Responses to Reviewer's Comments:

We appreciate the reviewer's comments and suggestions, which were very helpful in improving the overall quality of our manuscript. Basically, all the comments and suggestions were reflected in our revision. Our responses are listed below to each comment.

This article describes a study to compare different techniques for fusing aerosol optical depth products from two multi-spectral instruments viewing East Asia from geostationary orbit, GOCI and AHI, and evaluate the results during two field campaign periods. The topic is relevant given that these two instruments represent current state-of-the-art capabilities for diurnal aerosol observations from satellites. The hypothesis is that some type of ensemble mean will usually perform better than any individual member, and the paper compares 4 individual retrievals (2 each from GOCI and AHI), 4 different simple ensemble-mean combinations, and 3 different maximum-likelihood-estimate (MLE) combinations. In general, the hypothesis seems true, with the fused products generally overcoming different deficiencies in the individual products. However, the number of different permutations considered makes it difficult to focus on what attributes lead to the best improvements. The impact of bias correction should be clarified. Also, the current multi-panel figures make it difficult to see the differences. This work would make a valuable contribution to the literature if the clarity of the presentation can be improved. Specific suggestions are offered as follows.

SPECIFIC COMMENTS ————————————

The discussion of gap filling techniques starting at line 115 needs an introduction to provide context for the geostationary observations. While the need for gap filling in daily LEO observations is somewhat intuitive, it seems the simplest fusion of GEO products could produce a high yield without gap filling. Please clarify how the gap-filling applies to the current work.

- As reviewers commented, retrievals and applications using geostationary satellite observations cover many areas. Sorry for the lack of clarity in our originally submitted manuscript. Our aim was to provide optimized aerosol products from two different algorithm and two different instruments (GOCI and AHI). Therefore, this paper aims to produce the optimal fused AOD products where retrieved results are available, not to fill the gap where aerosol properties are not retrieved.

The organization of sections 2 and 3 was quite confusing to me; I had to re-read several times to understand what was being done. 2.1 is fine, simply describing the two AHI products. The first paragraph of 2.2 is fine, simply describing the two GOCI products. The second paragraph (beginning line 195) belongs in a separate section describing the different fusions, rather than in the GOCI algorithm section.

- *Thank you for your comment. This part has been moved to the beginning of section 3.*

Section 2.3 seems out of place; I suggest it be moved such that it is the last text before the Results section. Section 3 would greatly benefit from starting with a simple statement of the approach, e.g., we compare 4 different simple ensemble-mean combinations and 3 different MLE combinations. It is confusing to read about the MLE fusions in the Ensemble Mean section 3.2 (lines 242-244); I suggest you describe the FM entries of Table 1 in the next section about MLE method, instead of in this section about ensemble mean method. And at that point, note that the same members are included in F2 and FM2, F3 and FM3, F4 and FM1, as can be seen in Table 1. Further, I suggest you swap F1 and F4, so that the same members are included in F1 and FM1; this would make it easier for the reader to examine the differences in the figures. For example, adjacent panels (e) and (i) would be for the same members, similar to how adjacent panels (f,j) and (g,k) are for the same members, in Figs 2-3 and 5-8.

- *Thanks for your suggestion to reorganize, which improves the manuscript story flow. In the next session, we revised our manuscript to mention FM1-3, and moved Sec 2.3 to the last paragraph before the results section. Also, we swapped F1 and F4, per reviewer's suggestion to improve readability.*

Lines 256-261 (calculation of RMSE values) are confusing to me. If I understand correctly what you are doing, this could be explained much more clearly as follows. The locations of ground measurements are very sparse in comparison with the satellite coverage, so you choose to model RMSE as a function of NDVI. Then you bin all ground/satellite co-locations with respect to NDVI, AOD, and time, calculate the RMSE in each bin, and then apply this RMSE to every satellite pixel as a table lookup based on those 3 parameters.

- *Thank you for your advice. We revised these sentences.*

A similar comment applies to the description of bias correction technique (section 3.4, lines 267-272). It appears that bias correction is applied to the MLE fusions but not to the simple-mean fusions. It seems that bias-correcting the simple-mean fusions could easily be done and would provide a more direct comparison of the two techniques. And if you do this, you should be able to say something about the importance of bias correction in isolation.

- The simple average fusion field is an ensemble averaging technique, which utilizes the characteristic of finding a better value when multiple signals are averaged. In the main texts, we would like to mention that the accuracy becomes better (the less scattered), as we have more ensemble members. Our purpose was to show how well they matched the MLE fusion products through bias correction and pixel-based error fusion. However, to demonstrate the comments pointed out by the reviewers, the result of performing the bias correction is attached below

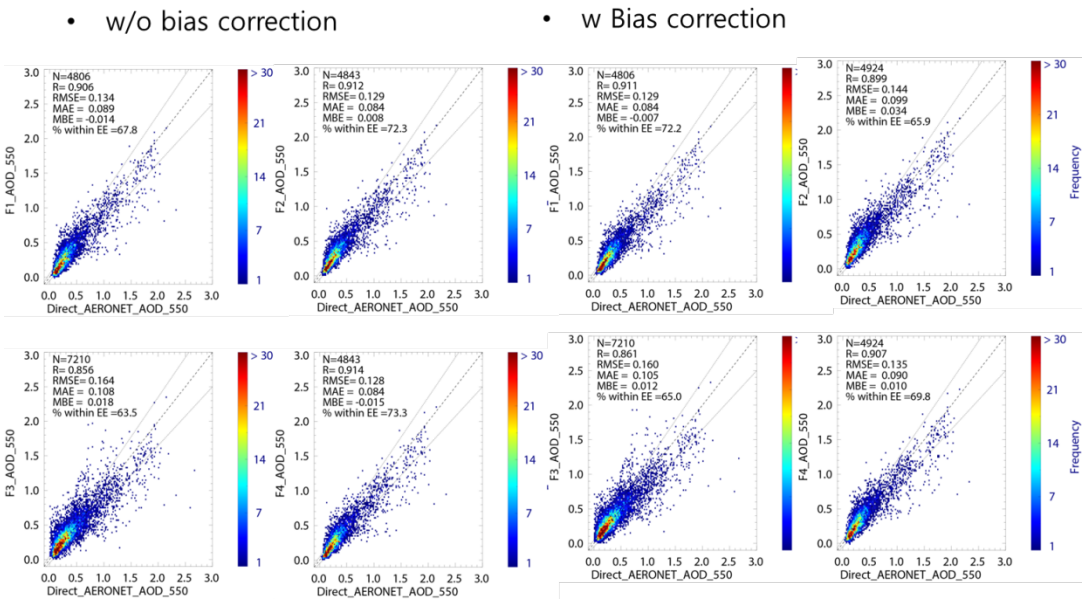

Figure 1. KORUS-AQ campaign

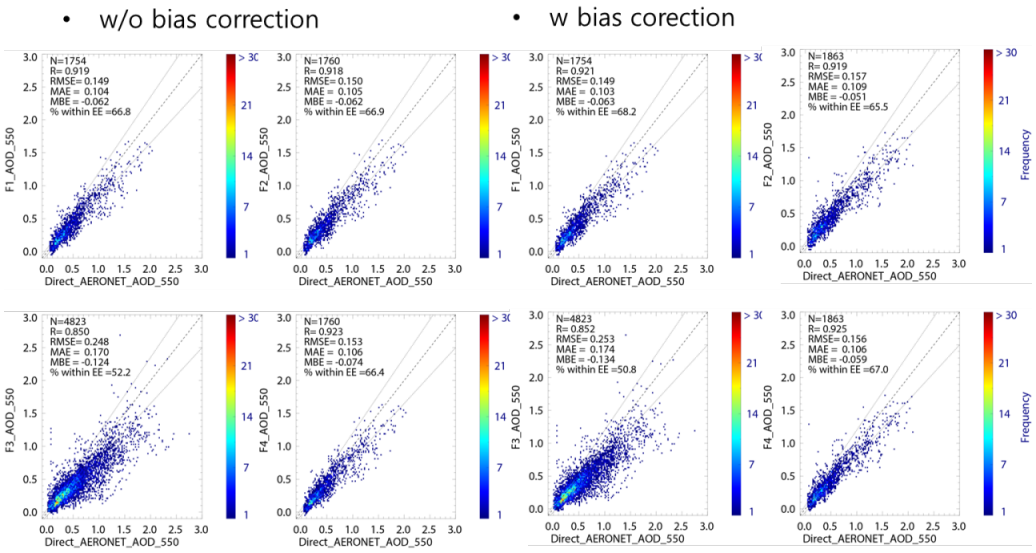

Figure 2. EMeRGe campaign.

- Looking at the effect of bias correction, F1 using all outputs shows an improvement results. Meanwhile, F4, which uses an ensemble member similar to F1, decreases in KORUS-AQ and increases %EE in the EMeRGe campaign. This may appear because the GV1's bias correction value is not accurate. Although mentioned in the text, the accurate correction may not be made using the RMSE and bias correction in this study for long-term analysis values. In general, if bias correction is performed and ensemble averaging is performed, MBE is improved in most cases, but the difference in EMeRGe F3 product is the greatest.

The panels in figures 2, 3, 6, and 8 are so small that it is very difficult to see the differences in any features. Also, the large-domain view is only relevant for 4 cases (AER, AMR, F3, FM3). I suggest splitting each of these figures into 2 figures. The first would be for the large domain and would only have 4 panels, so the panels could all be twice as large. The second would be for the small (GOCI) domain and although it would have the same number of panels, there would be much less wasted white space since you would be zoomed into the GOCI region. Even with these changes, it may be difficult for the reader to actually see the differences among the cases. Consider selecting one product to show a representative AOD distribution (perhaps the one you consider to be the "best") and plot all other cases as the difference from this reference; this would allow you to clearly highlight where the differences arise.

- Thank you for your suggestion. We revised the Figure 4, 5 and paragraphs with one representative average AOD (FM1), while the remaining products were modified to show

differences, mean (XX) – mean (FM1), for the same area as the GOCI's.

It is very hard to see the differences in Figure 4. I suggest the figure could be greatly improved by plotting AERONET and DAOD (the difference from AERONET) for all the other products. The left vertical axis could be AOD and the right vertical axis could be DAOD. I don't think you would need a log scale.

- Thank you for your suggestion. DAOD was added as the reviewer suggested. However, if the figure is not shown in log-scale, the variation of the low AOD does not appear well, so the symbol thickness has been modified to be thin.

Line 335: The statement "the fusion products have a value of 0.131, lower than the minimum value of various satellite products (0.161)" is not true for all cases; Fused3 is .163, essentially the same as the .161 value quoted as the minimum of the individual products.

- Sorry for the confusion. We corrected sentence based on the re-calculated results, and added the phrase excluding F3, and FM3.

Paragraph beginning on line 349: This discussion highlights what I see as a problem with the current analysis. The AHI-only fusion results could also be analyzed within the small (GOCI) domain. Then there would be direct comparisons of all the ensembles in the small domain and separate comparisons of the large-domain results. I noted that Figure 11 does do this, which is the best aspect of Fig 11, and find it strange that this isn't done more consistently throughout the analysis.

- Thanks for the suggestion. As mentioned above, analysis domain was set as the GOCI domain in Figure 4 and 5. And, we added to section 5.3 and table 2. This section and table 2 were shown two AHI AODs validation score within GOCI's observation area.

First paragraph of section 5.2: AHI_ESR is in all the ensembles, so it is not surprising that the ensemble results move toward the AHI_ESR behavior. It explains the KORUS-AQ results; AHI_ESR has a positive bias, the other 3 have negative or no bias, so the combinations will produce small biases. Similar behavior is seen in EMeRGe; the results tend to collapse toward the original

AHI_ESR values. Discussion of Fig 11, in particular paragraph beginning on line 394: It is really impossible to see that "results after fusion show slightly better than respective satellite product accuracy in terms of SD, RMSE, and EE values" since so many points are clustered essentially on top of each other.

All that can be clearly seen is that GV1, GV2, and the "all" points are distinct from the cluster of everything else. It appears that the best discriminator between points in the cluster is the %EE. Two suggestions could improve this figure. First, in the legend, color the symbols by the %EE for each case, then at least it will be easy to see that it improves for the ensembles. Second, consider adding an inset that zooms in on the cluster of points. This may be irrelevant though (and Fig 11 redundant), if the result is that there are only minimal differences between the various cases.

- Thank you for your suggestion. The legend is also shown as %EE, and only all site validation is shown in the Taylor diagram as suggested by the reviewer. The validation of broader products collocated with GOCI is summarized in Table 2.

Line 410: This could also be because EMeRGe was in a period of brighter surface reflectance.

- Thank you for your comments. We added this point per the reviewer's comment.

Summary and Conclusion: I think it is important to point out that the ensemble-mean and MLE techniques produce very similar results, based on the numbers in Figures 5 and 7.

- Thank you for your comment. We added this point per the reviewer's comment.

Sentence on lines 433-435: This appears to be marginally true for EMeRGe (fig 7) but is not true for KORUS-AQ (fig 5), where ensemble-mean actually appears to be best.

- Thank you for your comment. We revised this sentence.

TECHNICAL CORRECTIONS ————————————

The citations on lines 48-49 duplicate citations earlier in the paragraph.

- We removed duplicate citations.

Line 195: unclear meaning; instead of the word "for", do you mean "depending on"?

- *We revised this word.*

Line 197: Not sure what is meant by "the NDVI shows a negative bias". Isn't NDVI an independent variable? Do you mean, when AOD is analyzed as a function of NDVI, a low bias exists for all values of NDVI?

- *Sorry for the confusion. We revised this sentence.*

Throughout, be consistent between the convention used in the text, tables, and figures. The text mostly uses the "short" labels e.g. as defined in lines 231-232, while the figures use a "long" convention that is much easier for the reader to keep straight (e.g., AHI_MRM instead of AMR). As a reader, I preferred the longer conventions because of this ease of keeping things straight.

- *Thank you for your comments. We revised that the abbreviations of the text and pictures have been unified.*

Line 240: define what is meant by "wide area". It becomes clear when looking at the figures, however at this point in the paper it would be helpful to define it.

- *Thank you for your suggestion. We added domain information.*

Line 266: Typo, AESR should be AES

- *Thank you for your comment. We removed this sentence.*

Line 382: Start a new paragraph for the discussion of Figure 11.

- *Thank you for your suggestion. We revised.*

Throughout the manuscript, be consistent with terminology. Fusion is the generic term. Two fusion techniques are used, ensemble mean and MLE. E.g., the legend in Fig 11 should indicate ensemble-mean rather than fusion.

- *Thank you for your comment. We revised the all of the legends.*

Line 429-430: It seems the biases are not "due to" NDVI etc., but instead are represented as functions of NDVI, time of day, and AOD.

- Thank you for your comments. We revised this sentence.

Tables 1 and 3: The NDVI labels are identical for the 2nd and 3rd groupings. It seems one or both are typographical errors.

- Sorry for the confusion. We replaced the tables with figures.