The authors would like to thank the reviewer2 for all their suggestions. We have addressed the comments and provide a point-by-point response to the recommendations made by the reviewer2 below. The reviewer2 comments are in black, and our responses are in red.

Kumar et al.: **Correcting for filter-based aerosol light absorption biases at ARM's SGP site using Photoacoustic data and Machine Learning**, Atmos. Meas. Tech. Discuss. https://doi.org/10.5194/amt-2022-42, in review, 2022.

Review

General

The paper presents the use of a supervised ensemble Machine Learning (ML) algorithm for improving the constants in an algorithm used for calculating absorption coefficients from PSAP data and for calculating a new algorithm for the same purpose. The method interesting and useful since it can improve the accuracy of absorption meassurements. The paper is definitely worth publishing in AMT but I do have some suggestions and several questions that should be answered before that. They are all in the detailed comments and questions below.

The authors would like to thank the reviewer for providing comprehensive and insightful suggestions. We have made appropriate changes to the manuscript as detailed in the reviewer's specific comments.

Detailed comments and questions

L 45: Why not replace "Manufacturer's" with "Radiance Research"?
This statement is now replaced in the updated version of the manuscript as follows: "*Light absorption data by aerosols at the SGP site is collected using Radiance Research's 3-wavelength Particle Soot Absorption Photometer (PSAP)…*"

L66 - 79: You should also cite Müller et al. (AMT, 7, 4049–4070, https://doi.org/10.5194/amt-7-4049-2014, 2014) and note that it is based on much more rigorous theory of radiative transfer through the filter.
This paper is now cited in the manuscript.

L72, Eq. (1): First, I suggest you don't present the equation in the introduction, it would be much more logical to present it in section" 2.2 Correction algorithms". However, where ever you present it, you should define exactly what you mean by B_{PSAP}. Is it the B_{PSAP} presented in Eq. (3) of Bond et al. (1999)? If it is, you should keep in mind that it already includes one loading-correction function f. In other words, what is your "uncorrected filter-based absorption" all over the paper? The RR 3wl PSAP firmware calculates automatically absorption coefficients corrected with the Bond et al. correction excluding the scattering correction. With user-defined constants. So is that what you think that is the "uncorrected B_{PSAP}"? If so, that is not quite correct. The uncorrected B_{abs} shoud be B_{PSAP} divided by the loading correction function f that the PSAP firmware uses. Explain in more detail. And further, if you really have used the BPSAP calculated directly by the PSAP and assumed that it is the "uncorrected absorption" then you have to recalculate everything! I hope not. Recheck that!

Thank you for correctly pointing this out. As you advised, we have now used the uncorrected B_{abs} as B_{PSAP} divided by the loading correction function f(Tr). We have also recalculated everything and updated the numerical values in the text, tables, and figures. This causes a small non-uniform scaling of values as compared to the previous wrong "uncorrected Babs" values that we were using; hence it changes the numerical result values slightly and but does not affect the overall conclusions of the study. However, we have corrected and updated all the Figures in the manuscript affected due to this additional preprocessing. We also made changes in the manuscript text by defining "*the uncorrected aerosol absorption data as derived from PSAP ($B_{abs\_uncorrected\_PSAP}$)*" in equation form. We have now updated the terminology in the manuscript. Now, B_{abs_PSAP} means the absorption coefficient output from the PSAP firmware which includes the automatic B1999 correction and $B_{abs\_uncorrected\_PSAP}$ means the back calculated uncorrected PSAP-based absorption coefficient without the B1999 correction.

$$B_{abs\_uncorrected\_PSAP} = \frac{A_{PSAP}}{Q_{PSAP}\Delta t} ln\left(\frac{I(t)}{I(t+\Delta t)}\right) = \frac{B_{abs\_PSAP}}{f(Tr)} = \frac{B_{abs\_PSAP}}{\left(\frac{1}{1.317 \times Tr \; + \; 0.866}\right)} \qquad (1)$$

L86-88: "Our findings show … ". I suggest you move this to the conclusions. These are all results of the whole study. In the intro you should present the goals and in the conclusions the main result.

This paragraph is now shifted to conclusions and updated.

Section 2.1 You should write something about the inlets, flows, cutoffs and size ranges for the different instruments. These are not just for fun, they are important info to try to evaluate the sources of the differences of the absorption coefficients from the different instruments.

This information is now included in the manuscript in lines from 110-125 in Section 2.1.

In this section you should also tell, which filter material you used in the PSAP. That is important because the constants in the algorithms depend on the filter material.

This information is now updated in Section 2.1. "*The PSAP has been operated by ARM (and many others in the global community) for almost 25 years with the same filter media, Pallflex E70-2075W, which is composed of quartz fibers on a cellulose backing. All published corrections factors were developed and measured using the Pallflex E70 media*."

L110-111: In this preprocessing, did you divide the B_{PSAP} with the f(Tr) that is automatically calculated by the instrument firmware?

As mentioned in an earlier response, we took care of this preprocessing now and updated all the figures and affected text in the manuscript.

L116-117: The AAE from the PASS data in Table A1 are somewhat suspicious. Especially the ones that have the wl 532 included. The fact that AAE(405-532) < -0 and AAE(532-781) > 2 suggest that B_{ABS}(532) is overestimated. Then this would have important implications to the factors presented in Table 2. Discuss this.

To present the SGP data better, we now present Figure A1 in updated manuscript which includes descriptive statistics of measurements from all the instruments and the derived optical parameters like SSA, AAE and SAE obtained from the SGP data in the Appendix section.

In Table A1, the average AAE calculated using 405nm and 532nm wavelength was negative because, as it can be seen in Figure A1(d), there are many negative outliers for AAE_PASS_405_532 which was pulling the average below zero. Please note that the medians {green line} of all the AAEs calculated are positive.

We have now stated this in the manuscript as well – "*From figure A1(a), (d) and figure A4, however, we suspect that either the newly installed 532nm PASS laser could be slightly overestimating absorption, or that the old 405nm and 781nm lasers could be slightly underestimating absorption compared to their true values.*"

We agree that due to unavailability of highly accurate aerosol particle-phase light absorption data, we had to resort to assuming available PASS instrument's measurements as ground truth for absorption, which limits the accuracy of the revised Virkkula parameters shown in Table2.

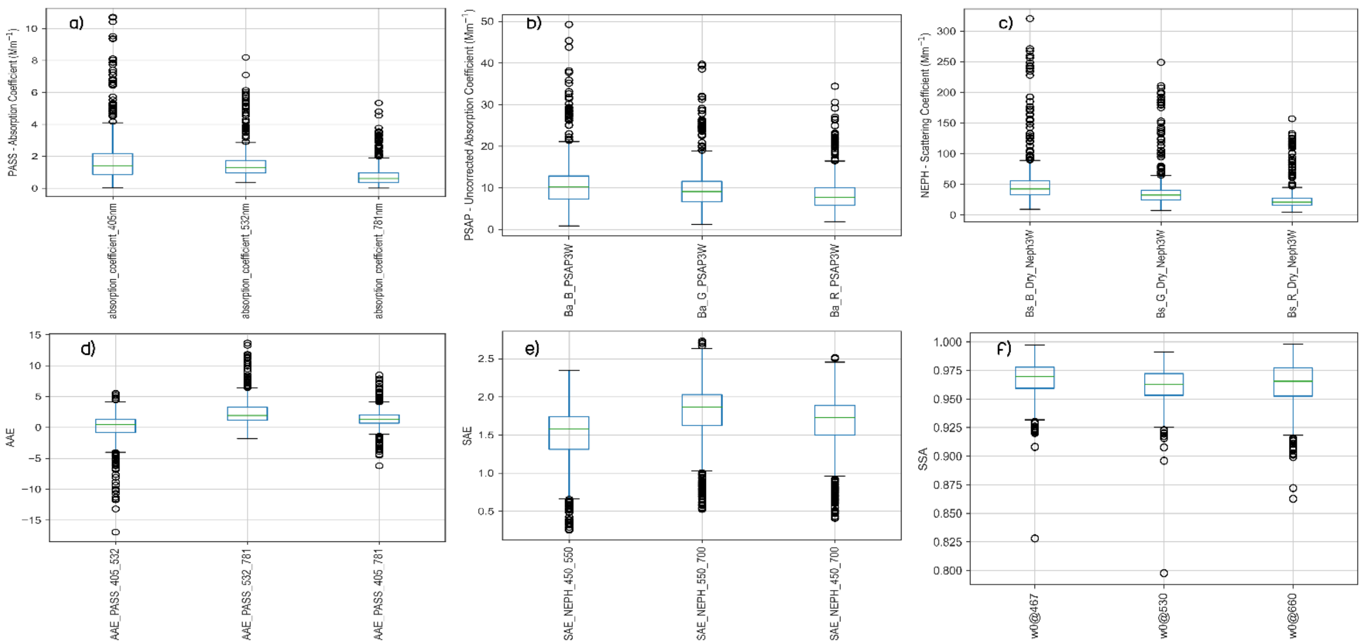

Figure A1: Summary of the SGP dataset. The boxplots of raw measurement data are shown as obtained from various instruments used in this study (a) PASS (b) PSAP{ $B_{abs\_uncorrected\_PSAP}$ } and (c) Nephelometer. The boxplots of parameters dervied from the raw data are also shown (d) AAE (e) SAE and (f) SSA. The green line is the median of the data. The bottom line of box is 25% percentile of data and top line of box is 75% percentile of data, therefore, the box represents the middle 50% of all the datapoints which is the core of the data.

Further about Table A1. I strongly suggest you add more information in it. You have not presented anywhere the descriptive statistics (ave, std, some percentiles) of the aerosol optical properties during the campaign. That would be very important because it would show the range in which your results are applicable. Present B_{abs}, B_{sca}, SSA, AAE from the different instrument and algorithm combinations. Especially AAE is important, from the different algorithms. Peope use them for source apportionment. And in my opinion this table is important, it would deserve to be in the main text.

Thank you for this suggestion. To present the SGP data better, we now present Figure A1 in the manuscript which includes descriptive statistics of measurements from all the instruments and the derived optical parameters like SSA, AAE and SAE obtained from the SGP data in the Appendix section.

L271-275: So, this is not just an adjustment of the factors in Virkkula (2010). Do I understand right: the RFR has given as an output an equation that was used for calculating the absorption coefficients. If so, you should present the equation so that other people can use it also! What are the parameters the new function depends on?

Yes, this is not an adjustment of factors in Virkkula (2010) here. However, since, RFR is a machine learning algorithm it cannot be expressed in a simple equation form. Machine Learning models can only predict the output using the input data provided to them once they are trained on both input and output data (training dataset).

L301-307: Please give the full functions so that other people can use and test them. What are the derived constants for wood burning and kerosene burning smoke?

We can only train machine learning models like RFR and test their output accuracy on unseen input data. Roughly speaking, tree-based algorithms can be thought of as a series of many if-else statements nested together, hence, we cannot obtain a simple looking mathematical equation out of a trained model. However, for practical applications, ML models can be part of post-processing of data at various sites to produce highly accurate outputs possible.

Fig. A2a: You do have also EC concentration data! I suggest you use those data also, not just in this plot. Excluding the obvious outliers in the data shown in A2a, how do the the absorption coefficients with the different methods correlate with EC? From the linear regressions you would get mass absorption coefficients. What would be the derived MAC(λ)? Anything close to published ones? These info would not be just for fun, they would be an additional support for the values derived from the different algorithms.

Unfortunately, we do not have EC concentration data for the period of focus in this paper (27th June-25th September). This period was chosen because it gave us access to good quality data, parallely, for all the required instruments for this study (PSAP, PASS, NEPH, ACSM), and a new high powered green laser PASS was upgraded at the SGP site in 2015 [1]. The only reason behind Fig. A3a (previously Fig. A2a) being present in the appendix of manuscript is to give the reader a glimpse of the SGP site's typical EC concentrations, since most of the results are using SGP data. The availability of the data can be cross-checked at the ARM's Data Discovery site [2].

References:

1) https://github.com/joshinkumar/Filter-correction-ML-code/blob/main/Dubey%20Poster%202015.pdf
2) https://adc.arm.gov/discovery/#/