Interactive comment on "Comparison of south Atlantic aerosol direct radiative effect overclouds from SCIAMACHY, POLDER and OMI/MODIS" *by* Martin de Graaf et al.

Martin de Graaf et al.

graafdem@knmi.nl

Reviewer #1
Interactive comment on **Comparison of south Atlantic aerosol direct radiative effect overclouds from SCIAMACHY, POLDER and OMI/MODIS** *by Martin de Graaf et al.*

This paper examines satellite retrievals of the radiative effect of absorbing aerosols that overlie clouds (here termed the DRE). Retrievals from OMI+MODIS, POLDER and SCIAMACHY are compared. The latter can observe at many different wavelengths, but

has low resolution. POLDER can observe the degree of polarization of the reflected light, which allows extra information about the aerosol and cloud to be obtained and minimizes the retrieval assumptions that need to be made. It is found that OMI+MODIS and SCIAMACHY agree reasonably well, but that POLDER produces larger DRE and cloud optical thicknesses (COT). Some of this difference is attributed to sampling issues (mainly arising from the different resolutions of the instruments) and some due to the larger optical depths retrieved by POLDER.

The study should be useful to other researchers since it would be useful to know how large this warming effect is (can it offset a significant amount of aerosol-cloud cooling?) and whether the models get it right. It also seems like the POLDER approach has some promise, particularly if it can be combined with more conventional instruments on e.g., the upcoming METOP-SG 3MI platform. As such I think it should be published after the suggested revisions.

However, the arguments are often a bit muddled and it would be good to see the reasons for the larger POLDER COT values explored a little more, as well as some more investigation into the effect of the low resolution retrievals from the other instruments. The paper talks a lot about 'sampling errors' for OMI+MODIS and SCIAMACHY, but this seems to assume that all such errors are just from averaging of the final DRE or COT values, whereas it seems likely that some retrieval errors may be introduced by the averaging effects of the reflectances to low resolution, particularly if the relationship between the reflectances and the retrieved quantities are non-linear. Such effects occur for MODIS retrievals of effective radius and COD for example (Zhang; doi:10.1029/2012JD017655, 2012). It would be good to discuss this and to look into this possibility. It would even be possible to test what effects the averaging of reflectances to lower resolutions might have using synthetic higher resolution reflectances. On a similar note – considering just the 'sampling effect' (i.e., just the effect of averaging the retrieved quantities, rather than the reflectances), it should be possible to quantify this effect by degrading the POLDER retrievals to the coarser grids, rather

than the other way round, as is currently done.

Section 3.2.3 needs some checking as some of the statements regarding the POLDER optical depth being smaller seemed to contradict the results. The explanations were also not clear.

The reviewer is thanked for the careful and thorough review of the manuscript. Many valuable suggestion were made, which were followed unless stated otherwise, in which a motivation is given. In particular, the regridding of POLDER data to the coarser grid of OMI was performed, to improve the comparison. This was not done the first time, because SCIAMACHY has the coarsest grid, and regridding to SCIAMACHY, and especially requiring SCIAMACHY collocation, yielded too sparse datasets. However, if only POLDER and OMI-MODIS are compared and collocated on the OMI grid, the analysis is much improved.

The manuscript was rewritten, to better distinguish between sampling issues and retrieval uncertainties. Sampling issues arise from the fact that different sensors in different orbits see different parts of the the atmosphere, and that different filter settings yield different pixels taken into account. This can be solved by requiring strict collocation of the considered pixels.

However, such collocation also requires resampling and regridding of data that are originally on different spatial resolutions. As the reviewer points out, nonlinear effects play a role here, and we have included a discussion on the role of the plane parallel bias for heterogeneous clouds. In our analyses MODIS radiances were added and resampled on the OMI footprint, while POLDER COT are averaged over the footprint. This has effects on the COT and CER averaging in a satellite footprint, and can account for the differences found between POLDER averaged COT and OMI-MODIS COT. We have tried to explain and quantify differences that we find. The resulting DRE differences are now explained in terms of the uncertainties in the AOT and COT retrievals. Additional

improvements of the measurements can then improve the DRE retrieval, but this is not the focus here.

All the issues raised by the reveiwer are addressed below:

Specific statements

p.1 L15 Aerosol-cloud-radiation interactions currently present the largest uncertainty in our understanding of Earth's climate (Boucher et al., 2013). The effects of atmospheric aerosols are especially uncertain.

The second sentence here reiterates the first and does not really make sense. It should be removed, or else made more clear what it is referring to. Do you mean that the effects of aerosols alone are especially uncertain (compared to cloud-aerosol interactions)? However, I think that it is hard to argue that this case.
Aerosol effect in global climate models are currently the largest uncertainty in global climate change attribution. However, the poor phrasing was also noted by reviewer #2, and the introduction was rewritten to better reflect the current state of aerosol climate science, and to clarify the text.

p. 1 L18 – 'The presence of clouds has a strong influence on the DRE from the light absorbing species in smoke at TOA.

It's hard to understand what you mean here. I think you mean something like this : – 'The DRE (at TOA) due to the light absorbing species in smoke is strongly affected by the presence of clouds.' Although maybe it would be good to introduce the idea of light absorption (rather than just scattering) affecting the DRE before this sentence. Or maybe this sentence isn't necessary given what follows?
I think an introductory sentence improves the paragraph, and the suggestion by the reveiwer was adopted as given.

p. 1 L20 – 'Over clouds, on the other hand, scattering by aerosols is negligible' – this is not quite correct I think. The scattering due to aerosols overlying cloud would be quite high – it is the cloud that is doing less scattering in this case because of this. I think you mean that the addition of aerosols above a cloud has negligible extra impact on scattering relative to that which the cloud is already causing.

Agreed. The text has been modified to read: 'Over clouds, on the other hand, scattering by aerosols hardly contribute to the upwelling radiation at TOA, since the scattering by clouds is dominant. However, the aerosols absorb radiation, lowering the planetary albedo, resulting in a positive direct effect (warming).'

p.5, L2 – 'CER was derived from collocated MODIS measurements.' Would it not be better for POLDER to retrieve the CER? Is this retrieval not possible? Could MODIS CER be biased by the overlying aerosol, or by inhomogeneous clouds, etc.?

POLDER does not have measurement in the near infrared. MODIS CER is retrieved primarily from the 2.1μm channel over the ocean. It can potentially be biased by the presence of aerosols above clouds. However, in the region of interest, the aerosols typically observed above clouds (i.e. biomass burning aerosols) are characterised by a large Ångström exponent. Therefore, their contribution to the signal at 2.1μm is expected to be negligible. This is the same argument that is used for the (OMI-)MODIS retrievals, except at 1.2μm. At 2.1μm the effect will be much smaller. Regarding the 3D effect, several filters are used on the POLDER AAC products in order to reject inhomogeneous clouds (Waquet *et al.*, 2013b, GRL)

p.5 L30 – 'MODIS, on-board the Aqua satellite, flies in formation with Aqua in the A-Train, leading Aqua by about 15 minutes' Should this be MODIS flies in formation with and leads Aura?

Correct, it should be (and is now): MODIS, on-board the Aqua satellite, flies in

formation with Aura in the A-Train, leading Aura by about 15 minutes.

p.6 L14 - 'Note however, that such an estimate is often missing, while methods other than DAA are moreover highly uncertain due to their dependence on the correct characterization of the spectral properties of the overlying aerosols.'
This doesn't quite make sense. Do you mean that often such an error estimate is not made in other studies (does this only apply to those that use DAA)? Please correct if so. The part after should probably be a separate sentence.

This is correct, an estimate on the individual measurements (of DRE in this case) is often missing. Many satellite products are delivered without uncertainty estimate on the individual measurements, e.g. relevant for this manuscript: OMI, MODIS, and CALIOP above cloud AOT. Uncertainty estimates are obtained from comparison with other datasets, like is done in this manuscript. However, we argue that error and uncertainty estimates can, and should, also be given on the basis of assumptions and uncertainties of the input parameters, which lead to measurement uncertainties. In that case, comparisons like the current one, can be performed in light of the uncertaities of the measurements.

Here, we have tried to quantify the uncertainties in aerosol DRE in terms of uncertainty estimates in above cloud AOT and COT for POLDER, and relate the difference between OMI-MODIS and POLDER DRE in terms of those uncertainties.

p.6 L16 – 'Other minor error sources for the DAA method are the uncertainty in input parameters, the influence of the smoke on the estimated cloud fraction, cloud optical thickness and cloud droplet effective radius, an uncertainty in the anisotropy factor (de Graaf et al., 2019), and the uncertainty of estimating the COT and CER at SWIR wavelengths.'
In Section 3.2.3 you say that the DRE depends very strongly on the COT. So, wouldn't

the COT uncertainty be likely to have a larger contribution to the error than indicated here? Also, this sentence needs to use semi-colons to make it clearer to become :-

'Other minor error sources for the DAA method are the uncertainty in input parameters; the influence of the smoke on the estimated cloud fraction, cloud optical thickness and cloud droplet effective radius; an uncertainty in the anisotropy factor (de Graaf et al., 2019); and the uncertainty of estimating the COT and CER at SWIR wavelengths.'

Yes and no. The (large) effect of the COT uncertainty on DRE is in the uncertainty estimate of the DRE by applying it to aerosol-free cloud scenes, which yields the rather large uncertainty of 13 Wm^{-2} for OMI-MODIS DRE. The additional cloud uncertainties investigated in De Graaf *et al.* (2012) are the effect *of smoke* on the cloud parameters, and uncertainty of estimating CER and COT in the SWIR *instead of in the visible*. However, it is agreed that this is not clear from the text. Futhermore, we show in this paper that the effects of COT and AOT are coupled and COT uncertainties will have larger effects at larger above clouds AOT, which was not estimated. This was added to the manuscript.

p. 8, L2 – 'the instantaneous aerosol DRE over clouds was normalized by dividing by the cosine of the solar zenith angle.'

Have you checked whether the DREs scale linearly with the cosine of the angle (presumably a proxy for the incoming SW)? This could be checked with a radiative transfer code. If not then this might introduce some bias. Presumably there is a lower limit for the solar zenith angle allowed?

The DRE is defined as the difference in upwelling flux at TOA for a cloud scene and a cloud with aerosol scene, which can be written as

C7

$$\mathrm{DRE_{aer}} = \mathrm{F}^{\uparrow}_{\mathsf{cld+aer}} - \mathrm{F}^{\uparrow}_{\mathsf{cld}} = \mu_0 \int\limits_{SW} E_0(\lambda)(A_{\mathsf{cld}} - A_{\mathsf{cld+aer}}) \, \mathrm{d}\lambda \qquad (1)$$

where A is the local planetary albedo, defined in terms of the reflectance $R = \pi I/\mu_0 E_0$ as $A(\mu_0, \lambda) = \frac{1}{\pi}\int_0^{2\pi}\int_0^1 R(\lambda; \mu, \phi; \mu_0, \phi_0)\mu d\mu_0 d\phi$. Neglecting the small effect on the planetary albedo (which is the effect of the BRDF of the cloud scenes, that has been treated in De Graaf *et al.* (2012) and De Graaf *et al.* (2019)), the fluxes can be noon-normalised by deviding DRE by μ_0 to cancel the effect of different solar incoming fluxes $\mu_0 E_0$ on the DRE computation during the overpasses of SCIAMACHY (10:00 LT) and OMI, MODIS and POLDER (around 13:30 LT).

The solar zenith angle has no lower limit, since this does not produce any problems. High solar zenith angles may introduce biases, but these do not occur, because the considered area of the south-east Atlantic basin is near the equator.

p. 8, L26 – 'The main reason for the much larger area-averaged POLDER DRE on 12 August 2006 is the smaller coverage of the area by POLDER, compared to that by OMI/MODIS, due to a smaller swath. This was illustrated in Figures 1d and 1e.'

This is part of the reason, but it seems that generally POLDER gives considerably higher values for the same regions. You could help demonstrate the magnitude of the differences caused by the different swath area vs that of POLDER values being higher by giving the collocated averages in Table 1. It would be better to change 'The main reason' to 'One of the reasons'.

All the sampling issues are resolved when the OMI-MODIS data are regridded to POLDER or vv, which is shown in the manuscript. The description here is merely to illustrate the main difference between the two datasets on 12 August 2006, which is not the day with the most extreme DRE values, but the day with the largest difference between the area-averages. The reason is the sampling, which is very different for

C8

OMI-MODIS and POLDER, as shown in Figure 1d–e. When the sampling issue is removed the average for that day reduced from 74 Wm^{-2} difference to 33 Wm^{-2} difference. So the sampling issue is the main reason for the large descrepancy.
The numbers are added to table 3.

p. 8, L33 – 'This means that a (dense) plume may be sampled once by a far off-center pixel, or by 15 nadir pixels, all of them receiving the same high values, depending on the satellite track.'

For this to have an effect on the average it would require that the values retrieved from the mean reflectances over the larger pixel did not produce the correct average DRE value – i.e., there is a non-linear relationship between reflectance and the retrieved products, so that the result is dependent on the averaging scale (pixel size). It would be worth nothing this here. Also, the sentence would be clearer without 'all of them receiving the same high values,'.
This unclear statement has been removed. The text has been changed to just state the different pixel sizes. The discussion has been changed to showing the effect of regridding, both SCIAMACHY and OMI-MODIS to the high-resolution POLDER grid, as POLDER to the OMI grid.

p.10, L20 – 'This issue could be resolved if all values were regridded to the coarsest available. However, since this is the SCIAMACHY grid, not many grid cells would remain.'

Although you could do it for the OMI grid vs POLDER, which would be useful?
Yes. The main addition to the new manuscript is the regridding of POLDER DRE to the coarser OMI grid. The text was rewritten to include this analysis. Without the SCIAMACHY collocation requirement the coverages of both datasets are still very good after collocation and sampling issues are removed. It shows that regridding to

POLDER grid or OMI grid does not change the results very much, only the removal of the SCIAMACHY collocation requirement ensures much better statistics.

p.11 L19 – 'A comparison of SCIAMACHY, OMI/MODIS and POLDER COT histograms (not shown) revealed a slightly higher COT from SCIAMACHY and OMI/MODIS compared to POLDER (up to 42 for POLDER and 48 for OMI/MODIS (Schulte, 2016)), but the maximum of POLDER is restricted due to LUT limits.'

It's not clear here where or when these histograms apply to. I see that it is likely to refer to the 19 August case (Table 3), but it needs to be mentioned in the text. Also, 'a slightly higher COT from SCIAMACHY' should be changed to 'a slightly higher maximum COT from SCIAMACHY' since it otherwise it sounds like you are referring to mean values. However, visually it looks from Figure 5 like POLDER has higher maxima in general? You should also explain the part about the LUT limits in the context of the statement on p.10 L24 ('The POLDER DRE is dependent on the retrieved AOT and COT, which in principle are both unbounded.').
The discussion on COT was completely rewritten and this text was removed. We believe the COT analysis is now much clearer, the relationship between POLDER COT and OMI-MODIS COT is now shown in Figure 6, with POLDER COT being systematically higher.

The statement on p.10 L24 is rephrased as: The POLDER DRE is dependent on the retrieved AOT and COT, which in principle are both unbounded (although in the LUT for POLDER THE COT is limited to 42).

p. 11 L28 – 'Even though the OMI/MODIS data are regridded to a high resolution grid, the values are obviously still more smoothed compared to the COT on the native high resolution POLDER grid. Therefore, even though POLDER COT and POLDER DRE are generally smaller than from OMI/MODIS on average, the extreme values and

averages are higher.'

The second sentence seems to contradict the rest of the paper – from the tables and figures POLDER has a generally larger DRE and COT?
The discussion on COT was completely rewritten and this text was removed.

Table 3 – it should be made clear in the table caption that the DRE was calculated using the POLDER AOT in both cases.
Agreed, this was added to the caption, and the value itself was placed in the center, so it is obvious it was not retrieved for OMI-MODIS.

p.12 L8 – 'It shows that the difference between these two quantities disappears completely for these instruments, and the slope is even reversed.'

- It has reduced a lot, but not disappeared completely! Plus, saying that the slope has reversed is a bit unclear. Perhaps better to say it went from <1 to >1.
The discussion on COT was completely rewritten. Only OMI-MODIS and POLDER data were used to examine the DRE disfferences. The relationship between POLDER COT and OMI-MODIS COT is now shown in Figure 6.

p. 12 L20 – 'The aerosol DRE from POLDER is completely independent. It correlates well with SCIAMACHY and OMI/MODIS DRE for moderate values, but is larger than SCIAMACHY and OMI/MODIS DRE for high values. This is caused by a larger COT retrieved by POLDER, and to a lesser degree by an underestimation of the aerosol DRE using DAA, which by definition assumes a zero AOT at SWIR wavelengths.

The largest contribution to the difference between SCIAMACHY, OMI/MODIS and POLDER DRE are sampling issues.'

- It seems that the last sentence contradicts the ones before where it says that larger

COT retrievals by POLDER are the cause. Is it the COT differences or the sampling issues that are most important? Or are they equally important? See also p.13 L8. Also, L9 in the abstract says that sampling issues are the most important – is this actually the case and can you point to the evidence that shows that it such errors are larger than the COT errors?
Sampling differences are the main source for DRE differences, which is now clearer from the case of 12 August 2006, see above. Only after sampling is removed an analysis of the differences in terms of the input parameters and assumptions makes sense. After removing sampling issues, COT differences are the largest contribution to the found differences. AOT differences have a smaller but still important effect, and AOT and COT influences are coupled. The new manuscripts makes these conclusions more apparent, thanks to the suggestions of the reviewers.

p.13 L5 – 'This approach removes issues related to selecting high positive DRE values by filtering on COT and CF, which introduce large differences in the average DRE.'

It's not clear what you are referring to here regarding filtering of COT and CF – is this a method that has been suggested in the literature (please say so and give a reference if so). Or from this paper – again this needs to me made clear.
The sampling by the three instruments is different. Not only because they sense different areas, but also because even if the same limits on permitted cloud fractions and COT are used, the results will be different, simply because the retrieved CFs and COTs may be different. Therefore, even perfectly aligned instruments with exactly the same filter settings will sample different parts of the Earth. These issues are also resolved with the collocation requirement used in the analysis.

A sentence explaining this was added to the manuscript: 'Even if the same filtering is used for the CF and COT for all instruments, different areas will be sampled, because the CF and COT retrieved by the different instruments may be different.'

p. 13 L13 – 'Normally, MODIS COT retrievals at 0.8 and 1.2 microns retrievals' - Doesn't the usual MODIS retrieval over oceans use the 0.86 and 2.1um bands?
Correct, this has been changed to 2.1 microns.

Figures *Fig. 2 – The linewidths of the monthly mean lines need to be quite a bit thicker for the colour and dash style to be visible.*
Agreed.

Fig.3 – the legend lines need to be thicker to be able to see the different colours.
Agreed

Typos

The word 'microns' is used a lot, but also the symbol 'μm'. I think that the latter is the ACP standard for units.
Agreed. All are changed to μm.

p. 7, L20 – 'when the comparison between the instrument is worst' -> 'when the comparison between the instruments are worst'
'instrument' was changed to 'instruments'. (the comparison ..) 'is' was retained.

p. 7, L30 – 'Here, we show the effect of ignoring the sampling effect, even of area averages of, in this case, aerosol DRE over clouds.'
The text was changed.

– this doesn't quite make sense. How about something more simple like 'Here we show the effect of ignoring the sampling differences between instruments'?

Agreed. I also don't understand this sentence. The reviewer's suggestion was adopted.

p.10 L23 – 'possibly' -> 'possible'.
Correct. Text was changed.

p. 11 L12 – 'This way, an AOT at 1.2μm can be found between 0.15 and 0.35' -> 'In this way an AOT at 1.2um of between 0.15 and 0.35 can be found'
Agreed.

p. 11, L22 – 'However, the spectral variation in COT is very small. Only for very small cloud droplets the COT at 0.87microns is about 4% smaller than the COT at 1.2microns for cloud droplet effective radii of 4 microns, and this reduces for larger droplets.'

- This would be better as :- 'However, the spectral variation in COT is very small and is only significant for very small droplets. For example, for cloud droplet effective radii of 4 microns the COT at 0.87microns is about 4% smaller than the COT at 1.2microns and this reduces for larger droplets.'
Gladly accepted.

p. 13 L23 – 'Comparing AOT over clouds POLDER with MODIS and CALIOP, showed POLDER to be high, but not necessarily overestimated' – insert 'from' between 'clouds' and 'POLDER'.
Agreed. Text was changed.

Reviewer #2
Interactive comment on **Comparison of south Atlantic aerosol direct radiative effect overclouds from SCIAMACHY, POLDER and OMI/MODIS** *by Martin de Graaf et al.*

This short study is a comparison of above-cloud aerosol direct radiative effects estimated by three methods applied to three satellite sensors or combinations of sensors (POLDER, SCIAMACHY, and OMI/MODIS). Looking at two days in August 2006 and at daily averages over 4 months in 2006, the authors find sizeable differences between the three sets of estimates, with POLDER retrievals producing significantly stronger radiative effects. Those differences are reduced when correcting for sampling differences. The remaining differences can be explained by differences in aerosol and cloud optical thickness, with cloud optical thickness being the dominant cause. The study is of interest to the wider aerosol community because aerosol modellers have now begun to use above-cloud aerosol retrievals to compare against their models, and large differences between observation-based estimates weaken observational constraints. This study is hopefully a first stage to eventually reconciling the different estimates. The paper is generally well-written, although language editing will help in places, and Figures and Tables illustrate the discussion well. My main criticism of the study is that it does not attempt to bring additional information to resolve the disagreement. The discussion can also be improved in places. I recommend major revisions because addressing my main comment will probably require additional analyses.

1 Main comment The study concludes that differences are mostly caused, once the effect of sampling has been accounted for, by differences in cloud optical thickness (COT) retrievals between the instruments. Differences in aerosol optical thickness (AOT) also play a role, especially at longer wavelengths. But it would be most useful to know which dataset does best. Retrievals of AOT in nearby clear-sky regions, or using CALIOP, or even nearer the sources by AERONET should help determine whether

the large AOTs (almost 2) retrieved by POLDER are realistic. Similarly, differences in retrieved COT are large enough to determine whether POLDER is realistic or not by comparing to CALIOP or passive retrievals, e.g. from SEVIRI. Adding such an analysis would make the study a more ambitious, and ultimately more useful, contribution.

The reviewer is thanked for the careful evaluation of the manuscript. We have changed many of the discussions, which were indeed sometimes vague, because the results were not always clear. After reanalysis of the sampling issues, by gridding POLDER to OMI and disregarding SCIAMACHY to get a good comparison between two datastes with a good coverage, the results are much clearer, which are -hopefully- clearer presented in the new manuscript.

After sampling, the remaining differences between POLDER and OMI-MODIS DRE are explained in terms of AOT and COT and the uncertainties in those. The uncertainties in COT retrieved by POLDER and OMI-MODIS for polluted clouds are difficult to establish, because there little to compare with, 'normal' COT retrievals being biased by overlying smoke. Both POLDER COT and OMI-MODIS COT retrievals show continuous behaviour from polluted to unpolluted areas. We show the difference in COT retrieval (9% on average, no extreme differences) and the effect on the DRE.

For AOT, the comparison with more established datasets is also difficult, because these are all in clear sky. However, an inspection of several AERONET sites showed high AOT during the biomass burning seasons, but never as high as 2 over the Atlantic. Ascension Island (almost 3000 km from the source) has no measurements during 2006, but other years AOT is measured up to 1 (UV). e.g. in 2016, which was also an anomalously extensive biomass burning season. St. Helena has few measurements, São Tomé has measurements in 2017 and 2018 up to 1.5 (UV). Only over Gabon, which is most likely the source region or in the path of the smoke towards the ocean, AOT at 340 and 280 nm of more than 2.0 was found in August 2016 (data start in 2014).

We have compared the above cloud AOT (ACAOT) from several sources that are currently available, from MODIS, OMI, and CALIOP, all of them science datasets, i.e. no proper validation has been performed for these datasets. High ACAOT up to 1.5 or 2 is common, with POLDER being in line with the highest retrievals. The discussion in the manuscript was extended with these numbers from the literature.

The DRE results and differences are explained in terms of these findings, but no conclusion was given on which dataset is 'best', because the truth is not known.

2 Other comments *Page 1, line 15: The statement 'The effects of atmospheric aerosols are especially uncertain' repeats the first sentence and can be deleted.*
Agreed. The introduction was rewritten to be more clear and correct.

Page 1, line 21: I acknowledge that the terminology of aerosol direct, indirect, and semi-direct effects is now well known by the wider atmospheric science com- munity, but I recommend defining them anyway for the sake of completeness.
Agreed. The introduction was rewritten to include this.

Page 2, line 2: 'which can be characterized relatively well' sounds like an in- stance of concluding too quickly!
We derive a direct effect of aerosols over clouds including an uncertainty estimate. I think this a relatively good performance, given that semi-direct effetcs are not even estimated at all from measurements.

Page 2, line 5: Caution: the use of 'forcing' in the sense of Forster et al. 2007 implies that the unperturbed values correspond to pre-industrial conditions. In the present study however, unperturbed values are for an aerosol-free atmosphere, so to avoid confusion I recommend avoiding the word "forcing".

In 2007, Forster *et al.* defined 'radiative forcing' as the net broadband irradiance change ΔF at a certain level with and without the forcing constituent, after allowing for stratospheric temperatures to readjust to radiative equilibrium, but with tropospheric and surface temperatures and state held fixed at the unperturbed values (chapter 2.2.). This was quoted in the manuscript, with 'radiative forcing' changed to 'radiative effect', because the terminology has changed since then in favour of 'forcing' as the change since per-industrial times, and 'effect' as the instantaneous change. However, one instance of 'radiative forcing' on p2,l7 was overlooked, and this was changed to 'radiative effect' in the new manuscript.

Page 2, line 33: Myhre et al. (2013) is not the correct reference for that statement, as that paper only refers to global averages and does not isolate cloudy- sky radiative effects. I think the authors mean Figure 2 of Zuidema et al. (2016) doi:10.1175/BAMS-D-15-00082.1 . The same comment applies to Page 13, line 28.
Agreed.

Page 3, line 15: 'Finally, the . . . using an RTM.' That has been said already.
Agreed.

Page 3, line 16: 'highest yet'. What do you mean? Over which period are you making that statement?
Over the south-east Atlantic in 2006, as stated in the manuscript.

Page 5, line 4: '(from models)'. Be more specific.
This has been removed.

Page 6, section 2.4: Isn't it possible to get an error/uncertainty for the POLDER prod-

uct?

The main source of error for the aerosol DRE over clouds from POLDER is the assumption on the aerosol refractive index. In the first step of the algorithm, an assumption on the refractive index is used in order to retrieve the above-cloud scattering AOT. In the second step, the imaginary part is modified in order to retrieve the absorption AOT from total reflectances, assuming the same real part of the refractive index as in the first step. The impact of the refractive index assumption on the DRE has been analysed in Peers et al (2015) and a maximum error of 10 Wm^{-2} has been observed. Finally, an error on the CER can cause a bias of up to 10% on the COT.

This was added to the manuscript.

Page 7, lines 3–4: How were the two cases selected?

The first case shows the situation during the largest difference between the datasets, and the second case the situation one week later, when the differences are small. During 2006 all instruments performed well, and August is the peak of the biomass burning season in southern Africa.

This was added to the manuscript.

Page 7, section 3.2: That section is confusing. It goes back and forth between case studies and monthly averages. I suggest starting with case studies, then discussing the implications for longer time averages.

The discussion now starts with the cases, and the discussion on the aera-averaged DRE goes back to the cases, which are part of the dataset. In Figure 2, the 12th and 19th are indicated more clearly, so the reader understands that Figure 1 and 2 are connected, and where.

Page 7, lines 30–31: 'even of area-averages': I do not understand that statement.

As stated, sampling effects are often treated by averaging. Here, we show that this is not sufficient for the sparse DRE over the Atlantic. The text 'even of area-averages' was removed.

Page 9, section 3.2.1: The comparison protocol is unusual. The usual method is to regrid higher resolution datasets on to the coarser grids. The reason for doing like that is that the higher resolution represents variability within the coarser grid- box, so it is safe to make an average. But the authors do the other way around, replicating coarser values to fill the higher-resolution grid. Why that choice?

The reason was to avoid the very coarse SCIAMACHY grid. However, another analysis was added without SCIAMACHY and with POLDER gridded to OMI, which is the 'normal' way, and still has a large coverage from both instruments. This improved the comparison considerably. Both the reason for the first choice, and the new comparison were added to the manuscript.

Page 11, line 3: 'it has been shown' requires a reference.

The statement was from the reference just before the sentence. The reference was moved to include this statement as well.

Page 11, section 3.2.3: Why not show 12 Aug 2006 on Figure 5? The DRE difference is even larger on that day, which should help identify differences in COT as the main cause.

Agreed, the figure was changed.

Interactive comment on Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2019-545, 2019.