



Comment on acp-2021-870

Anonymous Referee #1

Referee comment on "Assessment of NAAPS-RA performance in Maritime Southeast Asia during CAMP²Ex" by Eva-Lou Edwards et al., Atmos. Chem. Phys. Discuss.,
<https://doi.org/10.5194/acp-2021-870-RC1>, 2021

This paper evaluates the ability of the U.S. Navy's NAAPS-RA model in reproducing observations of extinction profiles and AOT in the vicinity of the Philippines during a recent field campaign. While both anthropogenic and biomass burning aerosols are transported into this region, the high frequency of cloudiness makes using satellite measurements of AOD problematic. The low frequency of AOD retrievals also means that NAAPS-RA has little information to constrain the simulation of aerosols in this region via data assimilation. Therefore, it is useful to take advantage of airborne HSRL-2 measurements to evaluate the performance of the model, in addition to other measurements that can be used to understand the factors affecting extinction profiles. Evaluation of extinction is complicated since errors can arise from many sources as pointed out by the authors. In general, the paper is well written, but the discussion of the evaluation methodology and the interpretation of the results need improvement.

General Comments:

1) The introduction provides some motivation of NAAPS-RA, which includes aerosol-cloud relationships. However, it is not clear how that can be accomplished with an offline model. The authors do not describe how the model results are used to compute CCN and/or aerosol-cloud-radiation interactions. The paper focuses on aerosol optical properties which are important for aerosol-radiation calculations and also indirectly affecting clouds by modifying heating at the surface and heating profiles. It does not make a connection between extinction and aerosol-cloud interactions. The paper does a nice job at quantifying the errors in simulated extinction and AOT, but the evaluation seems disconnected from the introductory material. In addition, are the errors significant for other potential uses? It is not clear "how good is good enough". It would be useful to describe in the conclusion/summary the implications of this work for NAAPS-RA applications.

Response: Good suggestion. We have altered the introduction to not include a discussion of CCN and/or aerosol-cloud-radiation interactions. We hope the reviewer finds the introduction has been streamlined to what is actually accomplished in the paper.

2) Apparently other sources of uncertainty in simulated extinction, such as aerosol mass, is left to a subsequent study. I am torn about that approach taken by the authors. While adding that component to this paper would increase its length and complexity, the paper seems

incomplete without it. After reading the paper, the main outcome is a straight- forward evaluation of the simulated aerosol optical properties (extinction and AOT) from NAAPS-RS for CampEx. Performing the sensitivity calculations with RH only partially explores the possible sources of uncertainty. Therefore, the reader is left with the perception that there are many uncertainties and conundrums (mass concentrations higher than observed while extinction is too low) that are unresolved. There is some comparison/evaluation with observed aerosol mass, but apparently there is not sufficient analysis in the present study since there are a lot of statements sprinkled in the results sections about material left to future studies. If the authors wish to leave the manuscript in its present form, they should more clearly articulate in the purpose of the present study versus a subsequent study.

Response: Great point. We have altered to the text so that it is very clear that the main objective of this paper is to examine relationships between model errors in RH and model errors in AOT and extinction.

For example, lines 122-124 states: “For this reason, we focus mainly on how replacing modeled RH profiles with dropsonde profiles affects NAAPS-RA simulations for AOT and extinction.”

Lines 268 – 269 says: “The main objective of this work is to investigate how correcting errors in simulated RH affects model outputs for AOT and extinction.”

There are also some aspects of the model assumptions that are not commented on. For example. The mass fractions shown in the supplemental material show relatively large mass fractions of sea-salt far above the surface. Does this seem realistic? There are lots of aerosol measurements from other aircraft campaigns that could be tapped into to at least in general confirm whether or not that seems to be a reasonable assumption. The authors mention in a couple of places that data assimilation may introduce some uncertainty in the mass fractions.

Response: To keep the paper focused on relationships between AOT, extinction, and RH, we have eliminated the plot that the reviewer is mentioning here. We no longer worry about modeled vertical profiles of modeled mass fractions nor do we use the HSRL-2 aerosol type product anymore as we agree that these distract from the main purpose of the paper.

3) The measurement and modeling comparison is rather complicated. I suggest the authors look over Section 2.4 – 2.9 to explain as best as possible the methodology and possible consequences on the results. Part of this is organization. Section 2.4 just provides a broad description of the strategy, which is talked about in more detail in Sections 2.5 – 2.9 and those sections really should be sub-sections within 2.4. The evaluation strategy complication seems to arise from the flight paths and sampling strategy used. Ideally two aircraft are needed: one to obtain the HSRL-2 extinction profile sampling while the other aircraft sampling aerosol mass that coincides with the HSRL-2 measurements as was done during TCAP (Berg et al. JGR, 2016).

Response: The methods section has been rewritten entirely. We believe that the flow is much clearer now (and we thank the second reviewer for his/her suggestions on how to structure the methods).

4) One of the conclusions of the paper is that simulated errors in RH were not the primary source of uncertainty in simulated extinction. This is not backed up with sufficient evidence. Other sources of errors seem to be left to another study and assumptions used in the aerosol water uptake calculations (which may be minimizing the sensitivity to RH) are not fully explored.

Response: We have adjusted the conclusions so that they only reflect findings from the paper and they do not speculate about things that we did not directly investigate.

Specific Comments:

Line 124: Reid et al. (2021) has not been submitted. Papers that are in preparation or in submission stage should not be cited. If the paper is published by the time this paper is accepted, it could be included. There seems to be sufficient discussion on the field campaign measurements used in this study.

Response: Great point, we have deleted this citation.

Line 151: Please include the time period of the campaign here. It is included later in lines 159-160, but it would be useful to include it up front.

Response: We have included the time period up front (in the first sentence of Sect. 2.1, Line 141).

Lines 151-159: The interests of the campaign could be applied to many regions of the world. What is needed here are some specifics as to the value of data collected around the Philippines. Perhaps some of the material in lines 168-170 could be moved here to provide a better motivation of the campaign for the reader.

Response: Good suggestion. The section has been rearranged.

Lines 175-177: MLH and aerosol classifications are derived products from the HSRL-2 measurements, but here they are put at the same level as the primary measurements (it measures backscatter to that is not mentioned). Most readers will not know that, so putting all this information at the same level is a bit deceiving. Perhaps including citations for HSRL-2 and the other products should be included here.

Response: We make it clearer that the MLH is a derived product (see Lines 201 – 202).

Lines 173-187: Where black carbon measurements available, i.e. from SP2? It would seem that those measurements would be useful in identifying anthropogenic and BB plumes, as well as aging of BB plumes (via coating of BC particles).

Response: Unfortunately the SP2 had a lot of issues during this field campaign due to the higher temperatures. Here is the exact text from the data archive:

“During CAMP2Ex flights, excessive heat caused YAG laser power to decrease sufficiently so incomplete incandescence was observed. These data are not recoverable and have been removed, sometimes resulting in large periods of -9999.”

For example, there are no data available for a large portion of research flight 9, which is considered the “smoke” flight of this campaign.

Line 186: Perhaps change “particles” to “cloud droplets and aerosol”. Just saying particles might imply to readers that only aerosols are measured.

Response: Done.

Line 211: Does NAVGEM include feedbacks between aerosols and meteorology via radiation and clouds? In areas of high aerosol concentrations, such as the biomass burning plumes examined in this study, aerosols can affect the meteorology which would then be used to drive NAAPS-RA.

Response: The version of NAVGEM that drives the NAAPS-RA does not include feedbacks (both radiative and microphysical) from aerosols onto meteorology. It is a one-way influence of meteorology onto aerosols through impacting aerosol emissions, transport, chemistry, hygroscopic growth, and removal. NAVGEM does have some capability of carrying interactive aerosols that include radiative feedbacks from aerosols in some model versions. However, these works were exploratory and computationally costly and didn’t end up in the operations (except for the fully coupled atmosphere-ocean-sea ice model, more on the climate side). Furthermore, aerosols are not in the NAVGEM-analysis that drives NAAPS-RA.

Line 219: Does the phrase “species-dependent mass scattering” mean that the model treats aerosols as an external mixture? If so, it might be useful to explicitly say that so that the reader better understands the assumptions in NAAPS-RA. Atmospheric particles are often a complex mixture of different species. Some models treat aerosol optical properties as internal mixtures which is the other extreme. In reality, aerosol populations often complex in that some regions may be more externally mixed and more internally mixed in other regions.

Response: Yes, external mixture. Stated in Lynch et al. (2016): “Aerosol particles in NAAPS are treated as external mixture of the aforementioned species and do not interact with each other.”

Lines 219 – 224 now state: “Lynch et al. (2016) provides a full description of NAAPS-RA, but in short it is a chemical transport model simulating the four-dimensional distribution of four externally mixed aerosol species: dust, sea salt (both of which are dominated by coarse mode [$> 1\mu\text{m}$] particles), open biomass burning smoke, and a combined anthropogenic and biogenic fine (ABF) species infrastructure that incorporates secondarily produced species such as sulfate and organics (both of which are dominated by fine mode particles [$< 1\mu\text{m}$]).”

Lines 237-239: Aerosol water significantly affects extinction in regions of relatively high RH, but is not included as a specie in NAAPS-RA. Instead, it seems that aerosol water is diagnosed when computing extinction. The issue I have is how MODIS AOT is used for assimilation. If the data assimilation process adjusts the four species to be close to the observed AOT when neglecting aerosol water, then the NAAPS-RA should always exceed the observed AOT once water uptake is accounted for. I am probably missing something important here that is not described.

Response: Data assimilation technique described in Lynch et al. (2016) in equations 13 – 15. Aerosol water is not neglected.

Lines 245 – 247 now state: “Corrections in τ_i are converted to changes in c_i using the optical properties for that species and the simulated meteorological conditions (e.g., RH).”

Line 338-340: It would seem that a more appropriate comparison is to average the 15-m HSRL-2 range gates within the model vertical grid cell rather than just take the points closest

to the mid-point of the model grid. I assume the model is assuming an average within its cell, so a coarse grid spacing will not resolve large gradients in extinction. So averaging the HSRL-2 data would seem to be a better approach, but it may not change the overall conclusions of this study. This can also be applied to the dropsonde comparison described starting on line 344.

Response: Excellent suggestion. We changed our entire method to average in situ values to the spatial resolution of the model. See Sect. 2.5.3.

Line 254: After reading Section 2.4, not using the other wavelengths from HSRL-2 seems to be a missed opportunity. Instead, the evaluation focuses only on 550 nm. Is that because NAAPS-RA does not account for aerosol size distribution? Atmospheric models that can account for aerosols in their radiation calculations simulate their effect on all wavelengths, not just at one. Or is AOD at 550 nm the primary purpose of NAAPS-RA? Some discussion would be useful to describe why the evaluation focusses only on one wavelength.

Response: Lines 194 – 200 now say: “This study focuses on retrievals at 532 nm to provide the most impactful model evaluation. As discussed in Section 2.4, NAAPS-RA is a bulk model, that can output AOT at over two-dozen wavelengths. Functionally these wavelengths are coupled to 550 nm, which is a widely-used wavelength in aerosol modeling and satellite remote sensing. Although we calculate model outputs at 532 nm, key findings from this work are still relevant to NAAPS-RA simulations at 550 nm. Given this, and that extinction and AOT are retrieved with the HSRL-2, we focus on the benchmark green wavelength in this study.”

Line 366: I assume that only FCDP measurements outside of clouds are used.

Response: Yes, as stated in lines 184-185 we only used data collected through the isokinetic inlet which was used in cloud-free air.

Line 397: This may be an overly broad statement. There are a wide range of aerosol models and the degree to which aerosols are parameterized varies. NAAPS-RA does have a simple treatment, since it only predicts bulk aerosols for four species. Other aerosol models are more explicit in predicting size resolved mass and number for a larger number of species.

Response: Lines 344-345 now read: “Aerosol models, such as NAAPS-RA, are heavily parameterized and are often challenged by the properties of individual air masses.”

Line 401: I understand trying to make the connection between the evaluation and CCN; however, the use of “representative” is misleading in this context. The authors are evaluating extinction, but CCN depends on aerosol size (which is neglected in NAAPS-RA) and hygroscopicity (via relative mass specie contributions). As noted by the authors in other places of the manuscript, extinction and AOT can have compensating errors – so how extinction alone relates to CCN is problematic.

Response: We have removed all discussion of CCN from the paper.

Line 400-403: This text is confusing. First they state that the performance focuses on the

ML (even though earlier they note 3 layers that are used for the evaluation), then they say evaluation of the performance in the PBL is the subject of another paper. Is there a difference in the ML and PBL, since these terms are often treated interchangeably? This seems to be the second area (in addition to aerosol mass?) that is left to another study?

Response: We completely understand the confusion as this was written poorly. The bulk comparison that includes modeled and retrieved AOT and extinction from every flight during CAMP²Ex examines all available altitudes. The case studies focus on the mixed layer only. We hope that this is clearer in Sects. 2.5.4 and 2.6.

Line 405: The authors mention one HSRL-2 profile is used in the 1 x 1 deg box. Why not horizontally average the extinction profiles within the 1 x 1 deg box? The authors do not show any time-height profiles of HSRL-2 extinction to know whether there are large spatial gradients or not.

Response: As mentioned above, we have altered our method entirely and we are now averaging HSRL-2 retrievals vertically and horizontally to match the spatial resolution of the model.

Line 406-407: I am confused by this statement. It sounds like the nephelometer, AMS, and FCDP measurements were usually not available in the 1 x 1 box. Is this because that box is chosen because of the dropsonde location which is made at high altitude? So you are using that data at lower altitudes (which may be in a different 1 x 1 box) for comparisons? This is obviously not ideal but one has to deal with the aircraft measurements you get. Ideally would be useful for the aircraft to also obtain an aerosol profile in the same column as the dropsonde – but I assume this rarely happened. It would be useful to reiterate the assumptions here. The comparison methodology is getting quite complex at this point.

Response: We agree that this was confusing. We have adjusted our method so that all data used in the case study must be from the same 1 x 1 grid.

Lines 352 – 355 now read: “Airborne observations from the AMS, FCDP, and nephelometers were filtered to isolate data collected below the average MLH for each case study flight. We identified the 1° × 1° grid with the most available data for these variables, and this became the grid used to represent a particular case study.”

Sect. 2.6.2 describes how HSRL-2 and dropsonde values were isolated for the same grid.

Line 465: The way this sentence is phrased implies the observed MLH is biased, but I assume the authors compared the model MLH to each of the three dropsonde methods and the HSRL-2 and the bias refers to the model. If this is not a comparison with the model, what does the bias in Table S2 refer to?

Response: We decided to only use the HSRL-2 MLH product to determine MLHs, so the table that this comment is referring to is no longer showing a column for bias.

Lines 493-499: What is missing in this paragraph is noting that while the correlation is reasonable, there is still a lot of scatter in Fig 2 with some differences as large as two orders of magnitude.

Response: We have changed the presentation of this figure entirely. The axes are no longer on a log scale.

Line 514: Does NAAPS-RS include wet-scavenging? I do not recall that being mentioned in the model description. 1 x 1 deg grid spacing is coarse, but I assume the parent meteorological model would simulate clouds and precipitation in some way that could be used for wet scavenging?

Response: Yes, NAAPS-RA includes wet scavenging. This is discussed in greater detail in Lynch et al. (2016) in Sect. 2.2.3 (titled “Sink Processes in NAAPS”). We have since removed the discussion that the reviewer is mentioning.

Line 523: Doesn't Table S2 contain the model bias in MLH?

Response: We simplified the method so that we are only using MLHs from the HSRL-2 product. We do not consider NAAPS-RA simulations of the MLH, so there is not a model bias discussion for MLH.

Lines 523-529: Seems that another explanation might be the assimilation of MODIS AOD and how that is handled in the vertical. Has past evaluations of NAAPS-RS provided any guidance on that? Although there are not many retrievals in this area, presumably aerosols from other regions (which would be subject to assimilation) would be advected over the Philippines.

Response: We have eliminated the part of the paper pertinent to this comment.

Line 537: Does this statement mean vertical variability between 145 and 500 m or horizontal variability of extinction in that layer. Not clear.

Response: We have eliminated the part of the paper pertinent to this comment.

Line 552: Change “Biases” to “Flight-averaged biases”. Figure 2 has the biases for each profile, but it looks like Fig. 3 averages them for each flight.

Response: We have eliminated this figure.

Lines 570-578: I wonder if Figures 5 and 4 can be combined in some way to highlight the differences which are difficult to see currently. Is it important to differentiate the flights in these plots? If not, the two figures could be combined showing the simulated AOTs using the model vs observed RH as different colors. Then the original figures could be moved to the supplemental information.

Response: We have combined these figures (see Fig. 2). We created 2D histograms in response to a comment from the second reviewer.

Line 576: I agree that other parameters in the model might be contributing to uncertainties in the simulated AOT, however, I would have expected changing RH would

have had a much larger effect. Figure 4 indicates there were some cases in which observed RH was 20% higher than simulated below 500 m – and the differences could be larger at higher altitudes. Since NAAPS-RA is using simplified techniques to represent aerosols – how good is its method of computing aerosol water uptake? There are aerosol box models available with complex thermodynamical representations that could be used to estimate aerosol water uptake and compare those results with the methodology in NAAPS- RS.

Response: We investigated this further. We learned that changes in extinction are most sensitive to the magnitude of the extinction coefficient and the value of the dropsonde RH being substituted into the model. So even though there are instances where there are large differences in simulated and measured RH, this may or may not correspond to a large change in extinction. We agree that comparing NAAPS-RA methodology to a more thermodynamically complex aerosol box model would be interesting, but we leave this for future work.

Line 620-627: I appreciate this discussion on the simulated mass concentrations with relation to observations. In line 626 the authors say that mass concentrations need to be increased, but the fine mode mass is similar to observed and coarse mode mass is higher than observed. So the authors are saying they would have to create another error to fix a current error in extinction. There is a mystery here, and it seems that another study would be needed to understand the true source of error(s) in the extinction calculation. I wonder if the source of uncertainties are the assumptions used in the simple treatment of hygroscopicity and/or aerosol water.

Response: This reviewer is completely correct. However, we cannot say with certainty that simulated fine and coarse mass concentrations are higher than “observed” values. There are many assumptions used in our calculations, especially for coarse particle mass concentrations. More work would need to be done to determine robust “observed” fine and coarse particle mass concentrations before the errors in extinction can be solely blamed on modeled water uptake for the four aerosol species considered in NAAPS-RA.

Figure 6 - 8: It would be better to plot a) and b) in the same panel to better see the differences between the simulated and observed RH. After reading the figure caption multiple times, I still do not understand what the vertical black line and gray shading is.

Response: Great suggestion, these profiles are now plotted on the same panel. We hope that the shaded areas are described more clearly. In short, they reflect the range of extinction coefficients the model outputs when a range of in situ gamma values are used. The shaded profile extends to the coefficients calculated with the mean gamma value MINUS one standard deviation to the left and the shaded profile extends to the coefficients calculated with the mean gamma value PLUS one standard deviation to the right.

Line 657: The authors mention the possibility of aerosol mass increasing with height. Why not use the AMS measurements confirm this? Are some of the contradictions (i.e. simulated aerosol mass larger than observed when simulated extinction is slightly lower than observed) due to the different boxes (Fig. S2) where aerosol mass and extinction profiles are compared? With smoke plumes, there could be large spatial gradients.

Response: We have eliminated this discussion on mass. We want to keep the paper focused on what we can actually do well, and speculating about vertical profiles in mass is not something that we want to include anymore.

Line 677: Again it would seem that the AMS could be used to evaluate the ABF and smoke species in NAAPS-RS. At the end of the paragraph, they state that more work is needed – so I presume this will be the subject of a future paper?

Response: It is true that the AMS could be used to verify ammonium sulfate and organic material, but this does not align with how NAAPS-RA categorizes particle types. For example, NAAPS-RA places organic material in both the ABF and smoke categories. We feel that another study needs to focus on the mass component because of complexities such as the one described.

Lines 781-782: I felt that only the authors only presented some preliminary speculation as to what the possible errors may be. There were no concrete conclusions here regarding what the specific errors for specific cases, so no tangible understanding is provided in this paper.

Response: This is understandable that this reviewer feels this way. We want to provide useful and tangible results, but it is difficult without a full picture of all of the variables affecting extinction. We hope that our conclusions are perceived as more useful now since we have focused the paper more on what we can study well (mostly effects of errors in model RH on extinction simulations).

Line 802: The conclusions are probably not applicable to the entire modeling community. It seems that the uncertainties are largely applicable to NAAPS-RA, and perhaps to other similar classes of aerosol models such as GOCART.

Response: Lines 794 – 796 now say: “Findings from this work can assist members of the modeling community to improve AOT forecasts in SEA and beyond.”

We hope that the use of the word “members” conveys that we are talking about a subset of the entire modeling community.

Lines 805-807: Are there other studies evaluating NAAPS-RA in other locations with other field campaign observations that might have urban and biomass burning sources? If so, it would be useful to compare the present work with results from those locales.

Response: Yes, Lynch et al. (2016) does a great job evaluating NAAPS AOT outputs around the world against AERONET sites and in various conditions discussed throughout the paper. We mention how work from Lynch et al. (2016) agreement compares to our study on Lines 790 – 794.

We write: “Fig. 12 in Lynch et al. (2016) shows R^2 values for comparisons between simulated (NAAPS-RA) and retrieved (AERONET) AOT values from around the world are slightly lower than our values for this study. Although we can see AOT agreement does not fluctuate too much across the globe, the driving forces behind disagreement in these locations are presumably uncertain.”

The line we have about biomass burning and East Asian pollution is more about NAAPS-RA incorrectly modeling the hygroscopicity of these air masses. There are many works evaluating NAAPS-RA AOT simulations in the presence of biomass burning, but to our knowledge, these works do not comment on the how well NAAPS-RA is modeling the hygroscopicity of the smoke particles.



Comment on acp-2021-870

Anonymous Referee #2

Referee comment on "Assessment of NAAPS-RA performance in Maritime Southeast Asia during CAMP²Ex" by Eva-Lou Edwards et al., Atmos. Chem. Phys. Discuss., <https://doi.org/10.5194/acp-2021-870-RC2>, 2022

The manuscript describes an evaluation of the performance of the NAAPS Aerosol Reanalysis over the Philippines region based on airborne measurements conducted in the framework of CAMP²Ex. The introduction is suitable to motivate the importance of information from aerosol reanalysis models in regions such as the Philippines where cloudiness poses a huge obstacle for regular satellite observations. The subsequent presentation of the work leaves much to be desired, though. Overall, the paper is lengthy and rather unfocussed. The authors focus on presenting all they have done in the manuscript and an additional 25 pages of supplement, rather than fitting it into an easy-to-follow story for the readers. This reviewer therefore suggest major revisions before the work could be considered for publication in ACP. Some specific issues are listed below:

The manuscript is written mostly in past tense. Please note that everything that still holds today, i.e. results show, should be in present tense

Response: This has been addressed. Thank you for the suggestion.

The Introduction is very broad and long given that the work deals with a relatively straightforward comparison of AOT and extinction coefficient from modelling and measurements and the effect of relative humidity and hygroscopic growth on the model output. Could the text be sharpened towards what is presented in the paper?

Response: The Introduction has been shortened and we tried to focus more on what this paper actually accomplishes.

Section 2 is too fragmented for my taste. First, it should really be "Data and Methods", Second, the description of the measurement data, the model and its output, and the comparison methodology should be clearly structures along the lines of, e.g., (i) measurement campaign, (ii) airborne in-situ measurements, (iii) airborne lidar measurements, (iv) NAAPS-RA model description (the relevant part that is needed for this study, and (v) comparison approach and model refinement.

Response: Great suggestion. We now follow this order.

Table 2: is there any source for these values?

Response: Yes, these are now cited in the caption for the table.

Mixed layer heights: Is there any conclusion on which method is used in the final comparison? I read about an abundance of methods, but later in the presentation of the findings, there's just the parameter MLH.

Response: We agree that the abundance of methods made this an unnecessarily difficult section to understand. We changed our method to only consider MLH from the HSRL-2 product (i.e., MLHs were no longer calculated using dropsonde data).

Section 2.6: The method to replace land-contaminated grid cells with the nearest neighbouring grid cell over open water is problematic as it leads to comparing apples and oranges. Better omit these data points rather than introducing comparisons that complicate the entire procedure. It might be better to instead relax the criterion that lidar profiles need to be in the vicinity of a dropsonde release to increase the number of comparison pairs. The findings later show that the shift to dropsonde RH has little effect on the modelled aerosol-optical properties.

Response: Very good point. We no longer consider land contaminated grids, but we still only consider grids containing dropsonde releases.

Later in the text the authors refer to 1 degree grid. It is not clear if this is a reference to an individual grid cell or a sub-grid within the cell.

Response: We only ever consider 1 x 1 degrees, and sub-grids are not considered. We hope that by specifying "1 deg. x 1 deg. grid" every time we talk about a grid that this will always be clear.

Section 2.8: It doesn't make any sense to me that the authors compare lidar extinction coefficients at altitudes closest to the mid-point of a model pressure layer to the modelled value. Lidar profiles can be very noisy and picking a value at just a single height risks to introduce all this noise of real-world data into the comparison. I'd suggest to work with a mean lidar extinction coefficient averaged over the width of the height layer covered by the corresponding pressure layer.

Response: We agree completely. Thank you for catching this mistake in the method. We now vertically and horizontally average all HSRL-2 data to match the resolution of the model. See Sect. 2.5.3.

Figure 2: please consider a different presentation of the data, e.g. as 2d histograms. It is really hard to extract useful information from the point clouds that are currently presented. The information regarding the research flight is not necessary here, as it is presented also in the next figure. In the context of this figure, I wonder if there is a minimum AOT that NAAPS-RA can represent?

Response: We agree presentation was not as good as it could have been in this figure. We adopted the reviewer's idea of 2D histograms and did not include flight-specific information. The minimum AOT is set by whatever we can verify with AERONET, so in that case it is around 0.01-0.02 (see citations in paper for Dubovik et al. [2000] and Eck et al. [1999] at the end of Sect. 2.4).

Figure 3: please revise into a format that allows to extract the information; or present the findings as a table?

Response: This figure (and flight-specific information such as this) has been eliminated from the paper. We do not feel it is necessary to tell the story anymore.

Figure 4: see comments on Figure 2. All I can see is noise.

Response: We changed presentation format to a 2D histogram as suggested above.

Figure 5: Omit or move to supplement as there is almost no change compared to Figure 2.

Response: We have combined with Fig. 2 and have changed the presentation format to a 2D histogram as suggested above.

Case studies: I understand that it is very interesting to assess the performance of the model under very different aerosol conditions. However, the current presentation of the case studies leaves much to be desired. Rather than shedding light on the model's performance, they are raising more questions than they answer. It is not sufficient to simply leave the deeper investigation of the issues raised by the case studies for later publications. The authors should at least formulate a solid hypothesis as to the nature of the inconsistencies.

Response: We agree that it can be frustrating that we cannot provide more explanation as to why modeled extinction is not in agreement with HSRL-2 retrievals. We adjusted our language to focus only on what we can study, and we are much less speculative on the mass concentration part of the case studies. We hope this reviewer finds our case study analyses more focused on the parts we can evaluate (error due to modeled RH and hygroscopicity).

Table 3 and approach to adjust gamma: Is it possible to revise the table for easier access to the findings; maybe by use of colour coding? Also, does it make sense to just adjust the modelled gamma to the value provided by the in-situ measurements? I reckon that the modelled value is the result of some mixing of the numbers in Table 2. Does this allow for a deeper view into the contribution of the individual aerosol types to gamma by adjusting the individual values that go into the mixing rule rather than exchanging the output? This might also lead to more consistent adjustments? Or are you doing it like this already?

Response: Table 3 has been simplified drastically. We now only focus on mixed layer AOT to evaluate a particular case study.

Sorry if it was not entirely clear before. We did not use in situ gamma values to adjust the gamma value of individual aerosol types (i.e., individual gamma values for smoke, ABF, dust, and sea salt) in the model. We take the in situ gamma value and use that for each species.

Line 406 – 407 states: “Here, we use the same in situ γ in Equation 2 for all four aerosol types.”

The modelled aerosol optical properties seems to be dominated by the fine mode. It is surprising to me that a modelled coarse-mode mass that is one order of magnitude larger than the measurements is supposed to have no effect. Or is this effect systematic as the same difference seems to be found for all case studies? I think that this topic should be explored deeper in the paper.

Response: We have eliminated this part of the analysis. We do not focus as much on modeled particle mass concentrations/fractions as we do not have the in situ data to really evaluate this. We eliminated it so that the paper appears more focused on what we can actually do well.

The work ends with a short list of conclusions that don't seem to warrant to amount of material presented in the manuscript and the supplement. I recommend that the authors either streamline the presentation into a short paper or make an effort in exploring all the questions raised by the case studies.

Response: We agree that the paper attempted to discuss more than we could actually support with the data available. We have streamlined the paper to focus as much as possible on relationships between errors in modeled RH and modeled extinction and AOT, as well as errors in modeled hygroscopicity for the case studies. The amount of material in the supplement has been greatly reduced in order to only present the most relevant information. We hope the reviewer finds that the conclusions are well-supported and the paper more streamlined.

List of relevant changes to the manuscript:

1. Method was completely changed in that we now average in situ observations to the resolution of the NAAPS-RA model both vertically and horizontally.
2. The analysis focuses on optical property evaluation at 532 nm instead of at 550 nm in order to keep HSRL-2 data at their native wavelength.
3. HSRL-2 extinction coefficients are calculated from the cumulative AOT product (instead of the extinction coefficient product) to reduce influence from noisy data.
4. NAAPS-RA data were interpolated to locations of in situ observations.
5. Mixed layer heights are now only used from HSRL-2 data (instead of being additionally calculated from dropsonde data).
6. We investigate how changes in RH affect NAAPS-RA simulations of AOT and extinction at a deeper level than in the first manuscript.
7. There is a greater emphasis on bias and RMSE as a metric for agreement instead of just R^2 .
8. Figures showing comparisons between simulated and observed parameters are now shown as 2D histograms (instead of scatter plots that were hard to extract information from).
9. The “Mixed” case study is now considered a “Biomass Burning with Additional Aging” case study.
10. We place far less emphasis on comparing simulated and “in situ” particle mass concentrations.
11. We place far less emphasis on agreement per flight and more on agreement for different categories based on the type of bias in simulated extinction and RH.