Reviewer: 2

The authors take 9 satellite aerosol optical depth (AOD) monthly mean data sets, and perform comparisons against each other and AERONET monthly mean data. These come from a variety of satellite instruments and algorithms. They look at similarities in spatial and temporal patterns. This research area is important because understanding aerosol influences on the Earth requires understanding the strengths and limitations of each data set.

This is a pretty big task and it is good to see it being tackled, because as the authors note there has not been a great deal of attention to data set choice in some satellite analyses. However, I think this version of the paper has problems. The statistics and analysis are very superficial, and the metrics used do not always make sense or are incorrect. For example, autocorrelation and false discovery rate are ignored, a level 2 error metric is used for level 3 analysis. The terminology has errors in some sections (e.g. “validation” when this is not a validation analysis). And in several places the authors omit relevant references and use out of date ones, or instead insert excessive self-citations. There is also a possible wavelength issue with the AVHRR product used.

I recommend major revisions and would like to review the revised version. This paper felt to me like the authors just downloaded a bunch of data and ran a bunch of statistical metrics against it, without thinking about what was being done or why. I suggest that when revising, they focus on what science question they are trying to answer, and then figure out the right tools to answer it and provide a detailed discussion. Otherwise this feels not like a scientific research paper but rather the output of some automated data processing software.

After writing this review, I read the other two comments currently posted on ACPD for this paper. I generally agree with the other reviewers’ comments.

Response: We appreciate the time and effort the reviewer spent on this manuscript, as well as their insightful and constructive suggestions. In light of your opinion, we have carefully revised our manuscript. The responses to the questions raised in your report are as follows.

My comments in support of my recommendation are as follows:

1. Line 20, and elsewhere: Operational is not the right word here. It implies something produces as part of routine agency operations while a mission is ongoing. Most of the products do not fit that definition; in fact, I think only MODIS and AVHRR do as they are produced with a few hours latency to support assimilation applications. I suggest deleting this word throughout.

Response: Thank you for your suggestion. We have deleted this word throughout the paper.

2. Title, line 30, line 35 and elsewhere: Terms like "significant inconsistencies" (or just "inconsistencies" alone), "seriously" are used a lot in this paper. But most of the time they are used as "weasel words", i.e. in a non-specific way which can lead people to get a certain impression which is not necessarily warranted. For example, "inconsistencies". Taken to an extreme, any two data sets will not be identical so are to some extent "inconsistent". The relevant question is, for any particular application, is the level of consistency between them sufficient? For example, if one wants to look at seasonal variations, AOD magnitude might not be as important as the pattern throughout the year. But if one wants to look at radiative effects, magnitude is more important. If one wants to see large-scale features, then a broader swath to improve sampling at the expense of some accuracy might be desirable. The point is that these are all different instruments with different characteristics. We expect them to not be identical. The wording in this paper (these examples and elsewhere) seems designed to send a message that aerosol remote sensing has big problems. In my opinion, that's an overly pessimistic assessment. There are differences but in general the reasons for those are understood. So which data set is best to use for a given study depends on the type of science question you are trying to answer. There is no "best" data set. This recent paper by Sayer et al in JGR (https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2018JD029465) covers some similar ground to the current study, in that part of it compares time series and maps of various over-water satellite AOD data sets. That paper goes into a lot of discussion about the differences between them and why they might be. So although there is a lot of diversity in the over-water AOD, the reasons are generally known. My personal opinion is that over much of the world, differences are probably more due to sampling differences (swath and pixel selection) than algorithm. I suggest the authors refer to that paper in their revised manuscript and go from describing things as "inconsistent" to try to for example make

recommendations as to which data sets might be better or worse suited for different applications. Recommendations like that, with evidence, are more useful than just declaring "inconsistency". Perhaps "comparisons" or "consistency assessment" is a better way to describe the analysis in title and text.

Response: Thank you for the constructive suggestion. We completely agree, and we carefully read the paper you mentioned above (Sayer et al., 2018). In the current version, we focused on describing the performances of multi-source aerosol products through comparisons with ground-based observations at different scales to determine the best product in terms of representing the temporal and spatial AOD variations. Moreover, we provide recommendations to users for the selection of these products for different applications according to your suggestion. We have also changed the title to "Inter-comparison in spatial distributions and temporal trends derived from multi-source satellite aerosol products".

3. Line 50: I think this should say 20th century, not 19th. I am not aware of any observation networks before the late 20th century. If there are, please provide references. Aerosol science didn't really start until John Aitken in the late 1800s.

Response: This term has been corrected.

4. Line 118: This should be "Holzer-Popp" not "Holzerpopp". The author's name is double barrelled.

Response: This name was corrected.

Line 121: This should be changed to indicate it is the NOAA AVHRR aerosol product. There is also a NASA GISS aerosol product (GACP), which is monthly-only and ocean-only, and a NASA Deep Blue aerosol product, which also covers land but is presently only available for limited time periods (I know 2006-2011 is available). It would be good to clarify what is used and why here. Deep Blue and GISS also provide 550 nm while NOAA AVHRR do not. Perhaps one of those could be added.

Response: Thank you for your suggestion. Due to mismatched wavelengths (630 nm) and missing land observations, we have abandoned the use of the NOAA AVHRR AOD product and replaced it with the newly updated NASA AVHRR AOD product (available from 2006 to 2011) in the revision.

5. Line 125: Authors should state more clearly here that they are using the 0.63-micron AOD (aot1 SDS), as it is important to note that this is different from the 550 nm AOD provided by most other data sets and would result in offsets dependent on aerosol type. The authors do not seem to mention this later in the paper (e.g. line 179 says the satellites are at 550 nm). Was the AVHRR AOD somehow extrapolated to 550 nm like the others? Or was it left at 630 nm and the wavelength dependence neglected?

Response: Per your previous suggestions, we have abandoned the use of the NOAA AVHRR AOD product due to mismatched wavelengths (630 nm) and missing land observations and instead used the newly updated NASA AVHRR AOD product (550 nm) in this revision.

6. Line 139: Authors are missing references for the version 23 algorithm they are using here. Martonchik/Kalashnikova are out of date. The water approach is discussed by Witek et al (2018): https://www.atmos-meas-tech.net/11/429/2018/ The land approach is discussed by Garay et al (2017): https://www.atmos-chem-phys.net/17/5095/2017/ I suggest authors read and cite these papers, since it appears they have been referring to older documents.

Response: Thank you for pointing out the out-of-date reference. We have carefully read these papers and cited them instead of the older documents in the paper according to your suggestions.

7. Line 155: Sayer et al (2014): https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2014JD022453 is a more complete reference for the DTB products than Levy et al (2013). It also provides a comparison for DB, DT, and DTB. It will also be useful for the authors' analysis since it provides similar discussion about the level of consistency between the data sets. All the papers cited here are about Collection 6 but I know the MODIS teams and they did not publish papers about Collection 6.1 yet (still in review).

Response: We have cited this more complete reference for the DTB products as well as a recently published paper from our team (Wei et al., 2019, AE) for the description of the Collection 6.1 aerosol products in the paper per your suggestion.

8. Line 163: I am not sure that the FM acronym for "forward model" is needed here. I don't think it is used later.
Response: We have removed this acronym from the paper.

9. Line 166: Somewhere in this section I would add a note to state that this is not a validation but a comparison, because the authors are using monthly data and not instantaneous data. So there are sampling differences contribution as well as retrieval quality. The authors are not performing a true validation exercise here.
Response: Thank you for this point. We have added these descriptions to the paper (Section 2.2) and replaced the term "validation" with "comparison" following the suggestions from two reviewers.

10. Line 189: The authors insert four self-citations for a one-line equation developed by other people something like 75 years ago. This seems a little excessive. Please remove these citations or replace with ones to the original work by Angstrom.
Response: We have removed these citations from the paper.

11. Line 190: I recommend the authors account for lag 1-month autocorrelation in the time series. This is commonly done in AOD trend analyses as the data can be significantly autocorrelated on these scales (because large-scale systems and seasonal patterns can persist for weeks to months). This will keep the same trend values but affects the estimated uncertainties on the trend. See Weatherhead et al (1998): https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/98JD00995 for examples how to calculate this.
Response: Thank you for the recommendation. We have accounted for the 1-month lag autocorrelation for all AOD time series analyses in the revision (Section 3.2 and 6.1) according to your suggestion.

12. Lines 197-200: I am not sure that "correct trend percentage" makes sense. If a trend is close to 0, you will end up with a lot of apparently "wrong" trends if the sign is wrong, even if the conclusion that there is almost no trend is correct. For example, if you had trends of +0.01 from AERONET and +0.1 from satellite the authors would say this is "correct" even though the difference is huge. But if you had

0.005 from AERONET and -0.005 from satellite the authors would classify it as “incorrect”, even though they are both small and probably statistically indistinguishable within trend uncertainties. A further problem is that this makes the implicit assumption that AERONET trends are perfect when of course they also have some measurement uncertainty and sampling uncertainty. I suggest that a better metric would be to report the “consistent trend percentage”. This could be calculated by checking whether the satellite and AERONET trends are consistent within each uncertainty or not. This is a more fair and statistically appropriate test. The authors could also report those situations in which the AERONET estimate is too uncertain to be useful. I doubt that five years is enough to estimate a trend robustly in many cases, due to significant annual variability. So quite possibly the uncertainty on the AERONET estimates even is quite high. I also wonder if seasonal trends would be better than annual, because we know that aerosol patterns show strong seasonal features (so trends in seasonal behavior could be masked in an annual trend analysis). The authors need to justify this more strongly.

Response: Thank you very much for your suggestions. We have modified and used the improved metric you mentioned to report the “consistent trend percentage” by checking whether the satellite and AERONET trends were consistent within each level of uncertainty (Section 3.3). We apologize for the misleading statements in the original version of the manuscript and have clarified this information in the revised version. All the trend analyses are based on the de-seasonalized time series of monthly AOD anomalies because the sample points at the annual and seasonal levels are not large enough to analyse the trends. Moreover, we have extended the study period from five years to eight years with approximately 96 monthly values, which is sufficient for long-term trend analysis according to previous studies. We have stressed on this in Section 3.2 in the revision.

13. Line 209: The subscripts are very long. I suggest replacing AOD_RETRIEVAL with AOD_R (for “retrieval”) or AOD_S (for “satellite”), and AOD_AERONET with AOD_A. This will make it more readable.

Response: These terms were changed per your suggestion.

14. Line 210: Correlation is not useful when the data range is small compared to the uncertainty on the data. You could have a great data set but still have a small correlation. For example, over the open ocean AOD does not change much, so a low correlation is scientifically not much of a problem for most scientific applications, as long as bias and RMSE are low. The authors should note this because a lot of the maps and discussion rely on correlation.

Response: Thank you for your suggestion. We agree with your opinion about the correlation, and we have removed most of the discussion related to the correlation from the revised version.

15. Line 211: This EE is an expected envelope for level 2 error over land only, not for level 3 and not for water. It is not meaningful for level 3 data, and it is misleading to apply it that way. There is at present no error estimate for satellite level 3 products. I suggest the authors remove this quantity because it is misleading. In my view the other statistics are enough. This also requires removing from the discussion later on. Either remove it or create and justify some metric for what an acceptable EE on the monthly data is. My feeling is that a monthly level 3 EE should be smaller than the level 2 one, because some error sources should cancel out.

Response: Thank you for pointing out this problem, and we completely agree with your opinion. We have removed the EE quantity throughout the paper according to your suggestions.

16. Line 219, 230, 468, Table 2, Figure 10, and elsewhere: No, this is not a validation, it is a comparison, because you are using monthly mean products and not level 2. Validation requires a ground truth. There is no ground truth for monthly data because there is no instrument sampling continuous monthly data. AERONET is only a validation for level 2 data. The authors should change the wording because it is misleading, and word choice matters. The analysis the authors are doing here is fundamentally different from the dozens of published level 2 validation papers, and it is important not to muddle the issue.

Response: We have replaced the term “validation” with “comparison” throughout the paper according to your suggestion.

17. Line 295: Again, it is not ideal to provide a single self-citation here when these issues have been documented by many algorithm teams for many years.

Response: We have removed this citation from the paper.

18. Line 362: Throughout section 6 the authors talk a lot about trend significance. However, something which has been overlooked is that since there is multiple hypothesis testing going on (many data sets and locations are being tested for trends), there could be a significant fraction of false positives. See e.g. Wilks (2006) for more on this: https://journals.ametsoc.org/doi/10.1175/JAM2404.1 So, the authors should make some quantification about the expected false discovery rate. Further, statistical significance is only one factor. Figures 11 and 14 are a prime example of this problem. Scientific significance is another. If you get a trend of 0.001 with an uncertainty of 0.0001, that is statistically significant but scientifically not important because it is so small. But if you get a trend of 0.1 with an uncertainty of 0.1, that is not statistically significant by traditional tests, but is potentially very important, because 0.1 is a large potential trend. The authors here seem to focus on statistical significance and sign rather than actually looking at the numbers. This is quite superficial. I would like to see the whole section reconsidered.

Response: Thank you for your suggestion. We agree with your opinion, and we have added the false discovery rate (FDR) test to exclude the fraction of false positives in our trend significance analysis (Section 3.2). Moreover, we have shifted our focus from statistical significance to actual significance, and we have mainly explained the possible reasons for the regions where the aerosols changed significantly in Section 6.

19. Line 496: “Goddard”, not “Godard”.

Response: This name was corrected.

20. Figure 7 (and associated discussion): I do not like annual mean maps in general because AOD patterns and sampling are strongly dependent on season. So in some areas there will be a difference just because data are coming from different months. And in some areas things could look to be in closer

agreement than they really are, if biases in different seasons are opposite and can cancel out. Annual mean AOD is also not meaningful for most applications. I would prefer to see this figure and discussion instead as a composite of four sets of seasonal plots. This would be a closer to apples to apples comparison, and also allow an examination of seasonal variability.

Response: We have replaced the annual mean AOD maps with the four sets of seasonal AOD maps in the figures and provided associated discussions in this revision according to your suggestions.

21. Figure 8, 12: Could this be redrawn to show coloured symbols instead of bars? In some cases, the bars are overlapping and so it is hard to tell which is. It can also give misleading impressions. For example, in land ENAM the black and pink are overlapped. I guess black was drawn first and pink second, so pink is on top. So, the impression is that black is lower than pink, because we can only see the bottom of black. But in reality, because so much of black is hidden, it probably means that black and pink are very similar. Coloured symbols instead of bars would be clearer and easier to tell.

Response: We have redrawn these figures with coloured symbols according to your suggestions in the revision.

22. Figure 9: since this is not a validation but a comparison, it would be better to say “offset” rather than “bias” here. Bias implies an offset with reference to a truth, and we have no truth. Word choice is important.

Response: This term was corrected per your suggestion.

Reviewer: 3

The authors provide a comparison of nine satellite-derived global AOD data sets, with ground-based AERONET (land) and MAN (ocean) AOD data as reference. They apply different statistical metrics and look at the data sets on different spatial scales: global, regional and per reference site. They also look at trends. Differences and agreements between data sets are described. The manuscript provides an interesting overview of AOD data sets available in the public domain, although some recent data sets like those from VIIRS are missing. Also, I wonder why for AVHRR only the over-ocean AOD is included and the recent over-land data sets described by Sayer and Hsu in JGR, 2017, were not included. It would be interesting to see how these data sets, retrieved from a sensor not designed for aerosol retrieval, compares to those from dedicated sensors like MISR and MODIS. Likewise, a comparison with PARASOL (POLDER) would have been interesting. As regards the title, I would recommend changing "inconsistency" to "Intercomparison", because not all and not always are the data sets inconsistent, they are often also consistent.

Response: We appreciate the time and effort the reviewer spent on this manuscript and the insightful comments and constructive suggestions. In light of your opinion, we have carefully revised our manuscript. The responses to the questions raised in your report are as follows.

Regarding the AOD product selection, we actually collected the VIIRS aerosol product. However, VIIRS was launched in 2012, and its common matching period with most other satellite products was short. In this revision, we have added the VIIRS monthly product for a simple comparison of the spatial coverage and distributions in Figure 1. Because of the similar sensor parameters and algorithms of VIIRS and Aqua MODIS, both data products have close monthly spatial coverages and mean AODs values throughout the time series. Therefore, we did not include the VIIRS products in the inter-comparison in the following analysis. In addition, we have added the AVHRR and POLDER products over both land and ocean in the revision according to your suggestions. Meanwhile, we have modified the title to "Inter-comparison in spatial distributions and temporal trends derived from multi-source satellite aerosol products".

Specific comments (line numbers refer to the pdf published online)

1. 46: suggest "composition and short life time of atmospheric aerosol particles"

Response: This phrase was modified per your suggestion.

2. 57: remove "observable"

Response: This term was removed per your suggestion.

3. 79: remove "seemingly"

Response: This term was removed per your suggestion.

4. 80: This sentence suggests that some studies have indeed focused on exploring : : : ; hence references to these studies are needed here

Response: Thank you. We have cited the main references in the revision according to your suggestion.

5. 85: suggest "evaluation and comparison"

Response: This phrase was changed per your suggestion.

6. 91: validation

Response: This term was corrected.

7. 103: ADV was first published by Veefkind et al., 1998a, for retrieval over land; Over Ocean ASV was first developed by Veefkind et al., 1998b

Response: We have modified and cited these references in the paper.

8. 112: A more recent reference for the Swansea algorithm is Bevan et al., 2012

Response: The relevant reference has now been cited in the paper.

9. 117: Holzer-Popp

Response: This name has been corrected.

10. 122: "AVHRR aerosol product is only available": this is NOT true, see my general comment and references to Sayer and Hsu Sect.

Response: Yes, according to the suggestions by two reviewers, we have replaced the ocean-only AVHRR data with the NASA AVHRR aerosol product as you mentioned (Sayer and Hsu), which is available over both land and ocean. We have also rephrased this paragraph in the paper.

11. 2.2: not only AERONET is used, AOD over ocean is provided by the Marine Aerosol Network (Smirnov et al., 2009)

Response: We have cited this reference and revised the statement in the paper.

12. 196: could you reword the text to make clearer how the lsq fit is applied

Response: We have rephrased this sentence and cited the main reference on the LSQ method in the paper.

13. 199: trend symbols: same direction of the trend Para starting at

Response: This has been corrected.

14. 230: An important indicator is also the EE, and the above and under EE which clearly indicate overestimation (e.g. for MODIS) and underestimation (e.g. for MISR). Here and in the next paragraphs, I do not understand how MISR can have a similar number of collocations as MODIS in spite of its much smaller swath; MISR should have an N similar to AATSR

Response: We apologize for the incorrect statistics, and we have corrected the sentence to "The Terra MISR product provides a sample size of 8418, which is smaller than the Terra MODIS sample size (N = 9196) and is possibly due to the narrower swath width" in the revised version.

15. 235: I am not sure that your judgement of ADV is completely fair, since indeed MAE and RMSE are worse, but not EE; looking at the statistics in Table 2, it seems that none of the sensors has the best statistics for all numbers, so it is hard to make such statements.

Response: We apologize for the improper description here, and we have removed the statement from the revision.

16. 236: smaller number of retrievals collected: I think this should be a smaller number of collocation pairs since less references data are available; again, how can MISR provide a similar number of data collections as MODIS?

Response: Yes, the small number of data is due to the limited availability of reference data. We have corrected the description in the revision.

17. 244: SeaWiFS is not improved, but its performance is better

Response: This information was revised per your suggestion.

18. 251: what is the statistical parameter indicating estimation uncertainty and accuracy?

Response: In this paper, the accuracy is represented by MAE, and the uncertainty is represented by RMSE and RMB, where RMB > 1.0 or RMB < 1.0 indicate the over- or under-estimation uncertainty. We have clarified this information in Section 3.3 in the revision.

19. 257 and 263 and 275-277: a high R does not imply that the performance is better: MODIS has high R, but figure 2 shows that MODIS overestimates, so actually it's performance in estimating AOD is not so good. This should be re-worded in the text.

Response: We agree with your opinion and revised this information in the results analysis. We mainly use the following indicators (i.e., MAE, RMSE and RMB) to describe the product performance and no longer use the correlation (R) according to the suggestions from two reviewers.

20. 266: RSA, typo and you mean ESA?

Response: This term has been corrected.

21. Sect. 5.2: there are very large differences in the mean AOD values; yet they all compare well with AERONET (Fig. 2 and 3): why are these differences not visible in the scatterplots?

Response: These figures have been corrected.

22. 304: suggest plotting the eight-year mean value in the figures

Response: This change has been implemented per your suggestion.

23. Sect. 5.3 title not clear: suggest changing the Section title to " Comparison of satellite- and AERONET- derived annual mean AOD at each site

Response: The title has been modified per your suggestion.

24. 340: this sentence is not accurate: you compare annual mean AOD for each satellite over an AERONET sites with the AERONET annual mean value

Response: We have revised the sentence as follows: "Furthermore, we also compare the annual mean AODs calculated from each satellite product and AERONET throughout the world from 2003 to 2010 (Figure 9)."

25. 375-376: I do not understand the sentence "Four : : : areas." Why are the first 4 similar and the other 2 consistent? What do you mean with that? MYD08 and SeaWiFs show quite some differences. Could you re-word so it is clearer?

Response: We apologize for the unclear description, and we have rephrased the descriptions to "On the other hand, the four ESA-CCI and MISR aerosol products are not significant in most ocean areas, even for the open seas. MODIS and SeaWiFS products have similar spatial patterns in most ocean areas, such as the significantly increasing trends observed over the Pacific and Indian Oceans" in the revised version.

26. 379: what do you mean with "treatment in neighboring pixels": did you describe that in the text?

Response: We apologize for the incorrect description, and we have removed this information from the revised version.

27. Sect. 6.4: Linear trends were fitted, so it may be that upward and downward trends are compensated over this long period of 18 years and thus the trends in Fig 14 are not representative. Could you please add a comment in the text?

Response: We have added this comment in the revised version according to your suggestion.

28. Figure Captions: 2 and 3: Density scatterplot of the monthly averages of satellite-derived AOD (operational products) versus AERONET AOD

Response: These captions have been corrected.

8: replace aerosols with AOD

Response: This term has been corrected.

10: ".. with annual mean AERONET AOD data for all sites : : :"

Response: This phrase has been corrected.

11: trends of AOD at 550 nm

Response: This phrase has been corrected.

12: replace "aerosol trends" with "trends of AOD at 550 nm"

Response: This phrase has been corrected.

13: I think you show trends of AOD at 550 nm, not annual mean aerosols?

Response: This phrase has been corrected.

15: remove “variations”

Response: This term has been removed.

References:

Bevan, S. L., North, P. R. J., Los, S. O., & Grey, W. M. F. (2012). A global dataset of atmospheric aerosol optical depth and surface reflectance from AATSR. Remote Sensing of Environment, 116, 119–210.

Hsu, N. C., Lee, J., Sayer, A. M., Carletta, N., Chen, S. H., Tucker, C. J., : : : Tsay, S. C. : Retrieving near-global aerosol loading over land and ocean from AVHRR. Journal of Geophysical Research: Atmospheres, 122. doi:10.1002/2017JD026932, 2017

Sayer, A. M., N. C. Hsu, J. Lee, N. Carletta, S.âAˇ RH. Chen, and A. Smirnov: Evaluation of NASA Deep Blue/SOAR aerosol retrieval algorithms applied to AVHRR measurements, J. Geophys. Res. Atmos., 122, doi:10.1002/2017JD026934, 2017.

Smirnov, A., et al. (2009), Maritime Aerosol Network as a component of Aerosol Robotic Network, J. Geophys. Res., 114, D06204, doi:10.1029/2008JD011257.

Veefkind, J.P., G. de Leeuw and P.A. Durkee (1998a). Retrieval of aerosol optical depth over land using two-angle view satellite radiometry during TARFOX. Geophys. Res. Letters. 25(16), 3135-3138.

Veefkind, J.P. and de Leeuw, G., 1998b, A new algorithm to determine the spectral aerosol optical depth from satellite radiometer measurements. Journal of Aerosol Sciences, 29, 1237-1248.

Reviewer: 4

The objective of the present study is the intercomparison of various spaceborne retrievals which are widely utilized in aerosol studies. The analysis has been performed at different spatial scales and for a long-term period thus increasing the robustness of the obtained findings. Nevertheless, the major weakness is that the interpretation of the results is poor without providing insight and sufficient answers about the potential reasons which can explain the apparent differences. More specifically, throughout the manuscript the authors are restricted just to a description of the figures which can be easily done by a reader without reading the text. Therefore, I strongly believe that the manuscript needs a major revision before it can be acceptable for publication in ACP. Below are listed my comments/questions which I hope will help the authors to improve their work.

Response: We appreciate the time and effort the reviewer spent on this manuscript and the insightful comments and constructive suggestions. In light of your opinion, we have carefully revised our manuscript. The responses to the questions raised in your report are as follows.

1. Which version of the AERONET data is utilized?

Response: We use the newly released AERONET Version 3 Level 2.0 monthly AOD observations in this study, and we have clarified this information in Section 2.2.

2. You have to provide a better description of the satellite datasets (version, spatial resolution, temporal resolution, temporal availability, where these data are stored, literature etc.).

Response: We have provided more detailed descriptions (including the data version, spatial and temporal resolution, temporal availability, scientific dataset, and literature) of the satellite-derived aerosol products in the revised version according to your suggestion. Meanwhile, the data acquisition addresses are provided in the acknowledgements.

3. Page 6 – Lines 177-179: This sentence is confusing for me. Are you using monthly means or daily retrievals which are used in order to calculate the monthly averages? What do you mean "…with

sufficiently high-quality…"?Are you applying any quality assurance flag or are you using the raw data as is?

Response: We apologize for the confusing sentence. In the paper, we did not apply any additional quality assurance and used the original monthly products for all analyses. The mentioned quality assurance flag is only a type of output control in the MODIS aerosol retrieval algorithm (Levy et al., 2013). We have removed this sentence from the revised version.

4. Page 6 – Lines 180-181: Please rephrase this sentence.

Response: We have rephrased the sentence "For multi-satellite aerosol products, the monthly retrievals at 550 nm are collected from the listed scientific dataset (SDS, Table 1) and used for the current analysis in this study" in the revision.

5. Figure 2: I cannot understand why the comparison versus AERONET is made for the periods where each dataset is available and not for the common period (Table 2 and 3). In the scatterplots, the EE dashed lines are common for all satellite data. This is not correct since each satellite sensor has different uncertainty limits (which are not stated in the text).

Response: We have removed the comparison for the period of each dataset and retained the common-period comparisons in the revision according to your suggestion. The problem regarding the EE dashed lines is explained below in the answer to question 7.

6. Page 8 – Lines 228-244: Is there any interpretation for these results? The authors must consider previous evaluation analyses in their discussion.

Response: We have compared our results with the results of previous studies on the four ESA-CCI products in the paper (Section 4.1). However, for the remaining aerosol products, we used the newest versions that have been released recently (e.g., MODIS C6.1 and AVHRR products available in October 2017; MISR V23 in November 2017; VIIRS V1 in February 2018). Meanwhile, most published studies focus on the validation of the instantaneous retrievals of Level 2 products against surface measurements.

Comparative studies on Level 3 monthly products are rare, and we did not find similar evaluation papers; thus, we did not make such comparisons in the current study.

7. Section 4.2: You have to repeat the analysis for EE using the corresponding limits for each satellite sensor. Moreover, you have to compare your results with other existing works.

Response: We have removed the EE quantity throughout the analysis due to its limitations for different satellite monthly aerosol products according to the suggestions from two reviewers.

8. Section 5.1: There are several points which must be discussed in Figure 7. For example, the differences among AATSR-ORAC, AATSR-SU, MODIS and SeaWIFS recorded across N. Africa. Likewise, in E. Asia, it seems that there is a strong diversity, in terms of AOD values, among the datasets. In AATSR-ORAC, there is an abrupt change of AODs between maritime and continental areas in the eastern tropical Atlantic Ocean as well as in the Arabian Sea. Finally, it would be useful to reproduce the maps by considering common points in all datasets separately over land (exclude AATSR in order to have available observations over Sahara and in the Middle East) and sea.

Response: We have added a discussion on this issue as "There is also strong diversity in the seasonal mean AODS over North Africa and East Asia among most datasets. This diversity is mainly due to the different aerosol algorithms applied over bright surfaces (i.e., desert and urban areas). Both high surface reflectance and complex underlying surfaces increase the difficulty of aerosol retrieval (Wei et al., 2018)" in the revision. There was a mistake when processing the AATSR-ORAC product, and we have corrected and fixed the problem you mentioned. Meanwhile, we have reproduced seasonal maps for land and ocean in the Supplement File (Figures S2-3) following your suggestion.

9. Figure 9: For the computation of the regional means based on the satellite observations are used all the grid cells of the domain of interest or only the pixels in which AERONET stations reside? Why there is an increasing trend for MODIS data in EAA as well as in EUR? On the contrary, in SAA the agreement between MODIS and AERONET improves gradually. Why this is happening?

Response: In the original manuscript, we used only the pixels located over each AERONET station. According to the comment from another reviewer, this is not a validation but a comparison because we use the annual averages, not the instantaneous values, which may be the main reason for these uninterpretable trends. The analysis makes little sense; thus, we have deleted this part in the revision.

10. Section 6.4: Are your results in agreement with other similar studies? In the global map, there are clear signals over wide areas of the planet which are not discussed appropriately in the text. Which factors regulate (meteorology, emissions, teleconnections, land use, etc.) the obtained pattern?

Response: Thank you for your suggestion. We have compared our results with the results of other studies and discussed the main factors regulating the present AOD spatial patterns in the revised version (Section 6.3).

11. Figure 1: First of all, there are mistakes on the region names. Please correct the European Coast as well as the South Africa (it is not in Asia!). Which is the domain for the European Coast? Replace Atlantic Ocean with South Atlantic Ocean.

Response: We apologize for these mistakes, and we have corrected them according to your suggestions. The European coast mainly includes the Eastern European Sea and Mediterranean Sea. To make the border clearer, we have replotted Figure 1 in the revision.

12. Figure 11: Replace 2017 with 2010.

Response: This information has been corrected.

13. Page 3 – Lines 64-77: In this part of the manuscript the authors are stating only studies representative for China. Satellite observations have been also used for other regions of the planet such as the Mediterranean, Europe, Atlantic Ocean etc.

Response: Thank you for your suggestion. We have enriched the introduction and added satellite-based AOD research over Europe, the Mediterranean Sea, Northern Africa, Topical Pacific, North and South Atlantic Oceans in the revised version.

Inter-comparison in spatial distributions and temporal trends derived from multi-source satellite aerosol products

Jing Wei[1], Yiran Peng[1*], Rashed Mahmood[2], Lin Sun[3], Jianping Guo[4]

[1] Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing, China
[2] Department of Atmospheric Science, School of Environmental Studies, China University of Geosciences, Wuhan Hubei, China
[3] College of Geomatics, Shandong University of Science and Technology, Qingdao Shandong, China
[4] State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing, China

Correspondence to: Yiran Peng (pyiran@mail.tsinghua.edu.cn)

Abstract

Satellite-derived aerosol products provide long-term and large-scale observations for analysing aerosol distributions and variations, climate-scale aerosol simulations, and aerosol-climate interactions. Therefore, a better understanding of the consistencies and differences among multiple aerosol products is important. The objective of this study is to compare ten global monthly aerosol optical depth (AOD) products, including the European Space Agency Climate Change Initiative (ESA-CCI) Advanced Along-Track Scanning Radiometer (AATSR), Advanced Very High Resolution Radiometer (AVHRR), Multi-angle Imaging Spectro Radiometer (MISR), Moderate Resolution Imaging Spectroradiometer (MODIS), Sea-Viewing Wide Field-of-View Sensor (SeaWiFS), and Polarization and Directionality of the Earth's Reflectance (POLDER) products. Aerosol Robotic Network (AERONET) Version 3 Level 2.0 monthly measurements at 308 sites around the world are selected for comparison. Our results illustrate that the spatial distributions and temporal variations of most aerosol products are highly consistent but exhibit certain differences in spatial coverage and regional performance. The SeaWiFS and AATSR-Dual View (ADV) products show the lowest spatial continuity with numerous missing values, while the MODIS products can cover most areas of the world. The highest spatial coverage and aerosol concentrations are found in June-July-August (JJA), and the lowest are found in December-January-February (DJF). The best performance is always observed in September-October-November (SON), but the worst is observed in JJA with large estimation uncertainties. Due to the influence from surface brightness and human activities, all the products perform unsatisfactorily over the Middle East, Africa, South Asia, and East Asia, and their coastal areas. In general, the Aqua MODIS product shows the best agreement with the AERONET-based AOD values at different spatial scales among all the products. All products can accurately capture the aerosol trends, especially in areas where aerosols change significantly. The MODIS products perform best in capturing the global temporal variations in aerosols. However, the aerosol trends are robust in only the Middle East, South

Asia, Europe and eastern North America and their coastal areas. These results provide a reference for users to select appropriate aerosol products for their particular studies.

1 Introduction

Atmospheric aerosols originating from both natural and anthropogenic sources have noticeable effects on the ecological environment, climate change, urban air quality, and human health; these issues also attract increasing attention from national governments and scientists (Cao et al., 2012; Guo et al., 2016, 2017; Li et al., 2011; Li et al., 2017; Pöschl, 2005). On the one hand, the increase in anthropogenic aerosols over the past century has significantly affected the radiation budget balance by scattering or absorbing solar radiation and by changing cloud microphysical properties (Ramanathan et al., 2001; Rosenfeld et al., 2008). On the other hand, fine particulate matter greatly endangers human health by causing various respiratory and cardiovascular diseases (Brauer et al., 2012; Bartell et al., 2013; Crouse et al., 2012). However, due to the complex sources, compositions and short lifetimes of atmospheric aerosol particles, large uncertainties exist in the estimation of aerosol-climate forcing and health effects. To better understand the spatial and temporal variability of aerosol distributions from regional to global scales, long-term data records with reasonable accuracy are needed as benchmarks to evaluate aerosol effects based on climate model simulations.

Since the 20th century, several aerosol ground-based observation networks, such as the worldwide Aerosol Robotic Network (AERONET), Interagency Monitoring of Protected Visual Environments (IMPROVE), European Monitoring and Evaluation Programme (EMEP), and Chinese Sun Hazemeter Network (CSHNET), have been established. The monitoring stations are sparsely distributed, and the observation periods at different sites vary across a large range due to instrumental or weather conditions. Therefore, ground-based observational data are limited to representing aerosol characteristics in long-term and large-scale studies. Since the 1990s, the continuous launch of satellite sensors has enabled the satellite remote sensing of aerosol measurements, which provides long-term data records with wide spatial coverage. Meanwhile, an abundance of mature aerosol retrieval algorithms has been developed according to the characteristics of different satellite sensors and atmospheric radiative transfer models, and these algorithms have been successfully applied to generate global-coverage aerosol products for over ten years. These satellite instruments include the Advanced Very High Resolution Radiometer (AVHRR), Total Ozone Mapping Spectrometer (TOMS), Advanced Along-Track Scanning Radiometer (AATSR), Multi-angle Imaging Spectro Radiometer (MISR), Moderate Resolution Imaging Spectroradiometer (MODIS), Sea-Viewing Wide Field-of-View Sensor (SeaWiFS), Polarization and Directionality of Earth's Reflectance (POLDER) and Cloud-Aerosol Lidar with Orthogonal Polarization (CALIPSO).

Based on these long-term space-borne aerosol products, numerous researchers have begun to explore the spatial and temporal variations in aerosols at regional and global scales as well as the potential climate effects of aerosols. For example, Guo et al. (2011) analysed the temporal and spatial distributions and trends in aerosol optical depth (AOD) over eight typical regions in China by combining TOMS (1980–2001) and Terra MODIS (2000–2008, Collection 5.1, C5.1) aerosol products. Hsu et al. (2012) explored the global and regional AOD trends over land and the oceans from 1997 to 2010 based on the SeaWiFS monthly aerosol products. Nabat et al. (2013) used different satellite-derived monthly AOD products (e.g., MODIS, MISR, and SeaWiFS) and model datasets to create a 4-D climatology of the monthly tropospheric AOD distribution and analyse the variations from 1979 to 2009 over Europe, the Mediterranean Sea, and Northern Africa. Zhao et al. (2013) analysed the AVHRR AOD datasets over the global oceans and explored the effects of subpixel cloud contamination on aerosol retrievals from 1981 to 2009. Floutsi et al. (2016) examined the spatiotemporal variations in the AOD, fine particle fraction and Ångström exponent over the Mediterranean Basin from 2002–2014 with the Aqua MODIS C6 aerosol products. Klingmüller et al. (2016) studied the aerosol trends over the Middle East and explored the effects of rainfall, soil moisture and surface winds on aerosols with Terra MODIS C6 aerosol products from 2000 to 2015. Mehta et al. (2016) presented the spatiotemporal AOD variations and their spatial correlations globally and over six subregions using the Terra MODIS (C5.1) and MISR monthly products from 2001 to 2014. Sayer et al. (2018) extracted and compared the AOD distributions and variations using multi-satellite monthly aerosol products (e.g., Visible-Infrared Imager-Radiometer Suite (VIIRS), Aqua MODIS, and MISR) over the main oceans (e.g., Topical Pacific and North and South Atlantic Oceans). Sogacheva et al. (2018) discussed the spatial and seasonal variations in aerosols over China based on two decades of multi-satellite observations using AATSR (1995–2012) and Terra MODIS (2000–2017, C6.1) aerosol products.

In most of the above studies, satellite-derived aerosol products are arbitrarily selected for research applications by simply following the usage in previous studies or based on data availability. However, noticeable inconsistencies exist among the aerosol datasets generated from different satellite sensors and aerosol retrieval algorithms. Few studies have focused on exploring the similarities and differences among aerosol datasets (Holzer-Popp et al., 2013; Naba et al., 2013; De Leeuw et al., 2015; Sayer et al., 2018). The selection of an accurate and appropriate aerosol product to represent the long-term aerosol variations and trends for their respective studies is of great importance for users, especially interdisciplinary scholars. Otherwise, problematic aerosol characteristics will inevitably lead to questionable conclusions.

The objective of this study is to comprehensively investigate the consistencies and differences in aerosol characteristics among multiple global aerosol products from satellites. For this purpose, a total of eleven of the most up-to-date global aerosol products are selected in this paper, including the European Space Agency's Climate Change Initiative (ESA-CCI) products: AATSR-Dual View (AATSR-ADV), AATSR Swansea University (AATSR-SU), AATSR-Oxford-RAL Retrieval of Aerosol and Cloud (AATSR-ORAC) and AATSR-ENSEMBLE (AATSR-EN), which cover the period from 2002-2012, AVHRR (2006-2011), MISR (2000-2017), Terra MODIS (2000-2017), Aqua MODIS (2002-2017), POLDER (2005-2013), SeaWiFS (1997-2010), and VIIRS (2012-2017) products. The newest AERONET Version 3 monthly AOD measurements at 308 globally distributed sites over land and the oceans are collected for comparison.

This manuscript is organized as follows: descriptions of the ten satellite global aerosol products and AERONET data sources are provided in Section 2. In Section 3, the matching methods for the comparisons, the calculation approaches for the aerosol distributions and trends, and quantitative evaluation metrics are presented. The statistical evaluation results for the monthly

AOD retrieval are presented in Section 4. In Section 5, the regional and global AOD distributions are analysed, and comparisons of the aerosol trends and their specific features over the last two decades are provided in Section 6. A summary and conclusions are presented in the final section.

2 Data description

2.1 Satellite-derived aerosol products

2.1.1 ESA-CCI aerosol products

Four typical ESA-CCI global-coverage aerosol products are selected, including the AATSR-ADV, AATSR-SU, AATSR-ORAC, and AATSR-EN. The AATSR-ADV product is generated using the dual view (ADV, Veefkind et al., 1998a) algorithm over land and the single view (ASV, Veefkind and de Leeuw, 1998b) algorithm over the ocean. The ADV algorithm uses the dual view feature and k-ratio approach to eliminate the contribution from the surface to the apparent reflectance. However, this approximation is not reliable over bright surfaces or in the presence of coarse mode aerosols. The ASV algorithm assumes the water is a dark surface at the near-infrared channel, and an ocean reflectance model is applied to correct for the effects of chlorophyll and whitecaps (Kolmonen et al., 2013). The SU algorithm employs a parameterized model of the surface angular anisotropy and estimates the surface spectral reflectance using the dual view feature over land. Over the ocean, the SU algorithm estimates the water-leaving radiance from the ocean at the red and infrared channels at both nadir and along-track view angles with a simple model (North et al., 1999; North, 2002; Bevan et al., 2012). The ORAC algorithm is an optimal estimation retrieval scheme for multispectral images (Thomas et al., 2009; Sayer et al., 2011; Poulsen et al., 2012), which uses a forward model to fit all the shortwave forward and nadir radiances through the DIScrete Ordinate Radiative Transfer (DISORT) model. Meanwhile, the retrieved errors for aerosol parameters are estimated by propagating the measurement and forward model uncertainties into the state space. The AATSR-EN product is integrated based on different ESA-AATSR aerosol products using likelihood estimate approaches (Holzer-Popp et al., 2013). In this study, the latest versions of the above four ESA-CCI products (Table 1) are collected.

2.1.2 MISR aerosol product

The MISR aerosol product provides aerosol distributions over both land and oceans. Over land, MISR is initially based on the dense dark vegetation (DDV) algorithm (Kaufman and Sendra, 1988, King et al., 1992) and uses spatial contrasts to explore an empirical orthogonal function of the angular variations in apparent reflectance. Then, the MISR product is used to estimate the scene path radiance and determine the best-fitting aerosol models. Additionally, the spectral and angular shapes of the reflectance function are assumed to be constant. The algorithm is continuously revised and developed to generate the AOD product with high spatial resolution (4.4 km) based upon the primary underlying physical assumptions (Garay et al., 2017). Over the ocean, water bodies are essentially assumed to be black at the visible and near-infrared wavelengths, and with an additional assumption of an ocean aerosol model, the aerosol retrieval is realized using the radiative transfer theory. MISR multi-angle radiances are used to improve the definition of aerosol models for aerosol retrieval. Recently, a new method was introduced to improve dark-water aerosol retrievals by considering the entire range of cost functions associated with each aerosol mixture, and a new aerosol retrieval confidence index was established to screen high-AOD retrieval blunders caused by cloud contamination or other factors (Witek et al., 2018). In this study, the latest MISR Version 23 monthly aerosol product was selected (Table 1).

2.1.3 MODIS aerosol products

The MODIS aerosol products are generated from three well-known algorithms, including the dark target (DT) algorithms over both the oceans and land and the deep blue (DB) algorithm over only land. Over the oceans, the DT algorithm considers the water as a dark surface from visible to longer wavelengths and neglects the water surface reflectance. Over land, the DT algorithm assumes that the surface reflectances in the visible channels exhibit stable statistical empirical relationships with the 2.1 μm apparent reflectance over the dark target surfaces (Kaufman et al., 1997; Levy et al., 2007). The aerosol retrieval can be realized based on the atmospheric radiative transfer model using the look-up table (LUT) approach. In contrast, the DB algorithm is designed to overcome the flaw in the DT algorithms and realizes aerosol retrieval over bright surfaces, where the surface reflectance in the visible channels is estimated based on the pre-calculated surface reflectance database using the SeaWiFS surface reflectance products (Hsu et al., 2004, 2006). Both algorithms have been continuously improved, and the second-generation operational DT (Levy et al., 2013) and the enhanced DB algorithms (Hsu et al., 2013) were used to generate the latest aerosol products. To increase the data coverage, a new combined DT and DB (DTB) dataset was recently generated according to the independently derived MODIS monthly normalized difference vegetation index (NDVI) products that leverage the strengths of the DT and DB algorithms (Sayer et al., 2014). In this study, the newly released Terra (MOD08) and Aqua (MYD08) Collection 6.1 (C6.1) monthly aerosol products with refinements and improvements made to the above aerosol retrieval algorithms (Wei et al., 2019) are selected (Table 1).

2.1.4 SeaWiFS, AVHRR and VIIRS aerosol products

The SeaWiFS, AVHRR and VIIRS aerosol products over land are generated from the same DB algorithm as MODIS (Hsu et al., 2013) but with some extensions and refinements (Hsu et al., 2017). Over the ocean, these products are based on the Satellite Ocean Aerosol Retrieval (SOAR) algorithm (Sayer et al., 2012; 2017) and include three phases: the selection of suitable pixels to exclude the sun glint, clouds, or suspect of excessively turbid water; pixel-level retrieval; and a post-processing stage (data downscaling and quality assurance). In the SOAR algorithm, the aerosol retrieval simultaneously retrieved the AOD at 550 nm, fine mode fraction (FMF) and the best fit aerosol optical model based on the linear interpolation of pre-calculated LUTs through the Vector LInearized Discrete Ordinate Radiative Transfer (VLIDORT) model. In this study, the newly released SeaWiFS Version 4, AVHRR Version 1 and VIIRS Version 1 monthly aerosol products are selected (Table 1).

2.1.5 POLDER aerosol product

The POLDER/PARASOL aerosol product is generated using the General Retrieval of Atmosphere and Surface Properties (GRASP) algorithm over land and ocean (Dubovik et al., 2011, 2014). The GRASP algorithm is based on the AERONET inversion algorithm and was developed for enhanced characterization of aerosol properties from spectral, multi-angular polarimetric remote sensing observations. POLDER is of great interest as it builds on the design of the forthcoming multi-viewing, multi-channel, multi-polarization (3 MI) instrument (Marbach et al., 2015). POLDER has provided a variety of aerosol characteristics, including spectral AOD, single scattering albedo (SSA), and Ångström exponent (AE); however, the data are available at only latitudes equatorward of 60°. The effect of this restriction on the global analysis is expected to be small because high latitudes are frequently unavailable due to clouds, snow, polar night, and continental land masses (Sayer et al., 2018). In this study, the latest POLDER Version 1.1 monthly aerosol products are selected (Table 1).

2.2 AERONET ground measurements

AERONET is a widely used ground-based observation network with long-term data records at numerous monitoring sites around the world. The AOD observations are available over a wide spectral range from visible to near-infrared channels (0.34–1.02 μm), and they are measured with a high temporal resolution of 15 min and a low bias of 0.01–0.02. The data quality has been divided into three levels (L): L1.0 (unscreened), L1.5 (cloud screened), and L2.0 (cloud screened and quality assured) (Holben et al., 1998; Smirnov et al., 2000; 2009). Meanwhile, the instantaneous AOD observations are further processed and released at daily and monthly levels. In the current study, the newly released AERONET Version 3 L2.0 monthly AOD observations (Giles et al., 2019) are collected to compare with the multi-source satellite-derived monthly aerosol products over land and ocean. The globe is divided into ten custom regions of land, four coastal areas, and four open ocean areas, as illustrated in Figure 1. Table 1 summarizes all the data sources used in this study.

[Please insert Figure 1 here]

[Please insert Table 1 here]

3 Methodology

3.1 Spatial comparison

For multi-satellite aerosol products, the monthly retrievals at 550 nm are collected from the listed scientific dataset (SDS, Table 1) and used for the current analysis in this study. Due to different spatial resolutions, all datasets are uniformly integrated into 1°× 1° grid cells using the bi-directional linear interpolation method. For comparison, monthly retrievals for diverse aerosol products are defined by the pixel centred on the AERONET site, and the corresponding monthly AERONET

AOD is regarded as the true value. Notably, the AERONET sites do not provide the AOD observations at 550 nm; thus, the AOD values at 550 nm are interpolated using the Ångström exponent (α) algorithm from 440–675 nm using the AERONET AOD measured at those wavelengths (Eq. 1). The annual mean AOD value is averaged from at least eight available monthly values over one year.

$$\mathrm{AOD}_{550} = \mathrm{AOD}_{\lambda}(550/\lambda)^{-\alpha} \quad (1)$$

3.2 Temporal trend

The satellite-derived and AERONET-measured monthly mean AOD values are selected for temporal variation and trend analysis; however, to remove the noticeable influence of the annual cycle, the data are first de-seasonalized by calculating the time series of the AOD anomalies. An anomaly is defined as the difference between the monthly mean AOD in one year and the monthly AOD average over all years. Then, the ordinary least squares fitting method (Lai and Wei, 1978; Zdaniuk, 2014) is selected to minimize the sum of residual squares of all observed values and obtain the coefficient of the linear regression slope that represents the temporal trend (AOD yr^{-1}, Eq. 2).

$$Y_t = aX_t + b + N_t,\ t = 1, \ldots, T \quad (2)$$

where Y_t is the AOD time series anomaly, a is the trend (AOD yr^{-1}), b is the offset term, and X_t is the annual time series (X_t = t/12, where t is the individual months in the time series). The term N_t represents the residuals in the time series. However, large-scale systems and seasonal patterns can persist for weeks to months and affect the temporal aerosol trend, and the 1-month lag autocorrelation in the time series is considered in the AOD trend analyses. The uncertainty (σ) in the estimated trend is approximated by the following approach (Weatherhead et al., 1998),

$$\sigma \approx \frac{\sigma_N}{N^{3/2}}\sqrt{\frac{R'}{1-R'}} \quad (3)$$

where σ_N is the standard deviation of the residuals N_t on the fit and R' is the autocorrelation coefficient. The mathematical value and uncertainty range of the AOD trend are represented by a±σ. The statistical significance of the trend is assessed using the two-side test approach, where *p* values less than 0.05 or 0.1 represent trends that are significant at the 95% or 90% confidence level, respectively.

Moreover, the false discovery rate (FDR) is also considered to exclude the fraction of false positives for multiple hypothesis testing (Wilks, 2006). The discovery refers to the rejection of a hypothesis, and a false discovery is an incorrect rejection of a hypothesis, and the FDR is the likelihood that such rejection occurs. The well-known Benjamini–Hochberg procedure is selected to calculate the FDR in this paper (Benjamini and Hochberg, 1995). This procedure begins by ordering the *m* hypothesis by ascending *p* values, where P_i is the p-value at the i_{th} position with the associated hypothesis H_i. Let *k* be the largest *i* for which:

$$P_i \leq \frac{i}{m}\alpha \ (4)$$

Reject hypotheses i = 1, 2, 3..., k. In this study, the FDR is controlled for all tests at the expected level (α = 0.05).

3.3 Statistical metric

To quantitatively evaluate the quality and uncertainty of the retrievals, four main metrics are calculated between the satellite-derived AOD (AOD_S) and AERONET-based AOD (AOD_A). The Pearson product-moment correlation coefficient (R) is selected to measure the linear correlation between the above two variables. The mean absolute error (MAE, Eq. 5) represents the overall estimation accuracy. The root mean square error (RMSE, Eq. 6) and relative mean bias (RMB, Eq. 7) represent the overall estimation uncertainty, where RMB > 1.0 or RMB < 1.0 indicate the over- or under-estimation uncertainty. Moreover, to quantify the performance of each satellite aerosol product in capturing aerosol trends, an additional correct-trend percentage (CTP) is defined as the percentage of sites where the satellite-derived and AERONET-based trends are consistent within each uncertainty or not.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\mathrm{AOD_S} - \mathrm{AOD_A}| \quad (5)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\mathrm{AOD_S} - \mathrm{AOD_A})^2} \quad (6)$$

$$\mathrm{RMB} = \frac{1}{n}\sum_{i=1}^{n}|\mathrm{AOD_S}/\mathrm{AOD_A}| \quad (7)$$

4 Performance of monthly aerosol products

4.1 Global-scale comparison

Figure 2 compares the monthly AOD_S values derived from ten satellite aerosol products and AOD_A values at a total of 268 available AERONET sites for the common period 2006-2010 throughout the world (VIIRS data are not discussed in Section 4 because they start in 2012). Due to the differences in aerosol retrieval algorithms and satellite observation conditions, the spatial coverage is not uniform among these products, which results in noticeable differences in the number of data collections (sample size, N). The four ESA-CCI monthly aerosol products show similar overall performance with comparable evaluation metrics. The AOD retrievals (N = 7938-9467) agree well with AOD_A (R = 0.7-0.8), with average MAE values ranging from 0.07 to 0.09 and average RMSE values ranging from 0.13 to 0.15. Among them, the AATSR-SU (AATSR-ADV) product shows the best (worst) performance with the smallest (largest) estimation uncertainties on the global scale. These results are consistent with those reported by a previous study (de Leeuw et al., 2015). The AVHRR AOD_S values (N = 8382) are well correlated with the AERONET AOD_A values with an average MAE and RMSE of 0.077 and 0.145, respectively. The Terra MISR product provides a sample size of 8418, which is smaller than the Terra MODIS sample size (N = 9196) and is possibly due to the narrower swath width. MISR AOD_S values are highly correlated with the ground-measured AOD_A values (R = 0.781), with an average MAE of 0.074 and RMSE of 0.127. The Terra MODIS product is generally better than the MISR product with a high correlation and low RMSE. Due to the afternoon imaging time, the Aqua MODIS product provides approximately 2% fewer data collections than Terra MODIS, but it exhibits superior performance in terms of most of the evaluation metrics (i.e., R = 0.868, MAE = 0.067, and RMSE = 0.107) among all ten products. In contrast, the POLDER product exhibits inferior performance with the largest MAE and RMSE errors among all the products, significantly overestimating the monthly aerosol loads (RMB = 1.287). This result could be partially attributed to the relatively low accuracy of cloud detection results in the current POLDER product, and an upcoming version of the POLDER product with an advanced algorithm will improve the AOD retrievals. The SeaWiFS product has the smallest sample size, which provides 33-44% fewer data collections than other products but exhibits overall good performance. In general, both MODIS and POLDER products overestimate and other products underestimate the monthly average aerosol loads, especially the MISR and AATSR-ADV products.

[Please insert Figure 2 here]

Table 2 summarizes the comparison of the AOD_S and AOD_A values from the ten products over land and ocean for the common period 2006-2010. Over land, the AVHRR and four ESA-CCI products show good performances with similar evaluation metrics. The MISR, MODIS and SeaWiFS products exhibit generally good performance with high correlations (R > 0.8) and low MAE (< 0.08) and RMSE (< 0.13) values. However, the SeaWiFS product provides the minimum number of matched samples. In general, Aqua MODIS yields the best performance with the highest estimation accuracy (MAE = 0.068) and lowest estimation uncertainty (RMSE = 0.110), showing only approximately 9% overestimations (RMB = 1.09). Over the ocean, MODIS AOD_S provides the maximum number of matching samples, the SeaWiFS product provides the minimum number, and the other products provide similar sample sizes with a small difference within 10%. The comparison of the AOD_A values over the ocean indicates that the AATSR-SU and MISR products underestimate the values, while the others generally overestimate the values, especially the Terra MODIS, POLDER, and AATSR-ORAC products. In general, the four ESA-CCI products perform similarly. POLDER and MISR products perform poorly with MAE (> 0.07) and RMSE (> 0.11) values that are larger than those of the other products. The AVHRR and MYD08 products are most accurate over the ocean, with the lowest estimation uncertainties (RMSE = 0.078 and 0.082) among the ten products.

[Please insert Table 2 here]

4.2 Continent-scale comparison

Aerosol characteristics over land are more diverse than those over the ocean due to complex surface structures, varying aerosol compositions, and influences of natural and human factors. Therefore, this section focuses on the comparison between monthly AOD_S and AOD_A at the continental scale over land. For this purpose, ten main customized continents (Figure 1) are considered, including eastern North America (ENA), western North America (WNA), South America (SAM), Europe (EUR), Africa (AFR), the Middle East (ME), South Asia (SAA), East Asia (EAA), Southeast Asia (SEA) and

Oceania (OCE). Figure 3 shows the continent-scale performance for ten AOD_S products for the common period 2006-2010 over land, and the statistical results are given in Table S1.

The results show some common features of the ten AOD_S products. In general, a large number of data samples are collected over Europe and North America due to intensive ground-based observation sites. In contrast, the sample sizes are small over the Middle East, East Asia, Southeast Asia, and Oceania due to the sparse observation sites and algorithm limitations over the high-brightness underlying surfaces. Most aerosol products exhibit good performances with low MAE and RMSE values less than 0.06 and 0.08 over Europe, North America, and Oceania. The main reason for this result is that the relatively high vegetation coverage and dark underlying surface allow for more accurate aerosol retrievals by different aerosol algorithms. However, poor performances with large MAE and RMSE values occur over South Asia, East Asia, Africa, and the Middle East. This result is mainly due to the complex and bright underlying surfaces (e.g., desert, bare land, and urban areas), as well as intense human activities, which increase the difficulty of aerosol estimation. Overall, most aerosol products overestimate the monthly AOD over North America and Oceania, while general underestimations occur over South America, Africa, and East Asia.

The performance of each AOD_S product is also distinct in each specific region. The SeaWiFS product has the smallest sample size, while the AATSR-ORAC, POLDER and MODIS products provide a larger number of data samples than the other products. In particular, the AATSR-ADV product provides fewer data samples over the Middle East than over the other regions because the ADV algorithm cannot be applied in bright desert areas. In terms of the retrieved AOD, all the products perform almost equally with similar evaluation metrics (e.g., MAE, RMSE) over North America, Europe, and Oceania, except for the POLDER product. In the other regions, large differences are found among the ten AOD_S products. In general, the MODIS and MISR products exhibit better performances (with low MAE and RMSE values) than the other products over South America, Africa, the Middle East, East Asia and Southeast Asia. The POLDER and MODIS products overestimate the monthly aerosol loads over most continents, especially America and Europe. In contrast, the AATSR-ORAC, AATSR-ADV and MISR products usually underestimate the monthly aerosol loads except for a few specific regions (i.e., western North America and Oceania).

[Please insert Figure 3 here]

4.3 Site-scale comparison

The global- and continent-scale comparisons show the overall performance of ten satellite aerosol products. However, the selected AERONET sites are unevenly distributed around the world, with most sites concentrated in densely populated land regions. Therefore, the site-scale comparison at a total of 308 available sites is performed in this section. For this purpose, four main evaluation metrics are calculated, including the sample size (N), MAE, RMSE, and RMB. For statistical significance, only those sites with at least half a year of observations (6 matchups) are used for analysis. Figure 4 shows the site-scale performance map for AOD_S against AOD_A, and Table 3 summarizes the percentages of the sites within a certain range of evaluation metrics for all AOD_S products in the common period 2006-2010.

Figure 4i illustrates the number of data collections for the different AOD_S products at each site over both land and ocean, where the black dots represent an insufficient number of matchups. Most products can provide enough data samples at more than 95% of the sites around the world, especially the AATSR and MODIS products. However, the SeaWiFS product has approximately 21% of the sites with no or few matchup samples, which are mainly distributed over North America, Europe, Asia, and Southeast Asia. The AATSR-ADV product has approximately 8% of the sites lacking matched samples, which are spread over North Africa, Southern Europe, the Middle East, and Central Asia. The main reason for this result is that the ADV algorithm cannot be adequately applied over bright surfaces. Moreover, the sites with no matched data samples from the POLDER product are concentrated in high-latitude areas because the POLDER algorithm is designed for aerosol retrieval between 60 degrees north-south latitude.

Figure 4ii and 4iii plot the MAE and RMSE errors between AOD_S and AOD_A at each site over the world. The MAE and RMSE maps have very similar spatial patterns for each aerosol product. Good performances are exhibited at most North American and European sites with low MAE and RMSE values less than 0.04 and 0.06, respectively. The sites with poor performances are mainly aggregated in North Africa, East Asia, and South Asia, where the MAE and RMSE values are generally greater than 0.16 and 0.20, respectively. This result indicates that the overall performance of the aerosol products at the site scale is spatially heterogeneous and highly dependent on the type of underlying surfaces and the impact of human activities. Among the ten aerosol products, the Aqua MODIS product shows the best performance, having a large percentage of sites (71% and 60%) with MAE and RMSE values less than 0.08 throughout the world. By contrast, the POLDER product performs the worst, having more than 31% and 47% of the sites with MAE and RMSE values greater than 0.12.

Figure 4iv shows the spatial distribution of the site-scale AOD_S bias. For the ten products, only 14~32% of the sites show good estimations with an average RMB between 0.9 and 1.1. The POLDER and MOD08 products overestimate at most sites, especially in North America and Europe, and more than 54% and 61% of the sites show significant overestimations (RMB > 1.2) according to the statistics in Table 3. The other products mostly underestimate at sites over Europe, Africa, the Middle East, and Asia and overestimate at sites over South America and Australia.

[Please insert Figure 4 here]

[Please insert Table 3 here]

5 AOD spatial coverage and distribution

5.1 Global and regional distribution

In this section, we compare the AOD distribution among the eleven aerosol products (VIIRS data are included). Figure 5 illustrates the global spatial coverage and mean value of all AOD_S products for their respective available periods from 1997-2017. There are several missing monthly data records for the AATSR-ADV, AATSR-ORAC, AVHRR and SeaWiFS products, which are given in Table S2.

All the aerosol products present a similar and obvious annual cycle, with high spatial coverage in September and October and low coverage in December and January (Figure 5a). In general, the MODIS and VIIRS products provide the largest spatial coverage, covering more than 64% of the area of the world. In contrast, the SeaWiFS and AATSR-ADV products have the lowest spatial coverage, with global averages of 46% and 51%, respectively. The AATSR-EN, AATSR-SU, AVHRR, MISR and POLDER products have similar spatial coverages, with an average of 51~58%. The spatial coverage decreased significantly as the SeaWiFS and POLDER satellite services approached their end stages. Figure 5b shows similar annual variations among the eleven AOD_S products, with the peak from July to September and the trough from November to January. The POLDER product exhibits the highest AOD values among all products, while the SeaWiFS and MISR products show the lowest values. The other products have relatively similar AOD_S values ranging from 0.13 to 0.18. Finally, we found that the VIIRS product is almost identical to the Aqua MODIS AOD_S, as shown in Figure 5, due to the similar satellite parameters and algorithms. Considering the relatively short data records of VIIRS, we will not include these data in the subsequent comparison and analysis.

[Please insert Figure 5 here]

Considering the remarkable seasonal variations, we plot the seasonal spatial distributions of the ten aerosol products for their common period 2003-2010 in Figure 6. Meanwhile, we also reproduce the satellite-derived global AOD_S maps considering the common points in all datasets separately over land and ocean (Figures S2-S3). Table 4 summarizes the average spatial coverage and AOD_S values in December-January-February (DJF), March-April-May (MAM), June-July-August (JJA) and September-October-November (SON) for each product. In DJF, the spatial coverage of AOD_S is the lowest for most aerosol products, especially the AATSR-ADV product (~54%). The missing data are mainly in the Northern Hemisphere in winter and in high-latitude areas with bright surfaces covered by snow and ice, where most of the retrieval algorithms cannot be implemented. For the spatial distribution of AOD_S, noticeable spatial heterogeneity occurs over land with low values in North America, Europe, and Australia and high values in North Africa, the Middle East, South Asia, and East Asia. Deserts, dry areas and their downwind regions have AOD_S peaks in spring (East Asia) or summer (North Africa and the Middle East) in accordance with the prevailing time of dust. Anthropogenic polluted regions exhibit peaks in high emission seasons, such as dry seasons in the savanna and Amazon due to biomass burning, summer in East Asia due to the formation of large amounts of fine particles and water uptake by hygroscopic particles. There is also strong diversity in the seasonal or annual mean AOD$_S$ over North Africa and East Asia among most datasets (Figure S2). This diversity is mainly due to the different aerosol algorithms applied over bright surfaces (i.e., desert and urban areas). Both high surface reflectance and complex underlying surfaces increase the difficulty of aerosol retrieval (Wei et al., 2018). For the spatial distributions over the ocean, the seasonal and annual mean AOD$_S$ values are generally lower than 0.1 in most areas, especially open seas (Figure S3). In coastal areas near Central and North Africa, Southern Middle East, Southern India and East China, the AOD$_S$ values are strongly influenced by the source regions. The seasonal mean AOD$_S$ values are generally high greater than 0.4, and the seasonal variation in AOD$_S$ in the downstream plume areas is consistent with that in the upstream land area.

[Please insert Figure 6 here]

[Please insert Table 4 here]

Figure 7 plots the seasonal spatial coverage and mean AOD$_S$ values over ten land and eight oceanic customized regions (see Figure 1) for each product during the common period 2003-2010. The results illustrate that the SeaWiFS and AVHRR products have much lower spatial coverage than the other products over most land regions, especially for South America, South Asia, and Southeast Asia. The range in spatial coverage of all AOD$_S$ products is greater in winter than in other seasons (especially over land regions). The AOD$_S$ products are more consistent and have higher spatial coverage over the ocean than over land; the average spatial coverage can even reach up to 100% in summer.

For the seasonal mean AOD$_S$, the POLDER product has the highest values, and the SeaWiFS product has the lowest values over most customized regions. The AATSR-ADV product exhibits the lowest seasonal AOD values in the Middle East due to a large amount of missing retrievals. For the remaining aerosol products, the range in the seasonal mean AOD$_S$ is greater than 0.2 over Africa, South Asia, East Asia, Southeast Asia, and the coastal areas of South Asia and East Asia. The main reason for this wide range could be the complex aerosol types from multiple sources (e.g., natural dust mixed with anthropogenic fine particles) that cannot be resolved by current aerosol retrieval algorithms. For the remaining land and ocean regions, the range in seasonal AOD values is generally within 0.1 among these aerosol products. The main reason for this result may be the differences in satellite scanning widths and pixel selection during the reprocessing of the monthly aerosol products.

[Please insert Figure 7 here]

5.2 Comparison between seasonal AOD$_S$ and AOD$_A$

Figure 8 compares the satellite-derived seasonal mean AOD$_S$ value for each satellite over AERONET sites with the ground-based AOD$_A$ values over land and ocean, and the statistical results are given in Table S3. The best performance with the smallest MAE (Figure 8b) and RMSE (Figure 8c) values are always found in SON. In contrast, the worst performances with the largest estimation uncertainties (i.e., MAE and RMSE) among the ten aerosol products are found in JJA. In general, the

MODIS and POLDER products overestimate, and the remaining seven aerosol products underestimate the aerosol loads in the four seasons (Figure 8d). The performance of the AATSR-ORAC and AVHRR products is poor with large estimation uncertainties in JJA but much improved in the other three seasons. The AATSR-SU product shows the smallest estimation bias (RMB = 0.95~1.05) in all four seasons among all products. In general, the Aqua MODIS product performs best with almost all the best evaluation metrics (e.g., N, MAE, and RMSE) compared to the other products on the seasonal level.

[Please insert Figure 8 here]

In Figure 9 we also compare the annual mean AOD_S values from each satellite product with the AERONET AOD_A values at available sites from 2003 to 2010. The results indicate that similar conclusions can be drawn for both seasonal and annual scales. The AATSR-SU product performs superior among the four ESA-CCI AATSR products. The AVHRR and MISR products show similar performance with close MAE (0.049 and 0.050) and RMSE (0.082 and 0.083) values but underestimate the annual mean AOD (RMB = 0.972 and 0.881). However, these products are overall better than the ESA-CCI AATSR products. The POLDER and SeaWiFS products exhibit poor performance due to the notable overestimation (RMB=1.307) and the smallest number of matchup samples, respectively. The MODIS products have noticeably high correlations with ground measurements (R > 0.92), but MOD08 shows an ~17% overestimation. In general, the MYD08 product has the best performance with the smallest estimation uncertainties (MAE = 0.047 and RMSE = 0.069) among all the aerosol products.

[Please insert Figure 9 here]

6 AOD temporal variation and trend

6.1 Global and regional AOD trend

In this section, we focus on the comparison of the temporal trends of global and regional AOD products. Because the AVHRR and POLDER products provide less than ten years of aerosol observations in this study, only the remaining eight long-term aerosol products are compared for a common observation period. To ensure that the long-term trend is not impacted by the trends of the aerosol products themselves, we calculated the autocorrelation coefficient of each product with a one-month lag (Figure 10). The results suggest that the magnitudes of the autocorrelation coefficient for most aerosol products are generally small and range from -0.3 to 0.3 over more than 90% of the world. This result indicates that the time series of AOD_S data are stable with weak self-influence and are suitable for long-term trend analysis.

[Please insert Figure 10 here]

The linear trends are derived from the de-seasonalized monthly anomaly of each AOD_S, and a two-sided test is conducted to present the statistical significance of the temporal trends, where the trends that are significant at the 95% confidence level ($p < 0.05$) are marked with black dots in Figure 11. Considering the multiple hypothesis testing (many data sets and locations are being tested for trends), there could be a significant fraction of false positives. Therefore, the FDR test at the 95% significance level (α = 0.05) is performed to address this issue. We see that the false positive points can be adequately eliminated after the FDR adjustment and that the statistically significant areas are more or less reduced (comparing Figure 11 with Figure S1). After these processes, the trends in Figure 11 are realistically able to represent the time evolution of aerosols.

The global AOD trend distribution shows similar overall spatial patterns among all aerosol products. Over land, significantly positive trends ($a > 0.01$, $p < 0.05$) are mainly found in the Middle East and South Asia, indicating increasing air pollution. In contrast, significantly negative aerosol trends ($a < -0.01$, $p < 0.05$) are mainly observed in eastern North America, Europe, and central Africa, indicating improved air quality. Large trends greater than 0.01 yr^{-1} but not statistically significant are found in a few areas of North Africa and East Asia. Strong negative but statistically nonsignificant trends are found in central South America and parts of Southeast Asia. The large trends indicate the importance of aerosol evolution, and the lack of significance may be attributed to the complex aerosol sources; thus, more attention should be placed on these areas to better understand the temporal variations in aerosols. The magnitude of the aerosol trend is generally small ($|a| < 0.005$) over the ocean. However, significantly decreasing aerosol trends ($a > 0.01$, $p < 0.05$) are observed along the west coast of South America, the east coast of North America and the east coast of Asia. A significant increase in aerosol trends ($a < -0.01$, $p < 0.05$) was observed along the Indian coast. On the other hand, the four ESA-CCI and MISR aerosol products are not significant in most ocean areas, even for the open seas. MODIS and SeaWiFS products have similar spatial patterns in most ocean areas, such as the significantly increasing trends observed over the Pacific and Indian Oceans.

[Please insert Figure 11 here]

Figure 12 compares the regional aerosol trends among the eight satellite AOD_S values, and Table S4 shows the statistics of the regional AOD_S trends and uncertainties. Over land, most small trends are not statistically significant, indicating unassured temporal trends over most land regions. However, most products show significantly increasing trends over the Middle East ($a = 0.0048$~0.0111 yr^{-1}, $p < 0.05$) and South Asia ($a = 0.0034$~0.0047 yr^{-1}, $p < 0.05$), confirming the robust enhancement of aerosols in these two regions. Some products also exhibit obvious decreasing aerosol trends over eastern North America, western North America, Europe and Southeast Asia. The robustness of the decreasing trends is credible in eastern North America and Europe but unsure in western North America. Over the ocean, the aerosol trends are generally small, especially for the three open ocean areas (i.e., Pacific, Indian and Atlantic Oceans in Figure 12b). However, the aerosol changes in the four coastal areas exceed 0.002 yr^{-1}. The downward trends on the eastern North American coast, European coast and the rising trend on the South Asian coast are robust. The temporal trend over the East Asian coast is unassured.

[Please insert Figure 12 here]

6.2 Comparison between AOD_S and AOD_A trends

The satellite-derived AOD_S trends are compared against the AERONET AOD_A trends from ground measurements. To ensure the statistical significance of the trend calculations, only the AERONET sites with at least five years (120 months) effective observations are selected. Figure 13 plots the AOD_S and AOD_A trends at all available sites for the eight satellite products. Most products can capture the AOD trends with the CTPs ranging from 40% to 45%. The SeaWiFS product has valid comparisons at only 59 sites due to the lack of retrieval over land, and the AOD_S trend exhibits the worst performance, with the largest MAE and RMSE values among all the aerosol products. Terra and Aqua show similar performance with almost equal CTPs of 42%, and the MODIS products capture the temporal AOD_S trend most accurately with the lowest MAE and RMSE.

[Please insert Figure 13 here]

6.3 AOD_S trend over the past two decades (2000-2017)

Based on the above conclusions and considering the time length, the Terra MODIS product is selected as a representative to study the aerosol variations over the past two decades. Figure 14 plots the global spatial distribution of the linear MOD08 AOD_S trends from January 2000 to December 2017 using the same approach as in Section 6.1, and Table 5 shows the regional AOD_S trends and uncertainties. Note that the upward and downward trends could be offset over such a long period of 18 years.

The MOD08 AOD_S trends are generally weak. The average trend over the entire land area is 0.0001 yr^{-1} and is not statistically significant. However, the trends in some specific land regions are worth noting. For example, fast-developing countries such as India in South Asia ($a = 0.0027 \pm 0.0010$ yr^{-1} and $p < 0.05$) and the North China Plain in East Asia show significantly increasing aerosol trends. The main reason for these trends is the acceleration of urbanization and increasing anthropogenic pollutant emissions caused by intense human activities (e.g., industrial pollution, fossil fuel combustion and straw burning), which have also been reported in previous studies (Lu et al., 2011; de Meij, et al., 2012; Suresh et al., 2013; Sogacheva et al., 2018). In dust dominant regions such as the Middle East, a significantly positive trend ($a = 0.0023 \pm 0.0012$ yr^{-1}, $p < 0.05$) is also observed due to enhanced dust emissions associated with unfavourable meteorological conditions (e.g., increasing temperature and decreasing relative humidity) (Hsu et al., 2012; Klingmüller et al., 2016). Meanwhile, the increasing trends in western North America and central Africa can be attributed to the biomass burning of forest fires (Edwards et al., 2006; Gavin et al., 2007; Kondo et al., 2011; Das et al., 2017). In contrast, significantly negative trends are found over eastern North America (-0.0009 ± 0.0004 yr^{-1}, $p < 0.05$), Europe (-0.0014 ± 0.0005 yr^{-1}, $p < 0.05$), central South America, central and southeast China, and Japan in East Asia (< -0.01 yr^{-1}, $p < 0.05$). These results are in good agreement with the results of other studies, and these negative trends are mainly due to the favourable climatic conditions and the decrease in pollution aerosols associated with government emissions control (Hsu et al., 2012; de Meij et al., 2012; Hu et al., 2017; Li et al., 2019).

Over most of the global ocean, MOD08 AOD_S shows an obvious increasing trend (0.0005 yr^{-1}, $p < 0.05$). At the regional scale, the Pacific Ocean ($a = 0.0009 \pm 0.0002$ yr^{-1}, $p < 0.05$), South Atlantic Ocean ($a = 0.0013 \pm 0.0003$ yr^{-1}, $p < 0.05$), Indian Ocean ($a = 0.008 \pm 0.0002$ yr^{-1}, $p < 0.05$), and coastal areas of South Asia ($a = 0.0042 \pm 0.0008$ yr^{-1}, $p < 0.05$) have notable positive trends. These results are comparable to the results of previous studies (Hsu et al., 2012; Sayer et al., 2018), and the main reason is the transport of mineral dust and smoke from biomass burning (Edwards et al., 2006; Das et al., 2017). In contrast, significantly negative trends are found over the coastal areas of eastern North America (-0.0019 ± 0.0004 yr^{-1}, $p < 0.05$), Europe ($a = -0.0011 \pm 0.0003$ yr^{-1}, $p < 0.05$), and western South America. The reduction in aerosols over these areas is mainly due to the decreased dust transport from the Sahara and the control/reduction of pollutant emissions by human activities (Hsu et al., 2012; Sayer et al., 2018). Overall, the temporal variations in global aerosol loads are strongly influenced by both natural and human sources, which need to be further investigated in our future studies.

[Please insert Figure 15 here]

[Please insert Table 5 here]

7 Summary and conclusion

This study focuses on the similarities and differences in the spatial variations and temporal trends of the current satellite-derived AOD products. For this purpose, eleven global monthly aerosol products at coarse spatial resolutions are collected and compared against the ground measurements from 308 AERONET sites throughout the world, including four products from the European Space Agency's Climate Change Initiative (AATSR-ADV, AATSR-EN, AATSR-ORAC, and AATSR-SU) and AVHRR, MISR, Terra and Aqua MODIS, POLDER, SeaWiFS, and VIIRS products. These data are evaluated in three ways: 1) direct comparison of monthly retrievals against the AERONET observations at global, continent, and site scales; 2) comparison of the global and regional AOD spatial coverage and distribution; and 3) comparison of the global and regional AOD temporal variations and trends. Our results may help readers to better understand the features of different satellite aerosol products and select a suitable aerosol dataset for their respective studies.

In terms of the performance of multiple products at different spatial scales, we show that the four ESA-CCI aerosol products show similar performance and are generally worse than the AVHRR and MISR products. The SeaWiFS product provides the smallest sample size despite an overall good performance. The seven abovementioned products underestimate the aerosol loads, especially the MISR and AATSR-ADV products. The POLDER product performs worst with the largest estimation uncertainties and significantly overestimates the aerosol loads. The MODIS products (especially Aqua MODIS) show superior performance among all products with small estimation uncertainties at most regions and sites but overestimate AOD overall. In general, most products exhibit consistently good performance over dark surfaces in Europe and North America but perform worse over bright and complex surfaces in South Asia, East Asia, Africa, and the Middle East.

In terms of the spatial distribution of aerosols, the SeaWiFS and AATSR-ADV products have poor spatial continuities with numerous missing values, while the MODIS products can provide almost full coverage throughout the world. Most products show the highest spatial coverage and aerosol concentrations in summer but the lowest concentrations in winter. In general, the seasonal aerosol spatial distributions over the ocean are more consistent among the different aerosol products. However, noticeable spatial heterogeneity and numerical differences are observed over land, especially over Africa, Asia, and some coastal areas, which are possibly due to the complex aerosol sources and the limitations of the different aerosol retrieval algorithms. In general, the best performance in describing the seasonal aerosol distributions is always observed in autumn, but the worst is observed in summer. The Aqua MODIS product performs best with almost all the best evaluation metrics (e.g., MAE and RMSE) among all the products at the seasonal and annual levels.

In terms of the temporal aerosol trends, most products exhibit similar spatial patterns throughout the world, where significantly positive trends are found over the Middle East, South Asia and South Asian coasts. In contrast, significantly decreasing trends are observed over eastern North America, Europe, and their coastal areas. In general, most products can capture the correct AOD trends at approximately 40% of the AERONET sites. The MODIS products show the best performance with the best evaluation metrics. The aerosol trends of the Terra MODIS product over the past two decades (2000-2017) show that the temporal variations in some land regions are unassured but important, which could be attributed to the complexity of the earth-atmosphere system and the interference from human activities. This finding should be investigated in detail in future work.

Author contribution

All authors made substantial contributions to this work. Y. Peng designed the research, and J. Wei carried out the research and wrote the initial draft of this manuscript. R. Mahmood, L. Sun and J. Guo helped review the manuscript. We declare no conflicts of interest.

Acknowledgements

The ESA-CCI AATSR monthly products are obtained from the ICARE Data and Services Centre (http://www.icare.univ-lille1.fr/cci). The MODIS, MISR, AVHRR and SeaWiFS monthly products are available at https://search.earthdata.nasa.gov/. The POLDER product is available at https://www.grasp-open.com/products/polder-data-release/, and the AERONET measurements are available from the NASA Goddard Space Flight Center (https://aeronet.gsfc.nasa.gov/). This work was supported by the National Natural Science Foundation of China (71690243, 41775137 and 41761144056) and the National Important Project of the Ministry of Science and Technology in China (2017YFC1501404). All authors made

substantial contributions to this work. Y. Peng designed the research, and J. Wei carried out the research and wrote the initial draft of this manuscript. R. Mahmood, L. Sun and J. Guo helped review the manuscript. We declare no conflicts of interest.

References

Bartell, S. M., Longhurst, J., Tjoa, T., Sioutas, C., & Delfino, R. J. (2013). Particulate air pollution, ambulatory heart rate variability, and cardiac arrhythmia in retirement community residents with coronary artery disease. Environmental Health Perspectives, 121(10), 1135.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological), 57(1), 289-300.

Bevan, S. North, P. Los, S. & Grey, W. (2012). A global dataset of atmospheric aerosol optical depth and surface reflectance from AATSR. Remote Sensing of Environment 116-210. doi:10.1016/j.rse.2011.05.024

Brauer, M., Amann, M., Burnett, R. T., Cohen, A., Dentener, F., & Ezzati, M., et al. (2012). Exposure assessment for estimation of the global burden of disease attributable to outdoor air pollution. Environmental Science and Technology, 46(2), 652-660.

Cao, J. J., Wang, Q. Y., Chow, J. C., Watson, J. G., Tie, X. X., & Shen, Z. X., et al. (2012). Impacts of aerosol compositions on visibility impairment in Xi'an, China. Atmospheric Environment, 59, 559-566.

Crouse, D.L., Peters, P.A., van Donkelaar, A., Goldberg, M.S., Villeneuve, P.J., Brion, O., et al., 2012. Risk of nonaccidental and cardiovascular mortality in relation to long-term exposure to low concentrations of fine particulate matter: a Canadian national-level cohort study. Environmental Health Perspectives. 120 (5), 708–714.

Das, S., Harshvardhan, H., Bian, H., Chin, M., Curci, G., Protonotariou, A. P., et al. (2017). Biomass burning aerosol transport and vertical distribution over the South African-Atlantic region. Journal of Geophysical Research: Atmospheres, 122, 6391–6415. https://doi.org/10.1002/2016JD026421

de Leeuw, G., Holzer-Popp, T., Bevan, S., Davies, W. H., Descloitres, J., & Grainger, R. G., et al. (2015). Evaluation of seven European aerosol optical depth retrieval algorithms for climate analysis. Remote Sensing of Environment, 162, 295-315.

de Meij, A., Pozzer, A., & Lelieveld, J. (2012). Trend analysis in aerosol optical depths and pollutant emission estimates between 2000 and 2009. Atmospheric Environment, 51, 75-85.

Dubovik, O., Herman, M., Holdak, A., Lapyonok, T., Tanre, D., Deuz ˙ e, J. L., et al. (2011). Stastically optimized inversion algorithm for enhanced retrieval of aerosol properties from spectral multi-angle polarimetric satellite observations. Atmospheric Measurement Techniques, 4, 975–1018. https://doi.org/10.5194/amt-4-975-2011

Dubovik, O., Lapyonok, T., Litvinov, P., Herman, M., Fuertes, D., Ducos, F., et al. (2014). GRASP: A versatile algorithm for characterizing the atmosphere. Newsroom: SPIE. https://doi.org/10.1117/2.1201408.005558

Edwards, D. P., Emmons, L. K., Gille, J. C., Chu, A., Attié, J.-L., Wood, S. W., et al. (2006). Satellite-observed pollution from Southern Hemisphere biomass burning. Journal of Geophysical Research, 111, D14312. https://doi.org/10.1029/2005JD00665

Floutsi, A. A., Korrascarraca, M. B., Matsoukas, C., Hatzianastassiou, N., & Biskos, G. (2016). Climatology and trends of aerosol optical depth over the mediterranean basin during the last 12 years (2002-2014) based on Collection 006 MODIS-Aqua data. Science of the Total Environment, 551-552, 292-303.

Gavin, D. G., Hallett, D. J., Hu, F. S., Lertzman, K. P., Prichard, S. J., & Brown, K. J., et al. (2007). Forest fire and climate change in western North America: insights from sediment charcoal records. Frontiers in Ecology and the Environment, 5(9), 499-506.

Giles, D.M., Sinyuk, A., Sorokin, M.G., Schafer, J.S., Smirnov, A., Slutsker, I., Eck, T.F., Holben, B.N., Lewis, J.R., Campbell, J.R., Welton, E.J., Korkin, S.V., Lyapustin, A.I. (2019). Advancements in the Aerosol Robotic Network (AERONET) Version 3 database – automated near-real-time quality control algorithm with improved cloud screening for Sun photometer aerosol optical depth (AOD) measurements. Atmospheric Measurement Techniques, 12, 169–209. https://doi.org/10.5194/amt-12-169-2019.

Guo, J. P., Zhang, X. Y., Wu, Y. R., Zhaxi, Y., Che, H. Z., & Ba, L., et al. (2011). Spatio-temporal variation trends of satellite-based aerosol optical depth in China during 1980–2008. Atmospheric Environment, 45(37), 6802-6811.

Guo, J., M. Deng, S. S. Lee, F. Wang, Z. Li, P. Zhai, H. Liu, W. Lv, W. Yao, and X. Li (2016). Delaying precipitation and lightning by air pollution over the Pearl River Delta. Part I: Observational analyses. Journal of Geophysical Research: Atmospheres, 121, 6472–6488, doi:10.1002/2015JD023257.

Guo, J., Su, T., Li, Z., Miao, Y., Li, J., Liu, H., Xu, H., Cribb, M. and Zhai, P. (2017). Declining frequency of summertime local-scale precipitation over eastern China from 1970 to 2010 and its potential link to aerosols. Geophysical Research Letters, 44(11), 5700-5708.

Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., et al. (1998). AERONET: A federated instrument network and data archive for aerosol characterization. Remote Sensing of Environment, 66, 1–16. https://doi.org/10.1016/S0034-4257(98)00031-5

Holzer-Popp, T., Leeuw, G. D., Martynenko, D., & Klüser, L. (2013). Aerosol retrieval experiments in the esa aerosol_cci project. Atmospheric Measurement Techniques, 6,8(2013-08-08), 6(8), 1919-1957.

Hsu, N. C., Gautam, R., Sayer, A. M., Bettenhausen, C., Li, C., & Jeong, M. J., et al. (2012). Global and regional trends of aerosol optical depth over land and ocean using SeaWiFS measurements from 1997 to 2010. Atmospheric Chemistry and Physics, 12(17), 8037-8053.

Hsu, N. C., J. Lee, A. M. Sayer, N. Carletta, S.-H. Chen, C. J. Tucker, B. N. Holben, and S.-C. Tsay (2017), Retrieving near-global aerosol loading over land and ocean from AVHRR, Journal of Geophysical Research: Atmospheres, 122, 9968–9989, doi:10.1002/2017JD026932.

Hsu, N. C., Jeong, M. -J., Bettenhausen, C., Sayer, A. M., Hansell, R., Seftor, C. S., . . . Tsay, S. -C. (2013). Enhanced deep blue aerosol retrieval algorithm: The second generation. Journal of Geophysical Research: Atmospheres, 118(16), 9296–9315. doi:10.1002/jgrd.50712.

Hsu, N. C., Tsay, S. C., King, M. D., & Herman, J. R. (2004). Aerosol properties over bright-reflecting source regions. IEEE Transactions on Geoscience and Remote Sensing, 42(3), 557-569.

Hsu, N. C., Tsay, S. C., M.D. King, & Herman, J. R. (2006). Deep blue retrievals of Asian aerosol properties during ace-Asia. IEEE Transactions on Geoscience and Remote Sensing, 44(11), 3180-3195.

Hu, K., Raghavendra, K., Kang, N. Boiyo, R., Wu, J. (2017). Spatiotemporal characteristics of aerosols and their trends over mainland China with the recent Collection 6 MODIS and OMI satellite datasets, Environmental Science and Pollution Research, 2018, 25, 7, 6909–6927.

Kaufman, Y. J., Wald, A. E., Remer, L. A., Gao, B. C., Li, R. R., & Flynn, L. (1997). The MODIS 2.1 mm channel correlation with visible reflectance for use in remote sensing of aerosol, IEEE Transactions on Geoscience and Remote Sensing, 35(5), 1286-1298.

King, M. D., Kaufman, Y. J., Menzel, W. P., & Tanré, D. (1992). Remote sensing of cloud, aerosol, and water vapor properties from the moderate resolution imaging spectrometer (MODIS). IEEE Transactions on Geoscience and Remote Sensing, 30(1), 2-27.

Klingmueller, K., Pozzer, A., Metzger, S., Abdelkader, M., Stenchikov, G., & Lelieveld, J. (2016). Aerosol optical depth trend over the Middle East. Atmospheric Chemistry and Physics, 16(8), 5063-5073.

Kolmonen, P., A.M. Sundstr?m, Sogacheva, L., & Rodriguez, E. (2013). Uncertainty characterization of aod for the aatsr dual and single view retrieval algorithms. Atmospheric Measurement Techniques Discussions, 6(2), 4039-4075.

Kondo, Y., Matsui, H., Moteki, N., Sahu, L., Takegawa, N., & Kajino, M., et al. (2011). Emissions of black carbon, organic, and inorganic aerosols from biomass burning in North America and Asia in 2008. Journal of Geophysical Research Atmospheres, 116(D8), 353-366.

Lai, T. L., & Wei, H. R. Z. (1978). Strong consistency of least squares estimates in multiple regression. Proceedings of the National Academy of Sciences of the United States of America,75(7), 3034-3036.

Levy, R. C., L. A. Remer, S. Mattoo, E. F. Vermote, and Y. J. Kaufman. (2007). Second generation operational algorithm: retrieval of aerosol properties over land from inversion of MODIS spectral reflectance. Journal of Geophysical Research Atmospheres,112, 1-21.

Levy, R. C., S. Mattoo, L. A. Munchak, L. A. Remer, A. M. Sayer, F. Patadia, and N. C. Hsu. (2013), The Collection 6 MODIS aerosol products over land and ocean, Atmospheric Measurement Techniques, 6, 2989–3034, doi:10.5194/amt-6-2989-2013.

Li, K., Jacob, D. J., Liao, H., Shen, L., Zhang, Q., & Bates, K. H. (2019). Anthropogenic drivers of 2013–2017 trends in summer surface ozone in China. Proceedings of the National Academy of Sciences, 116(2), 422-427.

Li, Z., Guo, J., Ding, A., Liao, H., Liu, J., & Sun, Y., et al. (2017). Aerosol and boundary-layer interactions and impact on air quality. National Science Review, 4 (6), 810–833. doi: 10.1093/nsr/nwx117.

Li, Z., Niu, F., Fan, J., Liu, Y., Rosenfeld, D., & Ding, Y. (2011). Long-term impacts of aerosols on the vertical development of clouds and precipitation. Nature Geoscience, 4(12), 888–894. doi:10.1038/ngeo1313.

Lu, Z., Zhang, Q., and Streets, D. G. (2011). Sulfur dioxide and primary carbonaceous aerosol emissions in China and India, 1996–2010, Atmospheric Chemistry and Physics, 11, 9839–9864, doi:10.5194/acp-11-9839-2011, 2011.

Marbach, T., Riedi, J., Lacan, A., & Schlüssel, P. (2015). The 3MI mission: Multi-viewing-channel-polarisation imager of the EUMETSAT polar system: Second generation (EPS-SG) dedicated to aerosol and cloud monitoring. In Proceedings of SPIE (Vol. 9613, 8 pp.) San Diego, CA. https://doi.org/10.1117/12.2186978

Mehta, M., Singh, R., Singh, A., Singh, N., & Anshumali. (2016). Recent global aerosol optical depth variations and trends — a comparative study using MODIS and MISR level 3 datasets. Remote Sensing of Environment,181, 137-150.

Nabat, P., Somot, S., Mallet, M., Chiapello, I., Morcrette, J. J., & Solmon, F., et al. (2013). A 4-d climatology (1979–2009) of the monthly tropospheric aerosol optical depth distribution over the Mediterranean region from a comparative evaluation and blending of remote sensing and model products. Atmospheric Measurement Techniques, 6(5), 1287-1314.

North, P. R. J. (2002). Estimation of aerosol opacity and land surface bidirectional reflectance from atsr-2 dual-angle imagery: operational method and validation. Journal of Geophysical Research Atmospheres,107(D12), AAC-1-AAC 4-10.

North, P. R. J., Briggs, S. A., Plummer, S. E., & Settle, J. J. (1999). Retrieval of land surface bidirectional reflectance and aerosol opacity from atsr-2 multiangle imagery. IEEE Transactions on Geoscience & Remote Sensing, 37(1), 526-537.

Pöschl, U. (2005). Atmospheric aerosols: composition, transformation, climate and health effects. Cheminform, 44(46), 7520-7540.

Poulsen, C. A., Siddans, R., Thomas, G. E., Sayer, A. M., Grainger, R. G., & Campmany, E., et al. (2012). Cloud retrievals from satellite data using optimal estimation: evaluation and application to ATSR. Atmospheric Measurement Techniques, 5(8), 1889-1910.

Ramanathan, V., Crutzen, P. J., Kiehl, J. T., and Rosenfeld, D.: Aerosols, climate and the hydrological cycle, Science, 294, 2119–2124, 2001.

Rosenfeld, D., Lohmann, U., Raga, G. B., O'Dowd, C. D., Kulmala, M., Fuzzi, S., Reissell, A., and Andreae, M. O.: Flood or drought. (2008). How do aerosols affect precipitation?, Science, 321, 1309–1313. doi:10.1126/science.1160606

Sayer, A. M., Hsu, N. C., Bettenhausen, C., Ahmad, Z., Holben, B. N., & Smirnov, A., et al. (2012). SeaWiFS ocean aerosol retrieval (SOAR): algorithm, validation, and comparison with other data sets. Journal of Geophysical Research Atmospheres, 117, D03206. doi:10.1029/2011JD016599.

Sayer, A. M., Munchak, L. A., Hsu, N. C., Levy, R. C., Bettenhausen, C., & M.-J. Jeong. (2015). Modis collection 6 aerosol products: comparison between Aqua's e-Deep Blue, Dark Target, and 'merged' datasets, and usage recommendations. Journal of Geophysical Research Atmospheres, 119(24), 13965-13989.

Sayer, A. M., Hsu, N. C., Lee, J., Kim, W. V., Dubovik, O., Dutcher, S. T. et al. (2018). Validation of SOAR VIIRS over-water aerosol retrievals and context within the global satellite aerosol data record. Journal of Geophysical Research: Atmospheres, 123, 13,496–13,526. https://doi.org/10.1029/2018JD029465

Sayer, A. M., N. C. Hsu, J. Lee, N. Carletta, S.-H. Chen, and A. Smirnov (2017), Evaluation of NASA Deep Blue/SOAR aerosol retrieval algorithms applied to AVHRR measurements, Journal of Geophysical Research: Atmospheres, 122, doi:10.1002/2017JD026934

Sayer, A., Hsu, N., Lee, J., Kim, W., Dubovik, O., Dutcher, S., Huang, D., Litvinov, P., Lyapustin, A., Tackett, J., Winker, D. (2018). Validation of SOAR VIIRS Over-Water Aerosol Retrievals and Context Within the Global Satellite Aerosol Data Record. Journal of Geophysical Research: Atmospheres, 123, 13496–13526.

Sayer, A.M., Poulsen, C. A., Arnold, C., Campmany, E., Dean, S., Ewen, G. B.L., et al. (2011). Global retrieval of ATSR cloud parameters and evaluation (GRAPE): dataset assessment. Atmospheric Chemistry and Physics, 11, 3913–3936. http://dx.doi.org/10.5194/acp-11-3913-2011

Smirnov, A., Holben, B. N., Eck, T. F., Dubovik, O., & Slutsker, I. (2000). Cloud-screening and quality control algorithms for the AERONET database. Remote Sensing of Environment, 73(3), 337–349.

Smirnov, A., Holben, B. N., Slutsker, I., Giles, D. M., McClain, C. R., Eck, T. F., et al. (2009). Maritime Aerosol Network as a component of Aerosol Robotic Network. Journal of Geophysical Research, 112, D06204. https://doi.org/10.1029/2008JD011257

Sogacheva, L., de Leeuw, G., Rodriguez, E., Kolmonen, P., Georgoulias, A. K., Alexandri, G., Kourtidis, K., Proestakis, E., Marinou, E., Amiridis, V., Xue, Y., and van der A, R. J.: Spatial and seasonal variations of aerosols over China from two decades of multi-satellite observations – Part 1: ATSR (1995–2011) and MODIS C6.1 (2000–2017), Atmospheric Chemistry and Physics, 18, 11389-11407.

Suresh, B. S., Manoj, M. R. , Krishna, M. K. , Gogoi, M. M. , Nair, V. S. , & Kumar, K. S., et al. (2013). Trends in aerosol optical depth over Indian region: potential causes and impact indicators. Journal of Geophysical Research Atmospheres, 118(20), 11,794–11,806.

Thomas, G. E., Poulsen, C. A., Sayer, A.M., Marsh, S. H., Dean, S. M., Carboni, E., et al. (2009). The GRAPE aerosol retrieval algorithm. Atmospheric Measurement Techniques, 2, 679–701. http://dx.doi.org/10.5194/amt-2-679-2009.

Veefkind, J.P., de Leeuw, G., and, Durkee, P.A. (1998a). Retrieval of aerosol optical depth over land using two-angle view satellite radiometry during TARFOX. Geophysical Research Letters. 25(16), 3135-3138.

Veefkind, J.P. and de Leeuw, G. (1998b). A new algorithm to determine the spectral aerosol optical depth from satellite radiometer measurements. Journal of Aerosol Sciences, 29, 1237-1248.

Weatherhead, C., Reinsel, C., Tiao, C., Meng, L., Choi, S., & Cheang, K., et al. (1998). Factors affecting the detection of trends: statistical considerations and applications to environmental data. Journal of Geophysical Research Atmospheres, 103(15), 1241-1255.

Wei, J., Li, Z., Peng, Y., and Sun, L. (2019). MODIS Collection 6.1 aerosol optical depth products over land and ocean: validation and comparison. Atmospheric Environment, 201, 428-440. https://doi.org/10.1016/j.atmosenv.2018.12.004

Wei, J., Sun, L., Peng, Y., Wang, L., Zhang, Z., Bilal, M., & Ma, Y. (2018). An improved high-spatial-resolution aerosol retrieval algorithm for MODIS images over land. Journal of Geophysical Research: Atmospheres, 123, 12,291–12,307. https:// doi.org/10.1029/2017JD027795

Witek, M. L., Garay, M. J., Diner, D. J., Bull, M. A., and Seidel, F. C. (2018). New approach to the retrieval of AOD and its uncertainty from MISR observations over dark water, Atmospheric Measurement Techniques, 11, 429-439, https://doi.org/10.5194/amt-11-429-2018.

Wilks, & D., S. (2006). On “field significance” and the false discovery rate. Journal of Applied Meteorology and Climatology, 45(9), 1181-1189.

Zdaniuk, B. (2014). Ordinary Least-Squares (OLS) Model. Springer Netherlands.

Zhao, X.-P, T., Chan, & Pui, K. (2013). A global survey of the effect of cloud contamination on the aerosol optical thickness and its long-term trend derived from operational AVHRR satellite observations. Journal of Geophysical Research Atmospheres, 118(7), 2849-2857.

Table 1. Summary of satellite-derived and ground-observed monthly aerosol products used in this study

Product	Version	Spatial resolution	Temporal Resolution	Temporal availability	Scientific Data Set	Literature
AATSR-ADV	V2.31	1°×1°	Monthly	2002.05-2012.04	AOD550_mean	Veefkind et al., 1998a, Veefkind and de Leeuw, 1998b
AATSR-SU	V4.3	1°×1°	Monthly	2002.05-2012.04	AOD550_mean	North, 1999; 2002; Bevan et al., 2012
AATSR-ORAC	V4.01	1°×1°	Monthly	2002.07-2012.04	AOD550_mean	Thomas et al., 2009; Sayer et al., 2011; Poulsen et al., 2012
AATSR-EN	V2.6	1°×1°	Monthly	2002.07-2012.04	AOD550	Holzer-Popp et al., 2013
MISR	V23	0.5°×0.5°	Monthly	2000.03-2017.12	Optical depth average (550 nm)	Garay et al., 2017; Witek et al., 2018
MOD08	C6.1	1°×1°	Monthly	2000.03-2017.12	AOD_550_Dark_Target_Deep_Blue_Combined_Mean	Sayer et al., 2014
MYD08	C6.1	1°×1°	Monthly	2002.07-2017.12	AOD_550_Dark_Target_Deep_Blue_Combined_Mean	Sayer et al., 2014
SeaWiFS	V4	0.5°×0.5°	Monthly	1997.09-2010.12	aerosol_optical_thickness_550_land	Hsu et al., 2013; Sayer et al., 2012
AVHRR (NOAA-18)	V1	1°×1°	Monthly	2006.01-2011.12	aerosol_optical_thickness_550_land_ocean_mean	Hsu et al., 2017
VIIRS	V1	1°×1°	Monthly	2012.03-2017.12	Aerosol_Optical_Thickness_550_Land_Ocean_Mean	Hsu et al., 2013; Sayer et al., 2012
POLDER	V1.1	1°×1°	Monthly	2005.03-2013.10	AOD550	Dubovik et al., 2011, 2014
AERONET	V3	site	Monthly	2003.01-2010.12	AOD	Giles et al., 2019

Table 2. Comparison between satellite-derived and ground-based monthly AOD_S values during 2006-2010 over land and ocean

Products	Land					Ocean				
Metrics	N	R	MAE	RMSE	RMB	N	R	MAE	RMSE	RMB
AATSR-ADV	6979	0.734	0.086	0.153	0.868	959	0.712	0.068	0.100	1.149
AATSR-EN	7739	0.745	0.082	0.140	0.941	1023	0.711	0.061	0.093	1.063
AATSR-ORAC	8401	0.713	0.081	0.143	0.896	1066	0.696	0.069	0.100	1.204
AATSR-SU	7503	0.766	0.081	0.140	0.997	1026	0.693	0.058	0.098	0.988
AVHRR	7331	0.743	0.082	0.152	0.970	1051	0.783	0.047	0.078	1.004
MISR	7464	0.795	0.074	0.128	0.869	954	0.587	0.070	0.118	0.977
MOD08	8108	0.875	0.074	0.113	1.162	1088	0.814	0.069	0.093	1.309
MYD08	7945	0.870	0.068	0.110	1.090	1088	0.812	0.056	0.082	1.191
POLDER	7956	0.733	0.108	0.162	1.292	1027	0.694	0.071	0.110	1.247
SeaWiFS	4516	0.819	0.072	0.117	0.920	775	0.746	0.057	0.088	1.053

Table 3. Percentage of sites within certain ranges of evaluation metrics for different satellite-derived monthly AOD_S products from 2006 to 2010

Products	N	MAE		RMSE		RMB		
	> 6	< 0.08	> 0.12	< 0.08	> 0.12	< 0.8	[0.9, 1.1]	> 1.2
AATSR-ADV	92	59	19	47	28	35	22	16
AATSR-EN	96	66	16	55	25	26	28	23
AATSR-ORAC	99	69	20	56	28	26	24	30
AATSR-SU	95	63	19	56	28	18	32	17
AVHRR	96	67	17	57	25	20	29	18
MISR	95	69	15	50	25	25	30	23
MOD08	99	67	12	52	23	9	14	54
MYD08	97	71	12	60	21	12	24	34
POLDER	93	35	31	21	47	2	17	61
SeaWiFS	79	56	14	46	21	12	27	24

Table 4. Seasonal statistics of spatial coverage and global means of satellite-derived AOD_S from 2003 to 2010

Products	Spatial coverage (%)				Mean AOD			
	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON
AATSR-ADV	54	66	67	62	0.16±0.10	0.17±0.13	0.17±0.12	0.16±0.10
AATSR-EN	68	75	77	75	0.13±0.08	0.16±0.13	0.15±0.11	0.14±0.09
AATSR-ORAC	75	76	80	79	0.15±0.09	0.16±0.10	0.16±0.10	0.16±0.08
AATSR-SU	70	77	79	78	0.12±0.09	0.15±0.14	0.15±0.13	0.13±0.09
AVHRR	69	74	76	73	0.13±0.09	0.14±0.14	0.14±0.13	0.13±0.09
MISR	73	77	79	78	0.12±0.08	0.13±0.12	0.14±0.11	0.12±0.08
MOD08	71	79	82	80	0.16±0.09	0.19±0.14	0.19±0.13	0.17±0.10
MYD08	71	79	82	80	0.15±0.09	0.17±0.14	0.17±0.12	0.15±0.09
POLDER	64	63	66	66	0.19±0.13	0.20±0.15	0.21±0.15	0.19±0.12
SeaWiFS	65	71	72	72	0.10±0.08	0.12±0.11	0.13±0.12	0.11±0.08

Table 5. Regional trends and uncertainties of the Terra MODIS AOD_S anomalies from 2000 to 2017, where ** and * indicate the trends significant at the 95% and 90% confidence levels, respectively.

Land		ENA		WNA		SAM	
Trend	Uncertainty	Trend	Uncertainty	Trend	Uncertainty	Trend	Uncertainty
0.0001	0.0004	-0.0009**	0.0004	0.0005	0.0006	-0.0009	0.0010
EUR		AFR		ME		EAA	
Trend	Uncertainty	Trend	Uncertainty	Trend	Uncertainty	Trend	Uncertainty
-0.0014**	0.0005	0.0002	0.0005	0.0023*	0.0012	-0.0012	0.0011
SAA		SEA		OCE		Ocean	
Trend	Uncertainty	Trend	Uncertainty	Trend	Uncertainty	Trend	Uncertainty
0.0036**	0.0010	0.0010	0.0018	0.0000	0.0002	0.0005**	0.0002
PAO		NAO		SAO		INO	
Trend	Uncertainty	Trend	Uncertainty	Trend	Uncertainty	Trend	Uncertainty
0.0009**	0.0002	0.0003	0.0004	0.0013**	0.0003	0.0008**	0.0002
ENC		EUC		SAC		EAC	
Trend	Uncertainty	Trend	Uncertainty	Trend	Uncertainty	Trend	Uncertainty
-0.0019**	0.0004	-0.0011**	0.0003	0.0042**	0.0008	-0.0013	0.0008

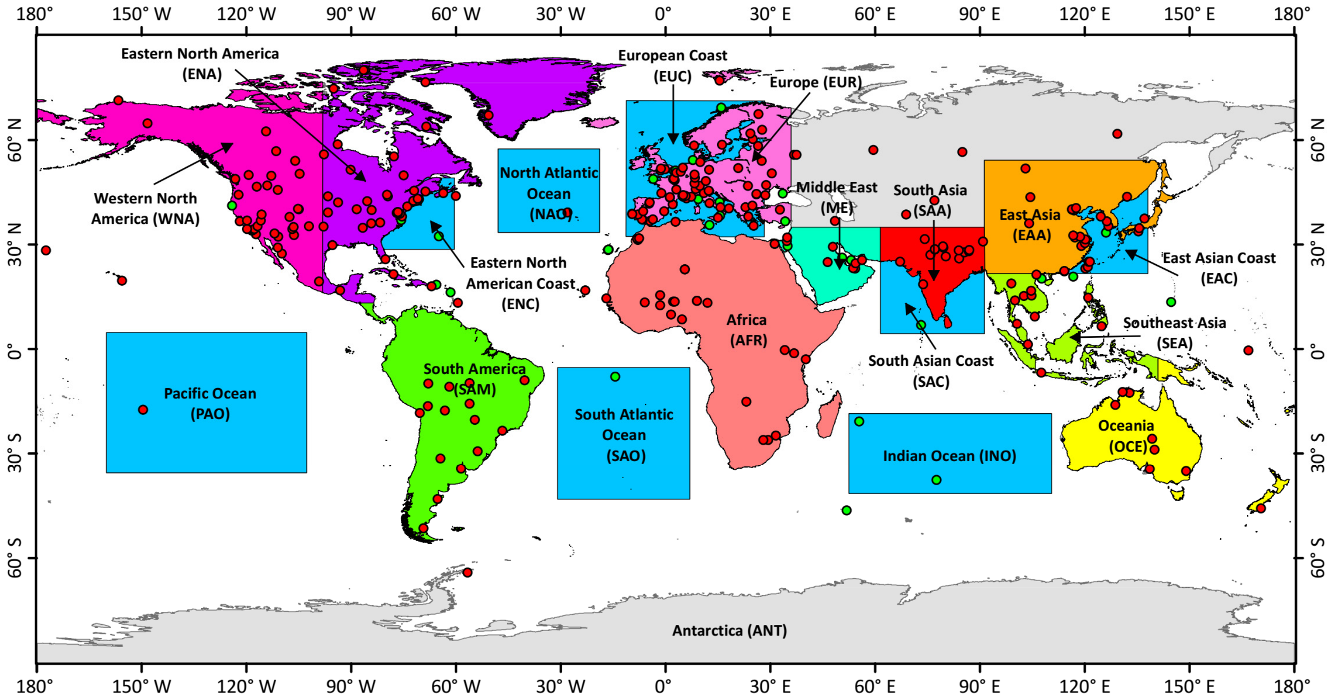
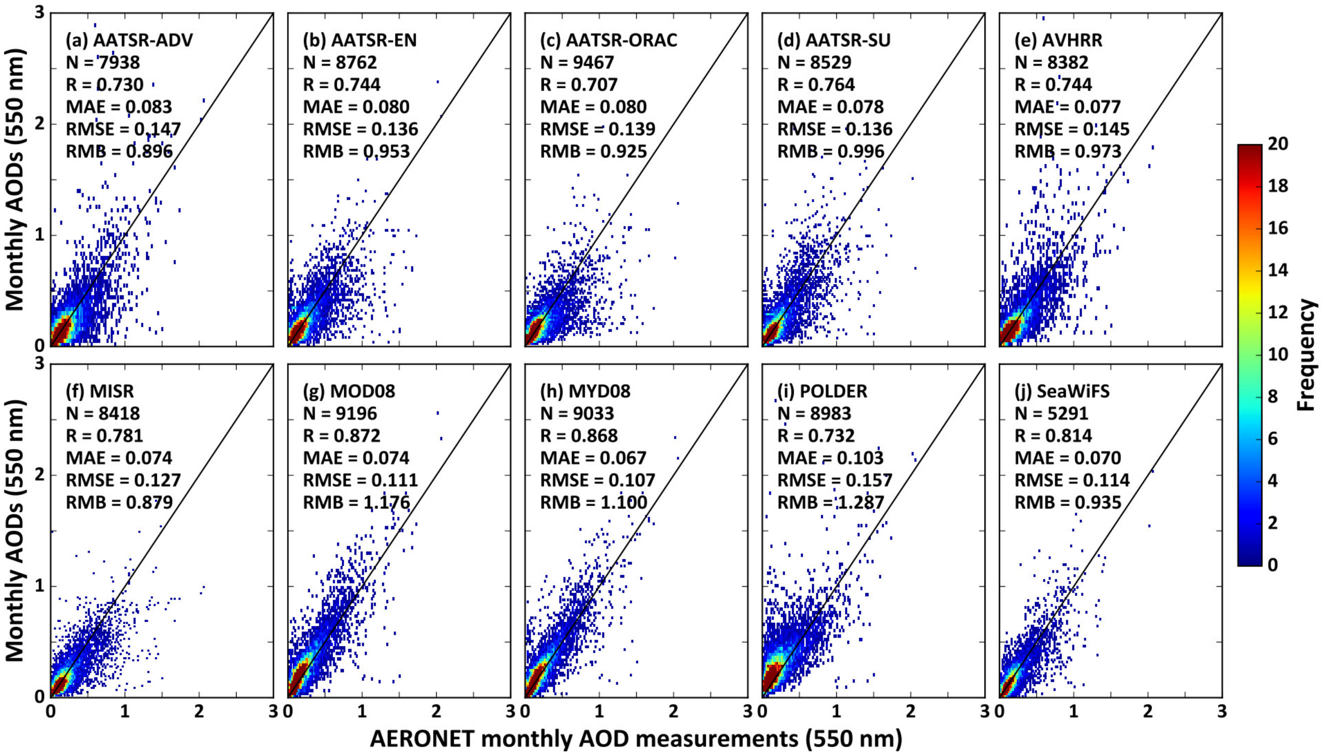


Figure 1. Locations of the AERONET sites and geographical bounds of the custom regions used in this study, where red and green dots represent land and ocean sites, respectively.



1305 Figure 2. Density scatterplots of the monthly averages of satellite-derived AOD_S versus AERONET AOD_A throughout the world

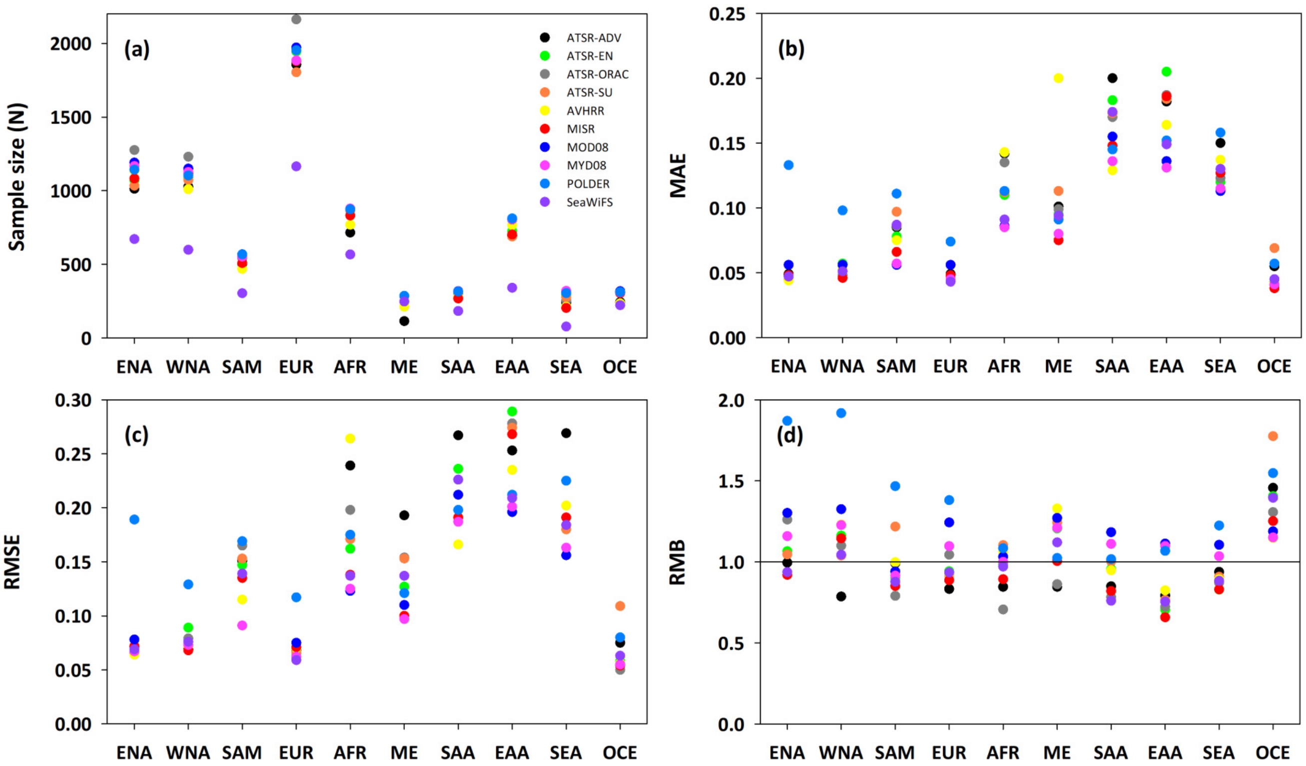


Figure 3. Continent-scale performance for satellite-derived monthly AOD_S against AERONET monthly AOD_A measurements from 2006 to 2010 in terms of (a) sample size (N), (b) MAE, (c) RMSE and (d) RMB

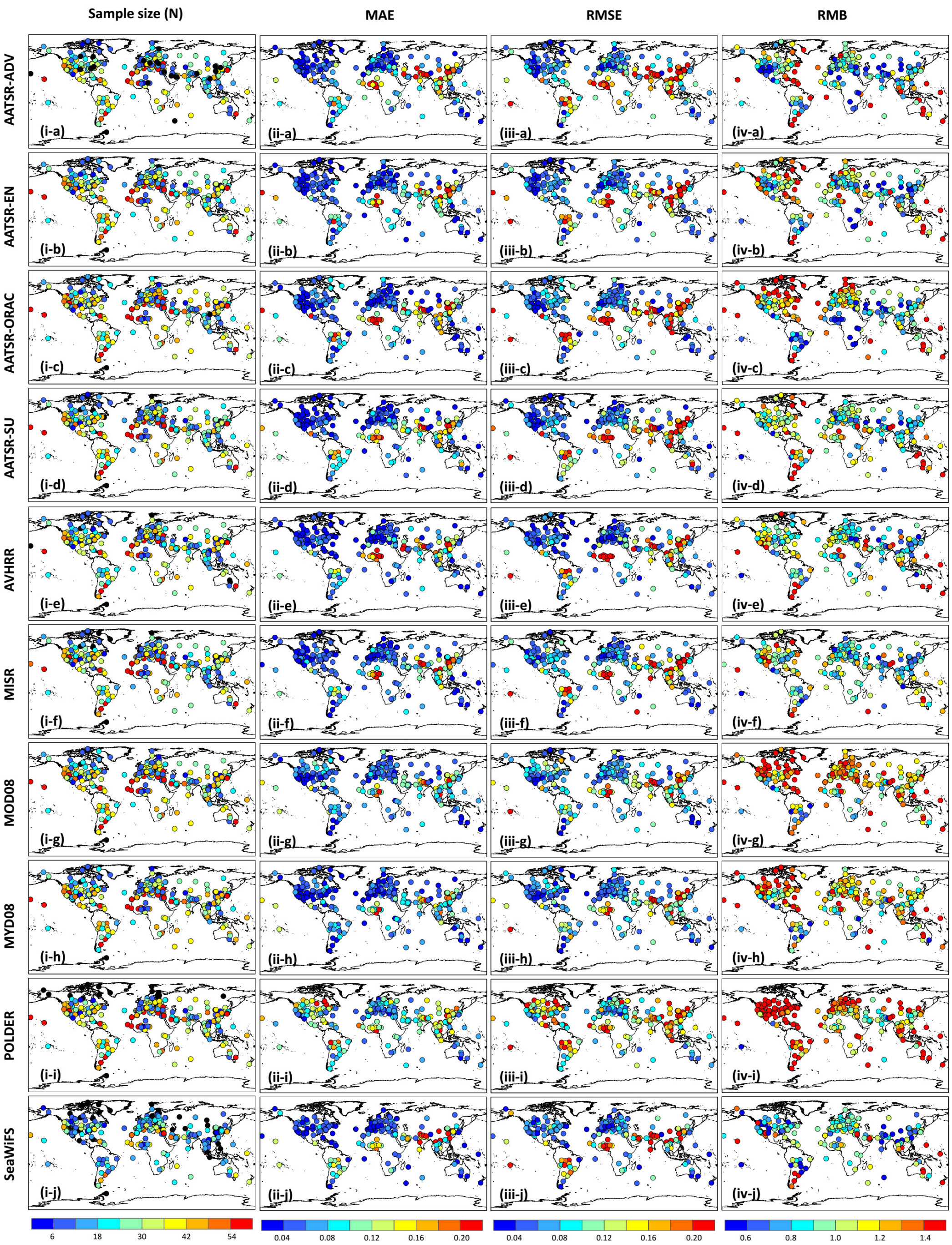


Figure 4. Site-scale performance map for satellite-derived monthly AOD_S against AERONET monthly AOD_A measurements from 2006 to 2010 in terms of (i) sample size (N), where black dots represent the sites with zero matchup samples, (ii) MAE, (iii) RMSE and (iv) RMB

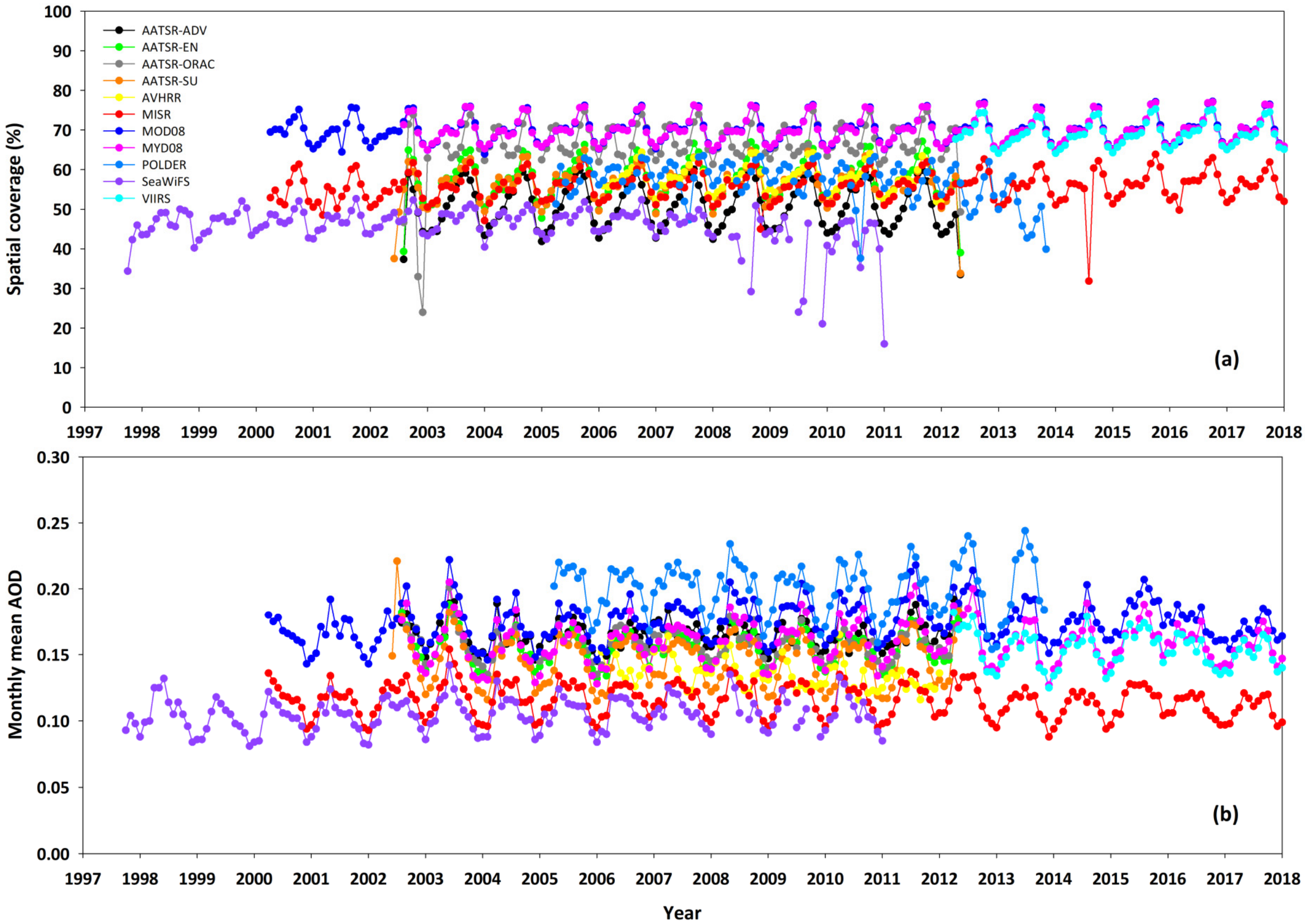


Figure 5. Time series of global spatial coverage and mean value of satellite-derived monthly aerosol products for their respective available periods from 1997-2017.

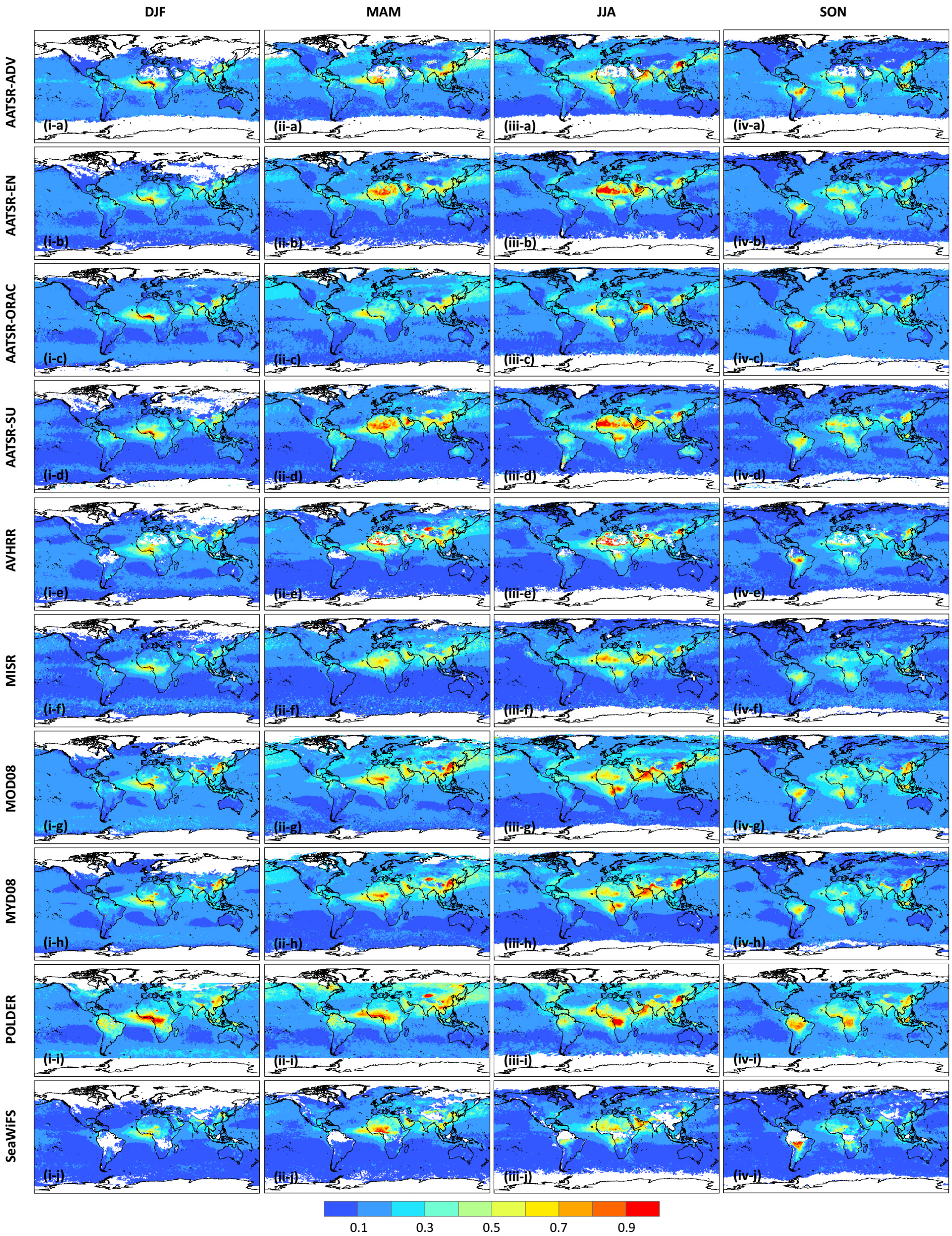
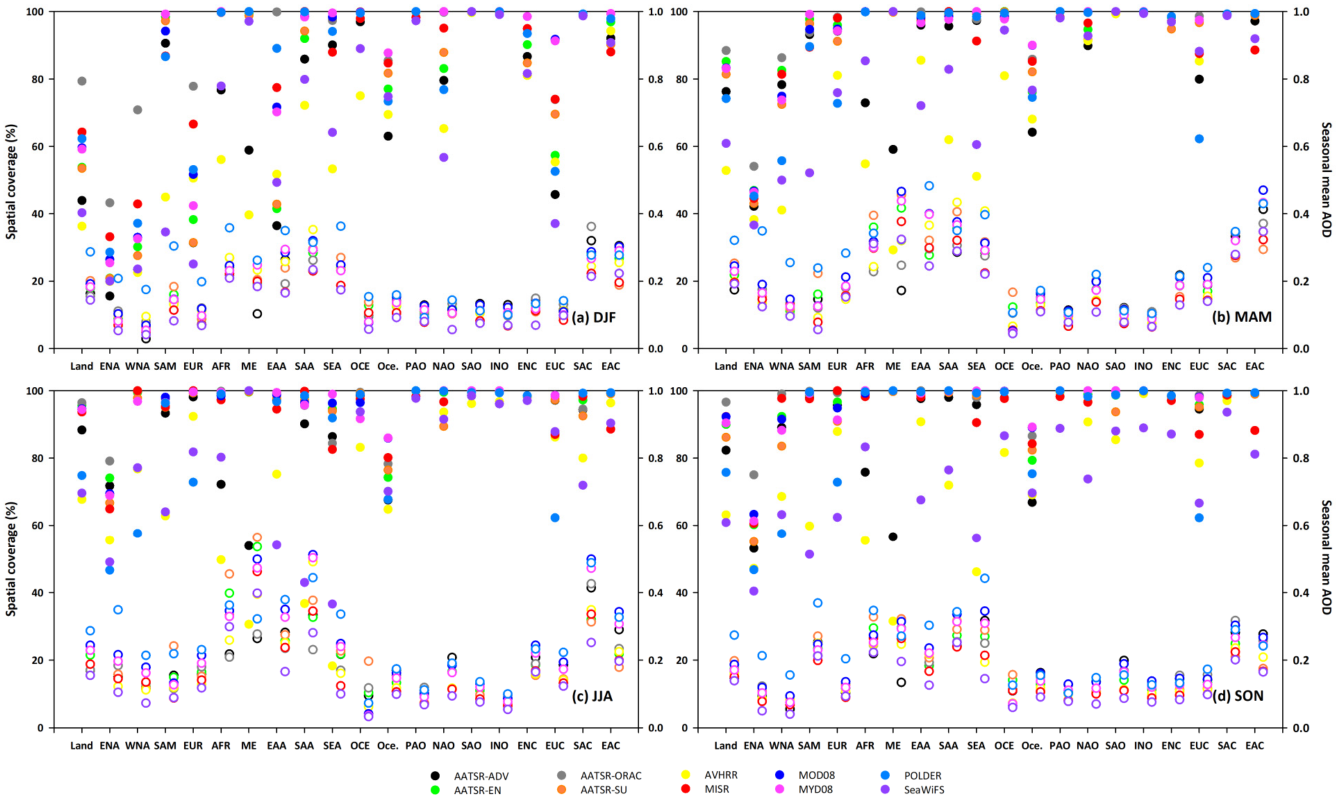
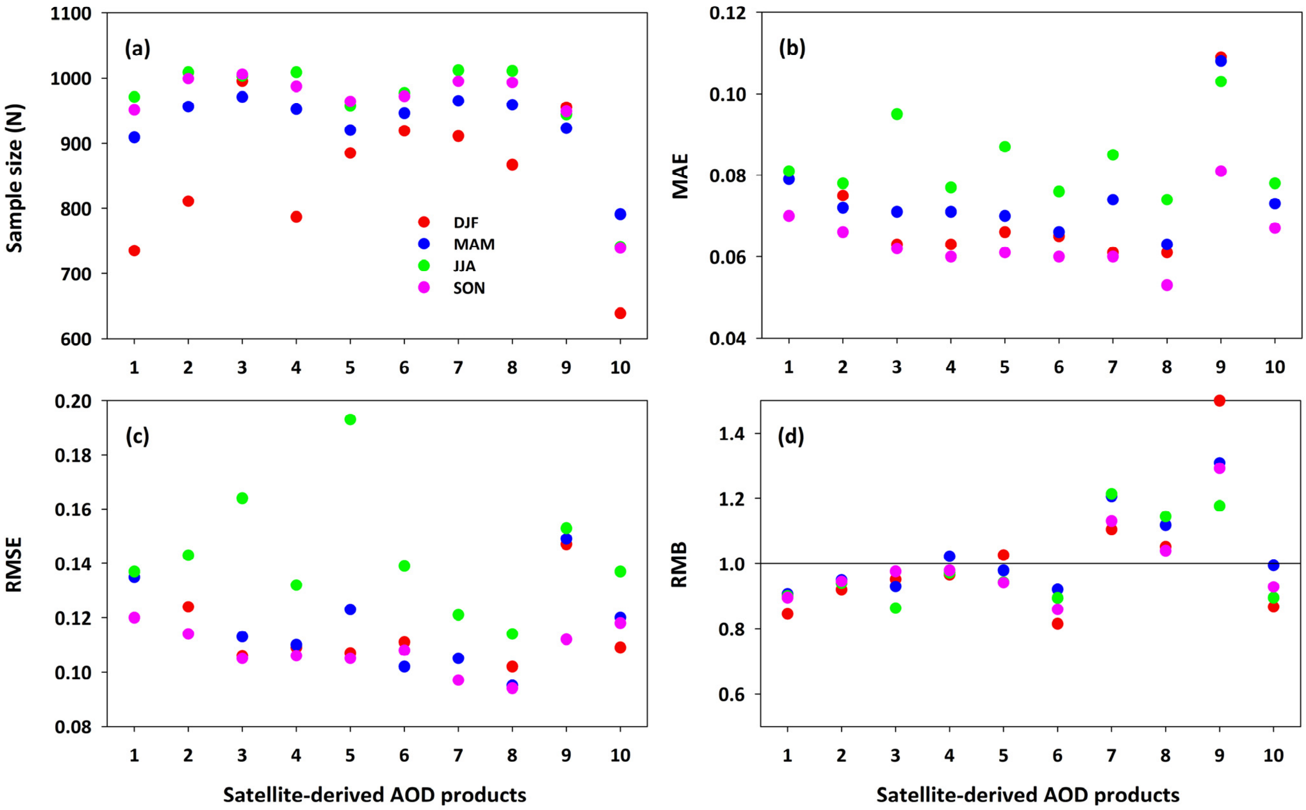


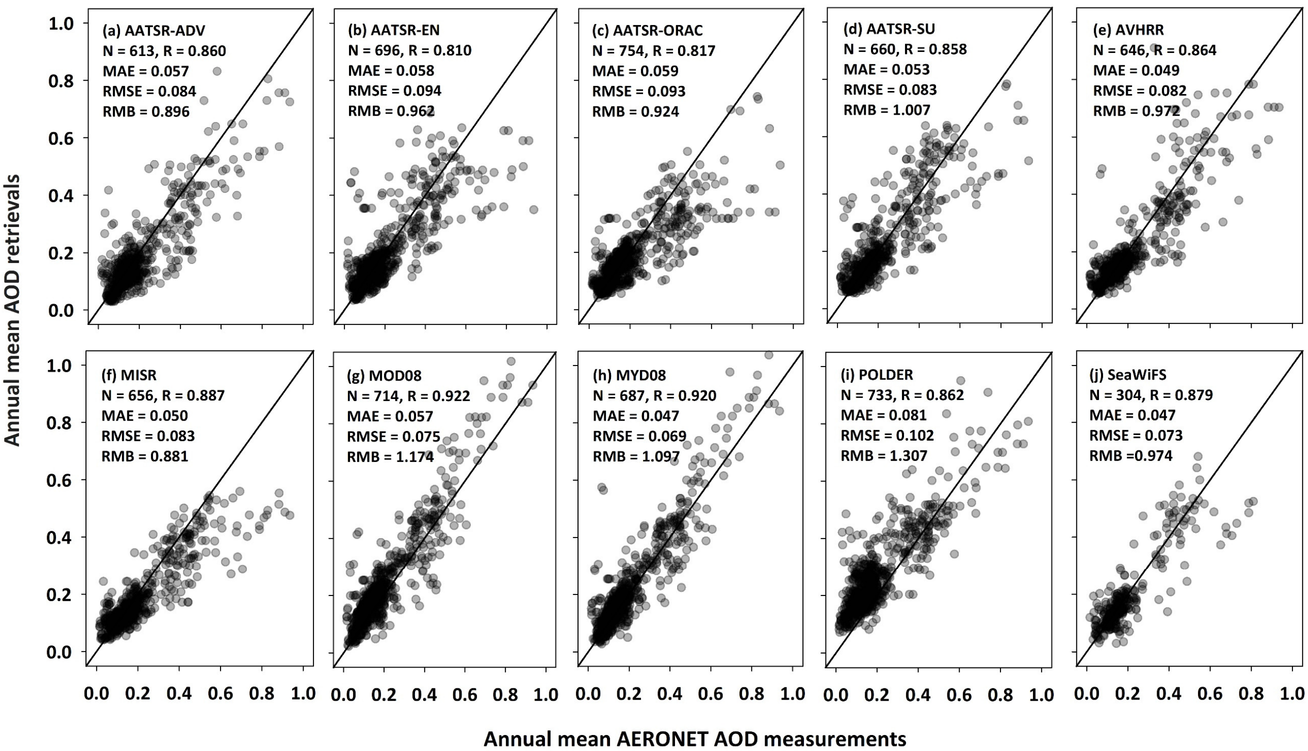
Figure 6. Satellite-derived global seasonal averaged AOD_S maps at 550 nm from 2003 to 2010



1325 Figure 7. AOD$_S$ spatial coverage (marked as solid circles) and seasonal mean (marked as hollow circles) for each customized region over land and ocean (refer to Figure 1) from 2003 to 2010.



1330 Figure 8. Seasonal performance for satellite-derived AOD_S against AERONET AOD_A measurements from 2003 to 2010 in terms of (a) sample size (N), (b) MAE, (c) RMSE and (d) RMB, where numbers 1-10 on the X-axis represent the AATSR-ADV, AATSR-EN, AATSR-ORAC, AATSR-SU, AVHRR, MISR, MOD08, MYD08, POLDER, and SeaWiFS products, respectively.



1335 Figure 9. Comparisons between the annual global mean satellite-derived AOD_S and AERONET-based AOD_A at 550 nm for all matchup sites throughout the world. The solid black line represents the 1:1 line.

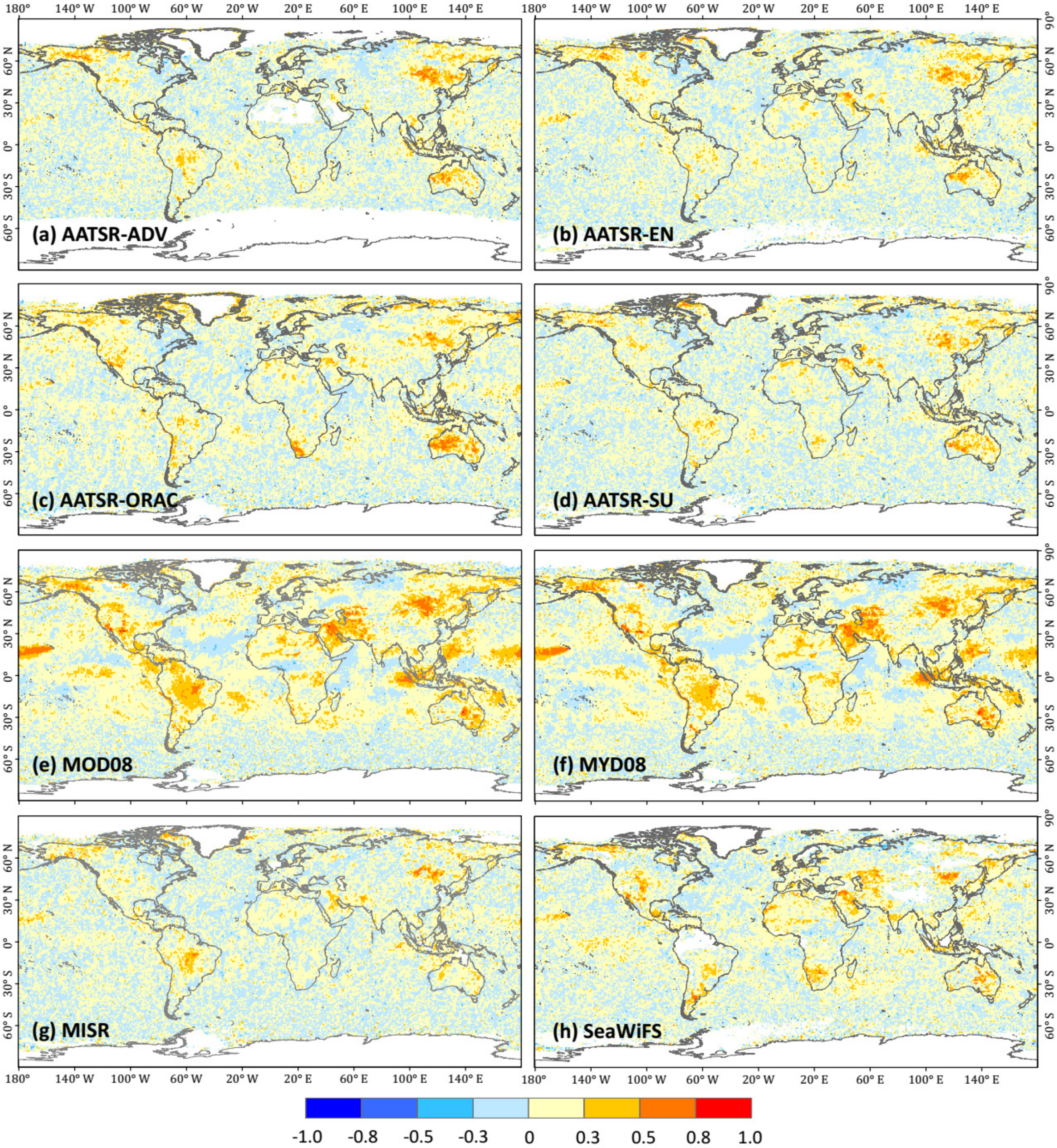


Figure 10. Spatial distribution of autocorrelation coefficient with a lag of one month based on de-seasonalized monthly AOD_S anomalies at 550 nm from 2003 to 2010.

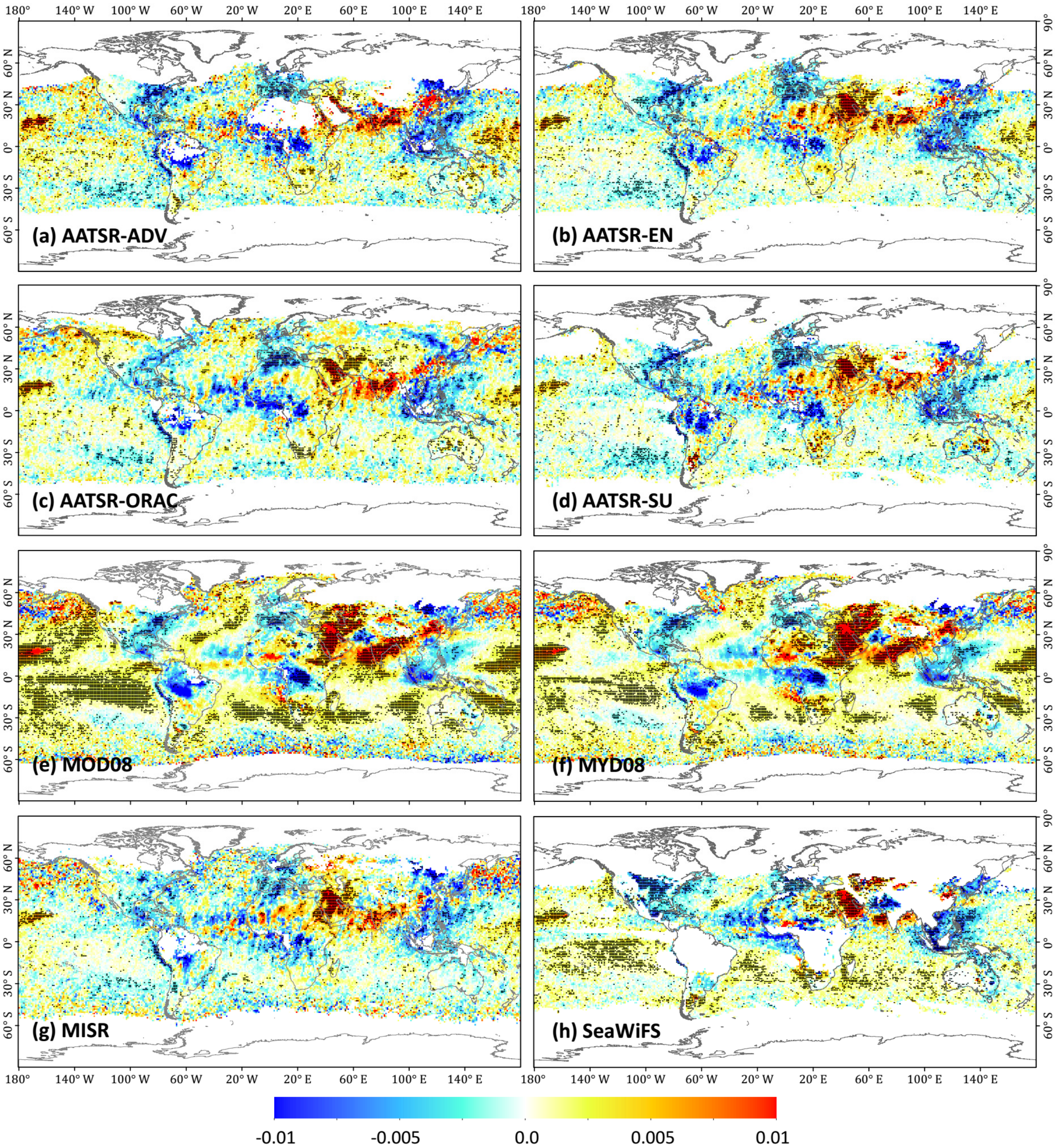


Figure 11. Linear trend based on de-seasonalized monthly AOD_S anomalies from 2003 to 2010. Units are AOD yr^{-1}. Black dots indicate a significant trend at the 95% confidence level ($p < 0.05$).

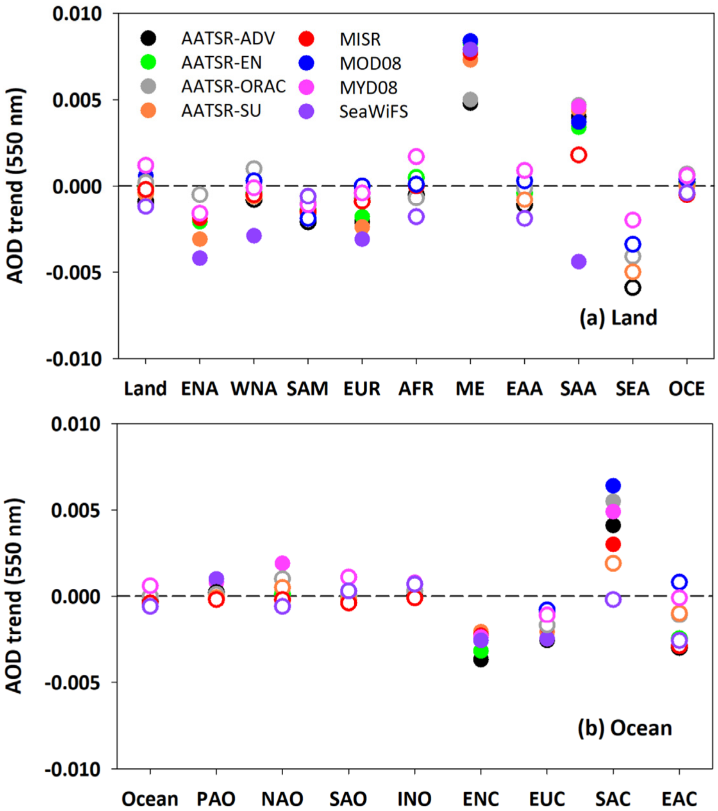


Figure 12. Regional linear trends based on de-seasonalized monthly AODs anomalies over land and ocean from 2003-2010, where the hollow and solid circles represent statistically nonsignificant and significant trends at the 95% confidence level ($p < 0.05$), respectively.

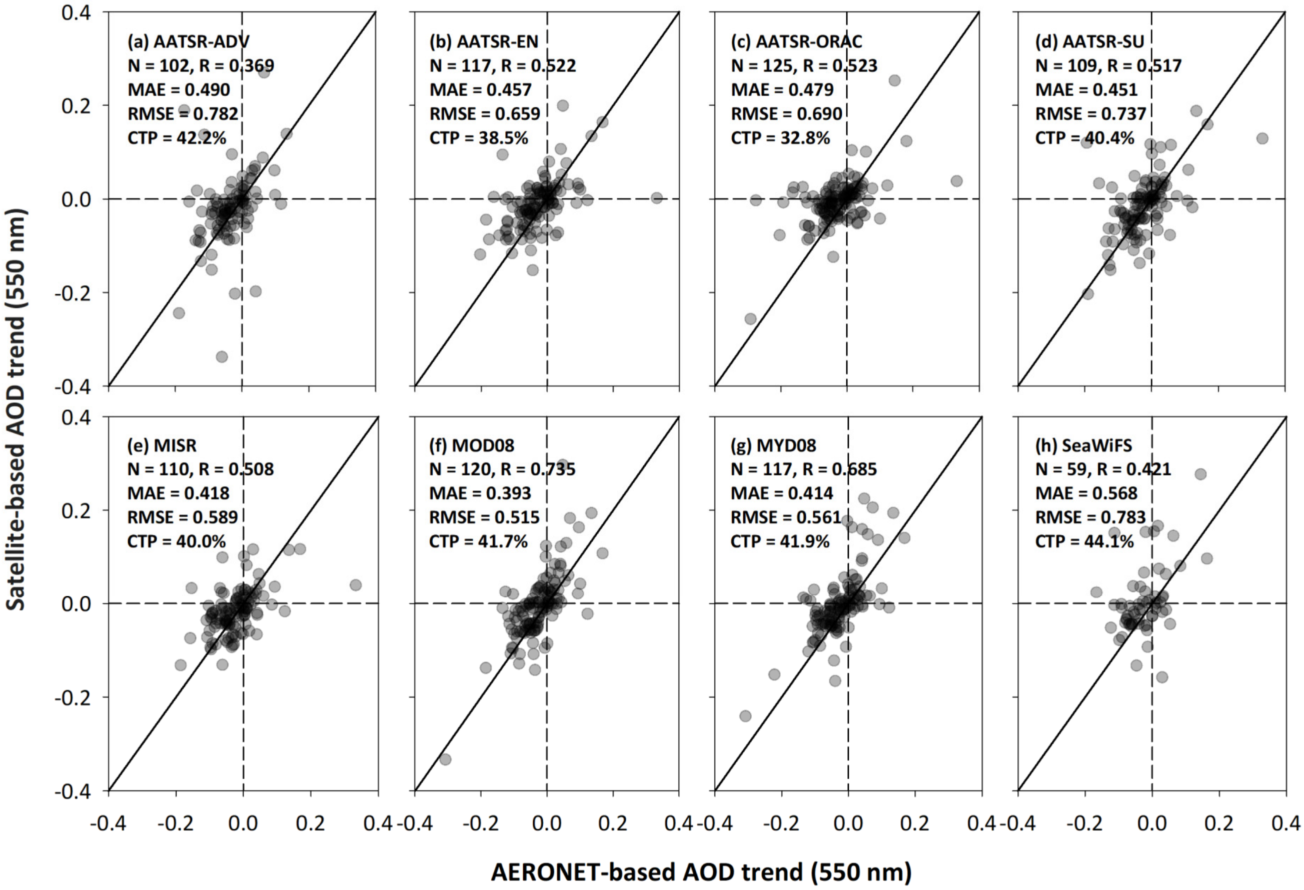


Figure 13. Comparisons between the linear trends based on the de-seasonalized monthly AOD_S anomalies from 2003-2010. Units are AOD decade^{-1}. The solid black line represents the 1:1 line.

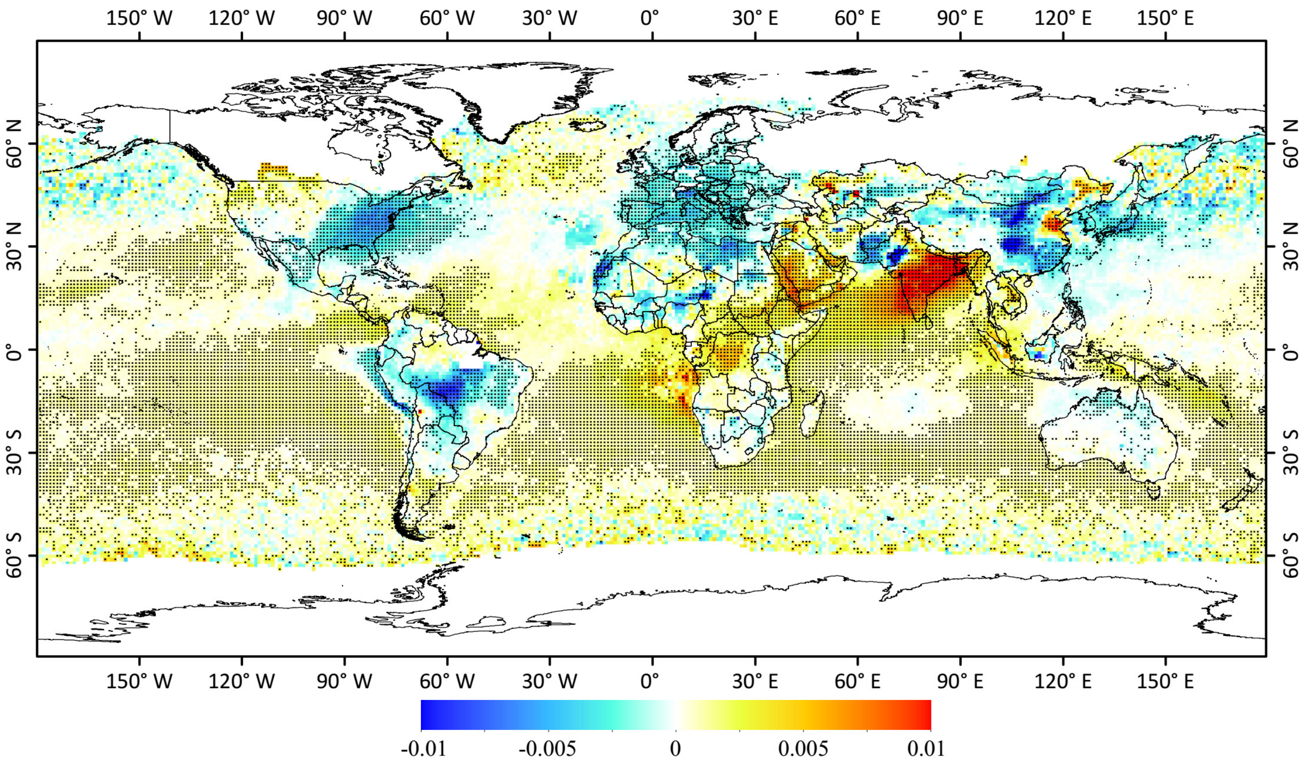


Figure 14. Linear trend based on the de-seasonalized monthly Terra MODIS AOD_S anomalies from 2000-2017. Units are AOD yr^{-1}. Black dots indicate that the trend is significant at the 95% confidence level ($p < 0.05$).