We would like to thank the Reviewer for his/her detailed review which helped us to essentially improve our manuscript. Point by point responses to the Reviewer's comments (in Italics) are provided below.

"*This manuscript by Gkikas et al. aims at describing the horizontal and vertical distribution of desert dust during 'intense events'. The methods followed to reach this aim include the use of satellite data from passive and active sensors. The work is largely an extension of a previous paper by Gkikas et al. (2013), and this makes it a bit lacking in originality. The manuscript could provide potentially interesting information. However, in its present form it shows some major weaknesses that substantially compromise the validity of the results reached. These weaknesses, as well as some additional comments, are detailed below. For what follows, I cannot recommend the paper for publication until these major weaknesses are properly addressed. I also recommend review of the language as the text is often confusing and difficult to follow.*"

We acknowledge that the present manuscript it is an extension and thus has some similarities with our previous work by Gkikas et al. (2013). Therefore, we can understand the worries of the Referee about its originality. Being aware of this, we had tried to highlight in the original manuscript the differences and steps forward made in the present with respect to our previous work. Nevertheless, obviously this was not achieved entirely, and hence, before presenting our point-by-point responses, we would like to emphasize the differences between the present analysis and Gkikas et al. (2013).
The main objective of the work by Gkikas et al. (2013) was the detailed description of the spatial and temporal variability of Mediterranean desert dust (DD) episodes. The primary focus in the present analysis is to study the vertical structure of the DD episodes, as stated in the title, at an extended spatial and temporal coverage, as not done before. However, we would like to note that in order to attempt this, one should ensure the quality of the utilized tool and derived DD episodes. It should be reminded that the episodes are identified with a satellite-based algorithm, thus having limitations and being constrained by the input satellite retrieved products. Such limitations were revealed in Gkikas et al. (2013) where a rather preliminary evaluation of the algorithm was made. For this reason, in the present study we emphasize the evaluation of the satellite algorithm through extensive and detailed comparisons against surface-based measurements from AERONET and PM_{10} stations. This is why a considerable part of the manuscript is devoted to this aim. On the other hand, the very short extent of section 4.1, dealing with the spatio-temporal regime of DD episodes, i.e. the main goal in Gkikas et al. (2013), shows the low priority given to this in the present paper.
Of course, one may consider that even this relatively short reference to the episodes' regime could be avoided here, as it was given in Gkikas et al. (2013). Nevertheless, we have chosen to include it for two reasons: (i) the temporal coverage of the episodes regime in Gkikas et al. (2013) was relatively short, namely from 2000 to 2007. Note that it was the first time that the Mediterranean DD episodes regime has been described at a complete spatial coverage in Gkikas et al. (2013) whereas no other study to date has done something similar. It is not given in any way that this regime remains unchanged with time, therefore it is valuable to re-assess it over longer time periods, and this is made in the present study where the period is almost double (2000-2013). (ii) obtaining the regime over the longer period, 2000-2013, is necessary in this study also in order to overlap with the period to which the examined vertical structure of episodes refers to. Note that the anyhow scarce available vertically resolved information is taken from CALIOP-CALIPSO, which is not available for years before 2006.
Finally, we would like to note that apart from what it has been reported above, we acknowledge that the part of the analysis dealing with the vertical structure of the DD episodes could, indeed, be a bit strengthened. In this view, an effort was made to further extend the relevant analysis in the revised manuscript, which is now enriched by the detailed study of specific dust outbreaks in the Mediterranean, also giving insight to the level of agreement between the algorithm outputs and surface-based measurements.

Major weaknesses of the work:

1) Methodological Problems.

"*Although the study follows a methodology almost identical to the one already published in Gkikas et al. (2013), in my view this methodology builds on assumptions that should AT LEAST be further commented and 'tested' to make the reader understand how reliable the derived results are. In particular a key point of the work is the identification of 'intense dust events' ant their separation into 'strong dust episodes' and 'extreme dust episodes' by the so-defined 'objective and dynamic satellite algorithm'. This selection is basically fully dependent on the AOD threshold chosen, defined as 'AOD Mean' in the text. Unfortunately I could not find any definition of this 'AOD Mean' in the text other that 'the mean (Mean) and the associated standard deviation (SD) are calculated for the whole study period' (page 27688, lines 18-20).*"

1.1) What does this 'Mean' mean?

"You use over 10 years of AOD data, so: are you getting a 1x1-resolved 'AOD Mean' by simply averaging the 1x1-resolved AOD time series corresponding to each pixel? (This is what I understood reading Gikaks et al. (2013), but this is not clarified in this manuscript)."

Yes, the mean AOD value is calculated for each grid cell over the whole study period, namely by averaging all the available daily AOD retrievals. In the revised manuscript we have modified the relevant text (lines 333-335) while in the caption of Figure 1 we have clarified that the methodology is applied to 1° x 1° grid each grid cell.

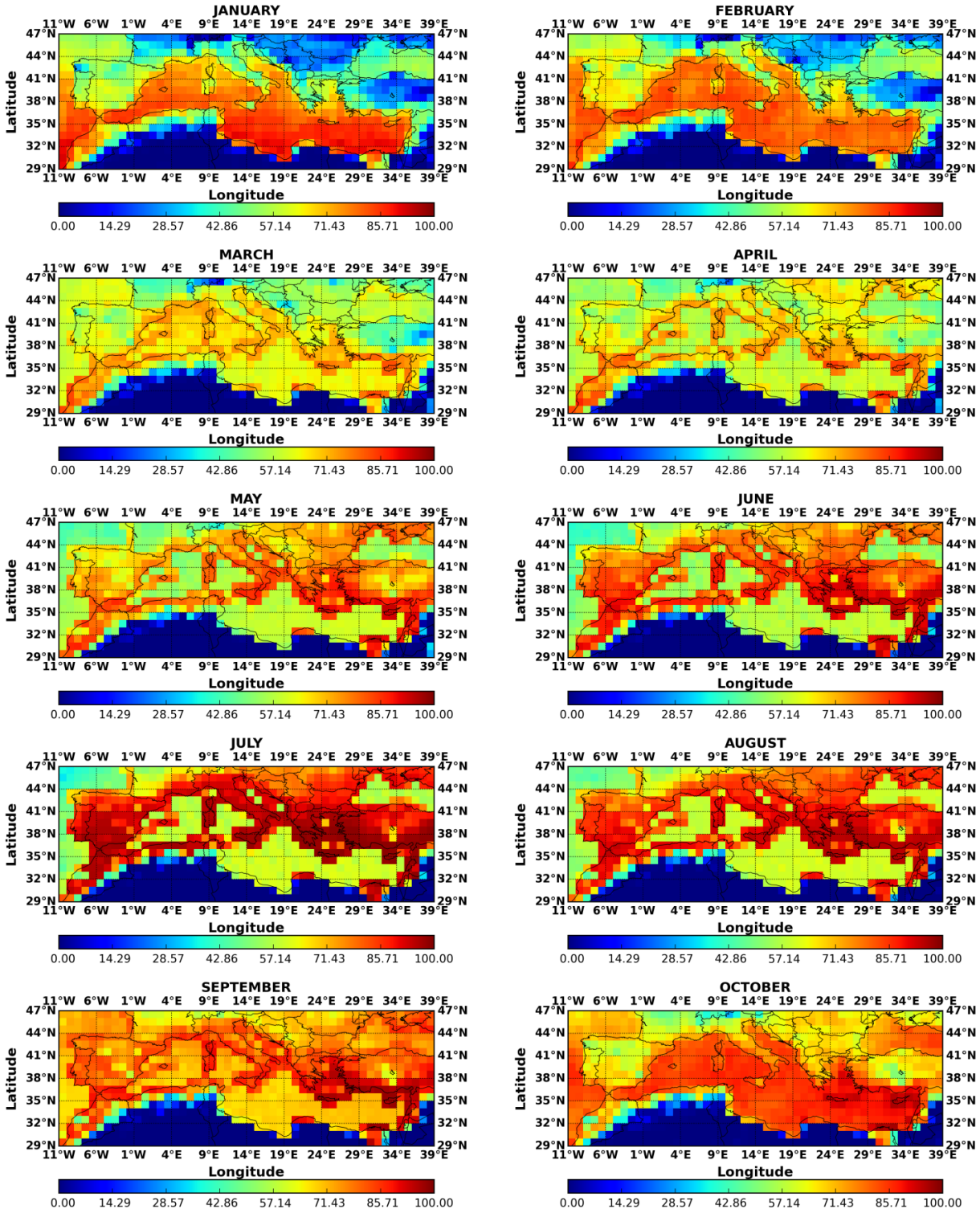
"What's the total number of AOD data points you have in each pixel? Are all years within the record equally represented? Are all seasons (or even months) in each year equally represented? You should show this is the case, otherwise the 'Mean' AOD value you get might be meaningless. In fact, given the inter-annual variability of AOD, and, above all, given its marked seasonal variability, an unequal data coverage of the different years/seasons could lead to a 'biased' Mean AOD. More explicitly: if, for example, the number of data points in winter is 50-60% of those in summer (this is likely due to enhanced cloud cover in winter) you derive a summer-biased 'Mean' AOD which alters all the subsequent analysis. Therefore, please provide a clear indication of the number of data points you are including in your statistics (for example with maps in an Appendix), and of their inter- and intra-annual distribution. If the monthly-resolved data coverage is not uniform (as I suspect), and if you still prefer to perform the whole analysis using a single, pixel-resolved 'AOD Mean' threshold value, then you should rather obtain it as an average of monthly-mean AODs. A map of the derived AOD Mean values would also be of help to better interpret the final results.

Although we see the point raised by the Referee, we would like to note 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. Nevertheless, 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. For brevity reasons, the analysis is made only for the period March 2000 – February 2013 using the MODIS-Terra AOD data. The results are presented and discussed below.

MONTHS

According to our results (Figure R1) the temporal dependence of AOD availability, i.e. its month by month variation, is dependent on the area of interest, if we except the desert areas of North Africa and Middle East, where retrievals of AOD are restricted by highly reflecting surfaces (note that Dark Target MODIS products are used in the present analysis). Apart from these areas, it is apparent a reduced data availability in winter months, especially in the northern continental parts of the study region, e.g. Balkans. The reduced availability

of AOD in winter, compared to other months, can really affect the representativeness (in temporal terms) of the calculated AOD mean value for the whole study period, as stated by the Reviewer. Even if, as stated before, this should not be a problem, we further like to note that its consequences for our analysis are minimized by the fact that due to the prevailing atmospheric circulation and the higher precipitation amounts (wet deposition) in this season, the probability for a desert dust outbreak to reach the areas with low AOD availability is rather small. In the revised manuscript (lines 350 - 360) we addressed the specific issue of month to month variation of AOD data availability raised by the Referee by adding a short paragraph.



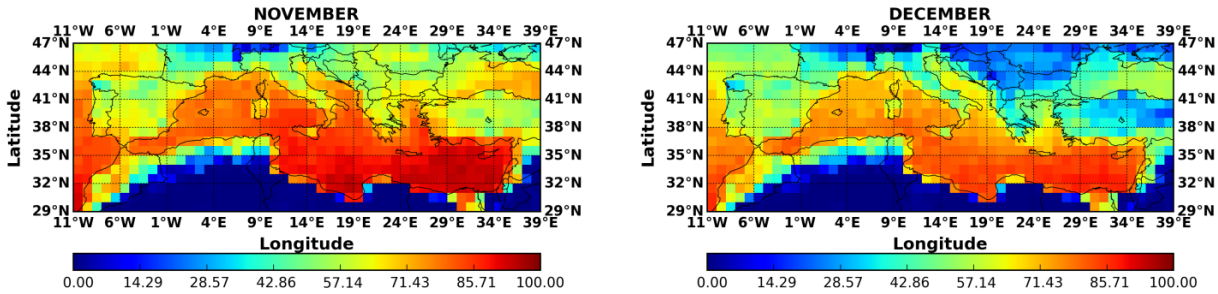


Figure R1: Monthly geographical distributions of the MODIS-Terra AOD_{550nm} retrievals availability (expressed in percentages), over the broader Mediterranean basin, for the period Mar. 2000 - Feb. 2013.

SEASONS

The geographical distributions of AOD data availability have been also reproduced on a seasonal basis, i.e. for winter, spring, summer and autumn, over the period 2000-2013. The results of Fig. R2 summarize those in Fig. R1 (monthly basis), namely that if we except the North Africa and Middle East deserts, where a low AOD availability is observed throughout the year, unequal seasonal low AOD availability is observed in the northernmost parts of the study region (e.g. Balkans), with numbers ranging from < 30% in winter to more than 80% in summer. For a discussion of possible consequences of this the Reviewer is referred to the discussion of monthly statistics above.

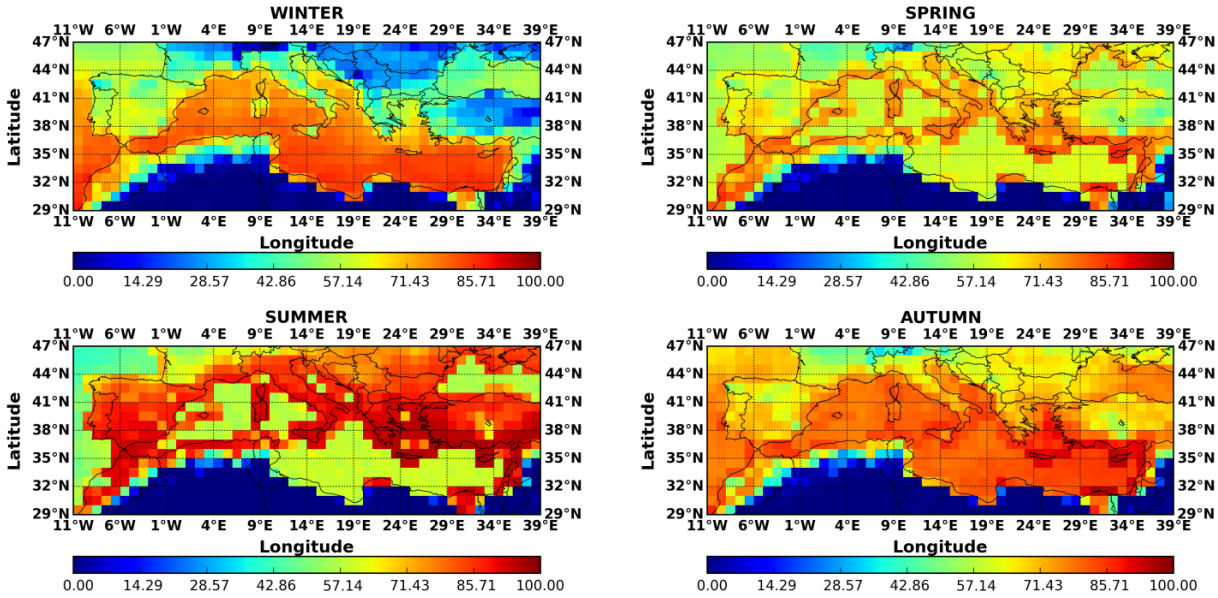


Figure R2: Seasonal geographical distributions of the MODIS-Terra AOD_{550nm} retrievals availability (expressed in percentages), over the broader Mediterranean basin, for the period Mar. 2000 - Feb. 2013.

ANNUAL

For the whole study period (2000-2013), the lowest percentages (about 40%) are observed across the northern parts of the study region. On the contrary, the highest percentages (75 - 95%) are recorded mainly across the coastlines and closed seas (e.g. Aegean and Adriatica). In all other areas, the AOD availability is satisfactory ranging between 65 and 80%.

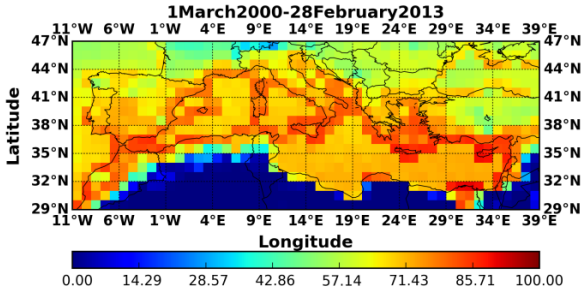
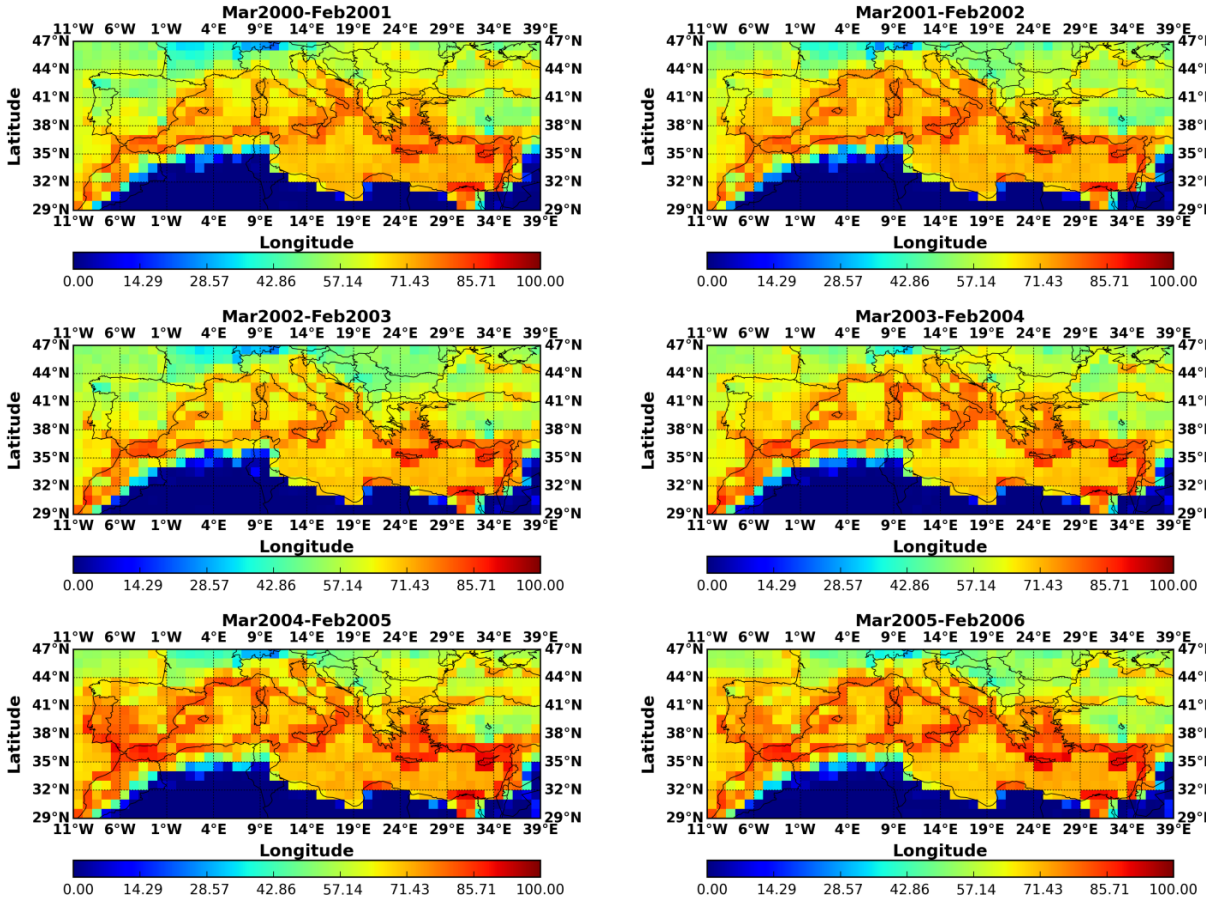


Figure R3: Geographical distribution of the MODIS-Terra AOD_{550nm} retrievals availability (expressed in percentages), over the broader Mediterranean basin, for the period Mar. 2000 - Feb. 2013.

YEAR BY YEAR

Finally, we have reproduced the corresponding geographical distributions of MODIS-Terra AOD retrievals' availability for each year (from March to February of next year) of the study period (2000-2013). It is evident, that the spatial patterns remain constant throughout the years revealing a stable regime in terms of aerosol optical depth data availability.



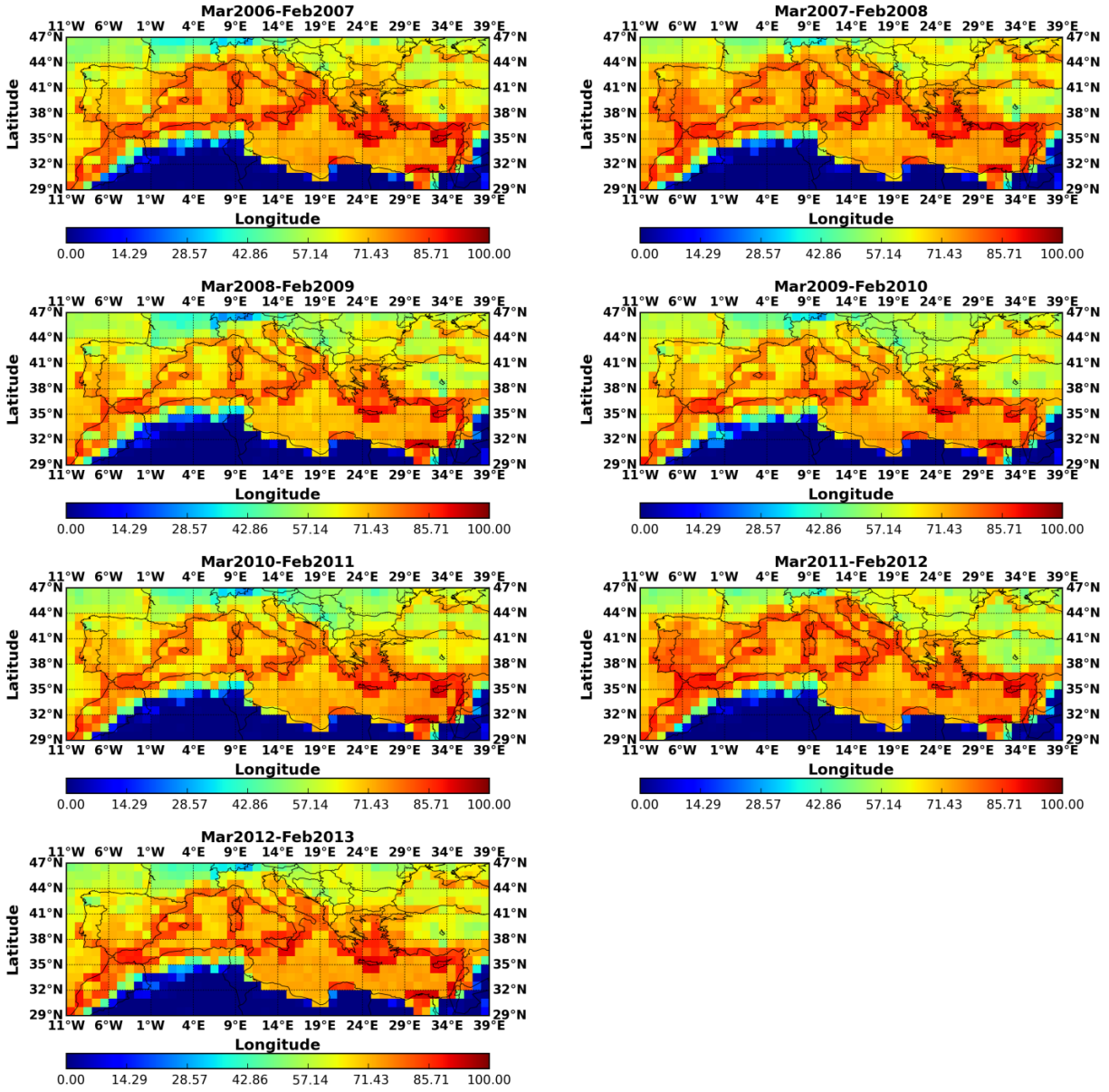
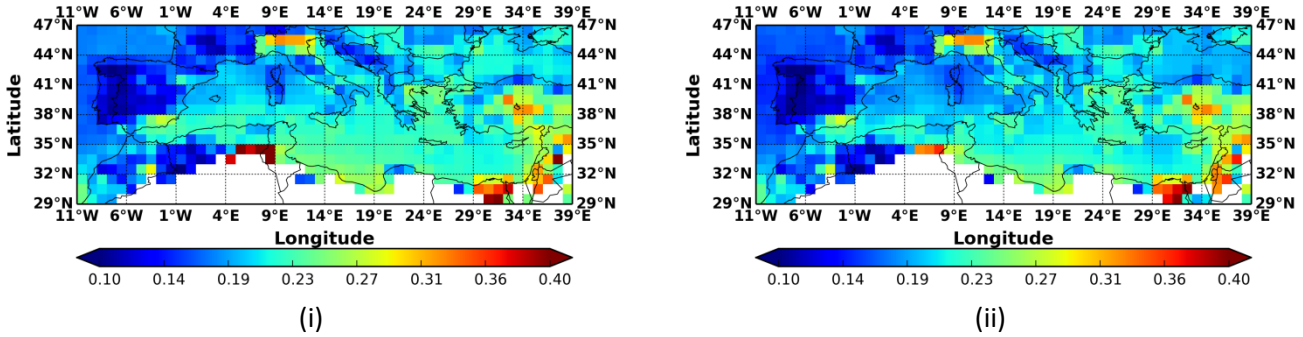


Figure R4: Annual geographical distributions of the MODIS-Terra AOD_{550nm} retrievals availability (expressed in percentages), over the broader Mediterranean basin, for each year of the period Mar. 2000 - Feb. 2013.

As suggested by the Reviewer, we show in Figure R5 the geographical distribution of the long-term averaged MODIS-TERRA (i) and MODIS-AQUA (ii) AOD at 550 nm, over the periods March 2000-February 2013 and 2003-2012, respectively.



(i) (ii)

Figure R5: Geographical distributions of the long-term averaged AOD at 550, over the broader Mediterranean basin, for the periods: (i) 1 Mar. 2000 - 28 Feb. 2013 (MODIS-Terra) and (ii) 2003 – 2012 (MODIS-Aqua).

1.2) *"I also see another problem in the computation of the 'Mean AOD'. Your algorithm firstly performs selection of the 'strong dust episodes' and of 'extreme dust episodes' based on the 'Mean AOD' threshold and AFTERWARDS uses the AI and FF information to exclude possible 'not-dust' cases. However, this means that if, by paradox, all the days of the time series are 'dust-affected', and assuming for simplicity a normal distribution of those AOD values around their mean, following your approach only a very limited fraction of them would be classified as 'strong' dust (the 2.2%) or 'extreme dust' (the 0.003%) (this is because in a normal distribution the 2.2% and the 0.003% of data exceed respectively the 'Mean + 2 s. d.' and 'Mean + 4 s. d.' thresholds you fixed). This is to say that, in my view, dust events should be firstly filtered out of the record to compute a sort of 'dust-free' mean AOD; otherwise, again, you get a biased 'Mean AOD' to which to compare the 'dust-affected' record. The way to do this could be based on a combined analysis including the other parameters you consider in your analysis (Angstrom, AI and FF). Obviously, this problem mostly affects those regions with higher dust-events frequency. Can you comment on 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.

The obtained results for the frequency of occurrence and the intensity of strong and extreme DD episodes, over the period 2000-2013, based on METHOD-B, are depicted in Figure R6. The spatial patterns, both for strong and extreme DD episodes, are quite similar with those displayed in Figures 2 i-a and 2 ii-a of the manuscript. The main difference, between the two methods, is that the frequencies are higher, both for strong (up to 13.3 episodes year^{-1}) and extreme (up to 8.1 episodes year^{-1}) DD episodes, when the algorithm operates based on METHOD-B. This is expected since omitting dust-affected cases in the computation of long-term mean AOD values results in lower AOD mean and thresholds. As it concerns the intensity, the AOD levels for the strong and extreme DD episodes are lower than the corresponding levels computed in our primary methodology (Figures 3 i-a and 3 ii-a). This is again expected, because the determined DD episodes with METHOD-B include cases with smaller AODs, which can result in loss of a "clear signal" of DD episodes determined with the new method. Indeed, the consequence of this can be seen in Fig. R6, 2ii-a, for the strong DD episodes, where there are not so distinct geographical patterns as in Fig. 3i-a (original manuscript, basic methodology). On the contrary, for the extreme DD episodes (stronger signal), the differences between the two applied methodologies are less remarkable in terms of spatial variability. Summarizing, when the algorithm operates based on METHOD-B the frequency of occurrence and the intensity of DD episodes increases and

decreases, respectively, without revealing remarkable differences, in spatial terms, with regards to the default methodology. Both facts can be explained by the reduction of the defined thresholds. Since there is not a single commonly accepted methodology in literature for the determination of DD episodes, and given that when applying the alternative suggested methodology the results do not change essentially, we prefer to keep the original methodology. However, the obtained results with METHOD-B are now provided in the supplementary material (Figs. S1 and S2) whereas the differences with the default methodology are discussed in the manuscript (lines 472-488).

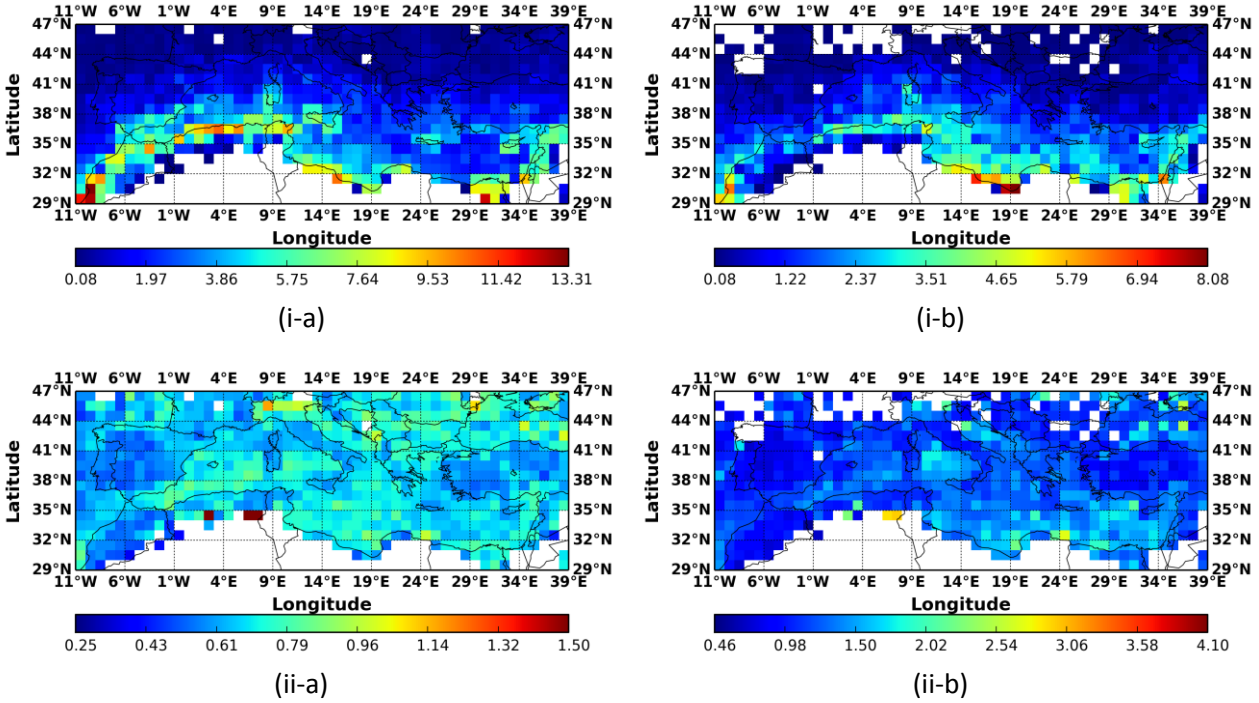


Figure R6: Geographical distributions of the: (i) frequency of occurrence (episodes/year) and (ii) intensity (in terms of AOD_{550nm}) for the: (a) strong and (b) extreme desert dust episodes, for the period Mar. 2000 – Feb. 2013 (MODIS-Terra), over the broader area of the Mediterranean basin. Both episodes' characteristics have been calculated based on METHOD-B.

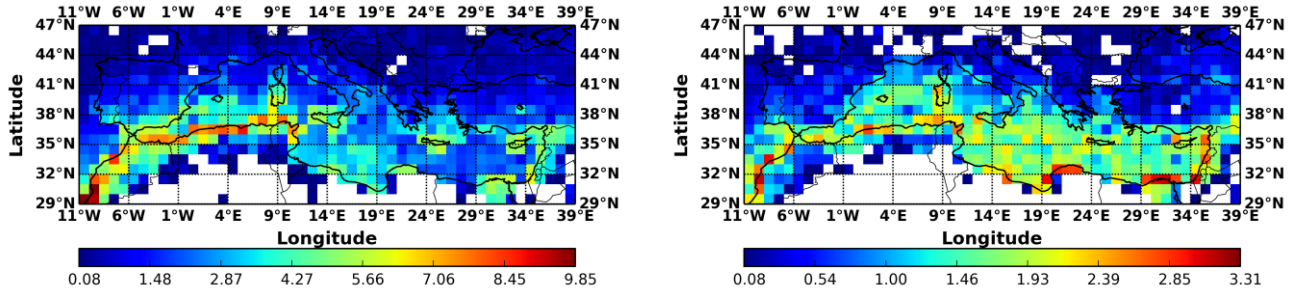
2) Dataset problems

"2.1) You make several efforts in the text to highlight the differences of this work with respect to the previous one (Gkikas et al., 2013) which, as mentioned above, is very similar to this in terms of methodology and structure. In fact, one of the differences with that paper is the extension in time and the use of the additional similar datasets from Modis-AQUA. So, one of the potentially interesting points of the study relates to the differences found between the TERRA and the AQUA-based results. Unfortunately, at the present stage, the validity of this comparison is completely jeopardized by the different time-periods used in the manuscript for the two Modis sensors. In those cases in which differences between the two outcomes are found and commented in the text, there is always the ambiguity whether those differences are due to a 'real' aerosol (mostly desert-dust in this case) diurnal variability, or rather to the different time-period covered by the two datasets (note that this point is also connected to my point 1). It is well know that AQUA has a shorter time-coverage with respect to TERRA, but I do not see the reason not to limit the analysis to the period 2003-2012, which is common to both sensors."

In order to answer to the Reviewer's comment, we are presenting here the geographical distributions for the DD episodes' frequency of occurrence (episodes/year) and intensity (in terms of AOD_{550nm}), in Figures R7 and

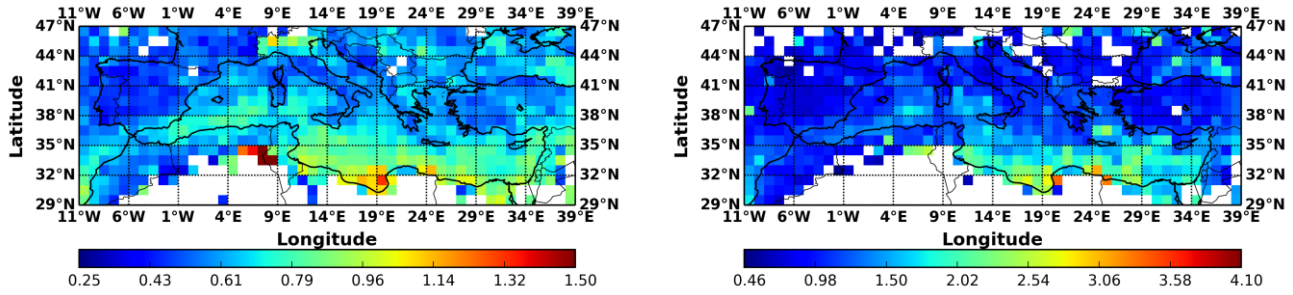
R8, respectively, over the period 2003-2012 based on MODIS-Terra retrievals. It is evident a great similarity, both for frequency and intensity, between the satellite algorithm's outputs when it operates with MODIS-Terra retrievals over the periods 2000-2013 (Figs 2 i-a, 2 ii-a of the revised paper) and 2003-2012 (Figs R7 and R8). This means that the "direct" comparison between MODIS-Terra and MODIS-Aqua, as it is described in the submitted manuscript, can be done even though the study periods are different. We agree with the Reviewer that a more detailed analysis it is required in order to highlight the differences (if any) between the two MODIS sensors, however this is not our primary scientific target in this paper. We have decided to present the results for the period 2000-2013 (MODIS-Terra) and not for 2003-2012 (MODIS-Aqua) for the following reasons:

1. MODIS-Terra provides data for 13 years instead of 10, which are available from MODIS-Aqua. The three extra years allow us to have more coincident measurements between the satellite algorithm's outputs and AERONET retrievals. Please, keep in mind that our algorithm requires satellite retrievals from different sensors (MODIS, OMI, TOMS) to be available concurrently and moreover in many AERONET stations there are gaps in data availability. Both facts can reduce the number of coincident observations and for this reason we prefer to keep the more extended study period (2000-2013) in order to have a larger sample of dust episodes.
2. The PM_{10} measurements are available from 2001 to 2011 (11 years). Through the implementation of MODIS-Terra retrievals (2000-2013), we can have more data (two years, 2001-2002) compared to MODIS-Aqua measurements, which are available since 2003. This means that the common period between MODIS-Terra/PM is 11 years (2001-2011) while between MODIS-Aqua/PM is 9 years (2003-2011).
3. As it concerns the part of the analysis referring to the satellite algorithm's outputs and CALIOP lidar profiles, we have found more coincident data between MODIS-Terra/CALIOP in comparison to MODIS-Aqua/CALIOP. Even though the main findings are similar, in the case of MODIS-Aqua/CALIOP the 3D plots are not so distinct, mainly for the latitudinal projections, as in MODIS-Terra/CALIOP (Figs 9 and 10 in the revised manuscript).



(i) (ii)

Figure R7: Geographical distributions of the frequency of occurrence (episodes/year) of: (i) strong and (ii) extreme desert dust episodes, for the period 2003 – 2012 (MODIS-Terra), over the broader area of the Mediterranean basin.



(i) (ii)

Figure R8: Geographical distributions of the intensity (in terms of AOD_{550nm}) of: (i) strong and (ii) extreme desert dust episodes, for the period 2003 – 2012 (MODIS-Terra), over the broader area of the Mediterranean basin.

"*In this respect, in Section 4 (page 27691) you specify that: 'In order to investigate this difference in detail we have also applied the satellite algorithm, over the period 2003–2012, i.e. that of Aqua, using MODIS-Terra retrievals as inputs. Through this analysis (results not shown here), it is evident that there is a very good agreement between the satellite algorithm's outputs, for the periods March 2000–February 2013 and 2003–2012, revealing a constant dust episodes' regime. Therefore, the discrepancy appeared between MODIS-Terra and MODIS-Aqua spatial distributions in Fig. 2, is attributed to the diurnal variation of factors regulating the emission and transport of dust particles from the sources areas.' Apart from the fact that you decide not to show in the text an element that would be fundamental in your analysis, in my view this sentence is by no means sufficient to justify the use of a different time-period coverage of the AQUA and TERRA Modis data in your study. I think the same period for both sensors should be used to strengthen the results reached.*"

We think that our previous answer explains adequately why we have decided to present the results for the extended period (2000-2013). Moreover, the fundamental element and main target of our analysis is to describe the annual and seasonal variation of the Mediterranean dust outbreaks' three dimensional structure and not to discuss in detail the possible inconsistencies between Terra and Aqua retrievals.

"*2.2) In the present form, description of the datasets used and of the way the different variables are matched is lacking. Description of the single datasets does not allow to get all the necessary information to understand their advantages and limits. I found some more details in Gkikas et al. (2013), but it is a bit annoying to always go to that paper to better understand the current one. Please provide more details to your Section 2 (e.g., 1- resolution of the AI from TOMS and OMI is a bit different, can you specify this? How do you match these values with the MODIS derived ones? 2- What's the exact meaning of 'Quality assurance-weighted' data?).*"

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. For this reason, we believe that it is better to keep the existing information provided in our text without extending our manuscript with many technical details. Nevertheless, we have added the required information about the spatial resolution of the EP-TOMS and OMI satellite retrievals, which was missing in the submitted paper, and a better explanation of the term "Quality assurance-weighted" is given in the final version of the manuscript. More specifically, the MODIS and OMI satellite retrievals are provided at 1^{o}x1^{o} spatial resolution while the coordinates are provided at the center of each grid cell. EP-TOMS data are provided at 1^{o}x1.25^{o} (lat x lon) spatial resolution and they have been converted in order to match, in terms of spatial resolution and colocation, with the other two databases (MODIS, OMI). After regridding the EP-TOMS data, all the satellite retrievals have common spatial resolution and geolocation information (coordinates).

MODIS data are available at Levels, apart from Collections, corresponding to different spatial (from meters to degrees) and temporal (from minutes to months) scales. Here, we are using the daily Level 3 data provided at 1^{o}x1^{o} spatial resolution. The MODIS team produces the aforementioned measurements from the Level 2 data (swaths of 5-min intervals at 10km x 10km spatial resolution). Each Level 2 retrieval, is flagged with a bit value (from 0 to 3) corresponding to confidence levels (No confidence: 0, Marginal: 1, Good: 2 and Very Good: 3). Based on this, the Level 3 QA-weighted spatial means are obtained by the corresponding Level 2 retrievals considering as weight their confidence level (bit value). In order to avoid any misunderstanding from our side we are providing the relevant part from the MODIS documentation (http://modis-atmos.gsfc.nasa.gov/_docs/QA_Plan_2007_04_12.pdf, at the bottom of Page 4):

"The MODIS Atmosphere L3 processing software makes use of the L2 Usefulness and Confidence flags by creating L3 QA-weighted mean and standard deviation statistics. The QA weighting is performed by weighting each L2 input pixel by its Confidence flag, so that non-fill no confidence data has a weight of 0x, marginal data has a weight of 1x, good data has a weight of 2x, and very good data has a weight of 3x in the statistical computation within the L3 one-degree grid box."

The web link is already provided in our manuscript, for those who want to find more details about the methodology, which is applied by the MODIS Team for the calculation of Level 3 grid cells spatial averages from the corresponding Level 2 retrievals. In the updated version of our manuscript, we have added a sentence explaining briefly the term "Quality assurance-weighted" as it has been asked by the Reviewer (lines 214-220).

"*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 this products suitable to be employed for scientific purposes?*"

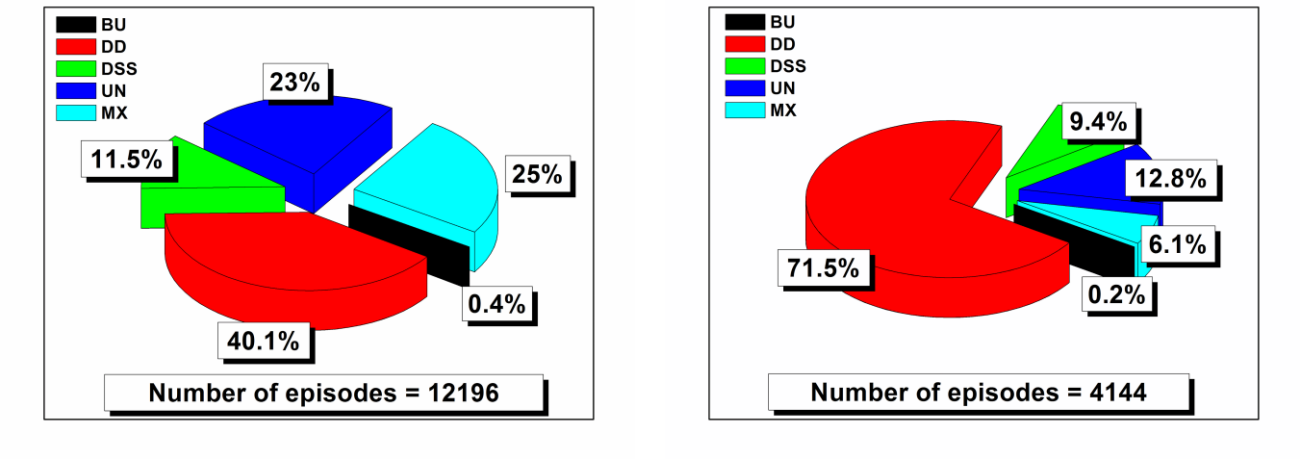
To our knowledge, we don't have enough information about the FF, Ångström exponent (or alpha), and effective radius retrievals' accuracies since those quantities have not been evaluated to the same extent that AOD has. FF and alpha data are derived from spectral information and their accuracy is determined 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.

"*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 a vailable 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?*"

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).

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). Here, we are presenting the results for the whole study region, as they have been described in Gkikas et al. (2016):



(i) (ii)

Figure R9: Percent contribution of BU (in black color), DD (in red), DSS (green), UN (in blue) and MX episodes (in cyan) to the total number of aerosol episodes over the sea surfaces of the broader Mediterranean basin as well as of its western, central and eastern parts. Results are separately given for: (i) strong and (ii) extreme aerosol episodes.

According to our results, 40.1% and 71.5% of the overall strong and extreme Mediterranean episodes, respectively, has been classified as DD episodes. Note, that desert dust aerosols can participate in MX and DSS categories, however in the present analysis we are keeping only the "pure" DD episodes. Moreover, the satellite retrievals which we are using are representative for the whole atmospheric column making thus impossible to quantify the contribution of mineral particles to the MX and DSS episodes. The performance of the algorithm with respect to the number of identified aerosol episodes is relatively satisfactory, since 77 and 87.2% of all strong and extreme episodes, respectively, have been classified. Nevertheless, 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.

3) Presentation of Results.

"*3.1) Results are reported in Section 4 which however includes a large body of material intended to provide a sort of 'validation' of the method followed (comparison with AERONET and in situ PM10 data). This is a bit confusing as, in my view, the logical sequence would be to present the methodology, then check/demonstrate its validity, and only at that point present the results obtained by that methodology. I would therefore rename the relevant sections accordingly.*"

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. In Section 3, it is described how the satellite algorithm operates. Then, depending on the analysis phase, the reader can easily follow our methodology, at each stage, about the comparison of satellite algorithm against AERONET/PM_{10} measurements (Section 4.2), how the CALIOP lidar profiles are used in order to describe the Mediterranean desert dust outbreaks' vertical structure

(Section 4.3) and how the dust outbreaks' vertical distribution can affect the level of agreement between columnar AOD retrievals and ground PM_{10} concentrations (Section 4.4).

"*Additionally, for its contents, the evaluation of the Method (Section 4.2) within Section 4 does not represent a real 'validation' but rather a 'comparison' with other datasets. Just to mention an example: the comparison of the AERONET AOD to the MODIS one (Figure 5) does not represent a validation of your 'objective and dynamic satellite algorithm' but rather a 'validation' of the Modis-AOD-retrieval algorithm. Therefore, I would avoid using expressions as 'the performances of the satellite algorithm are evaluated', widely used throughout the text, and rather refer to this material as: 'the results of the satellite algorithm are compared to...', which is completely different.*"

We agree with the Reviewer's comment. In the updated version of our algorithm the relevant parts have been modified accordingly.

"*3.2) In the same Section, I also believe the comparison with PM10 measurements has little validity in the context of this work. You want to report on the vertical structure of desert dust events, as clearly highlighted in the title of the manuscript. Your results show that several dust events do not reach the ground (see for example Figure 12), so: what do you expect to derive from the straightforward comparison of (columnar) AOD to (ground-level) PM10? I think this topic is of potential interest in general, but not in the form it is presented here. I would remove this part from the manuscript, as it does not add much to the text and rather makes it more confusing and weak.*"

In the revised manuscript, we have added a new section related to the investigation of specific dust events where satellite algorithm's outputs, ground PM_{10} measurements and CALIOP profiles are available concurrently providing an insight of possible factors which can lead to agreement or disagreement between MODIS AOD and ground PM_{10} concentrations. For this reason, to our view, it is required to sustain the part of the analysis related to in-situ measurements. Moreover, in the present analysis is provided information about the success score (a measure related to the performance of the satellite algorithm, Fig. 8-iii), dust contribution to the total PM_{10} (Fig. 8-iv) as well as the mean (Fig. 8-v) and median (Fig. 8-vi) PM_{10} concentrations. All the aforementioned results were not presented in Gkikas et al. (2013).

Other general comments

"*- Title: I would suggest to modify it as 'Mediterranean Intense Desert Dust Outbreaks from columnar and vertically–resolved remote sensing data'*"

We have changed the title to "Mediterranean intense desert dust outbreaks and their vertical structure based on remote sensing data" according to the suggestion of the Reviewer 3.

"*- Please define somewhere at the beginning of the manuscript the term 'intense dust events' you often refer to in the text and use the acronym IDD to refer to it. Specify clearly that, according to your classification, IDD events are divided into 'strong dust events' and 'extreme dust events' (and use respectively the acronyms SDD and EDD to indicate them throughout the text). This will improve its readability.*"

In order to sustain our terminology consistent with Gkikas et al. (2013) we prefer not to change the terms "strong" or "extreme" DD episodes to IDD and EDD, respectively, as it has been proposed by the Reviewer. The term "intense" is just a generic term trying to highlight that emphasis is given to intense dust outbreaks. It is repeated many times in the manuscript just to remind to the reader which dust outbreaks are considered in the present analysis.

"*- Please define somewhere at the beginning the exact study region considered.*"

Done (Lines: 332-333).

Minor Comments

"There are several minor revisions the manuscript would need. However, as major revisions are requested and the text will probably change a lot in its revised version, a not exhaustive list of minor comments is given below."

Section 2

"*Section 2.1.1. The ocean and land Angstrom exponents are computed using different wavelengths. Please explain why you use the same threshold for them in your algorithm.*"

The difference between Ångström exponents calculated in the spectral ranges where MODIS provides data over land (470-660 nm) and sea (550-865 nm) is very small (almost negligible). This is confirmed by the findings in Basart et al. (2009), who classified aerosols according to their type based on ground observations derived by AERONET stations located in northern Africa, Middle East and Mediterranean (http://www.atmos-chem-phys.net/9/8265/2009/acp-9-8265-2009.pdf). The authors reported that in the most dust affected Mediterranean sites δα (δα=α(440,675)–α(675,870)) values are almost equal to zero. Note that the wavelength pairs used for the calculation of Ångström exponent by MODIS sensor over land and sea are similar with those used in Basart et al. (2009). This indicates that common Ångström exponent thresholds over land and sea surfaces can be used in the satellite algorithm.

"*Page 27686, Lines 8-13. Confusing, please explain better*"

The CAD scores are bit values indicating the presence of aerosols (negative CAD scores) or clouds (positive CAD scores). As the CAD scores are getting higher, in absolute terms, the retrieval algorithm is more reliable either for aerosols or clouds. In our study we are using only the total backscatter retrievals associated with CAD scores between -100 and -20, as it has been proposed by Winker et al. (2013). Moreover, we are providing the weblink where the reader can find the relevant information in the CALIOP documentation. We have rephrased the relevant part in order to be clear to the reader (Lines: 271-272).

Section 2.2.2

"*Please specify better which data from AirBase and/or EUSAAR are you using (if this part will be kept in the revised version, which I discourage).*"

The data for the stations Montseny and Finokalia have been taken from the EUSAAR database while the corresponding data for the other stations have been downloaded from the Airbase database. Both have been added in the final version of the manuscript (Lines: 323-326).

Section 3

"*Page 27689 Line 22: It is not the quality of results to be improved but rather the quality of input data. Please rephrase.*"

We have rephrased the relevant part according to the Reviewer's comment.

"*Lines 25-28: '...in the present version of the algorithm are not implemented temporal filters, concerning the availability (masking out of AOD grid cells with less than 50% available data of the time-series) of raw AOD data, in contrast to Gkikas et al. (2009, 2013)'. This sentence is not clear to me. Do you mean in this work L3 pixels having less than 50% of L2 data are included? If so, do you consider this as an improvement of the methodology?*"

This sentence refers to the temporal availability of the L3 retrievals (i.e. the number of days over the period 2000-2013 or 2003-2012 where AOD data are available) and not in the number of L2 counts which are used for the calculation of the L3 retrieval. In order to avoid any misunderstandings we have removed it from the text since our analysis clearly shows that only few grid cells in the north parts of the study region have less than 50% available observations (Please, see our answer in comment 1.1).

"*Line 29 to the next page: Rephrase as this sentence is confusing. At the beginning I understood you performed the analysis only considering CF < 0.8, but then in Section 4 you mention (Page 27692, Lines 9-12): '...The analysis has been repeated (results not shown here) considering only AODs associated with cloud fractions lower/equal than 0.8,...' . I think you should restrict the analysis JUST to the cases with CF < 0.8, as AOD is not reliable above this threshold.*"

We have reproduced and show below the obtained geographical distributions of frequency of occurrence for the strong (Figure R10-i) and extreme (Figure R10-ii) DD episodes, when applying the same satellite algorithm but using only daily AOD values associated with cloud fractions (CF) lower than 0.8. The analysis has been accomplished for the period March 2000 – February 2013 using MODIS-Terra retrievals. For the strong DD episodes, it is evident that the spatial distribution reveals many similarities with the corresponding distribution obtained without applying the cloud filter (Figure 2 i-a of the revised manuscript). The main differences are encountered in the central parts of the Mediterranean Sea while the maximum frequency can reach up 11.9 episodes year^{-1}. On the contrary, for the extreme DD episodes the "zone" of maximum frequencies is restricted across the northern African coasts instead of the central Mediterranean Sea (Figure 2 ii-a of the manuscript).

The cloud filtration leads to the exclusion of possibly overestimated AODs retrievals, due to cloud contamination, from the dataset. Nevertheless, in the majority of the identified DD episodes without cloud filtering DD episodes the collocated AERONET AODs are relatively high indicating thus the occurrence of a dust outbreak (Figs 7-iii-a and 7-iii-b in the revised manuscript). This adds confidence to the general results (without applying the cloud filter). Taking also into account that cloud filtering diminishes the dataset (number) of DD episodes, we think that is better to use the raw AOD data (without cloud filtration).

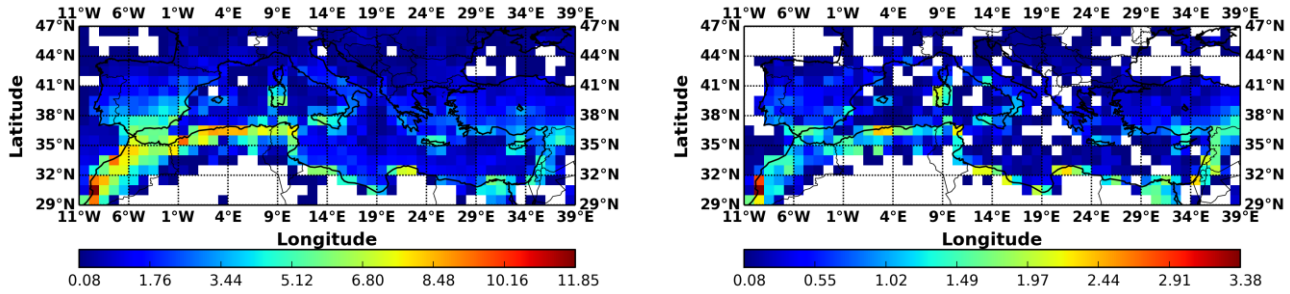


Figure R10: Geographical distributions of the occurrence frequency (episodes/year) of: (i) strong and (ii) extreme desert dust episodes, averaged for the period Mar 2000 – Feb 2013 (MODIS-Terra), over the broader area of the Mediterranean basin, when the satellite algorithm operates with AODs associated with cloud fractions lower than 0.8.

Section 4

"*Page 27690, Line 21-22: '..The obtained patterns are in a very good agreement with those presented by Gkikas et al. (2013),..'. This is just an example of similar comments you often insert in the text to comment your results. However my suggestion is to avoid repetition of this concept as it is rather obvious that results comparable to those presented in Gkikas et al. (2013) are obtained in this study, which follows a very similar methodology.*"

We have tried to reduce the number of similar statements in our manuscript as it has been proposed by the Reviewer. However, we would like to mention that it is reasonable to refer to Gkikas et al. (2013), only in the parts which are common but few, since the two different versions of the satellite algorithm (mainly different periods) are applied in each study. Moreover, these statements in the initial document are restricted only in Section 4.1 (geographical distributions) and in the part of the manuscript where the structure of the satellite algorithm is presented.

"*Page 27692, Lines 9-12: '...The analysis has been repeated (results not shown here) considering only AODs associated with cloud fractions lower/equal than 0.8,... ' . You are commenting something which is not shown and provide no explanation of the differences found. As already commented above for the same sentence, I think your study should be limited to the cases with CF < 0.8, avoiding most of the comments at the end of Section 4.1.*"

We think that we have already answered satisfactorily in our previous answer (two comments above).

"*Page 27693, Section 4.2: The title of this section is inappropriate for the reasons explained in my general comments 3.1 and 3.2.*"

The title has been modified according to the Reviewer's comment.

Page 27694

"*Lines 5-7: Do you mean you '...found at least one strong or extreme dust episode' over the whole period considered?*"

In each grid cell (1^{o} x 1^{o}), where an AERONET station is located within its geographical limits, it is found the number of DD episodes (strong or extreme) with coincident optical properties from the ground. It is clear that these calculations are done for each AERONET station at a pixel level and the number of coincident measurements depends on the availability both of satellite and ground-based retrievals.

"*Lines 5-19: This part is quite confusing. What's your aim here and how are you pursuing it? How do you define a 'clim' value? (what does it mean 'calculated from all the available retrievals'? how many data-points are used? Do these data cover the same period of your satellite dataset?). Additionally, when you refer to AERONET data please use 'ground-based' instead of 'ground' as this latter can be confused with 'in situ'.*"

We think that the relevant part of the document provides satisfactorily the information that we want to give to the reader. The climatological value is calculated from all the available AERONET retrievals, collected from all sites, during the period 2000-2013. Of course, this period is not common for each AERONET station individually since the availability of the AERONET data varies depending on the station. Here, the mean value is calculated from all these available retrievals derived by all AERONET stations located into the geographical limits of the study region (Mediterranean). Please note that the number of AERONET retrievals is provided in the caption of Figure S3 as well as in the legend of Figure 6 (revised manuscript). For 7 AERONET stations (depicted with cyan circles in Figure 4) we have identified the DD episodes based on ground-based observations and we are

comparing the outputs versus the satellite retrievals (Figure 7) for the time period where AERONET observations are available (see Table 1). However, we have rephrased a little bit the relevant part of the document in order to be clearer to the reader (Lines: 510-521). As it concerns the term "ground" we have changed it to "ground-based" according to the reviewer's suggestion.

Section 4.2.1

"*Subsections of Section 4.2.1 should be numbered. If I understand correctly, you are comparing the 1x1 degree satellite data with the AERONET data (measured in a specific site). Please at least comment on the expected spatial variability of AOD with in a 1x1 cell, and therefore on the validity of such an approach.*"

In the initial version of the paper, submitted to the ACPD, the subsections were numbered according to the Reviewer's suggestion (i.e. 4.2.1.x) but the Journal didn't accept it (not supported) due to its typesetting rules. As it concerns the second comment of the Reviewer, we would like to point out that the Figures 5 ii-a and 5 ii-b, and the relevant discussion in the text, address the issue of the sub-grid spatial representativeness and homogeneity, respectively, inside the Level 3 grid cells (1^{o}x1^{o} spatial resolution).

"*Page 27694, Line 22: '...346 pixel level intense DD episodes'. To understand the relevance of this number it would be important to mention somewhere how many pixels you have in your domain, how many of these are classified as intense DD (IDD).*"

Our study region consists of 900 Level 3 grid cells of 1^{o} x 1^{o} spatial resolution. At about 200-250 grid cells, located in the desert parts of the study region, AOD retrievals are not possible since we are using the MODIS Dark Target Algorithm products. Moreover, the number of grid cells with available AOD retrievals varies day by day, since many factors (e.g. clouds) can prohibit the satellite observations. During the period 2000-2013, the number of grid cells with available AOD data is equal to 2426303. Based on our algorithm, we have identified 22016 strong and 10619 extreme DD episodes, respectively, at pixel level, while the overall (strong+extreme) sample comprises 32635 DD episodes. The number (346) of coincident satellite algorithm's outputs and available AERONET retrievals corresponds to 1.06% of all DD episodes which have been identified based on our methodology. At a first glance, this percentage it seems quite low but can be easily explained. First, most of the identified DD episodes occur over maritime areas where there are not available AERONET observations. Furthermore, the station-by-station AERONET data availability varies a lot with regards to the satellite period (2000-2013) and also it is well know that there are gaps in the timeseries.

Section 4.3

"*How much do the Calipso-based result change if you use the overall calipso database and its aerosol type discrimination to investigate desert dust, not limiting only to those cases previously classified as IDD in your scheme?*"

In order to answer to the Reviewer's comment, we have reproduced the Figures 9-i and 9-ii considering only the dust and polluted dust records of the CALIOP-CALIPSO aerosol classification scheme without taking into account the satellite algorithm's outputs. The obtained results are available in the supplementary material (Fig. S6-ii and S6-ii) while in the last paragraph of Section 4.3.1 are discussed briefly the differences between average dust and intense dust outbreaks' conditions.

Figures:

Figure 1

"*The scheme of the work is exactly the same of Figure 2 in Gkikas et al. (2013), which is quite inconvenient, please modify highlighting differences from that work or remove the Figure.*"

We prefer to keep Figure 1 in our manuscript since it helps the reader to understand our methodology just reading this paper without going back to Gkikas et al. (2013).

Figure 2

"*- It should be enlarged as it is not very readable at the moment.*"
"*- As commented above, it should refer to the same period for Aqua and Terra.*"
"*- There is a clear discontinuity between land and ocean, can you comment?*"
"*- Change the color scale to more clear numbers (e.g. 0-10, top, 0 – 3 bottom)*"

-We have enlarged the size of the figures.
-We have already explained the reasons that we prefer to keep the MODIS-Terra period.
-As it has been addressed in our previous answers, the retrievals above land are less reliable compared to the corresponding ones over sea surfaces. This can affect the identification of DD episodes as well as their intensity. However, we don't think that this discontinuity is so pronounced as it has been stated by the Reviewer, particularly for the strong DD episodes.
-Done. For the strong and extreme DD episodes the ranges are 0-10 and 0-3.5, respectively.

Figure 3
"*- It should be enlarged as it is not very readable at the moment.*"
"*- In the i-plots, I see a problem of misclassification over the Po valley in Italy. Can you comment on that?*"

-Done
-Please note that the high intensities over the Po Valley are associated with relative small frequencies. This means that the intensity over the northern parts of the study region is computed from a small sample of DD episodes. In such a case, few relative high AODs, associated with long-ranged dust outbreaks, can exceed the defined thresholds in the aforementioned locations resulting thus to high intensity values. However, we must always keep in mind that possible misclassifications, especially over land, by the satellite retrievals or cloud contamination can affect the identification ability of the satellite algorithm.

Figure 12
"*- It should be enlarged as it is not very readable at the moment.*"
"*- Please specify the units of the backscatter values.*"

-We have enlarged furthermore the size of the figure.
-We have added the backscatter units in the captions of Figures 9 and 10.