Response to Reviewer

We thank the editor for reviewing the revised manuscript and related materials. We have carefully considered all the comments and revised our manuscript accordingly. We hope that we have addressed all the remaining concerns on the manuscript.

Here we provide the point-by-point responses to the reviewer's comments. All comments are in black, our responses are in blue, and corresponding changes in the revised manuscript are in red. We have also submitted a revised version of the manuscript and an alternative version that includes all the tracked changes.

We summarize our key messages as follows:

1. We address the possible impact of the cloud contamination and aerosol spatial gradient on the triple collocation analysis by repeating the analysis with minimal cloud contamination and low aerosol spatial gradient condition. And the error metrics of RSP and HSRL-2 are not affected.

2. We also address the possible impact of combining Terra and Aqua datasets on the triple collocation analysis if there is a bias between two sensors. Using a synthetic data approach, we demonstrate that adding bias to half of third dataset (randomly chosen) does not affect the error metrics of the first and second datasets.

3. We simplify the text in several places to make it easier to read.

Reviewer 3

Dear Leong Wai Siu and coauthors,

I am sorry for the slow review of your manuscript. I had difficulty finding reviewers and reviewer #2 is not responding to emails requesting a follow-up of his initial review. Accordingly, I have undertaken my own review of the manuscript and your responses to review #2. First let me note that it is clear to me that you put considerable efforts into your response and in most cases, I think your responses are fine. There are however, a few items which I think should be addressed before I can recommend the manuscript for publication. I have listed these below for your consideration.

Sincerely,
Roger Marchand

Comments:

- 1) As regards Reviewer #2's concerns about cloud contamination and gradients in the aerosol field.

  Specifically, Reviewer #2's comment was "Since you are collocating with MODIS, what does MODIS suggest about cloud-cover/fraction and all the issues that might make collocation (and retrieval) difficult? Also, what about gradients of aerosol nearby the collocation?"

  In response you have added a figure and material showing that the collocated cloud fraction and the Local COV in AOD is often quite small. These metrics certainly suggest these issues will not affect substantially your triple collocation analysis. But it would be better to show the impact. Specifically, it seem to me that you could simply reproduce the results given in Table #3, except where you have reduced the set of points used in the triple-analysis to those where MODIS collocated points have low cloud fraction (say 0 or 0 to 0.1), OR a Local COV < 0.1. I expect the values would be similar to those already given in Table #3, and would make very clear that these issues are not a problem for the analysis.

Thanks for your suggestion. We repeat the triple collocation analysis by only using MODIS quality flag $= 3$, MODIS cloud fraction < 0.1, and MODIS LCOV < 0.1 in Table R1. Unsurprisingly the error metrics of MODIS improve upon using stricter criteria but the error metrics of RSP and HSRL-2 do not change much. Therefore, it shows that the triple collocation analysis is not sensitive to the MODIS cloud cover, quality flag, and local aerosol spatial gradient.

We updated L281–305 and Table 3 in the revised manuscript.

Table R1: Triple collocation analysis of RSP, HSRL-2, and MODIS AOD.

		RSP		HSRL-2		MODIS	
	n_{triplet}	$\sigma_{\varepsilon_{\text{RSP}}}$	r_{RSP}	$\sigma_{\varepsilon_{\text{HSRL-2}}}$	$r_{\text{HSRL-2}}$	$\sigma_{\varepsilon_{\text{MODIS}}}$	r_{MODIS}
all available triplets	2344	0.0637	0.796	0.0273	0.926	0.0511	0.858
MODIS quality flag $= 3$	1592	0.0606	0.790	0.0259	0.922	0.0370	0.872
MODIS cloud fraction < 0.1	1349	0.0636	0.783	0.0222	0.945	0.0373	0.928
MODIS LCOV < 0.1	1210	0.0605	0.821	0.0272	0.930	0.0419	0.906

- 2) χ^2 and χ', What are the a priori?

  The additional material you added on the cost function (Lines 83-96) is problematic in that it introduces additional variables that are not defined. I suggest instead you simplify this text to read:

  "The algorithm uses an optimal-estimation method to minimize a cost function χ^2 that describes the difference between the observations and a forward model for the aerosol scattering, and includes a term that takes into account a priori knowledge of aerosol properties for the region. The cost function is developed and described in detail by Stamnes et al. (2018) and includes a metric describing the goodness of fit between the data and a solution, called χ'. Here we consider retrievals successful only when they have a χ' value less than 0.15. As regards the a priori knowledge we assume ..."

  To me the bigger question here is what did you assume for the a priori? Does this follow from Schlosser et al ?. If yes, you can simply state such. But the bottom line is you should provide enough information that another researcher could reproduce your results. If your a priori assumptions are not published you need to describe them.

  We adopted your suggestion to simplify the material related to the cost function. We follow the retrieval algorithm described in Stamnes et al. (2018). The a priori knowledge is obtained in a straightforward and conservative approach which averages the permissible range of each retrieval parameter, which is also employed as the first guess.

  We updated L83–89 in the revised manuscript.

- 3) Supplementary material

  This was a nice response to Reviewer #2 and a nice addition to the paper! But some reference to the supplementary material is needed in the main manuscript (otherwise, how does the interested reader know to look at the supplement). Perhaps note the analysis given in the supplementary material somewhere near line 265 (in the revised manuscript) where you first introduce the triple collocation technique.

  Thanks for your kind words and suggestion.

  We updated L292–294 in the revised manuscript.

- 4) Effect of Bias (between Aqua and Terra MODIS).

  Reviewer #2 asked "What if Terra or Aqua are biased differently?" You did not address the comment, directly. Perhaps you can use your same test approach (i.e. the supplemental material) to examine the effect of a bias? That is, add a bias of say 0.02 or 0.05 to a randomly selected 50% of data points. How does this affect the idealized results?

  Thank you for your suggestion. We use the same synthetic data approach to assess the potential influence of the biased data on the triple collocation analysis.

  Assume that the ground truth is given as one full cycle of sinusoidal function with mean = 0.1, minimum = 0, maximum = 0.2. We have 5000 data points stored in an array and we *randomly* add some bias to 50% of data points (i.e., 2500) to the third dataset. In python, the indices of these data points can be generated using numpy library with the following code,

```
rand = np.random.choice(5000, 2500, replace=False)
```

  We now have the experiments 1–3 with three synthetic datasets A, B, and C with a bias of 0.05 on 2500 data points of C (Table R2). The only difference is the SNR of dataset C (1, 5, and 25). All datasets are re-generated for each experiment.

Table R2: Summary of the synthetic experiments 1–3 with a bias of 0.05 on dataset C. Signal-to-noise ratio (SNR), error standard deviation σ, and error correlation coefficient r with respect to the ground truth are provided for each synthetic dataset.

Expt	Bias	A			B			C		
		SNR	σ_{ε_A}	r_A	SNR	σ_{ε_B}	r_B	SNR	σ_{ε_C}	r_C
1	0.05	5	0.0323	0.908	5	0.0315	0.914	1	0.0751	0.692
2	0.05	5	0.0315	0.913	5	0.0318	0.913	5	0.0401	0.871
3	0.05	5	0.0311	0.915	5	0.0317	0.913	25	0.0288	0.925

  All TC metrics of datasets A and B are highly consistent when the bias on dataset C is 0.05. While the true bias between Terra and Aqua is not known, it should not be too big given that the expected error of the MODIS Dark Target retrieval algorithm is $\sim$ 0.03–0.05 for the AOD range in the study region. Therefore, even if Terra and Aqua may be biased differently, the effect of this particular bias should be small.

  We expanded S1 in the supplement (in a separate file).

- 5) Figure 8 panel C.

  I think the vertical axis is mis-labeled and should be ”above-aircraft cloud mask fraction”. If not please explain how “below-cloud cloud cover” differs from “below-cloud cloud mask”.

  The vertical axis is not mis-labeled.

  We use the above-aircraft and below-aircraft cloud masks to detect the presence of cloud that leads to the contamination of aerosol retrievals. Cloud mask is a binary variable being either zero (absence) or one (presence), and is sampled at 1 Hz. To match with the HSRL-2 resolution, we compute the fraction of time if cloud mask is one (i.e., cloud is present) during each 30-s window of a 3-km HSRL-2 data point. Cloud fraction, in contrast, can range from 0 to 1.

  We include the panel C of Figure 8 because it is not necessary that the presence of cloud is related to high cloud cover fraction. The panel C shows that higher cloud mask fraction (more often presence of cloud) is associated with higher below-aircraft cloud cover.

- 6) ”Dynamic” cloud fraction

  Also, I don’t understand why the idea of “dynamic” cloud fraction has been introduced. This idea does not seem to be used anywhere? Or is the point that ALL of the cloud fraction analysis is based on pixels where there is no glint in the camera? If yes to the latter, then rather than introduce this term, perhaps simply states “we calculate cloud mask fraction using only pixels where there is no glint, where glint is defined as a scattering angle to the sun of $< \ldots$”.

  In an effort to minimize the influence of cloud contamination, an experimental cloud fraction/cloud mask dataset is produced to provide more information to data users. Because of the experimental nature, the current version actually provides three cloud fraction variables and we use the dynamic cloud fraction.

  We have adopted your suggestion to simplify the terminology as we exclude glint pixels in all cloud fraction analysis.

  We updated L345–346 in the revised manuscript.

- 7) Line 312.

  I think you wrote τ_c where you intended to write τ_f: That is $\tau_f < 0.7$?

  Fixed.

  We updated L317 in the revised manuscript.

- 8) Conclusion point three.

  While it is true that your filter improves the statistics given in Table #4, it also limits the maximum AOD. Thus, it would be highly problematic to use data filtered in this way to study, for example, large dust or fire events and effectively this removes the potential impact of these events on the overall statistics. In my view, this is OK as long a real events don’t have large optical depths in the region being studied, and this is an important caveat that should be discussed.

Also, can you comment, do larger RSP AOD retrieval errors (which you are filtering out) occur in situations where HSRL retrievals show aerosols are not the spherical particles (that you have assumed in the RSP retrieval)? Or is it mostly just cloud contamination?

Our final choice, the cost function filter, does not limit the maximum AOD. It only filters out RSP data points with cost function larger than 0.05 (i.e., inferior data quality). The range of RSP AOD before and after filtering can be inspected from Figure 3a and Figure 7a, respectively, in the revised manuscript. The removed RSP AOD data points are those with the largest bias with HSRL-2 (off the diagonal).

When depolarization ratio deviates from zero, the aerosol becomes more non-spherical. The range of column-maximum HSRL-2 depolarization ratio does not change much with RSP cost function, and neither do mean and median of depolarization ratio (Figure R1c). It may not be surprising because the AOD bias between RSP and HSRL-2 is not very sensitive to the column-maximum depolarization ratio as shown in Figure 9d in the revised manuscript. But we noticed that the AOD bias is more sensitive to the near-surface maximum depolarization ratio in the Figure 9e in the revised manuscript only when relative humidity is below a certain threshold. Therefore, we conclude that non-sphericity does not play a big role on the RSP retrieval errors all the time.

On the other hand, above-aircraft cloud contamination is common regardless of cost function, i.e., retrieval errors (Figure R1a). Below-aircraft cloud contamination increases with cost function (Figure R1b). The result suggests that below-aircraft cloud contamination often affects the quality of retrievals.

We updated L353-355 and L398-399 in the revised manuscript.

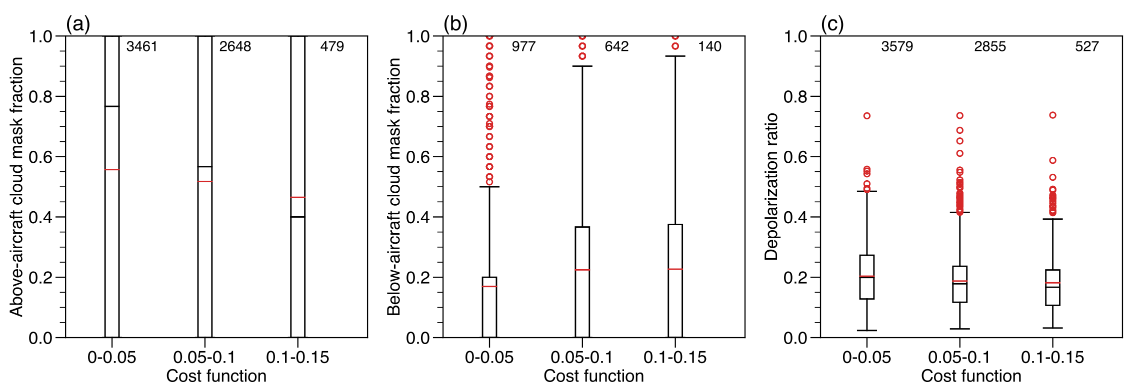

Figure R1: Boxplots of aerosol and cloud parameters stratified by cost function. (a) Above-aircraft cloud mask fraction. (b) Below-aircraft cloud mask fraction. (c) HSRL-2 column-maximum depolarization ratio. Numbers at the top indicate the number of each data group. The box and whiskers indicates the variability of each group (box, first and third quartiles; black line in box, median; red line in box, mean; whiskers, a distance of 1.5×IQR beyond the first and third quartiles; red circles, outliers).

Reference

Stamnes, S., Hostetler, C., Ferrare, R., Burton, S., Liu, X., Hair, J., Hu, Y., Wasilewski, A., Martin, W., van Diedenhoven, B., Chowdhary, J., Cetinić, I., Berg, L. K., Stamnes, K., and Cairns, B.: Simultaneous polarimeter retrievals of microphysical aerosol and ocean color parameters from the "MAPP" algorithm with comparison to high-spectral-resolution lidar aerosol and ocean products, Applied Optics, 57, 2394, doi: 10.1364/ao.57.002394, 2018.