Authors' response to the reviews of: Evaluation of satellite-based aerosol datasets and the CAMS reanalysis over ocean utilizing shipborne reference observations

Jonas Witthuhn, Anja Hünerbein, Hartwig Deneke

February 7, 2020

Dear Editor and Reviewers,

We thank the editor and the three reviewers for their detailed reviews and thoughtful suggestions. We largely agree with their comments and have tried to address their concerns in the revised paper. In the following text, we give a point-by-point reply to the reviewer's comments. If changes are given in the answers with line, figure or table numbers, those numbers refer to the discussion article. Also the latexdiff file highlights the changes between the discussion article and the revised manuscript.

In order to separate the reviewer's comments and the author's response, we have printed the comments in black, and our response in blue.
We highly appreciate the detailed comments and suggestions, which have helped to improve the manuscript.

Sincerely, on behalf of all authors

Jonas Witthuhn

e-mail: jonas.witthuhn@tropos.de

Overview of changes made to the manuscript:

- Sect. 1:
  - The first three paragraphs have been merged to provide a shorter and more comprehensive introduction about why spectral aerosol observations over ocean are necessary.
- Sect. 2.1:
  - Included short description of MAN Microtops measurement protocol.
  - Included note of calibration procedures for GUVis and Microtops.
  - Included note of post processing in the GUVis introduction and restructured segments.
  - Rephrased the COMB dataset description.
- Sect. 2.2:
  - Added the wavelengths of the MODIS aerosol product.
  - Added note about referring to *MxD04* (MODIS) or *SEV_AER-OC-L2* (SEVIRI) when writing about MODIS or SEVIRI aerosol products.
  - Added a note on the increased side-scatter effect of non-spherical particles.
- Sect. 2.3:
  - Question accuracy of CAMS RA AOD under cloudy sky conditions.
- Sect. 3.1:
  - Change the thresholds of the aerosol classification method as suggested by Stefan Kinne. This effects all figures and tables related to aerosol type by changing number of datapoints but do not change the main conclusions.
  - Emphasize that the presented aerosol classification is an estimate of the dominant aerosol type of the current (mixed) aerosol situation.
- Sect. 4.2:
  - This section receives a major rework to account for changes in aerosol classification and the presentation of the statistics in Table 5., as well as to avoid several repetitions.
  - Added a note about the incompleteness of the analysis of the AOD variation between MODIS overpasses, and the additional value of high temporal resolution observations from SEVIRI, since morning and evening hours with potential aerosol growth are omitted.
- Sect. 4.3:
  - This section receives a rework due to Tables 6 and 7 are omitted or merged to Table 5 in the revised paper.
- Sect. 5:
  - Added a sentence that it has been shown, that the bias of SEVIRI AOD is dependend on AOD.

  - Reworded the paragraph about benefits of SEVIRI temporal resolution to emphasize more the targeted applications such as studies about aerosol plumes or frontal zones.

- Table 2 is updated due to the changes made to the aerosol classification.
- Table 4 is updated since outliers are no longer omitted from the calculation of GUVisE.
- Table 5 receives a major update as suggested by Stefan Kinne.
  - Results are presented for the comparison to MIC only.
  - The table focuses on 550 nm only.
  - The table includes the information about aerosol type and in case of CAMS RA additional information with and without AATSR.
- Figure 1 shows all aerosol types.
- Figure 3 shows all aerosol types.
- Figure 4 and similar Figures: added solid lines to connect the median values of each bin for clearer visualisation of the change in bias with increasing wavelength.
- All figures and tables are updated after changing the aerosol classification as suggested by Stefan Kinne.
- Minor changes and corrections to wording, grammar and typos throughout the manuscript as suggested by the reviewers.

Response to RC1 from Referee #4:

The paper discusses aerosol optical depth observations from ships. The calibration and retrieval techniques for a shadowband radiometer are revised to improve agreement with simultaneous sun-photometer measurements, removing a fractional bias. The new dataset is used to evaluate products from the MODIS and SEVIRI sensors. The former is, unsurprisingly, found to be more precise, but both overestimate the Angstrom exponent, with SEVIRI being out by up to an order of magnitude. The CAMS aerosol reanalysis is similarly evaluated, finding it eliminates much of the bias in the underlying MODIS data.
The paper is suitable for publication in this journal. Current satellite retrievals show significant disagreement as to the average AOD over remote ocean and the data provided by this study should be invaliable in resolving that discrepancy. I hope the authors can place their data in a publicy available repository — I am eager to use it in the evaluation of my own satellite products!

- Thank you for your recognition of the work in this paper. The processed data of all Polarstern cruises with the shadowband radiometer have been published at the PANAGEA data platform, where it can be freely accessed (without the need of a login): `https://doi.pangaea.de/10.1594/PANGAEA.910535`. We have also updated the link in the 'Data availability' section.

I have a few minor few comments that warrant the authors' attention:

- *Though I am fond of your analysis in Fig. 9, I disagree with the scope of your conclusions with respect to the information provided by SEVIRI.*
  - *By using the name of the sensor to refer to a specific dataset, you imply that your conclusions apply to all SEVIRI aerosol products. If one could produce a more accurate aerosol product from SEVIRI, that would provide useful information. You don't present sufficient evidence that all possible SEVIRI products provide minimal additional information.*
  - You are completely right. To be more specific, we have added the following text at the beginning of Sect.2.2 after the introduction of used satellite products:
    "In the following text unless otherwise stated, the terms "MODIS aerosol products" or "MODIS retrieval" refer to the MxD04_L2 *and* MxD04_3K *products, and similarly, the term "SEVIRI aerosol product" to the ICARE* SEV_AER-OC-L2 *aerosol product."*
  - *Your wording is fairly definitive: 'only offer minor benefits compared to the use of polar-orbiting satellite platforms'. The circumstances where aerosol changes rapidly, such as plumes or the passing of a frontal system, are scientifically very interesting and exactly the sort of circumstances that geostationary imagery are absolutely vital in understanding. Geostationary observations might not add much to our understanding of the climatology of AOD, but this doesn't mean that they only provide minor benefits; they provide targeted benefits.*
  - Thanks for pointing this out. We did not mean to disparage the value of geostationary measurements. As you have pointed out, this section was written with climatology studies in mind. To clarify this aspect, we have now reworded the paragraph in the conclusions section as follows:
    "Hence, the better time resolution of SEVIRI and other geostationary satellite sensors offers minor benefits for climatological studies compared to the use of polar-orbiting satellite platforms, given its increased uncertainties. The SEVIRI AOD product provide valuable information on the temporal evolution of AOD when the aerosol changes rapidly. Specific cases with high temporal variability are dust storms, plumes of volcanic ash or the passing of frontal systems."

  - *You only evaluate the representivity of observations between the two MODIS overpasses. This omits the periods of boundary layer growth and collapse in the morning and evening, which current polar orbiting satellites do not observe.*
  - That's indeed a limitation which deserves to be mentioned. For that reason, we have adopted your wording and added the following text at the end of Sect.4.2:
    "We are aware that the analyses presented here do not provide a complete picture of the AOD variability over the full diurnal cycle. It was only possible to analyse the variability between daytime overpasses of MODIS. Continuous evaluation of the daily cycle of AOD are only possible with geostationary satellites such as SEVIRI."

  There is no need to perform additional analysis, but your conclusions should be reworded to be clearer about their breadth.

- *I am surprised by the repeated implications that laboratory lamp calibration is inadequate. Calibration in a controlled environment is usually held up as the gold standard of observational atmospheric science. Did the authors mean to imply that such calibrations are insufficient to produce a scientifically valid product(e.g. 'limited accuracy')?*

  I would find that difficult to believe. I suspect what was meant is that there is an intrinsic difference between what a sun-photometer and shadowband radiometer measure. That limits the extent to which they could ever agree without additional correction methods, such as those outlined in this paper.

- You are right, the repeated use of the term 'lamp-based calibration' suggested that a calibration using a lamp was insufficient. This is indeed not true and not what we wanted to express in the text. Therefore we have removed the 'lamp-based' specification and now simply refer to 'calibration' in section 3.4 and 4.1. In addition, we now give a short description of the calibration procedures of GUVis and Microtops in section 2. Nevertheless, several previous studies state that Langley-calibrations have a lower uncertainty than lamp-based calibartions. Hence, we think that it would be beneficial to also calibrate the GUVis instrument using the Langley-technique since this should lower the calibration uncertainty compared to a laboratory lamp-calibration [Schmid and Wehrli, 1995, Alexandrov et al., 2002].

- *I'm not convinced by the explanation in §4.2 of the narrow, highly biased observations of AOD $\simeq$0.3 in Fig. 4 as I can't see why the choice of aerosol type would only affect one range of AODs. Are there an anomalously small number of collocations in those conditions or are they clustered in a small area? If you loosen your quality control conditions, does the distribution more closely resemble the typical behaviour?*

- Thanks to your comment, we recognize that the wording of these two paragraphs is misleading. We wanted to draw attention to the overestimation of satellite AOD during situations with low AOD values <0.4, where AOD $\simeq$0.3. Panels (a) and (b) of Figure 4 show this overestimation most prominently. We referred to aerosol type with regards to GUVisE and COMB data, since both datasets consist mainly of cases with maritime and desert dust. Therefore the overestimation of AOD on the satellite side is strongly visible in panels (a) and (b) since we attribute it to limitation in the satellite retrivals, which requires parameterization of the surface reflection properties. We replaced these two paragraphs with the following text:

  "The SEVIRI retrieval shows an even stronger tendency to overestimate AOD in comparison to the MIC reference dataset. The bias of satellite AOD also shows a dependence on the magnitude of the AOD. A positive bias (overestimation) is mostly found in situations with AOD values below 0.5, and decreases for larger AOD. This behavior is most evident in Fig. 4 panel (a)

and panel (b) as the reference datasets are GUVisE and COMB. A similar behavior also appears in the comparison to Microtops (panel (c)), although it is far less pronounced. Since the satellite instruments measure reflected radiance the reflecting properties of the ground used in the retrievals influence the retrieved AOD. Especially for clean atmosphere e.g. low AOD the influence of such parameters (e.g. surface albedo) is strong, since the values measured reflectance at TOA are close to the values of surface reflectance. For larger AOD values the uncertainty of those characterizations shrinks, therefore the overestimation of AOD decreases. Since GUVisE and COMB datasets contain more maritime and desert dust cases than the MIC, this behavior is strongly visible."

- *Are the outliers identified at page 11 (line 344) excluded from further analysis? That seems statistically suspect, as we expect large deviations to occur occasionally by random chance.*

- We indeed initially excluded outliers to produce the COMB dataset, where GUVis and MIC are combined. We saved the outliers for further analysis, but regarding the low number it does not change the results. You are right, the outliers should not be excluded for the regression to produce GUVisE. We have run the calculations again with included outliers. The results did not change much and do not affect our conclusions. You can check the differences by looking at Table 4 in the latexdiff file.

- *I found it strange that Fig. 1 implies that only maritime and dust aerosols were observed while Fig. 2 shows that mixed and continental were occasionally encountered as well.*

- In the discussion paper, we had restricted Fig.1 to show only maritime and desert dust, since the plot is already very crowded. Based on your comment, we have now updated Figure 1 to also show mixed and continental aerosol.

- *Fig. 7 is a compelling way to present the limitations in the retrieval of Angstrom exponent. In a future paper, it would be interesting to see a study of the implications of your results on the Aerosol Index, which is widely used as a proxy for cloud condensation nuclei in studies of aerosol-cloud interactions.*

- Thank you for this suggestion, a study considering the uncertainty of the AI would indeed be interesting.

- *At L540, is an increase from 0.90 to 0.92 really evidence of a 'clearly superior' product? That doesn't seem a particularly significant shift*

- We wanted to point out that CAMS RA performance is better than the performance of the SEVIRI product in all statistical measures considered here, even if 'clearly superior' might be misinterpreted to indicate a larger difference. We have now modified this sentence to: "*The CAMS RA outperforms the SERVIRI aerosol dataset in all presented statistical measures at least slightly (e.g., correlation 0.90 versus 0.92 or LOA of 0.13 versus 0.15).*"

- *On page 21, the EarthCARE lidar isn't itself that 'unique'. It's unique that said lidar is being flown collocated with an imager and radar.*

- An additional unique aspect is the high-spectral resolution lidar, which will be able to directly measure extinction, in contrast to CALIOP. We have now reworded the sentence as follows: "*The combination of both instruments on a single satellite and the use of a high-spectral resolution lidar enabling direct observations of the aerosol extincition at 355nm is a unique feature and will benefit scientific studies targeting aerosols including their radiative effects.*"

- *Your discussion about CAMS in §4.3 would be improved if you mention that the inputs to a reanalysis system must be bias corrected before input to ensure as table assimilation of the data. Hence, the reduction in bias is to be expected (but remains evidence of the utility of the CAMS product).*

- Thank you, we have now reworded the paragraph slightly to include this statement:

  " *Further, the dependency of the bias on AOD reduced for the CAMS RA product as it shows low bias values for both low and high AOD values. This is expected since the MODIS AOD bias must be corrected before assimilation into the reanalysis product.*"

- *In Fig. 3, is the sharp transition from maritime to dust aerosol at 0.18 a true feature of your data (which would be concerning) or a feature of plotting the orange points over the blue ones? If the latter, perhaps add some transparency, so the transition is easier to see?*

- The points overlap only slightly, since the AOD on the x-axis is an average value. The sharp transition originates from the aerosol classification method we are using. Based on Toledano et al. [2007], and the suggestions of Stefan Kinne (see reviewer comment RC3) we chose the following thresholds to identify aerosol types: maritime background (AOD < 0.15), mineral dust transport (AE < 0.5, AOD > 0.15), continental transport (AE >1, AOD > 0.15), and mixed (0.5 < AE <1, AOD > 0.15) type. Therefore, the sharp transition is a result of the method. We are aware that this empirical classification of aerosol type has limitations. While a better aerosol classification based on more complex methods could be developed, we content that for the purpose of our study, this simple empirical approach is sufficient.

- *In point (ii) of the appendix, you change the method for filtering perturbed observations. What motivated this choice? In undergraduate labs, I teach my students to throw out any observation for which the method was suspect as making a correction involves a number of assumptions. Why do you feel the need to keep some corrupted observations here?*

- We found that the former filter criterion was too strict in excluding data from the whole sweep. Since we are only interested in the direct irradiance, it is sufficient to have good observations of the global irradiance before and after the sweep, and the moment when the shadowband fully shades the diffusor, and right before or after the shadow falls on the diffuser. A full sweep needs 40 seconds and during most of the time, the shadowband shades some part of the sky nowhere close to the sun. In the former (strict) filter method, the whole sweep was skipped if too large variations occured during a sweep. Now the algorithm focuses only on the relevant parts of the sweep (when the shadowband is close to the sun) to assess whether it should be skipped or not. We have found that this new filter works well for identifying perturbed scenarios while providing stable data even during short sunny periods.

I also include some technical comments and corrections. P1L2 means line 2 of page1.

- *P1L16 similar performances for both datasets*

- Done.

- *P2L39 e.g. from ships are available*

- Done.

- *P2L57 Does 'earth' need to be capitalized?*

- The AMT guidelines on capitalization leave the decision in particular for 'earth' to the authors, as long as it is consistent. So we leave it as is. Instead we corrected 'Earth' > 'earth' at L21, L31 and L33.

- *P3L79 complex, non-spherical shape*

- Done.

- *P4L109 These findings are understood in the context of the results found for the SEVIRI aerosol product to observe*

- Done.

- *P4L119 Add a space after the comma.*

- Done.

- *P4L124 are publicly available*

- Done.

- *P5L133 reference: the sunphotometer*

- Done.

- *P8L251 of ±30min have been used*

- Done.

- *P9L257 distance angle less than 0.2°*

- Done.

- *P9L272 the analyses are*

- Done.

- *P10L303 Perhaps add 'to ensure the' after 'compensated for'? It means something slightly different but is what I think you meant to say here.*

- Done, thanks for the suggestion.

- *P12L373 Add a space after 'Table'.*

- Done.

- *P16L513 The wrong style of reference is used.*

- Done.

- *P16L518 I don't know what you meant to say by 'follow up'.*

- What is meant hear are overlapping pixels of MODIS images from consecutive Terra & Aqua scans. In the text we replaced 'follow up' with this description:

  "To further investigate this point, MODIS collocations with the shipborne datasets are used to serve as random samples to study the AOD variability between successive overpasses. For each pixel of a MODIS image the corresponding SEVIRI AOD for every available SEVIRI image between overlapping MODIS images of consecutive Terra and Aqua overpasses was acquired to calculate the AOD variation."

  We did this also to other occurrences of 'follow up' in this paragraph.

- *P20L627 has channels only at*

\- Done.

- *P20L640 products can provide a*

\- Done.

- *P21L672 with the next few years collocated with an imager and radar.*

\- Done.

- *Tab.5 collocated data points. Listed*

\- Done.

- *Fig.1 The aerosol classification method*

\- Done.

- *Fig.8 requirement for simultaneous . . . figure also shows CAMS RA*

\- Done.

Response to RC2 from Referee #3:

This paper aims to evaluate the satellite (MODIS, SEVIRI) and reanalysis (CAMS) retrievals of AOD and Angstrom exponent over ocean by comparing them with moving ship-borne observations using Microtops sunphotometers and multi-spectral shadow-band radiometer GUVis-3511 during several cruises in the Atlantic Ocean. The results are re-evaluated for defined aerosol types, mostly maritime and desert dust.
Overall, the manuscript is well written and organized, although some improvements may be attained in the discussion of the results.
However, the manuscript is rather long enough and some parts may be significantly shortened without any effect in the general discussion and importance of the results, since there are several repetitions throughout the manuscript.

- *There is a rather long discussion of aerosol direct and indirect effects in the beginning of the Introduction that is beyond the scope of the current research. I understand that authors initially discuss the role of aerosols on global climate and the necessity of accurate measurements of them, in a way to reduce the uncertainty in their climate response, but this part may be shortened in one paragraph (for example the first three paragraphs could be shortened and merged into one).*

- You are right. We have now merged the three paragraphs as suggested. The text is reduced to:

  "*Aerosol particles directly influence the earth's radiation budget through their interaction with solar and terrestrial radiation, and indirectly by modifying the optical properties of clouds (Boucher2013). Studies of aerosol effects on the climate system are based on radiative transfer models. Therefore, knowledge about the spectrally resolved optical properties of different aerosol types is essential. Over ocean, sea spray (Bellouin2005, Loeb2005, Yu2006, Myhre2007) and desert dust (e.g., Tegen2003, Christopher2007, Nabat2015) are the major contributors to the direct radiative effect of aerosol. Observations of aerosol load and optical properties with global coverage are required to improve our understanding of climate-relevant aerosol processes.*"

- *Although a very analytic description is provided for satellite products, GUVis measurements, collocation procedures and so on, there is lack of information about uncertainties in the Microtops-II AOD retrievals, which may be high if the instrument is not exactly oriented to the sun's disk. Usually, 3-5 sets of measurements are taken from Microtops in order to select the best one via techniques described in previous papers (e.g. [Sharma et al., 2014]; Tiwari et al. 2018, Environ. Science Pollution Res.).*

- The Microtops data is processed within the Maritime Aerosol Network (MAN) framework within AERONET. There is a detailed description of the procedure available in [Smirnov et al., 2009], indeed the series data consists of the average of >5 consecutive scans. In the paper we simply state the uncertainty estimate of ± 0.02 and cite the general article [Smirnov et al., 2009]. For clarification, we added a short description in section 2.1.:

  "*The Microtops is a hand-held sunphotometer, which has to be pointed manually at the sun. To minimize uncertainties arise from manual pointing, more than five consecutive scans are averaged to form one measurement [Smirnov et al., 2009]. The Microtops instrument measures the incident direct normal solar irradiance with a field of view of 2.5° [Porter et al., 2001]. The MAN Microtops sunphotometers are calibrated against an AERONET master Cimel sunphotometer, which in turn is calibrated using the Langley-technique. [...] The uncertainty of Microtops AOD is estimated to be within ± 0.02 [Smirnov et al., 2009].*"

- *In addition, during the W-ICARB cruise campaign over the Bay of Bengal, there was a comparison between Microtops-II and MODIS AODs revealing a very good agreement between them, which may be mentioned in the paper and discussed against the current findings [Kharol et al., 2011]*

- Thank you for pointing out this relevant study, we now mention this reference in Sect.4.2:

  "Nevertheless, the correlations found here agree well with the findings of Levy et al. [2013] considering the MODIS C6.1 aerosol product (0.937) and the 550 nm channel. A smaller dataset of Microtops observations was compared to MODIS aerosol products by Kharol et al. [2011], where a general overestimation of AOD, and a high correlation was found similar to our results."

- *Section 4.2 is composed of numerous relatively short paragraphs, whose meanings are not so distinguishable. This creates some difficulties in reading and understanding exactly the major issue (spirit) of each paragraph. Taking also into account the several repetitions,this becomes more problematic. What I recommend is to merge the paragraphs into longer ones discussing a define issue, for example results of the presented figures and tables and/or discussion on these results.*

- We have rewritten this section and refined the structure to increase the clarity of the text.

- *Special care should be taken throughout the manuscript on avoiding several repetitions. Some of these are emphasized below.*

- As we have rewritten this section we have removed several repetitions.

Minor comments/corrections

- *Line 51: Levy et al. (2013) estimated...*

- Done.

- *Line 73: Double use of "system" at the end of this sentence does not make good sense and should be revised.*

- Done.

- *Lines 205-206. I recommend to remove this sentence from this part of the manuscript. In case the reader would expect a better accuracy from MODIS, what's the reason to read the results of this study?*

- Done.

- *Lines 362-364 and lines 380-381. These sentences are just a repetition and one should be removed.*

- We stripped the first sentence to its bare minimum rather than deleting it, since it is used to describe the expectation. The second sentence now provide the explanation for the expectation.

- *Line 447....is presented here.*

- Done.

- Repetitions:

  - *Line 476-480. Since the data...MIC data. Such statements have been repeated several times in the manuscript and may be removed or significantly shortened.*
  - *Line487. This sentence, even rephrased has been stated several times in the manuscript.*
  - *Line493-494. Since the MODIS...accurately. Similarly, this has been stated several times in the manuscript.*

- As we have reworked section 4.2 we have removed several repetitions.

• *Line 541. This emphasizes...*

- Done.

• *Line 562. A slight increase...*

- Done.

• *Line 589. This is a similar statement as in line 571.*

- We deleted the first sentence (L571).

Response to RC3 from Stefan Kinne:

Evaluation of satellite-based aerosol datasets and the CAMS reanalysis over ocean utilizing ship-borne reference observations by J. Witthuhn et al.
Positives

- *needed demonstration of (MAN) reference data over ocean*
- *efforts to evaluate more than just the aerosol column amount (AOD)*

Concerns

- *shadowband (GUV) data seem to lack processing maturity to serve as reference*
- *comparisons to two satellite data and one assimilation lack interpretation*
- *aerosol typing via AOD and Angstrom should be redone - more tailored of oceanic reg.*

General comments

The paper evaluates the aerosol properties over oceans of two satellite retrievals (MODIS, SEVIRI) and one modeling effort (CAMS) that assimilated MODIS data. The evaluation is based on matches (in time and location) to shipborne samples of the direct solar attenuation and of solar scattering during multiple latitudinal Atlantic crossings with the German POLARSTERN research vessel. The accuracy of (handheld MICROTOPS) direct solar attenuation data is out of question and serves as the main reference in the evaluation. In contrast, the usefulness of the complementary shadow-band radiometer (GUV) appears more limited (completely cloud-free situations are needed) and the retrieval algorithm to retrieve the direct solar spectral irradiance component as not reached the needed maturity (still many ad hoc adjustments are needed to match solar attenuations of simultaneous MICROTOPS measurements). Thus recent efforts to improve the shadow-band references (GUV, GUVisE) are more something for the Appendix. Focusing on the sunphotometer data there are many nice aspects addressed (although other satellite data sets and models could be included – possibly in an additional paper). I am not so happy about the chosen aerosol classification (especially since many source types like biomass are unlikely to be observed over oceans). Based on the sunphotometer data for AOD and Angstrom value (use AE only if AOD550 is larger than 0.15) I would separate into 4 categories (see comments to Figure 1) and work from there. I really like the plots that investigate biases are function of AOD. That said I am really disappointed that this detail is missed in the statistical tables, as common features of (1) AOD overestimate at lower AOD and (2) AOD under-estimate at larger AOD lead the authors to claim via a linear fit 'overall good agreement' and 'low biases'. The paper is great contribution but I think it needs a major revision prior to a publication.

Thank you for your valuable and detailed review of our manuscript. We largely agree with your suggestions, which motivated the following changes we made to the manuscript:

- Changes of the thresholds used for the aerosol classification.
- Table 5 has been changed, now covering the bias at different magnitudes of the AOD and for different aerosol classifications. Therefore, Tables 6,7 and Tables in the Appendix have been omitted.
- Section 4.2 has been modified according to the changes of the aerosol classification to present the statistics given in Table 5.

Detailed comments

- *1/17 if AE is overestimated then offer an explanation such as that large mineral dust size-events are missed or that the compositional mix is constrained by model assumptions*

- Yes, we reworded the sentence to:

  "When considering aerosol conditions, an overestimation of AE is found for scenes dominated by desert dust for MODIS and SEVIRI products versus the shipborne reference dataset. As the composition of the mixture of aerosol in satellite products is constrained by model assumptions, this highlights the importance of considering the aerosol type in evaluation studies for identifying problematic aspects."

- *1/21+ the radiative effect/forcing introduction (although I am not sure if it is needed) could/should be sharper: In terms of radiative forcing (that is the impact by anthropogenic aerosol at all-sky conditions at TOA) the indirect effect RFaci overall has larger contributions than the RFari. And anthropogenic aerosol impacts are primarily caused by extra smaller aerosol sizes so that (mostly natural components) of seasalt and dust which are mainly observed over oceans are almost irrelevant even for RFari. The impact of these two components on Raci (seasalt to increase CCN, dust to increase IN) has been demonstrated but the overall impact strength remains unclear (for more on radiative effects and forcing have a look at the MACv2 paper)*

- You are right. Since RFari and RFaci are not further considered in the manuscript, we have decided to shorten and merge the three introduction paragraphs into one, also based on the commments RC2 by reviewer #3. The new paragraph only gives a general introduction of aerosol direct and indirect radiative effects and emphasizes the need of spectral aerosol observation over ocean for climatology studies. The text now reads as follows:

  "Aerosol particles directly influence the earth's radiation budget through their interaction with solar and terrestrial radiation, and indirectly by modifying the optical properties of clouds (Boucher2013). Studies of aerosol effects on the climate system are based on radiative transfer models. Therefore, knowledge about the spectrally resolved optical properties of different aerosol types is essential. Over ocean, sea spray (Bellouin2005, Loeb2005, Yu2006, Myhre2007) and desert dust (e.g., Tegen2003, Christopher2007, Nabat2015) are the major contributors to the direct radiative effect of aerosol. Observations of aerosol load and optical properties with global coverage are required to improve our understanding of climate-relevant aerosol processes."

- *2/37+ if you mean CERES (broadband radiation) data then say so, even though then the link of aerosol retrievals (they make a compositional and size assumptions to yield AOD estimates) to broadband fluxes is not straight forward and need further rad.transfer modeling.*

- No, this statement does not refer to CERES data specifically. This paragraph is just a statement about the capability of satellite observations to provide continuous and global coverage for climatological studies. As well as the need to validate such satellite retrievals.

- *2/39 Exactly, this is why you want the ship data: to evaluate and to test (retrieval) model assumptions.*

- Yes.

- *2/55 you may also consider to evaluate to (the newly re-processed) MISR data, to VIIRS data (designated as MODIS follow on), to SLSTR (the ATSR follow-on) data and to GRASP-type (e.g. MERIS) data.*

- While we in agree in principle that it would be very interesting to study a wider range of aerosol products from different satellite instruments, we chose here to evaluate "only" two satellite products, as representative approaches of a simple (two spectral channels) and a more advanced (seven spectral channels) retrieval of a geostationary and polar orbiting satellites respectively.

- *2/76 A separate evaluation for fine (and mostly anthropogenic) and coarse-mode AOD (as offered for the MICROTOPS data via the SDA approach) - in place of an Angstromparameter evaluation - would elevate a marine aerosol retrieval or modeling evaluation: Most global models and satellite retrieval models use bi-modal (by size) scheme, while Angstrom parameters depend spectral choice and becomes extremely noisy and probably meaningless at low AOD values in already one of the two spectral bands needed.*

- We acknowledge your point here, and yes, the AE evaluation for maritime aerosol is extremely noisy. We have therefore modified the AE comparison figures, as you suggest later, to filter out situations with low AOD. We have decided to keep the AE evaluation, as for other types than maritime aerosol, it is still a widely used method to extrapolate AOD to other wavelengths.

- *5/131 I would have used a different description of the Angstrom parameter : ANG = -ln(AOD1/AOD2) / ln(wavel1/wavel2)...as the negative spectral slope in ln/ln-space*

- As the Ångström relation is commonly known, it is a matter of taste how to present equation (1). We prefer the presentation we have chosen, as it clearly shows the relation of AOD and wavelength.

- *5/138 The AOD shadow-band radiometer requires more actions (leveling corrections) and is more restrictive (completely cloud-free over the time of a scan).*

- We have added one sentence describing the shadowband radiometer post processing (e.g., leveling correction):

  "*The shadowband radiometer GUVis utilizes an entrance optic with a global field of view combined with a shadowband that performs a 180° sweep, while the global irradiance is measured at a high temporal frequency of 15 Hz. Several corrections are applied as post-processing to correct the influence of the ship motion, and to retrieve the direct spectral irradiance for later AOD calculation, as is described later. The measurement principle of the shadowband radiometer can be described as follows: While the global irradiance is observed with the shadowband in its lowest position between sweeps, the shadowband blocks a fraction of the incoming diffuse irradiance during its rotation, and will occlude the direct irradiance at a specific angle determined by instrument orientation and sun position. From the irradiance time series measured during the sweep, the global, diffuse and direct irradiance components can be inferred [Witthuhn et al., 2017].*"

  It is however not entirely true that a completely cloud-free situation is needed for a successful retrieval. The direct irradiance can be inferred from the data even if clouds are present, unless they cover or are close to the sun. So I would prefer "sun-free" instead of "cloud-free", which is a similar requirement for the Microtops measurement protocol.

- *6/180 The empirical correction is covered in detail but only applied to the shadowband(GUV) instrument (and not to the sunphotometer). Considering that the GUV match statistics is so much sparser then for the sunphotometer (with useful data evaluation data at this stage only through sunphotometer matches) so that in the end only sun-photometer data are used in the evaluation, I wonder if all that GUV detail (table and figures) is not better added to the*

Appendix. In terms of the forward scattering correction I am worried about a linear approach - which ignores a dependence on aerosol size (over oceans certainly at larger AOD) size.

- We partly agree with this point. Since we have adapted the GUVis retrieval within the scope of this study, we have to provide the details of the processing and correction. This is not necessary for Microtops, since it is already well documented. In an internal draft version we had moved all the GUVis detail to a section in the Appendix. But since one key point of the results section is the inter-comparison of both shipborne instruments, we have decided to provide the details of the correction and the cross-calibration in the main section, and only describe the processing-update in the appendix.

- *6/179 COMB is a subset of GUV, where GUV outliers (inconsistent to MIC) removed -correct?*

- Not entirely. COMB consists of the mean of the collocated GUVisE and MIC products. Outliers were removed previously, but are now included in the analysis given in the revised manuscript. To clarify this aspect, we have rephrased this paragraph to:

  "*The enhanced GUVis dataset (GUVisE) has been combined with the Microtops dataset to obtain a merged surface product, to test whether the combination can lead to further improvements in accuracy. This combined surface dataset (COMB) serves as third reference dataset for the evaluation of the satellite products. COMB consists of the mean of the collocated GUVisE and MIC AOD for this purpose. As shown in Table 2, the total amount of data points decreases to 1033 due to the collocation requirement.*"

- *7/194 mention also the AOD (I assume 550) wavelength of MODIS*

- We used the full dataset of *Effective_Optical_Depth_Average_Ocean*, so we added all seven channel wavelengths in the text:

  "*Satellite based aerosol datasets over ocean considered here are obtained from both the MODIS and SEVIRI satellite instruments. The MODIS Collection 6.1 (C6.1) level-2 aerosol products* MxD04_L2 *[Levy et al., 2015] and* MxD04_3K *[Remer et al., 2013] are used from both the Terra and Aqua satellites. This dataset includes the AOD at 470, 550, 660, 860, 1240, 1630 and 2130 nm.* "

- *7/208 there are three major solar retrieval problems for dust outflow:*

  (1) dust is solar spectrally flat (like clouds) and large AOD events may be removed as clouds -> retrieved dust AOD too low

  (2) dust size is underestimated and with it the size related 'dust absorption' (with low SSA values) -> retrieval AOD will be too high

  (3) non-sphericity has increased side-scatter so that a retrieval model with coarse-mode and fine-mode spheres will interpret the extra side-scatter artificially with extra fine-mode spheres→wrong composition and likely size underestimate (see under 2).

- Thank you for this insightful comment. As we don't evaluate the cloud masking of the satellites, the key property of desert dust influencing a valid AOD measurement is its non-sphericity, as it is stated in the text. We rephrased this sentence to mention the increased side-scatter as the result of the non-sphericity:

  " *A degraded accuracy for aerosol properties in the presence of desert dust in both satellite products is expected, since dust particles are non-spherical. This leads to an increased side-scatter effect compared to spherical particles which are assumed in both retrievals.*"

- *7/219...but cloud coverage may be wrong in global modeling...aside that the AOD might be different in clear-sky and cloud-sky regions of modeling*

\- Yes, therefore we added a sentence to this paragraph:

"*The advantage of utilizing CAMS RA over satellite observations is the availability of aerosol properties independent of factors such as cloud coverage or satellite orbit. Albeit the accuracy of AOD under cloudy sky conditions in the model might be questionable.*"

- *7/226 in CAMS only AOD is assimilated and in its bias correction the compositional mixture of the forecast model is maintained...so the composition may get worse (in comparison to a forecast without assimilations). By the way, the use of the (relatively sparse) ATSR data added little in a combined (ATSR & MODIS) assimilation - as the volume of MODIS data dominated*

\- This is true, we found only slight differences in the comparison MIC & CAMS with and without AATSR (see Sect.4.3). We slightly rephrased the sentence regarding assimilation of MODIS products to emphasize, that only AOD is assimilated:

"*It relies on the assimilation of global observational datasets into the Integrated Forecast System (IFS) from various satellites to provide a global picture. In terms of aerosol properties, the AOD from the products of the MODIS C6 from both Terra and Aqua are assimilated, while the composition mixture is maintained as given from the IFS.*"

- Comments on the aerosol classification
  - *8/228 the offered aerosol classification (based on AOD and ANG) is rather general and prone to failure if one of the two needed AOD values becomes low. For those events any ANG value becomes meaningless. And with respect to aerosol typing aerosol is always a mixture of many components. And unless surface winds are very calm, even marine aerosol has Angstrom parameters between 0.7 and 1.0. An ideal thing would be if simultaneous info on aerosol absorption would be available (as demonstrated in MACv2). To extract info an components, I would start with an AOD separation into fine-mode (r<0.5um) and coarse-mode (r>0.5um) contributions via the fine-mode AOD fraction (easily derived from the AOD spectral dependence) to separate seasalt/dust from pollution/wildfire aerosol (in a quantitative way!). The MAN data-base (where MICROTOPS data are stored) offers such derived (fine-mode-fraction) estimates!*
  - *8/240 aerosol is always a mixture of many components and what the simple 'Toledano' scheme does is simply identifying a likely dominant component in a qualitative way. I disagree though that marine aerosol has Angstrom parameters larger than 1.2 unless there are contributions from pollution and or biomass burning. AE values larger than 1.2 at should be linked fine-mode aerosol with contribution from sulfate / nitrate (less absorption), pollution and aged biomass burning (medium absorption) and fresh biomass burning (highly absorption and AE >1.5).*
  - *8/245 all classes are 'mixed'...but in terms of the mixed category here, how useful is a 'mixed' category if it could be a 10/90% or a 90/10% mixture?*

\- You are right, the used aerosol classification is very simple and serves only as a rough estimate of the dominant type of aerosol. Of course all classifications are mixed to some extent, which implies that the mixed category has no obvious dominant aerosol type and could therefore also be named "unclassified". We rephrased and added a sentence to this paragraph:

"*The pair of AOD and AE values is checked against empirical thresholds to identify the dominant aerosol type of the current situation as being one of maritime background (AOD < 0.15),* *mineral dust transport (AE < 0.5, AOD > 0.15), continental transport (AE >1, AOD > 0.15), or mixed (0.5 < AE <1, AOD > 0.15) type. It should be noted, that all categories are expected to cover mixed aerosol types to some extent. Therefore, the mixed category consists of a mixture of aerosol without a dominant type."*

For the purpose of this study, the presented aerosol classification enables the investigation of the performance of the satellite and CAMS RA products with regard to the predominant aerosol situation such as maritime or desert dust. As a future improvement, benefits of using a more complex aerosol classification including absorption, fine- and coarse mode aerosol should be evaluated. As you have suggested also in later comments, we have restricted the AE validation to AOD(870nm)>0.05, and changed the thresholds of the aerosol classification based on your recommendation.

• *8/230 the AE overestimate in CAMS can be just related to the improper component mixture (too much fine-mode contributions) in the forecast model, which an (total) AOD assimilation cannot fix .. and in case of AOD adjustment during the assimilation process will pull away from likely expectations (such that mix of assimilated products is often worse than that of the forecast model)*

- We have rewritten this paragraph taking into account your comment:

  "A first validation presented within Inness et al. [2019] emphasizes the high quality of AOD in the CAMS RA system, as judged by a comparison to AERONET stations around the world. However, an overestimate of AE was shown during desert dust events, and was attributed to the fixed component mixture (e.g., less dust in CAMS RA) in the forecast model. Further evaluation with a focus on individual aerosol components as well as aerosol properties over ocean has been recommended [Inness et al., 2019]."

• *9/262 why only so few matches between MIC CAMS-RA (as modeling has AOD data at any location and time)?*

- CAMS RA data has a temporal resolution of 3 h, which limits the available datapoints from one CAMS RA grid cell during daylight to about 4 or 5. If the ship is able to move to another grid cell (one cell is 80x80km) within $\pm$30min of a CAMS timestep, the number of possible collocation could increase. But since the top speed of RV Polarstern is about 30 km/h its unlikely that there are more than two collocation per CAMS timestep. If we assume very clearsky conditions and a fast moving ship, the number of collocation samples per day might reach up to 8. From all cruises ever done within MAN until 12.2016 (CAMS data was only available until this date), there are only 1590 days consisting of at least one valid AOD measurement (valid implies here that none of the AOD values of the channels 380,440,500,675 and 870nm are NaN). Therefore a total number of 2474 valid collocated datapoints seem reasonable.

• *10/298 Well it is sad that GUV retrieved AOD are often inconsistent to (trusted) MICROTOPS data – as apparent more algorithm work is needed.*

- Sure, since the methods of measurement are considerably different, it is hard to compare the direct sun products without extensive radiative transfer modeling. On the presented database, we tried to produce a fair comparison, trusting the Microtops data. As also emphasized in the manuscript, the deviation of both instruments on a majority of cases is within the uncertainty limits of both instruments. We are aware that the linear dependence found in the bias is most likely not applicable to all situations with varying aerosol type. The correction with the linear factor is only feasible to provide a fair comparison, closing the gap of different methods/field-of-view influences within the presented study. Longer term, we would like to come up with an improved correction for the GUVis data, which might be situation-dependent.

- *10/306 To achieve GNU vs MICROTOPS consistency a rather complex procedure is applied to GNU data based on comparisons to almost simultaneous MICROTOPS samples. Hereby the described method seems a bit arbitrary to me. To account for the forward scattering into the GNU 15deg field of view (isn't 15 deg just the width of the band so in effect the missed diffuse radiation is even larger than 15 deg?) a constant Cis defined...while in reality this forward scattering contribution depends (at completely cloud-free conditions) on aerosol size (!) and multiple scattering potential (AOD, air-mass)...so a constant seems a poor choice. And then there is even a second scaling factor for a temporal (do you mean sun-elevation?) correction!*

- Actually, the constant Ci is applied for a calibration correction to compensate for differences in Microtops and GUVis calibration for the sake of comparability. This is done separately for each cruise, since the instruments are calibrated between the cruises and instrumental response might drift over time. We have implemented here the method used for MFRSR shadowband radiometers introduced by Alexandrov et al. [2002]. This corrections is done on the irradiance measurements (Eq.5) and therefore appears as the term $\mu_0\, c_i$ in Eq. 10. In contrast, S_i is introduced as a constant linear scale factor of the GUVis AOD in order to correct for the discrepancy attributed to forward scattering. di Sarra et al. [2015] introduced this method for shadow band radiometers, also suggesting a quadratic fit. Based on our results, a quadratic term does not lead to significant improvement in our case, so we chose to utilize the linear model to reduce the complexity (as stated in L338). As such, we disagree that the corrections are arbitrary.

- *12/376 a fixed bias of -0.02 (with the new processing) seems strange as I would expect a bias related to missed forward scattering to be function of AOD and size..? Actually Figure 3 shows such an AOD dependence of the bias. Generally: the entire correction explanations of 3.4 and of 4.1 (till line 400) including the comparisons of Table 4 are technical details which seem better suited for the Appendix as the focus should be on the comparison of satellite (and CAMS) data to trusted references.*

- Yes, you are right, the bias is not fixed, and is proportional to AOD in our case. This paragraph refers to the statistics calculated in Table 4, where we calculate the mean bias (which in this case is -0.02) and its dependency on AOD as the correlation factor R(D) (in this case about -0.7, which emphasize the linear dependence also seen in Figure 3). As you rightly explain, we have attributed this to forward scattering, knowing the difference in "field-of-view" of both instruments. Therefore we have chosen to correct this linear dependent bias to produce the GUVisE dataset.

  As the comparison of MIC and GUV is part of the discussion, and we think that the automatic nature of GUV measurements offers advantages over the manual MIC measurements, we have decided to not move this sections to the appendix.

- *12/383...the forward scattering missed radiation problem is more an issue for larger size and AOD (e.g. especially relevant for larger mineral dust sizes). You actually show in Figure 3a that this bias is near zero at low AOD and stronger at larger (likely dust) AOD...in that sense it is misleading to talk about a -0.02 bias (but I said this before).*

- This comment is covered by the answer of the previous comment.

- *13/412 how was the SEVIRI AOD at 550nm derived (was the MICROTOPS Angstromparameter used?)*

- The ICARE SEVIRI product (*SEV_AER-OC-L2*) is based on the wavelengths 630 and 810nm. The dataset is delivered with AOD at 500nm and AE. We used the AE reported for the SEVIRI product for the calculation of AOD at 550nm. We added "*Since the SEVIRI dataset does not provide AOD at 550 nm, it was calculated with Eq. (1) using the AOD of 630 nm and the AE from the SEVIRI dataset.*" to this paragraph.

- *13/417 The larges MODIS AOD outliers in the 0.3 bin vs GUVisE may be less meaningful since the sample number is fairly low. This raises the question, why the GUVisE dataset is so small and why this data-set only addresses selective dust and marine cases. Does it mean that the retrieval and correction method very frequently fails?*

- No, until now we only have data of five cruises with RV Polarstern, which are about 4-5 weeks of consecutive measurements, so about 170 days of measurements on the ocean. This means there have been about 340 overpasses (Terra & Aqua) during daytime to collocate the data to. Subtracting days in port or days with some clouds, a number of about 200 collocated datapoints does sound reasonable. I compared the number of collocated datapoints of GUVis and MIC from the five cruises which operated the the GUVis (see table 1). The table shows that more collocations are achieved with GUVis. This is expected, since GUVis measures automatically and continuous, while Microtops is operated manually with individual breaks.

Table 1: The table presents the total amount of datapoints collocated from satellite to the shipborne datasets of GUVis, Microtops, within a 30 and 5 min timeframe. The table compares data collected during certain cruises with *RV* Polarstern (PS83, PS95, PS98, PS102 and PS113).

Reference	GUVis	MIC
total	10412	1572
Collocation [$\pm$min]	30 / 5	30 / 5
CAMS	141 / 90	157 / 67
SEVIRI	1126 / 996	697 / 467
MxD04_L2	147 / 80	115 / 34
MxD04_3K	210 / 82	141 / 29

- *14/424 I do not see major differences between 4a and 4b unless low (and less meaningful) statistics is considered. the low GUV sample number remains a mystery to me (a few 'marine' and 'dust' cases).*

- Yes, there is no more a clear improvement from GUVisE to COMB. The paragraph was written with a comparison to GUVis (no forward scattering / calibration correction) as panel a) and we overlooked this detail in the internal review process. This paragraph has been omitted/merged in the new version of the manuscript with the leading and following paragraph addressing Figure 4.

- *14/435 you could compare if the SEVIRI AOD550 data would be different if the MICROTOPS 440/870 Angstrom had been used instead (also the SEVIRI band are spectrally much broader with some water vapor contamination especially in 810nm band).*

- Yes, as expected, the SEVIRI performance increases with the usage of AE from an external source like Microtops. Below in Table 2 we show the results as in Table 5 in the manuscript, but SEVIRI AOD at 550nm is calculated once with its original AE and once with MIC AE. With the usage of MIC AE the performance of SEVIR at 550nm is similar to the comparison at 630nm (lower bias, less proportionality of bias on AOD). Nevertheless, as it is the goal of the study to show certain shortcomings of the satellite products, we have calculated all statistics using only data of these datasets itself. High uncertainty of SEVIRI AE is expected can be shown in this study.

Nevertheless, we have reworked Table 5 also to your recommendations later, including the usage of MIC AE for SEVIRI AOD at 550nm (see Sect.4.2 and Table5 in the manuscript).

Table 2: Statistics as in Table 5 SEVIRI versus MIC, but SEVIRI AOD 550 is calculated with AE from SEVIRI product (original) and with MIC AE for comparison.

Instrument	N	channel	R	linear regression	R(D)	bias ± LOA	G1	G2
-	-	nm	-	-	-	-	%	%
SEVIRI original	10060	550	0.90	Y = 0.81 X-0.00	0.21	0.03±0.15	56	68
		630	0.89	Y = 0.87 X-0.00	0.06	0.02±0.14	63	73
SEVIRI MIC AE	10055	550	0.88	Y = 0.85 X+0.00	0.08	0.02±0.15	60	71
		630	0.89	Y = 0.87 X-0.00	0.06	0.02±0.14	63	73

- *14/439 I would say Figure 5 shows (opposite to your assessment) that SEVIRI AOD is usually higher than MODIS (at low to median AOD) and that at high AOD SEVIRI likely has a low bias, as (large AOD dust events with no solar spectral dependence)are probably removed during cloud screening.*

- We did not draw this conclusion from figure 5 but from figure A3, where SEVIRI AOD (630nm) is indeed slightly lower in all cases. But as we reworded this section this statement has become obsolete. Figure5 was updated to show the comparison of 550 and 630nm. As you have mentioned, SEVIRI overestimation is larger than MODIS for lower AOD values and turns into an underestimation at higher AOD bins.

- *14/445 Can you say this when samples of the two different MODIS products are different?*

- Our collocation method ensures that all pixels within a 0.2° radius are considered. The MODIS products may have different spatial resolutions, but are both built on the same radiance observations using the same algorithm and lookup tables. Because of their resolution both products differ only due to their method of combining pixels of the radiance measurement. Over ocean this means the number of good pixels for a valid retrieval is reduced for the 3km product (5 instead of 10 for 10 km product) and obviously the size of the pixel array of the retrieval box (6x6 instead of 20x20 for the 10 km product). Nevertheless, as the difference is negligible, this sentence is omitted in the new version of this manuscript.

- *15/471 Why did you apply SEVIRI Angstrom and not MICROTOPS Angstrom data to interpolate from 630 to 550nm?*

- As we want to evaluate the satellite products, we compared AOD and AE from the satellite to the reference data. As expected this leads to uncertainties for AE of the SEVIRI product, which is calculated based on the measurements of only two spectral channels. With the revision of this section, we have calculated AOD 550 for SEVIRI both ways (SEVIRI AE and MIC AE) to show the limitations of the SEVIRI AE.

- *15/474 Why would I want to use COMB if I have MIC (and much better statistics)*

- COMB ensures the comparability of both GUVis and MIC products, since both instruments utilize different methods and agree within their uncertainty limits. As the true values of AOD are unknown, COMB can serve as a best estimate of the surface measurements. Unfortunately, the statistics are restricted to situations when both instruments are operated together, which means data are available only for four cruises and with a time resolution determined by MIC. As MIC was operated since 2004 in numerous cruises, statistical significane is obviously better. Extending this study in the future will lead to better statistics, and will also offer additional insights on the accuracy of both shipborne instruments.

- *15/478 the MICROTOPS 440/870 Angstrom is more reliable than the 670/870 Angstrom. (The label in Figure 7 says 440/870...but the captions say differently?). Anyway, you do make sure not to include AE data if AOD at 870nm is below 0.05 ... otherwise derived Angstrom parameters are likely close to 'garbage'.*

\- Thank you. We corrected the label of Figure 7. AE is calculated for 440/870nm. Also we have redone the figure 7 and similar figures and restrict AOD870>0.05.

- *16/489 I agree that at low AOD satellite tend to overestimate AOD due to a more relaxed or insufficient cloud-screening (sometime there are many low clouds which are hardly visible)*

\- Thanks.

- *16/495 the CAMS forecast model (and thus also MODIS AOD data assimilations) have a too high fine-mode fraction for the dust outflow region. I am not sure if the problem of the satellite retrievals is related to non-spherical shapes of dust: Assuming spherical shapes in retrieval models the higher side-scattering of non-spheres is compensated (then incorrectly) by an extra fine-mode fraction.*

\- While this is an interesting interpretation, we decided to not include this in the manuscript, as we do not investigate fine and coarse mode separately.

- *17/530 Frequently the AOD over oceans is quite stable...as long as there are no changes in near surface winds and unless you enter/exit a big plume, as the Saharan dust outflow. Then AOD variations are much stronger than suggested here (possibly these cases were removed in this study a mixed aerosol type to not acceptable aerosol type). In particular for process studies these aerosol gradients (as few as they may be)are of interest.*

\- Yes, we reworded the paragraph slightly to emphasise such process studies as an application for high temporal resolution products. In Sect.4.2:

" *Continuous evaluation of the daily cycle of AOD are only possible with geostationary satellites such as SEVIRI. Nevertheless, with the high temporal resolution, the SEVIRI product may be needed for many applications, such as case studies of dust or smoke plume development, where high variability of AOD is expected.*"

In Sect.5:

"*Nevertheless, the SEVIRI AOD product provide valuable information on the temporal evolution of AOD fields where aerosol changes rapidly. Specific cases with high temporal variability are dust storms, plumes of volcanic ash or the passing of frontal systems.*"

- *17/549 the positive bias at low AOD is still there! You cannot use negative biases at higher AOD to suggest a lack of bias.*

\- To cover this we reworded Sect.4.2 and recalculated Table 5 as you have recommended.

- *18/558 The (to MODIS) added use of ATSR data had a negligible impact, since MODIS data were dominating in volume.*

- Yes. Still, we see a slight difference favouring data with ATSR. But as the number of datapoints is low, this small differences are not that meaningful. As emphasized in L563.

- *18/575 I would not expect significant differences between MODIS and CAMS except that the aerosol typing (e.g. AE) becomes worse in CAMS, as AE in CAMS in model prescribed.*

- Yes, we found no significant differences. But we cannot verify if AE is worse in CAMS, it just looks more restricted.

- *19/610 this is not too convincing as along as a MICROTOPS reference is needed to assure accuracy.*

- In order to assure the comparability we have to correct the influence of forward scattering on the shadowband radiometer. Still, the shadowband provides good performance for irradiance measurments. Therefore, we think that the GUVis works fine as an standalone instrument, but one has to keep the limitations for aerosol retrieval in mind, such as larger influence of forward scattering compared to a narrow field of view sunphotometer.

- *19/620 an important result is that AOD biases change with AOD strength*

- Thank you for pointing this fact out. We added this sentence to the paragraph: "*Also it was shown that the bias of SEVIRI AOD is dependent on the AOD, as the bias for AOD < 0.4 is larger than for the MODIS product, and the bias for AOD $\geq$ 0.4 turns negative (underestimation).*"

- *20/629 have you determined the MODIS Angstrom (if so which wavelengths) or was simply the available product from MODIS used?*

- The MODIS AE is determined from the MODIS AOD. For the AE comparison to MIC we have used the 470 and 860nm channels.

- *20/641 I disagree. There are biases and AOD retrievals make assumptions, which are not validated and often can be poor - affecting in turn the assigned AOD.*

- While we acknowledge and generally agree with the concerns raised in this comment on satellite retrievals, our results do confirm that the satellite products agree with the shipborne reference observations within the expected error limits. Hence, we believe our statement “our results confirm that satellite products can provide a global view of AOD” is well-founded, at least as long as the error limits are properly taken into account. To address the concern, we have modified the text as follows:

  "*Our results confirm that satellite products can provide a global view of the spatiotemporal aerosol distribution e.g for climate studies or model assimilation, as long as their error limits are properly taken into account, and spectral extrapolation of products is avoided.*"

- *21/644 I agree that AERONET sky capabilities and the use of available inversion data would be great. The MFRSR type GUV however is a different approach. AERONET inversions show that sky data offer aerosol size (-distribution) detail, while info on aerosol absorption requires large AOD cases. Anyway, those inversion capabilities with GUV need to be demonstrated first.*

- Yes, we plan to continue working on the GUVis processing and want to investigate further inversion methods in comparison to AERONET / MAN.

- *22/...A nice touch to mention EarthCare...although 3MI might be an aerosol satellite dedicated satellite sensor, which likely will launch and provide data sooner.*

- Yes, that might indeed be true. We mentioned MTG, EarthCARE and ESP-SG in no particular order.

Figures and tables

- *table 1 nice*

- Thanks.

- *table 2 I would rename the categories: maritime -> maritime background, mixed and continental -> continental transport, desert dust -> mineral dust transport. I would get rid of this biomass (near sources) component since aged transported biomass is more like continental transport. I also wonder why there are so many 'no class' events for GUVis – apparent outliers as they do not appear in COMB. It is disappointing that COMB matches are only 5% compared to the MIC matches*

- We now applied the aerosol classification as you suggested in the manuscript. Yes, the low number on datapoints do not lead to meaningful statistics, but this is expected, as we only have five cruises on the Atlantic ocean where we can collocate MIC and GUVis. As shown in the answer to your earlier comment to 13/417, the total amount of MIC datapoints during this five cruises is 1572.

- *table 3 the C-values without a few exceptions are very close to 1.0, so in effect only the lab calibration factors seem to matter (I am not sure how this V to radiance conversion was done, though). And S to correct for the forward scattering into the field of view should be a function of AOD and size (which is also a function of AOD since large dust AOD usually carry larger sizes).*

- This table presents the lab calibration factors k for the GUVis instrument along with the C-values to show the difference of calibration of GUVis and Microtops. All C-values are close to one, so both instruments are well calibrated (and the bias is therefore not caused by different calibrations). S is basically the slope for the linear correction of the bias caused by the forward scattering (see Eq.10). Therefore the correction is dependent on AOD. On our case study we found a simple linear correction model feasible, but we are aware, that this might not be true for cases with different aerosol type or mixtures.

- *table 4 I would move this into the Appendix.*

- As the comparison of MIC and GUV is part of the discussion, we have decided to not move this table to the appendix.

- *table 5/6/7 Maybe all the data can be compacted into a single table. I would work with 550nm statistics only (and use Microtops data to convert SEVIRI to 550nm). I would focus on MIC data and I would separate by 3 regions (dust outflow, continental outflow and maritime background) and there into lower and higher AOD. Also I would focus on bias and I do not know how to judge correlation. With low AOD cases (and more noise)correlations are meaningless.*

- We see the benefits of this condensed table you suggest. Therefore we have updated Table5 as you recommend and drop table 6/7 and the appendix tables. Nevertheless, we kept the correlation in, as it is widely used in validation studies and can therefore be compared to them. We are aware, that the correlation is meaningless for certain selections, such as maritime aerosol.

- *figure 1 great overview. But I cannot believe a dust domination in the southern hemisphere (as for the last cruise near 8S) - unless close to the Namib. I also would separate into transported dust (AE <0.5, AOD >0.15), maritime background (AOD <0.15),continental transport (AE >1, AOD >0.15) and mixed (AE 0.5 to 1.0, AOD >0.15) unless close to the Namib. For a clearer presentation I would compare the different reference data in one plot and MIC vs satellite and CAMS in another plot*

\- We applied your suggested thresholds and wording of the aerosol classification and updated all figures and tables and implemented the following sentences to Section 3.1:

"*The pair of AOD and AE values is checked against empirical thresholds to identify the dominant aerosol type of the current situation as being one of maritime background (AOD < 0.15), mineral dust transport (AE < 0.5, AOD > 0.15), continental transport (AE >1, AOD > 0.15), or mixed (0.5 < AE <1, AOD > 0.15) type. It should be noted, that all categories are expected to cover mixed aerosol types to some extent. Therefore, the mixed category consists of a mixture of aerosol without a dominant type.*"

- *figure 2 redo with the suggested typing of Figure 1 comments*

\- Done.

- *figure 3 /4 something for the Appendix (why are counts of COMB sometimes higher than for GUVisE in figure 4?)*

\- Counts for individual bins may differ between GUVisE and COMB occasionally when AOD values are close to the bin borders, since COMB consists of the means values between GUVisE and MIC. But as we looked at this detail, we found that the datapoint numbers presented in the Figure 5 and table 5 are lower as shown in Table 2. We found a bug in our code, which omits all datapoints from GUVisE which are not collocated to MIC. After the fix, we calculated Figure5 and table 5 again with the correct number of collocated datapoints.

- *figure 5 nice detail...but just because SEVIRI has likely (and missed dust event related) biases at higher AOD, while most cases (at low AOD) are larger than MODIS you cannot conclude that the SEVIRI overestimate (compared to MIC) is lower*

\- We did not draw this conclusion from figure 5 but from figure A3, where SEVIRI AOD (630nm) is slightly lower in all cases. We admit, that this is easily overlooked, since the paragraph mainly discusses the findings from figure 5. We added "*(see Fig. A3 and Table 5)*" to the sentence L436-L438 for clarification.

- *figure 6 seeing this plot, who wants to use the Angstrom parameter from satellite remote sensing and even modeling? SEVIRI seems to use a very simple Aerosol model without allowing for large and small Angstrom parameters. Similarly CAMS is constrained by composition, which does not allow for large Angstrom parameters. On the other hand very high Angstrom parameters with MICROTOPS may be associated with low AOD. Thus I recommend to show Figure 6 only for MIC data, when AOD at 870 is at least 0.1.*

  Actually I call a redo of Figure 6 via a Figure 5-type plot where Angstrom data are compared as a function of AOD bins.

\- We have restricted the comparison to AOD(870nm)>0.05, as you have recommended before. But we opt to stick with the scatter type of the figure, since it nicely shows the limitations of aerosol products in terms of usage of simple models or compsition restrictions as you have analysed. Nevertheless, the following Figure shows the AE comparison in a Figure5 type plot to compare AE as a function of AOD bins. Apart from conclusions already drawn from Figures 7/8 and 9 in the manuscript we see no improvement choosing this representation style.

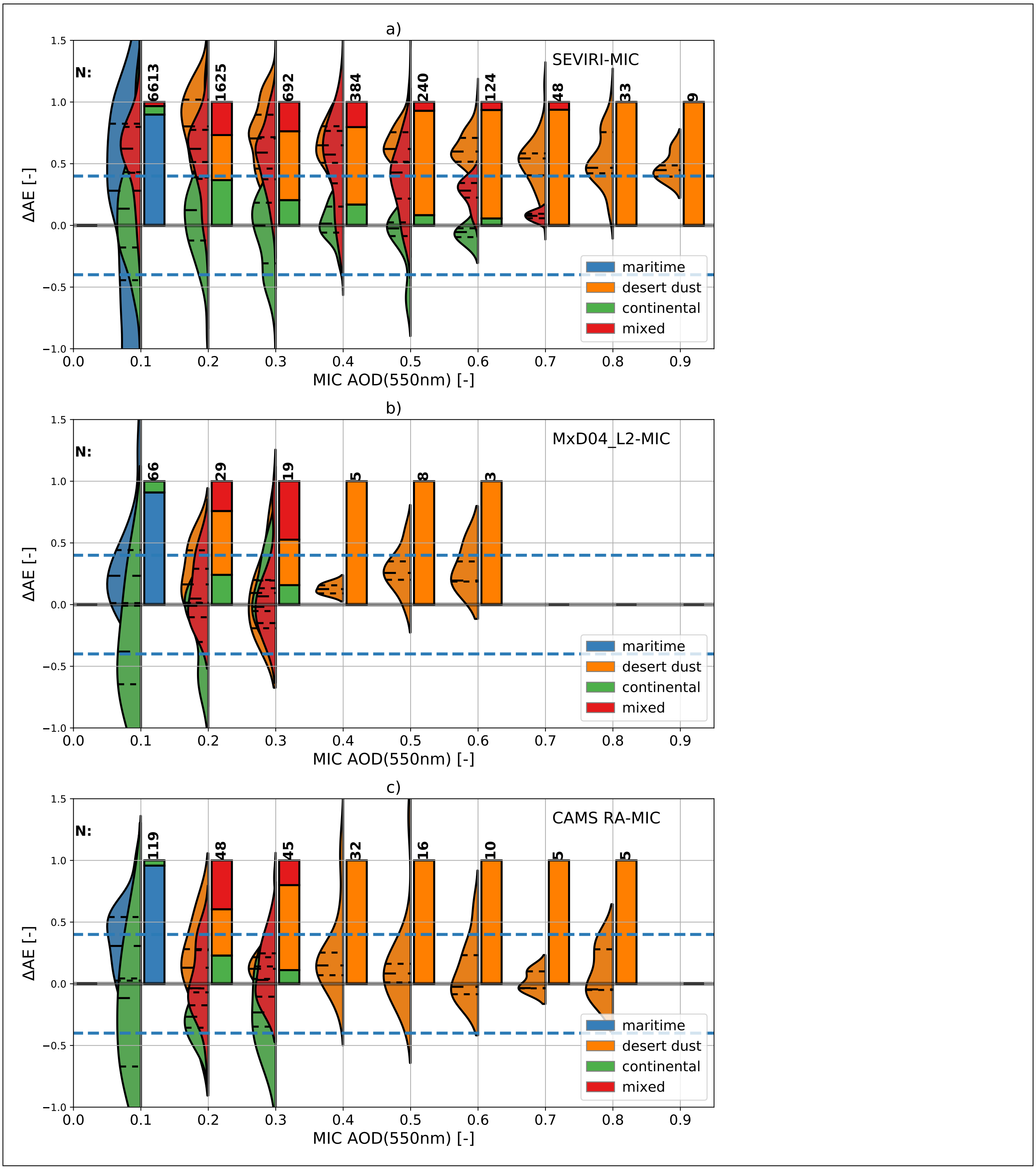

Figure 1: A plot like Figure 5 but with AE difference and aerosol type

- *figure 7/8/11 I would redo the aerosol types – see comments to Figure 1 (I assume all component samples fit within circles)*

- Done.

- *figure 9 interesting aspect*

- Thanks.

- *figure 10 nice*

- Thanks.

resource

I placed relevant monthly 1x1 lat/lon gridded data of the MACv2 aerosol climatology on ftp in ascii and netcdf (for details on file naming look at README) `ftp://ftp-projects.zmaw.de/aerocom/climatology/MACv2_2018/550nm/for_tropos/` *there you find expected monthly average for the 450nm/850nm Angstrom parameter and aerosol single scattering properties at 450, 550, 650 and 850nm ...maybe this helps as general reference*
Within the scope of the current study, we focus on the instantaneous agreement of satellite products with the shipborne observations. Hence, we did not find a direct application of these data for this paper at this stage. Nevertheless, we thank You for providing these datasets, as in the future, we are planning to investigate the radiative effects of aerosols based on the shipborne measurements. A comparison with the monthly climatology and the additional information provided on the single scattering albedo might be interesting contributions of the MACv2 climatology to these plans.
We have therefore added the following sentences to the Outlook:

- L647: "*It also offers the chance of an evaluation of the*" => "*to evaluate the direct ...*"
- and add => "*and these observations could contribute towards improving climatological estimates of the aerosol radiative effect (e.g. Kinne [2019]* "

References

Mikhail D. Alexandrov, Andrew A. Lacis, Barbara E. Carlson, and Brian Cairns. Remote Sensing of Atmospheric Aerosols and Trace Gases by Means of Multifilter Rotating Shadowband Radiometer. Part I: Retrieval Algorithm. *Journal of the Atmospheric Sciences*, 59(3):524–543, feb 2002. doi: 10.1175/1520-0469(2002)059<0524:rsoaaa>2.0.co;2.

Alcide di Sarra, Damiano Sferlazzo, Daniela Meloni, Fabrizio Anello, Carlo Bommarito, Stefano Corradini, Lorenzo De Silvestri, Tatiana Di Iorio, Francesco Monteleone, Giandomenico Pace, Salvatore Piacentino, and Sergio Pugnaghi. Empirical correction of multifilter rotating shadowband radiometer (MFRSR) aerosol optical depths for the aerosol forward scattering and development of a long-term integrated MFRSR-Cimel dataset at Lampedusa. *Applied Optics*, 54(10):2725, mar 2015. doi: 10.1364/ao.54.002725.

Antje Inness, Melanie Ades, Anna Agustí-Panareda, Jérôme Barré, Anna Benedictow, Anne-Marlene Blechschmidt, Juan Jose Dominguez, Richard Engelen, Henk Eskes, Johannes Flemming, Vincent Huijnen, Luke Jones, Zak Kipling, Sebastien Massart, Mark Parrington, Vincent-Henri Peuch, Miha Razinger, Samuel Remy, Michael Schulz, and Martin Suttie. The CAMS reanalysis of atmospheric composition. *Atmospheric Chemistry and Physics*, 19(6):3515–3556, mar 2019. doi: 10.5194/acp-19-3515-2019.

S. K. Kharol, K. V. S. Badarinath, D. G. Kaskaoutis, A. R. Sharma, and B. Gharai. Influence of continental advection on aerosol characteristics over bay of bengal (BoB) in winter: results from w-ICARB cruise experiment. *Annales Geophysicae*, 29(8):1423–1438, aug 2011. doi: 10.5194/angeo-29-1423-2011.

Stefan Kinne. Aerosol radiative effects with MACv2. *Atmospheric Chemistry and Physics*, 19(16): 10919–10959, aug 2019. doi: 10.5194/acp-19-10919-2019.

R. Levy, C Hsu, et al. Modis atmosphere l2 aerosol product. nasa modis adaptive processing system, goddard space flight center, usa, 2015.

R. C. Levy, S. Mattoo, L. A. Munchak, L. A. Remer, A. M. Sayer, F. Patadia, and N. C. Hsu. The collection 6 modis aerosol products over land and ocean. *Atmospheric Measurement Techniques*, 6(11): 2989–3034, 11 2013. ISSN 1867-8548. doi: 10.5194/amt-6-2989-2013.

John N. Porter, Mark Miller, Christophe Pietras, and Craig Motell. Ship-based sun photometer measurements using microtops sun photometers. *Journal of Atmospheric and Oceanic Technology*, 18 (5):765–774, 2001. doi: 10.1175/1520-0426(2001)018<0765:SBSPMU>2.0.CO;2.

L. A. Remer, S. Mattoo, R. C. Levy, and L. A. Munchak. MODIS 3 km aerosol product: algorithm and global perspective. *Atmospheric Measurement Techniques*, 6(7):1829–1844, jul 2013. doi: 10.5194/amt-6-1829-2013.

Beat Schmid and Christoph Wehrli. Comparison of sun photometer calibration by use of the langley technique and the standard lamp. *Applied Optics*, 34(21):4500, jul 1995. doi: 10.1364/ao.34.004500.

Manish Sharma, Dimitris G. Kaskaoutis, Ramesh P. Singh, and Sachchidanand Singh. Seasonal variability of atmospheric aerosol parameters over greater noida using ground sunphotometer observations. *Aerosol and Air Quality Research*, 14(3):608–622, 2014. doi: 10.4209/aaqr.2013.06.0219.

A. Smirnov, B. N. Holben, I. Slutsker, D. M. Giles, C. R. McClain, T. F. Eck, S. M. Sakerin, A. Macke, P. Croot, G. Zibordi, P. K. Quinn, J. Sciare, S. Kinne, M. Harvey, T. J. Smyth, S. Piketh, T. Zielinski, A. Proshutinsky, J. I. Goes, N. B. Nelson, P. Larouche, V. F. Radionov, P. Goloub, K. Krishna Moorthy, R. Matarrese, E. J. Robertson, and F. Jourdin. Maritime aerosol network as a component of aerosol robotic network. *Journal of Geophysical Research: Atmospheres*, 114(D6), 2009. ISSN 2156-2202. doi: 10.1029/2008JD011257.

C. Toledano, V. E. Cachorro, A. Berjon, A. M. de Frutos, M. Sorribas, B. A. de la Morena, and P. Goloub. Aerosol optical depth and ångström exponent climatology at el arenosillo aeronet site (huelva, spain). *Quarterly Journal of the Royal Meteorological Society*, 133(624):795–807, 2007. ISSN 1477-870X. doi: 10.1002/qj.54.

J. Witthuhn, H. Deneke, A. Macke, and G. Bernhard. Algorithms and uncertainties for the determination of multispectral irradiance components and aerosol optical depth from a shipborne rotating shadowband radiometer. *Atmospheric Measurement Techniques*, 10(2):709–730, 2017. doi: 10.5194/amt-10-709-2017.

Evaluation of satellite-based aerosol datasets and the CAMS reanalysis over ocean utilizing shipborne reference observations

Jonas Witthuhn[1], Anja Hünerbein[1], and Hartwig Deneke[1]

[1]Leibniz Institute of Tropospheric Research, Leipzig, Germany

Correspondence: jonas.witthuhn@tropos.de

Abstract. Reliable reference measurements over ocean are essential for the evaluation and improvement of satellite- and model-based aerosol datasets. Within the framework of the Maritime Aerosol Network, shipborne reference datasets have been collected over the Atlantic ocean since 2004 with Microtops sun photometers. These were recently complemented by measurements with the multi-spectral shadowband radiometer GUVis-3511 during five cruises with the research vessel *Polarstern*. The AOD uncertainty estimate of both ship-borne instruments of ±0.02 can be confirmed, if the GUVis instrument is cross-calibrated to the Microtops instrument to account for differences in calibration, and an empirical correction to account for the broad shadowband ~~and~~ as well as the effects of forward-scattering is introduced. Based on these two datasets, a comprehensive evaluation of aerosol products from the Moderate resolution Imaging Spectroradiometer (MODIS) flown on NASA's Earth Observing System satellites, the Spinning Enhanced Visible and Infra-Red Imager (SEVIRI) onboard the geostationary Meteosat satellite, and the Copernicus Atmosphere Monitoring Service reanalysis (CAMS RA) is presented. For this purpose, focus is given to the accuracy of the aerosol optical depth (AOD) at 630 nm in combination with the Ångström exponent (AE), discussed in the context of the ambient aerosol type. In general, the evaluation of MODIS AOD from the official Level-2 aerosol products of C6.1 against the Microtops AOD product confirms that 76% of datapoints fall into the expected error limits given by previous validation studies. The SEVIRI-based AOD product exhibits a 25% larger scatter than the MODIS AOD products at the instrument's native spectral channels. Further, the comparison of CAMS RA and MODIS AOD versus the shipborne reference ~~show similar performances of~~ shows similar performances for both datasets, with some differences arising from the assimilation and model assumptions. When considering aerosol conditions, an overestimation of AE is found for scenes dominated by desert dust for MODIS and SEVIRI products versus the shipborne reference dataset. ~~This~~ As the composition of the mixture of aerosol in satellite products is constrained by model assumptions, this highlights the importance of considering the aerosol type in evaluation studies for identifying problematic aspects.

1 Introduction

Aerosol particles directly influence the ~~Earth~~earth's radiation budget through their interaction with solar and terrestrial radiation~~. The direct radiative effect of aerosol (RE_{ari}) is defined as the change of radiative fluxes caused by aerosol particles (Boucher et al., 2013). Quantifying RE_{ari} on a global scale is crucial for understanding the role of aerosol in~~ , and indirectly by modifying the optical properties of clouds (Boucher et al., 2013). Studies of aerosol effects on the climate system ~~, and in~~

particular the climate forcing by aerosol in the context of anthropogenic climate change. To quantitatively estimate RE_{ari} are based on radiative transfer models. Therefore, knowledge about the spectrally resolved optical properties of different aerosol types is essential.

The annually averaged shortwave RE_{ari} over cloud-free oceanis estimated to lie in the range of -4 to -6 $\mathrm{W m^{-2}}$, which can mainly be attributed to Over ocean, sea spray (Bellouin et al., 2005; Loeb and Manalo-Smith, 2005; Yu et al., 2006; Myhre et al., 2007) . Besides maritime aerosol originating from sea spray, mineral dust emitted from the large desert areas (e.g. the Sahara desert) is a major contributor to both the short- and longwave RE_{ari} over ocean (e.g., Tegen, 2003; Christopher and Jones, 2007; Nabat et al. .

Aerosol also influences the Earth's radiation budget indirectly through its role in cloud formation and by modulating the radiative effects of clouds through a number of possible pathways (e.g., by altering cloud albedo or cloud lifetime) (Denman et al., 2007). All indirect effects on the Earth's radiation budget due to aerosol–cloud interactions can be combined into the effective radiative forcing arising from aerosol(Boucher et al., 2013)and desert dust (e.g., Tegen, 2003; Christopher and Jones, 2007 are the major contributors to the direct radiative effect of aerosol. Observations of aerosol load and optical properties with global coverage are required to improve our understanding of climate-relevant aerosol processes, to investigate their role in anthropogenic climate change, and to improve quantitative estimates of this effective forcing.

Satellite remote sensing provides global observations of aerosol properties and the radiation budget (Chen et al., 2011; Kahn, 2012). These observations are key to quantify RE_{ari}direct radiative effects of aerosols, in particular over ocean, where only limited surface observations (e.g. from ship) are available (Haywood et al., 1999). Due to the sensitivity of the retrievals to factors such as instrumental calibration and retrieval assumptions however, a critical evaluation of the accuracy of the resulting satellite datasets is essential for understanding their quality and limitations, e.g. by comparing these products with well-calibrated ground-based reference observations.

The most widely used satellite-based aerosol products are based on the Moderate resolution Imaging Spectrometer (MODIS) instrument flown on the polar-orbiting Terra and Aqua satellite platforms, which were launched in 1999 and 2002, respectively, by the National Aeronautics and Space Administration (NASA), and continue operations to this day. These products were evaluated in numerous studies in their evolution from Collection 4 (C4) (e.g., Remer et al., 2005; Kleidman et al., 2005) to C5 and C5.1 (e.g., Levy et al., 2010; Bréon et al., 2011; Misra, 2015) and finally to C6 and C6.1 (e.g., Munchak et al., 2013; Levy et al., 2013; Livingston et al., 2014). Validation of the product quality over ocean was more limited compared to that over land, and has mostly relied on coastal or island sites with sun-photometer measurements (e.g., Abdou et al., 2005; Bréon et al., 2011; Shi et al., 2011; Anderson et al., 2012; Wei et al., 2019). Ship or airborne reference observations were utilized less frequently (e.g., Smirnov et al., 2011; Adames et al., 2011; Schutgens et al., 2013). Levy et al. (2013) estimate estimated an error of aerosol optical depth (AOD) over ocean within the error limits of $[+(0.04+0.1\mathrm{AOD}), -(0.02+0.1\mathrm{AOD})]$ for the C6 products. Considered in this paper are the products from both the MODIS Aqua and Terra instruments and refer to them as *MxD04_3K* and *MxD04_L2* for the high resolution 3km Remer et al. (2013) and lower resolution 10km Levy et al. (2015) swath products, respectively.

In addition to the widely used aerosol datasets available from the MODIS instruments, datasets based on geostationary satellite observations are of high potential interest for scientific applications. In particular, their high temporal resolution combined with their fixed field of view on the earth enables studies of the diurnal cycle and the temporal evolution of aerosol plumes. Hence, the aerosol product of Thieuleux et al. (2005), which is based on the Spinning Enhanced Visible and Infra-Red Imager (SEVIRI) onboard the geostationary Meteosat second generation (MSG) satellites operated by the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT), is also taken into consideration in this evaluation. It is available at a temporal resolution of 15 min. Compared to the MODIS aerosol products, some limitations arise from the instrumental characteristics of the SEVIRI instrument, and thus have to be taken into account: the spatial resolution of SEVIRI is 3 km in nadir versus 1 km for MODIS, and only two spectral channels (630 and 810 nm) are utilized in the retrieval. A smaller set of 12 aerosol models is used as basis for the retrieval, and the product has received far less validation efforts (e.g., Bréon et al., 2011; Bernard et al., 2011). To our knowledge, it has not been validated previously with shipborne observations.

For many research purposes, aerosol properties from model-based reanalysis datasets are a promising alternative to the direct use of satellite-based aerosol products. In contrast to satellite products, aerosol properties from a reanalysis are available independent of cloud cover and satellite overpass time. The Copernicus Atmosphere Monitoring Service reanalysis (CAMS RA) is the latest global reanalysis of atmospheric composition produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) and provides global information on aerosol optical properties. It relies on the data assimilation of satellite observations into ECMWF's Integrated Forcasting System (Inness et al., 2019). In the case of aerosol, it has to be realized that MODIS datasets are assimilated into CAMS RA, so that differences between both datasets are expected to be relatively small and will mainly show the influence of model assumptions and the assimilation system of ~~the CAMSsystem~~CAMS.

In this study, two independent datasets of shipborne aerosol products are compared and used for an evaluation of both satellite products and the CAMS RA over ocean, with an additional focus on aerosol type. There is still a lack of shipborne spectral radiation measurements for this purpose (Brando et al., 2016). Furthermore, by separating the evaluation according to aerosol type, more insights can be gained into the limitations of the current satellite products. Also, further validation of the CAMS RA aerosol products with respect to aerosol type is needed (Inness et al., 2019). While the optical properties of maritime aerosol are considered to be relatively well understood, the optical properties of mineral dust are still the topic of ongoing research due to their complex, ~~non-spheric~~ non-spherical shape (Dubovik et al., 2006; Mishchenko et al., 1999), which introduces significant uncertainty in their optical properties and remote sensing.

Compared to observations on land, shipborne observations are more challenging due to the continuously moving nature of the observational platform caused by waves. Observations of aerosol optical properties were established within the framework of the Maritime Aerosol Network (MAN) as a sub-project of the Aerosol Robotic Network (AERONET), based on the sunphotometer technique. Global observations from MAN are available since 2004, and utilize the hand-held Microtops II sunphotometers (referred to as Microtops in the following text). It thus relies on the skill of human observers to compensate for the ship movement (Smirnov et al., 2009). An automatic approach to derive aerosol optical properties over ocean using the shadowband radiometer technique was established within the framework of the OCEANET project (Macke, 2009). The shadowband radiometer GUVis-3511 (referred to as GUVis in the following) built by Biospherical Instruments Inc. was operating alongside other OCEANET instruments to provide observations during five Atlantic transit cruises of the German research vessel Polarstern since 2014 (Witthuhn et al., 2017).

Observations from both the GUVis and Microtops instruments on a number of Polarstern ship cruises over the Atlantic ocean are utilized in this study. The GUVis aerosol product has received a substantial update since the version presented in Witthuhn et al. (2017). The improvements are briefly discussed in the appendix section (Sect. A). The comparison of these shipborne datasets to aerosol products from MODIS C6.1 and SEVIRI as well as the CAMS RA aerosol datasets is presented here, which were collocated to the ship's position along these cruises.

This paper has three principal goals:

1. Inter-comparison of both shipborne aerosol products in terms of their accuracy, with a particular focus on the verification of the uncertainty estimate of the GUVis dataset, and the usability of both datasets for the validation of satellite retrievals.

2. Evaluation of the satellite aerosol products from SEVIRI and MODIS over ocean using these shipborne datasets. A specific question is whether SEVIRI can offer additional information ~~due to~~ on the diurnal cycle and temporal evolution of aerosol.

3. Evaluation of the CAMS RA as an alternative source of aerosol information to MODIS and SEVIRI for research purposes.

The two shipborne datasets serve as reference for the subsequent validation study. Since they are based on different techniques, an inter-comparison is presented first to point out their individual strengths and weaknesses. In this context, focus is given in particular to their suitability for satellite validation.

Within the second and third point, the estimated error limits proposed previously for the MODIS AOD products are investigated compared to the deviations found in this study. These findings are ~~put into context to~~ understood in the context of the results found for the SEVIRI aerosol product ~~,~~ to observe how the limitations of the SEVIRI sensor influence the retrieval accuracy. Further, the benefit resulting from the increased time resolution of SEVIRI is investigated. Besides the accuracy of the AOD, the estimate of Ångström exponent (AE) is investigated, in particular in the context of characterizing the aerosol type. Both AOD and AE from the CAMS RA are compared to the satellite and shipborne datasets to identify differences due to the satellite retrievals, and to evaluate its performance during different aerosol situations.

The paper is structured as follows: First shipborne instrumentation and reference datasets are introduced (Sect. 2.1). A description of the satellite products and the CAMS RA are shown in Sect. 2.2 and Sect. 2.3. The methods utilized for aerosol classification, satellite data collocation and statistical measures for evaluation as well as the GUVis cross-calibration and aerosol forward scattering correction are reported in Sect. 3. The inter-comparison of the shipborne data and the comparison of the satellite products versus the shipborne reference is given in Sect. 4. Finally, the evaluation results are discussed in the conclusions and outlook sections (Sect. 5 and Sect. 6). In the appendix section (Sect. A) the update of GUVis irradiance processing algorithm is described.

2 Instruments and datasets

This section gives an overview of the shipborne instruments and reference datasets (Sect. 2.1) as well as the satellite (Sect. 2.2) and model reanalysis dataset (Sect. 2.3). All datasets are ~~publically~~ publicly available, see the section on data availability at the end of the article.

In this study, focus is given to the aerosol optical depth (AOD) and Ångström exponent (AE), the latter quantifying the dependency on wavelength λ of the former quantity. Specifically, the AOD at $\lambda = 440\,\mathrm{nm}$ (for inter-comparison of the shipborne datasets) and at $\lambda = 630\,\mathrm{nm}$ (for comparison of shipborne and satellite data) are mainly considered here, while the AE α is calculated from the AOD τ_A at $\lambda_1 = 440\ \mathrm{nm}$ and $\lambda_2 = 870\,\mathrm{nm}$ based on the Ångström relation as follows, unless otherwise noted:

$$\frac{\tau_{A,\lambda_1}}{\tau_{A,\lambda_2}} = \left(\frac{\lambda_1}{\lambda_2}\right)^{-\alpha}. \qquad (1)$$

2.1 Shipborne instruments and datasets

Two aerosol datasets based on shipborne observations are considered here as ground-based reference: ~~on~~ the sunphotometer Microtops (Microtops II manufactured by Solar Light Inc.), and the shadowband radiometer GUVis (GUVis-3511 plus BioSHADE accessory manufactured by Biospherical Instruments Inc.).

Both instruments are well-suited for operation on moving platforms such as ships. Their measurement principles however are rather different. The ~~Microtops is a~~ technical specifications of the Microtops and GUVis instruments are summarized in Table 1. The configurations of both instruments allow a direct comparison of all spectral channels of the Microtops versus corresponding GUVis observations.

The Microtops is a hand-held sunphotometer, which has to be pointed manually at the sun. ~~It~~ To minimize uncertainties arise from manual pointing, more than five consecutive scans are averaged to form one measurement (Smirnov et al., 2009). The Microtops instrument measures the incident direct normal solar irradiance with a field of view of 2.5° ° (Porter et al., 2001). The ~~shadowband radiometer~~ MAN Microtops sunphotometers are calibrated against an AERONET master Cimel sunphotometer, which in turn is calibrated using the Langley-technique.

The shadowband radiometer GUVis utilizes an entrance optic with a global field of view combined with a shadowband that performs a 180° ° sweep, while the global irradiance is measured at a high temporal frequency of ~~10~~15 Hz~~.~~

Hz. Several corrections are applied as post-processing to correct the influence of the ship motion, and to retrieve the direct spectral irradiance for later AOD calculation, as is described later. The measurement principle of the shadowband radiometer can be described as follows: While the global irradiance is observed with the shadowband in its ~~low~~ lowest position between sweeps, the shadowband blocks a fraction of the incoming diffuse irradiance during its rotation, and will occlude the direct irradiance at a specific angle determined by instrument orientation and sun position. From the irradiance time series measured during the sweep, the global, diffuse and direct irradiance components can be inferred (Witthuhn et al., 2017).

Prior to the processing of the GUVis sweeps, the measured irradiance data has to be corrected, to compensate for the motion of the ship and the imperfect cosine response of the instrument. The actual cosine response of the entrance optic is measured by the manufacturer during lab calibrations, and can be corrected by applying correction factors depending on the spectral channel and sun elevation, if the orientation angles of the ship are known. The motion correction utilizes the method of Boers et al. (1998) based on the ship motion angles to correct the direct and diffuse irradiance components. The GUVis instrument has been calibrated in a laboratory at regular two-year intervals using a 1000 W FEL standard calibration lamp as absolute reference. The correction and processing of GUVis irradiance data as well as the calculation of AOD is described in detail in Witthuhn et al. (2017).

Given the direct normal irradiance obtained from both instruments and a given spectral band, the AOD can be calculated using the well-known Lambert-Beer law, and by subtracting optical depth contributions from Rayleigh scattering and gas absorption.

The concept of the "field of view" of a sun photometer is not directly applicable to a shadowband radiometer. Instead, there is the "shading angle" as described in Witthuhn et al. (2017), which is the minimum angle between the edges of the shadowband as viewed from the center of the global entrance optic. For the GUVis, the shading angle is about 15° (depending on shadowband position), and thus relatively large in comparison to the Microtops field of view.

The wide angle of the shadowband of the GUVis causes an underestimation of AOD caused by the influence of the forward scattering of the aerosol (Russell, 2004). The GUVis processing algorithm has received a substantial update (see Sect. A) to compensate at least partially for this effect. The reduction of measured irradiance during the shadowband sweeps is stronger in situations with increased aerosol forward scattering. Besides some other refinements, an offset was introduced for estimation of the blocked diffuse irradiance as part of the processing algorithm update in order to compensate for this effect (see Sect. A).

The technical specifications of the Microtops and GUVis instruments are summarized in Table 1. The configurations of both instruments allow a direct comparison of all spectral channels of the Microtops versus corresponding GUVis observations.

Given the direct normal irradiance obtained from both instruments and a given spectral band, the AOD can be calculated using the well-known Lambert-Beer law, and by subtracting optical depth contributions from Rayleigh scattering and gas absorption. In the following, an overview of the shipborne datasets based on both instruments is presented:

(i) As the first dataset, all observations conducted during numerous cruises with the Microtops II sunphotometer in the framework of AERONET MAN since 2004 to 2018 in the area of the Atlantic ocean are used, holding . The uncertainty of Microtops AOD is estimated to be within ±0.02 (Smirnov et al., 2009). The datasets include a total number of 19250 valid data points. This dataset (referred to as MIC in the following text) also provides the diversity needed to investigate aerosol type–related effects for the evaluation of satellite products, and for the comparison with CAMS RA.

(ii) The second reference dataset (GUVis) is based on the GUVis shadowband radiometer. Observations with the GUVis were conducted within the framework of OCEANET (Macke, 2009) during Atlantic transect cruises with the German research vessel *Polarstern* operated by the Alfred-Wegener Institute since 2014. Until now, five cruises including the shadowband radiometer observations have been performed, namely PS83, PS95, PS98, PS102 and PS113. A Microtops instrument from MAN has also been operated in parallel on all these cruises. This offers the opportunity to directly compare both datasets. A direct comparison of Microtops and GUVis AOD product has already been presented for PS83 in Witthuhn et al. (2017). For both shipborne AOD datasets, the The AOD uncertainty is estimated to be within ±0.02 (Smirnov et al., 2009; Witthuhn et al., 2017)(Witthuhn et al., 2017)

. Following the same procedure, this comparison is extended to all available cruises, with some minor changes to obtain more meaningful results. The total number of valid GUVis observations is 10412.

In order to improve the agreement of the aerosol products of both instruments to acceptable limits, it has been found necessary to introduce a cross-calibration to the MIC instrument, and an empirical correction for aerosol forward-scattering, to account for differences arising from the limited accuracy of lab-based instrumental calibration, and the broad shadowband of the GUVis instrument. The correction is done fitting a linear regression curve (Eq. (10)) to the GUVis AOD (see Sect 3.4), similar to the approaches adopted by ~~(di Sarra et al., 2015) and (Wood et al., 2017)~~di Sarra et al. (2015) and Wood et al. (2017). This enhanced dataset is denoted as GUVisE in this study.

(iii) The enhanced GUVis dataset (GUVisE) ~~is~~ has been combined with the Microtops dataset to obtain a merged surface product, to test whether the combination can lead to further improvements in accuracy. ~~For this purpose, the mean of all GUVisE and Microtops AOD retrievals which are not flagged as outliers is calculated.~~ This combined surface dataset (COMB) serves as ~~the~~ third reference dataset for the evaluation of the satellite products. COMB consists of the mean of the collocated GUVisE and MIC AOD for this purpose. As shown in Table 2, the total amount of data points decreases to ~~1006~~ 1033 due to the ~~combination~~collocation requirement.

2.2 Satellite aerosol products

Satellite based aerosol datasets over ocean considered here are obtained from both the MODIS and SEVIRI satellite instruments. The MODIS Collection 6.1 (C6.1) level-2 aerosol products *MxD04_L2* (Levy et al., 2015) and *MxD04_3K* (Remer et al., 2013) are used from both ~~the~~ Terra and Aqua satellites. This dataset includes the AOD at 470, 550, 660, 860, 1240, 1630 and 2130 nm. The AOD(500 nm) and AE obtained from the SEVIRI instrument onboard the Meteosat Second Generation satellite ~~introduced by Thieuleux et al. (2005)~~ provided in *SEV_AER-OC-L2* product (Thieuleux et al., 2005) of the ICARE data center is also considered. In the following text unless otherwise stated, the terms "MODIS aerosol products" or "MODIS retrieval" refer to the *MxD04_L2* and *MxD04_3K* products, and similarly, the term "SEVIRI aerosol product" to the ICARE *SEV_AER-OC-L2* aerosol product.

Both aerosol retrievals are based on the inversion of the measured reflectance at top of atmosphere to estimate the AOD at the instrumental spectral channels, using lookup tables of radiative transfer calculations. The accuracy of these estimates critically depends on realistic assumptions about the optical properties of aerosols assumed in the calculations. A larger number of channels enables a more accurate choice of aerosol type used by the retrieval, and is thus expected to increase the overall accuracy. In addition, factors such as the spatial resolution of the sensor, the viewing geometry, sensor calibration, as well as the accuracy of cloud screening will influence the overall accuracy. While the SEVIRI retrieval is based on only two wavelengths (630 nm and 810 nm) (Thieuleux et al., 2005), the MODIS retrieval utilizes seven spectral channels. In addition, it is continuously monitored with ground-based observations at AERONET stations (Levy et al., 2015). ~~Therefore, it is expected that the AOD and AE obtained from MODIS will have a better accuracy than from SEVIRI when compared to the shipborne AOD and AE datasets. Further, a~~ A degraded accuracy for ~~these~~ aerosol properties in the presence of desert dust in both satellite products is expected, since dust particles are ~~nonspherical, contrary to the assumption of sphericity made~~ non-spherical. This leads to an increased side-scatter effect compared to spherical particles which are assumed in both retrievals.

Besides the retrieval differences, MODIS and SEVIRI products are also different due to their satellite platform characteristics. MODIS is operated on both the Terra and Aqua satellites, which fly in a polar orbit. For studies targeting aerosol properties at a specific location, MODIS observations are only available for the two overpasses during daylight, compared to SEVIRI with a time resolution of 15 minutes. On the other hand, the geostationary orbit of MSG leads to lower spatial resolution of nadir 3 km for SEVIRI versus a 1 km nadir resolution of MODIS. In order to avoid cloud contamination in the aerosol product, the MODIS retrievals consider multiple pixels together with a strict cloud mask, leading to a decrease of the spatial resolution to 3 km for the high resolution aerosol product (*MxD04_3K*), and to 10 km for the standard aerosol product (*MxD04_L2*).

2.3 CAMS RA aerosol product

CAMS RA is the latest global reanalysis dataset of atmospheric conditions produced by ECMWF (Inness et al., 2019). Amongst other atmospheric constituents, it contains the spectral AOD at a temporal resolution of 3 h on a global grid of 0.7° (corresponding to a T255 ~~spectral~~ spatial resolution). The advantage of utilizing CAMS RA over satellite observations is the availability of aerosol properties independent of factors such as cloud coverage or satellite orbit. Albeit the accuracy of AOD under cloudy sky conditions in the model might be questionable.

CAMS RA was developed based on the experiences gained with the former Monitoring Atmospheric Composition and Climate (MACC) reanalysis and the CAMS interim analysis (Inness et al., 2019). It relies on the assimilation of global observational datasets into the Integrated Forecast System (IFS) from various satellites to provide a global picture. In terms of aerosol properties, the AOD from the products of the MODIS C6 from both Terra and Aqua are assimilated, while the composition mixture is maintained as given from the IFS. Before its failure in March 2012, retrievals from the Advanced Along-Track Scanning Radiometer (AATSR; Popp et al. (2016)) flown aboard the Envisat mission were also being assimilated. The influence of this additional source of information for data assimilation on the accuracy is investigated in Sec. 4.3. Currently, the dataset covers the period 2003-2016, and will be extended in the following years. For the evaluation of the CAMS RA aerosol dataset an accuracy close to MODIS aerosol product is expected.

A first validation presented within Inness et al. (2019) emphasizes the high quality of AOD in the CAMS RA system, as judged by a comparison to AERONET stations around the world. However, an overestimate of AE was shown during desert dust events, and was attributed to ~~problems in realistically representing the fine- and coarse-mode fractions for dust particles in the aerosol formulation of~~ the fixed component mixture (e.g., less dust in CAMS RA) in the forecast model. Further evaluation with a focus on individual aerosol components as well as aerosol properties over ocean has been recommended (Inness et al., 2019).

3 Methods

This section gives an overview of the methods used for aerosol classification (Sect 3.1), for collocation of satellite and shipborne measurements (Sect 3.2), and presents the statistical measures used for evaluation (Sect 3.3), as well as the correction approach adopted for the GUVis aerosol product for better comparability to MIC AOD (Sect 3.4).

3.1 Aerosol classification

Our study aims to compare shipborne and satellite AOD products also with respect to the role of aerosol types. A satellite–independent aerosol classification is applied, which is based on the empirical method presented in Toledano et al. (2007) for Cimel instruments from AERONET. This method is also applicable to the Microtops AOD product, as it contains all required parameters. The aerosol classification is done by comparing the AOD at $\lambda = 440\,\mathrm{nm}$ with the AE calculated based on the 440 and 870 nm channels (Eq.(1)). The pair of AOD and AE values is checked against empirical thresholds to identify the ~~aerosol type~~ dominant aerosol type of the current situation as being one of maritime ~~, desert dust , continental, biomass burning or mixed type. For example, AOD(440~~background (AOD nm~~)~~ < 0.15), mineral dust transport (AE < ~~0.2 and AEbetween zero and two will identify the data point as maritime type(Toledano et al., 2007)~~0.5, AOD > 0.15), continental transport (AE >1, AOD > 0.15), or mixed (0.5 < AE <1, AOD > 0.15) type. It should be noted, that all categories are expected to cover mixed aerosol types to some extent. Therefore, the mixed category consists of a mixture of aerosol without a dominant type.

In the following text we shorten the aerosol type description from maritime background (maritime), mineral dust transport (desert dust, since over Atlantic ocean most dust cases originate from the Sahara desert), continental transport (continental). All results shown in this study separated by aerosol type (maritime, desert dust, continental, ~~biomass burning,~~ mixed) are based on this aerosol classification method. It should be noted that the shipborne observations at wavelengths of 440 and 870 nm are utilized for classification, even if figures and tables present AOD and AE for different channels (e.g., Fig. 7).

3.2 Collocation criteria

As common practice for spatiotemporal collocation with MODIS, a window size of 50x50 km and a time window of one hour is recommended by Ichoku (2002) for sunphotometer observations. For the MODIS C6 validation by Levy et al. (2013), a spatial radius of 25 km and a temporal window of $\pm 30\,\mathrm{min}$ ~~has~~ have been used. Both the *MxD04_L2* and *MxD04_3K* products have been validated using a window of 5x5 pixel by Munchak et al. (2013), resulting in different window sizes of $50\,\mathrm{km}^2$ and $15\,\mathrm{km}^2$, respectively. The following collocation technique is utilized here to find the appropriate pixel of the satellite dataset, and to compare it to the shipborne data obtained at a certain position. First, eligible satellite images are selected using a time-frame of $\pm 30\,\mathrm{min}$ around observations, and checking if the ship position is located within the field of view of the satellite image. Then, the distance angles of all pixel coordinates to the ship position has been calculated. The satellite AOD is finally calculated as the median of all non–cloudy pixel values with a distance angle ~~equal and below~~ less than 0.2°.

Choosing a distance angle threshold of $0.2°$ for the collocation of all satellite and model datasets to the shipborne observation assures that the same area around the reference observation is chosen regardless of satellite or model product, spatial resolution, and projection, and ensures comparability of results. This threshold results in a spatial radius of about $22\,\mathrm{km}$.

Applying the collocation strategy introduced above to the 19250 MIC data points results in a total number of remaining 1517 data pairs for *MxD04_L2*, 1448 for *MxD04_3K*, 10061 for SEVIRI and 2474 for CAMS RA, as shown in Table 2.

After collocation with the GUVis dataset, consisting of a total number of 10412 data points, the resulting number of data pairs is 147 for *MxD04_L2*, 210 for *MxD04_3K*, and 1126 for SEVIRI. The collocation with CAMS RA results in 141 data pairs. The number of collocated data pairs is rather small, limiting the statistical significance of the comparison results.

The number of data pairs per aerosol type classified based on the shipborne reference data as described in Sect. 3.1 is given in Table 2. Since the observations are performed across the Atlantic ocean, the dominant aerosol conditions are mainly maritime or desert dust originating from the Sahara desert (see Fig. 1 and Table 2).

3.3 Limit of agreement method

To assess the agreement of two measures (X,Y) of the same quantity such as AOD from Microtops versus GUVis, or the shipborne dataset versus satellite products, linear regression statistics and the Pearson product-moment correlation coefficient R (referred to simply as the correlation in the following text) are calculated. Further, the ~~analysis~~ analyses are extended with the so-called "limits of agreement" (LOA) method first introduced by Bland and Altman (1986). This method considers the mean of the differences of both quantities X - Y (i.e. the bias), and the LOA defined as the 95% confidence interval for those differences as additional parameters. As not stated otherwise, Y denotes the reference dataset for comparisons presented in this study.

For the evaluation of the uncertainty estimates for the shipborne observations, the method of Knobelspiesse et al. (2019) is adopted, weighting the difference X - Y (D) with their uncertainty estimate (σ_X,σ_Y):

$$(\mathrm{X}-\mathrm{Y})/\sqrt{\sigma_{\mathrm{X}}^2+\sigma_{\mathrm{Y}}^2}. \tag{2}$$

Thus, utilizing the LOA method together with the weighted difference, the uncertainty estimate can be confirmed if the uncertainty–weighted difference lies within the range of ± 1.96 for the 95% confidence interval (see Fig. 3). The percentage of outliers exceeding the limits of ± 1.96 is used as quantitative measure for the validation.

For the evaluation of the satellite products and CAMS RA, the bias and LOA (95% confidence interval) are used as a measure for the agreement to the shipborne reference datasets. Additionally, *Gfrac* defined as the percentage of data lying within expected error (EE) limits is calculated, in order to be consistent with other validation studies (e.g., Bréon et al., 2011). Expectations of the error are met, if 67% of data points of the satellite or model product fall into the EE range compared to the shipborne reference (Levy et al., 2013). Two EE limits are chosen here, originally presented for the MODIS aerosol product based on former validation studies e.g. by Abdou et al. (2005); Remer et al. (2008) and Livingston et al. (2014):

$$\mathrm{EE1} = \pm(0.03 + 0.05\mathrm{AOD}) \tag{3}$$

and more recently in Levy et al. (2013):

$$\mathrm{EE2} = [+(0.04 + 0.1\,\mathrm{AOD}), -(0.02 + 0.1\,\mathrm{AOD})] \tag{4}$$

EE1 is a general measure of agreement, since the boundaries are equally distributed around the reference dataset. EE2 has been specialized for the MODIS aerosol product, since a known overestimation is considered via different intercepts.

3.4 Cross-calibration and empirical correction of AOD

The relatively large differences originally observed in the comparison between Microtops and GUVis (Sect. 4.1), and their changes from one cruise to another, lead to the hypothesis that ~~the lamp-based~~ calibration of the GUVis instrument might introduce significant uncertainties and be responsible for the differences, given the importance of calibration for the AOD accuracy (see Alexandrov et al., 2002; Witthuhn et al., 2017).

Despite the fact that the deviation of AOD between Microtops and GUVis due to forward scattering effects of aerosol is partially compensated by the processing update of GUVis (see Sect. A), a remaining linear dependence of the bias has been observed (Sect. 4.1), which can most likely be attributed to the wide shadowband of the GUVis instrument, and the resulting difference in the field of view of both instruments. If AOD increases, this effect increases due to the enhanced circum-solar radiation. Although this effect does not have a major impact on the correlation in the direct comparison of Microtops and GUVis AOD datasets of this study (see Sect. 4.1), it introduces a substantial relative bias, and needs to be compensated ~~for~~ to ensure the consistency of both shipborne datasets, and for the comparison to the satellite and model datasets. The compensation is done using a linear scaling factor for measured AOD (S), as is explained later in this section.

To improve the consistency of the GUVis and MIC datasets, the following approach has been adopted to both transfer the calibration from the MIC instrument to the GUVis instrument, and to empirically correct for the effects of forward scattering. The first correction is accomplished following the method introduced by Alexandrov et al. (2002) for the Multi-Filter Rotating Shadowband Radiometer (MFRSR). The spectral direct irradiance measured by the GUVis can be represented by the equation:

$$I_i = C_i\, I_i^0 \exp\left(-\frac{\tau_i}{\mu_0}\right), \tag{5}$$

where I_i^0 and I_i are the spectral direct irradiance at top of atmosphere and surface, respectively, for a spectral channel i. The inverse of the airmass is denoted by μ_0, the cosine of the solar zenith angle. τ_i is the atmospheric column extinction optical depth for a spectral channel i. Following Alexandrov et al. (2002), a correction factor C_i for the calibration is introduced.

The absolute calibration of GUVis spectral channels is carried out in the laboratory to obtain the channel-specific calibration factors (k_i, [$\mathrm{V\,W^{-1}m^2nm^1}$]) for the conversion of the measured voltage (V_i) to spectral irradiance (I_i):

$$I_i = \frac{V_i}{k_i}. \tag{6}$$

The relation of the calibration factor k_i and the correction C_i can be obtained from Eq.(5) as:

$$ks_i = k_i\, C_i, \tag{7}$$

where ks_i denotes a corrected calibration factor.

τ_i can be expressed as the sum of AOD $\tau_{A,i}$, and remaining contributions to the atmospheric optical depth ($\tilde{\tau}_i$) from Rayleigh scattering and gaseous absorption as:

$$\tau_i = \tau_{A,i} + \tilde{\tau}_i. \tag{8}$$

The AOD can now be obtained from Eq.(5) and Eq.(8) as:

$$\tau_{A,i} = -\mu_0 \ln\left(\frac{I_i}{I_i^0}\right) + \mu_0 \ln(C_i) - \tilde{\tau}_i. \tag{9}$$

This equation shows that the calibration correction factor C_i introduces a change in AOD which is proportional to the product of the cosine of the solar zenith angle, and the logarithm of the correction factor. Introducing also a linear scaling factor S_i for the AOD to account for the effects of aerosol forward scattering (see Wood et al., 2017; di Sarra et al., 2015), the following correction equation is used here in a ~~bilinear~~ bi-linear fit, using μ_0 and the GUVis-based AOD $\tau_{\mathrm{GUV},A,i}$ as dependent variables, and the MIC-based AOD $\tau_{\mathrm{MIC},A,i}$ as independent variable:

$$\tau_{\mathrm{MIC},A,i} = \mu_0\, c_i + S_i\, \tau_{\mathrm{GUV},A,i}. \tag{10}$$

Thus, the scaling factor S_i and the calibration correction factor $C_i = \exp(c_i)$ can be obtained simultaneously from this ~~bi linear~~ bi-linear fit.

In the approach adopted for this study, the factor C_i has been determined independently for each of the five *Polarstern* cruises (PS83, PS95, PS98, PS102, PS113), in order to account for potential temporal changes in calibration between the different ship cruises, while a single constant value is assumed for S_i. The correction factors obtained by multi-linear regression based on Eq. (10) (C_{ij}, S_i) are listed in Table 3. Excluding individual cruises from the regression has been found to cause only negligible influence on the remaining coefficients, confirming the stability of this correction approach. In addition, adding either a constant or quadratic correction term such as used by di Sarra et al. (2015) does not lead to a significantly improved fit quality, and has thus not been used.

The final procedure adopted here for the correction of GUVis AOD is done in the following steps:

(i) First, the closest GUVis and MIC data points regarding time of measurement are selected for comparison within a time frame of 30 min.

(ii) If the deviation of the AOD pair exceeds the uncertainty estimate of ±0.02 of both instruments, the data pair is flagged as an outlier.

(iii) The fit coefficients (C_{ij}, S_i) are calculated based on Eq. (10) from the GUVis and MIC AOD~~without considering outliers~~. In this fit, multiple values of C_{ij} are obtained for separated cruises j, whereas a single value of S_i is assumed for all data.

(iv) Based on both correction coefficients, a corrected AOD is calculated from the GUVis measurements.

The cross-calibrated and scaled dataset is denoted in the following text as *enhanced* dataset GUVisE.

4 Results and discussion

This section presents and discusses the results of this study. First, the shipborne reference datasets are compared (Sect. 4.1). Second, the satellite aerosol products are evaluated against the shipborne reference datasets (Sect. 4.2). Lastly, the evaluation of the CAMS RA aerosol data is presented in Sect. 4.3.

4.1 Shipborne datasets comparison

An evaluation of the AOD product of the GUVis shadowband radiometer compared to the Microtops sunphotometer as reference was previously described by Witthuhn et al. (2017), considering one cruise of the research vessel (RV) *Polarstern* (PS83). This study extends the comparison to include four additional cruises with the RV *Polarstern* (comprising PS83, PS95, PS98, PS102 and PS113) (see Fig. 2). Regarding the comparability of both datasets, certain shortcomings are expected, as already mentioned. (i) Since the radiometers of both instruments utilize different calibration methods, and the spectral response of comparable channels might slightly differ, a deviation due to calibration is expected. (ii) Due to the different measurement methods of the sunphotometer and shadowband radiometer ~~, and in particular the wide shadowband and the resulting differences in the field of view,~~ an underestimation of AOD is expected for the GUVis instrument~~, related to forward scattering of aerosols (Russell, 2004)~~.

Given the importance of calibration for the AOD accuracy of the GUVis (see Witthuhn et al., 2017), only the calibration difference of both instruments is corrected for first, based on the method presented in Alexandrov et al. (2002) (see Sect. 3.4). The correction factor C_{ij} for each spectral channel i and each cruise of RV *Polarstern* j is given in Table 3. The Microtops ~~observations can serve as a reliable reference for calibration , given that the lamp-based calibration of the GUVis instrument might introduce significant uncertainties and the Microtops calibration~~ calibration is considered as consistent and trustworthy, due to ~~tracebility~~ traceability to the mature AERONET retrieval and calibration process. Therefore, it serves as the calibration reference. In the following, all versions of the GUVis datasets are calibration–corrected towards the Microtops by the method presented in Sect. 3.4. The correction of AOD with the linear scaling factor S is only applied to GUVisE.

The extended comparison is presented in the top part of Table 4. The GUVis irradiance data is first processed with the original algorithm used in Witthuhn et al. (2017). The correlation (R>0.95) found for all spectral channels comparing GUVis (old processing) and MIC generally confirms the findings of Witthuhn et al. (2017). However, the goal of an outlier ratio below 5% (see Table 4) as well as the weighted LOA within ±1.96 to verify the uncertainty estimate of ±0.02 for GUVis is missed with the old processing algorithm. As expected, an underestimation of AOD measured by the GUVis is reflected in the negative bias of ~~-0.02~~ -0.03. Since the observations are performed over ocean, the dominant aerosol conditions are maritime or desert dust (mineral dust transport) from the Sahara desert (see Fig. 1 and Table 2), which significantly differ in their forward scattering behaviour. Comparing sunphotometers with a narrow field of view to measurements with shadowband radiometers with a wide shading angle the influence of the forward scattering of the aerosol causes an underestimation of AOD of the shadowband radiometer (Russell, 2004). This has previously been confirmed by di Sarra et al. (2015) for the MFRSR as well as for the autonomous marine hyperspectral radiometers presented by Wood et al. (2017).

The GUVis processing algorithm has received an substantial update to improve the data quality and to compensate for the underestimation of aerosol forward scattering. This update is described in detail in the appendix section (Sect. A). The GUVis AOD data of all cruises with the RV *Polarstern* have been reprocessed with the new algorithm, and the resulting improvement of the measured AOD compared to the Microtops is also shown in Table 4. The correlation of GUVis AOD compared to MIC increases from >0.954 to >0.988 for all channels, indicating that any non–linear deviations due to aerosol forward scattering and other effects (di Sarra et al., 2015) have been substantially reduced. The underestimation of AOD is still present, indicated by a negative bias of -0.02. The uncertainty estimate of ±0.02 can be verified for spectral channels with wavelengths larger than 500 nm, since the statistics show a weighted LOA within ±1.96 (see Sect. 3.3). The uncertainty of GUVis AOD increases with decreasing wavelengths, indicated by the increase of outlier percentage and LOA.

The difference (D) of GUVis and MIC AOD shows a higher linear correlation as |R(D)| increases from >0.4 to >0.6 going from the old to new processing. As also shown by Fig. 3 panel (a), the underestimation of GUVis AOD increases linearly with increasing AOD. This linear dependence is here attributed to the difference of field of view of both instruments. If AOD increases, the effect of the circum-solar radiation due to differences in the field of view will increase. Since GUVis utilizes a broad shadowband resulting in a shading angle of 12° to 15° compared to the field of view of Microtops of 2.5°, this effect results in an underestimation of AOD for the GUVis radiometer.

The lower part of Table 4 shows the results of the comparison of the enhanced GUVis dataset GUVisE and MIC, which includes both the calibration and forward-scattering corrections. The expected uncertainty of ±0.02 is verified again, as the values of the weighted LOA are all within ±1.96 (see Sect. 3.3). This is also shown in Fig. 3 panel (b), where the LOA falls within the uncertainty limits. In addition, the outlier percentage is close to zero, indicating a close agreement of MIC and GUVisE. The correlation increases to >0.992 for all comparable channels. ~~The MIC and GUVisE datasets are thus~~ While the corrections introduced here need to be reconfirmed based on future observations, they are able to reconcile the observed differences between the MIC and the GUV products for the currently available observational data. Hence, we consider the MIC and GUVisE datasets as consistent, due to their strong linear correlation, and their agreement within the individual uncertainty limits. Therefore, the datasets are used as reliable ground–based reference datasets in the following.

4.2 Satellite product evaluation

~~The MODIS aerosol products have been extensively validatedin a large number of previous studies (e.g., Abdou et al., 2005; Bréon et al., 20~~ ~~. In contrast to most of these studies, the satellite observations are here compared over ocean to collocated shipborne observations of the Microtops sunphotometer.~~ In the following the comparison of the two MODIS and the SEVIRI aerosol products to collocated shipborne observations is shown.

(i) First, the satellite AOD at 550 nm validated. The wavelength of 550 nm is chosen as this channel is mainly used in previous validation studies (e.g., Abdou et al., 2005; Bréon et al., 2011; Shi et al., 2011; Anderson et al., 2012; Wei et al., 2019) and EE limits are defined for it (Levy et al., 2013). Since the SEVIRI dataset does not provide AOD at 550 nm, it was calculated with Eq. (1) using the AOD of 630 nm and the AE from the SEVIRI dataset.

Figure 4 ~~visualizes the comparison , showing results grouped in bins of 0.1 in AOD, and for an~~ shows the comparison of *MxD04_L2* and the SEVIRI AOD at 550 nm to the shipborne aerosol datasets within AOD bins of 0.1. The validation with respect to the ~~EE2~~ EE limits shows that the MODIS aerosol product meets the goals ~~set by the 67~~% ~~confidence interval~~ of both EE1 and EE2 compared to the Microtops dataset. As expected, the SEVIRI aerosol product shows a higher deviation versus Microtops than MODIS ~~, as the SEVIRI product utilizes a less complex scheme of aerosol models and only two spectral channels. The~~ and only meets the goal of 67% for EE2~~limits have been adjusted to account~~ , since it accounts for a general overestimation of ~~0.02 in AODby the MODIS product, which is also observed here~~satellite AOD. The SEVIRI ~~AOD~~ retrieval shows an even stronger tendency to overestimate AOD in comparison to the MIC reference dataset.

~~Comparing the MODIS- and SEVIRI-based AOD values to the GUVisE dataset, an overestimation of satellite-based AODversus GUVisE is observed.~~ The bias of satellite AOD also shows a dependence on the magnitude of the AOD. A positive bias (overestimation) is mostly found in situations with AOD values below 0.5, and decreases for larger AOD. This behavior is most evident in Fig. 4 panel (a) ~~shows a positive bias of satellite AOD for an AOD value of 0.3, turning towards a negative bias for higher AOD. This could be an artifact of the correction method for the GUVisE dataset, but a~~ and panel (b) as the reference datasets are GUVisE and COMB. A similar behavior also appears in the comparison to Microtops (panel (c)), although it is far less pronounced. ~~This suggests that the reasons for this behavior lies on the satellite side, and might be an artifact arising from the utilization of different aerosol models. While the Microtops dataset is rather diverse in observed aerosol conditions, the GUVisE dataset contains mostly maritime and desert dust situations. This is one potential explanation why the non-linearity is more pronounced when compared to the GUVisE dataset.~~

~~Comparing both satellite products to COMB (Fig. 4 panel (b)), a clear improvement is found compared to the GUVisE data. MODIS and SEVIRI agree equally well with COMB, and show the same dependence of~~ Since the satellite instruments measure reflected radiance the reflecting properties of the ~~bias as pointed out in panel (a). Since COMB is the combined product of MIC and GUVisE , it also contains only~~ ground used in the retrievals influence the retrieved AOD. Especially for clean atmosphere e.g. low AOD the influence of such parameters (e.g. surface albedo) is strong, since the values measured reflectance at TOA are close to the values of surface reflectance. For larger AOD values the uncertainty of those characterizations shrinks, therefore the overestimation of AOD decreases. Since the GUVisE and COMB datasets contain more maritime and desert dust ~~situations like GUVisE. Additionally, the COMB dataset consists of less data points than GUVis. Hence, it lacks statistical significance, so the difference in the results needs to be considered with caution~~cases than the MIC, this behavior is strongly visible.

~~The comparisons displayed in~~ Fig. ~~4 show individually collocated satellite data versus Microtops. Fig.~~ 5 presents the same comparison as Fig. 4, but simultaneous availability of data from all datasets (SEVIRI, MODIS, MIC) is required to preclude differences arising from a different sampling of cases. Therefore, the accuracy of SEVIRI and MODIS are directly comparable with respect to the MIC reference. This comparison shows that both AODs retrieved from SEVIRI and MODIS agree well with the shipborne reference, although the non-linear behavior of over- and underestimation is more pronounced for the SEVIRI retrieval. Since 550 nm is not a native spectral channel of SEVIRI, increased deviations in AOD are expected~~due to~~ , since the uncertainty of ~~the AE calculated with the two spectral channels 630 and 810~~ nm ~~and used for extrapolation~~AE calculated from SEVIRI native channels is high as shown later. Therefore, a strong improvement of agreement is found comparing the

630 nm AOD to the shipborne reference as the SEVIRI AE is not used for calculation in Fig. ??. Considering the AOD at 6305 panel nm for the comparison, the SEVIRI performance is similar to MODIS, and even slightly better when compared to COMB and GUVisE, since the SEVIRI AOD is generally slightly lower than the (b). At lower AOD values SEVIRI AOD is close to the MODIS AOD.

The results shown in Fig. 4, Fig. 5 and Fig. ?? are also presented in Table 55 summarizes the results of the AOD validation at 550 nm. For the statistics presented in this table the SEVIRI AOD is calculated using MIC AE since using the native AE leads to high uncertainties. Generally, the satellite-based AOD is higher compared to the shipborne reference datasets, as is reflected in the bias > 0 for all comparisons. selections except for larger AOD values, where the SEVIRI bias turns negative. The MODIS aerosol products show the highest linear correlation , which also slightly exceed those for the CAMS RA. (0.93 for *MxD04_L2*, 0.95 for *MxD04_3K*correlation is slightly higher than that) and lowest values for LOA. LOA values are even slightly less than for the *MxD04_L2* product, and the LOA is nearly equal for both MODIS products. Thus, this finding does not confirm the expectation of higher noise in the 3 km versus the 10 km product of MODIS expressed in Levy et al. (2015). Our comparison of *MxD04_3K* and *MxD04_L2* indicates a slightly lower AOD calculated in the former product, which would presumably represent maritime conditions better. The analysis has also been repeated separately for the MODIS datasets based on the Terra and Aqua satellites (not shown), but only minor differences in the evaluation statistics for the individual satellites was found. Thus, only the combined MODIS dataset from Terra and AQUA are Aqua is presented here. It also has to be stressed that the considered dataset is still relatively small compared to other validation studies, and should be repeated if more reference data becomes available. Nevertheless, the correlations found here agree well with the findings of Levy et al. (2013) for considering the MODIS C6.1 aerosol products (0.93) and for product (0.937) and the 550 nm channel, and exceed the findings of Bréon et al. (2011) for the SEVIRI aerosol product nm channel. A smaller dataset of Microtops observations was compared to MODIS aerosol products by Kharol et al. (2011), where a general overestimation of AOD, and a high correlation was found similar to our results. For SEVIRI the findings exceed the values found in the study of Bréon et al. (2011) at 630 nm over ocean (0.795 versus 0.9 0.88 in this study), indicating a significantly better performance over ocean than over land. The results for SEVIRI and MODIS show a similar agreement of the AOD compared to the reference data, but with a larger scatter of ±0.140.15 LOA for SEVIRI versus ±0.110.13 LOA for MODIS AOD. Also, SEVIRI AOD shows higher bias values for AOD < 0.4 and negative bias for AOD values ≥ 0.4.

Besides the AOD (ii) Second, the AE of calculated from the satellite products are evaluated as a quantity characterizing the spectral dependence of AOD, as shown in is validated. We chose to calculate the AE for each aerosol product from the channels matching closest to the wavelengths 440 nm and 870 nm with Eq. (1). This obviously leads to increased uncertainties for the SEVIRI product but also demonstrates its limitations.

Fig. 6 . The panels display shows the comparison of the difference of AE versus the different shipborne reference datasets as a scatter plot, indicating the bias, as well as the LOA and EE limits. The EE for AE is estimated to be ±0.4 for the MODIS products (Levy et al., 2010, 2013), and the same EE is applied for the SEVIRI AE product. In general, the MODIS AE agrees with this estimate of the EE limits, but shows a tendency to overestimate the shipborne AE, as reflected by the positive bias. The bimodal behaviour of AE of the MODIS products found for C5 in Levy et al. (2010) is not reproduced here, which agrees

with the findings for C6.1 presented in Levy et al. (2013). Also, MODIS AE meets the expected Gfrac of >67% for an EE of ±0.4, as was already found in Levy et al. (2013).

The results for the SEVIRI AE show a general overestimation versus MIC, indicated by the positive bias. Furthermore, a bimodal behaviour of AE is found similar to that reported for the C5 MODIS products in Levy et al. (2010). The SEVIRI-based AE mostly lies close to two values: AE close to zero is associated with the models of oceanic and maritime aerosol used by the retrieval (O99, M99, (Shettle and Fenn, 1979)). Another large fraction of the dataset is related to purely tropospheric aerosol models (type T99, T90, T50, (Shettle and Fenn, 1979)), covering AE from 1.29 to 1.61. Another frequent assignment of aerosol model are those for very small particles which cover the AE range from 1.8 to 2.4 (Thieuleux et al., 2005). Therefore, it can be concluded that SEVIRI retrieval of AE cannot realistically capture the variability in the AE which is observed from shipborne products, likely due to the limitation of using only two relatively closely spaced spectral channels of SEVIRI. Thus, calculation of AOD extrapolated to wavelengths outside the range of the native channel 630 nm and 810 nm from the SEVIRI productbased on the AE may lead to high uncertainties.

which is expected given the limitations of the product. It should be noted that the results comparing AE from satellite to MIC can be reproduced with the COMB dataset (Fig. 6 panel (a) and panel (b)), although the number of collocated measurements is small. Thus, this study should be extended with data from additional cruises in future.

After the general evaluation of (iii) Third, the representation for different aerosol conditions within the satellite-based AOD and AE products , their representation for different aerosol conditions is investigated. Since the data has been acquired over the Atlantic ocean, the most prominent conditions contain maritime or desert dust aerosol. To examine the representation of AOD and AE with respect to aerosol type, the layout presented in Toledano et al. (2007) is used for example in Fig. 7 and Fig. 8. Instead of the AOD at 440 nm (as chosen by Toledano et al. (2007)), the wavelength of 630 nm is chosenchoosen, to match the SEVIRI channel at 630native SEVIRI channel. Also we restrict the comparison choosing only datapoints where AOD at 810 nm >0.05 to avoid uncertainties from calculating AE from low AOD values. The aerosol type is classified by applying the Toledano et al. (2007) scheme to the MIC data . based on the MIC data (see Sect. 3.1). Points related to a certain aerosol type are combined in the form of a covariance ellipse which spans 67 % of the related data points.

Fig. 7 shows that overall, the AOD of the different products and instruments lie very close together, with a slight tendency to overestimate the AOD for desert dust (only MODIS products) or maritime aerosol types. In general, the satellite-based datasets overestimate the AE. These results confirm the statistics discussed before. The satellite ellipses are tilted compared to the MIC ellipses, as a result of the assumed relation of AOD and AE in the retrieval, which is determined by the choice of aerosol model. This effect is strongly visible for SEVIRI in desert dust situations situations other than maritime, because the AE is calculated from the 630 nm and 810 nm native channels only. Satellite SEVIRI AOD in maritime conditions exhibits a stronger overestimation as also shown in Table ?? and Table ??5. This effect might be related to the coarser spatial resolution of the satellite pixels or undetected cloud contamination (see Sect. 2.2). The spatial-mean mean AOD inferred from satellite pixels can deviate from the AOD which is retrieved from slant transmission in case of MIC, due to the mismatch of spatial scales.

The most prominent feature of Fig. 7 is the deviation in AE for desert dust conditions, and to a lesser extend for the mixed type aerosol. When compared to the shipborne products, satellite products show an AE for desert dust situations which is more than two times larger for MODIS, and more than three times larger than for SEVIRI. This relates to a lack of realistic mineral dust models in the satellite retrievals. ~~Since the MODIS product uses a larger set of spectral channels and aerosol models, it is able to estimate AE more accurately.~~ This emphasize, that the Ångström behaviour is not applicable for desert dust conditions, at least with a limited set of spectral channels.

Although, the AE is still the method of choice for extrapolating the AOD at the desired wavelengths to validate or increase observation capabilities, as is done in several studies (Kleidman et al., 2005). ~~Therefore, one should be aware that during desert dust conditions, only the AOD from satellite products is accurate at available spectral channels only and that extrapolating AOD at other wavelengths using AE may lead to unexpected high uncertainties.~~

Figure 8 confirms above findings. In this figure, each dataset has been collocated individually to the MIC reference to increase the diversity of conditions. No noteworthy discrepancies are found for ~~other aerosol typesthan maritime and desert dust~~continental aerosol types. With the exception of the positive bias in AOD especially at lower values of AOD, and the overestimation of AE in particular for desert dust and therefore to some extend also mixed aerosol, the satellite aerosol products are found to agree closely to MIC, in particular for continental ~~as well as biomass burning~~ aerosol.

The previous statistics confirm that the AOD retrieved from satellite agrees well with the shipborne reference, but slightly overestimate AOD in general and especially at low AOD. AE is also overestimated for maritime and especial desert dust aerosol. Therefore, AOD is only represented well for the native spectral channels of the satellite instruments. The estimation of the spectral behaviour of AOD remains challenging, due to the lack of realism of the aerosol models (MODIS and SEVIRI), or the number of spectral channels available (SEVIRI). These findings are in particular applicable for conditions dominated by mineral dust.

Last, we investigate the value of increased temporal resolution within the SEVIRI aerosol product versus MODIS products. While the MODIS aerosol product is clearly the product of choice for many applications, i.e. for data assimilation and climate studies, due to its accuracy, availability and global coverage, the SEVIRI aerosol product is still of scientific interest due to its high temporal resolution of 15 min ~~Bréon et al. (2011)~~(Bréon et al., 2011). The high temporal resolution however ~~only~~ adds information compared to products from polar-orbiting satellites, if the temporal variations of aerosol properties since the last overpass of a polar-orbiting satellite exceed the error limits of the retrieval. Thus, it is not clear how much information can actually be gained from the higher temporal resolution of SEVIRI, as it is expected that AOD variation are generally small on the time scale of hours. To further investigate this point, MODIS collocations with the shipborne datasets are used to serve as random samples to study the AOD variability between successive overpasses. For each pixel of ~~follow up MODIS images~~ a MODIS image the corresponding SEVIRI AOD for every available SEVIRI image between ~~both MODIS images~~ overlapping MODIS images of consecutive Terra and Aqua overpasses was acquired to calculate the AOD variation. Relative to the linear regression line of the ~~follow up MODIS AOD~~ MODIS AOD of each overpass, the standard deviation (STD) of SEVIRI AOD was calculated. Therefore, STD is a measure of the additional variation of AOD which cannot be seen in a MODIS only AOD product. Fig. 9 shows the STD calculated for different time intervals between ~~follow up~~ consecutive MODIS images. The mean STD of AOD within six hours is slightly larger than 0.02. The STD is compared to the mean EE2 calculated using Eq.(4) and mean SEVIRI AOD, indicated by the green dashed line in Fig. 9. If the STD is larger than EE2 it points out a situation where

AOD variation cannot be captured by MODIS and can be called significant. SEVIRI aerosol measurements add information to the general AOD monitoring only if the AOD variation is significant. As Fig. 9 reveals, this is only true for slightly above 8% of all situations, knowing that, in general, Terra and Aqua satellites overpass the same region every three hours. This emphasises that, in terms of climate studies or data assimilation, the significant higher temporal resolution of SEVIRI does not lead to improvements for the majority of situations, unless the accuracy of this product could be significantly improved. In fact, such an improvement ~~could in future be possible~~ will be possible in future with the third generation of Meteosat (MTG). ~~Nevertheless, with~~ We are aware that the analyses presented here do not provide a complete picture of the AOD variability over the full diurnal cycle. It was only possible to analyse the variability between daytime overpasses of MODIS. Continuous evaluation of the daily cycle of AOD are only possible with geostationary satellites such as SEVIRI. With the high temporal resolution, the SEVIRI product ~~may be~~ are needed for many applications, such as ~~case studies of~~ extreme events such as dust or smoke plume development, where high variability of AOD is expected.

4.3 CAMS RA evaluation

Alongside the evaluation of satellite aerosol products described above, results for the CAMS RA AOD are presented in Table 5.

In comparison to MIC as reference dataset, the Table 5 shows that CAMS RA AOD agrees closely to MIC, since the correlation is 0.92 and the bias is about zero. The LOA of ±0.13 is similar to the one found for the products of SEVIRI and MODIS compared to MIC, with values ranging from ±0.12 to ±0.15. The ~~correlation of the AOD difference R(D) is close to zero, indicating that the difference of~~ CAMS RA ~~and MIC AOD is not linearly related. Compared to the statistics calculated for the satellite products, shows that CAMS RA performance is clearly superior to the SEVIRI dataset, since the correlation is increased from 0.90 to~~ outperforms the SEVIRI aerosol dataset in all presented statistical measures at least slightly (e.g., correlation 0.88 versus 0.92 ~~as well as the LOA from 0.15 to~~ or LOA of 0.13 versus 0.15). Further, the ~~overestimation of AOD found for the satellite products, indicated by the bias of about 0.03, is not present in the~~ bias of AOD and its dependency on AOD is reduced for the CAMS RA ~~AOD. This emphasize that although~~ product as it shows low bias values for both low and high AOD values. This is expected since the MODIS AOD ~~is assimilated, the overestimation of AOD is compensated in CAMS RA~~bias must be corrected before assimilation into the reanalysis product. This effect is clearly shown in Fig. 10, together with a tendency of CAMS RA towards an underestimation of AOD for larger values of AOD. ~~As the evaluation shown in Table 5 compares each aerosoldataset individually against each reference dataset, it offers the largest amount of collocated data and is best for obtaining the individual statistics. The performance of each product is however not directly inter-comparable among the datasets, due to potential differences arising from different sampling of conditions. To address this point, Table **??** is given, which shows the same statistics as in Table 5, but requires simultaneous availability of all datasets for enabling a direct inter-comparison. Table **??** confirms the conclusions drawn from Table 5. (i) CAMS RA compensates for the overestimation of AOD provided by MODIS aerosol products. The bias calculated for~~ For maritime aerosol, CAMS ~~RA equals zero in this dataset. (ii)~~ RA AOD has lowest correlation of 0.7. The values of LOA are lowest for maritime, which is expected since this measure favors lower AOD and maritime aerosol situations are generally connected to low AOD values. A slight overestimation of 0.01 is shown by the bias considering only maritime aerosol situations, which is lower than the bias found for the MODIS products. For desert dust conditions, the correlation of CAMS ~~RA AOD shows larger scatter, indicated by larger LOA and less correlation than the MODIS products. Therefore, in terms of consistency, the MODIS aerosol products show better performance than~~ RA to MIC (0.85) is similar to the one found for *MxD04_L2* (0.87) in Table 5, although the correlation of *MxD04_3K* is largest with 0.92. As for maritime aerosol, the overestimation of AOD is compensated in the CAMS RA ~~. (iii)~~ aerosol product. This emphasizes that the CAMS RA ~~performance is clearly superior to the SEVIRI dataset, with an increased correlation and no overestimation of AOD~~aerosol product is comparable in accuracy to the MODIS products in maritime and desert dust situations.

In terms of assimilated aerosol observations, data from MODIS and AATSR are used by the IFS for CAMS RA starting from the year 2003 until March 2012, when the ENVISAT mission ended due to loss of contact to the satellite. After March 2012, only the MODIS AOD is used (Inness et al., 2019). Table ~~**??**~~ 5 shows the evaluation of CAMS RA versus MIC for the different time periods, to investigate potential differences in quality. ~~Since CAMS RA can provide AOD regardless of cloud cover and satellite orbit, a comparison of CAMS RA AOD to MIC conditioned on the availability of collocated MODIS data is also shown. The results show no significant difference in terms of correlation, bias or LOA.~~ Without AATSR, MODIS is the only contributor for data assimilation in terms of AOD. Comparing the results of CAMS RA with and without AATSR, the performance of CMAS RA with additional AATSR data is increased, indicated by increased correlation from 0.87 to 0.90 and lower LOA, dropping from 0.18 to 0.14. This shows that the AATSR observations lead to an improvement of the representation of aerosol in CAMS RA. Inness et al. (2019) suspected ~~an~~ a slight increase of CAMS RA AOD without AATSR, which cannot be observed in this study. As the analysis presented here is based on a limited number of data points, it is unclear whether these findings are statistically significant, and the discussed tendencies should be considered with caution.

~~To evaluate CAMS RA AOD with respect to the representation of aerosol type, Table **??** lists evaluation statistics calculated only for maritime and only for desert dust aerosol type, in comparison to the statistics of the whole dataset of CAMS RA versus MIC. For maritime aerosol, CAMS RA AOD has lowest correlation of 0.7. The values of LOA are lowest for maritime, which is expected since this measure favors lower AOD and maritime aerosol situations are generally connected to low AOD values. A slight overestimation of 0.01 is shown by the bias considering only maritime aerosol situations. Compared to MODIS for maritime aerosol type, the overestimation found for CAMS RA AOD in Table **??** is less significant as shown in Table **??**. This emphasizes that CAMS RA exceeds the MODIS accuracy for maritime aerosol conditions in terms of AOD. For desert dust conditions, the correlation of CAMS RA to MIC (0.85) is similar to the one found for *MxD04_L2* (0.86) in Table **??**, although the correlation of *MxD04_3K* is largest with 0.92. As for maritime aerosol, the overestimation of AOD is compensated in the CAMS RA aerosol product. This emphasizes that the CAMS RA aerosol product is comparable in accuracy to the MODIS products in maritime and desert dust situations.~~

Inness et al. (2019) reported an overestimation of AE in CAMS RA compared to AERONET stations of about 5-20%. From the comparison to MIC presented in Fig. 6 (panel (f)), the same conclusion can be drawn based on our dataset, showing a positive bias of 0.17. Compared to MODIS, similar values are found for Gfrac, but while the MODIS AE scatters more equally around the reference AE, CAMS RA AE is clearly distributed above zero. Also, the AE difference of CAMS RA and MIC shows a increased linear dependency indicated by increased correlation R(D). This indicates that similar to the SEVIRI product, certain aerosol models are favored in the processing. Nevertheless, the overall scatter of AE indicated by the values of LOA is lower for CAMS RA.

As stated by Inness et al. (2019), the overestimation of AE results from a deficit in the handling of the coarse dust fraction in the model. The total AOD calculated in CAMS RA is composed of less dust than in its predecessor versions, which explains the higher overall AE. Nevertheless, the comparison of CAMS RA AE with respect to aerosol type in Fig. 11 reveals that the AE for desert dust agrees best with the MIC reference, compared to the satellite products. The slightly better representation compared even to MODIS indicates that the representation of the spectral dependence of AOD for dust is most realistic in CAMS RA. For maritime aerosol, CAMS RA AE shows a similar overestimation compared to the satellite products, but with less scatter. This emphasises a more consistent representation of maritime aerosol in CAMS RA as compared to satellite products. The CAMS RA AE representation versus MIC in Fig. 8 shows a close agreement for all aerosol types, except an overestimation of AE in maritime conditions, and a tendency for overestimation during dust conditions with low AOD. In general, the AE and AOD of CAMS RA is similar or in some instances even exceeds the accuracy of the satellite retrievals including MODIS in comparison to the reference data presented in this study.

5 Conclusions

Within this paper, a comprehensive evaluation of MODIS and SEVIRI AOD products as well as the representation of AOD in the CAMS reanalysis has been presented with shipborne reference datasets. For this purpose, available Microtops observations from MAN across the Atlantic ocean were utilized, and complemented by a unique set of shipborne aerosol products collected during five Atlantic transit cruises of RV *Polarstern* with the multi spectral shadowband radiometer GUVis-3511.

Three separate aspects have been investigated within the study:

(i) First, the two shipborne datasets were inter-compared to verify their consistency. Extending the comparison presented in Witthuhn et al. (2017), the AOD derived from the GUVis and Microtops instruments from five cruises with RV *Polarstern* were compared. A substantial update of the GUVis processing algorithm is shown to address several shortcomings identified in the prior version. To improve upon the lamp-based instrumental calibration of the GUVis, the method of Alexandrov et al. (2002) has been applied to obtain a cross-calibration based on the MIC observations. In addition, an underestimation of AOD by the GUVis instrument compared to the MIC has been observed, which is related to strong forward scattering by aerosol, and arises from the broad shadowband and the much wider effective field of view, compared to the MIC observations (Russell, 2004). Combining the cross-calibration with an empirical correction following the approach of di Sarra et al. (2015) and Wood et al. (2017), a correlation >0.992 is found for all spectral channels. The uncertainty estimate of ± 0.02 for the GUVis AOD is shown to be valid after applying these two corrections.

Compared to the manually operated Microtops instrument, an important advantage of the GUVis dataset is its high temporal resolution as well as the uniformity of sampling. These automated shipborne measurements lead to a larger collocated dataset for satellite evaluation, which in turn leads to more robust evaluation statistics. They also offer the chance to conduct such observations on more cruises, as they greatly reduce the amount of effort to operate the instrument.

(ii) Second, the shipborne datasets have been utilized to evaluate the MODIS *MxD04* and SEVIRI AOD products. The satellite products differ in temporal and spatial resolution as well as in number of spectral channels available from the satellite instruments. The AOD has been compared at 550 nm (used in previous validation studies) and at 630 nm, the latter being a native channel of the SEVIRI instrument, enabling a consistent and fair evaluation of the aerosol products from both satellite sensors. For non-native channels, interpolation of AOD based on the AE using the Ångström relation have been used. The AOD at these two wavelengths have been compared with collocated Microtops measurements and show similar agreement, although the comparison to SEVIRI AOD shows larger scatter (about 25%) and therefore less correlation than the one to MODIS. Also it was shown that the bias of SEVIRI AOD is dependent on the AOD, as the bias for AOD < 0.4 is larger than for the MODIS product, and the bias for AOD ≥ 0.4 turns negative (underestimation).

Previous evaluation studies of the MODIS aerosol products have utilized the EE limits at 550 nm (defined as spanning at least 67% of the data) of $\pm(0.03{+}0.05\mathrm{AOD})$ (EE1) (e.g., Abdou et al., 2005; Remer et al., 2008; Livingston et al., 2014) and $[+(0.04+0.1\,\mathrm{AOD}), -(0.02+0.1\,\mathrm{AOD})]$ (EE2) by Levy et al. (2013) over ocean. The EE1 limits are missed slightly by the 67% criterion, while the EE2 limits are confirmed by this study. The EE2 limits account for a general overestimation of AOD by the MODIS satellite products of about 0.02, which is close to the value of 0.03 found here. The SEVIRI aerosol product also meets the EE2 limits for the interpolated AOD at 550 nm.

Since SEVIRI has ~~only channels~~ channels only at wavelengths of 630 and 810 nm, which lie relatively close together, the accuracy for calculating the AE is significantly degraded. The representation of the spectral dependence of AOD therefore is superior in the MODIS products, due to the large set of available spectral channels, combined with the mature set of aerosol models used in the retrievals. This manifests itself in a consistent accuracy of AOD for all available spectral channels utilizing the AE for the MODIS product, which is not the case for SEVIRI.

Evaluating the satellite products with a focus on aerosol type reveals that the main challenge arises from the identification of realistic aerosol models for use in the retrieval (for both MODIS and SEVIRI), and from the limited number of spectral channels (for SEVIRI). Therefore, the AOD and AE from the SEVIRI product should not be used to extrapolate the AOD to wavelengths outside the available spectral channel range. Given the large number of channels, the MODIS AOD at non-native wavelengths is significantly more accurate than that of the SEVIRI product, but still relies on the underlying aerosol model, which can introduce uncertainties depending on aerosol conditions. In particular, the AE calculated from satellites during aerosol conditions dominated by mineral dust aerosol shows values which are two times (MODIS) and three times (SEVIRI) larger than the AE from shipborne products.

~~Avoiding spectral conversions, our~~ Our results confirm that satellite products can ~~serve well to~~ provide a global view of ~~AOD~~ the spatiotemporal aerosol distribution e.g ~~.~~ for climate studies or model assimilation, as long as their error limits are properly taken into account, and spectral extrapolation of products is avoided. This finding is consistent with results of former validation studies for the MODIS instrument (e.g., Munchak et al., 2013; Levy et al., 2013; Livingston et al., 2014). The quality of MODIS products is continuously monitored over land by comparison with products from worldwide AERONET stations.

In most situations, temporal variations of AOD within a window of six hours are smaller than the uncertainty limits of the satellite products. Hence, the better time resolution of SEVIRI and other geostationary satellite sensors offers ~~only minor~~

~~benefits~~ minor benefits for climatological studies compared to the use of polar-orbiting satellite platforms, given its increased uncertainties. ~~Nevertheless, the~~ The SEVIRI AOD product ~~can still~~ provide valuable information on the temporal evolution of AOD ~~fields for specific~~ when the aerosol changes rapidly. Specific cases with high temporal variability ~~such as dust storms~~are dust storms, plumes of volcanic ash or the passing of frontal systems.

(iii) Finally, the aerosol fields obtained from the CAMS RA have been evaluated versus collocated Microtops measurements. The performance of CAMS RA is rather close to that of the MODIS product. The differences of MODIS and CAMS RA arise mainly from the model handling of different aerosol types: while an overestimation of AOD observed for MODIS for maritime and desert dust aerosol is compensated in CAMS RA, the overall consistency of MODIS AOD exceeds CAMS RA AOD, indicated by larger correlation of MODIS AOD to the reference datasets.

Finally, it has to be noted that the evaluation presented here is still based on a relatively small set of collocated shipborne and satellite observations. For more meaningful results, a significantly larger shipborne dataset would be desirable.

6 Outlook

Ground-based and shipborne observations will continue to play an important role for monitoring and investigating aerosols at a global scale. Applications range from the evaluation and monitoring of satellite products to independent studies targeting radiative closure and aerosol processes, which cannot be resolved by satellite datasets. Shipborne observations of aerosol optical properties with the Microtops sunphotometer will continue within MAN, and will be complemented by the GUVis shadowband radiometer on future OCEANET cruises. The GUVis measures direct and diffuse irradiance simultaneously. It is thus well suited to extend the aerosol products by additional parameters such as single scattering albedo and asymmetry parameter utilizing the diffuse to direct ratio as outlined by Herman et al. (1975) and applied in a number of previous studies (e.g., Petters et al., 2003; Kassianov et al., 2007). It also offers the chance ~~of an evaluation of~~ to evaluate the direct radiative effect of aerosol and these observations could contribute towards improving climatological estimates of the aerosol radiative effect (e.g., Kinne, 2019).

Current efforts are also directed to operate state-of-the art Cimel sunphotometers on shipborne platforms (Yin et al., 2019). While the automatic operation of those instruments on a moving platform still poses a significant challenge, the high accuracy offered by sunphotometers combined with recent advances in navigation and alignment sensors make this seem a promising approach for the future. Fully automated Cimel sunphotometer observations on ship including the capability of sky scans will open up the full potential of the well-developed AERONET aerosol ~~product~~ products for studies over ocean. Nevertheless, the shadowband principle used by the GUVis instrument is less complex, and thus might contribute in parallel towards improved data availability of aerosol observations over ocean.

Alongside these ongoing effort in shipborne observations, a number of promising new satellite missions will be launched within the next ~~years~~few years collocated with an imager and radar, whose validation will increase the demand for reliable reference datasets.

(i) With the launch of Meteosat third generation (MTG) operated by EUMETSAT, SEVIRI on MSG will be replaced by the Flexible Combined Imager. This will lead to observations with an increased spatial resolution comparable to that of MODIS, but with the benefits of the geostationary satellite perspective (in particular in terms of temporal resolution). The set of available spectral channels will also increase, including channels at 440 and 510 nm wavelength. Since the AOD retrieved at SEVIRI's native spectral channels already has been shown to have satisfactory accuracy here, the availability of MTG observations should increase the accuracy of aerosol products to the level of MODIS, including significant improvements in aerosol model selection and AE calculation within the retrievals. The high temporal resolution of MTG will thus provide novel information on the spatio-temporal distribution of aerosols with MODIS-like accuracy, which will be valuable for studies targeting air quality or aerosol transport. These data are also expected to be useful for data assimilation into CAMS RA, and can provide information on temporal changes beyond the current time resolution of CAMS RA.

(ii) The Earth Cloud Aerosol and Radiation Explorer (EarthCARE) satellite will be launched by ESA in ~~2021.~~ 2022. This polar orbiting satellite mission utilizes a combination of instruments including the Multi-Spectral Imager (MSI) spectral radiometer system, which utilize spectral channels in the visible and near infrared region similar to the SEVIRI instrument. In addition, the Atmospheric Lidar (ATLID) will provide vertical profiles of extinction at 355 nm, and thereby reveal new information on the vertical distribution of aerosols and thin clouds. ~~This unique feature will benefit for~~ The combination of both instruments on a single satellite and the use of a high-spectral resolution lidar enabling direct observations of the aerosol extincition at 355nm is a unique feature and will benefit scientific studies targeting aerosols including their radiative effects. The synergy of the MSI and ATLID instruments will open up new opportunities for the retrieval and classification of aerosol properties, and will provide new insights on the vertical distribution of aerosol optical properties. Based on our findings, it seems particularly important to combine MSI and ATLID information to constrain the spectral dependence of aerosol properties, due to the limitations reported here arising for the SEVIRI wavelengths.

(iii) Following on from the EUMETSAT Polar System program (Metop), the second generation of European polar orbiting spacecraft (EPS-SG) will continue the meteorological observations in the morning orbit from 2022 onward. The Multi-Viewing Multi-Channel Multi-Polarisation Imaging (3MI) instrument on board this mission utilizes 12 spectral channels from 410 to 2130 nm and with a nadir resolution of 4 km. Together with information on light polarization, these observations will provide unique observations for the estimation and characterization of aerosol optical properties at a global scale.

With all these upcoming satellite observations, the consistency of the different aerosol products will become an important aspect for future analyses, in particular with respect to aerosol type. Reliable ground-based reference datasets will continue to play an important role for their evaluation, and for reconciling the unavoidable discrepancies between datasets.

Data availability. The datasets from all instruments used in this study are public available:

GUVis data of AOD and is available from the Pangaea database (https://doi.org/10.1594/PANGAEA.872377).

Microtops data is available over the MAN AERONET website (https://aeronet.gsfc.nasa.gov/new_web/maritime_aerosol_network.html).

MODIS aerosol products called *Effective_Optical_Depth_Average_Ocean* are available through the Level-1 and Atmosphere Archive & Distribution System (LAADS) Distributed Active Archive Center (DAAC) (https://ladsweb.modaps.eosdis.nasa.gov).

The SEVIRI aerosol products are available from the ICARE Data and Services Center (http://www.icare.univ-lille1.fr).

CAMS RA data can be acquired from the ECMWF public dataset catalogue (https://apps.ecmwf.int/datasets).

Data availability. The datasets from all instruments used in this study are public available:

GUVis data of spectral irradiance and AOD and is available from the Pangaea database (https://doi.pangaea.de/10.1594/PANGAEA.910535).

Microtops data is available over the MAN AERONET website (https://aeronet.gsfc.nasa.gov/new_web/maritime_aerosol_network.html).

MODIS aerosol products called *Effective_Optical_Depth_Average_Ocean* are available through the Level-1 and Atmosphere Archive & Distribution System (LAADS) Distributed Active Archive Center (DAAC) (https://ladsweb.modaps.eosdis.nasa.gov).

The SEVIRI aerosol products are available from the ICARE Data and Services Center (http://www.icare.univ-lille1.fr).

CAMS RA data can be acquired from the ECMWF public dataset catalogue (https://apps.ecmwf.int/datasets).

Author contributions. Jonas Witthuhn developed and implement processing scheme and conducted AOD calculation from the GUVis-3511 shadowband radiometer. Jonas Witthuhn calculated the evaluation statistics and prepared the manuscript. Anja Hünerbein and Hartwig Deneke contributed expertise about satellite remote sensing, provided helpful advice on the work and paper, and contributed to writing the manuscript.

Competing interests. The authors declare no competing interests.

Acknowledgements. We thank Stefan Kinne and Alexander Smirnov for their effort in maintaining and organisation of Microtops observations on RV *Polarstern*, as well as all scientists operating the Microtops during the cruises.

Thanks are also due to the Alfred Wegener Institute for Polar and Marine Research (AWI) for the opportunity to operate the GUVis instrument during the research cruises PS83, PS95, PS98, PS102 and PS113 across the Atlantic Ocean on RV *Polarstern*.

We thank the ICARE Data and Services Center for providing access to the data used in this study.

The Terra/MODIS and Aqua/MODIS Aerosol 5-Min L2 Swath 10 km and 3K Swath 3 km datasets (Levy et al., 2015) were acquired from the Level-2 and Atmosphere Archive & Distribution System (LAADS) Distributed Active Archive Center (DAAC), located in the Goddard Space Flight Center in Greenbelt, Maryland (https://ladsweb.nascom.nasa.gov/).

This paper contains modified Copernicus Atmosphere Monitoring Service information [2019], neither the European Commission nor ECMWF is responsible for any use that may be made of the information it contains.

We thank the reviewers for the detailed comments and thoughtful suggestions to improve this manuscript.

References

Abdou, W. A., Diner, D. J., Martonchik, J. V., Bruegge, C. J., Kahn, R. A., Gaitley, B. J., Crean, K. A., Remer, L. A., and Holben, B.: Comparison of coincident Multiangle Imaging Spectroradiometer and Moderate Resolution Imaging Spectroradiometer aerosol optical depths over land and ocean scenes containing Aerosol Robotic Network sites, Journal of Geophysical Research: Atmospheres, 110, https://doi.org/10.1029/2004JD004693, 2005.

Adames, A. F., Reynolds, M., Smirnov, A., Covert, D. S., and Ackerman, T. P.: Comparison of Moderate Resolution Imaging Spectroradiometer ocean aerosol retrievals with ship-based Sun photometer measurements from the Around the Americas expedition, Journal of Geophysical Research, 116, https://doi.org/10.1029/2010jd015440, 2011.

Alexandrov, M. D., Lacis, A. A., Carlson, B. E., and Cairns, B.: Remote Sensing of Atmospheric Aerosols and Trace Gases by Means of Multifilter Rotating Shadowband Radiometer. Part I: Retrieval Algorithm, Journal of the Atmospheric Sciences, 59, 524–543, https://doi.org/10.1175/1520-0469(2002)059<0524:rsoaaa>2.0.co;2, 2002.

Anderson, J. C., Wang, J., Zeng, J., Petrenko, M., Leptoukh, G. G., and Ichoku, C.: Accuracy assessment of Aqua-MODIS aerosol optical depth over coastal regions: importance of quality flag and sea surface wind speed, Atmospheric Measurement Techniques Discussions, 5, 5205–5243, https://doi.org/10.5194/amtd-5-5205-2012, 2012.

Bellouin, N., Boucher, O., Haywood, J., and Reddy, M. S.: Global estimate of aerosol direct radiative forcing from satellite measurements, Nature Sci, 438, 1138 – 1141, https://doi.org/10.1038/nature04348, 2005.

Bernard, E., Moulin, C., Ramon, D., Jolivet, D., Riedi, J., and Nicolas, J.-M.: Description and validation of an AOT product over land at the 0.6 μm channel of the SEVIRI sensor onboard MSG, Atmospheric Measurement Techniques, 4, 2543–2565, https://doi.org/10.5194/amt-4-2543-2011, 2011.

Bland, J. M. and Altman, D.: STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT, The Lancet, 327, 307–310, https://doi.org/https://doi.org/10.1016/S0140-6736(86)90837-8, originally published as Volume 1, Issue 8476, 1986.

Boers, R., Mitchell, R. M., and Krummel, P. B.: Correction of aircraft pyranometer measurements for diffuse radiance and alignment errors, Journal of Geophysical Research: Atmospheres, 103, 16 753–16 758, https://doi.org/10.1029/98jd01431, 1998.

Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., Kerminen, V.-M., Kondo, Y., Liao, H., Lohmann, U., Rasch, P., Satheesh, S., Sherwood, S., Stevens, B., and Zhang, X.: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, in: Climate Change 2013, edited by Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., chap. 7, Clouds and Aerosols, Cambridge University Press, https://www.ipcc.ch/pdf/assessment-report/ar5/wg1/WG1AR5_Chapter07_FINAL.pdf, 2013.

Brando, V., Lovell, J., King, E., Boadle, D., Scott, R., and Schroeder, T.: The Potential of Autonomous Ship-Borne Hyperspectral Radiometers for the Validation of Ocean Color Radiometry Data, Remote Sensing, 8, 150, https://doi.org/10.3390/rs8020150, 2016.

Bréon, F.-M., Vermeulen, A., and Descloitres, J.: An evaluation of satellite aerosol products against sunphotometer measurements, Remote Sensing of Environment, 115, 3102–3111, https://doi.org/10.1016/j.rse.2011.06.017, 2011.

Chen, L., Shi, G., Qin, S., Yang, S., and Zhang, P.: Direct radiative forcing of anthropogenic aerosols over oceans from satellite observations, Advances in Atmospheric Sciences, 28, 973–984, https://doi.org/10.1007/s00376-010-9210-4, 2011.

Christopher, S. A. and Jones, T.: Satellite-based assessment of cloud-free net radiative effect of dust aerosols over the Atlantic Ocean, Geophysical Research Letters, 34, https://doi.org/10.1029/2006GL027783, l02810, 2007.

Chu, D. A.: Validation of MODIS aerosol optical depth retrieval over land, Geophysical Research Letters, 29, https://doi.org/10.1029/2001gl013205, 2002.

Denman, K., Brasseur, G., Chidthaisong, A., Ciais, P., Cox, P., Dickinson, R., Hauglustaine, D., Heinze, C., Holland, E., Jacob, D., Lohmann, U., Ramachandran, S., da Silva Dias, P., Wofsy, S., and Zhang, X.: Couplings Between Changes in the Climate System and Biogeochemistry., in: Climate Change 2007: The Physical Science Basis, edited by Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., M.Tignor, and Miller, H., Cambridge University Press, https://www.ipcc.ch/publications_and_data/ar4/wg1/en/ch7.html, 2007.

di Sarra, A., Sferlazzo, D., Meloni, D., Anello, F., Bommarito, C., Corradini, S., Silvestri, L. D., Iorio, T. D., Monteleone, F., Pace, G., Piacentino, S., and Pugnaghi, S.: Empirical correction of multifilter rotating shadowband radiometer (MFRSR) aerosol optical depths for the aerosol forward scattering and development of a long-term integrated MFRSR-Cimel dataset at Lampedusa, Applied Optics, 54, 2725, https://doi.org/10.1364/ao.54.002725, 2015.

Dubovik, O., Sinyuk, A., Lapyonok, T., Holben, B. N., Mishchenko, M., Yang, P., Eck, T. F., Volten, H., Muñoz, O., Veihelmann, B., van der Zande, W. J., Leon, J.-F., Sorokin, M., and Slutsker, I.: Application of spheroid models to account for aerosol particle nonsphericity in remote sensing of desert dust, Journal of Geophysical Research: Atmospheres, 111, https://doi.org/10.1029/2005JD006619, d11208, 2006.

Ge, J. M., Su, J., Fu, Q., Ackerman, T. P., and Huang, J. P.: Dust aerosol forward scattering effects on ground-based aerosol optical depth retrievals, Journal of Quantitative Spectroscopy and Radiative Transfer, 112, 310–319, https://doi.org/10.1016/j.jqsrt.2010.07.006, 2011.

Haywood, J. M., Ramaswamy, V., and Soden, B. J.: Tropospheric Aerosol Climate Forcing in Clear-Sky Satellite Observations over the Oceans, Science, 283, 1299–1303, https://doi.org/10.1126/science.283.5406.1299, 1999.

Herman, B. M., Browning, R. S., and Luisi, J. J. D.: Determination of the Effective Imaginary Term of the Complex Refractive Index of Atmospheric Dust by Remote Sensing: The Diffuse–Direct Radiation Method, Journal of the Atmospheric Sciences, 32, 918–925, https://doi.org/10.1175/1520-0469(1975)032<0918:DOTEIT>2.0.CO;2, 1975.

Ichoku, C.: A spatio-temporal approach for global validation and analysis of MODIS aerosol products, Geophysical Research Letters, 29, https://doi.org/10.1029/2001gl013206, 2002.

Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., Dominguez, J. J., Engelen, R., Eskes, H., Flemming, J., Huijnen, V., Jones, L., Kipling, Z., Massart, S., Parrington, M., Peuch, V.-H., Razinger, M., Remy, S., Schulz, M., and Suttie, M.: The CAMS reanalysis of atmospheric composition, Atmospheric Chemistry and Physics, 19, 3515–3556, https://doi.org/10.5194/acp-19-3515-2019, 2019.

Kahn, R. A.: Reducing the Uncertainties in Direct Aerosol Radiative Forcing, Surveys in Geophysics, 33, 701–721, https://doi.org/10.1007/s10712-011-9153-z, 2012.

Kassianov, E. I., Flynn, C. J., Ackerman, T. P., and Barnard, J. C.: Aerosol single-scattering albedo and asymmetry parameter from MFRSR observations during the ARM Aerosol IOP 2003, Atmos. Chem. Phys., 7, 3341–3351, https://doi.org/10.5194/acp-7-3341-2007, 2007.

Kharol, S. K., Badarinath, K. V. S., Kaskaoutis, D. G., Sharma, A. R., and Gharai, B.: Influence of continental advection on aerosol characteristics over Bay of Bengal (BoB) in winter: results from W-ICARB cruise experiment, Annales Geophysicae, 29, 1423–1438, https://doi.org/10.5194/angeo-29-1423-2011, 2011.

Kinne, S.: Aerosol radiative effects with MACv2, Atmospheric Chemistry and Physics, 19, 10 919–10 959, https://doi.org/10.5194/acp-19-10919-2019, 2019.

Kleidman, R. G., O'Neill, N. T., Remer, L. A., Kaufman, Y. J., Eck, T. F., Tanré, D., Dubovik, O., and Holben, B. N.: Comparison of Moderate Resolution Imaging Spectroradiometer (MODIS) and Aerosol Robotic Network (AERONET) remote-sensing retrievals of aerosol fine mode fraction over ocean, Journal of Geophysical Research, 110, https://doi.org/10.1029/2005jd005760, 2005.

Knobelspiesse, K., Tan, Q., Bruegge, C., Cairns, B., Chowdhary, J., van Diedenhoven, B., Diner, D., Ferrare, R., van Harten, G., Jovanovic, V., Ottaviani, M., Redemann, J., Seidel, F., and Sinclair, K.: Intercomparison of airborne multi-angle polarimeter observations from the Polarimeter Definition Experiment, Applied Optics, 58, 650, https://doi.org/10.1364/ao.58.000650, 2019.

Levy, R., Hsu, C., et al.: MODIS Atmosphere L2 Aerosol Product. NASA MODIS Adaptive Processing System, Goddard Space Flight Center, USA, https://doi.org/10.5067/MODIS/MYD04_L2.006 and 10.5067/MODIS/MOD04_L2.006, 2015.

Levy, R. C., Remer, L. A., Kleidman, R. G., Mattoo, S., Ichoku, C., Kahn, R., and Eck, T. F.: Global evaluation of the Collection 5 MODIS dark-target aerosol products over land, Atmospheric Chemistry and Physics, 10, 10 399–10 420, https://doi.org/10.5194/acp-10-10399-2010, 2010.

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

Livingston, J. M., Redemann, J., Shinozuka, Y., Johnson, R., Russell, P. B., Zhang, Q., Mattoo, S., Remer, L., Levy, R., Munchak, L., and Ramachandran, S.: Comparison of MODIS 3 km and 10 km resolution aerosol optical depth retrievals over land with airborne sunphotometer measurements during ARCTAS summer 2008, Atmospheric Chemistry and Physics, 14, 2015–2038, https://doi.org/10.5194/acp-14-2015-2014, 2014.

Loeb, N. G. and Manalo-Smith, N.: Top-of-Atmosphere Direct Radiative Effect of Aerosols over Global Oceans from Merged CERES and MODIS Observations, Journal of Climate, 18, 3506–3526, https://doi.org/10.1175/JCLI3504.1, 2005.

Macke, A., ed.: The expedition of the research vessel "Polarstern" to the Antarctic in 2008 (ANT-XXIV/4), Berichte zur Polar- und Meeresforschung (Reports on Polar and Marine Research), vol. 591, p. 64, Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, hdl:10013/epic.32648.d001, 2009.

Mishchenko, M. I., Hovenier, J. W., and Travis, L. D.: Light Scattering by Nonspherical Particles - Theory, Measurements, and Applications, Academic Press, Amsterdam, Boston, https://doi.org/10.1088/0957-0233/11/12/705, 1999.

Misra, A.: Validation of Version 5.1 MODIS Aerosol Optical Depth (Deep Blue Algorithm and Dark Target Approach) over a Semi-Arid Location in Western India, Aerosol and Air Quality Research, https://doi.org/10.4209/aaqr.2014.01.0004, 2015.

Munchak, L. A., Levy, R. C., Mattoo, S., Remer, L. A., Holben, B. N., Schafer, J. S., Hostetler, C. A., and Ferrare, R. A.: MODIS 3 km aerosol product: applications over land in an urban/suburban region, Atmospheric Measurement Techniques, 6, 1747–1759, https://doi.org/10.5194/amt-6-1747-2013, 2013.

Myhre, G., Bellouin, N., Berglen, T. F., Berentsen, T. K., Boucher, O., Grini, A., Isaken, I. S. A., Johnsrud, M., Mishchenko, M. I., Stordal, F., and Tanré, D.: Comparison of the radiative properties and direct radiative effect of aerosols from a global aerosol model and remote sensing data over ocean, Tellus B, 59, 115–129, https://doi.org/10.1111/j.1600-0889.2006.00226.x, 2007.

Nabat, P., Somot, S., Mallet, M., Michou, M., Sevault, F., Driouech, F., Meloni, D., di Sarra, A., Di Biagio, C., Formenti, P., Sicard, M., Léon, J.-F., and Bouin, M.-N.: Dust aerosol radiative effects during summer 2012 simulated with a coupled regional aerosol–atmosphere–ocean model over the Mediterranean, Atmospheric Chemistry and Physics, 15, 3303–3326, https://doi.org/10.5194/acp-15-3303-2015, 2015.

Petters, J. L., Saxena, V. K., Slusser, J. R., Wenny, B. N., and Madronich, S.: Aerosol single scattering albedo retrieved from measurements of surface UV irradiance and a radiative transfer model, J. Geophys. Res., 108, https://doi.org/10.1029/2002jd002360, 2003.

Popp, T., de Leeuw, G., Bingen, C., Brühl, C., Capelle, V., Chedin, A., Clarisse, L., Dubovik, O., Grainger, R., Griesfeller, J., Heckel, A., Kinne, S., Klüser, L., Kosmale, M., Kolmonen, P., Lelli, L., Litvinov, P., Mei, L., North, P., Pinnock, S., Povey, A., Robert, C., Schulz, M., Sogacheva, L., Stebel, K., Zweers, D. S., Thomas, G., Tilstra, L., Vandenbussche, S., Veefkind, P., Vountas, M., and Xue, Y.: Development, Production and Evaluation of Aerosol Climate Data Records from European Satellite Observations (Aerosol_cci), Remote Sensing, 8, 421, https://doi.org/10.3390/rs8050421, 2016.

Porter, J. N., Miller, M., Pietras, C., and Motell, C.: Ship-Based Sun Photometer Measurements Using Microtops Sun Photometers, Journal of Atmospheric and Oceanic Technology, 18, 765–774, https://doi.org/10.1175/1520-0426(2001)018<0765:SBSPMU>2.0.CO;2, 2001.

Remer, L. A., Kaufman, Y. J., Tanré, D., Mattoo, S., Chu, D. A., Martins, J. V., Li, R.-R., Ichoku, C., Levy, R. C., Kleidman, R. G., Eck, T. F., Vermote, E., and Holben, B. N.: The MODIS Aerosol Algorithm, Products, and Validation, Journal of the Atmospheric Sciences, 62, 947–973, https://doi.org/10.1175/JAS3385.1, 2005.

Remer, L. A., Kleidman, R. G., Levy, R. C., Kaufman, Y. J., Tanré, D., Mattoo, S., Martins, J. V., Ichoku, C., Koren, I., Yu, H., and Holben, B. N.: Global aerosol climatology from the MODIS satellite sensors, Journal of Geophysical Research: Atmospheres, 113, https://doi.org/10.1029/2007JD009661, d14S07, 2008.

Remer, L. A., Mattoo, S., Levy, R. C., and Munchak, L. A.: MODIS 3 km aerosol product: algorithm and global perspective, Atmospheric Measurement Techniques, 6, 1829–1844, https://doi.org/10.5194/amt-6-1829-2013, 2013.

Russell, P. B.: Sunlight transmission through desert dust and marine aerosols: Diffuse light corrections to Sun photometry and pyrheliometry, Journal of Geophysical Research, 109, https://doi.org/10.1029/2003jd004292, 2004.

Schutgens, N. A. J., Nakata, M., and Nakajima, T.: Validation and empirical correction of MODIS AOT and AE over ocean, Atmospheric Measurement Techniques, 6, 2455–2475, https://doi.org/10.5194/amt-6-2455-2013, 2013.

Shettle, E. P. and Fenn, R. W.: Models for the aerosols of the lower atmosphere and the effects of humidity variations on their optical properties, Tech. rep., Environmental Research Paper Air Force Geophysics Lab., Hanscom AFB, MA. Optical Physics Div., 1979.

Shi, Y., Zhang, J., Reid, J. S., Holben, B., Hyer, E. J., and Curtis, C.: An analysis of the collection 5 MODIS over-ocean aerosol optical depth product for its implication in aerosol assimilation, Atmospheric Chemistry and Physics, 11, 557–565, https://doi.org/10.5194/acp-11-557-2011, 2011.

Smirnov, A., Holben, B. N., Slutsker, I., Giles, D. M., McClain, C. R., Eck, T. F., Sakerin, S. M., Macke, A., Croot, P., Zibordi, G., Quinn, P. K., Sciare, J., Kinne, S., Harvey, M., Smyth, T. J., Piketh, S., Zielinski, T., Proshutinsky, A., Goes, J. I., Nelson, N. B., Larouche, P., Radionov, V. F., Goloub, P., Krishna Moorthy, K., Matarrese, R., Robertson, E. J., and Jourdin, F.: Maritime Aerosol Network as a component of Aerosol Robotic Network, Journal of Geophysical Research: Atmospheres, 114, https://doi.org/10.1029/2008JD011257, 2009.

Smirnov, A., Holben, B. N., Giles, D. M., Slutsker, I., O'Neill, N. T., Eck, T. F., Macke, A., Croot, P., Courcoux, Y., Sakerin, S. M., Smyth, T. J., Zielinski, T., Zibordi, G., Goes, J. I., Harvey, M. J., Quinn, P. K., Nelson, N. B., Radionov, V. F., Duarte, C. M., Losno, R., Sciare, J., Voss, K. J., Kinne, S., Nalli, N. R., Joseph, E., Moorthy, K. K., Covert, D. S., Gulev, S. K., Milinevsky, G., Larouche, P., Belanger, S., Horne, E., Chin, M., Remer, L. A., Kahn, R. A., Reid, J. S., Schulz, M., Heald, C. L., Zhang, J., Lapina, K., Kleidman, R. G., Griesfeller, J., Gaitley, B. J., Tan, Q., and Diehl, T. L.: Maritime aerosol network as a component of AERONET – first results and comparison with global aerosol models and satellite retrievals, Atmospheric Measurement Techniques, 4, 583–597, https://doi.org/10.5194/amt-4-583-2011, 2011.

Tegen, I.: Modeling the mineral dust aerosol cycle in the climate system, Quaternary Science Reviews, 22, 1821–1834, https://doi.org/10.1016/S0277-3791(03)00163-X, Loess and the Dust Indicators and Records of Terrestrial and Mari ne Palaeoenvironments (DIRTMAP) database, 2003.

Thieuleux, F., Moulin, C., Bréon, F. M., Maignan, F., Poitou, J., and Tanré, D.: Remote sensing of aerosols over the oceans using MSG/SEVIRI imagery, Annales Geophysicae, 23, 3561–3568, https://doi.org/10.5194/angeo-23-3561-2005, 2005.

Toledano, C., Cachorro, V. E., Berjon, A., de Frutos, A. M., Sorribas, M., de la Morena, B. A., and Goloub, P.: Aerosol optical depth and Ångström exponent climatology at El Arenosillo AERONET site (Huelva, Spain), Quarterly Journal of the Royal Meteorological Society, 133, 795–807, https://doi.org/10.1002/qj.54, 2007.

Wei, J., Li, Z., Peng, Y., and Sun, L.: 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, 2019.

Witthuhn, J., Deneke, H., Macke, A., and Bernhard, G.: Algorithms and uncertainties for the determination of multispectral irradiance components and aerosol optical depth from a shipborne rotating shadowband radiometer, Atmospheric Measurement Techniques, 10, 709–730, https://doi.org/10.5194/amt-10-709-2017, 2017.

Wood, J., Smyth, T. J., and Estellés, V.: Autonomous marine hyperspectral radiometers for determining solar irradiances and aerosol optical properties, Atmospheric Measurement Techniques, 10, 1723–1737, https://doi.org/10.5194/amt-10-1723-2017, 2017.

Yin, Z., Ansmann, A., Baars, H., Radenz, M., Jimenez, C., Engelmann, R., Seifert, P., Herzog, A., Ohneiser, K., Hanbuch, K., Blarel, L., Goloub, P., Dubois, G., Victori, S., and Maupin, F.: Aerosol measurements with shipborne sun-sky-lunar photometer and collocated multiwavelength Raman polarization lidar over the Atlantic Ocean, Atmos. Meas. Tech. Discuss, https://doi.org/10.5194/amt-2019-132, in review, 2019.

Yu, H., Kaufman, Y. J., Chin, M., Feingold, G., Remer, L. A., Anderson, T. L., Balkanski, Y., Bellouin, N., Boucher, O., Christopher, S., DeCola, P., Kahn, R., Koch, D., Loeb, N., Reddy, M. S., Schulz, M., Takemura, T., and Zhou, M.: A review of measurement-based assessments of the aerosol direct radiative effect and forcing, Atmospheric Chemistry and Physics, 6, 613–666, https://doi.org/10.5194/acp-6-613-2006, 2006.

Table 1. Technical specifications of the GUVis and Microtops instruments.

Characteristics	GUVis-3511	Microtops II
spectral channels	18x [310-1640 nm]	5x [380-870 nm]
	1x unfiltered	1x 940 nm (water vapor)
FWHM	10 nm	10 nm (4 nm at 380 nm)
Measurement frequency	15 Hz	-
Sweep period	40 s	-
Time resolution	1 min	variable (>10 min)
viewing/shading angle	13-15°	2.5°
weight	5 kg	0.6 kg
dimensions	24x24x36 cm	10x20x4 cm

Table 2. Number of available data points in the MIC, GUV, and COMB datasets, given as total number, and separated by aerosol type. The type classification follows the scheme of Toledano et al. (2007). Collocated data points for the comparison with satellite and model datasets (see text for collocation criteria) are also given.

Class	MIC	CAMS	*MxD04_L2*	*MxD04_3K*	SEVIRI
maritime background	~~12749~~ 10635	~~1478~~ 1216	~~1090~~ 936	~~1064~~ 912	~~7218~~ 6134
~~desert dust~~ mineral dust transport	~~5085~~ 4492	~~730~~ 579	~~296~~ 257	~~261~~ 243	~~2023~~ 1719
~~biomass burning~~ continental transport	~~225~~ 2141	~~44~~ 388	~~14~~ 194	~~6~~ 181	~~116~~
~~continental~~	~~931~~	~~174~~	~~92~~	~~93~~	~~599~~ 1288
mixed	~~252~~ 1982	~~48~~ 291	~~23~~ 130	~~22~~ 112	~~100~~ 920
~~no class~~	~~8~~	~~0~~	~~2~~	~~2~~	~~5~~
total	19250	2474	1517	1448	10061
Class	~~GUVis~~ GUV	CAMS	*MxD04_L2*	*MxD04_3K*	SEVIRI
maritime background	~~6179~~ 5210	~~93~~ 79	~~92~~ 80	~~131~~ 118	~~793~~ 704
~~desert dust~~ mineral dust transport	~~2552~~ 3120	~~31~~ 34	~~27~~ 49	~~46~~ 68	~~240~~ 277
~~biomass burning~~ continental transport	~~0~~ 282	~~0~~ 4	~~0~~ 2	~~0~~ 3	~~0~~
~~continental~~	~~152~~	~~1~~	~~0~~	~~0~~	~~24~~ 35
mixed	~~32~~ 1800	~~2~~ 24	~~0~~ 16	~~1~~ 21	~~2~~ 110
~~no class~~	~~1497~~	~~14~~	~~28~~	~~32~~	~~67~~
total	10412	141	147	210	1126
Class	COMB	CAMS	*MxD04_L2*	*MxD04_3K*	SEVIRI
maritime background	~~698~~ 607	~~84~~ 77	~~73~~ 64	~~86~~ 76	~~355~~ 319
~~desert dust~~ mineral dust transport	~~278~~ 310	~~23~~ 26	~~16~~ 28	~~13~~ 27	~~88~~ 122
~~biomass burning~~ continental transport	~~0~~ 10	~~0~~	~~0~~	~~0~~	~~0~~
~~continental~~	~~5~~	1	0	0	~~3~~ 4
mixed	~~0~~ 106	~~0~~ 8	~~0~~ 1	0	~~0~~ 24
~~no class~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~
total	~~981~~ 1033	~~108~~ 112	~~89~~ 93	~~99~~ 103	~~446~~ 469

Table 3. Per–channel coefficients obtained for the cross-calibration of the GUVis to the MIC instrumental channels (determined per cruise), expressed as relative correction C to the most recent laboratory calibration, ~~and as absolute calibration coefficient k,~~ together with the empirical scaling coefficient S to correct the AOD for the forward scattering contribution in the GUVis observations (determined for all cruises). See text for details on their estimation. For comparison the absolute calibration coefficient k for each cruise is shown. The channel of 630 and 810 nm are non native spectral channels and therefore marked with (i) as they are interpolated using the AE with Eq. (1).

Channel	$C_{\mathbf{PS83}}$	$k_{\mathbf{PS83}}$	$C_{\mathbf{PS95}}$	$k_{\mathbf{PS95}}$	$C_{\mathbf{PS98}}$	$k_{\mathbf{PS98}}$	$C_{\mathbf{PS102}}$	$k_{\mathbf{PS102}}$	$C_{\mathbf{PS113}}$	$k_{\mathbf{PS113}}$	S
nm	-	$\frac{\mathrm{V\,m^2\,nm}}{\mathrm{W}}$	-	$\frac{\mathrm{V\,m^2\,nm}}{\mathrm{W}}$	-	$\frac{\mathrm{V\,m^2\,nm}}{\mathrm{W}}$	-	$\frac{\mathrm{V\,m^2\,nm}}{\mathrm{W}}$	-	$\frac{\mathrm{V\,m^2\,nm}}{\mathrm{W}}$	-
380	1.02	1.42	0.97	1.44	1.03	1.32	0.96	1.31	0.99	1.29	1.12
440	0.98	7.52	0.93	7.48	0.96	6.90	0.92	6.87	0.95	6.79	1.13
500	1.02	20.91	0.99	21.33	1.03	19.55	0.97	19.63	1.00	19.83	1.14
630	0.99	~~43.57~~ (i)	0.98	~~43.83~~ (i)	1.01	~~39.90~~ (i)	0.96	~~39.95~~ (i)	0.98	~~40.09~~ (i)	1.13
675	1.02	51.64	1.01	51.96	1.05	46.74	1.01	46.90	1.01	47.32	1.13
810	0.97	~~48.86~~ (i)	0.98	~~48.43~~ (i)	0.99	~~43.47~~ (i)	0.95	~~43.60~~ (i)	0.94	~~43.92~~ (i)	1.11
870	0.96	48.86	0.98	48.43	1.00	43.47	0.96	43.60	0.94	43.92	1.10

Table 4. Statistics comparing the old and new processing algorithms of the GUVis observations, as well as the calibration– and forward-scattering–corrected GUVis dataset versus the MIC dataset. Availability of Microtops and GUVis channels are indicated with (x: available; i: internally interpolated; - : interpolated using AE). Statistics include the number of datapoints N, Pearson correlation coefficient (R), the Pearson correlation coefficient of difference (R(D)), and bias plus fraction of outliers based on the limit of agreement (LOA) method for 95% confidence interval (see Sect. 3.3 for explanation of R(D), bias and LOA).

Channel	availability		N	R	R(D)	bias ± LOA	bias ± LOA	Outlier
[nm]	MIC	GUVis	-	-	-	-	weighted	%
			GUVis data (old processing)					
380	x	x	993	0.954	-0.43	-0.03±0.10	-1.03±3.49	19.94
440	x	x	993	0.965	-0.58	-0.03±0.09	-1.12±3.23	17.42
500	i	x	993	0.965	-0.66	-0.03±0.09	-1.07±3.32	13.39
630	-	-	993	0.969	-0.67	-0.03±0.09	-0.97±3.13	12.19
675	x	x	993	0.968	-0.67	-0.02±0.09	-0.86±3.15	10.98
810	-	-	993	0.970	-0.62	-0.03±0.08	-0.92±2.89	10.78
870	x	x	993	0.970	-0.62	-0.03±0.08	-0.91±2.86	10.47
			GUVis data (new processing)					
380	x	x	1061	0.988	-0.66	-0.02±0.06	-0.78±2.23	11.12
440	x	x	1061	0.989	-0.69	-0.02±0.06	-0.87±2.16	10.74
500	i	x	1061	0.990	-0.73	-0.02±0.06	-0.69±2.07	9.52
630	-	-	1061	0.991	-0.74	-0.02±0.06	-0.67±1.96	9.43
675	x	x	1061	0.992	-0.75	-0.02±0.05	-0.56±1.86	8.11
810	-	-	1061	0.992	-0.68	-0.02±0.05	-0.59±1.71	7.26
870	x	x	1061	0.992	-0.67	-0.02±0.05	-0.55±1.66	6.69
			Enhanced GUVis data (GUVisE)					
380	x	x	~~1006~~ 1061	~~0.992~~ 0.988	~~-0.13~~ -0.12	0.00±~~0.04~~ 0.05	~~0.04~~0.03±~~1.42~~ 1.78	~~0.99~~ 3.86
440	x	x	~~1006~~ 1061	~~0.993~~ 0.989	-0.12	0.00±~~0.04~~ 0.05	0.03±~~1.32~~ 1.65	~~0.70~~ 2.73
500	i	x	~~1006~~ 1061	~~0.994~~ 0.990	~~-0.14~~ -0.13	0.00±~~0.03~~ 0.04	0.04±~~1.20~~ 1.53	~~0.10~~ 1.60
630	-	-	~~1006~~ 1061	~~0.995~~ 0.991	~~-0.17~~ -0.15	0.00±~~0.03~~ 0.04	0.06±~~1.07~~ 1.42	~~0.00~~ 1.32
675	x	x	~~1006~~ 1061	~~0.996~~ 0.992	~~-0.13~~ -0.12	0.00±~~0.03~~ 0.04	0.03±~~0.96~~ 1.33	~~0.00~~ 1.32
810	-	-	~~1006~~ 1061	~~0.995~~ 0.992	-0.11	0.00±~~0.03~~ 0.04	0.03±~~0.95~~ 1.32	~~0.10~~ 1.32
870	x	x	~~1006~~ 1061	~~0.996~~ 0.992	~~-0.11~~ -0.10	0.00±~~0.03~~ 0.04	0.03±~~0.93~~ 1.29	~~0.00~~ 1.04

Table 5. Statistics comparing ~~the GUVisE, COMB and~~ MIC reference AOD at ~~wavelengths of~~ 550~~and 630~~ nm versus ~~the~~ CAMS, SEVIRI and *MxD04* ~~AOD~~ aerosol products. The comparison is shown for the collocated datasets with different selections based on aerosol type (maritime, desert dust and continental) and for CAMS RA with and without available AATSR measurements. Further, the comparisons are separated regarding MIC AOD value to all, AOD higher and AOD lower 0.4. N denotes the number of collocated data ~~point~~points for all selections. Listed are also the correlation (R) ~~, coefficients of a linear regression, followed by the correlation of the difference (R(D))~~ and the bias/fraction of data based on the limit of agreement (LOA) method for 95% confidence interval. G1 and G2 indicate the percentage of data points laying within the expected error limits EE1 and EE2.

dataset	selection	N	R	bias ± LOA	G1	G2	N	bias ± LOA	N	bias ± LOA
-	-	-	-	-	%	%	AOD < 0.4		AOD ≥ 0.4	
CAMS RA	all	2472	0.92	0.00±0.13	59	66	2174	0.01±0.10	298	-0.01±0.27
	maritime	1214	0.66	0.01±0.07	72	75	1214	0.01±0.07	0	-
	desert dust	579	0.85	0.01±0.22	45	60	333	0.02±0.15	246	-0.01±0.28
	continental	388	0.81	-0.01±0.13	46	52	357	-0.01±0.12	31	-0.04±0.21
	with AATSR	941	0.90	-0.00±0.14	57	63	848	-0.00±0.11	93	-0.01±0.28
	no AATSR	190	0.87	0.00±0.18	49	61	157	0.01±0.14	33	-0.04±0.32
SEVIRI	all	10055	0.88	0.02±0.15	60	71	9392	0.03±0.13	663	-0.01±0.35
	maritime	6130	0.44	0.03±0.11	64	74	6130	0.03±0.11	0	-
	desert dust	1719	0.84	0.02±0.26	49	62	1173	0.04±0.18	546	-0.02±0.36
	continental	1287	0.85	0.01±0.11	59	69	1237	0.01±0.10	50	0.03±0.19
MxD04_3K	all	704	0.95	0.03±0.12	65	78	640	0.02±0.08	64	0.07±0.30
	maritime	447	0.64	0.02±0.07	68	81	447	0.02±0.07	0	-
	desert dust	163	0.93	0.05±0.21	53	63	103	0.03±0.13	60	0.08±0.31
	continental	35	0.89	0.02±0.10	60	80	34	0.01±0.08	1	-
MxD04_L2	all	924	0.93	0.03±0.13	65	76	841	0.02±0.09	83	0.07±0.34
	maritime	563	0.67	0.02±0.07	70	81	563	0.02±0.07	0	-
	desert dust	210	0.87	0.05±0.24	51	60	134	0.04±0.14	76	0.07±0.35
	continental	63	0.88	0.00±0.10	68	83	62	0.00±0.08	1	-

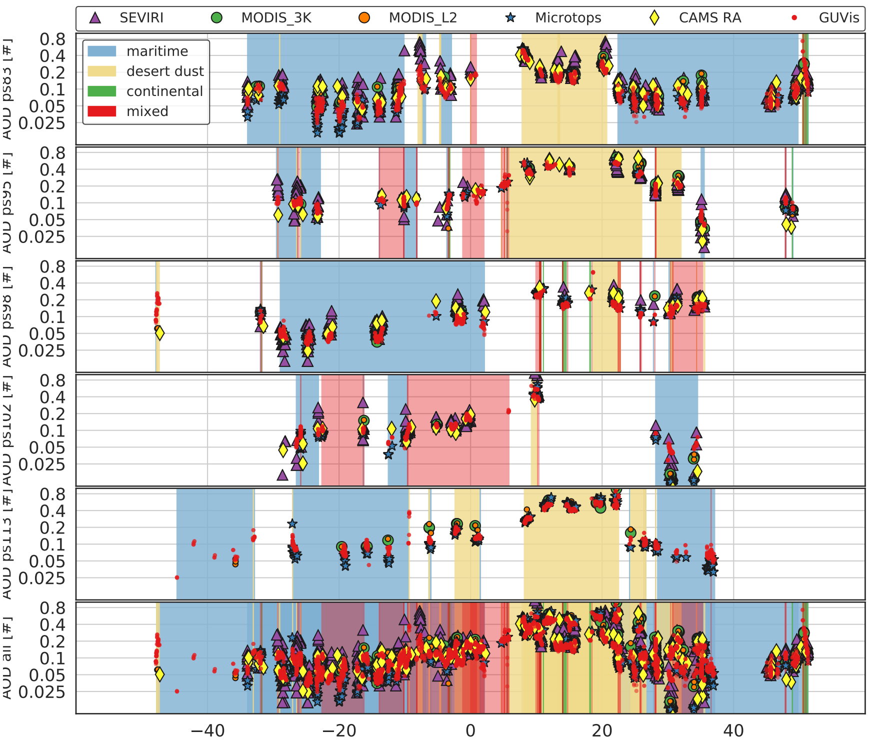

Figure 1. Zonal cross-section of AOD at 630 nm estimated from GUVis measurements during the *Polarstern* cruises PS83, PS95, PS98, PS102 and PS113, together with collocated AOD obtained from Microtops, satellite products and CAMS RA. Along this cross-section across the Atlantic ocean, the dominant aerosol type is either maritime (blue shaded region) or desert dust (yellow shaded region) while passing the Sahara desert. The ~~Aerosol~~ aerosol classification method is based on the method of Toledano et al. (2007) and GUVis products.

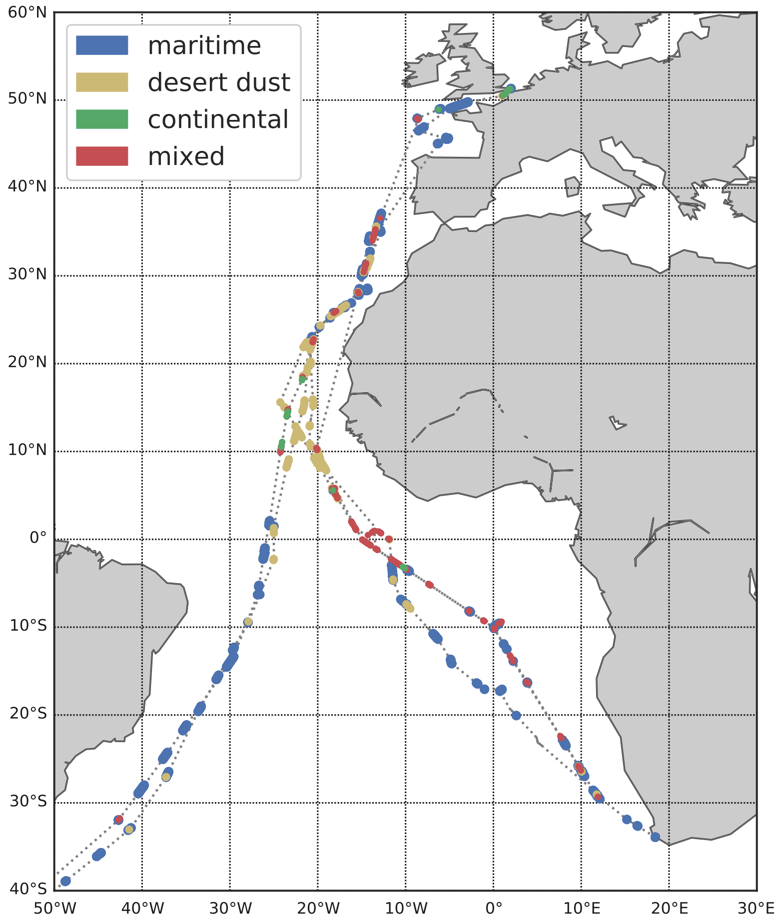

Figure 2. Cruise tracks of the Polarstern during PS83, PS95, PS98, PS102 and PS113. The resulting aerosol type classification obtained from the GUVis observation is shown by color coding.

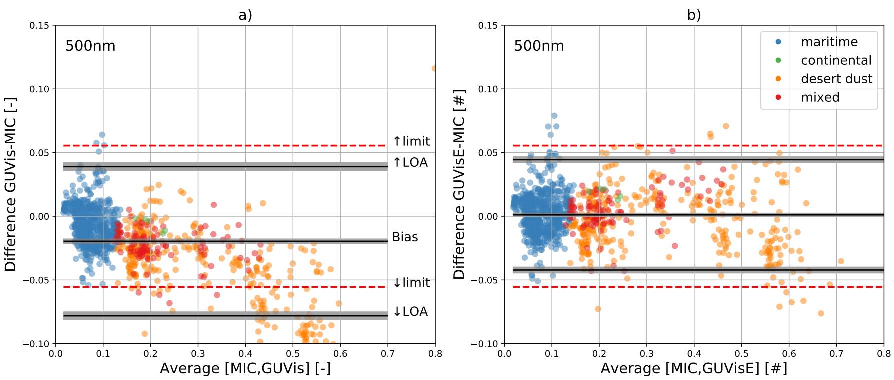

Figure 3. (a) Difference of the AOD from GUVis - MIC datasets, plotted versus their mean. (b) Diffence of AOD from GUVisE - MIC datasets, plotted versus their mean. Blue and orange dots indicate maritime and desert dust respectively. This data is considered as valid, while red dots are flagged as outliers (which exceed the uncertainty estimate in at least one of the considered spectral channels). The black lines indicate the bias and the upper and lower limit of agreement (LOA), which should contain 95% of data points (see Sect. 3.3). The gray–shaded areas indicate the uncertainty estimate (95% confidence limit) of bias and LOA.

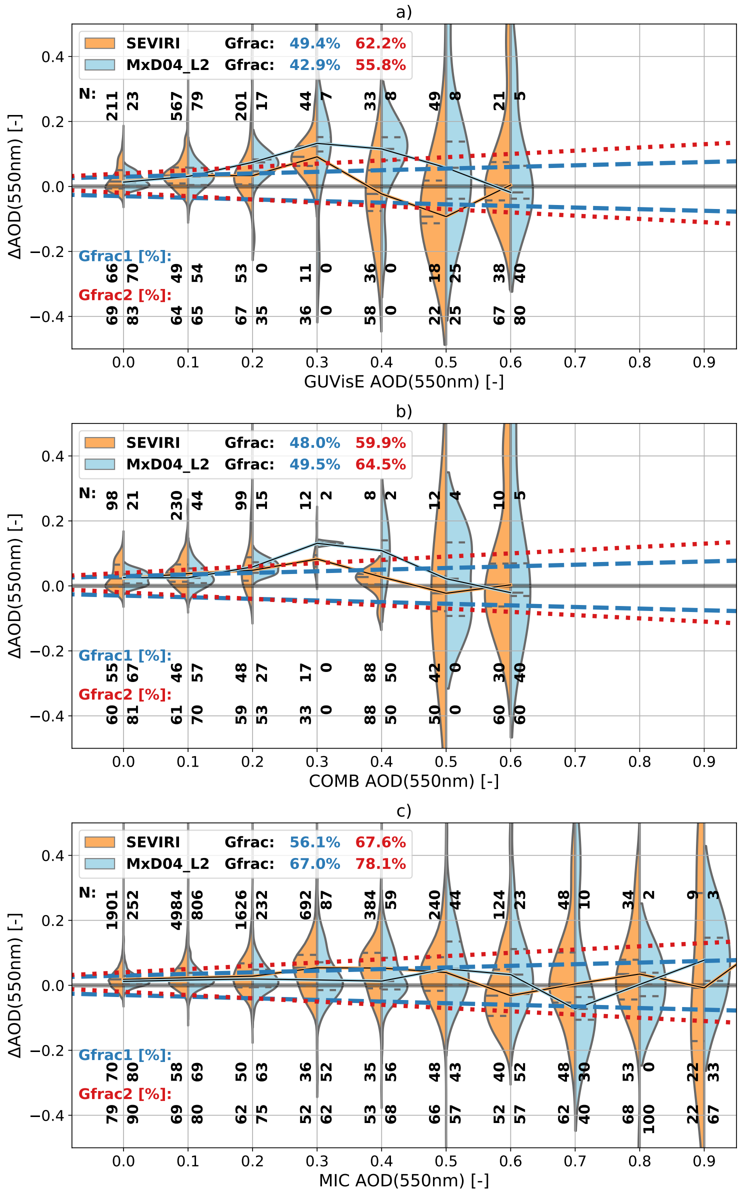

Figure 4. Comparison of AOD at 550 nm from the (a) GUVisE, (b) COMB, and (c) MIC reference datasets versus the SEVIRI and MODIS AOD products. The two sided violin-plots indicate the distribution of the difference for bins of 0.1 in AOD. The median of each bin is connected with a solid line to visualize the development of the bias. The blue dashed and red dotted lines indicate the expected error limits for the MODIS AOD products. It is expected that at least 67% of data points fall into the expected error limits. Gfrac1 and Gfrac2 are the actual percentage of data points lying within the error limits, calculated for each bin, and as total for both the SEVIRI and MODIS AOD products.

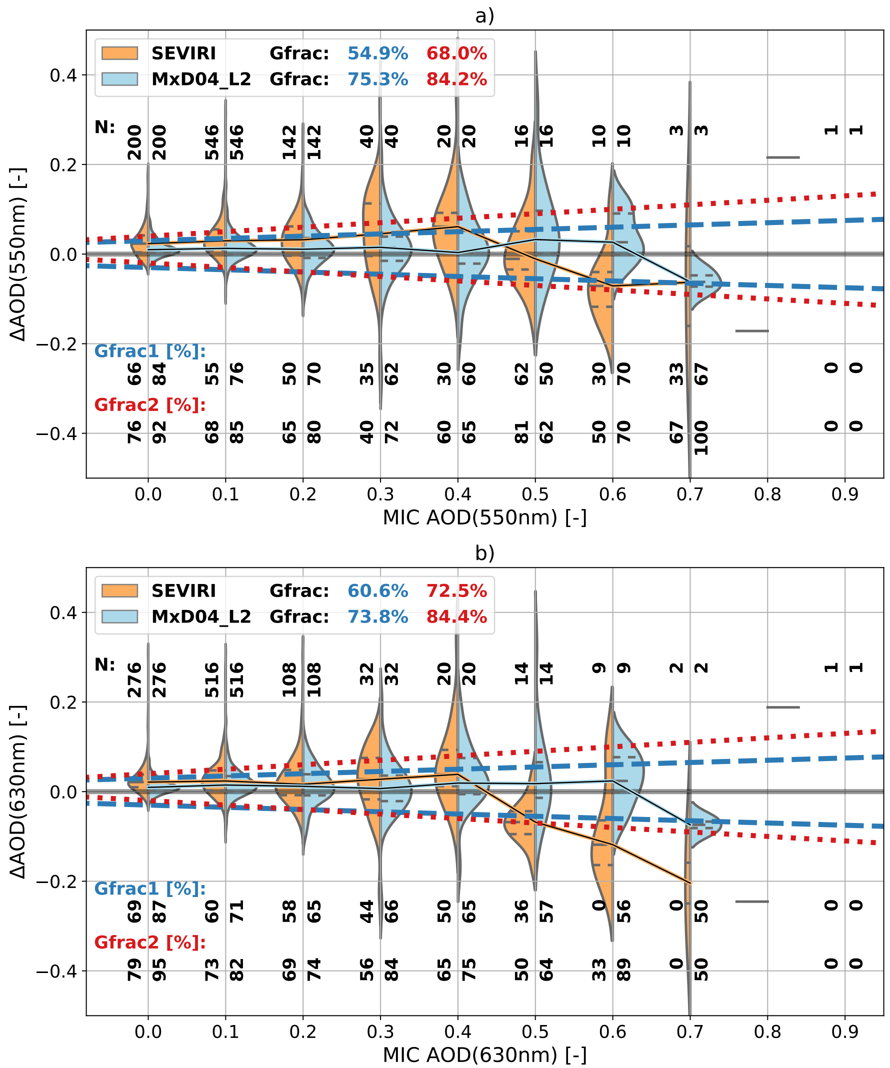

Figure 5. Same as Fig 4, but only using MIC as reference dataset. The comparison is presented for AOD at 550 nm (panel (a)) and 630 nm (panel (b)). As additional constraint, availability of data from both the SEVIRI and MODIS datasets is required.

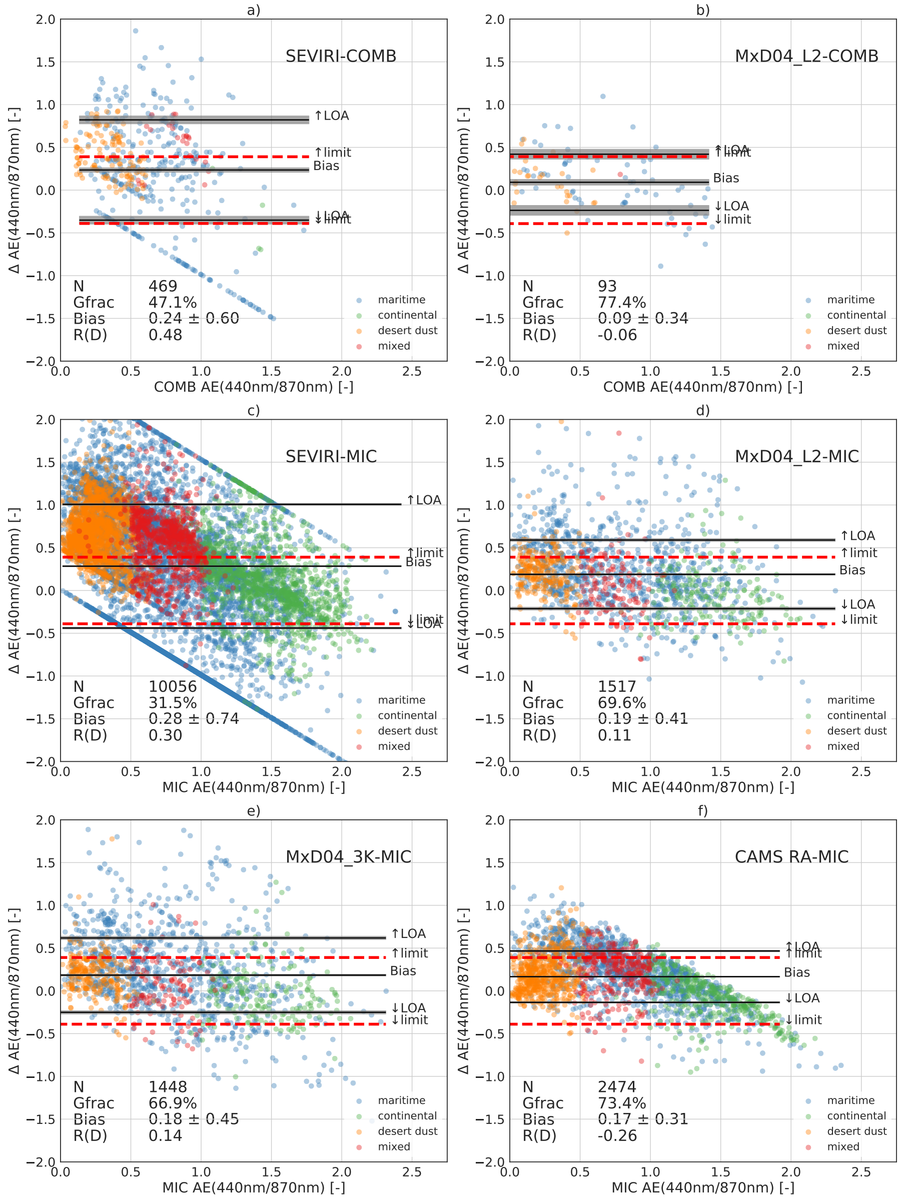

Figure 6. Comparison of AE from the COMB or MIC reference datasets versus the SEVIRI, MxD04 and CAMS RA datasets. The aerosol type classified with the reference dataset is indicated by the color of each point. Red dashed lines indicate the estimated error limits for AE (±0.4) of the MODIS products (Chu, 2002). 67% of AE data points are expected to fall into these limits. LOA (outer black lines) are based on 67% confidence intervals. The bias is given by the middle black line, and is calculated as the mean of the difference. The statistics state the number of measurements (N), percentage of data within expected error limits (Gfrac), bias ± LOA, and correlation of the difference (R(D)).

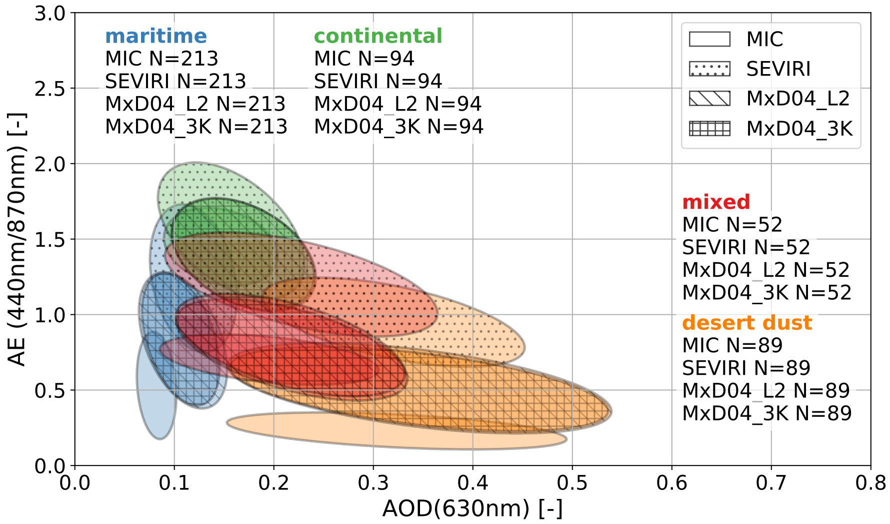

Figure 7. Comparison of AE calculated from AOD at the wavelengths of ~~630~~440 nm and 810 nm versus AOD at 630 nm, calculated from the Microtops, SEVIRI and MODIS products. Simultaneous data availability from satellites and MIC is required, so that each instrumental data points has a corresponding counterpart from the other instruments. The data points are grouped by aerosol type (classified with MIC), and visualized as covariance ellipsoids for a 67 % confidence interval.

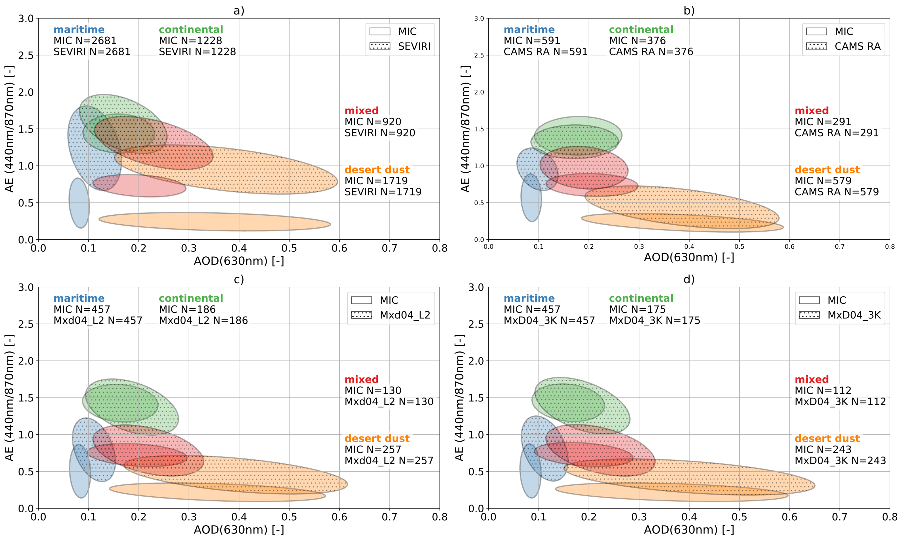

Figure 8. Same as Fig. 7, but the requirement for simultaneous data availability from all data points was dropped, and the figure also shows CAMS RA data points.

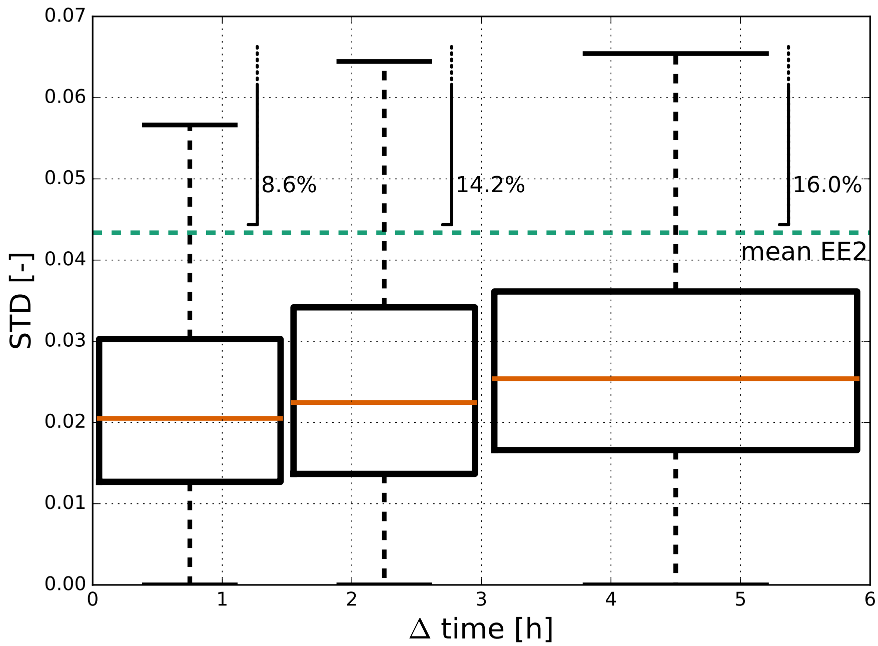

Figure 9. Temporal variation of AOD expressed as the standard deviation (STD) of SEVIRI AOD between MODIS overpasses, as a function of time lag between retrievals (Δ time). The mean of the expected error limits (EE2) of MODIS $[+(0.04+0.1\,\mathrm{AOD}), -(0.02+0.1\,\mathrm{AOD})]$ are calculated from the mean SEVIRI AOD of all datapoints and shown as the green dashed line. A variation in AOD can be considered significant, if the magnitude of STD exceeds the error limits. The precentage of significant situations are denoted for each time interval.

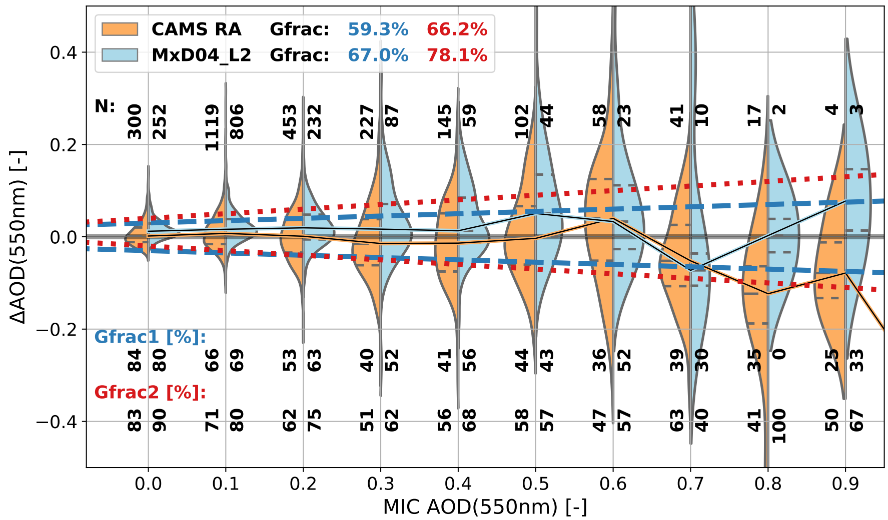

Figure 10. Same as Fig 4 but comparing CAMS RA and MODIS *MxD04_L2* AOD to MIC AOD as reference.

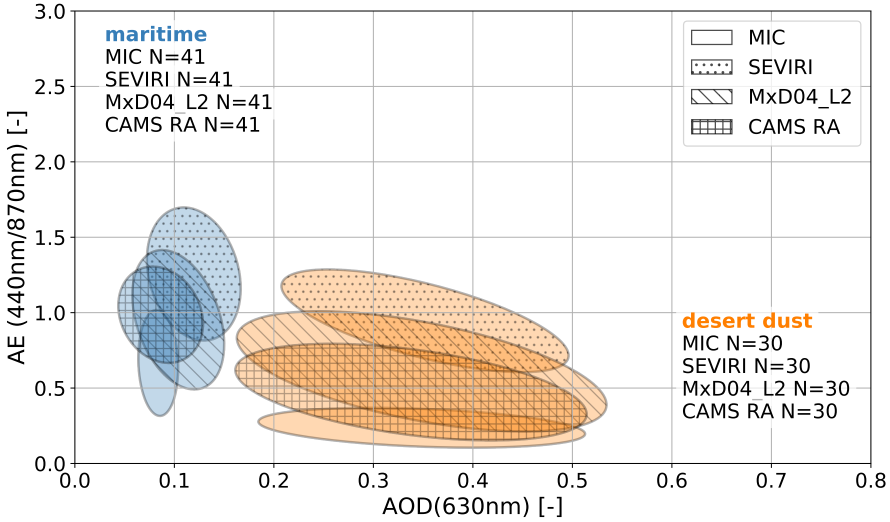

Figure 11. Same as Fig. 7, but comparing CAMS RA AOD instead of *MxD04_3K* AOD.

Appendix A: GUVis processing update

In this appendix, the improvements of the GUVis processing algorithm of the shadowband sweep irradiance data is described.

The processing of the sweep time-series of the GUVis instrument is required to extract the global, direct and diffuse components from the measured spectral irradiance components. As described in detail in Witthuhn et al. (2017), the accurate estimation of the blocked diffuse irradiance while the direct sun is also blocked by the shadowband (i.e. the shadow of the band falls onto the detector) during the sweep is a fundamental challenge which has to be solved by the processing algorithm. Figure A1 and A2 illustrate an idealized and a measured shadowband sweep together with the processing algorithm. Between shadowband sweeps, the band is stowed in a parking position out of sight of the hemispheric field of view of the sensor, so the global irradiance (F_{glo}) is observed by the radiometer (periods (a) and (g) in the figures). The amount of blocked diffuse irradiance can be directly inferred from the sweep data while the sensor is not shaded from direct sunlight by the band, by considering the reduction in measured irradiance (periods (b) and (f) in the figures). While the sun is partially or completely blocked by the band due to the shadow falling onto the sensor (periods (c) to (e) in the figures), the reduction of the irradiance recorded by the shadowband compared to the global irradiance consists both of the blocked direct irradiance component (F_{dir}), plus a blocked fraction of the diffuse irradiance ($F_{dif,b}$). It is necessary to separate both parts in order to be able to calculate F_{dir}. The relation of the irradiance components is as follows:

$$F_{glo} = F_{dir} + F_{dif}, \tag{A1}$$

with F_{dif} being the diffuse irradiance component which is partially blocked by the shadowband during the sweep, and can be separated into a blocked ($F_{dif,d}$) and a non-blocked part contributing to the observations ($F_{dif,o}$):

$$F_{dif} = F_{dif,b} + F_{dif,o}. \tag{A2}$$

As $F_{dif,b}$ cannot be inferred directly from the measurement, it is estimated by linear extrapolation of the measured irradiance data during the sweep while the sun is not blocked by the shadowband (periods (b) and (f)). As the 30 samples before and after the shadow of the band transitions across the sensor are used, accurate knowledge of the time when the shadow starts to shade the sensor is required (termed point of contact from here on). The identification of the point of contact is accomplished in our processing algorithm by considering the slope of the measured irradiance data using empirical thresholds. It has to be realized that the change of slope before and after reaching the points of contact (thus at the transition from (b) to (c)) depends strongly on the present atmospheric situation, shape of the circum-solar radiation and shadowband geometry. In particular, this change is not as sharp as indicated by Fig. A1, but shows a smooth transition as visible in Fig. A2.

The GUVis processing algorithm has received a substantial update compared to the version introduced in Witthuhn et al. (2017) to address several shortcomings. The following improvements were made:

(i) The identification of the point of contact is done by considering the slope of the measured irradiance during the complete shadowband sweep. Since the measured irradiance drops sharply once the sensor is partially shaded by the shadowband, a threshold can be used for the slope to identify the point of contact. This threshold was choosen by using a constant absolute value in the old processing, which sometimes resulted in an inconsistent identification of the point of contact, in particular during low sun or high AOD situations. In the revised processing, a relative threshold is used, which is calculated relative to the difference of measured global irradiance and the minimum measured irradiance of the sweep. This leads to a more reliable identification of the point of contact, as well as less scatter of the irradiance components of successive sweeps and during the daily cycle.

(ii) The measured irradiance during one sweep sometimes contains high-frequency variations of the irradiance, e.g. caused by small clouds or the smoke plume of the ship. Affected sweeps are identified by a pre-processing filter and excluded from further processing. The pre-processing filter is applied by calculating the variance for the selected interpolation data (30 data points before and after the points of contact). The variance is compared to an fixed threshold value of 0.002^2. In the old processing, if the threshold was exceeded either using the data before or after the points of contact, the sweep is dropped completely. For the updated processing algorithm, if the threshold is exceeded, the data point with the largest deviation are removed from the interpolation. Data points are removed either until the variance criterium is met and the processing continues, or the number of data points used for interpolation is less than 21, in which case the sweep is excluded.

(iii) As mentioned before, the BioSHADE accessory for the GUVis utilizes a broad shadowband with a shading angle of about 15°. The shading angle is comparable to the field of view of a sunphotometer, which has a field of view of about 2.5° for the Microtops instrument (Porter et al., 2001). Comparing AOD based on the GUVis and Microtops, the difference in the field of view will lead to an underestimation of AOD retrieved with the GUVis, as has been reported also for the multi-filter rotating shadowband radiomeer (MFRSR) in a comparison to AERONET sunphotometer in the study of di Sarra et al. (2015). The underestimation is attributable to the forward scattering contribution of aerosol scattering as investigated by Russell (2004). The underestimation of AOD is substantial for large shadowband shading angles, and especially large for aerosol particles with strong forward scattering (e.g. desert dust) (Ge et al., 2011). To at least partly compensate for this effect, an offset has been introduced in the linear extrapolation of the blocked diffuse irradiance. The offset depends on the slope of the interpolation data before and after the points of contact, which is steeper during stronger forward scattering aerosol situations. The offset is calculated from the difference of the irradiance at the point of contact, and the extrapolated irradiance using the interpolation data at the time of the point of contact. Therefore, the offset is larger during strong forward scattering aerosol situations, since the increase in forward scattering leads to a steeper drop before the point of contact. Thus, the offset compensates for the underestimation of AOD due to aerosol forward scattering.

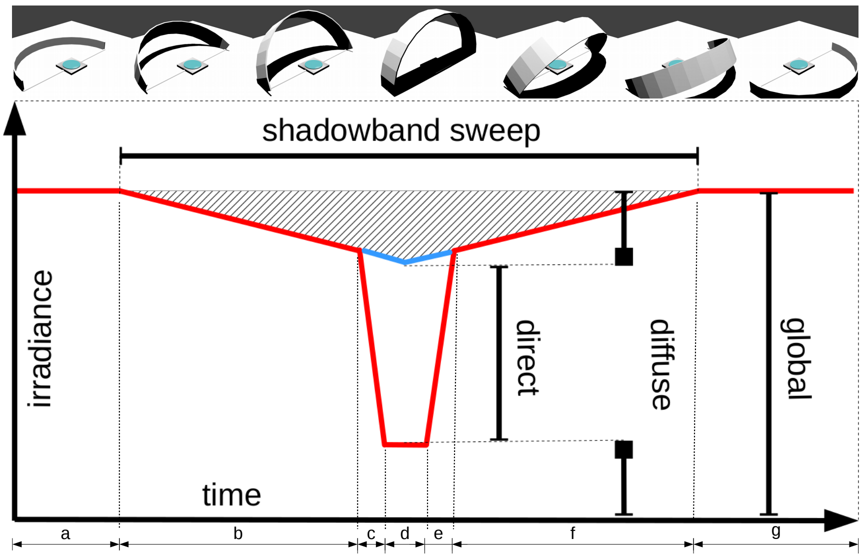

Figure A1. Schematic illustration of a shadowband sweep measured by the GUVis shadowband radiometer. The red line indicates the mearued irradiance. The figures on top illustrate the shadowband position relative to the sensor during the sweep. The hatched area indicates the diffuse irradiance blocked from the sensor during the sweep. The blue line indicates the unknown blocked diffuse irradiance when the direct irradiance is at least partially blocked by the shadowband. It has to be estimated by the processing algorithm in order to accurately estimate the direct irradiance. The letters a to g indicate different periods during the sweep as follows: (a, g) shadowband in parking position, out of sight of the hemispheric field of view of the sensor; the measured irradiance corresponds to the global irradiance. (b, f) the shadowband is moving, but the direct irradiance of the sun is not blocked from the sensor. (c, e) the direct irradiance is partially blocked by the shadowband, as the band shades the sensor. (d) the direct irradiance is completely blocked by the shadowband.

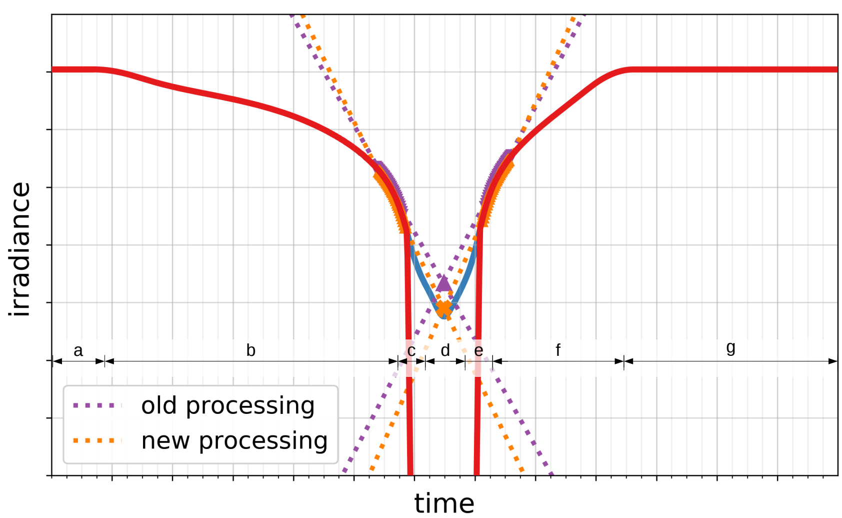

Figure A2. Like Fig. A1, but with a measured irradiance time series (red line) of a shadowband sweep. The blue line indicates the unknown amount of blocked diffuse irradiance, which has to be estimated by the processing algorithm. The dotted lines indicate the extrapolation lines of the old (purple) and new (orange) processing algorithm to estimate the blocked diffuse irradiance.