Review of the revised manuscript acp-2015-558 by Gkikas et al.

[To facilitate the reading, the following color code is used when necessary: blue is the text in the author's reply to me, red are sentences inserted in the revised manuscript.]

I do appreciate the efforts the authors made to address the weaknesses raised in my first revision. I'm aware they did additional work trying to address my points. However, and unfortunately, I still believe the manuscript is not ready for publication in ACP in the revised form presented. The reasons are listed below.

General

Many answers to my questions are given as 'reply to the reviewer', but not (or only partly) inserted into the body of the manuscript itself. It was obviously the intent of my revision to provide hints to strengthen the manuscript, and make it more legible. I did not need 'personal' answers, my comments were rather oriented to make the authors provide any potential reader with all the needed/useful information and to show/demonstrate the robustness of the methodology proposed to achieve the results. Thus, some of my comments were not fully understood/addressed in the text and this intent of mine was therefore not achieved within the revised version submitted.

My personal impression in reading the revised version is that all the efforts done by the authors in revising the text translated at last into some additional material inserted within the original manuscript in a rather 'patchy' way, with no substantial improvement of its structure and no real embedding of the given suggestions in its core concept.

The overall result is a long text, focusing on (too) many different aspects. The authors should make an effort to decide which is the core objective of the manuscript and limit the length of the other sections. Some suggestions on which parts could be removed (and possibly addressed in a companion study) were already given in my first review, and are further indicated below.

Specific (only the most important objections are given here)

1. METHODOLOGICAL PROBLEMS

My review point 1.1:

The authors reply that: '*in our opinion, an unequal distribution of available AOD data to the various months and seasons is not a problem in our methodology. We think that the AOD thresholds should be simply determined/computed from the entire dataset, since any climatic, health or other effect of aerosols is not dependent on season or month.*'

I fully disagree with this statement since:

a) the general sentence '..*any climatic, health or other effect of aerosols is not dependent on season or month'* is a at least questionable…(not to say incorrect).

b) an unequal distribution of AOD data within the whole dataset considered DOES HAVE A CRITICAL IMPACT on the pixel-resolved 'AOD Mean values' being derived and THEREFORE on the final outcome of their algorithm. Proving there is a rather uniform distribution of the number of single, daily-resolved AOD data points in the dataset is essential to demonstrate your method builds on robust basis. Should this not be the case, i.e., should the number of AOD data-points be unbalanced (in the different months or in the different years within the long-term period covered), other strategies should be adopted to compute reasonable 'AOD Mean values' (i.e., the threshold values used to drive the empirical algorithm).

This said, the authors add that: '*in order to dispel concerns of the Referee about this issue, we have investigated the AOD data's availability, in terms of percentages, at different temporal scales, namely monthly, seasonal, and yearly, and we have reproduced the corresponding geographical distributions.*'

However, this point is only partially clarified/solved by their additional R1-to-R4 Figures.

In fact, in Figure R1-to-R4 the authors show the data in terms of 'percentages', without reporting the formula these percentages represent. Anyway, if I understand correctly, this 'percentage' means that for example in the month of January (Figure R1) a 100% value (dark red) is obtained in pixels when they get 31/31 daily data points for each year of the period considered (i.e., they get (31 days x 13 years)/(31 days x 13 years) values, covering all days of January 2001-2002-2003-2004-2005-2006-2007-2008-2009-2010-2011-2012-2013).

Is this interpretation correct?

If no, the authors should please define what's the meaning of 'percentage' here.
If yes, then by similarity your Figure R2 shows that in winter they get more than 85% of the data over sea and less than about 50% over land (excluding bright areas having no data), while the opposite occurs in summer. This clearly reveals the seasonal unbalance I suspected, plus a land/sea unbalance in data availability I did not expect. This does critically affect their results.

In the revised manuscript the following paragraph is added to comment on this point:
'It should be noted that the representativeness of the calculated mean levels is possibly affected by the availability of the AOD retrievals and particularly by the way these data are distributed at different temporal scales. To this aim, we have calculated the percentage availability of AOD retrievals on a monthly, seasonal and year by year basis, over the period 2000-2013 (results not shown here). Seasonal differences of AOD availability are mainly encountered in the northern parts of the study region, with lower values (20 to 40 %) from December to February against 50-85% for the rest of the year. This is attributed to the enhanced cloud coverage prohibiting the satellite observations. Nevertheless, this does not essentially affect the algorithm outputs since these regions, being far away from the dust sources, are not so frequently affected by dust outbreaks, especially given the significant wet removal of aerosols during this most rainy season of the year.…'

I think the results of Figure R1-R2 would merit more attention.

Why should cloud cover in winter affect land and not sea (Mediterranean) areas?

Are the two different MODIS retrievals over land and ocean playing a role in this land/ocean difference in the number of data availability?

It is true that the DD impact in winter at higher latitudes is expected to be low, but what about the different data availability over the Med sea among seasons? This clearly affects the results over the Med areas closest to the North African coasts where the maximum impact of dust is rather expected. The authors should comment on that.

In brief, although I'm aware the same methodology has been used in several other publications, to my opinion the different data coverage in time and space is a major critical element not properly addressed by the algorithm currently adopted.

[Further comment: good that this unbalance of Figure R1 and R2 somehow 'compensates' in the annual Figure R3 and that this compensation keeps similar for the different years considered (Figure R4). I believe Figure R5 I suggested to produce is indeed interesting and useful. I think the authors should insert it at least in the supplementary material].

My review point 1.2:

It is a good idea to provide a sort of 'sensitivity' of your results to the chosen algorithm by using an alternative METHOD-B, particularly because their 'objective and dynamic algorithm' is used in several other publications. Unfortunately, I could not fully understand how the Method-B is implemented, and a scheme similar to that of Figure 1 would have been helpful. In their reply they state that: '*In order to answer to the Reviewer's comment we have applied our algorithm according to his/her suggestion followed a new methodology, so-called METHOD-B. According to this, from the raw AOD retrievals we have excluded the 'pure' DD cases, defined based on the defined thresholds for Ångström exponent, Fine Fraction, Aerosol Index and Effective radius (available only over sea). From the remaining non-DD AOD retrievals we have calculated the mean and the associated standard deviation values for the whole study period as well as the defined thresholds, in each grid cell.'*

Does this mean in METHOD-B the identification of 'pure' DD cases uses the thresholds for Ångström exponent, Fine Fraction, Aerosol Index and Effective radius BUT NO threshold for AOD? The authors should please clarify this point, as the same ambiguity is found in the text added to the revised manuscript (lines 371-383).

2. DATASET PROBLEMS

My review point 2.1:

I understand the authors reasons to keep a different time range for AQUA and TERRA, exploiting the longer TERRA dataset. Still, I think showing the results I asked for (Figures R7 and R8) in the supplementary is important because then (and only then) the authors can substantiate the statement that differences between TERRA and AQUA are likely due to daily cycle effects rather than to the different long-term period addressed.

My review point 2.2:

I know that '*the satellite data which we have been used in our analysis have been thoroughly described and presented in numerous studies related to aerosol research in the past*', and I thank the authors for the details on the MODIS L2 and L3 data they are providing me in their reply. I believe IT WAS IMPORTANT TO PROVIDE THE SAME DETAILS TO THE POTENTIAL READERS OF THE POTENTIAL PAPER, as now done in the text.

To my question : "*In Section 2.1.1 you only give the expected accuracy of the AOD data used. Which is the accuracy of the other MODIS-derived parameters employed in the study? How does this accuracy change above land and ocean? Is it sufficient to make these products suitable to be employed for scientific purposes?*"

the authors reply that: 'To our knowledge, we don't have enough information about the FF, Ångström exponent (or alpha), and effective radius retrievals' accuracies since those quantities have not been evaluated to the same extent that AOD has. FF and alpha data are derived from spectral information and their accuracy is determined by very sensitive spectral dependent factors such as errors in the surface model or sensor calibration changes. Over land, these factors play an important role and for this reason the accuracy of the aforementioned observations is lower compared to the corresponding ones over maritime regions. Over sea, the size parameters (FF and alpha) are strongly dependent on wind conditions. According to our analysis, it seems that for strong AOD signals, such as the case of intense desert dust outbreaks, the results reveal a satisfactory agreement between satellite and ground measurements.'

A similar text is inserted in the revised manuscript (lines 211-214). Given the 'uncertainties' on the accuracy of these parameters, some 'sensitivity' on the use of the relevant thresholds in the algorithm would have been opportune, as well as on the real need for such a multi-parameter dataset.

To my question : "*The algorithm uses the information on Angstrom Exponent (AE), AI and FF (plus reff over sea) to select 'strong' dust and 'extreme' dust events. However, there is very little information in the text on HOW the matching between AOD, AE, FF (plus reff) and AI is operatively done at the pixel level. In particular the manuscript lacks in describing the statistics of the coincident multi-parameter dataset. I guess you do not always have ALL the parameters available at the same time. What happens in case you do not have coincident datasets? How frequent these cases are? What's the impact of this on the final outcome of your study?*" the authors reply that: All the aerosol optical properties retrievals have common spatial and temporal resolution. Therefore, it is straightforward how the collocation, in spatial and temporal terms, is done.

Following the Reviewer's comment, we have added in our manuscript that all the defined criteria must be fulfilled concurrently in order to be clear to the reader (lines: 364-366).

It was rather clear that all the aerosol optical properties retrievals have common spatial and temporal resolution and indeed it is rather straightforward how the collocation, in spatial and temporal terms, is done. But this was not my question. My question was about *the statistics of the coincident multi-parameter dataset.* And specifically: *'what happens in case you do not have coincident datasets (i.e. in case of any missing value)? How frequent these cases are? What's the impact of this on the final outcome of your study?'*

The questions were pertinent, because, as the authors specify later in their reply: '.. 23% and 12.8% of the strong and extreme episodes, respectively, is unclassified (UN) mainly due to missing AI data. Our algorithm has been constructed in such way that makes possible the identification and classification of the aerosol episodes when all the satellite retrievals are available (coincident), trying to avoid any "guess" (UN episodes) which can be ambiguous'.

This was the information I was asking for in my question. This information needs to be added to the text, together with the explanation on how these 23% and 12.8% percentages are obtained.

Additionally, in their reply to this comment of mine, the authors add: 'As to what happens with the algorithm when there are not coincident data, that is when the criterion for a dust episode is not fulfilled, we would like to note that the applied algorithm in the present analysis is a branch of a unified algorithm which identifies and characterizes not only DD episodes, but also four other types of aerosol episodes, namely biomass-urban (BU), dust/sea-salt (DSS), mixed (MX) and undetermined (UN). The relevant results, for the period 2000-2007, over the Mediterranean Sea, are discussed thoroughly in Gkikas et al. 2016 (http://www.sciencedirect.com/science/article/pii/S135223101530563X).'

This information is very interesting, and it is indeed a bit strange no mention to that Gkikas et al. 2016 paper (or manuscript in press) was done in the first version of the manuscript nor in the current revised one (the paper is now published). It should obviously be included in the reference list and mentioned in the introduction to put the present study into a more general prospective.

3. PRESENTATION OF RESULTS

My review point 3.1:

To my comment 3.1 the authors reply: *'We strongly believe that it is better to keep the existing paper's structure since it helps the reader to follow our approach without going back and forth on the text'.*

However my comment was about re-arranging Section 4, and renaming some sub-sections, not to change the structure of the paper.

At this stage, my point 3.1 still holds.

Again, my intent was (is) to make the manuscript clearer. So I would not suggest anything forcing the reader (and myself) going back and forth on the revised text.

My review point 3.2:

As mentioned in my general comment at the beginning, this manuscript addresses a lot of different aspects of DD events (DD algorithm issues, horizontal scale, vertical scale, comparison to AERONET, comparison to PM10) and seems to lack a main focus. This also translates into a very long and inhomogeneous text that does not help the reading. In my former comment 3.2, I suggested to leave the comparison with PM10 data to a further (interesting) investigation. I'm still convinced of this, and even more now that the revised version further added a 3 pages-section (4.4) on this aspect, with the study of specific 'desert-dust cases'.

Some minor comments

- Some of the information provided in your answers to my minor comments needs to be inserted in the text to improve its clarity and completeness.

- Declared aim of the study in the abstract is now '*to describe the vertical structure of the intense Mediterranean dust outbreaks*' but this is just a part of the work, the horizontal structure over the basin, although less original, is also important within the manuscript and should be also emphasized in the abstract (see also my general comment on the need to better clarify which is the focus of the work and to re-arrange the whole text accordingly).

- The authors use the two metrics DD 'episode frequency' and 'intensity' to characterize DD over the Mediterranean. From their description (Section 3), it seems that an 'episode' (per pixel) is a single day of dust, while quite often in the literature the term 'episode' is used for a dust event lasting more than 1 day. This is quite important to clarify, also to allow comparison with other literature data. For example Pey et al., 2013 show a dust frequency (% over annual days) over Southern Europe up to about 37% in Sicily (Italy), and about 30% in Southern Spain and Greece. Note that these percentages in Pey et al. (2013) are obtained ONLY limiting to the cases in which Saharan dust is observed at the ground, thus should correspond to a subset of the DD statistics from columnar measurements as in the present work. Here Gkikas et al get maximum frequencies for 'strong' events of about 10 episodes/year (i.e. less than 3% over annual days) close to the African coasts. How do the authors reconcile these numbers? Could they please comment on that? Note that this aspect is somehow connected to the definition of the 'AOD thresholds', from which all the subsequent analysis and definition of 'strong' and 'extreme' events derives.

- Figure 1: The authors can keep it if adding a second panel showing the equivalent scheme for the METHOD-B, this making the figure different from the one in their former publication.

- Figures 9 and 10 are not very much readable and should be improved.

- I did not have the opportunity to fully read the Gkikas et al (2016) paper on Atmospheric Environment, but why are the Figures 2 and 3 in this manuscript different from Figures 4 and 5 in that paper (for those plots referring to the same DD analysis on the same MODIS Terra dataset)? It would also be important to comment on these differences, given that this manuscript follows that other one from the authors.

- I noticed the following reference is reported twice in the reference list, with two different years of publication:

- Gkikas, A., Houssos, E. E., Lolis, C. J., Bartzokas, A., Mihalopoulos, N. and Hatzianastassiou, N.: Atmospheric circulation evolution related to desert-dust episodes over the Mediterranean. Q.J.R. Meteorol. Soc., 141: 1634–1645. doi: 10.1002/qj.2466, 2015.
- Gkikas, A., Houssos, E. E., Lolis, C. J., Bartzokas, A., Mihalopoulos, N. and Hatzianastassiou, N.: Atmospheric circulation evolution related to desert-dust episodes over the Mediterranean. Q.J.R. Meteorol. Soc., doi: 10.1002/qj.2466, 2014.