Response to Reviewer3

We would like to thank the Reviewer who helped us to improve our paper through his/her report. Below are listed our detailed responses (regular font) to each comment raised by the Reviewer (bold font).

The paper presents an interesting study for calculating DRE with the use of the NNMB-MONARCH model (NNMB). It is a well written paper which with the following revisions it could be published in the ACP journal. My main comments are:

- In order to accept the results of such a study, a more comprehensive validation of the presented outputs using real measurements and an analysis of the uncertainties introduces in several phases of the method have to be presented.

In the revised manuscript we have made a more detailed comparison between MODIS-NMMB for the cases (dust outbreaks) which are analyzed here. Regarding the validation of radiation and temperature fields, the discussion has been updated whenever is needed. Please see our responses to your comments below.

- A major aspect of the paper is not clarified. The abstracts talks about DRE and as the authors point out this is mostly aerosol optical depth (AOD) dependent. MODIS retrieves total (dust + other types) AOD while NNMB only dust AOD (that is what is shown throughout the text and in e.g. figure 3). So the authors have to clarify if they talk about Dust DRE or DRE. If someone assumes that these 20 events are purely dust events, an AOD comparison of MODIS AOD and NNMB AODs have to be included (not quantitevily as in fig. 3), in order try to assess the model results.

We have changed the title of our paper from "*Direct radiative effects of intense Mediterranean desert dust outbreaks*" to "*Direct radiative effects during intense Mediterranean desert dust outbreaks*" so that the goal of our study is more clear. This modification has been made since based on the configuration of the model the amount of dust aerosols is simulated dynamically (online) while for the other aerosol types the GOCART climatology is used (Lines 340-342). Moreover, the DREs are computed for days in which intense dust outbreaks prevail over the greater Mediterranean basin. Under such conditions, and over places where Saharan dust is transported, dust predominates and is the main contributor of AOD, even in MODIS AOD retrievals. Of course, in such cases all aerosol types exert a perturbation of the radiation budget, but the impact of mineral particles is predominant. A quantitative comparison between MODIS and NMMB has been made (suggested also by the Reviewer 2) and the obtained results are presented in Figure S2 (supplementary material) and discussed in Section 5.1.

- A major issue of the paper is the link between the NNMB results and the Gkikas et al., methodology (GM) for identifying dust episodes. Some questions that have to be clarified on the manuscript are the following:

(i) Are the domains seen in figure 3 and 1 have been used in the GM for all the episodes that are presented in the table 1? Is there a mix of surface and sea Modis pixels used?

The identification of DD episodes through the implementation of the satellite algorithm is made only for the Mediterranean Satellite Domain (MSD, red rectangle in Figure 1) as stated in the manuscript (see lines 192-194). The structure, methodology, and operational phases of the satellite algorithm have been presented in detail by Gkikas et al. (2013, https://www.atmos-chem-phys.net/13/12135/2013/). Briefly, the algorithm operates separately over land and sea surfaces by taking into account the MODIS AODs obtained by the dark target land and ocean retrieval algorithms.

Therefore, the number of DD episodes presented in Table 1 corresponds to the number of grid cells (1° x 1° spatial resolution) where a desert dust (DD) episode has been recorded/identified within the geographical limits of the MSD. Please see lines 207-210 and the caption of Table 1 in the revised manuscript.

(ii) When GM identifies an episode (e.g. example of figure 3) are the DRE calculations of NNMB account only the relative (episodic) modis pixels? I think the answer here is no but it has to be clarified. So, If the answer is no (thus the whole domain (e.g. MSD) is used for NNMB) then the importance of the GM episode identification is only partially valid. (e.g. a lot of white in fig. 3 are used based only on NNMB and not on GM). As identifying an episode in a limited area in the MSD domain does not mean that this is valid for the whole domain.

We think that it is clear that the NMMB DREs calculations are made all over the Mediterranean basin and not only over the episodic MODIS pixels. This does not limit the validity and importance of GM dust episode identification. It is self evident that when talking about a dust episode over the Mediterranean, not the entire basin but just a significant part of it is expected to be dominated by dust, which is adequately ensured by GM. Therefore, having "a lot of white in Fig. 3" is not strange, unreasonable or problematic, but on the contrary it is expected and sound. Nevertheless, this does not prevent us from talking about Mediterranean dust episodes and radiative effects (DREs). The only issue that might be relevant to this comment, is averaging regionally over the Mediterranean, where dust and no dust dominated areas are considered together, but even in such cases DRE computations are meaningful. In the revised manuscript, the calculations of the regional DREs have been made taking into account all the grid points and therefore the spatial representativeness is consistent at each forecast step and among the studied cases.

(iii) If the whole domain is used are results of table 1 dependent in addition to dust AOD to the spatial extension of the event? Can a number of different episodes with different spatial extends and AODs, averaged (table 2)

In the revised manuscript (Table 1, lines 213-218) it is explained that the frequency and regional intensity, i.e. AOD of 20 dust outbreaks, is calculated from the total pixel-level DD episodes, therefore the results of Table 1, more specifically the intensity, are not dependent on the spatial extend of the episodes. As already answered in the previous comment, regional DREs can be computed for every dust episode. Therefore, as it concerns the second part (sentence) of this comment (e.g. Table 2 results) we believe that averaging DREs over the 20 different dust episodes is meaningful and representative of DREs during Mediterranean dust outbreaks.

Another example of the last point above is that modis GM detects a plume (high AOD) covering very few pixels in the western part of MSD (for example last row of figure 3). Then based on GM the whole MSD domain is considered as the one that will provide the DRE. In this case the link on GM used as a proxy in this work is very week as it covers only a small part of the domain, plus AODs are not compared. So also a number of episodic pixels should be included in these GM dust episode restrictions. Or simply dust outbreak identification can be based on NNMB spatial and NNMB-AOD absolute criteria as now the link with GM is really weak.

Most of the content of this comment has already been answered. However, we would like to note that of course intense dust outbreaks are not supposed to cover the entire Mediterranean, on the contrary, they always cover a part of it, this is logical. However, this does not prevent us of talking about dust episodic days over the Mediterranean basin whenever such dust outbreaks occur. And, moreover, it also does not prevent us from computing DREs all over the Mediterranean basin, even averaging over it. Therefore, there is not any problematic link in our concept and methodology combining the detection of dust outbreaks with GM and the DRE computation with NMMB. Concerning the last sentence and suggestion of the Referee, of course this is an option, i.e. dealing with detection of dust outbreaks and computing the associated DREs solely using the NMMB model. However, this would be purely theoretical. On the opposite, detecting intense dust outbreaks based on an observational approach, i.e. using MODIS products, is more appropriate. Finally, as already stated in our responses to this Referee's previous comments, a comparison of AODs has been made and it is discussed in the revised version of the manuscript.

In addition, in this case (and others e.g. west domain of fig. 3b) NNMB dust pixels cover less than 50% of the MSD. When averaging the 20 cases this percentage of pixels varies a lot. In the end you are averaging and provide a result e.g. SW = -9.7. So some of the outbreaks contribute much more and some others not, based on the dust coverage on the MSD only. Where can such statistics be used?

First, we would like to state that in the revised manuscript the regional DREs have been calculated considering all the grid points without setting any criterion on the simulated dust AOD or on clouds (this approach was initially followed). Therefore, at each forecast step and among the 20 desert dust outbreaks the number of grid points is constant. This ensures that the spatial representativeness of the regional DREs does not vary in time and among the studied cases (Figure 5).

To summarize, if GM is not used for AOD validation and GM identifies as "dust episodic pixels" only a fraction of the pixels used finally from NNMB for calculating dust DRE, then its use becomes not important for this study. So if someone trusts NNMB for DRE calculations, then it is much more easy to trust it also for dust outbreak identification.

We think that our previous responses give a sufficient answer to the reviewer's summary comment.

- There are more than 100 references and a lot of discussion about aerosol effects and model applications, but very few about NNMB validation on e.g. AOD retrievals. And only one (Ohmura) on BSRN radiation related validation. I think it is more essential to prove the validity of AOD NNMB output (e.g. radiation) and intermediate parameters (e.g. AOD), than a numerous studies cited here, with a very theoretical link to the paper.

It is not the first time that NMMB is used, so validation of its AOD has already been done. In our paper, we have included all the available studies regarding the evaluation of the simulated AODs relied on the same NMMB version which is used here (Lines 294-306). Moreover, in the revised manuscript we are providing the weblink of the SDS-WAS System (https://sds-was.aemet.es/forecast-products/forecast-evaluation) in which is presented the forecast evaluation of NMMB AODs, among other aerosol models, utilizing ground-based (AERONET) and satellite (MODIS) retrievals as reference.

Concerning BSRN, we would like to remind and underline that it provides just reference radiation measurements. The BSRN is considered the best global network of quality radiation measurements. There is a very high number of scientific papers (http://bsrn.awi.de/other/publications/reviewed-scientific-papers-referring-to-bsrn/) or reports (http://bsrn.awi.de/other/publications/other-related-reports-and-papers/) referring to BSRN, so there is no need to make further reference to it than to the key paper of Ohmura et al. (1998) which is commonly used as reference for BSRN data. The validity of NMMB radiation fluxes is exactly proved through their comparison against BSRN measurements.

- The validation using BSRN is incomplete. In the document and in the abstract you are talking about this validation and 8 stations. Then in the manuscript only one station is shown. And from that only 4 days. In order to validate the results a more comprehensive analysis of long

term periods of these 8 stations is needed. Probably Ohmura has answered some of the validation related questions, but this paper focuses on "intense dust outbreaks", and a specific model, so results might differ from the Ohmura related ones.

We would like to point out that the calculated biases (NMMB-BSRN) over the hindcast periods, for each case and for each station (6 in total), are given already for the SW and LW radiation in Tables S2 and S3, respectively and discussion (lines 882-887) refers to their results. In the main text, we have decided to present just as an example the obtained results for the SW (first row in Figure 10) and LW (second row in Figure 10) radiation for two dust outbreaks (22/2 -25/2/2004 and 21/4-24/4/2007) that affected the Sede Boker station, for which concurrent AERONET retrievals were available. This allows us to give a better insight regarding the factors that can affect the level of agreement between model and ground observations. We agree with the reviewer that a long-term evaluation is valuable (i.e. identification of systematic errors) but for our purpose focus is given only on specific desert dust outbreaks trying to investigate if the inclusion of dust-radiation interaction in the numerical simulations can improve the forecasting skills of the NMMB-MONARCH model.

- There are several issues that have to be clarified/commented on the input parameters of the model:

(i) Optical properties proposed in figure 2. Have been validated?

The optical properties have not been validated. The model dust optical properties are based on single-particle optical properties derived by the GOCART model (Chin et al., 2002) and refractive indices from the Global Aerosol Data Set (GADS) (Koepke et al.,1997). Both datasets are very well known and very much often used and cited in literature, and therefore we believe that there is no need for further validation here.

(ii) Water vapor, carbon dioxide, ozone, methane and oxygen. Where do you find these inputs?

Water vapor comes from the model simulations. We used a fixed value of CO_2 (350 ppm), methane (1.5 ppm) and oxygen; and a seasonal climatology for ozone.

(iii) Differences in dust optical properties of Sahara and middle East sources. What did you use and how much uncertain are they? and what is the contribution of this uncertainty in the final DRE budget?

The dust single-particle optical properties and the emitted size distribution are constant throughout the simulation domain without discriminating between different dust sources (Sahara, Middle East). At each forecast step, the aerosol optical depth (AOD), the single scattering albedo (SSA) and the asymmetry parameter (ASYM) have been produced based on the formulas presented in Pérez et al. (2006) utilizing the simulated mass concentration, the GOCART single-particle optical properties and the refractive indices from the Global Aerosol Data Set (GADS) which have been modified according to Sinyuk et al. (2003), as it has been described in Pérez et al. (2011) (lines 331-336). Regarding the last question of the reviewer, in order to be give an accurate answer a sensitivity analysis is required. More specifically, it must be investigated how the variation of key aerosol optical properties (AOD, SSA and ASYM) will affect the perturbations of the radiation budget and subsequently the associated impacts on dust AOD, dust emission, meteorological variables and radiation. This is something that has not been done in the present paper but it will be considered in a future work dedicated to all the aforementioned aspects considering also other parameters (e.g., dust layer vertical extension) which can affect DREs.

- BSRN and model differences in wavelength integrals of solar radiation. You mention: "These differences might contribute to the level of agreement between model and observations;

however, are not discussed in our evaluation analysis". I think this is an important issue that have to be clearly discussed if a proper validation is included.

For solar radiation, the NMMB-BSRN SW flux departures, attributed to the different spectral coverage and integrals, are minor, varying from 1 to1.5 % (higher values for the model), therefore they do not affect substantially the agreement (in terms of biases) between model and measured fluxes.

- As already mentioned AOD comparisons from MODIS and NNMB could add value to this work. "The model's ability to reproduce correctly the spatial patterns and values of dust AODs is crucial for a successful computation of the dust DREs, since DREs are determined to a large extent by AOD". In addition you are mentioning modis uncertainty in section 2. Is this getting high (e.g. ~ 0.5) for both sea and mostly surface retrievals when you examine AODs in the order of 2-3 based on the table 1? And is this uncertainty already important for such outbreaks for the GM and indirectly for the DRE related uncertainty?

Actually, the uncertainty of C051 MODIS AOD retrievals is not reported in section 2, where only the detection of dust outbreaks is described. The uncertainty of MODIS AOD retrievals over ocean is ±0.03±0.05*AOD (Remer et al., 2002) while over land is higher and equal to ±0.05±0.15*AOD (Levy et al., 2010). The maximum MODIS retrieved AOD, over both continental and maritime areas, do not exceed 5, which means that the AOD uncertainties above sea and land, in absolute terms, are smaller than 0.28 and 0.8, respectively. In our cases, but also in general, these maximum AOD uncertainties are locally restricted and not recorded frequently (see Figures 3 and S1) while uncertainties are generally smaller, and thus do not affect the GM. Moreover, they neither affect DREs, since as already explained in our previous responses and in the manuscript, the DREs have been computed via the NMMB simulations without setting any constrain depending on MODIS retrievals (i.e., availability, magnitude).

- Table 2. These statistics are not referring to the model uncertainty but is an averaging of the episodes provided by the GM. NNMB DRE uncertainty is much more useful for any future user of these results. For example a systematic bias can not be identified here. This is also because the GM thresholds are mostly subjective as:

(i) Mean AOD values on dust related areas do not have an important statistical meaning due to the non normal distribution of AOD. It is clear that this is a published work and I have tried to follow the previous work by Gkikas et al and the relative open discussion, describing the method. However, as this is an open public statement I have to comment that AOD does not follow necessarily a normal distribution so using the mean is not absolutely correct. Moreover, dust outbreaks related pixels/locations can be characterized more from a bimodal distribution of AODs when another (than dust) important AOD source is rarely present (e.g. most of the marine grids of Mediterranean domain).

First of all, as stated by the Referee, we would like to remind that the GM method has already been published (Gkikas et al., 2013; 2016) just after the discussion that took place concerning the way of computation of AOD thresholds, i.e. geometric versus arithmetic mean AOD values, which implies its validity against similar arguments cited in this comment. Nevertheless, we can remind the following. We agree with the Reviewer that AOD follows a log-normal rather than a Gaussian distribution, and that the arithmetic mean and standard deviation are not probably the best metrics for the calculation of the AOD thresholds, even though both primary statistics are widely applied in numerous aerosol studies. During the review process of Gkikas et al. (2013), following a similar comment raised by one of the referees, proposing to calculate the AOD thresholds based on the geometric mean and geometric standard deviation, we recomputed the AOD thresholds and compared them to the typical ones already used (based on arithmetic mean and standard deviation). Although there were found some differences in the thresholds' magnitude, in general, the geographical patterns of AOD thresholds were similar for both strong and extreme DD episodes. As for strong episodes, those differences were rather small, for example typical AOD thresholds varied within the range 0.4-1.2 and the geometrical thresholds ranged from 0.4 to 1.6. On the other hand, larger differences existed for extreme DD episodes, with the typical thresholds ranging from 0.6 to 2.2 while the geometric ones varying from 1 to more than 10. However, such extremely high AOD values are extremely rare and using them would be unrealistic from the physical point of view. For these reasons, it was decided to rely on GM methodology of Gkikas et al. (2013).

(ii) GM: By definition high mean AOD values per pixel are closer to dust sources. That makes possible that a pixel with high (in an absolute sense) AOD close to a dust source to be considered non episodic and a pixel with lower AOD, away from the sources to be considered episodic. This is ok, as it is just a matter of definition. But it gets more important when it is used for DRE calculations. So, the latest can be problematic when you calculate DRE in dust outbreaks or filter the outbreaks, as for the first example pixel (high AOD) it is not an outbreak and for the second (lower AOD) it is characterized as an outbreak. The results using this method for DRE calculations become not easily useful and applicable.

The issue of the identification method of DD outbreaks based on pixel-level AOD values, has already been addressed in our previous papers using the GM methodology, following similar comments to the one made by the Referee here. It has been shown that any differences in terms of AOD thresholds and dust outbreaks features (frequency, intensity) were not substantial.

However, the most important concerning the rest of Referee's comment referring to possible effects of this issue on computed DREs here, we would like to clarify again that DREs are computed by NMMB and have nothing to do with the AOD thresholds. It should be clear and kept in mind that GM methodology is only used for the determination of days with intense dust outbreaks for which NMMB then operates and makes computations of DREs all over the domain.

- Last but very important, the paper is very long and in various cases the discussion includes a lot of details that in the end confuse the reader on what is the important findings here and which are not. Even for scientists in the field it becomes difficult to read. Authors have to try to reduce the length of the manuscript keeping the important aspects of the results presented. Basically for section 5 I would suggest to try to take out a lot of information that are secondary and to focus on the important results.

In the revised manuscript, following the suggestion of the Reviewer, we made an effort and reduced the paper length by removing some parts which can be considered as secondary information. However, at the same time, also following the Reviewers' suggestions, we added a discussion about the quantitative intercomparison between MODIS and NMMB as well as about the potential improvements on short-term forecasts of the temperature fields by the model. Therefore, the final length of the revised manuscript is similar to that of the original manuscript. We believe that any further shortening of the manuscript would be at the expense of its quality and scientific content.

Minor comments:

Line 141: it has already mentioned previously.

It has been modified.

Line 173: developed – improved

Done.

Table 1: episodes = grid cells

We think that is already clearly stated in the caption.

The overall approach of this paper is valuable and worth publishing. I strongly believe that after the above revisions, corrections and additional analysis it will be essentially upgraded and then it could be published in the ACP journal.