Manuscript: amt-2023-80
Title: Deep-Pathfinder: A boundary layer height detection algorithm based on image segmentation
Revision: round 2

Dear Reviewers:

Thank you for your time and effort to carefully assess the revised version of the manuscript. In this second round, the manuscript has been thoroughly revised based on all remaining comments.

The following pages provide our detailed responses.

Kind regards,
The authors

Reviewer 1

Summary: *After the review, this paper improved significantly. I recommend to publish.*

Thank you very much for your feedback.

Reviewer 2 (report #3)

General comment: *The Authors explain a lot to the reviewer, but not to the reader of the paper, i.e. they do not put their arguments in the revised paper. Several changes are mentioned several times in the reply. This is of course ok if it fits, but one may get the impression that a lot of changes happened. Most of the changes are adaptions within the sentences, often just a few words.*

I am mostly satisfied with the adaptions around the description of the algorithm and the training, fine-tuning and validation. What has been also clarified is what part of the dataset has been used for what. And i think the reorganization of section 2 is also helpful.

Thank you very much for your feedback.

A point-by-point response to your remaining feedback is provided below.

1. Clouds (point 5,):

   I suggested that the authors should repeat the Deep-Pathfinder - STRATFinder comparison with no or few clouds to investigate whether clouds have an impact on the retrievals. This could give insights how to deal with clouds in BLH retrievals. The authors provide a longer detailed discussion to the reviewer based on the case of fig.6e, but instead of explaining this in the articles text they add just one sentence to the manuscript. The authors miss the opportunity to prove or disprove on a statistical basis that clouds might pose a problem.

The suggested analysis has now been performed. For each day in the test set, the percentage of time was computed that clouds were present in the input image. This is referred to as the cloud overcast fraction. Subsequently, Deep-Pathfinder and STRATfinder performance was compared for different ranges of cloud cover.

The following text and table with statistics have been added to the manuscript:

“These daily fluctuations can be partly explained by the amount of cloud cover. To illustrate this point, the daily cloud overcast fractions were computed for all dates in the test set, looking only for clouds below 2245 meters (i.e., the vertical range captured by our model). Table 3 shows that on days with no or few clouds the Deep-Pathfinder and STRATfinder algorithms were more closely aligned, based on Pearson correlation and mean absolute difference statistics.” (see lines 309–312)

Table 3. Comparison of Deep-Pathfinder and STRATfinder estimates for different ranges of cloud cover in the test set.

		Cloud overcast range		
	Overall	0–10%	10–30%	30–100%
Number of days	161	27	24	110
Pearson correlation	0.706	0.811	0.819	0.632
Mean absolute difference (m)	189.0	141.6	176.7	205.3

Finally, we indeed did not list all changes we made to the manuscript during the first revision in the response to comment #5. This could have been stated more clearly. For completeness, the changes describing the most important observations related to the case of Fig. 7e are listed below.

“In case of multiple cloud layers, our annotations typically followed the lower layer, while STRATfinder followed the higher layer. For example, this behaviour can be observed in Fig. 7e.” (see lines 322–323)

“The example of Fig. 7e shows that for multiple cloud layers Deep-Pathfinder and STRATfinder typically followed a different layer. Hence, in case of multiple cloud layers, users should be aware that the methods may produce different MBLH estimates.” (see lines 298–300)

2. Comment number 6:

   This is my remark that //there is no night time mixing layer// which was also brought forward by reviewer 1. I repeated this at several places in my comments. The Authors somehow agreed on that - but do not mention the RH/backscatter problem in the revised manuscript. And they refused to drop the data nor change the naming ('...mixing layer...') because they wanted to be "consistent with former papers." In other words they agree that it has been done wrong in the past but they want to go on with that.

   They discussed at no place the argumentation that with high relative humidity aerosol particle grow in size, backscatter increases and that this may result in a layer detection where no layer is. They could say that this mechanism exists and may result in faulty layer attribution - but they did not. In my opinion this argument must go in the paper. It should be noted that this high humidity layer does not even necessarily falls together with the stable nighttime surface inversion.
   If they want they can state in the paper that further investigation is beyond the scope of the paper - as they wrote somewhere in their reply. But this in an essential problem with backscatter based layer detection and it must be suspected that also artificial intelligence and computional vision is not able to get around this problem. The authors should discuss this.

First, we would like to mention that as part of the first revision, we incorporated in the manuscript that there is no mixing at night. This now reads as follows:

“During nighttime, the pollution rich layers may drop to very low altitudes into the incomplete-overlap region of the CHM15k ceilometer and vertical mixing in fact predominantly ceases to exist. However, the concentration levels of pollutants remain layered and, therefore, Cabauw mast measurements were used to aid in the identification of the presence and height of nocturnal layers.” (see lines 176–179)

The following has now also been added to the manuscript:

“Finally, with high relative humidity aerosol particles grow in size, leading to increased backscatter which may result in a layer detection where no layer is (i.e., faulty layer attribution). The high humidity layer also does not necessarily coincide with the stable nighttime surface inversion, meaning MBLH retrieval during night by use of ceilometer backscatter data can be strongly biased. Further investigation of these mechanisms is beyond the scope of this paper. For further reading, refer to Kotthaus et al. (2023, section 3.3.2 and references therein) or Collaud Coen et al. (2014).” (see lines 54–59)

References

Kotthaus, S., Bravo-Aranda, J. A., Collaud Coen, M., Guerrero-Rascado, J. L., Costa, M. J., Cimini, D., O’Connor, E. J., Hervo, M., Alados-Arboledas, L., Jiménez-Portaz, M., Mona, L., Ruffieux, D., Illingworth, A., and Haeffelin, M.: Atmospheric boundary layer height from ground-based remote sensing: a review of capabilities and limitations, Atmospheric Measurement Techniques, 16, 433–479, https://doi.org/10.5194/amt-16-433-2023, 2023.

Collaud Coen, M., Praz, C., Haefele, A., Ruffieux, D., Kaufmann, P., and Calpini, B.: Determination and climatology of the planetary boundary layer height above the Swiss plateau by in situ and remote sensing measurements as well as by the COSMO-2 model, Atmospheric Chemistry and Physics, 14, 13 205–13 221, https://doi.org/10.5194/acp-14-13205-2014, 2014.

3. Figure R1:

   The authors provided a figure R1 in the response showing the course of backscatter, Temperature, specific humidity etc. from the Cabauw tower for one of the cases they presented in fig.6. I suggested in my review that they should investigate these parameters because I hoped that it would become obvious to them that their night time BLH is not correct. They present the plot and say in the response just very short:

   > "we can see that in this example the manual annotation was mostly in line with thermodynamic definitions.
   > Note that this figure has not been added to the manuscript."

   No arguments or explanation why they think that this plots support their retrieval. They changed the text, but they just changed one single word i.e. //correctly// to //as intended// (why do they talk here about their 'manual annotation' and not about the retrieval ?)

   But there is more than a missing argument.
   The temperature axis in this figure is annotated with:

   > "Potential Temperature * [K] *(my own estimative using metpy) "

   I.e. the author was not sure whether it is correct. I am sure he can find a colleague at KNMI who can support him in finding an exact instead 'estimative' approach.

If you ask me that cannot be Potential Temperature (Tpot) but instead it is rather air temperature (Tair) in Kelvin.

In the first hours 00-03 UTC , i.e. in the middle of the night, the temperature at 2m (grayish) is by more than 1K warmer than at 200m (blue). Even during daytime in summer under clear sky that would be an exceptional unstable stratification. If I would believe that this is Tpot I would expect a *mixing*layer to reach several thousand meter, but not only 300m as found by the retrievals. If you assume it is just Tair in K and add as an estimate deltaT from the adiabatic gradient (deltaT = +z*1K/100m) to calculate Tpot you get stable stratification with a stronger gradient at the surface - as you would expect during night. At about 3:30 this gradient inverts (advection of cold and dry air) - from then on T(2m) (grayish) is colder than T(200m) (blue) and T(140m) (orange) - i.e. if we take this temperature as Tpot we have stable stratification, no mixing and thus mixing layer height = 0m. And if we assume that this is not Tpot and add deltaT the gradient becomes even larger. But the BL retrievals find a BL of 150-170m - way higher.

I could go on with arguments like this for the whole day ...

So: the authors cannot use this plot as an argument that their retrieval works well - not if it would be Tpot and not if it is, as i suspect, Tair.

It is obvious that Temperature presented in figure R1 can not be potential temperature, but is rather something at least close to air temperature. Accordingly their statement that "...manual annotation was mostly in line with thermodynamic definitions. " is not supported by this figure. It rather supports the reviewers argument that a boundary layer height (BLH) retrieval during night by use of ceilometer backscatter data can be strongly biased. The reason for this is probably the RH-aerosol growth-backscatter mechanism described by the reviewer.
They should discuss this argument in the manuscript.

Figure R1 indeed contained a mistake related to the potential temperature. We thank the reviewer for noticing this and for the opportunity to correct this figure. For the record, a corrected version of this figure is added below (figure R1_v2).

This figure has not been updated in the manuscript, as it does not appear in the manuscript (i.e., it only appeared in the response to reviewers document).

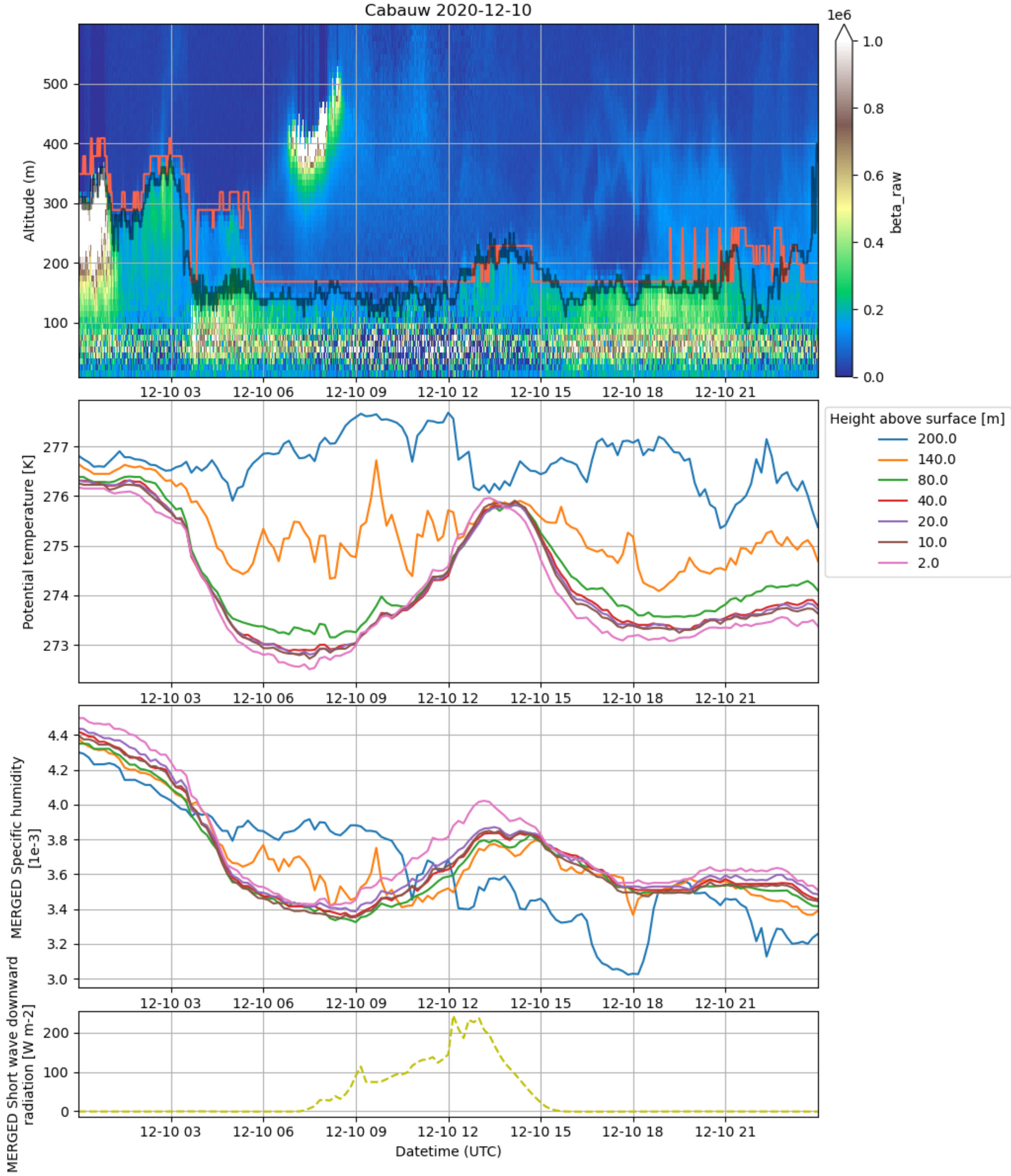

Figure R1_v2: Case study at Cabauw on 2020-12-10. From top to bottom: RCS data with Deep-Pathfinder (black line) and STRATfinder (orange line) retrievals, potential temperature, specific humidity, and short wave downward radiation.

To resolve this comment, the discussion of Fig. 7f in the manuscript has been updated as follows:

"When a clear CBL was not apparent (e.g., Fig. 7f), Deep-Pathfinder and STRATfinder obtained similar estimates, although in Fig. 7f both were far above the stable nighttime surface inversion." (see lines 300–301)

Finally, the reviewer suggests to discuss that a boundary layer height retrieval during night by use of ceilometer backscatter data can be strongly biased, and that this may be caused by the RH-aerosol growth-backscatter mechanism.

As also noted in the response to comment #2, the following has now been added to the manuscript:

"Finally, with high relative humidity aerosol particles grow in size, leading to increased backscatter which may result in a layer detection where no layer is (i.e., faulty layer attribution). The high humidity layer also does not necessarily coincide with the stable nighttime surface inversion, meaning MBLH retrieval during night by use of ceilometer backscatter data can be strongly biased. Further investigation of these mechanisms is beyond the scope of this paper. For further reading, refer to Kotthaus et al. (2023, section 3.3.2 and references therein) or Collaud Coen et al. (2014)." (see lines 54–59)

References

Kotthaus, S., Bravo-Aranda, J. A., Collaud Coen, M., Guerrero-Rascado, J. L., Costa, M. J., Cimini, D., O'Connor, E. J., Hervo, M., Alados-Arboledas, L., Jiménez-Portaz, M., Mona, L., Ruffieux, D., Illingworth, A., and Haeffelin, M.: Atmospheric boundary layer height from ground-based remote sensing: a review of capabilities and limitations, Atmospheric Measurement Techniques, 16, 433–479, https://doi.org/10.5194/amt-16-433-2023, 2023.

Collaud Coen, M., Praz, C., Haefele, A., Ruffieux, D., Kaufmann, P., and Calpini, B.: Determination and climatology of the planetary boundary layer height above the Swiss plateau by in situ and remote sensing measurements as well as by the COSMO-2 model, Atmospheric Chemistry and Physics, 14, 13 205–13 221, https://doi.org/10.5194/acp-14-13205-2014, 2014.

Reviewer 3 (report #2)

Summary: *I believe the authors have answered most of the reviewers' comments, and the revised manuscript is substantially improved with respect to the first version.*

Thank you very much for your feedback.

A point-by-point response to your comments is provided below.

1. However, I feel the authors have not clearly addressed comment #7 of Reviewer 2.

   In particular, I see a potential contradiction in their breakout answers, i.e. "We do not claim that we apply an overlap correction", and then "with instrument-specific overlap correction".

   The authors should remove the potential contradiction and clarify their arguments.

The cause of this confusion is that there are two types of overlap corrections. First, the CHM15k instrument has a built-in overlap correction. This first overlap correction is not perfect, as also noted by reviewer 2 in comment #7. Second, some studies therefore use an additional overlap correction when processing the data from the ceilometer (e.g., Hervo et al., 2016). For clarification: we use the built-in overlap correction, but no additional overlap correction.

To clarify our initial response to comment #7, this should be rephrased as: "<u>Besides the built-in overlap correction</u>, we do not claim that we apply an <u>additional</u> overlap correction"

Further, the following sentence has been inserted in the corresponding paragraph of the manuscript for clarification:

“Note that prior research has indicated that the built-in overlap correction of the CHM15k is not perfect (Hervo et al., 2016).” (see lines 126–127)

Reference:

Hervo, M., Poltera, Y., and Haefele, A.: An empirical method to correct for temperature-dependent variations in the overlap function of CHM15k ceilometers, Atmospheric Measurement Techniques, 9, 2947–2959, https://doi.org/10.5194/amt-9-2947-2016, 2016.

Minor comments:

2. In addition to Milroy et al. (2012), I would suggest the authors refer to Collaud et al. 2014 for insightful discussion on the features of MBLH estimated from different instruments (openly accessible at: www.atmos-chem-phys.net/14/13205/2014/)

The suggested reference has been added to the manuscript:

“These methods are complementary, as each approach has different advantages and limitations for capturing certain features of the MBLH (Collaud Coen et al., 2014).” (see lines 37–39)

Reference:

Collaud Coen, M., Praz, C., Haefele, A., Ruffieux, D., Kaufmann, P., and Calpini, B.: Determination and climatology of the planetary boundary layer height above the Swiss plateau by in situ and remote sensing measurements as well as by the COSMO-2 model, Atmospheric Chemistry and Physics, 14, 13 205–13 221, https://doi.org/10.5194/acp-14-13205-2014, 2014.

3. Table 3: In addition to UTC, I suggest to explicit local standard time (LST) as well, as the latter is more meaningful for the diurnal cycle.

The following has been added to the caption of this table:

“Time of day is stated in UTC; note that the local standard time at Cabauw is UTC+1 or UTC+2 (daylight saving time).” (see page 16)