

***Interactive comment on* “On the estimation of boundary layer heights: A machine learning approach” by Raghavendra Krishnamurthy et al.**

Anonymous Referee #1

Received and published: 18 February 2021

Review for AMT-2020-439 Title: “On the estimation of boundary layer heights: A machine learning approach” Authors: R. Krishnamurthy, R. K. Newsom, L. K. Berg, H. Xiao, P.-L. Ma, D. D. Turner

=====
Major comments: The manuscript presents a machine learning approach (random forest) used for the computation of the boundary layer height from Doppler lidar measurements. The algorithm uses as input some parameters derived from Doppler lidar observations as well as different surface meteorological measurements. The topic is very interesting as boundary layer heights are a key parameter in many aspects. I had to re-read it several times because things were not explained clearly along the way. I am still confused on what are the variables from the Doppler lidar that are really used

as input for the RF method. I think that the manuscript will be ready for publication after the authors will address some of the concerns discussed below.

1) The main question that I have is why the authors decided to use the Tucker method for comparison and not the Bonin et al. method, which is more refined and accurate than the Tucker method. Bonin et al. (2018) presents a technique that blends all the data together from multiple scans to determine a unified measurement of the MH and the uncertainty of the estimate, but it can also be used with limited inputs, such as only data from zenith stares, so I am not sure why the authors did not choose it. Moreover, the clear limitation of the Tucker method (visible in Fig. 2, 3, 6, 7) is that even in the case of loss of signal, due to the low signal to noise ratio of the Doppler lidars (Fig. 6 is a clear example of this) it still provides an estimation (biased low), while the MH is clearly above that value. Fig. 3 clearly shows that the y-axis estimations (Tucker method) do not go often above 2 km, while the radiosonde ones (all three methods) have a lot of estimations above 2 km. Maybe there is a good reason for their choice, such as some limitation determined by the dataset, but I think it was not well addressed in the manuscript. As it is presented now, the RF Machine learning method seems to be used as a bias correction method applied to the Tucker method. I.e., they use the Tucker method z_i as input, knowing that it is biased low when the boundary layer grows tall; they include surface observations; and they use the radiosonde estimates to train the RF method to correct it for the low bias it has during well-developed boundary layers.

2) The machine learning algorithm is somehow presented in Section 3, but in a generic context that I am not sure make it reproducible by an interested reader. I think Section 3 could be expanded and clarified in the revised version of the manuscript.

3) Another concern is that the results presented in Section 4 use a dataset to train the RF model with no missing data (which 'could' be OK), but also verify it on a dataset (future features) with no missing data, which is an ideal situation, that does not happen often in reality. Therefore, the results presented in Section 4 are on the optimistic side.

I am curious on why they just don't use the verification dataset in 2019 as is.

4) Section 6 "Case Study: Preliminary Model Comparisons" is a very, extremely preliminary test, with no quantitative results, due to the very limited number of days available to the analysis. For this reason, I am not sure I find it very useful in this study. Just my opinion.

5) Finally, I strongly suggest the authors to look for grammatical mistakes, as I found many, some of which, but not all, reported below.

=====
Specific comments:

Page 1, line 12: Please, clarify the meaning of "long-term data"

Page 1, line 19: "using Doppler lidars only."

Page 1, line 19: "improvements . . . were observed"

Page 1, lines 20-21: This sentence does not read well: "where a 50% improvement in mean absolute error compared to lidar-only zi estimates and provided an R2 of greater than 85%."

Page 1, line 28: "the top of the PBL is that the turbulence is near zero". What about cloudy conditions?

Page 2, line 34: "de facto"

Page 2, line 47: The Bonin et al. (2018) reference should be placed here.

Page 2, lines 51-52: "Alternatively, velocity information from a Doppler lidar can be used to estimate zi." I think this sentence is not well explained and possibly not well positioned.

Page 3, line 77: "Bianco and Wilczak 2002 and Bonin et al., 2018" are referenced to in the wrong place. They employ fuzzy logic-based methods to estimate boundary layer

heights (as stated in their titles), so, no ML used there.

Page 3, line 91: “, and data are compared”. Replace with “, and observations are compared”

Page 5, Table 1: Instead of column 2 that now presents the data stream names (which could be rather included in an Appendix), I think it would be wiser to have a column introducing the height (or ranges of heights) of each measurement used in the RF machine learning method. As presented now I am confused why the vertical velocity, range-corrected attenuated backscatter, signal to noise ratio variance go from surface to 800 m AGL, as well as the average eddy dissipation rate (also why do they stop at 800m?), but no range of heights are specified for the height-resolved vertical velocity variance, wind speed and direction.

Page 6, line 114-116: What is the vertical range of measurements for the Doppler Lidar (min, max)? It is not mentioned in the text nor in Table 1. Also, it would be interesting to see the % data availability with height of the Doppler lidar used in the study.

Page 6, lines 124-125: “Estimates of eddy dissipation rate were computed between 100 to 800 m AGL”. Is it only eddy dissipation rate that is computed between up to 800m? In Table 1 you mention that also all the other variables (Vertical velocity, range-corrected attenuated backscatter, signal to noise ratio (SNR) variance) are measured “from surface to 800 m AGL”, but this cannot be right. Can you clarify?

Page 7, lines 155-157: It seems a large source of error the fact that even with a reduction of sensitivity during the hottest portion of the day the algorithm still provided estimates, which are of course biased low.

Page 6, lines 129-130: Since radiosonde launches are at ~0530, 1130, 1730 and 2330 UTC each day (local time = UTC - 0600 hours), and you are only evaluating daytime performances of your ML method, are you only using the radiosonde launches at 1730 and 2330 UTC to verify it?

[Printer-friendly version](#)[Discussion paper](#)

Page 7, line 163: “radiosonde-derived zi are assumed to be the best guess zi estimate and is used to calibrate”. Please, correct the grammar.

Page 8, lines 165-166: “1785 days with daytime clear and shallow cumulus conditions”. What dataset are you using here? 1785 days are almost 5 years of data.

Page 8, lines 173-174: “In this paper, the primary focus is on evaluating the daytime zi estimates from RF models.” How do you determine the “daytime” start and end times in your study? Do they change according to the time of the year?

Page 8, lines 183-184: “Bootstrap aggregation (bagging) is used so that each tree can randomly sample from the dataset with replacement, while only a random subset of the total feature set is given to each individual tree”. I don’t understand what you are doing. Could you try to rephrase this sentence? Page 10, lines 208-209: Check the grammar.

Page 10, line 226: “order of magnitude in their variability”. Do you mean “in their value”?

Page 10, line 226: The word “data” is plural. Please correct here and elsewhere in the manuscript.

Page 10, lines 226-229: This whole sentence is very convoluted. Please try to rephrase it. Also, please clarify what you mean by “standardized”.

Page 11, lines 256-257: “Therefore, in this analysis, the model is trained with no missing data, and no imputation is done on the data (either input or future features) to accurately test the efficacy of the RF model.” This is a main concern to me. If I understand correctly, but maybe I am wrong, here you are saying that for the results presented in Section 4 you use to train and verify your model only using data with no missing features. Nonetheless, earlier in the text you stated that “It is critical for the RF model to deal with missing values in its training phase”, which I strongly agree with, as in real life missing features can happen. If this is true (again, I might have misunderstood), I think your results are more representative of a best-case real-time scenario, without

Printer-friendly version

Discussion paper



ever missing features.

Page 11, line 262: “where in a model developed at a given site is tested”, “in” can be removed.

Page 12, line 283: I see that you define “daytime” here. Maybe you could specify it earlier in the text, when you first talk about it.

Page 13, line 307: “with an R2 of greater than 0.85”, remove “of”

Page 13, line 315: “estimates that used to calibrate the model”, replace with “estimates that are used to calibrate the model”.

Page 13, line 320: Please reword “2 annual cycles of data”?

Page 14, Figure caption: Please specify that c) and d) are for “daytime and nighttime”.

Pages 15, Fig. 6: It seems that the RF method can provide estimated where there are no Doppler lidar measurements. Is that correct?

Pages 15 and 16, Figs. 6 and 7: Please keep the colors the same in both figures (i.e.: Lidar zi should be red in Fig. 7).

Page 16, line 359: “are shown in Figure 8a”

Page 16, Fig. 7 caption: I think here you could simply say “As in Fig. 6, but for June 22, 2019.”

Page 16, line 363: “a standard bias correction would not always improve zi estimates from the Tucker method”. Actually, the bias in the Tucker method seems pretty constant in Fig. 8a. . . Which is confusing because before you said that “the Tucker method generally works well at tracking the height of the mixed layer during its initial development phase” and this does not reflect in the gray line in Fig. 8a.

Page 17, Figure caption: You say that in panel a) there is the “Tucker method zi”, but you call it “Lidar-only zi” in the label. Sometimes you refer to it as the “Lidar-only”

Printer-friendly version

Discussion paper



method and sometime as the “Tucker method”, here and in other places in the text. Also, how is the “Lidar-only zi” determined at night-time? Here it seems that it cannot be the lowest range-gate of the Doppler lidar as it seems higher than the one in Fig. 6. Also, please, specify what the error bars represent.

Pages 18 and 19, Tables 4 and 5. Very interesting analysis and results. Would it be possible in your future research to include land use type? Could this give you the possibility to include a variable to distinguish between different sites/seasons in your future analysis?

Page 19, line 419: This should be Eq. (4), not (1).

Page 19, Table 5: Is this “Lidar-only zi” simply the lowest range-gate of the Doppler lidar?

Page 20, line 442: “that the parameters shown to be important are with respect to the RF model, are features that successfully. . .”. Remove one “are”.

Page 20, lines 447-448: “In this research, we have mostly analysed using standard processed data from SGP instruments as an input into the RF model.” This sentence is incomplete.

Page 21. Fig. 9: Should y-label include (%)?

Page 22: The Equation numbers are wrong.

Page 23, line 496: What are “lidar false alarm rates”?

Pages 24 and 25: I find this whole section not very interesting. The models are described in great detail, but the analysis is very poor, due to the very limited dataset. So, I don’t know if it adds much to the manuscript.

Page 24, line 517: “horizontal resolution”. Would it be more accurate to say, “horizontal grid spacing”?

[Printer-friendly version](#)[Discussion paper](#)

Page 24, lines 530-531: “Nocturnal zi is not estimated using model data”. Why is that?

Page 24, line 535: “RF model provides zi estimates at a much finer temporal resolution than radiosondes”. What is the RF model temporal resolution?

Page 26, line 566: “the mean absolute error of boundary layer height estimated by RF model reduced”, “is” is missing.

Page 26, lines 577-578: “in convective velocity scale estimates when used Tucker method.” Replace with “in convective velocity scale estimates when the Tucker method is used.”

Page 26, lines 583-585: The whole sentence is poorly written, please rephrase.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2020-439, 2020.

Printer-friendly version

Discussion paper

