Anonymous Referee #2

This is a good, valuable, and important manuscript describing a massive comparison of satellite aerosol retrievals in the case of a major dust event. Given the importance of the aerosol forcing problem, it is essential to quantify the actual uncertainties in our knowledge of aerosols derived from satellite observations. This manuscript contributes substantially towards achieving this goal. It should be published after a relatively minor revision.

Specific comments

1. The title is a bit misleading since the authors compare spatially and temporally averaged retrieval results rather than pixel-level retrievals.

Title changed in 'Intercomparison of desert dust optical depth from satellite measurements'

2. Abstract, line 4. Since pixel-level retrievals are not analyzed, it is problematic to "identify and understand the differences between current algorithms".

'The aim is to identify and understand the differences between current algorithms, and hence improve future retrieval algorithms. '
Changed in:
' The aim is to identify the differences between current datasets.'

3. Abstract, line 6. ". . .hence HELP improve. . ."

this part of the sentence is now is deleted, see above comment.

4. Abstract, final sentence. This statement is not substantiated by the main text since the authors have not analyzed separately the effects of sampling and the effects of retrieval-algorithm differences.

'These differences are partially due to differences in the algorithms, such as assumptions about aerosol model and surface properties. However, in this comparison of spatially and temporally averaged data, at least as significant as these differences are sampling issues related to the actual footprint of each instrument on the heterogeneous aerosol field, cloud identification and the quality control flags of each dataset. '
Have been changed with:
'However, in this comparison of spatially and temporally averaged data, it is important to note that differences in sampling, related to the actual footprint of each instrument on the heterogeneous aerosol field, cloud identification and the quality control flags of each dataset can be an important issue. '

5. Page 718, line 5. Mishchenko et al. (2007) is missing in the reference list.

Added

6. Page 721, lines 20-22. The statement after the comma is not generally true. In the case of MODIS and MISR, fully collocated pixel-level comparisons are possible, in which case sampling, cloud-screening, and quality control data cut issues are completely avoided. This allowed the introduction of the concept of fully compatible MODIS and MISR pixels in

Liu, L., and M. I. Mishchenko, 2008: Toward unified satellite climatology of aerosol properties: direct comparisons of advanced level 2 aerosol products, J. Quant. Spectrosc. Radiat. Transfer 109, 2376-2385.

Direct comparisons of level-2 MODIS and MISR aerosol products have revealed differences comparable to those reported in this manuscript, with quality flags playing a minor role; see Mishchenko et al. (2010) (cited in the manuscript).

Furthermore, the left-hand upper panel in Fig. 3 in

Mishchenko, M. I., I. V. Geogdzhayev, L. Liu, A. A. Lacis, B. Cairns, and L. D. Travis, 2009: Toward unified satellite climatology of aerosol properties: What do fully compatible MODIS and MISR aerosol pixels tell us? J. Quant. Spectrosc. Radiat. Transfer 110, 402-408. Correction: J. Quant. Spectrosc. Radiat. Transfer 110, 1962 (2009)

reveals large differences between long-term spatial averages of AOT for fully compatible MODIS and MISR pixels, including those over areas affected by dust. These averages are large despite the fact that the corresponding Level-2 MODIS and MISR pixels were fully collocated in time and space.

Therefore, it is incorrect to attribute the observed differences in the spatial and temporal averages to sampling issues or quality flags only.

We agree with the referee that the differences in the AOD values in a specific area are due in part to algorithm, instrument, and aerosol model differences.

We did mean that the spatial coverage between the different dataset is remarkable in both daily dataset and monthly mean (see fig 1 and 2), and our sentence was misleading.
'There are remarkable differences in the monthly means obtained with the individual satellite datasets, and this is mainly due to differences in satellite coverage (overpass time, swath) and quality control data cuts. '
Changed in:
'There are remarkable differences in the spatial coverage of the daily products obtained with the individual satellite datasets, and this is due to differences in satellite coverage (overpass time, swath) and quality control data cuts (the last one is apparent from the plot of different dataset for the same instrument in fig 1.). And, even if a pixel by pixel comparison has not been performed in this paper, the differences in spatial coverage have been found to make important contributions to the monthly mean differences.'

Anonymous Referee #1

The manuscript compares a large number of satellite aerosol retrievals for a mineral dust event in Africa. The paper is well written and present interesting results in well chosen figures. The focus on monthly means is particularly commendable. However, I feel that the story is not complete. The conclusion is too short and lacks the discussion

that the abstract promises. More user guidance would also be welcome.

I recommend major revisions to address the main comments below

1 Main comments

- The paper does not live up to its goal. The abstract states that the “aim is to identify and understand the differences between current algorithms”. The identification of differences is there, but the understanding is lacking. The conclusion reads "for SEVIRI ORAC the main issue is a bias over desert in clean conditions, which is attributed to error in the modelling of surface properties". This is interesting and gives an idea of how an algorithm can be improved, but what about the other sensors?

The referee is right and the abstract was possibly overselling the paper and we correct it.
The kind of study we have performed cannot go much deeper, partly because there are too many algorithms involved and partly because the approach employed (analysing spatially/temporal averages) is better suited to identification of discrepancies between a wide range of datasets than diagnosing the specific reason for those differences (which benefits from more targeted study at the individual retrieval level). Nerveless we think there are enough results in the paper that it can be interesting for the scientific community.
We changed the sentence in the abstract (see previous referee) and modified the introduction accordingly.

In addition, the introduction states (page 694, line 13) that “the interpretation of [visible and near-infrared measurements] becomes difficult over bright surfaces such as deserts. To overcome these difficulties more recent algorithms make use of additional information available from certain instruments.” Later, page 721 line 18, the conclusion mentions the “need to improve dust optical properties and surface characterization over land”. Those two statements at opposing ends of the paper strongly suggest that the additional information provided by other instruments has not been enough. A discussion on that point would be very useful.

The reviewer is correct that not *all* the difficulties can be overcome using existing information.
We added this sentence in the conclusion (before line 18 at page 721):
‘The additional information (as for example extending the spectral range used to UV or IR channel), that allow the retrieval of some datasets to be performed over bright surface, at the same time introduce the problem of the variability of other parameters that affect the retrievals (as adequate dust optical model and broad spectral range for surface properties).
There is a need to improve dust optical properties and surface characterization over land, and to extend the comparisons to the retrieved aerosol models in a future study.’

A possible line of discussion is synergies, as mentioned page 722, line 6. The results of this study suggest that a simple synergy made of some combination of satellite products is not enough (see comment on the averaged dataset below). What seems to be needed is synergy at the algorithmic level, where inputs from several sensors help to better constrain a single retrieval scheme. The authors are well placed to give recommendations on this subject.

The reviewer is right but it cannot be done. Sensors are not on the same platform and they don't have the same resolution. The benefits provided by sensors combination are

cancelled by the additional errors/uncertainties resulting from the time difference, cloud screening, non-uniformity of the scene over the different pixel sizes, and cross-calibration. It's the reason why there are projects for a single instrument combining directional, polarization, spectral (solar) informations.

Another useful discussion is to give guidance to a potential user. Let's imagine a MODIS aerosol retrieval user who has grown used to not having retrievals over the Sahara. Reading this paper, he may think that other sensors do not bring much to the table: poor or incomplete retrieval of aerosols over that region, lack of coverage, large differences between instruments. Why should our user be interested in the datasets described here, and how should they choose the dataset they need?

The primary intent of this paper is to investigate the differences between datasets. Guidance for potential users depends on the user's application. For example, if a user wants AOD data to identify dust sources, good daily coverage can be the main consideration, whereas if the user wants the data for climatological study, it might be preferable to choose a dataset that fits better statistically with AERONET measurements and covers all aerosol types. For dust-specific studies, it might be advantageous to choose a dataset that fits the AERONET observations best under dusty conditions. The temporal sampling of different instruments is reported in Table 1, one example of daily coverage is reported in fig 1.
We have added a brief summary of the instrument-AERONET statistical comparison in Section 4, and an example of time series of the dust plume over ocean and over land in the section 6.

- The description of the different datasets is useful, but should be made more consistent between datasets. For example, the current manuscript describes cloud-screening procedures for some datasets, but not others. Same for quality flags. I suggest having a common structure for each description, starting with techniques and assumptions (e.g. prescribed parameters, prior aerosol models), followed by cloud screening and quality flagging, and ending with a list of parameters retrieved for those algorithms that retrieve more than the AOD only. It seems unnecessary to mention validation results at this stage.

Algorithm descriptions have been slightly modified, in particular the missing information have been added, and table 2 has been added to summarise the important points of all the dataset.

- Section 6. I'm not sure what the authors conclude with this averaged dataset. Is it a good idea, in spite of the standard deviation suggesting it is not? Does the land/sea continuity happen purely by chance or is there some underlying reason why it should happen?

The main scope to compute the combined dataset was to identify the region with major spread of data, where there is needed of future improvements.
The averaged dataset is possibly the more reasonable estimate the AOD field for the period considered. This under the assumption that the errors from different algorithm can be somehow smoothed in the averaging with so many dataset. The continuity between ocean/land is qualitatively confirming this as we will expect that the 'true' AOD field is not affected by land/ocean discontinuity. From the other side the single datasets could be

affected, by land ocean discontinuity and moreover the number and the type of dataset available over ocean and land are different.

2 Other comments

- Page 694, line 4: Mineral dust is not a good CCN, unless it is heavily coated with other materials. However, it is a good ice nucleus, and can affect precipitation and atmosphere dynamics through its semi-direct effect alone.

'Dust also has indirect radiative effects by acting as cloud condensation nuclei (CCN) and modifying precipitation. '
Changed in:
'Dust also have semi-direct radiative effect, that can affect atmospheric dynamics, and can act as ice nucleus modifying precipitation. '

- Page 696, line 9: It may be useful to give the AIRS retrieval wavelength as 11 microns.

done

- Page 707, line 14: Any reason why MODIS is not represented here by both its standard inversion algorithm over dark surfaces and the Deep Blue algorithm over more reflective surfaces?

In retrospect it could have make sense to include standard MODIS, however the comparison was based solely on database provided by co-authors in a standard format. Standard MODIS was not one of the dataset submitted. Because of the clear benefit of deep blu for desert over desert we had not make the special effort to add standard MODIS.

- Page 716, line 17: This statement is unclear. Are the authors saying that mineral dust flows around Tamanrasset, and dust conditions are therefore very different there?

Yes.
'Note that Tamanrasset is at an altitude of 1000 m, which could explain why its observations are biased compared to nearby satellite observations, particularly in the case of a desert plume flowing close to the surface.'
Changed in:
'Note that Tamanrasset is at an altitude of 1000 m, which could explain why its observations are biased compared to nearby satellite observations. In particular in the case of a desert plume flowing close to the surface (as typical during winter time) a dust can flow around Tamassaret and result in a significant amount of dust below the AERONET site.'

- Page 718, line 3: A more likely reason why comparisons against AERONET are more successful than comparisons against other instruments is that AERONET measurements are only used in coincident and cloud-free conditions where satellite retrieval algorithms are likely to do a good job. This is a weakness of validations

against AERONET: the more numerous cases where satellite measurements are cloud-contaminated are essentially not validated.

We agree with the referee.
(e.g. in the conclusion we had: Discrepancies between satellite datasets are larger than this agreement with AERONET would imply. This is possibly due to the fact that AERONET itself provides a stringent quality control.)

“The comparisons with AERONET (section 4) show better results than the satellite-to-satellite AOD inter-comparisons. This has been previously documented for MISR and MODIS (Mishchenko et al (2007), Mishchenko et al. (2010), Tanre 2010). It is explained by considering AERONET AOD as ground truth. [...](Kahn et al 2009a).”
Modified in:
“The comparisons with AERONET (section 4) show better results than the satellite-to-satellite AOD inter-comparisons. This has been previously documented for MISR and MODIS (Mishchenko et al (2007), Mishchenko et al. (2010), Tanr\'{e} 2010). It could be explained by considering AERONET AOD as ground truth. [...](Kahn et al 2009a).
Nerveless the more likely reason why the comparisons with AERONET show better results, than the satellite-to-satellite, is presumably due to the fact that AERONET itself provides a stringent quality control and the resulting AERONET vs. satellite comparisons are in cloud free conditions, while the the satellite-satellite results itself can be cloud contaminated. “

- Page 718, line 26: which transport model? What do the authors mean by "assimilate"? Data assimilation requires a matrix of observational errors, which the authors say is not available for the averaged dataset.

With assimilation we means last square fitting between model and measurements, the error for a single dataset are not available, but the SHAMAL model have been assimilate with the DRI combined dataset and use the combined dataset standard deviation as measurement error.
“Moreover, the average of all the datasets has been used to compare/improve/assimilate the transport model (Banks et al. 2009, Banks 2010).”
Changed in:
“Moreover, the average of all the datasets has been used to assimilate the transport model SHAMAL (Banks et al. 2009, Banks 2010) to better understands dust transport.”

3 Technical comments

- Page 702, line 15: Typo: "extrapolated".

Corrected

- Page 710, lines 6 and 21: The subsection numbering should be 2.7.3.1 and 2.7.3.2.

Corrected

- Page 716, line 7: Typo: "Banizoumbou".

Corrected

- Page 717, lines 1 and 5: Typo: "than".

Corrected

- Page 717, lines 14: CC and RMSD have already been used at line 7 and should be defined there. No need to define them again at line 16.

Corrected

Caption of Figure 13: Typo: "than".

Corrected

Other addition of the final version of the paper:

In the discussion of fig 5 (satellite datasets v.s AERONET) we added:

“Satellite retrievals in dust-dominated regions tend to overestimate instantaneous AOD in the low AOD range and underestimate it in the high range compared to AERONET (as already documented for MISR in Kahn et al. 2010). A greater diversity of dust optical models is needed to better represent different desert source regions, though other factors might also be involved.”

At the end of the paragraph 3 we added more discussion on the results of individual’s dataset.

“The AIRS AOD retrieval algorithm is extremely sensitive to the assumed height of the dust layer. In addition over land, since the algorithm uses window channels, the emissivity of the underlying land can impact the retrieval for cases of low optical depth.
One limitation of MODIS Deep Blue Collection 5 is that the surface reflectance database is static and this can be source of regionally/seasonally-dependent error; elevated terrain can also lead to biases as pressure is not accounted for explicitly.
AATSR ORAC and SWA have a significant better coverage, compare with the same instrument dataset from AATSR-GLOB, due to the better representation of the surface reflectance.
Limitation of the MISR dataset is the smoothing mask used in the current version of land retrievals that eliminates high AODs leading to AOD underestimation at high aerosol loading (a problem for heavy dust events).
POLDER retrievals are limited over land by the week sensitivity to the coarse-mode leading to the impossibility to estimate the total AOD, but nevertheless present a good CC with AERONET data.
OMI-KNMI is a more complex algorithm then OMI-NASA that make use of a wilder spectral range and fits several aerosol parameters in the retrieval. It has a better coverage of the dust plume compared to OMI-NASA and in comparison with the other dataset OMI-KNMI tends to give higher AOD in the southern part of the region considered in this comparison.
MERIS, SEAWIFS, SEVIRI-IMP-VIS, all use visible channels to retrieve the aerosol loading making it difficult to overcome the problem of dust retrieval over bright surface, so these dataset are applied only over ocean. Nerveless MERIS and SEAWIFS tend to miss the dust plumes due to presumably too restrict quality control while SEVRIR-IMP-VIS is

able to follow them.
SEVIRI-GLOB is using VIS-NIR channels and is applied both over land and ocean but does not cover any bright surface and also tend to miss the thicker part of the dust plume due to quality control.
SEVIRI-IMP-IR is applied over land and use only the infrared channels, it works best if the dust loading is relatively large and is less certain when there is little dust in the atmosphere because is more dependent on meteorological data.
SEVIRI-ORAC is a first attempt to overpass the problem of bright surface using together VIS and IR channels, but due to simple treatment of surface emissivity, the main issue is an overestimation of AOD over desert in clean conditions, which is attributed to errors in the modelling of surface properties."

At the end we added:

“All the datasets used in this comparison are available at:
ftp://ftp.atm.ox.ac.uk/pub/user/elisa/DRI/”