

Interactive comment on “Pathfinder: Applying graph theory for consistent tracking of daytime mixed layer height with backscatter lidar” by Marco de Bruine et al.

Anonymous Referee #2

Received and published: 9 December 2016

Manuscript: amt-2016-327

Manuscript: amt-2016-327

Recommendation

Major revisions (mainly because the modifications would require quite some work, not because the paper is bad).

1

General

- The paper presents a new method to derive mixed layer height (MLH) from profile data from backscatter lidars. The novelty of the method lies in the use of information from profiles that are adjacent in time. As the MLH is a variable that has a continuous evolution in time, it will only vary by a limited amount from time to time. The use of this temporal information is implemented using a graph method: the most plausible evolution of MLH in height and time is determined.
- The method is tested for a short data set (a 12-day campaign) for which a number of cases are studied in more detail in order to highlight the strengths and weaknesses of the method. For this period a range of reference methods is used. Furthermore, the method is tested against a full year of MLH observations that were derived manually from the same backscatter lidar data.
- The paper is written clearly, in terms of language, structure, figures and description of method and results.

However, I do have some comments:

- a. The authors explicitly decide to apply the method to a time window of 15 minutes, and at a time resolution equal to that of the underlying data (30 seconds). This implies that their method tracks short-term, turbulent, variations of the MLH. This in itself provides interesting data that would be beneficial to study the dynamics of boundary layer turbulence. But for most applications that the authors refer to the *slow* variation of the MLH (at time scales above turbulent time scales) is more important. What I miss is a thorough analysis of the length scales and time scales involved in the turbulent and diurnal evolution of the MLH, and how these scales relate to the various design choices that were made

2

in the method.

(NB: I expect that the use of temporal information would be more beneficial at those time scales: since MLH evolves slowly, more narrow bounds can be used for plausible evolution. Hence, I wonder why the authors choose to look at these short time scales. Perhaps it would be beneficial to first use longer term averages of lidar data (e.g. 10 minute averages) in the same Pathfinder method to track the MLH within a single window spanning the entire day (that would yield a window of about 100 points for a full day, computationally still doable). Subsequently, based on the envelope provided by the full-day estimate, the short term variations could be tracked).

- b. The main statistic that is used by the authors to judge the skill of their method is the coefficient of determination R^2 . However, for data that share a common causal variation like the diurnal cycle or the yearly cycle, this statistic carries little information: anything that increases during the day will correlate well with something else that increases during the day. Therefore, I would like to ask the authors to include other statistics as well (e.g. mean bias and mean absolute error).
- c. The sensitivity analyses only show that the results in MLH indeed do differ when certain parameters are changed (which I would consider a trivial result). However, it remains unclear whether changes in parameters improve, or deteriorate, the MLH estimates. E.g. if a mean positive bias were observed in the validation, it would be informative if a sensitivity analysis would show that a change in a certain parameter would lower the mean MLH. This would at least suggest that this change in parameter, although not tested with validation data, could solve the bias.

Below I will provide detailed comments

3

Note: in the comments below, the comment is preceded by the page number - line number.

Detailed comments

1. 2-1: somewhere in the introduction you need to make clear at which time scale (record length) and time resolution you intend to derive MLH.
2. 3-15: could you provide a reference for the lidar equation. I appreciate that below the equation you explain a number of the components, but it would be helpful to the reader to have a place where he/she could read about the other components of the equation.
3. 3-27: 'parts of the data will be excluded as MLH'. At this stage the matrix that is shown in figure 2 is not yet in the readers mind. Hence, it is unclear what you mean with 'data' here. Is it a time slot (but a time slot *is* not a MLH)? I guess what you intend to say is that you imagine a single lidar profile and blank out certain parts of that profile where you know/think that the MLH cannot be located.
4. 4-5, section 2.3.1. This section is missing theoretical backing of the different restrictions. The motivation of the values of height intervals etc. seem to come out of the blue. Please provide independent support for the choices you make here.
5. A number of suggestions to clarify section 2.3.1:
First, I would suggest to add a figure here that sketches the MLH with the different disturbing items in it (residual layer, cloud layers etc.). Then it is easier to visualize and explain the various restrictions.

4

Secondly, figure 1 (which is very informative by the way, well done!) mentions the detection of residual layer as a restriction. However, I could not find a specification of this restriction in section 2.3.1.

Finally, the text in section 2.3.1. could be shortened (and made more informative) if the various restrictions would be summarized in a table. Then details of numbers could be removed from the text. And the various 'thresholds' mentioned in figure 1 could easily be tracked down (I found it somewhat confusing that the figure simply mentions 'threshold' where probably each of these thresholds has a different value/meaning). Only when I arrived at the results section, line 8-20 I have to discover that there is indeed such a table. I don't understand why it is not referred to in section 2.3.1, where all these numbers are mentioned in the text as well.

6. 5-25 and further. I would like see a more thorough analysis here of the distinction between mean growth rates and turbulent growth rates. Apart from the growth rates themselves, also the time scale related to the distinction between what is *mean* and what is *turbulent* is important here. Now this distinction seems to be related/coincident with the 15 minute window size. This discussion is related to my first major comment (a).
7. 6-1: Here you apply a weight to the points (where one would expect a large weight to be related to a large number). Later on you are talking about costs (the term that probably comes from graph theory). I think it would be better to start talking about costs from the outset. Then the reader has to make the mental jump from 'important = low number' only once.
8. 6-9: Apparently, the first point in each time window is determined independently of the information that could be derived from the previous window. To me this seemed rather wasteful, as the continuity of the MLH is only retained for 15 minutes. Only in line 6-19 it becomes clear that indeed the windows are

5

coupled in time (initialization from final value of previous window).

However, this still poses a problem, when it comes to application of the method to a full day of data: the graph method is applied within the window (those points are treated as one dataset) and the subsequent windows are only connected by one point). Hence, the question is to what extent the full day evolution depends on positioning of the windows in time. It is unclear to me if the start-time sensitivity tests does this test on a full-day evolution or only on a within-window evolution.

9. 6-17: the size of the time window is discussed here, based on computational costs. But, in fact the choices also imply a decision on which time scales in MLH variation the authors want to resolve (see comment related to line 5-25).
10. 7-1: it is unclear which datasets have been used. Only in the results it becomes clear that there are 3 sets: the 12 day campaign in 2008, May 2010 and the entire year of 2010. A clear motivation for using these data is missing. Please include a specification (which dates) of the used data sets in this section.
11. 7-18: the ceilometer is mentioned here (and ceilometers were mentioned in the introduction, see remark there), but the data are not actually used (as you mention in line 23). So please remove this.
12. 7-25: The wind profiler uses very different information on the ML than the backscatter lidar. However, it is unclear whether/why the MLH retrieved from wind profiler data is superior to other methods. And hence it is unclear whether the profiler data are considered as a reference, or rather as an alternative method. Please clarify.
13. 8-16: the STRAT2D algorithm appears here at once, without earlier introduction in the methodology section. It is unclear why STRAT2D would be an important reference. Is it the *de facto* standard to derive MLH? What are the general

6

characteristics of the STRAT2D algorithm? In which way is it similar and in which ways different than the Pathfinder method?

14. 8-11: Please note that Vogelezang and Holtslag (1996) only tested the the Richardson number method for stable conditions. So please provide references that show that the method also works for neutral and convective conditions.
15. 8-18: The manual estimates of MLH have not been introduced and defined before. They should have been defined in the methodology section, rather than at the place where the results are presented.
16. 8-19: Linked to an earlier comment on the absence of a table in section 2.3.1: the fact that table 2 is referred to as a table of 'tuning parameters' of the instrument suggests that some of the 'guiding restrictions' would be instrument dependent. However, the values related to the guiding restrictions are not presented as such 2.3.1. In order to show to other researchers (with different instruments) that the method is versatile, it would be good to clarify (in section 2.3.1) which parameters are considered instrument specific, and which could be more general.
17. 10-28: The value of R^2 is probably dominated by the strong diurnal cycle in MLH (see major comment (b)). Therefore, I find the remark that the agreement is within 100 m 'almost for the complete period' more informative. Subsequently, the most interesting question is in which cases the deviation is *more* than 100 m. Apart from the type of statistic used, it is important to know whether the authors consider the wind profiler derived MLH as a reference against which the lidar results are judged.
18. 11-2: The R^2 mentioned here is more meaningful (than those based on the full diurnal cycle), since now the full range of possible MLH's is only determined by the variation of 16 UTC MLH between the 12 days. This range is probably much

7

smaller than the range of values occurring during the full diurnal cycle and hence the R^2 is more meaningful. The 5 UTC values probably suffer from the fact that the spread in MLH between days is small early in the morning. Hence, it is more difficult to obtain a meaningful coefficient of determination. The success for the 16 UTC data is due to the fact that the MLH is accumulating possible small differences between the days (in terms of insolation, partitioning between sensible and latent heat flux, subsidence). Due this accumulation effect small differences can lead to significant variation in MLH by the end of the day.

19. 11-24: The reason mentioned here for the failure of the Richardson method is related to the dilution of a rising parcel. But the Richardson method is *not* a parcel method: it only looks at the local Richardson number.
20. 11-26: 'that the radiosonde ...': clarify here that you are talking about the De Bilt sondes.
21. 11-29: please stress here that the underlying data for the manual estimate and the Pathfinder are identical. Although one could consider this a weakness (there is no real independent check), you could also consider it a strength (you only check the algorithm, not the data).
22. 12-3: I'm a bit puzzled here: is the quality criterion mentioned here part of the algorithm or not? It seems to be separated, since it can be included or excluded. Please clarify.
23. 12-13, section 4.5.1: it is unclear here whether the different 15 minute windows of a day were shifted independently, or that the whole train of 15 minute windows for a given day were shifted forward and backward. Furthermore, I think that it is more important to know how sensitive the method is to the *phase of ML development* it is started at than the location within a 15 minute window

8

(in case of ML development we are looking at shifts of much more than 15 minutes, e.g. an hour).

24. 12-24: I don't see the link between an R^2 of 0.96 and a match with the reference run for 93% of the runs. Furthermore, it is not fully clear to me whether the correlations have been determined per day, and that these correlations have been averaged subsequently (what was the order of processing?). The same remark as made earlier applies here: determining a regression between two variables that share the same diurnal cycle is not very informative. Hence, focusing on e.g. the 16 UTC values only would be a more fair comparison.
25. 12-27: I suppose that with an increasing window size the overall maximum growth rate (of 1 m/s) also was still applied to the entire window, see line 5-26.
26. 13-1: I assume that the correlation mentioned here is between the MLH with the default window size, and the alternative window sizes. Of course this correlation is below 1, since you changed something in the algorithm. But the real question should be: does the change in window size make the MLH estimates *better*. Since the analysis with the May 2010 data does not include data from a reference method for MLH, this improvement (or lack thereof) with varying window size cannot be judged.
27. 13-19: are you referring to high spatial resolution or temporal resolution (or both). Please specify what you consider high resolution (this means different things to different people).
28. 14-4: 'climatology is used'. I had the impression that for the first window of the day the height with maximum gradient was used. And that for the subsequent windows the result from the previous window was used. If climatology is used, please specify what this climatology is based on.

9

29. 14-19 to 26: this short summary of the method is not very clear. I think it would help to verbally draw the picture of a grid with height slots in the vertical and time on the horizontal axis.
30. 14-27: in the discussion of results it is insufficiently clear what length of period the quoted correlation coefficients refer to (12 day period, May 2010, full year).
31. 14-30: the $R^2 = 0.21$ is a rather random quote to show the success or failure relative to radiosonde data. You compare a value obtained by regressing values at one time of day (0.21), to the result of a regression based on data that span the full diurnal cycle (0.90). Furthermore, the 16 UTC data gave a better regression. Please paint a *complete* picture in terms of numbers, or just describe the results in words.
32. 20: the figure suggests that in one time step MLH can only move one level up or down. Does that mean that the vertical resolution of the grid is fixed to the 75 meter? I had the impression that the vertical resolution of the data is higher. So in reality more levels can be jumped? Perhaps the reason is that the figure is merely meant to show the logic of the method (which it does very well!). But at least clarify the vertical resolution of the actual grid used (in the text, section 2.3.2).

Very detailed comments

1. 1-24: the link between MLH and the surface energy balance is through the surface fluxes that *drive* the turbulence in the ML (that is: the link between surface fluxes and turbulent processes was missing in your causality chain).

2. 2-6: here at once the ceilometer occurs, which was not part of your enumeration in the sentence before. If you want to treat ceilometers and backscatter lidars as one type of instruments (the start of line 10 with *backscatter lidar* suggests that), then better use one term for them.
3. 5-28: 'However', the contrast with the previous sentence is not fully clear. I guess that what you want to say is the 0.278 m/s is valid for the *mean* growth whereas you intend to follow the *turbulent* growth.
4. 5-29: 'These': it is not clear to which conditions this refers back.
5. 10-18: to clarify the difference between sections 4.3 and 4.4, I would suggest to change the title to 'Comparison with other methods – 12 day period' (then 4.4 would become 'Full year analysis of midday MLH').
6. 10-24: 'overview lidar' → 'overview of lidar'
7. 11-8: extinguished → distinguished
8. 11-20: remove 'data used in this study'
9. 12-23: 'correlation ... of ... with' → 'correlation ... between ... and'
10. 13-10 to 14: I think the 'This ...' in the last sentence refers to the '... to jump ...' at line 10. But for this connection to be clear, the sentence 'In this particular ...' interferes. Perhaps better move this sentence ('In this particular ...') to the end of the paragraph (after '... the correct solution.'). Furthermore, at line 13-12 I would suggest to insert 'until' between 'exists or' and 'the guiding'.
11. 14-6: check sentence 'with if'
12. 20, figure 2: please indicate which panel is a, b, c, d.

11

13. 21: what does the solid line in the figure mean. Is this the result of the Pathfinder method?
14. 25: figure 10: Too small to interpret. Perhaps you could make it a full page figure with 2 times 6 figures?