

Comment on acp-2021-578

Anonymous Referee #2

Referee comment on "How well do the CMIP6 models simulate dust aerosols?" by Alcide Zhao et al., Atmos. Chem. Phys. Discuss., <https://doi.org/10.5194/acp-2021-578-RC2>, 2021

"How well do the CMIP6 models simulate dust aerosols?" by Zhao et al. (**acp-2021-578**)

This article analyzes the dust aerosol cycle in AMIP-style simulations by 16 CMIP6 Earth System Models (ESMs). The analysis is based upon a comparison to two reanalysis products containing dust, MERRA2 and CAMS, mainly over the period 2005-2014. The CMIP6 multi-model ensembles (MEMs) are used for studies of future impacts and mitigation, so it is important to evaluate their fidelity (in this case, of their dust cycle). The article is clearly written. However, the interpretations and significance of the article are questionable because of two problems. First, the authors treat the reanalysis products as an observational standard, despite substantial uncertainties and disagreements between the two. Second, the authors compare total emission (and deposition) without accounting for the varying range of particle size represented by the models. This neglect of size challenges any interpretation of emission diversity between the models. I suggest that the authors give less emphasis to model variables like emission and deposition that are strongly size-dependent. I also suggest that the authors compare to actual observations and retrieval products where possible, while giving more emphasis to the difference between the dust cycle calculated by the MERRA2 and CAMS reanalysis models.

Major Comments:

1. The MERRA2 and CAMS reanalysis products have substantial biases, as evidenced by the twofold contrast of global load (23 Tg for MERRA2 v. 12 Tg for CAMS: line 259). Both reanalysis products rely upon models that assimilate total AOD. The problem is that the contribution of dust to total AOD (the dust optical depth or DOD) is strongly model-dependent. The reanalysis models, like the ESMs themselves, make a number of assumptions. The article notes that both reanalyses compute emission using a scheme taken from Ginoux et al. (2001). There are many admirable features of the Ginoux study, but the calculation of emission has a particle size dependence that is now recognized to give unphysical emphasis to smaller particles (as discussed by Legrand et al. GMD 2019 <https://doi.org/10.5194/gmd-12-131-2019>). Moreover, both models weight emission

using the Ginoux topographic source map. Other regional weightings (i.e. erodibility maps) are used in some ESMs. These maps are equally plausible but emphasize different regional sources. (See Fig.1 of Cakmur et al. JGR 2006.) In summary, the MERRA2 and CAMS reanalyses are heavily dependent upon modeling assumptions, just like the ESMs, which undermines the use of the reanalyses as an observational standard. There is a nice comparison of MERRA2 and CAMS by Xian and Klotzbach et al. (ACP 2020 <https://doi.org/10.5194/acp-20-15357-2020>), who instead recommend a multi-reanalysis composite, while emphasizing the resulting uncertainty. The authors of the present article also use the ModIs Dust AeroSol (MIDAS) product for ESM evaluation. The problem is that they construct DOD given the retrieved AOD combined with the MERRA2 ratio of these two variables. In other words, the "observed" DOD in fact is dependent upon the contribution of dust compared to the total aerosol extinction as calculated by a single model. In the end, the effect of differing reanalysis model assumptions means that their output is a highly uncertain standard with limited influence of actual observations. (Again, as evidence, note the difference in global load between the reanalyses.)

In fairness to the authors, one challenge of evaluating a dust model is that instrument retrievals, which are spatially detailed and available for as long as two decades, do not differentiate between different aerosol species. Their use introduces uncertainty into any evaluation of a model dust cycle. However, the reanalyses are highly uncertain for the same reason. The authors need to address this uncertainty rather than just talk about model diversity and biases. (Neither reanalysis product is named in the abstract nor are any of their disagreements noted.) The uncertainty and lack of consensus among the reanalyses has to be an explicit part of the study and given emphasis in the abstract and conclusions.

2. Another problem with the article is its limited consideration of particle size range. While the authors note that "dust particle size range represented differ significantly between models" (line 102), their discussion of diversity of model emission does not account for this varying range. In the Abstract (line 10), they write "For example, global dust emissions, primarily driven by model-simulated surface winds, vary by a factor of 5 across models, while the MEM estimate is double the amount in reanalyses. " Not all of this diversity is a result of uncertain representation of the physical processes controlling emission. Some of it is simply based on a somewhat arbitrary decision by each modeling group about the maximum particle size to represent. The authors cite a multi-model ensemble mean (MEM) emission of 3.5 Pg/yr (line 189), but this average results from the combination of emission from models with varying size ranges and does not solely reflect our imprecise knowledge of emission physics. On line 191, the authors refer to the "observationally-constrained estimates of $\sim 5 \text{ Pg yr}^{-1}$ (Kok et al., 2021), but Kok et al. explicitly consider only particles with diameters of less than 20 μm .

This incomplete characterization of model emission extends to the analysis of deposition and lifetime. The authors write that "dust is predominantly removed by dry deposition (60-86%) in most models," (line 274) but this statement strongly depends upon the represented size range in each model. Models with larger maximum particle diameters will remove more of their dust using dry deposition and have shorter particle lifetimes (Figure 8) even among models that represent the physical process of deposition identically.

To be sure, variations of particle-size range among models are imposed upon the authors because the CMIP6 archive records only size-integrated emission, load and other variables. Still the analysis in this article would be much more useful if the authors distinguished the impact of uncertain model representations of emission and deposition physics from the varying ranges of particle size. One (uncertain) way of addressing the effect of size is to plot for each model its emission (or its logarithm) vs. the largest particle size. I would expect that models with larger particles will generally be associated with larger emission. This will help untangle (albeit imperfectly) the influence of physics and model size range upon the diversity of emission and deposition.

Finally, in the abstract and conclusions, the authors should note the challenge imposed by the absence of size-resolved emission and deposition in the CMIP6 archive and strongly recommend the addition of this dependency for CMIP7. In the abstract, the authors recommend that future MIPs request "More detailed output" (line 19) without providing an example.

Dust load and dust AOD are also subject to this limitation, but to a lesser degree because the larger particles (that might cause the greatest discrepancies in model emission and deposition) make smaller contributions to the former variables.

Minor Comments

42 "as they become a larger fraction of the total aerosol burden" We expect that air-quality regulations will reduce the concentrations and impacts of anthropogenic aerosols, but whether the global dust load becomes larger and thus "has a greater role in shaping future climate variability" is uncertain.

57 "while their year-to-year changes were poorly constrained compared to observations" This was an ill-posed test. Evan et al. (2014) compared interannual dust variations to CMIP5 twentieth century historical simulations that were initialized in 1850. These models cannot be expected to reproduce observed interannual dust variability any more than the models can be expected to successfully predict the weather on any given day in the late-twentieth century.

59 "satellite-observed and CMIP5 models' simulated decadal variabilities of dust emissions..." Satellites do not observe dust emission.

68 "featured amplified uncertainties" How is uncertainty defined here?

110 "the intermediate horizontal resolution" Please specify this resolution explicitly.

154 "we made leaf area index and soil moisture redundant" Do you mean that you deleted them from your regression model?

185 "dust emission hotspots" What is meant by 'hotspots' in this sentence?

199 How is "North China" defined?

205 "CMIP6 models also feature diversities in the global surface area of dust emissions" This is an interesting metric!

226 "MEM and most individual models are much larger (up to 10 times) than those in CAMS and MERRA2" Can you estimate how much of this difference is due to contrasting choices of particle size range by each model?

235 "Surface wind speed is shown as the dominant driver of dust emissions in all the models and CAMS." Here, it should be noted that the regression model is based upon monthly mean winds while a disproportionate fraction of emission comes from strong winds on shorter time scales that may not always be well-correlated with monthly means.

Figure 3: add a vertical scale to each panel?

305 "This highlights the inconsistent behaviour of CMIP6 models in simulating the optical depth of different aerosol species." This is an interesting result, but doesn't this inconsistency call into question the derivation of DOD from MIDAS retrievals combined with the fraction of dust from a single model (in this case MERRA2)?

309 Here, you should add the Ridley et al. ACP 2016 observationally constrained estimate of $DOD = 0.03 \pm 0.01$ for PM₂₀ dust.

310 "significant biases in the MEM-simulated DOD magnitudes at regional scales." Do these biases really reflect unrealistic physics of some of the models? How much of this bias is due to the reanalyses using the Ginoux JGR 2001 topographic erodibility map rather than the other (physically reasonable) maps used by some models?

Figure 9e, f and g. Please label the axes on the plot.

351 typo "Out key findings are"

388 "more detailed output relating" Please make specific recommendations here. (e.g. add size-resolved variables)