The study deals with the use of the Matlab regression learner software to explore various regression methods applied to several large observational airborne lidar (HSRL) data sets. These data were collected during large field campaigns over major urban areas in the USA, South Korea, Philippines, Taiwan, and Thailand. Goal was to investigate what combination of lidar information (backscatter, extinction, lidar ratio, depolarization ratio at single or multi wavelengths) allows a good estimation of PM2.5 and PM10 at heights close to the surface. In this machine learning (ML) studies, dense sets of network in situ observations of PM2.5 and PM10 were used. It was found that the Exponential Gaussian Process Algorithms consistently showed the best performance. 12 different lidar configurations (models 1-12) were defined. However, in the result section only the optimum model (model 11) was applied.

1. This is an excellent and well elaborated study done by experienced researchers and lidar experts!

Response: We thank the reviewer for these comments.

I have only minor remarks. As a reviewer, my role is to be critical and to criticize if I find something that should be mentioned. All the positive aspects remain widely uncommented.

2. The Abstract may be too long. According to the AMT/ACP rules the abstract should not exceed 250 words.

Response: We have not reduced the size of the abstract because:

1) We could not find any mention of specific abstract length in the instructions for preparing the paper provided at https://www.atmospheric-measurement-techniques.net/submission.html.
2) We have also found several papers with abstracts longer than 250 words that have been recently published in AMT.
3) The paper covers several topics and results which should be included in the abstract and summary of the paper.

3. Lines 93-111: What about all the ground-based lidars and lidar networks? Why are they not mentioned? All the multiwavelength Raman polarization lidars, EARLINET? Ground-based lidars are ideal to monitor the diurnal, weekly, and seasonal cycle of the aerosol pollution state in urban areas, and this, in contrast to airborne and satellite lidars, continuously! Airborne field campaigns are very useful, no doubt, but they are snapshots! Spaceborne lidar observations provide

global coverage, however, also snapshot-like. To my opinion, in such a general introduction one should provide a more general overview on the available lidar techniques and networks, MPLNET, ADNET, EARLINET.

Response: In the context of providing detailed aerosol vertical information, it is true that ground-based lidars can also provide measurements of the fine scale vertical aerosol structure in addition to airborne and space-based lidars; consequently, the text (lines 93-103) has been modified to be more general and avoid the impression that only airborne and space-based lidars provide such detailed aerosol vertical information. The text has also been modified to mention ground based lidar networks in this context.

Note that the paper (line 103) already refers to (Xiang et al., 2020) for an example of how ground-based lidars can help provide estimates of $PM_{2.5}$ concentrations. We have also added another reference (line 100) (Li et al., 2017)) regarding the use of ground-based lidar measurements to infer $PM_{2.5}$ concentrations.

However, please note that the paper is not meant to give a review or survey of all lidars and lidar techniques used to provide vertically resolved aerosol information. Rather, the introduction is meant to provide some background on how such lidar measurements have been specifically used to estimate or retrieve $PM_{2.5}$ concentrations and how airborne measurements such as those described and used in this paper are particularly relevant. The revised text (lines 93-103) now reads

"Lidar measurements can help alleviate these issues. Through their ability to provide the fine scale vertical structure of aerosols as well as by constraining aerosol type (Burton et al., 2012), lidar measurements are very helpful for evaluating and improving models. Ground-based lidars such as those operated in the EARLINET (Pappalardo et al., 2014), ADNET (Sugimoto et al., 2016), and MPLNET (Welton et al., 2018) networks are valuable for providing data for climatological studies, long-range transport events, and model evaluation. Lidar measurements of aerosols at or near the surface avoid uncertainties associated with using column aerosol measurements (i.e., AOT) to infer aerosol concentrations near the surface. Lidar measurements of aerosol extinction near the surface have been used to help derive estimates of surface $PM_{2.5}$ concentrations. For example, Li et al. (2017) used an empirical model based on the regression between $PM_{2.5}$ and the near-surface backscatter measured by ceilometers to derive surface $PM_{2.5}$ at a few selected sites. Coincident ground-based lidar and surface $PM_{2.5}$ measurements were also used to derive a linear model to relate aerosol extinction to surface $PM_{2.5}$ concentration (Xiang et al., 2020)."

4. Lines 139-150: To continue with my general comment: I was surprised that the Raman lidar technique was not mentioned at all, although the first author Rich Ferrare grew up as an aerosol Raman lidar specialist. The use of the robust and very stable Raman lidar technique is, to my opinion, the optimum approach for long term monitoring of aerosol pollution, even at low heights of 100-200 m above ground (by using near range receiver units). Meanwhile, rotational Raman channels allow coverage of the lower part of the atmosphere even at daytime.

Response: It is true that Raman lidars, particularly ground based systems, are well suited for long-term continuous measurements at a particular location. However, as mentioned above, this paper is not a review paper to discuss or debate the advantages and disadvantages of various lidar techniques for atmospheric aerosol measurements used for long-term monitoring of aerosol pollution. Instead, the paper discusses the use of airborne HSRL data to derive PM concentrations during episodic field missions. Therefore, the paper is focused on why such measurements are valuable and how they are used in this manner. If there have been long-term, airborne Raman lidar aerosol measurements that have been used to retrieve PM concentrations in a similar manner as the airborne HSRL measurements discussed in this paper, we would reference these as well; however, we are unaware of any such measurements.

5. To avoid misunderstanding. The development of all the different airborne HSRL lidars at LARC, NASA is unique! The lidar team as a whole did a fantastic job during the last 10-15 years.

Response: Thank you for these comments.

6. Back to the manuscript. Later on, I was also surprised that none of the defined models 1-12 covers the CALIOP lidar configuration. I think that should be improved. Or does it make no sense at all, when there is no lidar-ratio information? The CALIOP model would be model 7 without information on 532 nm lidar ratio and 532-1064 nm depolarization ratio. The comparison of model 7 (without the lidar ratio and 1064 nm depol ratio information) and model 12 would be the perfect opportunity to demonstrate the big step forward in spaceborne lidar development from CALIOP to ATLID (EarthCARE lidar)!

Response: The following text was added (lines 396-409):

"It is also interesting to note the performance of Model 5. Of all the models discussed and shown in Table 2, Model 5, which includes measurements of near-surface aerosol extinction, near-surface aerosol depolarization, and AOT and excludes measurements of the lidar ratio and the ratio of aerosol depolarization (532/1064 nm), probably provides the most appropriate configuration to estimate the performance associated with CALIOP. Figures 5 and 6 show that the retrieval performance of Model 5 is not that much worse relative to other models (e.g. Models 7-11) that included additional parameters as well as Model 12, which corresponds to the ATLID lidar configuration. However, it is important to note that these airborne HSRL systems measure both total attenuated backscatter and total attenuated molecular backscatter to enable direct retrievals of unattenuated aerosol backscatter and aerosol extinction. Since CALIOP did not measure total attenuated molecular backscatter, CALIOP retrievals of aerosol extinction and unattenuated aerosol backscatter typically rely on estimates of the lidar ratio based on inferences of aerosol type. As a result, the CALIOP measurements of near-surface aerosol extinction and, to a lesser extent, near-surface aerosol depolarization can have significant systematic uncertainties due to the uncertainties in the assumed lidar ratios. Consequently, we expect the actual performance of Model 5 as applied to CALIOP retrievals would be worse than the performance as applied to the airborne HSRL systems shown in Figures 5 and 6."

7. Line 216: Table 1! It is not easy to find out what the HSRL 2 (the main lidar in all these field campaign discussed in this paper, model 11) can measure. A better, clear overview of the different systems would be helpful.

Response: Table 1 and the Table 1 caption have been modified to show more clearly (in violet) the additional parameters at 355 nm measured by HSRL-2.

8. Line 218: What do you mean with self-calibration. In the backscatter coefficient retrieval, you always need to assume a reference backscatter value at the reference height.

Response: As stated in this text, unlike standard backscatter lidars, these airborne HSRL systems do not require an assumed, aerosol-free region for calibration but rather use the internal calibration technique described in section 5 in Hair et al., (2008). As described by Hair et al., (2008), the gain ratios required for this internal calibration are measured during each flight. The gain ratio between the molecular and total scattering channels are measured using a movable iodine filter. The depolarization calibration between parallel and perpendicular channels is measured by matching the polarization axis of the outgoing light is matched to the receiver; this process is automated and performed each flight. The polarization gain ratio calibration is performed by rotating the transmitted polarization 45° to the receiver analyzer, such that both polarization channels measure equal components of the parallel and perpendicular backscattered return. These calibration procedures are explained in detail by Hair et al., (2008).

The following lines (Lines 223-224) were added to the text "As described in detail by Hair et al. (2008), through careful system design and operations, the channel gain ratios and filter transmission are accurately measured to enable this self-calibration."

9. Line 290: So, the basic goal was to use 193 flights (conducted from 2010-2024) over major metropolitan regions to explore various machine learning regression models for deriving PM concentrations. The result section is, however, mainly contains HSRL-2 observations and applications of model 11.

How many flights were conducted with the HSRL-2?

Response: There were 93 flights conducted with HSRL-2. The number of flights conducted by each airborne HSRL is shown in the legend in Figure 1. The caption for Figure 1 has been edited to also indicate this.

10. When using these 193 flights over urban areas then you investigated the link between lidar observations and in situ observations for only ONE aerosol type, even if PM2.5/PM10 ranged from 0.1 to 0.9? Please comment on that!

Response: Although these flights occurred predominantly over urban areas, the aerosols encountered over these areas correspond to several different aerosol types (urban, smoke, maritime, dust, etc.) as derived from the HSRL aerosol typing algorithm. Therefore, there were no restrictions on aerosol type nor the $PM_{2.5}/PM_{10}$ ratio when comparing PM concentrations derived from the airborne HSRL measurements and those derived from surface or airborne (i.e. DC-8) in situ measurements.

The following text has been added (Lines 362-365) "Although these flights occurred predominantly over urban areas, the aerosols encountered over these areas correspond to several different aerosol types (e.g., urban, smoke, maritime, dust, etc.) as derived from the HSRL aerosol typing algorithm. There were no restrictions on aerosol type nor the $PM_{2.5}/PM_{10}$ ratio when assessing PM concentrations retrieved from the airborne HSRL measurements."

As stated in section 3.5, when comparing mass extinction efficiencies derived from HSRL-2 and the DC-8 in situ measurements, the data were restricted to cases where the $PM_{2.5}/PM_{10}$ ratio was greater than 0.5 and the aerosol type was one of four types (smoke, urban, dusty mix, and maritime).

11. To cover the entire globe (in the case of global observations with CALIOP or ATLID) would that mean we need global sets of in situ PM observations in the machine learning studies?

Response: Yes, as discussed at the end of section 2.3 (Lines 464-475), the results obtained in this study using airborne HSRL data apply to a certain range of PM values and aerosol parameters. The specific training set used here would be suitable for retrieving PM estimates for PM values and HSRL aerosol measurements in these ranges but would not be suitable for retrieving PM significantly higher than the maximum values listed. Since these values would certainly be exceeded in some regions of the globe, a revised training set that included a larger range of PM values would be desired for global monitoring of PM from space.

12. Line 316: Is there a good reference available so that the reader can learn more about the Exponential Gaussian Process Algorithm?

Response: The references provided in the paper could be a starting point. This includes the Williams and Rasmussen (2006) (line 325) (which can be found at https://gaussianprocess.org/gpml/chapters/RW.pdf) as well as the references provided in the footnotes. Another reference would be the Matlab documentation provided at

https://www.mathworks.com/help/stats/gaussian-process-regression-models.html

13. Line 329, Table 2: Model 11 has the most crosses and is obviously the best model in this study. Model 12 is the EarthCARE model! Why is there no CALIOP model? … model 7 (without lidar ratio and 1064 nm depol ratio information)? Please comment on that!

Response: The focus of the paper is on the measurements acquired by airborne HSRL systems and not on previous measurements made by a less capable backscatter lidar such as CALIOP. Since ATLID provides HSRL measurements at 355 nm like the airborne HSRL-2 system, it was straightforward to also include a model that represents the ATLID configuration. To accurately represent the CALIOP system would require developing and using models that use total attenuated backscatter and not unattenuated backscatter measured by HSRL systems. This goes beyond the scope of the paper.

As mentioned in our response to the earlier comment, and in the additional text now added to the paper in section 2.3, of the models discussed in this paper, Model 5 is probably the most appropriate to study the CALIOP configuration.

14. In the case of models 7-11: Either BSC or EXT, but always LR is used! Does that mean: When BSC and LR is included, automatically the information about EXT is available, andis not needed? Please explain why a model that uses BSC plus EXT plus LR makes no sense!

Response: Correct, models 7-11 include the lidar ratio (LR) and either aerosol backscatter or aerosol extinction. Since the lidar ratio is the ratio of aerosol extinction to aerosol backscatter, including two of these parameters is sufficient; including all three would be redundant.

The following sentence has been added (Lines 336-337) "Models 7-11 include the lidar ratio and either aerosol backscatter or aerosol extinction since the lidar ratio is the ratio of aerosol extinction to aerosol backscatter. "

15. Line 351: Figure 4 shows models 1, 2, 3, and 11! I think one should show model 7 in this figure!

Response: The point of Figure 4 is to show how the GPR $PM_{2.5}$ retrieval performance improves as additional information is included. Therefore, the Figure starts with Model 1, which has the least amount of information and progresses to Model 11, which includes the most information. In addition, as shown in Figure 5, there are relatively large differences in performance among models 1, 2, 3 and 11, so it is easiest to understand these differences when viewing Figure 4. In contrast, there is very little difference in performance between models 7 and 11 so it would not be very informative to add Model 7 in addition to Model 11 in Figure 4.

16. Line 366, Figure 5: I am surprised that the use of BSC gives better results than the use of EXT. The extinction coefficient (overall scattering effect) is closely linked to the cross section of the particles, and PM is also well correlated with the particle cross section and thus with EXT. Is the reason that BSC is the better parameter related to the fact that the study only concentrates on the urban-haze aerosol type (mostly fine-mode aerosol)?

Response: As discussed in the response to the earlier question, the study does not concentrate on a single aerosol type. There was no restriction on aerosol type when constructing the training sets for retrieving PM concentrations. The reasons why the models that use aerosol backscatter provide better reasons than the models that use aerosol extinction are mentioned in this paper in this section (Lines 368-380). These reasons are:

1) aerosol extinction profiles generally have vertical resolutions of between 150-300 meters. To avoid contamination from the surface return, the lowest aerosol extinction value is restricted to be generally 200-350 meters above the surface. In contrast, the near-surface aerosol backscatter can be derived considerably lower, generally around 60-100 m above the surface, and so will likely be better correlated to surface network measurements.

2) HSRL measurements of aerosol backscatter are computed from the ratio of the total (aerosol+molecular) signal to the molecular signal, in contrast to the measurements of aerosol extinction, which are computed from the derivative of the molecular signal (Hair et al., 2008). Computing this derivative leads to more uncertainty than computing the ratio; hence the near-surface aerosol backscatter is less noisy and has generally smaller uncertainties than near-surface aerosol extinction.

17. Section 3: The result section shows interesting results and the full potential of airborne aerosol HSRL observations to quantify the pollution state close to the ground. I have no questions here!

Response: Agreed. The results are interesting.

18. Figure 8: The in-situ observations (EPA surface stations) are not easy to see. Maybe a bit larger symbols will help?

Response: In Figure 8 as well as in Figures 9, 10, and 11, the symbols representing the surface station measurements have been enlarged and outlined in black to make them easier to see. These figures, as well as several others, have been revised to improve the accessibility for readers with color vision deficiencies as per the AMT submission directions at https://www.atmospheric-measurement-techniques.net/submission.html.

19. Figures 12-15 show convincing (excellent) results. But as a critical reviewer my question would be? Can we use the developed approach if we have totally independent data sets, e.g., lidar observations over Beijing, Shanghai, Wuhan, Pearl River Delta in China, or over polluted Cairo, Egypt, Dakar, Senegal, Nairobi, Kenia, or over Paris and London in Europe, or Tomsk in Siberia or Fairbanks in Alaska? Any comment on that would be fine! Do we need always complex data sets of lidar and in situ observations in ML efforts for each region of the world, before we can make trustworthy use of lidar observation?

Response: This is a bit complicated. Based on this comment and the reviewer's earlier comments regarding ground-based lidars, it sounds like the reviewer is asking about the use of ground-based lidars in each of these locations to derive PM concentrations. If the desire is to use lidars at these fixed locations to derive profiles of PM concentrations in the manner described in this paper, it probably would be best to use a single training set based on measurements of many different aerosol types when analyzing these data and when attempting to interpret the results. The main reason would be that many of these locations would be likely to see a variety of different aerosol types with different levels of PM so a training set that was developed using a variety of measurements corresponding to different aerosol types and levels would be needed. This training set could be developed using measurements from these lidars assuming that coincident and collocated measurements of PM were available for training. It would also be important that measurements from these lidars had the same or very similar temporal and vertical resolutions and averaging characteristics.

The following text has been added to the Summary and Conclusion section (lines 780-785):

"The ML methodology described here can also be applied to ground-based and space-based lidars. For the former, it probably would be best to use a single training set based on measurements from several locations to include a wide variety of different aerosol types with different levels of PM. This training set could be developed using measurements from several of these lidars assuming that coincident and collocated measurements of PM were available for training. It would also be beneficial that measurements from these lidars had the same or very similar temporal and vertical resolutions and averaging characteristics."