obtained are comprehensively presented in the manuscript. The paper is of limited originality with respect to the methods applied and the relevance for progressing measurement techniques. It relies on the application of methods adopted from literature, not all of them are considered as appropriate. Findings regarding the strengths and weaknesses of the different methods to derive the ABLH from radiosonde and MWR are in line with previous studies reported in the literature. Data Analysis of the ERA-5 data is, to my opinion, the stronger part of the paper. The observations part could be shortened, e.g., I am not sure whether the case studies section is really needed and whether all the experimental methods need to be discussed extensively considering their (known) limitations and methodical differences to the approach used in the ERA-5 analysis. In addition, I see a number of specific deficits and minor issues the authors should address, before the manuscript might be finally accepted. These are specified below.

Specific comments

1. A few more experimental and methodical details should be given, e.g.
    - Which type of radiosondes was used during the 2014 experiment? I am not aware of any radiosonde that measures dry and wet bulb temperature.
    - Was the 12 UTC radiosonde really released at 12 UTC? (in the operational praxis of Met Services the so-called 12 UTC sounding is often performed an hour earlier or so to ensure that a profile covering the whole troposphere is available at the reference time even in case of a necessary second ascent if the first one fails at low altitudes) – this would affect the discussion in line 625ff.
    - The authors mention two sonics at 23 m and at 46 m, but it remains unclear which of the sonics was used to derive stability?
    - Why did the authors use a temperature from within the canopy as a reference temperature?
    - The definition of the stability class ranges in lines 178-180 (and also on p. 10) differs from what is indicated in Figure 6, 7 and in Tables 3, 4.
    - The authors work with quite a number of different stability estimates (from the sonic, from the LL10 method, from the MWR output) – all these methods rely on different variables and represent different scales – have the results from these methods ever been compared and critically assessed?
    - For the temporal assignment it appears a bit irritating in several places that UTC is translated to local legal time instead of local solar time since ABL dynamics is related to the latter rather than to the former.

2. A critical discussion on the limitations of the measurements and methods is largely missing – just a few examples:
    - According to Figure 1, the sonic is positioned at the edge of an about 1:12 slope with a lake and some larger clearings just a few hundred meters to the west. Does this have any influence on the representativeness of the sonic data? How can it be justified that a local z/L value close to a forest canopy is appropriate to characterize atmospheric stability up to a height of several hundred meters. Did the authors perform any comparison of, e.g., the stability derived from the eddy covariance measurements compared to the one derived from the soundings in the frame of the LL10 method – see above?
    - Both the parcel and Ri number methods are quite sensitive to the surface temperature value used in the analysis, and it is always a matter of discussion which data to use here. A near-surface temperature measured independently in a shelter often does not fit to the radiosonde profile, on the other side the first temperature values from the radiosonde profile may suffer from the sonde being in the hands of the observer before being released and not experiencing the ventilation as during the free flight. In some studies, an excess temperature is introduced or a limit for a possibly superadiabatic near-surface lapse rate is set. These issues should be addressed in the methods description.
    - In lines 522-523, the authors bring in another ABLH estimate which is based on a completely different variable than all estimates considered in this paper. I suggest to omit this,

otherwise it would call for a critical discussion of ceilometer ABLH estimates which is a delicate task itself.

- Figure 9 indicates quite a substantial part of missing data, even in summer. For winter the authors argument with icing of the sensor, what are the possible reasons in summer? Are these missing data assumed to introduce a bias in the distribution of stability classes?

3. The authors in their analysis mix methods that are based on the consideration of different physical processes, they thus compare apples and pears naming all these derivatives ABLH without a critical discussion on that. This necessarily will result in some dis-agreement when comparing the results. It is obvious that Heffter's method is basically suited to obtain an estimate of the convective ABLH, and previous studies have already demonstrated the tendency of this method to overestimate ABLH. From the MWR, in stable conditons, they derive a kind of inversion or stable layer upper boundary, that is purely based on an analysis of the temperature profile which contains limited information on shear-induced turbulence, if at all. For the ERA-5 analysis, a Ri number method is considered as appropriate, it would thus be consistent to put emphasis on those experimental data which provide a comparative ABLH estimate.

4. The presentation of results in Sections 5 and 6 is widely descriptive, I would love to see some discussion of the findings already here, even if it is partly given in Section 8, e.g.,
   - Why is ERA-5 in better agreement with Ri(0.5) from radiosoundings for the very unstable case, but with Ri(0.25) for the weakly stable case?
   - Why are very shallow BLs less common than moderately shallow BLs in winter?
   - Why is the ABL height variability so large in April and May?
   - How to explain the annual cycle in the occurrence frequency of the weakly stable class (line 620ff)?

5. Deficits exist concerning the discussion of the authors work with respect to previous studies on, e.g., ABLH climatology at higher latitudes. How do the authors results behave when compared to the cited studies of Liu and Liang (2010) and Seidel et al. (2010, 2012)? I would also like to bring to the author's attention the work of Lotteraner and Piringer (Boundary-Layer Meteorol. 161 (2016), 265-287 – concerning the deficits of Heffter's method), as well as of Beyrich and Leps (Meteorol. Z. 21 (2012), 337-348 – concerning the analysis of radiosonde data: many of the authors results on methods comparison, annual variability and uncertainties have been reported there as well).

6. For the Figures, I noticed some issues concerning either the interpretation in the text or the methods chosen to display the data:
   - Figure 2 (related to text line 295ff): It appears not really possible to reproduce the ABLH values given in the text from Figure 2. At 00 UTC, I would not see any of the values at 307 m – all the lines appear to be well below 1/4 of the first height axis tick label at 1000 m. The same is with the inversion at the 06 UTC plot which appears to be well below 500 m. And at

12 UTC, the lowest plotted estimate is definitely below 1000 m whereas the text gives the minimum value with 1073 m.

- Figures 5-7, 9, 11: I am not sure whether the choice to plot the whiskers at Q1 – 1.5 IQR and Q2 + 1,5 IQR is a clever one given the skewed distributions of the variables, because this does not consider the real distribution of data points in the lowermost and uppermost quartiles. It may thus happen, that – especially the lower – whiskers cover a range where no data points occur. This becomes obvious by the many whiskers for ABLH ending at z = 0 m. I probably would have chosen to represent the 10% / 90% percentiles by the whiskers.
- Figure 10: It does probably make not much sense to average the mean diurnal cycle of the stability classes occurrence frequencies over the whole year as the seasons behave very different.
- Figure 12: I am not sure whether the correlation coefficient is a suitable measure for this type of analysis, at least if the authors chose this as the only parameter considered. From the seasonal distributions we have learned that derived ABLH values vary over a wider range in summer than in winter, this automatically gives higher correlation coefficients if at least the general tendencies are the same in the two data sets considered. Wouldn't it be an idea to consider absolute or relative differences here as well, the more since for practical applications it might be of higher relevance to know the area over which ABLH does not differ by more than a certain absolute (say 100 m or 200 m) or relative value (say 20 %), rather than to know correlation coefficients. Similarly, I doubt whether just from the value of the correlation coefficient one may judge about the spatial variability (lines 680f).

Minor Issues

- The abbreviation SMEAR is introduced in the title already and it occurs in the abstract again before it is finally explained in line 73.
- The abstract could be shortened a bit. E.g. the first two sentences might be deleted, they would well fit in an introduction section, but are unnecessary in an abstract. I also consider it for unusual in an abstract to write "for example …" – either the result is so relevant that it should be stated or it might be omitted.
- Line 36-37: Here the authors state that vertical profiles of temperature and wind speed are required to determine the ABLH, then they write that "these profiles can be obtained from … ceilometers." I do not know any ceilometer that would provide wind and temperature profiles.
- Line 45-46: Another limitation of remote sensing instruments has to be seen in the fact that often more than one instrument is needed to measure the relevant variables, while tower measurements or sondes may provide the full set of essential thermodynamic variables.
- The authors do not seem to be friends of using commas to structure a sentence, this makes it a bit difficult to read in some places (e.g., line 74, 632, 709, 737, and others)
- Line 119 / line 129: If M is the "main SMEAR station location" (line 110) in Figure 1, I would see the location R to the Southwest of M, not to the Southeast.

- Line 207: “Stable boundary layer are not characterized by inversions.” – This is a statement I simply do not understand.
- Line 246: Isn’t it a bit dangerous to consider just the immediate neighboring levels below AND above a given height to diagnose a low-level-jet. Often the jet nose may be not that sharp that there is a significant change of wind speed both below AND above the height of the maximum?
- Line 269: The use of z0 is a bit misleading here, z0 is normally taken as the aerodynamic roughness length.
- Line 278-280: The Vogelezang and Holtslag method additionally considers a correction term to account for surface friction / shear.
- Line 312: It is not clear to me why – “due to the inclusion of the wind profile … the Ri number methods diagnose shallower SBLs”. Additional consideration of the wind profile would add shear-induced turbulence. Parcel methods will give ABLH where $\theta h > \theta 0$ while Ri methods will lift this up before Ri exceeds a value of Ri = 0.25 or Ri = 0.5. In fact, parcel methods are equivalent to Ri methods with Ricrit = 0, and with critical Ri values above zero they should give higher ABLH values.
- Line 327-328: I cannot follow this argumentation. The wind speed differences with respect to the near surface values are larger in the measurements when compared to the ERA-5 profiles up to a height of about 2.5 km, consequently the denominator in the Ri definition gets larger, and hence Ri gets smaller. It thus appears not logical that the threshold of Ri = 0.25 should be exceeded at lower heights.
- Line 370: I would prefer to call this RMSD instead of RMSE. To use the word “error” would imply that the truth is known based on a well-calibrated method.
- L406-L408: How to explain the outliers in Figure 4 where Ri(0.5) < Ri(0.25)?
- L458: I would not a-priori expect that the Ri(0.25) methods radiosonde data must be in better agreement than the Ri(0.5) method with the ERA-5 ABLH estimate, even if the latter uses the same threshold: Note that the profiles have different vertical resolution and the reanalysis profiles are usually much smoother than any measured profile which will strongly influence the level of exceedance of fixed threshold values.
- Line 466-467: 23 % is a quite accurate number, I would not call that “approximately”.
- Line 609-610: This seems to be a bit too general, the very stable cases do occur more frequently in summer.
- L643: “In winter cold conditions are required for shallow boundary layers to develop” – This statement calls for some further explanation.
- L669: It appears to be a subjective view whether Figure 12 really represents a “considerable part of Northern Europe”.
- L724f: This conclusion does a bit contradict to the discussion 457-463 (“high vertical resolution … sufficient“ vs. “limited vertical resolution” explaining disagreement)
- L735: This is not a suggestion but a rather logical consequence of the stability classification with the LL10 method as described in Section 3.1.
- L742-743: Again, a minor inconsistency: Here the authors, to my opinion, correctly point at the difficulties to determine the ABLH during the evening transition both from the

measurements and in ERA-5, in Section 5.2 (line 496) they solely "blame" ERA-5 for the deficits.

- L767f: I also see the widely missing diurnal cycle (no pronounced daytime heating) as an additional reason for the reduced occurrence of very stable situations in winter, except of the more frequent occurrence of clouds.
- Figure 2: Relative humidity is not a good choice to plot vertical humidity profiles for illustrating the ABL structure, absolute or specific humidity would be better here. In essence, the humidity profiles are not even needed to be displayed here, at least they are not discussed in the text.
- Figure 4: The diagrams give the impression that some lower boundary values exist for the LL10 method (no stable values below 40 m, no unstable values below ca. 150 m) – nothing has been mentioned about that when describing the method.
- Figure 5: The black points are black crosses.
- Figure 8: It appears unnecessary to explain the legend again in the text of the Figure caption.
- Figure 9, caption: Just to make sure: panel (b) is stability from the sonic or from ERA-5? The data period represented by the graphs should probably be mentioned in the Figure caption, the more if it is not the same (ERA-5 40 years vs. sonic 19 years, as in Figure 10?)
- Table 3, caption: NMSE must be nRMSE, in the Table NRMSE should be nRMSE according to the definition given in the text (but see my remark on RMSE vs. RMSD)

And a few misprints:

- Line 34: assumption
- Line 104: years
- Line 185: delete "can"
- Line 294: suggests
- Line 326: value
- Line 437: in → is
- Line 451: methods
- Line 459: a systematic bias
- Line 582: delete "the" at the beginning, and delete "monthly" after "month"
- Line 659: BLs instead of BLS
- Line 735: This suggests …
- Line 767: clouds
- Caption Table 1, 2nd line: coefficients
- Figure 6, legend: Heffler must be Heffter

Recommendation: Major Revision