We thank the referee for comments and questions. They help us to explain better several aspects of the paper. Here we are indicating our answers in boxed frames after each point raised by the reviewer and our changes/actions in the manuscript within a green colour box.
1 General comments

This manuscript presents the results of five Earth System Models simulations of the global dust cycle, emissions, dry and wet deposition, optical depths, and surface concentrations comparatively to satellites and in situ observations. The authors explore global and regional variability between models in three different simulated experiments: PD (calculated winds), PDN (reanalysed winds), and PI (prescribed chemistry and aerosols). Overall, the manuscript is well written and provides ample content. Having said that, this manuscript is quite extensive and important information is left for the reader to find in the supplement. The content of this manuscript could be divided in two different publications. In the first one, you could explore the differences between the five models, and then, in the second, you could explore more deeply the differences between the three simulated experiment scenarios.
The reason to have a single publication for the mineral dust evaluation is that we have a publication plan within CRESCENDO-ESMs: one study per aerosol species. This also explains why we are including here the PI experiment (with prescribed chemistry and aerosols precursor emissions to pre-industrial values). In the case of mineral dust, we don’t expect important differences between PI and PD experiments. For other aerosols however, we expect larger differences, and within this scope it is convenient to compare all three experiments systematically. As commented to the first reviewer, other analysis based on diagnostics per bin-size, vertical structure of dust concentrations or long dust transport have not been added to this study although a few of these results have been shown in conferences already.

At least, I would include a figure of the particle size distributions used in each model into the main text.

During the manuscript preparation we considered representing the dust particle size distribution (DPSD) as suggested by the reviewer. However, the shape of DPSD depends on the localization of the grid cell of the model and the time. We have two options: a representation at emission at specific locations, or a kind of global mean particle size distribution.

- For the first option, we have to deal with the several kinds of dust emission schemes. For those based in the brittle fragmentation, it is possible to have a representation of a normalised size distribution (like Figure 1 in Di Biagio et al. (2020)), as far as, the different modes have a prescribed mixing factor. However, several models proceed with a sectional emission scheme that is mapped into a multimodal log-normal size distribution, and it is not easy to have a unique multimodal dust size distribution at emission to compare with.
• For the second option we did not find a suitable average in space and time that can help us in the discussion.

For this reason we have decided to include a detailed description of the several dust emission schemes rather than plot a qualitative DPSD.

We have introduced two classifiers related with the modelling of coarse/large particles to improve the description and discussion of those aspects related with the DPSD (more information in our answers to reviewer 2).

2 Specific Comments

• Page 6, Line 7-31: For this part of the text, it would be very useful to have a plot overlapping the particle size distributions in each of the 5 models. This would be similar to what you have in the bottom panel of Figure 4.

As commented above we did not find how to include a plot of the overlapping DPSD without being specific of a location at emission, or without defining an average method that can not be representative of the several microphysical processes in different regions.
• Page 7, Line 11: “Therefore those optical properties are representative for the global mineralogical composition rather than a description of the soil-type dependence of the mineralogy that would imply local differences on emitted optical properties.” The point you are raising here is important, but it is still not clear what optical properties you have actually used. For instance, do all five models use exactly the same spectral complex refractive indices? Which databases/references are you using in each model?

Thanks for this comment. We agree that this information is important in the paper. In CRESCENDO each model is using a different refractive index. Also, each model implements slight differences in the pre-computation of lookup tables for each optical parameter.

We have added this information in the Table 1.

• Page 3, Table 1: It would be useful for the reader to include the specific particle sizes ranges used in each model simulation in column DPSD. Alternatively, you could move Tables S.MD.8 and S.MD.9 from the supplement to the main text. Please, also include the meaning of PD, PDN, and PI in the title of Table 1.

Thank you for pointing it out. We have included new information in the main manuscript and finally the Tables S.MD.8 and S.MD.9 remain in the Supplementary information for further reference.

We have added:
  – Two columns in Table 1 to classify the modelling of large dust particles by the different models.
  – The meaning of PD, PDN and PI in the Table caption.
  – A new table with the information of the several model experiments.
• Page 8, Line 2: "A last simulation where aerosols and chemistry emissions are prescribed for 1850 (named PI)". Why? Could you add a few words explaining why such simulation is relevant and how do you use its results specifically in this study, covering the years between 2000 and 2014?

We have explained the initial motivation above. For mineral dust the hypothesis to be tested is that the differences between PD and PI are small and the general behaviour is the same. Any kind of indirect effect is much smaller than the role of wind fields (PD vs PDN). If, as a consequence of prescribed emissions at 1850, we have slight differences in clouds or precipitation this potentially could have a slight effect on dust cycle. Our comparison suggests that those effects are not conditioning the global behaviour of dust global cycle.

We have added to Section 2.1: The comparison between the PD and PDN experiments inform about the role of wind fields to explain model diversity. The difference between PD and PI dust emissions allow us to evaluate whether the effects in the climate system due to non-dust emissions have a discernible impact on the global dust cycle (as both PD and PI have been prescribed with the same SSTs). A summary of the properties of the model experiments is given in Table 2.

• Page 12, Line 3: Could you please clarify the criterion used to select optical depths? What does "all – aer" mean here?

Thank you. $\tau_{all-aer}$ means the total optical depth of all aerosols in contrast with $\tau_{dust}$ that is only for dust. We have added this information in the main text.
We added a better description in the main text:
\[ \tau_{dust}^{440} > 0.5 \tau_{all-aer}^{440} \] for all models and all the months of the year, where \( \tau_{all-aer}^{440} \) refers to optical depth at 440 nm of all aerosols and \( \tau_{dust}^{440} \) is the optical depth of mineral dust aerosols at 440 nm.

• Page 15, Figure 4: What do you mean by "samples are the marks on x-axis"? Does the "sample" correspond to a given year? Which are the years you consider here? Include the time-period in the figure caption.

Thank you for this comment. We improved the sentence. In the Figure 4, each sample (the value for a year) is represented by grey-dots (for the top panel) and by coloured-marks just over the x-axis. The idea is that each year is a sample, and the plot is more informative with these dots and marks.

We improved the caption by explaining that:
- The grey-dots (top-panel) over the box-plot represent each of the annual values.
- In the bottom panels our sample values per model are represented by the coloured vertical marks just above the x-axis.
- For both, the models and the observations (MISR and MODIS), the estimates are for time-period 2000-2014.

• Page 22, Line 21: Could you clarify if the model simulations were sampled at MODIS and MISR times in Figure 4? What is the main reason for having UKESM’s AOD so high?
Thank you for this comment. Yes, the years are the same for MODIS and MISR in Figure 4 bottom right panel. We have added this information in the caption. In the context of the paper, we know that the large values of UKESM are not due to mineral dust, so it is related with other aerosol species. In principle, most of the models (in other multi-model evaluations like AEROCOM) show global values of aerosol optical depth smaller than MODIS. However, the reason for that is an active research topic at this moment. In our figure the estimates of AOD are also in general smaller than MISR. The best scenario is a progressive convergence between models, but also between observations. Although only with two satellite platforms, the Figure 4 also suggests that the diversity of total aerosols optical depths is larger in models than in observations. The inter-annual variability on AOD by MODIS is also the largest one.

We have added to the Section 5.1: The bottom right panel of Figure 4 indicates model discrepancies in the magnitude of the inter-annual variability (as measured by the width of the distribution) and an overall underestimation of AOD at 550 nm with respect to satellite platforms.

• Page 22, Line 29: Do you know why the EC-Earth and the NorESM have MEE values that differ from the other models? If so, I suggest you to discuss the main reasons in the text.
We think that MEE values of the EC-Earth and the NorESM differ (from other models) by a combination of factors. Both models have the lowest dust loadings, in the case of NorESM, also with the smallest inter-annual variability. In the EC-Earth the dust scheme has a cut-off at $8 \mu m$ and NorESM an accumulation mode with smaller particles than the equivalent mode of IPSL or EC-Earth so there are differences in modelling of the particle size distribution. In the case of NorESM the imaginary part of the refractive index is also the largest.

- Page 23, Line 7 to 24: The plot overlapping all size distributions would again be helpful here.

Thank you. We have introduced additional information in Table 1 and improved the text.

- Page 28, Line 7: “On the side the modelled wind surface friction velocity and speed agree better with actual meteorological conditions, and on the other side the description of the soil surface properties has become more accurate.” These are important points. Could you provide references of experimental studies that support these two statements?

Thank you. These are mostly general assessments about the improvement of the climate models in the boundary layer with respect to wind fields, and in the land surface. These aspects are not restricted to mineral dust science, and belong more to the evolution of climate models.
We have added two additional references in our revised manuscript (for more general discussions good references are: (Shao, 2008; Knippertz and Stuut, 2014)):

On the one hand the modelled wind surface friction velocity and speed agree better with actual meteorological conditions, and on the other hand, the description of the soil surface properties has become more accurate due to both, improvements in the soil texture databases, and the use of satellite retrievals to better describe the roughness length, e.g Prigent et al. (2005); Menut et al. (2013).

• Page 30, Line 26: “. . . they indicate that although there are important differences between PD and PDN experiments in terms of total emissions, . . .” . It is difficult to see those differences here.

Thank you. The important differences are in the global emissions, for example, CNRM-6DU has 3450 Tg/yr for the PD experiment and 1278 Tg/yr for the PDN experiment. However, these large differences in global emissions are not discernible in the normalized emission maps.

• Although it might be out of the scope of this paper, I think it would be interesting to comment on how the re-analyzed wind fields in PDN differ from the calculated wind fields in PD near dusty regions. Are there significant differences between the wind fields in PD and PDN? How much would be that difference?
Thank you. This is an interesting comparison. We expect differences that explain the decrease on the model diversity of mineral dust when we have consistency for the wind fields. Several previous studies have been done for a single model. We think that the ideal situation is to compare at several temporal resolutions that can be relevant for dust related processes to actually explain the improvement in agreement between models for the diagnostics analysed. It is an interesting suggestion for future evaluations with global climate models, but the analysis would be too extensive to include in this already long paper.

• Page 35, Line 22: “All the other models underestimate total depositions fluxes over stations where fluxes exceed 100gm-2yr-1.” What do you think is the main reason for that?

This fact is probably correlated with the largest emissions of UKESM model and its modelling of particles with diameters larger than 20 µm.

We added this interpretation to the discussion.

• Page 44, Line 7 to 30: Could you comment on the temporal resolution of the surface concentration observations? Are those monthly means based on continuous daily observations? Are these observations for one or the average value for multiple years? Include the years of the observations and the simulations in the figure caption.
There are few available observations of dust surface concentrations. The general treatment is to consider them as **climatological dataset** (although we know that it is an approximation) and compare their values with the values for our 15 years of simulation (from 2000 to 2014). The same hypothesis has been applied by Huneeus et al. (2011) and Albani et al (2021). We further investigate possible discrepancies due to different sampling years of measurements. In this regard, we added those of the reference Cheng et al. (2008) to the Table S.MD.4 in the Supplementary information to visualize the role of annual sampling. We observe that measurements from different years are only slightly different but not discernible in the figures of the paper, and the conclusions are the same. The values are included in the table S.MD.4 in brackets. We only had access to the raw INDAAF dataset where the observations of PM10 are measured with high-temporal resolution (several measurements per day).

We refer better to Table S.MD.4 in brackets and explained the interpretation as a climatology dataset.

### 3 Technical Corrections

- **Page 8, Figure 1**: The aspect ratio of the Figures 1a and 1b seems strange.

  Thank you. This figure has been improved.

- **Page 14, Line 27**: Replace “has” by “have”.

  Thank you. We have changed the text of the paragraph.
• Page 16, Line 25: Replace “in the main paper” by “this paper”.

Thank you. We replaced the word "main" by "this".

• Page 17, Table 6: Explain the meaning of “in and out”.

The definitions are in the Appendix A. But we have added a less mathematical and more visual definition in the Table 6: the method in add each specific mode to a case without any mode of dust, the method out remove that specific mode to a case with all the modes of dust. The DRE estimates give different values for each method.

• Page 33, Line 11: “Figures Dep.11 and Dep.12 show . . .”. I cannot find these figures.

Thank you. We included the right names of the figures of the supplement.

• Page 44, Line 25: “sitation”

Thank you. It should be "situation".

References


