Article AMT-2020-78 – Response of Referee 2's comments

Iteration #2

Note: every response is followed by corresponding changes (CC) in the manuscript. Changes are referenced with lines numbers from the marked-up version to make them easier to identify. Changes visible in marked-up version that are not referenced as referee's response are for English language improvement.

General comments

In the article "Mixing height derivation from aerosol lidar using machine learning: KABL and ADABL algorithms", two machine-learning algorithms presented for the definition of the atmospheric boundary layer height, which is a principal parameter for the atmospheric modeling and air pollution dispersion.

I appreciate author answers to questions and suggestions of reviewers and I found that the second revision of the manuscript is definitely better than the first revision.

The authors are pleased to read that their efforts were appreciated and they would like to thank the reviewer for participating to the improvement of the manuscript by making thoughtful comments.

Detailed comments

Nevertheless, I found that the structure of the article could still be improved. For example,

1. Line 223: "– classif_score: The internal score used to automatically choose the number of clusters (only used when n_clusters='auto')." I suggest adding a reference to a following appropriate section.

The authors are not sure to correctly understand this comment because a reference to the section 3.4, where the classification scores are described, is already given line 250. These scores are also listed in the Table 1, therefore a reference to the Table 1 was added too.

Corresponding Changes (CC): add a reference to Table 1 next to the reference to Sect. 3.4 at line 250 (marked-up version).

2. Line 203: "1. Initialization: K centroids". A question arises how this K is defined.

Yes, the definition of K is added in the sentence.

CC: add "where K is the number of clusters specified by the user", line 227.

There are a few problems with equations and definitions:

3. Lines 270, 271 after equations, E and sigma are not defined; E is presented lately in line 312. Even if the expectation and the standard deviation are well known, they should be once defined in the manuscript.

The definition of E[.] and \sigma(.) are added in the text.

CC: add "and E[.] denoting mathematical expectation, \sigma(.) denoting the standard deviation", line 302.

4. Line 311: Please verify the equation as an expectation is a constant, and the variance of a constant is always zero.

The expectation is a constant, but the conditional expectation with respect to random variable (or a σ-algebra) is a random variable, by definition. Therefore, the variance of the conditional expectation is not always zero.

CC: none.

5. Line 284: N and K are not defined in the equation

The corresponding paragraph was heavily edited. The definition of N and K were added in the text at this occasion.

CC: definition of N and K added line 319-320.

Other questions:

6. Line 188: "The accuracy was estimated by group K-fold, ... ", do you mean K-fold cross-validation? In my opinion, the cross-validation is a key word.

Yes, adding the key word "cross-validation" makes the sentence clearer, therefore it will be added. Group K-fold is a variant of K-fold where the K parts of the dataset are taken among Kg>K groups and the groups are specified by the user. Here, groups where chunks of four consecutive hours. It was the chosen solution to ensure that the cross-validation is done properly in the time series (as previously mentioned by the referee in its previous comments).

CC: add "cross-validation" after "K-fold", line 213.

7. Line 192: "An independent validation set was not used here because the accuracy was only used to discriminate between the classification algorithms." Do you mean that the aim was to compare classification algorithms using cross-validation accuracy?

Yes, supervised classification algorithms were compared according to the accuracy estimated by group K-fold cross-validation.

CC: "accuracy" → "cross-validation accuracy", line 217.

8. Line 252: "Finally, we look for the first change in the cluster attribution, starting from the ground level." The justification of the algorithm should be somehow presented. I suggest adding an example of KABL classification as one as in previous author's answer (question 21. Line 209) to illustrate the problem.

The justification for taking the first change is the definition of the boundary layer as the "layer of atmosphere connected to the ground". However, the suggested figure is also included as the Figure 12 in the new version of the manuscript. The authors did not judge the Section 3.3 is the most appropriate to comment on that figure and preferred the Section 5.2 to do so.

CC: add "by definition of the boundary layer as the layer directly influenced by the ground...", line 279.

9. Line 434: "They occur with KABL because clusters do not always have vertical persistence (some points are identified as a free atmosphere in the middle of the boundary layer)". As in the previous question, I suggest adding an example of KABL classification as one as in previous author's answer (question 21. Line 209).

The mentioned figure (question 21, line 209 of previous authors' answer to review) was added as Figure 12 in the new version of the manuscript. It is used to illustrate the problem of vertical persistence of the clusters, as well as the different options available to solve it in the current version of the code. The Sect. 5.2 was modified to include the description of this Figure.

CC: add Figure 12 and heavy edition of the Section 5.2.

10. Line 301: "The most relevant configurations were retained and tested on the two-year dataset." Do you mean the two-year radiosonde dataset?

No, as KABL apply to lidar data, we refer here to the two-year lidar dataset. However, the lidar dataset is also covered by radiosondes twice a day, therefore we did not think of precising more than "two-year dataset".

CC: add precision "lidar dataset", line 340.

11. Line 302: "There are height parameters..." , I believe that it is eight (8) parameters.

Yes, sorry for the typo.

CC: remove the "h".

Article AMT-2020-78 – Response of Referee 3's comments

Iteration #2

Note: every response is followed by corresponding changes (CC) in the manuscript. Changes are referenced with lines numbers from the marked-up version to make them easier to identify. Changes visible in the marked-up version that are not referenced as referee's response are for English language improvement.

General comments:

1. The paper needs to undergo English proofreading. For example, some small grammar mistakes exist throughout the paper, such as in line 72: "one of the six sensor". Another example, the use of punctuation also requires to be improved. Please check the words: therefore, however, and so on.

The manuscript was submitted to English proofreading and it appeared that there were many grammatical errors introduced into the text, which supports the reviewer's observation that that the document needed English editing. There were also subject-verb problems, spelling problems, missing articles, and missing punctuation. We hope this English proofreading has corrected all errors and now matches the journal's requirements on language.

Corresponding Changes (CC) : many corrections throughout all the manuscript.

Some detailed comments:

2. The motivation of using machine learning algorithms is not clearly indicated in the abstract.

The main motivation is the ability of machine learning to reproduce human expertise, which is acknowledged as the best way to derive boundary layer height. Major changes in the abstract to clearly motivate the use of machine learning algorithms. To keep it short, references to secondary elements were removed or shortened.

CC: major changes in the abstract.

3. Also, when performance was mentioned in the abstract (e.g. outperform, discrepancy etc.), the used metrics should be stated.

The metrics used to discriminate KABL, ADABL and the manufacturer's algorithm are RMSE (the lower the better) and Pearson correlation coefficient (the higher, the better) with respect to colocated radiosondes. The word "results" was replaced by explicit reference to these metrics.

CC: "discrepancy in the results" → "discrepancy in terms of RMSE and correlation with RS", lines 13 and 558 (marked-up version).

4. The abstract conclusion stating that ADABL is a promising algorithm, should be accompanied with key metrics evidences.

The RMSE values at the Trappes site were added to support this conclusion.

CC: add for ADABL "(RMSE of 550 m at Trappes, 800 m for manufacturer)" + Kabl "(RMSE of 800m at Trappes)", lines 15-16.

5. The motivations of the need of derivation of BLH needs to be improved. For example, Some sentences are not explained properly. For example, line 29: "Others methods are based on derivatives.". The word of "derivatives" here are not clear, even after we read the next sentences. Another example, line: 34-35, the abbreviations of STRAT and CABAM are also not expanded and explained. Also, in line 38: "PathfinderTURB".

Additional end-users of the BLH parameter (physical processes research) where included in the introduction. Joined with the effort to clarify the paragraph that follows, we hope the motivation of the need of BLH derivation is improved enough. We acknowledge that the use of "derivative" is misleading here. It refers to the calculation of the derivative function, while, at many occasions in the paper, the word "derive" is used as a synonym of "estimate". Therefore, this specific use of the word derivative has been clarified. We did not find explicitly any specific details about *pathfinderTURB* name in its reference article. It seems that it's just the name of the new and extended version of the *pathfinder* algorithm.

CC : add a sentence at lines 26-28; expand algorithm acronyms for STRAT and CABAM, line 47-49; "derivative" → "calculation of the derivative function", line 41.

6. The paragraph starting by line 40 should be used to explain AI first briefly, as some readers in journal may not be familiar with AI definition.

We agree with this comment and we added the definition of AI in the sentence introducing the acronym.

CC: "(AI)" → "(AI), as the set of techniques aiming to reproduce...", line 55.

7. The authors need to describe clearly supervised and unsupervised learning concept, which can be done between lines and 55.

An additional sentence on the distinction between supervised and unsupervised algorithms was included. More generally, the introduction of acronyms was entirely proofread.

CC: add the sentence "Algorithms classifying from a reference dataset (as ADABL) are called supervised algorithms while algorithms classifying without reference dataset (as KABL) are called unsupervised.", lines 70-72.

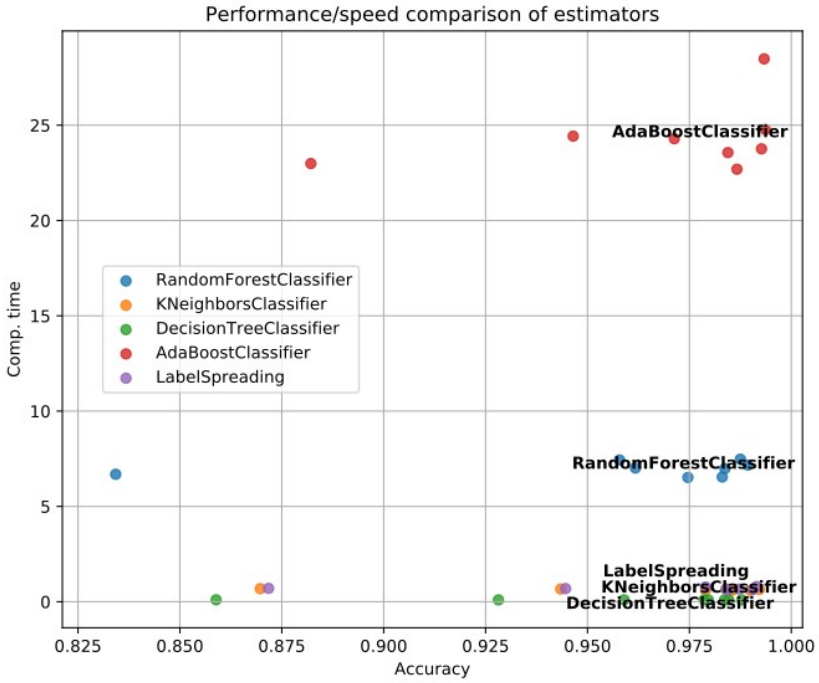
8. Some words are not known by wider audiences. For examples, line 95. The word row211 is not known well by the readers.

raw2l1 is the name of the software used to prepare the lidar data. We re-organized the sentence to make it clearer, also we changed the word "routine" by "software" which seems more explicit. The use of italic is reserved to software which are named with a common name.

CC: we set footnote anchor at the first occurrence of raw2l1 and put what remains of the sentence in the footnote, to clarify "was developped" is replaced by "is maintained", line 115.

9. There are numerous supervised learning algorithms, the authors did not state clearly why boosting algorithms were chosen. Instead, the authors just stated that "the boosting algorithms are a very powerful family of algorithm".

Several algorithms were put in competition to maximize accuracy (estimated by group K-fold cross-validation, as explained in Sect. 3.1.3). The results of the competition (not shown in the paper) are the following:



AdaBoost has the best accuracy in average on all cross-validation splits. Its computing time is, although the highest, still affordable. Therefore it was chosen to make the supervised classification.

CC: add a sentence to precise that several algorithms were tested, line 154-155.

10. The KABL flowchart needs to be shown as a pseudo-code. The reviewer is also wondering, why the flowchart was made only for KABL, not for AdaBoost algorithm. If the paper attempts to compare both algorithms. They should be described fairly.

The Figure 6 was modified to highlight the common steps between KABL and ADABL (all but apply_algo). As suggested, a pseudo-code was added in stead of the "Data" information. In the pseudo-code, we focused on the parameters of KABL because they are subject to a sensitivity analysis in the Sect. 4.1. The many abbreviations were also precised in the caption.

CC: new Figure 6

11. The notations are not fully described in equations (1) and (2).

The subsections 3.4.1 and 3.4.2 where re-organized in order to clearly introduce all notations, number all equations and improve coherence between these two subsections.

CC: editing of Sect. 3.4.1 and 3.4.2.

12. The equations in subsections 3.4.2 are not numbered. The notations must also be explained properly. The reviewer understands that these are from textbook, but if the authors decided to include those in the manuscript. The notations must be clearly described.

Same as previous comment.

CC: editing of Sect. 3.4.1 and 3.4.2.

13. Table 1 has no explanation. All of items inside the table must be described.

The caption of Table 1 is above the table. It has been expanded with a reference to the paragraph were the items inside are described.

CC: caption of Table 1, line 336.

14. The fonts in Figure 9 are not readable.

The fonts of Figure 9 were enlarged.

CC: new Figure 9.

15. Caption for Figure 11 needs to be expanded a bit.

The caption of the Figure 11 was expanded with a description of the different elements in the figure and the expansion of the acronym “INDUS”, which was not previously introduced.

CC: caption of Figure 11, line 461.

16. Any idea, what is manufacturer algorithm does mathematically?

The reference given by the manufacturer is Brooks, 2003. We mention it in the manuscript at line 409 (marked-up version). From what we understand, the basis of the manufacturer algorithm is a Haar wavelet covariance transform method. It consists in calculating the convolution of the lidar backscatter signal with a Haar wavelet. The maximum of the convolution is identified as the boundary layer top. We do not have any specific details about the implementation of the method. For example, we do not know how the dilation parameter of the wavelet is chosen.

CC: none

17. Please state future direction of the research and this work.

Future directions are given in the Section 5 which was renamed "Discussion and prospects" to make it more explicit. We chose to put the discussion and future directions before the conclusion because the conclusion includes material from the discussion and we would like to keep our conclusion concise.

CC: change of Section 5 title to "Discussion and prospects"