

***Interactive comment on “Evaluation of
satellite-based aerosol datasets and the CAMS
reanalysis over ocean utilizing shipborne
reference observations” by Jonas Witthuhn et al.***

Jonas Witthuhn et al.

jonas.witthuhn@tropos.de

Received and published: 7 February 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.

C1

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

C2

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.

C3

- 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 depend 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 GUViE.
- Table 5 receives a major update as suggested by Stefan Kinne.

C4

- 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.

C5

Response to RC3 from Stefan Kinne:

Evaluation of satellite-based aerosol datasets and the CAMS reanalysis over ocean utilizing shipborne 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 MICROTOS) direct solar attenuation data is out of question and serves as the main reference in the evaluation. In contrast, the usefulness of the complementary shadowband radiometer (GUV) appears more limited (completely cloud-free situations are

C6

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 as 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.

C7

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 RF_{aci} overall has larger contributions than the RF_{ari}. 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 RF_{ari}. The impact of these two components on R_{aci} (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 RF_{ari} and RF_{aci} are not further considered in the manuscript, we have decided to shorten and merge the three introduction paragraphs into one, also based on the comments 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 climate

C8

tology 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.

C9

- 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 -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).

C10

- 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 (Witthuhn2017)."
 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 than 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

C11

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 (Levy2015) and MxD04_3K (Re-

C12

mer2013) 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."

C13

- 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 on components, I would start with an AOD separation into fine-mode ($r < 0.5 \mu\text{m}$) and coarse-mode ($r > 0.5 \mu\text{m}$) 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!

C14

- 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

C15

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 (Inness2019) 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 (Inness2019)."
- 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 data-points 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 ± 30 min 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

C16

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) MICROTOS 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 MICROTOS consistency a rather complex procedure is applied to GNU data based on comparisons to almost simultaneous MICROTOS 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 C_i is 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 C_i is applied for a calibration correction to compensate for differences in Microtops and GUVis calibration for the sake of comparability. This

C17

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 (Alexandrov2002). 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. (Sarraf2015) 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

C18

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 MICROTOSPS 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

C19

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 ship-borne 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 total	GUVis 10412	MIC 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 MICROTOSPS 440/870 Angstrom had been used instead (also the SEVIRI band are spectrally much broader with some water vapor contamination especially in 810nm band).

C20

- 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 \pm LOA	G1	G2
-	-	nm	-	-	-	-	%	%
SEVIRI original	2*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	2*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

C21

son 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 MICROTUPS 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

C22

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 significance 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.

C23

- 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.

C24

- 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 measurements. 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 ≥ 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.

C25

- 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

C26

- 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.*

C27

- 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

C28

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*

C29

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 composition restrictions as you have analysed. Nevertheless, Figure 1 (see last page) 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 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.

C30

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. Kinne2019)"*

Please also note the supplement to this comment:

<https://www.atmos-meas-tech-discuss.net/amt-2019-321/amt-2019-321-AC3-supplement.pdf>

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2019-321, 2019.

C31

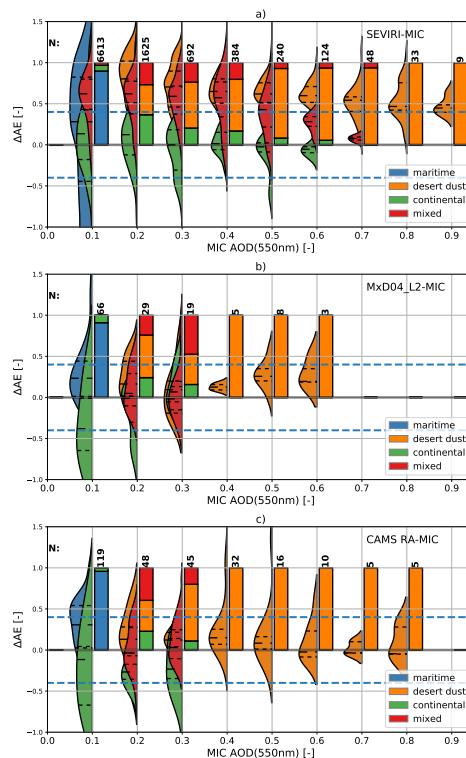


Fig. 1. A plot like Figure 5 but with AE difference and aerosol type

C32