**Point-by-point response to the reviews**

**Referee 1**

We thank this reviewer very much for the so detailed constructive comments on this work. We have made changes to the manuscript accordingly. We colored our response in blue. Text from the manuscript is quoted with double quotation marks and new text is shown in *italics*.

**General comments**

This article presents multiple developments included in the dust cycle representation within the CAM6.1 model and assesses their impact on relevant variables, such as the dust surface concentration, deposition, size distribution, optical depth and direct radiative effect. The work conducted provides relevant information beyond the dust modeling community, as dust has impacts on different features of the atmospheric dynamics and chemistry, the climate and the Earth System. As such, I believe this article is well within the scope of the Geoscientific Model Development journal, it presents novel results, and it deserves publication.

Many thanks for the positive comments.

However, in my view, in its current form the reader has to put in a considerable effort to follow the details of the massive amount of work presented.

Thanks for the comments and time in reviewing the manuscript.

The authors present nine different experiments: five defining dust as a bulk species and four experiments considering speciated dust. This involves a duplication of experiments in which one (or several) of the new developments are tested, and adds an additional variable to the analysis, making it harder to focus on the specific impact of the new aspects included in the model.

With respect to the experiments design, the authors could better clarify the criteria used to include the new features in the tests. Instead of relying on a baseline (e.g. CAM6.1), and adding separately to that configuration the different developments (on the emissions scheme, dry deposition, size, or asphericity), the authors combine multiple developments in the different experiments. I believe these combinations could hinder a clean comparison of the effect of each development (e.g. looking at Table 4 it is difficult to know which pair of experiments allows disentangling the effect of shape and deposition changes). This issue is accentuated by the fact that the experiments are referenced along the manuscript by different names or acronyms, which further complicates tracing them.

There are a couple of different methods to estimate the effect of each development, such as the one we used and the one the reviewers suggested. Strictly speaking, either method cannot totally exclude the possible influence of the parametrizations that had already been included and can affect the dust cycle modeling in the base model, such as the advection scheme and cloud processing. The reason is that there likely exists a nonlinear “interaction” between the existing parameterizations and the newly introduced one, which seems weak though. We acknowledge that adding one by one seems clearer than the original experiment design, but it requires more simulations and thus more computational resources while yielding a similar estimate of the impact of each development (Fig. R1), compared to what we had presented based on our own’s experiment. We had selected our own’s set of experiments, because adding the modification on top of the previous change can help understand how the simulated dust cycle evolves while updating the model (MINE_BASE) toward the most advanced one (CAM6.α_MINE).

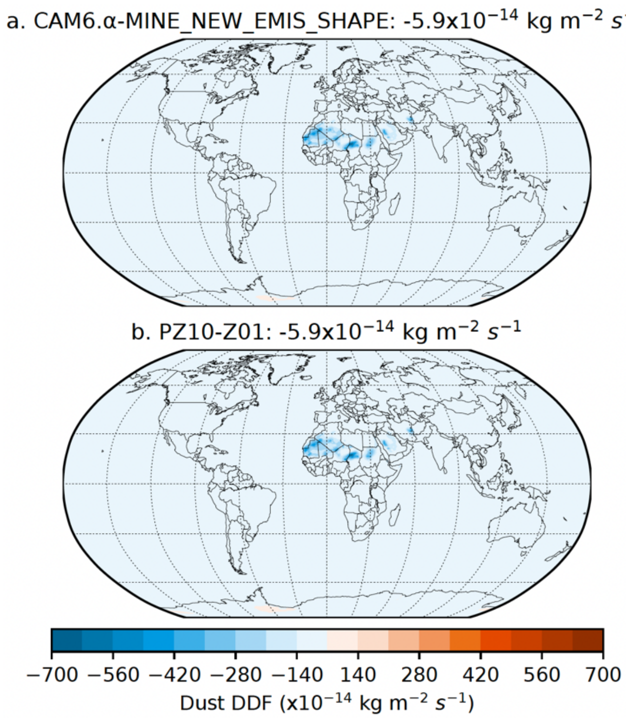



**Fig. R1.** Influence of changing to PZ10 on the simulated dry deposition fluxes in the dust-speciated model (change to the global annual mean of dry dust deposition: ~70 Tg) based on our experiment (a) and the suggested experiment by the reviewers (b). Quantified change to the global annual mean of dry dust deposition equals ~70 Tg by either method.

The BULK runs were constructed to investigate how the mistakenly set dust size distribution influences the dust cycle modeling and the estimate of dust DRE. This inappropriate size distribution has been employed in studies using the officially released BULK CAM6 and not in any study using the dust-speciated CAM. So, we do not have a good reason to perform size sensitivity tests in the MINE runs. What’s more important is that quantifying the impact of individual processes, based on the base CAM6.1 that uses an inappropriate dust size distribution, seems not that meaningful: it makes more sense to make such quantification using models with the "correct" size distribution. That is why in all the MINE runs designed for that purpose we revert the narrow coarse-mode size distribution to the broad one. Also, following the reviewer's design would change little to the results obtained from our experiments on the dust cycle modeling. This is because the offline dynamics and the dust tuning employed ensures quite similar dust cycles modeled by BULK and MINE with different developments (Fig. R2 and Fig. R3), if the size distribution is also set identical, since the sum of the mass fraction for each of the eight minerals always equals unity. We had pointed this similarity out in our originally submitted manuscript: "It is worth noting that with the dust tuning applied toward the similar global mean DOD ~0.03 the modeled dust cycle (i.e., burdens, concentrations, loadings, deposition fluxes) would be similarly comparable between the bulk and speciated dust models using the same offline dynamics and dust size distribution". Repeating the set of simulations using BULK instead to quantify the impact of each altered process would then yield similar results that we presented in the manuscript.

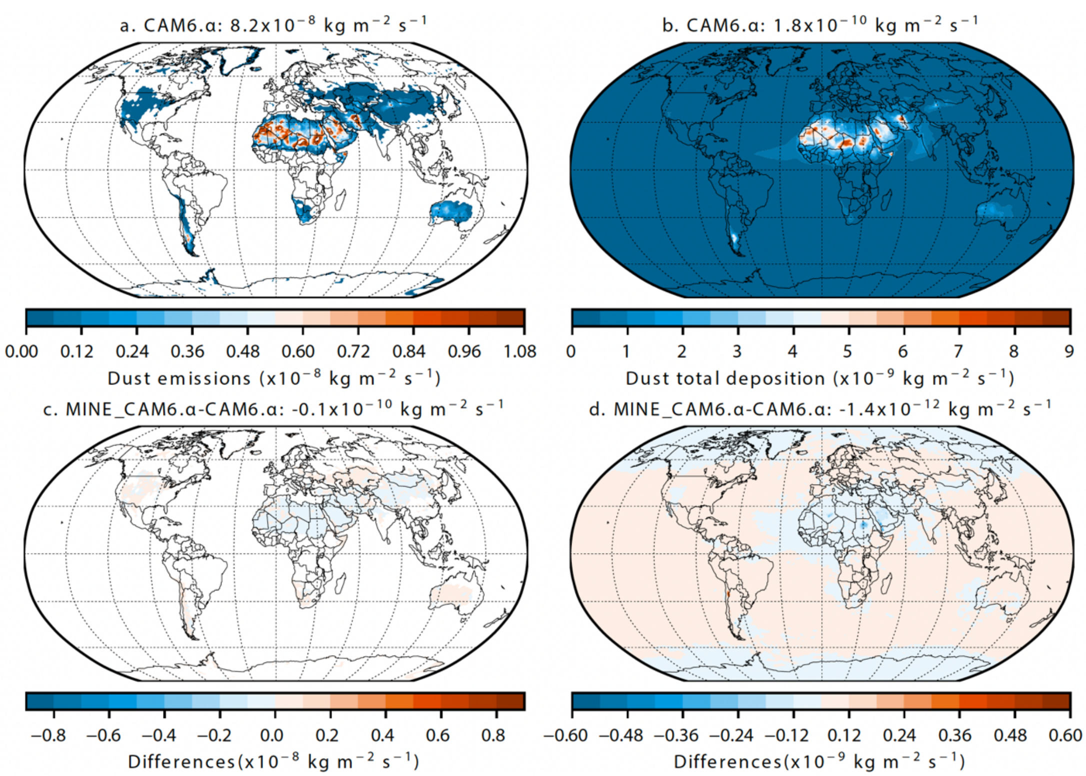



**Fig. R2.** Surface dust emissions (a; global annual mean=2891 Tg) and deposition fluxes (b; global annual mean=2893 Tg) simulated by CAM6.α and their differences (c and d; both global annual mean=22 Tg) between MINE_ CAM6.α and CAM6.α.

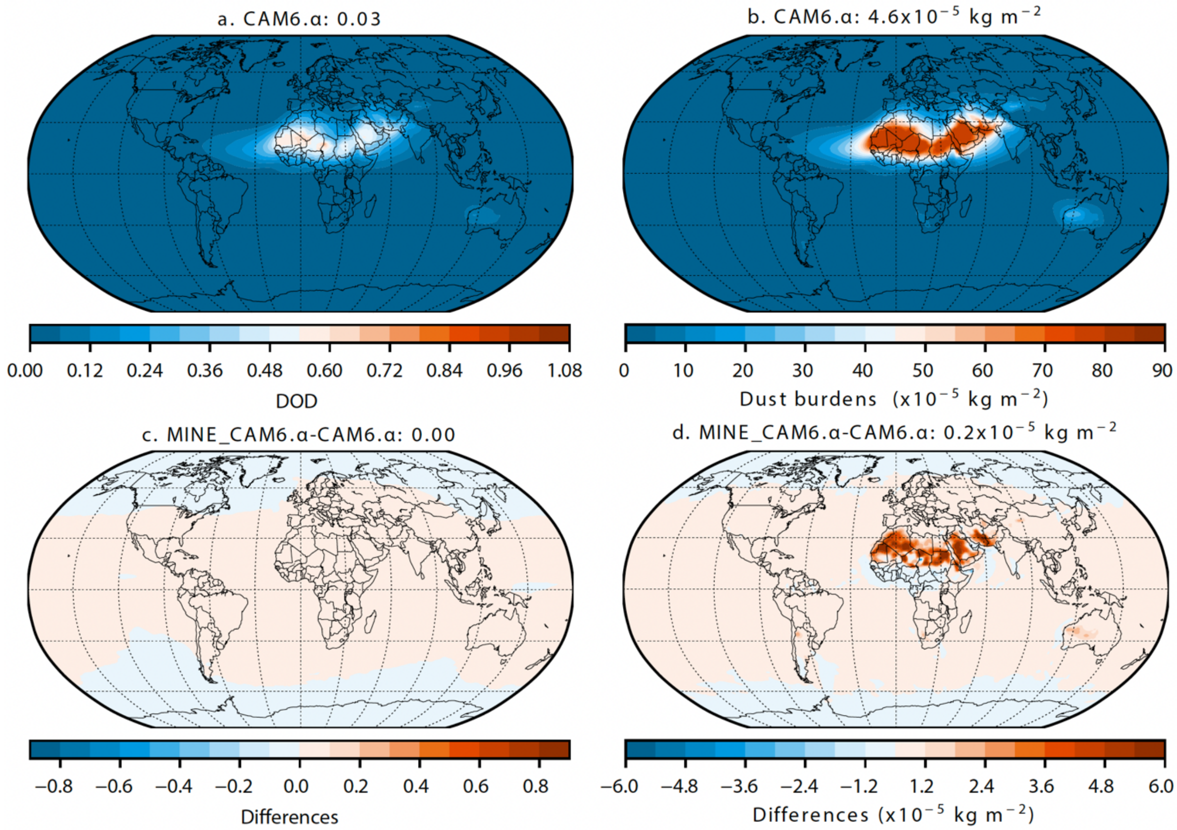



**Fig. R3.** The same as Fig. R2 but for DOD (a: global annual mean=0.030 and c: global mean difference=0.001) and dust burdens (b: global annual mean of dust mass=24 Tg and d: global mean difference$\approx$0 Tg), respectively.

NEW_EMIS and MINE_NEW_EMIS appear like a duplication of experiment for testing the new dust emission scheme. But the estimate of dust DRE differs considerably between the two experiments.

To reflect the Reviewer's suggestion, we added the following in the section "2.6 Experiment design**"**:

"*We investigate how the mistakenly set dust size distribution influences the dust cycle modeling and the estimate of dust DRE in the bulk-dust model rather the speciated-dust model, because this inappropriate size distribution has been employed in previous studies using the officially released bulk-dust CAM6 only and not in any study using the speciated-dust CAM. It is also reasonable to make all the quantifications in the model that use an appropriate dust size distribution. Therefore, we reverted the dust size distribution in all the speciated-dust runs to that configured in CAM5.*"

"*It is worth noting that with the dust tuning applied toward the similar global mean DOD ~0.03 the modeled dust cycle (i.e., burdens, concentrations, loadings, and deposition fluxes) would be* *similarly comparable between the bulk- and speciated-dust models that nudged toward identical offline dynamics and using the same dust size distribution (see Sect. 6). The quantified effect of each of the modifications would thus be similar if using the bulk dust model instead (Fig. S2: R1 in this report), but the modeled dust optical properties (e.g., single scattering albedo) by the bulk and speciated dust models differ considerably, resulting in considerably different dust DRE (Scanza et al., 2015) and DRE efficiencies between NEW_EMIS (CAM6.α) and MINE_NEW_EMIS (CAM6.α_MINE).*"

"*A comparison of the BULK and MINE models on simulating dust DRE had been previously documented (Scanza et al., 2015). This study includes the MINE runs because we want to check as well if the updates help improve reproducing the observed dust DRE efficiency in a model that may more reasonably represent the regional variation of dust optical properties.* Note that there are many ways to conduct sensitivity studies, which could lead to slightly different results. *We added the modification on top of the previous change to understand how the simulated dust cycle evolves while updating the model (MINE_BASE) toward the most advanced version (CAM6.α_MINE). This may not hinder a clean comparison of the effect of each development, since the 'interaction; between the existing and the newly introduced parameterizations seems weak (Fig. S2: R1 in this report).*"

To clarify how we quantify influence of each development, we added two columns in the Table 4 pointing out the size distribution used and purpose of each experiment and added the following text in the "Experiment design" section:

"We separately compared the performance of PZ10 to Z01, aspherical to spherical dust, and BRIFT to DEAD on the simulated dust cycle and quantified influence of each of those modifications on the climatic-effect estimate *by comparing the modeled dust cycle in the paired simulations CAM6.α _MINE vs MINE_NEW_EMIS_SHAPE, MINE_NEW_EMIS_SHAPE vs MINE_NEW_EMIS, and MINE_NEW_EMIS vs MINE_BASE, respectively.*"

To easily trace the experiments, we now refer to them using their case names instead of EXP# all through the text.

Finally, we added a separate new section to compare results from BULK with those from MINE:

"***6. Bulk- versus speciated-dust model***

*The bulk (CAM6.α) and dust-speciated model (CAM6.α_MINE) simulate a similar dust cycle with the difference between the two types of models orders smaller than those simulated by the former (Fig. 12 and 13: R2 and R3 in this report, respectively). This similarity results from the dust tuning toward the global mean DOD of 0.03, the same meteorology dynamics both models* *were nudged toward, and the design of the dust-speciated model that summing the mass fraction of each dust species equals unity. With the same reasons, the influence of each of the modifications on the modelled dust cycle quantified using the bulk model instead of the dust-speciated model, as this study used, would be similarly comparable. The modelled dust optical properties, however, differs remarkably (i.e., dust SSA; Table 6) with the simulated global mean dust SSA by CAM6.α_MINE (0.896) lower than by CAM6.α (0.911) at the visible band centered at 0.53 μm. Note the dust DRE is sensitive to variation of the dust SSA. This lower dust SSA obtained here in the dust speciated model than in the bulk dust model is consistent to the finding of a previous study (Scanza et al., 2015) using an early model version (CAM5). Correspondingly, CAM6.α_MINE yields a reduced dust cooling (Table 6) and DRE efficiency (Fig. 9) than CAM6.α. For dust DRE efficiency (Fig. 9), speciating dust in CAM6 tends to reduce the RMSE while retaining the horizontally spatial correlation in either SW (CAM6.α: RMSE=11; R=0.26 versus CAM6.α_MINE: RMSE=10; R=0.20) or longwave (CAM6.α: RMSE=4; R=0.86 versus CAM6.α_MINE: RMSE=3; R=0.84) or both spectral ranges (CAM6.α: RMSE=7; R=0.93 versus CAM6.α_MINE: RMSE=6; R=0.92). This comparison suggests that modeling dust as component minerals with the dust size distribution in coarse mode of MINE_NEW_EMIS_SIZE helps improve the model performance relative to modeling dust as a bulk to reproduce the retrieved dust DRE efficiency (Fig. 9a). The improvement, however, could be artificial because of the combined use of imaginary complex refractive index of hematite volume (see Fig. 1b of Li et al., 2021) and the volume mixing used in the dust speciated model to compute the bulk-dust complex refractive index (Li et al. in prep.), leading to artificially more absorptive dust than in the bulk dust model (Fig. 9a and Table 6).*”

Then, I believe that a fundamental piece of this article is the variety of observations, retrievals, model-derived products and model results that are used for the model evaluation. The modelling community could greatly benefit from the effort done here to compile that information and produce a benchmark for dust properties evaluation at the global scale (in present climate). Unfortunately, these are only presented in the article in a summarized manner (through a table). I would recommend adding in the manuscript at least a discussion on the variables available, their usefulness for modelled dust evaluation and their limitations.

We moved the supplementary sections to Section 3 in the revised main text and added more descriptions accordingly following this suggestion.

Added subsections in Section 3 include (please see contents of each of these subsections in the revised manuscript):

“*3.1 Surface dust concentrations and dust aerosol optical depth from AERONET*”,

“*3.2 Surface dust deposition fluxes*”,

"*3.3 Size distributions of dust aerosol*",

"*3.4 The direct radiative effect efficiency of dust*",

"*3.5 Other datasets*",

and, a section to describe the metrics used for model assessment

"*4 Model assessment metrics*".

We also oriented the readers to the discussion section 7 for in-common limitations before Section 3.1:

"*Due to limitations in precisely matching the period and locations between model results and data, the evaluations focus on checking if models can capture overall features of the measured/observed/retrieved dust cycle and the corresponding dust DRE efficiency. We summarize limitations going beyond this mismatch on period and location and common in all the model-data comparisons in Sect. 7.*"

In order to lighten up the contents of the paper, I would recommend splitting the results in two different articles, one focusing on the current developments and their impact on the bulk dust cycle, and another focusing on those improvements that potentially have an impact on the mineralogy (e.g. the changes on the emission scheme).

We had tried to do this but found that having the potential impacts on the mineralogy by changing to the new dust emission scheme is not enough for a separate paper. Instead, we added a new section briefly documenting results from the MINE runs, such that it makes more senses to have both BULK and MINE runs in this article.

"***6. Bulk- versus speciated-dust model***

*The bulk (CAM6.α) and dust-speciated model (CAM6.α_MINE) simulate a similar dust cycle with the difference between the two types of models orders smaller than those simulated by the former (Fig. 12 and 13). This similarity results from the dust tuning toward the global mean DOD of 0.03, the same meteorology dynamics both models were nudged toward, and the design of the dust-speciated model that summing the mass fraction of each dust species equals unity. With the same reasons, the influence of each of the modifications on the modelled dust cycle quantified using the bulk model instead of the dust-speciated model, as this study used, would be similarly comparable. The modelled dust optical properties, however, differs remarkably (i.e.,* *dust SSA; Table 6) with the simulated global mean dust SSA by CAM6.α_MINE (0.896) lower than by CAM6.α (0.911) at the visible band centered at 0.53 μm. Note the dust DRE is sensitive to variation of the dust SSA. This lower dust SSA obtained here in the dust speciated model than in the bulk dust model is consistent to the finding of a previous study (Scanza et al., 2015) using an early model version (CAM5). Correspondingly, CAM6.α_MINE yields a reduced dust cooling (Table 6) and DRE efficiency (Fig. 9) than CAM6.α. For dust DRE efficiency (Fig. 9), speciating dust in CAM6 tends to reduce the RMSE while retaining the horizontally spatial correlation in either SW (CAM6.α: RMSE=11; R=0.26 versus CAM6.α_MINE: RMSE=10; R=0.20) or longwave (CAM6.α: RMSE=4; R=0.86 versus CAM6.α_MINE: RMSE=3; R=0.84) or both spectral ranges (CAM6.α: RMSE=7; R=0.93 versus CAM6.α_MINE: RMSE=6; R=0.92). This comparison suggests that modeling dust as component minerals with the dust size distribution in coarse mode of MINE_NEW_EMIS_SIZE helps improve the model performance relative to modeling dust as a bulk to reproduce the retrieved dust DRE efficiency (Fig. 9a). The improvement, however, could be artificial because of the combined use of imaginary complex refractive index of hematite volume (see Fig. 1b of Li et al., 2021) and the volume mixing used in the dust speciated model to compute the bulk-dust complex refractive index (Li et al. in prep.), leading to artificially more absorptive dust than in the bulk dust model (Fig. 9a and Table 6).*"

Finally, I would recommend modifying the organization of some of the contents, and re-writing or improving some parts of the text. Also, in some sections, the authors rely excessively on external references, making it difficult to follow the discussion with the information provided in the paper itself. My recommendation would be to restructure or adapt the article contents, such that:

(1) the previous status of the model is clearly defined and the motivation to improve or change the specific dust representation is justified.

We slightly re-structured the introduction, but did not add more contents, since the Reviewer #2 also thinks the introduction is highly readable (please see their general comment: Line 19). Please see our detailed response below (Line 320-373).

(2) the new developments are described in the current paper in a comprehensive manner (i.e. not trusting excessively on the reader to go and check the external references).

We introduced briefly the key formulas used in the parameterizations, so, the readers now do not have to check those references.

(3) the evaluation methodology is explained before the presentation of results, for instance adapting current section 3. It would be particularly useful to identify the multiple metrics that are going to be used for the model evaluation and their purpose (i.e. regional variability, temporal variability, etc.), comment on the dust tuning methodology and its impact on the evaluation metrics (if any), as well as to merge the description of the observations with the

comments on section 5 about the limitations of the datasets. Section 5 could be kept to provide an overall assessment of the observations limitations on the main conclusions of the article.

We added a new section briefly describing the metrics used to assess the model performance and we keep the original Section 5 (#7 in the revised manuscript) as it was but orient the readers to it in this section before Section 3.1:

"*Due to limitations in precisely matching the period and locations between model results and data, the evaluations focus on checking if models can capture overall features of the measured/observed/retrieved dust cycle and the corresponding dust DRE efficiency. We summarize limitations going beyond this mismatch on period and location and common in all the model-data comparisons in Sect. 7.*"

The new section reads as:

**"4 Model assessment metrics**

Metrics used to evaluate the model performance against observations include the root mean square error (RMSE) and correlation efficient (Kendall's τ or Spearman's Correlation). Both the Kendall's τ and Spearman's Correlation are non-parametric methods which do not require a distribution of the data, such as Gaussian or normal. For dust deposition, loadings correlations calculated are to assess how well models reproduce both their regional climatology mean or one-time observation and the seasonal cycles. However, because of a lack of reliable monthly data, assessments for the dust DRE efficiency, DOD from Rideley et al. (2016), and percentages of wet deposition in the total deposition are on spatial variability based on the regional climatology mean or one-time observations. We tested the correlation significance of the metrics at the statistical confidence level of 95%. For the dust DRE efficiency and percentages of wet deposition, some domains only have a range available, such as, Sahara Desert (15º-30ºN, 10ºW-30ºE) in the longwave spectral range. For those domains, a mean of the low and high boundaries of the range is used in the calculation of the Spearman's Correlation and the corresponding significance test."

Comments on the dust tuning methodology are now given in the "Experiment design" section, such as:

"by modifying a CAM namelist variable, dust_emis_fact, such that the simulated global mean DOD is ~0.030 at the visible band…".

Values for the tuning parameters are given in the revised Table 2.

I believe that with these changes, the article would be much easier to follow and it would reach a broader audience.

Thanks for the constructive suggestions!

**Introduction**

I believe this section could be slightly re-structured, particularly to better clarify the current model status, justify the need for improvement in the specific aspects that are dealt with in this work, and briefly explain how these are going to be approached.

We re-structured the introduction to reflect these excellent suggestions:

"*As one of the widely used climate models, the Community Atmosphere Model (CAM) contains several weaknesses of modeling the dust cycle. For example,*

*1) the default scheme in CAM6.1 (Zender et al., 2003; Dust Entrainment And Deposition DEAD model, referred as DEAD) relies on an empirical geomorphic dust source function, created based on satellite retrievals of dust source regions, to model dust emissions;*

*2) the current default CESM2.1 is using the dry deposition scheme Zhang et al. (2001; Z01 hereafter) developed for particle deposition over smooth and non-vegetated surfaces. This scheme, however, underemphasizes the interception loss, the mechanism of which is less influential over the other surfaces such as grassland. The use of the Z01 in the current default CESM2.1 is, thus, very likely overestimating the dry deposition velocity of fine-sized aerosols (diameter < 1.0 μm; referring to the geometric diameter herein unless stated otherwise) and slightly underestimating that of coarse-sized aerosols (diameter > 5.0μm) (Wu et al., 2018), especially over non-vegetated surfaces (Petroff and Zhang et al., 2010);*

*3) one of the changes from CAM5 to CAM6.1 was replacing the size distribution of aerosols in the coarse mode in CAM5 with the one that has a much narrower width in CAM6.1 (Table 1). This change was to accommodate stratospheric aerosols in the coarse mode (e.g., volcanic sulfate) compared to an early officially released version of this model (Mills et al., 2016). A recent model evaluation against satellite retrievals (Wu et al., 2020) suggest that CESM2.1-CAM6.1 worsened the dust cycle representation and stands out in simulating the relative importance of wet to dry deposition, compared with the other global climate models or model versions, such as CESM1-CAM5, due partially to the narrow coarse geometric standard deviation;*

*4) dust aerosol are typically aspherical particles in shape. The dust asphericity could lengthen the dust lifetime by ~20% compared to modeling dust as spherical particles (Huang et al., 2020).*

*Still, CAM6.1 simulates dust as spherical particles, though the impact of dust asphericity on optical depth and resulting radiative effect of dust (Kok et al., 2017) has been previously introduced to CAM6.1 (Li et al., 2021).*

*Correspondingly, this paper describes several updates to the dust representation in CAM6.1 on the four aspects and evaluates whether and for what conditions they improve the dust model comparison to observations in the present climate. Specifically, we*

*1) replace DEAD with a new more physically based dust emission scheme, Kok et al., (2014a) previously developed for the climate models within the framework of DEAD. This scheme performs well against observations in CESM-CAM4 (Kok et al., 2014b) without the aid of the empirical geomorphic dust source function;*

*2) replace Z01 by the dry deposition scheme Petroff and Zhang et al., (2010) developed (PZ10 hereafter) to mediate the overestimation of the dry deposition velocity of fine-sized aerosols;*

*3) revert size distribution of dust aerosol particles in the coarse mode to the one previously employed in CAM5;*

*4) account for the lifetime effect of dust asphericity by decreasing the modeled gravitational settling velocity.*

*These updates are based on up-to-date knowledge of the dust cycle and are thus more physically realistic than the default dust parameterizations in CAM6.1/Community Land Model (version 5; CLM5).*"

**2. Model descriptions**

I would recommend starting by describing the aerosol representation in CAM6.1, as it affects both bulk dust, speciated dust and other aerosols simulated in the model.

Excellent suggestion. Per this specific comment, in the revised manuscript, we created a new section titled as "*Aerosol representation*" ahead of Section 2.2 "Bulk dust modeling" and moved text relevant to the general aerosol representation from the "Bulk dust modeling" section to the new section.

Please, see my general comment above. Which is the added value of conducting two set of simulations (with bulk and speciated dust) for the purpose of this article (assessing changes due to deposition, emission, size distribution and shape)? If this is not justified, I would focus on this article in the bulk dust experiments, and present the speciated dust experiments elsewhere.

Thanks for this question. Please see our response to that general comment on the experiment design (Line 41-135).

### 2.3. Dust optical properties and radiation flux diagnostic

Please, take advantage of this section to explain aspects related to the calculation of optical properties and/or radiative variables that are currently explained in the results section (see my comments below on sections 4.3 onwards.

Done. We moved up text from Section 4.3 (5.3 in the revised version): "We augmented the longwave radiative effect from the model by 51% to account for the dust scattering (Dufresne et al., 2002)", and defined the DRE efficiency in this section: "*The DRE efficiency, which we used to evaluate the model performance on simulating the dust optical properties, is defined as the ratio of dust DRE to dust optical depth (DOD) under clear conditions*".

### 2.4.2. Dry deposition schemes

The original dry deposition scheme is partly described here and partly in the introduction. I would use this section to describe the details on both the previous and the new proposed scheme. At least, I would include here the references to both schemes, and clarify if the empirical coefficients are updated in the new scheme.

Added the reference to the default Z01 scheme. These two schemes greatly differ from each other. For example, PZ10 considers additional processes that are not in Z01, such as the turbulent impaction and vice versa, and accounts for more morphological characteristics of the canopy than Z01. Even for processes described in both schemes, the parameterizations are very different, such as the aerodynamic resistance (See Equation 4 of Petroff and Zhang, 2010 vs Equation 4 of Zhang et al., 2001) and Brownian diffusion (See Equation 4 of Petroff and Zhang, 2010 vs Equation 6 of Zhang et al., 2001). Consequently, these two schemes are employing two different sets of empirical coefficients.

We now provide key formulas for both parametrizations in the revised text including and descriptions of the coefficients, such that the readers do not have to check external references. Please see "2.5.2 Dry deposition schemes".

### 2.4.3. Dust asphericity

Being this one of the developments listed in the article, it would be worth to include in this section at least the main characteristics of the development (e.g. factor varying according to the source region, and ranging from X to X).

We thank the reviewer for their help in improving the readability of the manuscript. Although such information was presented in the supplementary, to make it clearer, we moved some text to this section and made a revision as follows.

"*In this calculation, we assume that the dust shape parameters are independent of the size of dust aerosol particles. Therefore, a constant revision of the dust gravitational settling velocity (the calculated value in the model by default is for spherical aerosols) due to dust asphericity by multiplying the velocity by γ was applied to dust species in the three modes that contains dust aerosol (Aitken, accumulation, and coarse). The size independency assumption of dust asphericity follows the recent observational evidence that there does not exist a statistically significant relationship between the shape parameters (aspect ratio and height-to-width ratio) and dust sizes (Huang et al., 2020). Because of highly limited measurements of dust shape parameters, we subjectively divided the dust coverage into "close-to-source", "short-range", and "long-range" zones and calculated the asphericity factor γ for each of the zones, the global map of which is shown in Fig. S1, ranging between 0.82 and 0.93. We acknowledge limitation of the methodology here to account for the lifetime effect of dust asphericity, anticipating improvements on modeling this effect when more high-quality dust shape measurements available.*"

Also, similarly, we now provide key formulas used in our calculations in the revised text. Please see "2.5.3 Dust asphericity".

Also, the authors mention the impact of the dust asphericity on optical properties (line 119). In section 4.2.3, they state that CESM2 does not include the enhancement in mass extinction efficiency due to asphericity, but that it is considered in this study (section 4.2.3). I believe the approach used to consider asphericity in the mass extinction efficiency should be clarified and described in this section (2.4.3).

We moved relevant text from the result Section 4.2.3 (5.2.3 in the new version) to Section 2.4.3 (2.5.3 in the new version).

We also added the following in the "Experiment design" section for clarity:

"*The enhancement of the mass extinction efficiency of aerosol particles by dust asphericity is included in all the simulations, since we do not attempt to quantify how this enhancement impacts the simulated dust cycle.*"

## 2.5. Experiment design

Please, see my general comments related to the experiments' design.

I would recommend to describe first the common model configuration amongst experiments (i.e. configuration of the model components, spatial resolution, period simulated, etc.), and then identify the experiments designed to test the different developments.

Reordered the description to reflect this suggestion.

## 4. Results

Please, review and re-structure this section, see my general comment above.

We think that either doing what the reviewer suggested or keeping as what it was would be fine. In the drafted manuscript, we had tried doing the same as the reviewer suggested but then reordered the result section taking the "principle" that "the most important goes first", since the manuscript is lengthy. In any order, the conclusions of this article would remain unchanged.

I believe using the same set of experiments to discuss all the modifications (either bulk or speciated dust) would help.

Please see our response to the comment on BULK versus MINE runs (Line 41-135).

In addition, a discussion focusing on the different variables, combining the multiple datasets used as a reference, rather than a separate explanation for each comparison could be of benefit. Another strategy to make easier the discussion for the reader, could be to "qualify" the sites / observations by their characteristic trait when explaining the details, e.g. source region, remote station, etc., rather than leaving it to the reader to figure out where the site is or its characteristics.

All the variables share some shortcomings in common. That is why we have a separate section "7 Limitation in the model-observation comparison" to discuss the model-data comparison. To reflect the suggestion and to make the discussion in the result sections lighter, we described more the variables in the section "3 Observational datasets for model evaluations".

Added subsections in Section 3 include (please see contents of each of these subsections in the revised manuscript):

"*3.1 Surface dust concentrations and dust aerosol optical depth from AERONET*", "*3.2 Surface dust deposition fluxes*",

"*3.3 Size distributions of dust aerosol*",

"*3.4 The direct radiative effect efficiency of dust*",

"*3.5 Other datasets*",

and, a section to describe the metrics used for model assessment

"*4 Model assessment metrics*".

We also oriented the readers to the discussion section 7 for in-common limitations before Section 3.1:

"*Due to limitations in precisely matching the period and locations between model results and data, the evaluations focus on checking if models can capture overall features of the measured/observed/retrieved dust cycle and the corresponding dust DRE efficiency. We summarize limitations going beyond this mismatch on period and location and common in all the model-data comparisons in Sect. 7.*"

**4.1.1. Dust emissions**

Why compare the total emission burden with model estimates that go beyond CAM6.1 simulated size range? I believe it would be useful to include comparisons with models that use the same range (e.g. some of the AEROCOM phase I models, Huneeus et al. 2011).

Good point, though not all models participated in AEROCOM use the same size range. As we pointed out the different size range between ours and that of Kok et al. (2021a), it would be fine to keep this small signpost: the estimate of Kok et al. (2021a).

The revised sentence reads as:

"To achieve the global mean DOD of ~0.03, CAM6.α requires a dust emission of 2891 Tg $a^{-1}$ (Table 6), which falls below the estimate of 3400-9100 Tg $a^{-1}$ by Kok et al. (2021a; their Table 1) that accounts for dust between 0.1-20 μm in diameter *and above the median, 1123 Tg $a^{-1}$, reported in AEROCOM phase I (Huneeus et al., 2011).*"

**4.1.2. Climatology annual means of [...]**

The discussion here will greatly benefit from a previous definition of the statistics, metrics, and evaluation, which I would suggest including in Section 3. In that way, the authors could make the discussion in this section lighter.

Good point. A definition of these is now included in Section "4 Model assessment metrics".

"*Metrics used to evaluate the model performance against observations include the root mean square error (RMSE) and correlation efficient (Kendall's τ or Spearman's Correlation). Both the Kendall's τ and Spearman's Correlation are non-parametric methods which do not require a distribution of the data, such as Gaussian or normal. For dust deposition, loadings correlations calculated are to assess how well models reproduce both their regional climatology mean or one-time observation and the seasonal cycles. But because of a lack of reliable monthly data, assessments for the dust DRE efficiency, DOD from Rideley et al. (2016), and percentages of wet deposition in the total deposition are on spatial variability based on the regional climatology mean or one-time observations. We tested the correlation significance of the metrics at the statistical confidence level of 95%. For the dust DRE efficiency and percentages of wet deposition, some domains only have a range available, such as, Sahara Desert (15º-30ºN, 10ºW-30ºE) in the longwave spectral range. For those domains, a mean of the low and high boundaries of the range is used in the calculation of the Spearman's Correlation and the corresponding significance test.*"

The authors mention the tuning as a factor affecting the comparison of modelled DOD to MODIS and Ridley et al. (2016) products; however, this is not taken into consideration when AERONET information is used as a target. Could the tuning also have an effect on those results?

Good point. The reviewer is right. We added the following in this pagraph.

"*This overestimated DOD in the model near the source regions resulting from the tuning method may also partly explain the imperfect match between the modeled and AERONET-based DOD (Fig. 1a).*"

Does the dust wet vs. dry deposition balance in their model change with the improvements on size distribution? Could this partly be explained by an overestimation of the finer dust fractions? Or is the representation of modal internal mixtures more relevant to this process?

We added two more columns showing results from the size tests and the following sentence. To better quantify the assessment, we also added RMSE and correlation efficient for each simulation shown in the revised Table 7.

“*The models tend to overestimate the observed percentages of the wet deposition (Table 7). This overestimation could be due partly to the internal mixing assumption of dust aerosol with sea salts which increases hygroscopicity of the aerosol mixture during transport. Correcting the coarse-mode distribution, following we suggest (Table 1), does not help improve the model performance (Table 7).*”

“Table 7. Percentage (%) of wet deposition. Observations compiled by Mahowald et al., (2011b) from data at Bermuda (Jickells et al., 1998), Amsterdam Island, Cape Ferrat, Enewetak Atoll (R.Arimoto et al., 1985), Samoa; New Zealand sites (Arimoto et al., 1990); North Pacific sites (Uematsu et al., 1985); Greenland Dye 3 (Hillamo et al., 1993), Coastal Antarctica (Wagenbach et al., 1998), and Dome C of Antarctica (Wolff et al., 2006). *RMSE: root mean square error; R: Spearman’s Correlation.”*

| Location | CAM6.1 *[RMSE=39; R=-0.38]* | *NEW_EMIS [RMSE=39; R=-0.52]* | *NEW_EMIS_SIZE [RMSE=37; R=-0.63]* | CAM6.α *[RMSE=37; R=-0.31]* | MINE_BASE *[RMSE=34; R=-0.45]* | MINE_NEW_EMIS *[RMSE=35; R=-0.29]* | CAM6.α_MINE *[RMSE=36; R=-0.38]* | Observations |
|---|---|---|---|---|---|---|---|---|
| Bermuda [32ºN, 65ºW] | 92 | *91* | *81* | 87 | 81 | 85 | 87 | 17-70 |
| Amsterdam Island [38ºS, 78ºE] | 88 | *88* | *73* | 81 | 78 | 80 | 83 | 35-53 |
| Cape Ferrat [43ºN, 7ºE] | 92 | *94* | *89* | 86 | 87 | 84 | 86 | 35 |
| Enewetak Atoll [12ºN, 162ºE] | 79 | *73* | *52* | 66 | 58 | 56 | 64 | 83 |
| Samoa [14ºS, 152ºW] | 91 | *91* | *83* | 86 | 83 | 81 | 85 | 83 |
| New Zealand [35ºN, 173ºE] | 89 | *92* | *82* | 87 | 80 | 85 | 88 | 53 |
| North Pacific[a] [4º-28ºE, 162º-158ºW] | 62-90 | *71-91* | *48-80* | 53-85 | 46-80 | 48-80 | 56-84 | 75-85 |
| Greenland [65ºN, 44ºE] | 82 | *87* | *82* | 86 | 75 | 86 | 84 | 65-80 |
| Coastal Antarctica [76ºN, 25ºW] | 96 | *92* | *68* | 93 | 82 | 87 | 88 | 90 |
| Dome C. Antarctica[b] [75ºN, 123ºE] | 97 | *97* | *95* | 96 | 88 | 89 | 91 | 20[b] |

a shown are minimum and maximum of the annual wet percent among the four sites
b Non sea salt-sulfate

**4.1.4. Size distribution of transported dust**

Why is the comparison with AERONET presented in the supplement?

Thanks for the comment. We moved the figure in the supplement to the main text as Figure 5:

“*Figure 5. Modelled and observed atmospheric size-resolved dust mass in the geometric diameter range of 1-10 μm at AERONET stations. Numbers in each plot indicate the Kendall's τ coefficient between model and observations (blue bars). The model runs here include the one using the old model with the mode size parameters from CAM6 by default (CAM6.1 in cyan) and the other one using the new model with the mode size parameters from CAM5 (CAM6.α in black). Both runs were using the offline dynamics.*”

**4.2.1. Dust emission schemes**

Please, avoid relying on excessively on external references to explain features observed among the experiments (e.g. lines 561 to 563), summarize them directly here.

We provided key formulas for the new and old parameterizations in the revised text (please see Section 2.5) and cited them accordingly here instead of relying on external references, such as:

"…the dust emission coefficient in BRIFT (*Eq. 10*) and the new method of calculating the threshold gravimetric water content of the topsoil layer (*Eq. 9; see values for the tuning parameter "b" in Table 2*) shifts the main dust emission in…"

What is the impact of the dust tuning on the results? According to section 2.5, both EXP06 (MINE_BASE) and EXP07 (MINE_EMIS) were tuned to match a global DOD of around 0.03. Was that not the case? What does the re-scaling of the DOD mentioned on line 591 refer to?

The dust tuning, via a namelist variable, is to ensure that the global mean of the simulated DOD equals 0.030, which is one of the "best" estimate of the global dust quantities. The dust emission shown in this section is required in the model with different dust emission schemes to reach that criterion. To make this clearer and the discussion here lighter, we added the following in the revised "Experiment design" section.

"*We prefer to tuning the model to reproduce the global mean DOD, 0.030, because DOD is currently the best estimate of global dust quantities, compared to the others (i.e., dust concentrations). It turns out that doing so can also reasonably reproduce the other quantities with no need of a regional tuning. MINE_NEW_EMIS requires the dust tuning to use a much larger tuning parameter (dust_emis_fact=3.6; Table 2), than MINE_BASE (dust_emis_fact=1.6), because, otherwise, if using the same dust_emis_fact as in DEAD, the dust emissions in BRIFT would lead to an unrealistically high global mean DOD (>~0.5).*"

On Line 591 (original manuscript), the global DOD in BRIFT is lower than in DEAD (0.035 versus 0.029), because we did not retune the model to have the global DOD equal exactly 0.030. So, if re-scaling the dust deposition and loadings such that both global DOD equal exactly 0.030, then the difference of the dust deposition and loadings between the two experiments would be smaller.

To make this clearer, we revised the sentence a little bit:

"*…differences between the global annual mean dust deposition in BRIFT and DEAD would become smaller, if we rescaled the global annual mean dust deposition and loadings offline using factors to make the global mean DOD in the two experiments exactly equal 0.030.*"

**4.2.3. Dust asphericity**

The authors state that the dust asphericity could mediate the overestimated dust emission from source regions, is this shown in their experiments?

No, this is not directly shown in the experiment, but the result is indicative of the probably mediated effect. Since the dust tuning is to have global mean DOD ~0.03, introducing the lifetime effect of dust asphericity to the model is expected to have the potential to reduce the dust emission level. We added the following text to explain it a little bit.

"…dust asphericity could potentially mediate the overestimated dust emission from source regions (e.g., North Africa)*, because dust asphericity could enlengthen the lifetime in the atmosphere and thus it takes less amount of dust to have same amount of dust loadings and DOD as spherical shape assumption does.*"

Does the asphericity factor affect differently fine vs coarse particles?

No, the asphericity factor is the same over the three modes. This is based on the finding of Huang et al. (2020) that there is no statistically significant dependence of dust asphericity on the dust size. To clarify this, we revised a sentence in the "Dust asphericity" section:

"… *the asphericity factor γ (defined as the ratio of the gravitational settling velocity of aspherical dust to that of spherical dust) offline, which is independent of the dust size, based on* …"

**4.2.4. Dust size representation**

This section is difficult to follow, please, revise.

Please see the revised Section 5.2.4 below:

"The removal rates of dust aerosol particles by both dry and wet deposition highly depends on their size (Mahowald et al., 2014). Since most of dust loadings are in the coarse mode, changing parameters of the coarse-mode size distribution (σ, initialized GMD, and the prescribed minimal and maximum boundaries within which the modeled GMD can vary, Table 1) from σ=1.2 to 1.8 halved the lifetime of dust (lifetime=4.9 days versus 2.4 days; Table 6). This reduced dust lifetime is primarily due to the change in σ of the coarse mode (Fig. 8b) rather than the initialized GMD and its boundaries, as we obtained almost the same dust lifetime (~2.4 days) between *experiments with different parameters for dust size distribution but identical σ=1.8* (NEW_EMIS_SIZE versus NEW_EMIS_SIZE_WIDTH; Table 6). *We also notice a different DOD simulated by NEW_EMIS_SIZE (DOD=0.013) and NEW_ EMIS_SIZE_WIDTH (DOD=0.019). The prescribed GMD boundaries does not affect the simulated dust loadings and DOD, because* *the predicted GMD in the model varies little. We can, therefore, derive that the initialized GMD itself, is also relevant to simulated DOD, but its influence (absolute relative change=20%) is second to that of changing the coarse-mode σ. Thus, it is the increased σ of the coarse mode that explains the reduced dust loadings (22 versus 11 Tg in NEW_EMIS and NEW_EMIS_SIZE, respectively; Table 6; Fig. 8b) and DOD (0.030 versus 0.013 Tg in NEW_EMIS and NEW_EMIS_SIZE, respectively; Table 6). This impact of changing the coarse-mode σ is also greater than that of the other modifications (e.g., speciating dust or changing the dust emission scheme from DEAD to BRIFT) on the simulated dust lifetime which appears trivial (e.g., dust lifetime increased by 0.6 days only by changing to the new emission scheme). Correspondingly, given a similar emission rate, changing the coarse-mode σ affects DOD most, compared to the other modifications we made.*"

**4.3. Dust direct ratiative effect.**

Details such as the LW increase by 51% could be explained in section 2.3. I would only mention this again here if the approach used in the different experiments would differ, and thus affect the comparison.

Mentioned it there now.

"*We augmented the longwave radiative effect from the model by 51% to account for the dust scattering (Dufresne et al., 2002).*"

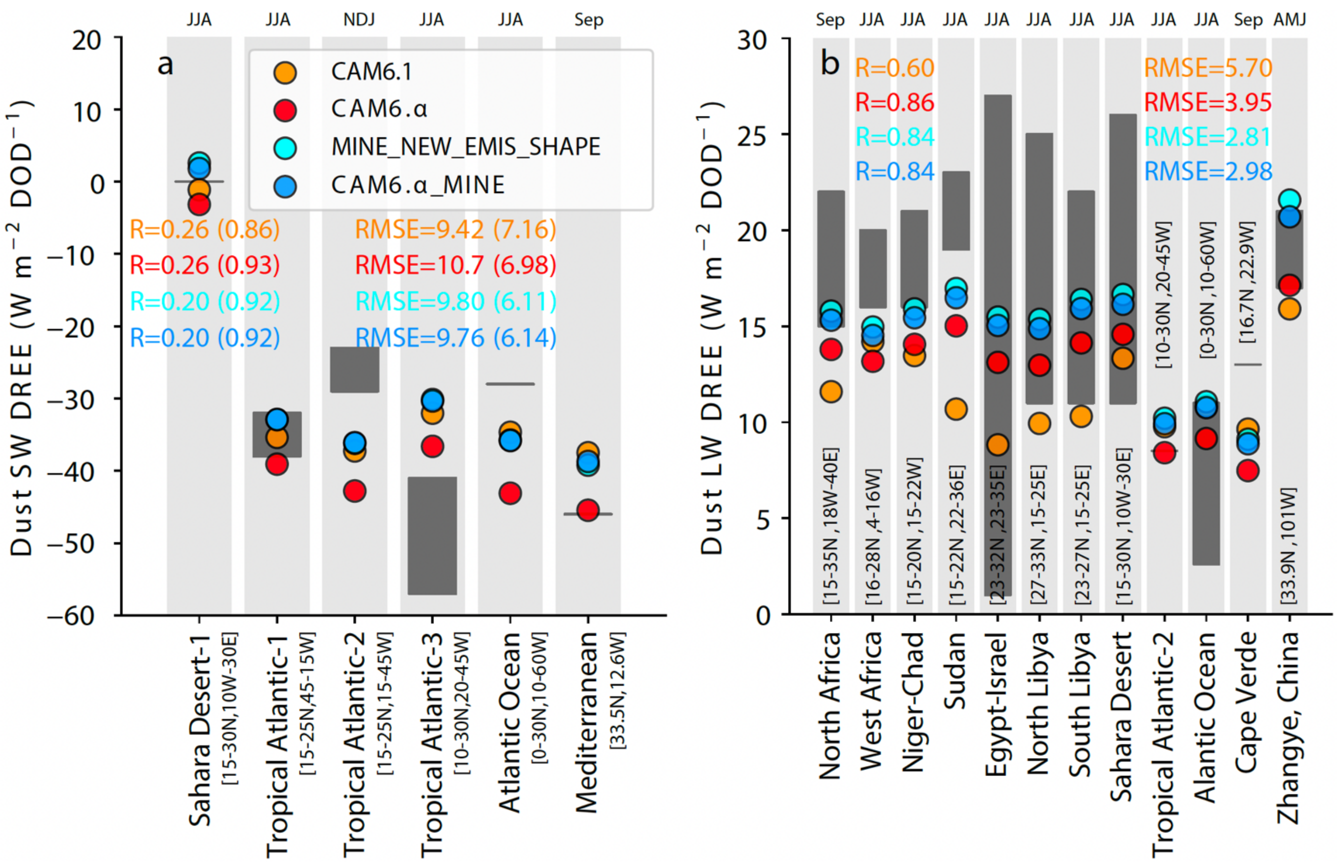
**4.3.1. Dust direct radiative effect efficiency.**

Please, use also section 2.3 to define the net DRE efficiency.

Done: "*The DRE efficiency, which we used to evaluate the model performance on simulating the dust optical properties, is defined as the ratio of dust DRE to dust optical depth (DOD) under clear conditions.*"

What is the metric used here to define the model performance?

There are only several points included in this comparison, and, for some, only ranges are provided in the corresponding reference. So, we had not used any statistic metric to measure the distance between model and observations. But as a response, we included the correlation coefficient and RMSE with the assumption made for points where there is only a range that the mean could be used in the calculations.




"Figure 8. Modelled and observed dust direct radiative effect efficiency in the shortwave (SW)/longwave (LW) spectral ranges under clear conditions at the TOA over the sub-domains (shown in the inserted map and location described below) in summer, fall, and September for the 2000s climate. The radiative effect efficiency is defined as the ratio of the radiative effect to DOD, so has units of W m-2 τ-1. Included cases from left are CAM6.1, CAM6-α, MINE_NEW_EMIS_SHAPE, CAM6.α _MINE. *The field value/range are from references listed in Table 3. Colored numbers show correlation coefficient (R) and the root mean square error (RMSE) between the model and retrievals in the SW (a) / LW (b) spectral ranges or in both spectral ranges (numbers in parenthesis in Panel a).*"

The difference between the experiments with speciated and bulk dust is not exclusively dependent on the developments presented here, but, as the authors mention, attributed to the resulting optical properties for the different representation on the dust.

Please see our response (Line 41-135) to the general comment on bulk dust versus dust-speciated model. We added a new section ("6. Bulk- versus speciated-dust model": see Line 142-169 above) to compare results from the two types of models.

Does the model diagnose all sky or clear sky DRE (line 730)? Please, clarify this in section 2.3.

We had stated that this is DRE under all sky condition in that section: “The direct radiative effect by dust aerosols is then determined by calculating the difference of the net radiative flux with and without dust at the top of the atmosphere *under all-sky conditions*”.

To make it clearer, we revised this sentence a little bit, so now it reads as

“The direct radiative effect by dust aerosols *under all-sky conditions (here and hereafter unless stated otherwise)* is then determined by calculating the difference of the net radiative flux with and without dust at the top of the atmosphere under all-sky conditions”.

**Conclusions**

The authors mention the effect of dust asphericity on mass extinction efficiency as one of the aspects that produces a larger change in the results, as mentioned above, it is unclear to the reader which is the approach followed to introduce this in the model and/or if it’s introduced at all.

As stated in the original manuscript, all simulations here, including the base CAM6.1 and MINE_BASE, have considered the enhancement of dust asphericity on the mass extinction efficiency. Such effect has also been well documented previously. Thus, we do not aim at investigating it in this study. To avoid possible confusion, we remove relevant statements in the conclusion and added a sentence in the “Experiment design” section:

“*The enhancement of the mass extinction efficiency of aerosol particles by dust asphericity is included in all the simulations, since we do not attempt to quantify how this enhancement impacts the simulated dust cycle.*”

I believe it would be useful to include a brief discussion on the implications of reverting the standard deviation changes in the coarse mode for the stratospheric aerosols. If the change was initially introduced to better accommodate those, which would be the recommendation of the authors for the model version to be issued?

The solution could be to have a coarse mode for dust separate from the stratospheric aerosols. Then, for this separate coarse dust mode the model developer could use the broad standard deviation.

We revised relevant contents as the following in response to this suggestion.

“*Our analysis suggests that the defaulted 1.2 for the geometric standard deviation of the transported dust size distribution (coarse mode) may be too narrow to simulate the dust lifetime.*

*In the next released model version, we recommend reverting the geometric standard deviation to 1.8, as in CAM5, which may require a split of representation of dust and the stratospheric aerosols.*"

The authors comment on potential ways of improving further the dust cycle, however, it is unclear for the reader if those stem from the work performed in this article. I would recommend to highlight the weaknesses detected in this study concerning the dust cycle representation (even after all the improvements included), and link to the appropriate suggested next step to solve that issue.

We revised this part a little bit to connect it with what we present in previous sections:

"1) *for the dust emission parameterization,* the threshold friction velocity calculated in both BRIFT and DEAD does not account for…"

"2) …in the northern high-latitude regions *(Sect. 5.1.1)*,…"

and added more text:

"*3) comparisons with the constrained global dust size distribution and measurements downwind of North Africa suggest that the model underestimates dust aerosols in the coarse mode with the geometric diameter > 5 μm and misses aerosol particles with the geometric diameter > 10 μm (Fig. 6). The former happens may be due to an underestimate of dust aerosol particles in that size range upon emissions and/or the removal rate of those particles being too high during transport in the model (Adebiyi and Kok, 2020b), the reason for which is still under exploration. For the latter, extending the dust size range to include particles with the geometric diameter > 10 μm in CAM6 is a worthy endeavor, such as in Ke et al. (2022).*

*4) as noted in (Wu et al., 2018), some of the variables in the dry deposition parameterizations could vary in different seasons for certain land cover and use types, such as the roughness length, $Z_0$, in Z01 and the displacement height of the canopy, h, in PZ10, for which a fixed climatological mean is used in the models. How accounting for the seasonal variation of those variables in the model can affect the dust cycle modeling deserves further exploration.*

*5) compared to bulk dust, modeling dust aerosol as component minerals could potentially help better reproduce the observed spatiotemporal variability of dust optical properties and thus the dust DRE efficiency (Fig. 9), while retaining the accuracy of modeling the dust cycle with the offline dynamics in present day. But the current atlas of soil mineralogy and the optical properties of key minerals (i.e., iron oxides) contain large uncertainties which should be better quantified in* *the future, such as that planned in the Earth Surface Mineral Dust Source Investigation (EMIT) and in our ongoing work (Li et al., in prep), respectively.*"

**Technical corrections**

Please, find below a list of technical corrections that could be applied to the current manuscript version.

Thanks a lot for these technical corrections. We made corresponding changes in the revised manuscript.

L19. Either refer to the CAM6 model in the abstract (as it is in the article title) or change the title to include the CESM model.

We mentioned the CAM6 model in the abstract.

"*The Community Atmosphere Model (CAM6.1), embedded in the Community Earth System Model (CESM; version 2.1), simulates the lifecycle (emission, transport, and deposition) of mineral dust…*"

L23-24. If possible, outline the main changes included in the different parameterizations (emission, dry deposition, size distribution and dust particle shape).

We mentioned these in the revised text.

L26-27. Is it the effect of the size distribution change as large as the change in the dust emission scheme?

Great point. It can be larger than changing to use the new dust emission scheme, for dust lifetime, burden, and DOD, for instance (see Fig. 10), no matter if we retune the model to have the simulated global dust AOD ~0.3. We added some words to reflect this comparison.

"In comparison, *the other modifications induced small changes to the modeled dust cycle and model-observation comparisons, except the size distribution of dust in the coarse mode, which can be even more influential than that of replacing the dust emission scheme*."

L46. Is shape also a factor affecting the uncertainty in dust direct radiative effect?

The primary influence of shape is on the dust asymmetry factor and extinction. In our global model, we tune dust emissions to a level at which the global mean DOD is around 0.03. Since, the direct radiative effect roughly linearly depends on DOD, the irregular shape of dust particles would not impose influence as big as those we stated on this line. To make that statement more scientifically rigorous, we slightly revised this sentence saying those are primary factors.

"*These uncertainties in the dust cycle modeling, as well as uncertainties in optical properties due primarily to dust size and mineral composition*…"

L63-64. Is it necessary to mention the previous CAM and CESM versions?

In the revised manuscript, we deleted this paragraph.

L71. Why do the authors mention now the Community Land Model version 5 (CAM6.1/CLM5)? Please, use the same acronym/naming convention all along the article, either CAM6.1 or CAM6.1/CLM5, or at least, mention the full name the first time it appears and explain that from then on it will be referenced as CAM6.1.

Thanks for questioning this. CAM6.1 refers to the atmosphere component only of CESM, while CLM5 refers to the land component. Correspondingly, when mentioning solely CAM6.1, it means only modifications to the atmosphere component, for example, to incorporate the new dry deposition scheme. Incorporating the new dust emission scheme requires modifications to both components, therefore, we need to mention both CAM6.1 and CLM5 to be scientifically rigorous.

L102 (Table 1 caption): MAM4 is mentioned for the first time. Why use two abbreviations for the standard deviation, remove extra dot after CAM6.1 in L103.

Thanks! In the revised version, we spelled MAM4 out, deleted GSD, and removed the extra dot after CAM6.1 on that line.

L108: Homogenize the naming of the sections, either Sect. or Section.

Well. This seems a requirement by the journal: when beginning with the word, Section, one should use the full name, but in a sentence, one should use Sect. to refer.

L109: I would substitute semi-observation by more specific term(s).

We now specify both the observation and semi-observation as "*measurements, retrievals, and model-observation integration*" which brackets all the data used in this work.

L117: Is it CESM2.1 or CESM2? Please, keep consistency in the naming of the model versions along the document.

We now use CESM2.1 all through the manuscript.

L125: Why is the iron solubility mentioned here?

This is redundant information, so deleted. In the original version, we also included modeling of iron from dust, fire, and so on, but we had decided to delete them from this manuscript, since it is already a long article.

L126: I would state in the introduction that the tests are to be conducted under present climate conditions, this will already justify using observations for the same period and then the clarification on the pre-industrial will not be needed here.

Great suggestion. We very briefly mentioned this in the revised introduction.

"*…and for what conditions they improve the dust model comparison to observations in the present climate…*"

"*…and the experiment we conducted (Sect. 2.6) under present climate conditions to…*"

L138: Please, change "models" by model.

Done.

L139: Please, remove "generally".

Done.

L141: CESM2, CESM2.1? CLM?.

Changed CESM2 to CESM2.1. As stated in previous comment on CAM6.1/CLM, here we think CLM is better to be kept as it was.

L153: Please, rephrase to specify the variable that is independent of the friction velocity (rather than the theory itself).

Good point. The revised statement now reads as:

"*The size distribution of the emitted dust is derived using the brittle fragmentation theory developed by Kok (2011b) distributing 0.1%, 1.0%, and 98.9% percentage of dust mass into*

*Aitken, accumulation, and coarse modes, respectively, independent of the friction velocity upon dust emissions (Kok, 2011a).”*

L154-155: As it is expressed now, the improvement in CAM4 size distribution is not informative to the reader. Please, either remove the part about the improvements or to briefly explain the difference between the approaches in previous CAM4 PSD and that derived from Kok (2011).

Deleted.

L156: Please, remove “other”, and “of aerosols”.

Done.

L179: Please, remove “the so-called”.

Done.

L178: As mentioned above, please, select just one acronym for the standard deviation.

Using only one now.

L189: Please, change “their ranges”, by “its ranges”.

Done.

The reference to Scanza et al. (2015) was already included.

Reference deleted.

L221: Is the vertical transport modified per se? Or is it indirectly affected by changes in emission/size?

We did not modify it per se, so, the change is indirect due to changes in dust emissions and size. In the very first version of this manuscript, we also perturbed the vertical layers which can affect the vertical transport more efficiently, but we deleted that part after. In response to the reviewer’s question, here we removed “vertical transport” to avoid possible confusions.

L231: What do the authors mean by “although even dust modeling with BRIFT can be improved if optimized against observations”, is that optimization relevant for this specific study?

There is a typo which caused this confusion and was corrected in the revised manuscript. We replaced BRIFT here by DEAD. It means that improvements could still be likely achieved using other methods, such as statistical optimizations (Kok et al., 2021) rather than employing the new dust emission scheme.

L328: Please, avoid repeating references unnecessarily (e.g. remove described in Sect. 2.2).

Repeated references removed.

L333: There are two references for Kok et al. (2021), please, specify a or b.

Done.

L338: Please, change “could change”, by the appropriate: does or does not change the model performance?

Paragraph removed in response to the next comment.

L338-343. May not be necessary to explain again the content of each sub-section.

Removed the navigation paragraph.

L358: Please, explain what the binned method is.

This is a terminology that has been widely used in aerosol modeling: it has been frequently used without definition. This study does not employ the binned method as well. So, we believe it may be fine without explaining it here.

L466 (and other locations in the text): Please, refer to the different experiments as such, instead of mentioning the models. If preferred by the authors, they could use model versions.

We changed “models” to “*all experiments*” here and at other locations in the text as well.

L369: Please, identify the reference with a or b.

Done. Changed to “Kok et al. (2021a)”.

L432: Typo: averages.

Corrected. Thanks!

L439: Change “to the low” by “to the *lower*”.

Done. Thanks!

L475: Have the authors information on the precipitation evaluation for their own model?

No, but CAM6 had been fully evaluated over aspects including precipitation.

L524: Please, include the coordinates of both stations or none.

The coordinates of both stations included.

L543: Does the super coarse dust start at 10 um? or larger diameters?

There is no clear boundary between “coarse” and “super coarse”. Here we refer to particles >10 μm in diameter not including the 10 μm. Since the coarse dust is well defined in Table 1 as dust in the coarse mode, and there is no “super coarse dust” clearly defined, we removed “super coarse” in avoid of possible confusions. So, now only keep expressions like “dust coarser than 10 μm in diameter” here and elsewhere in the text.

For example, we changed “the super coarse dust particles are also…” to “*dust particles in this size range are also*…”.

L614: Hematite and illite have a high iron content, feldspars not much. The sentence could be rephrased as “, including hematite and illite, and feldspar”

Rephrased, but following both reviewers’ suggestion, we moved contents of minerals in the companion paper. Thanks!

L636: I believe the increase is in wet deposition (not dry), please, verify.

Fig. 6c suggests the increase is in dry deposition. This increase could probably stem from release of fine-mode particles by evaporation of the cloud-borne dust. We revised the statement, such as it reads now as:

“…*which then become cloud-borne. The increased cloud-borne particles in turn increase the possibility of horizontal transport and release of particles by the cloud droplet evaporation, leading to an increase of*…”

L655: Please, include the full reference and then in parenthesis the values.

Done. It reads now as:

"*…between the global mean DOD in Aerosol Comparisons between Observations and Models (AEROCOM; median: 0.023) (Huneeus et al., 2011) and that in Ridley et al. (2016) (0.03±0.005) near the visible band.*"

L683: Why is the calculation explicitly included there? It makes the text more difficult to read. I would avoid it (here and in other locations in the text below).

We removed this kind of expressions everywhere.

L699: The sentence "where the dust emission occurs in transport" is difficult to understand, please, clarify.

Changed it to "*the importance of accurately simulating convergence-related convection (i.e., haboob) (Marsham et al., 2011) and where the dust emission occurs for dust transport modeling…*"

L869: Substitute "new model" by the appropriate model version name.

The new model version name inserted.

Table 1: I would order the modes from smaller to larger in size. I believe this table could be included in the supplement and leave in the text exclusively the default and new configuration for the coarse mode.

The order in the table 1 is the same as that in the model which lists the accumulation mode ahead of the Aitken mode. We included this table in main text, because we wanted to inform the readers about the mode information, for which they may search while reading through the main text, especially considering that the mode change is one of the important changes we made to the model.

Table 4: Why is the dust SSA for NEW_EMIS_SIZE missing?

When designing and performing simulations, we did not attempt to address impacts of these changes on dust radiative effect. So, we had not requested model output for this variable in that single experiment. According to dust SSA from the other experiments shown in this Table, we speculate a value around 0.90 for this experiment. But, since we did not present model-data comparison of dust SSA, we believe the missing of dust SSA in this single experiment may not influence the overall merit of this work.

Please, homogenize the naming convention for the different experiments, here tagged in Table 4 as NEW_EMIS, NEW_EMIS_SIZE, etc. In Table 2 and sections 2.1 and 2.2 they were listed also as EXP01, EXP02, etc. In Table 4 caption CAM6S5 and CAM6S6 are mentioned, which were not identified nor described before.

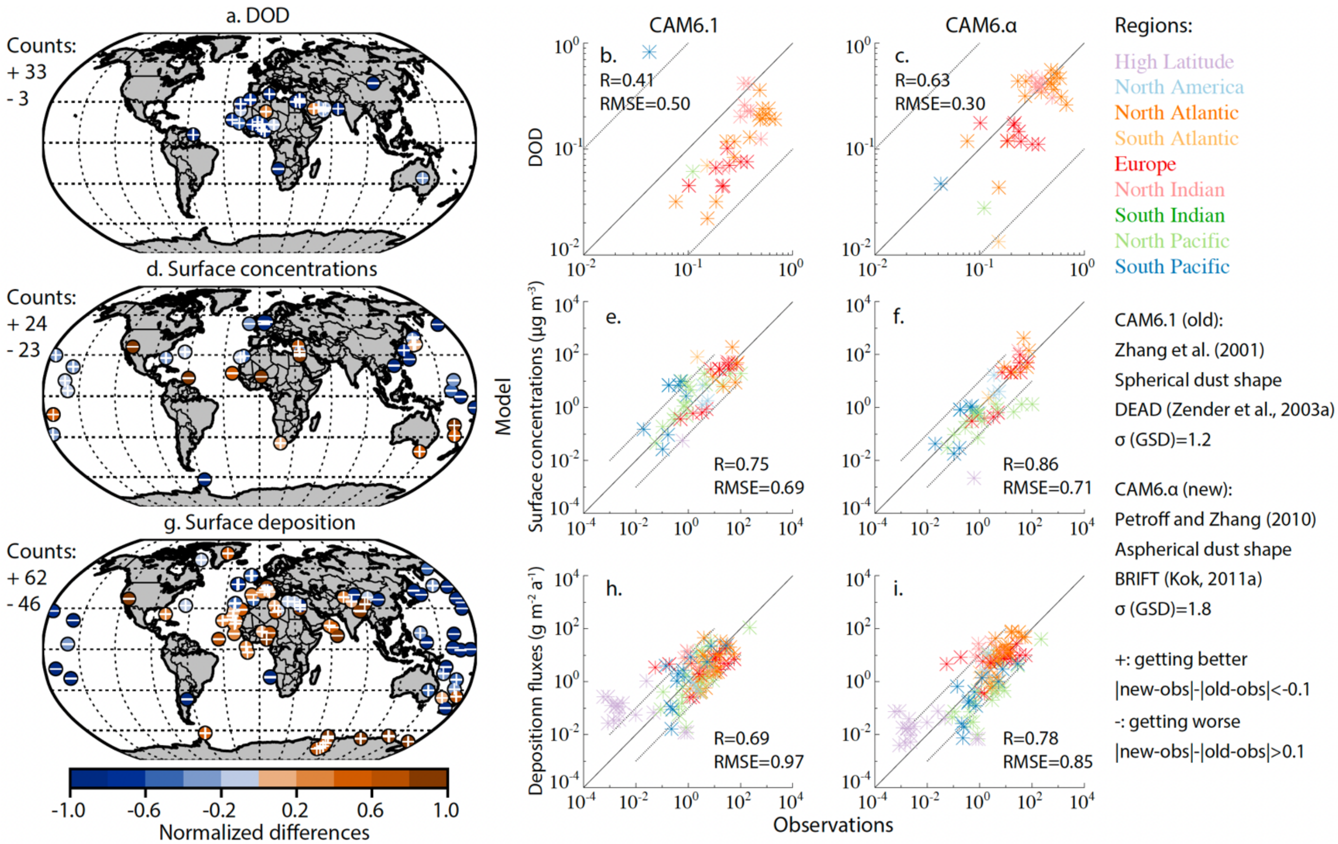
CAM6S5 and CAM6S6 deleted. The case names are all consistent through the text now. We revised the “size” column in Table 2 as well, since those notions are no longer used.

Table 5: Could the locations be represented in a map, together with the other observations location?

We provided such information in the revised table (first column), but we did not show that for each set of the observations in a map together with location information of the other observations, since otherwise the map would be super busy and very confusing, considering the huge number of observations we have included in this work.

Figure 1: Which is the metric used to define the improvement (+) or worsening (-) of the comparison? Remove the comment on Figure S3 from the caption, and if needed, clarify in the text (line 392) the information presented in main paper and in the supplement.

The citation of Figure S3 removed, and the metric used clarified in the figure.




"Figure 1. Model-observation (AERONET) comparison for DOD (dust optical depth) at the visible band centered at 0.53 μm (a, b, and c), dust surface concentrations (d, e, and f), and surface deposition fluxes (g, h, and i). Colored dots in a, d, and g show the difference between the proposed new model (CAM6.α) and observations. White symbols indicate the new model CAM6.α improves (plus sign) or worsens (minus sign) the model-observation comparison over that between the default model (CAM6.1) and observations *with the metric included in the bottom right-hand corner of the figure*. Numbers listed in a, d, and g are counts of the number of improved or worsen stations. The spatial correlation coefficients between model (CAM6.1: b, e, and h; CAM6.α: c, f, and i) and observations were calculated based on the annual mean values in log space (the log of each model and observational value was taken before calculating the correlation coefficient, since the values span several orders of magnitude except DOD). Dash lines in the scatter plot show 10:1 or 1:10 lines."

Figure 2: Could the re-scaling factors now explained in the caption be included also in the figure legend (e.g. above each map)?

Added. Please see Fig. 2 in the revised manuscript.

Figure 5: Please, review the caption: remove "and" in the third line, remove "for the abbreviation for other models", either explain them there or leave just the reference, specify what do we

understand by semi-observations. Please, do not refer to other figures in figure captions unless they are needed to understand the figure contents.

Removed. We also added the following in the caption:

"…*semi-observations: DustCOMM (black line) inverted based on an integration of a global model ensemble and quality-controlled observational constrains on the transported dust size distribution, extinction efficiency, and regional DOD*"

Figure 6: What do the maps represent? Is it the ratio? Or the differences over the reference?

We believe the caption for Panel a-h is clear on this. "Figure 6. Impacts of the dust emission scheme (a and b: **ratio of BRIFT to DEAD**), aerosol dry deposition scheme (c-f: **ratio of PZ10 to Z01**), and dust shape (g and h: **ratio of ellipsoidal to spherical dust**) on the modeled dust deposition (total: a, d, and g; fine mode: c), and dust loading (total: b, f, and h; fine mode: e)."

Figure 7: Please, use the same naming convention for the different experiments along the manuscript, otherwise is very confusing.

Done.

Figure 8: Homogenize the experiment names with the rest of the document, review the seasons listed in the caption, the inserted map below is not shown in this document version.

Changed relevant text to:

"Figure 8. Modelled and observed dust direct radiative effect efficiency in the shortwave (SW)/longwave (LW) spectral ranges *under clear conditions at the TOA over the sub-domains (location described as [lat, lon]) in April-June (AMJ), summer (JJA), fall (NDJ), and September (Sep) for the 2000s climate.* The radiative effect efficiency is defined as the ratio of the radiative effect to DOD, so has units of W $m^{-2}$ $\tau^{-1}$. Included cases from left are CAM6.1, CAM6.α, MINE_NEW_EMIS_SHAPE, CAM6.α _MINE. The field value/range are from references listed in Table 3. *Colored numbers show correlation coefficient (R) and the root mean square error (RMSE) between the model and retrievals in the SW (a) / LW (b) spectral ranges or in both spectral ranges (numbers in parenthesis in Panel a).*"