We would like to thank the Reviewer for his/her thorough report that helped us improving the quality of our study. Through his/her constructive comments and suggestions the submitted manuscript has been updated significantly. Below are given point-by-point replies (regular font) to the comments (bold font) raised by the Reviewer.

Reviewer #1

This study describes a new dust optical depth (DOD) data set, MIDAS, which is derived by taking MODIS (Aqua only) satellite-based aerosol optical depth (AOD) and the aerosol speciation from the MERRA2 reanalysis. CALIOP and AERONET are used for evaluation.

The manuscript is in scope for the journal, though would be a closer fit to the other Copernicus journal ESSD because it is mostly a data set description paper. The material is important because speciated AOD is one of the next frontiers for better climate and air quality applications of data sets. The quality of language and visuals is satisfactory overall, though some edits are needed, and figures 5, 7, 9 would benefit from labels being increased in font size (hard to read without zooming in). Some of the content in the Supplement should be in the main paper. Overall, I recommend major revisions and would like to review the revision.

We agree with the Reviewer that our study fits well also with the scope of ESSD. Actually, the manuscript had been submitted to ESSD. The problem with ESSD was the long delay (3 months +) finding an editor. For this reason, we took the decision to withdraw the paper and resubmit it to AMT. Regarding the quality of the figures, we have reproduced all of them and their illustration has been improved.

General comments:

The main technical weak point of this study is that all the observational data sets used are out of date: MODIS Collection 6 instead of 6.1; CALIOP version 3 instead of version 4; AERONET version 2 instead of version 3. So this affects the AOD source used (MODIS), the optical properties used for matching (AERONET), and the data sets used for evaluation (AERONET, CALIOP). Some of the differences between old and new versions are systematic. So it is not clear to me how different the derived data set, or the evaluation results, would be if the newest data versions were used. To my best knowledge all of these latest data versions have been available for 1.5 years or so (i.e. they are not that new), so it is unfortunate that outdated versions were used when this analysis was done. It sounds like the authors are using a post-processed CALIOP product from another group (LIVAS?) rather than the official NASA CALIOP data products, so maybe that can't be changed. But, if the authors intend for others to use MIDAS for scientific analyses, it would really be best to use the most up to date inputs. I know that this means more work downloading files and rerunning code but this should mostly be computer time if the analysis code has already been written. So my main recommendation is to do that. I guess it is up to the authors and editor to decide what is most reasonable here. The 2007-2016 time period could also possibly be extended, I see no reason why it couldn't cover more of the Aqua record. Longer time series are of course more beneficial for things like trend analyses.

We decided to follow the reviewer's suggestion and in the revised manuscript we have used the MODIS-Aqua C061 data as well as the AERONET Version 3 retrievals. Moreover, the temporal availability of the MIDAS dataset has been extended from 10 (2007-2016) to 15 years (2003-2017). Therefore, the major comment raised by the Reviewer has been addressed adequately to our opinion. For the evaluation of the MDF we have used the CALIOP data which have been post-processed from our group and are provided via the LIVAS database (Amiridis et al., 2015). In the submitted manuscript, we stated (Lines 248 – 250) the published works describing the methodology for the derivation of the pure dust product (accounting for dust plus its portion from dust mixtures; Amiridis et al., 2013) as well as the series of filters applied in order to analyze only the quality assured CALIOP profiles (Marinou et al., 2017). The aforementioned techniques are also briefly discussed in our manuscript (Section 2.3). The in-house developed LIVAS database has been built using

CALIOP V3 data and its temporal availability spans from 2007 to 2015. Currently, the group responsible for the ESA-LIVAS database is working on the development of an updated version, covering the entire CALIPSO CALIOP observational period, in which the CALIOP V4.2 profiles are used. We acknowledge that there is a confusion to the reader regarding the terms "CALIOP" and "LIVAS" which has been addressed in the revised document following the recommendation made also by the Reviewer 2.

Numbers are often given to too many significant digits. One example I'll mention again later includes referring to an offset as 4.264%. Including all these digits gives an unrealistic impression of the precision of these estimates: can you really say that the true population offset is 4.264% and not 4.265%? Is it important that it is 4.264% and not 4.265%? If the answer to either of these is no, this is an indication that there are too many significant digits being reported. The authors should consider all numbers presented in this manuscript. For this case, for example, I'd probably just say 4.3%. This will also make the paper more readable.

We agree with Reviewer. We have kept only one digit in all numbers mentioned in the text.

I downloaded some MIDAS data from the link in the paper to have a look. The contents of those files seemed as described. I have four suggestions based on looking at these files:

1. **I didn't see a MIDAS file version identifier, but the Readme file notes that some things are in testing or will be added in a future version. So it would be good to add a MIDAS version number somewhere in the filenames so the user can be sure which version of MIDAS they have (and which version of MIDAS technical documents such as this refer to). It may be unclear for the data user otherwise.**

   We agree with the Reviewer that it was an omission from our side not including a file version identifier. We have changed the filenames by adding the MIDAS Version (V1) while the necessary notification is given in the new README file.

2. **One issue with is that the files contain some negative AOD values, which are unphysical. This is a result of the Dark Target land AOD algorithm which allows small negative retrievals. However since this is unphysical I recommend that in the next version, the authors set these values to 0. This is one issue with the source data which is easily fixed.**

   We prefer to keep the negative values and give the option to the user to decide how he/she will treat them. For example, the inclusion of negative AOD values (reducing the positive biases in low-AOD conditions according to previous evaluation studies) in the calculation of long-term averages will give more "accurate" results. On the other hand, for the calculation of temporal or spatial, median or geometrical mean values the negative AODs can be replaced with very small positive values as it has been done in Sayer and Knobelspiesse (2019).

3. **It would also be useful to add an uncertainty estimate to each pixel. There is extensive discussion in the middle of the paper about uncertainty estimates, but these don't appear to have made it through to the data set itself, based on the files I looked at.**

   The pixel-level DOD uncertainty has been added in the netcdf files.

4. **Finally, the files seem to contain some data fields inherited directly from the MODIS aerosol product, e.g. Angstrom exponents. As these are for total AOD and not dust AOD, I wonder if it would be better to remove these. Or, combine the Deep Blue land Angstrom exponent with one of the Dark Target ocean ones. There's also solar and sensor zenith angles, but I'm not sure what these are in there for. This would decrease the size of the archive to be downloaded somewhat.**

As correctly stated by the Reviewer, the Ångström exponents are related to AOD and not to DOD. We are storing them in the MIDAS netcdf files in case where a user would like to work only with AODs (also available in the netcdf files) and use these size parameters in parallel for a discrimination between coarse- or fine-particles dominant conditions. We think that it is better not to merge ocean and land Ångström exponents because they are provided at different wavelength pairs and this might confuse the users. The solar and sensor zenith angles are required for the estimation of the air mass factor (AMF, Eq. 6) according to Sayer et al. (2013) as clearly stated in the manuscript. Each MIDAS daily file has a size of ~10MB which is not "prohibitive" for a fast downloading.

My more specific comments are as follows:

Line 133: should the word "conclusions" be added before "are drawn"?

The missing word has been added.

Lines 143-147: I suggest rewording this sentence. The Dark Target algorithms are really two different approaches as they have different bands used and completely different assumptions between them. Also, Deep Blue is over all snow-free land, not just bright deserts. So really it is one water algorithm (Dark Target ocean) and two land algorithms (Dark Target land, and Deep Blue). It is probably worth acknowledging that there are other MODIS aerosol algorithms too (e.g. MAIAC), they are just not included in those files.

The sentence mentioned by the Reviewer has been rewritten in the revised manuscript as follows:

"*The derivation of AOD is achieved through the implementation of two retrieval algorithms based on the Dark Target (DT) approach, valid over oceans (Remer et al., 2002; 2005; 2008) and vegetated continental areas (Levy et al., 2007a; 2007b; 2010) but relying on different assumptions and bands, or the Deep Blue (DB) approach (Hsu et al., 2004; Sayer et al., 2013) over arid and semi-arid surfaces.*"

We don't see the point of mentioning other MODIS aerosol algorithms in Section 2.1 since it is discussed only the standard product which has been processed in our analysis.

Line 153: I think the authors mean either "increasing pixel size" or "decreasing pixel resolution" here. Not "increasing pixel resolution", which is the opposite.

We have corrected the sentence as suggested.

"*Each swath is composed by 203 x 135 retrievals, of increasing pixel size from the nadir view (10 km x 10 km) towards the edge of the satellite scan (48 km x 20 km), in which a Quality Assurance (QA) flag is assigned (Hubanks, 2018).*"

Line 206 and 212-215: note that the MODIS aerosol product is not assimilated. Rather, it is a neural network retrieval based on MODIS radiances that is assimilated. Not a neural network bias correction based on the MODIS retrieval. So the MODIS information going into MERRA2 is not the same as is being used as the main AOD data set here.

We have modified the relevant part of the text. Below is given the paragraph in the revised document.

"*For aerosol data assimilation, the core of the utilized satellite data is coming from the MODIS instrument multichannel radiances in addition to observational geometry parameters, cloud fraction and ancillary wind data. Over oceans, AVHRR radiances are used as well, from January 1980 to August 2002, and over bright surfaces (albedo > 0.15) the non-bias-corrected AOD (February 2000 – June 2014) retrieved for the Multiangle Imaging SpectroRadiometer (MISR; Kahn et al., 2005) is assimilated. Apart from spaceborne radiances and retrievals, the Level 2 (L2) quality-assured AERONET retrievals (1999 – October 2014; Holben* *et al., 1998) are integrated in the MERRA-2 assimilation system (Goddard Aerosol Assimilation System, GAAS) which is presented in Randles et al. (2017; Section 3). The cloud-free MODIS (above dark target continental and maritime areas, Collection 5) and AVHRR (above oceanic regions) radiances are used for the derivation of bias-corrected AODs, via a neural net retrieval (NNR), adjusted to the log-transformed AERONET AODs.*"

Line 333: note that Levy reference is only for Dark Target over land. For discussion of Deep Blue Angstrom exponent over land, see the Sayer et al (2013) paper that is cited later in the manuscript.

The sentence has been rephrased to:

"*Previous evaluation studies (Levy et al., 2013; Sayer et al., 2013) have shown that size parameters acquired by MODIS are highly uncertain, particularly over land and at low AOD conditions.*"

Line 357: I am not sure it makes to take the quadrature sum of DT and DB uncertainties when they are merged. This means the overall uncertainty is worse than either DT or DB. If you think the uncertainties on these algorithms are independent, then you are effectively averaging two observations which means the uncertainty in the sum should be divided by sqrt(2). Since the retrieval is the average of two algorithms then the uncertainty should represent the uncertainty on that average.

The merged AOD uncertainty in the revised document is calculated based on the following formula:

$$\Delta(AOD_{DTDB-Land}) = \pm\frac{\sqrt{[\Delta(AOD_{DT-Land})]^2 + [\Delta(AOD_{DB-Land})]^2}}{2}$$

which is the uncertainty of the mean of DT and DB using the quadrature.

Line 386-389: If the output is at 0.1 degrees, then there should only be 1 retrieval in each (as the MODIS product is 10x10 km at nadir), so I don't understand this part about decreasing uncertainties when you have multiple retrievals. Or is this about when there are overlapping retrievals at the edge of swath from consecutive orbits? If so, that should be stated. If this is about averaging to a coarser space/time scale, then I don't think it makes sense to use the root n factor here because we know there is high spatial correlation in the errors because the errors are mostly not noise.

In this sentence we are describing the calculation of the DOD uncertainty at each pixel and at various temporal scales (i.e., monthly, seasonally, annually). Therefore, we are dealing with a time-series in which the applied algorithm (when possible) over land through time (i.e., from day-to-day) can switch from DT to DB depending on the NDVI threshold (see Sayer et al., 2014) while over oceans the AODs are retrieved always via the DT-Ocean algorithm.

Section 4.1: I am not sure that it is useful to compare total MERRA2 and MODIS AOD in this way. Or at least, the framing of the purpose here is not right. If there is a systematic disagreement, then that tells you that there might be an error in the derived MERRA2 dust fraction as well. Would it not be more meaningful for the present analysis to compare MERRA2 and MODIS dust AOD rather than total? Or to report summary results of the evaluation of MODIS AOD against AERONET (from DT/DB team studies)? As written, section 4.1 doesn't fit well with the rest of the paper.

We agree with the Reviewer and we have removed Section 4.1 from the revised manuscript. A similar comment regarding the usefulness of comparing MERRA-2 AOD versus MODIS in the current study has been raised also by the Reviewer 2. The intercomparison between MERRA-2 and MIDAS (MODIS) DODs, along with LIVAS (CALIOP), is already presented in Section 4.4.

Section 4.3: The authors here frame the differences as if MIDAS is in error. However, unlike the direct-Sun AERONET AOD data, the AERONET almucantar scan retrievals used here have non-negligible uncertainties (which are not necessarily random). So some of the discrepancies and biases might in fact come from uncertainties in the AERONET DOD estimates. This was not directly discussed beyond a mention that the AERONET DOD estimates made here neglect fine-mode dust, although I think with the AE filtering this is likely to be a negligible effect in most cases.

The comparison here includes two aspects. The use of the direct sun AERONET retrievals for AOD and the coincident inversions (using the SSA). Regarding AOD, the uncertainty is reported to be between 0.01 and 0.02, Eck et al. (1999). For the SSA, based on the results of Sinyuk et al. (2020) for the Mezaira site (i.e., predominance of dust aerosols), its uncertainty (being lower than 0.06) decreases significantly for increasing AODs. Since the $SSA_{675} – SSA_{440}$ difference (i.e. positive values) is used as a criterion for the discrimination of dust from sea-salt particles, the obtained SSA uncertainties, particularly those at 440nm, can affect the spectral signature of SSA and subsequently dust identification. Therefore, in some cases the AERONET DODs can be misclassified.

So to summarize, still AERONET AODs have lower uncertainty than the MODIS retrievals. For the case of the use of AERONET inversions, spectral SSA related uncertainties can lead to a misclassification of such cases.

Lines 666, 667: here the authors say that MERRA2 has "biases" and "overestimates" compared to CALIOP. It would be better to refer to positive and negative "offsets" or "differences" instead, because "bias" and "overestimate" imply a problem and that CALIOP is the truth. Really none of the data sets are the truth and we are only making comparisons and not diagnosing errors. So more neutral language like "offsets" should be used here (and throughout), and terms like "bias" and "overestimate" should be avoided unless it involves a comparison with something that can be considered a reference truth. I mentioned only these examples although there are others in this section and through the paper where these or similar terms are used (and there are places where the wording is ok as well).

We agree with the Reviewer's comment and we have made the appropriate modifications throughout the paper.

Line 716: authors should check and clarify which of the data sets corresponds to which number here. For example the wording implies that 4.264% is more than 9.405% which is obviously backwards.

Thanks for the correction!

Sections 4.2 to 4.5 were honestly a little hard to read because it's a large amount of text which is basically describing several figures and providing references. This also comprises about 11 of 28 pages of body text in the paper. I wonder if this can be streamlined a bit. The authors write that there will be a follow up paper looking at this same material in more detail as well. So I wonder if here it is best to just show figures and highlight where the data sets do not agree well (and maybe try to figure out why), as these are areas to focus future study on. That type of approach (figure out where and why there are differences) would also make the paper fit better in AMT.

We have made an effort to reduce the length of the text. As it concerns the interpretation of our findings we believe that we are providing all the necessary explanations without just describing plots. For example, there are statements about issues that can affect CALIOP (e.g., lidar ratio, total attenuation of the laser beam, cloud screening) and MODIS (e.g. surface reflectance) performance as well as MERRA-2 reliability (e.g., consideration only natural dust sources).

Table 1: This is a bit of a sea of numbers. It is difficult to easily pull out the main message here. What is the main message here? Or is this just for reference? Given it relies on regional acronyms, it would be better to present the map defining the regions in the main paper rather than in the supplement.

We are providing the long-term annual averages as well as their margins during the study period in order to have an overall view among the three DOD products at planetary scale, for the northern and southern hemisphere as well as for each sub-region. Figure S7 has been moved to the main text in order to help the reader. We think that it is useful for a scientist wanting to use such an information for a particular area or globally to have an easy and direct way of using these numbers.

Figures 2,3: most of the world here is in the 0-0.2 range in Figure 2, which is very hard to distinguish visually because it is different tones of blue. Figure 3 solves this but as it's a separate figure, it is card to glance back and forward. Also, I am not sure how helpful it is to show an annual map here because dust seasonality is strong. So I suggest making seasonal maps instead of annual to replace these figures, this would give more insights. I am also not sure whether the maps of FB and FGE are needed for this panel. Maybe just replace this with an 8-panel figure: left column is seasonal MDF, right column is seasonal MDF minus CALIOP dust fraction (i.e. mean bias)? Then Figure 3 could be 4 panels showing seasonal correlation coefficients? I think they are the most crucial metrics to show here because they show the level of consistency in typical dust fraction and the variation captured by MERRA2, which are what inform the DOD uncertainty here. The other panels could maybe move to the Supplement if the authors think they are useful. I know there are a few seasonal maps in the Supplement but think the maps discussed above should be in the main paper.

Following the suggestion made by the Reviewer, we have added in the Supplement a panel of figures presenting the biases and the correlation coefficients obtained on a seasonal basis. Also, the most important findings are briefly discussed in the main text. The FB and FGE metrics are less affected by outliers with respect to bias and serve as a complementary diagnostic tool.

Figure 4 (and text discussion): I don't think the linear regression is appropriate here, so it should be removed. Since the uncertainty on DOD is proportional to total AOD, it is likely that the assumptions of regression are violated. Also I don't think a global regression is useful because it is likely there are regional differences in the errors, meaning that the global regression line is not informative. Same comments apply to Figure S5 in the Supplement.

The way we define the uncertainty here is exactly the one that MODIS is using for AOD. It is a common practice to compare ground-based and spaceborne AODs (DODs in our case) through scatterplots. Figure 4 shows the overall comparison of MIDAS and AERONET DODs at global scale. Of course there are regional differences which are presented in the calculated metrics at station level in Figure 5. Therefore, all the necessary information is included. Figure S5 has been removed from the revised supplementary material.

Figure 5: the circles are all too small to see.

We have increased the size of the circles.

Figure 8 (iii): this is the mean of the DOD uncertainties, right? Or is it the uncertainty on the mean DOD? This needs to be stated more clearly.

We acknowledge that the description in the submitted document was not clear to the reader and for this reason we have modified accordingly the revised text. Figure 8-iii in the submitted document shows the uncertainty of the DOD average while in the revised text depicts the mean of the DOD uncertainties over the study period. Below is given the relevant part of the revised text.

"*Depending on the selected MODIS algorithm, the appropriate combination between AOD (Eqs. 4, 5, 6 and 7) and MDF (Eq. 8) uncertainties is applied to calculate the Δ(DOD) (Eq. 3) on each measurement (i.e., DOD) and at each grid cell. These pixel-level DOD uncertainties are averaged over the entire study period as well as for each season and the obtained findings will be discussed along with the global spatial patterns (Section 4.5) of dust optical depth in order to provide a measure of the reliability of the derived MIDAS DOD product.*"

Figure S4: this illustrates a problem I have with the validation methodology. A 4-hour averaging window is pretty huge! And the time variation of AERONET DOD in that window can be much bigger than the AERONET uncertainty. So some of the disagreement seen in Figure 4 is due to this time mismatch. For this example, the range of DOD in this window is about 0.09, or 40% of the average. This makes it hard to assess the performance of MIDAS. This is something that shouldn't be buried in the Supplement; I didn't see the mention of a 4-hour window in the main paper (if it is there, it is not clear) so the reader may not realise how big it is. Probably a smaller window is needed, and some filter based on AERONET time homogeneity. I know this will decrease the data volume, maybe a lot, but with such a big time variation in DOD within the window it makes the AERONET comparison a lot less useful for MIDAS evaluation.

It is true that the 4-hour time window is not the optimum and it would be better to be reduced down to ±30 minutes (a temporal margin applied in many evaluation studies). However, there are reasonable arguments which can support our approach. Please note that we are using the almucantar retrievals which have substantial less amount of data with respect to O'Neill retrievals or to sun-direct measurements. This volume of ground-based data is further suppressed when we are applying the criteria for the "determination" of AERONET DOD. By adding a time homogeneity criterion (which probably would be arbitrary), as suggested by the Reviewer, then more data are masked out from our sample. In our case, we had identified the MODIS-AERONET common pairs based on the ±30 min and ±1 hour time-window frames but the number of coincident observations, derived mainly at desert stations, was very small.

We believe that the Figure S4 should remain in the supplement rather than move it to the main text because it is just an illustration of the collocation method. Likewise, we would like to clarify that the MODIS map and the AERONET timeseries both refer to AOD and not DOD. The treatment of both datasets for the derivation of DOD is described sufficiently in the relevant sections of the paper.

On a non-scientific note, I thought the use of MIDAS as an acronym was amusing and a good choice.

Thank you!