

Review of “How well do the CMIP6 models simulate dust aerosols?” by Alcide Zhao, Claire L. Ryder, and Laura J. Wilcox

This paper examines the performance of 16 CMIP6 models in simulating dust emissions, deposition, burden, lifetime, and dust optical depth (DOD) for the present day climatology (2005-2014). The AMIP-type of model results are compared with reanalyses (CAMS and MERRA-2) and satellite retrievals. It is found that models in general capture the spatial patterns of dust emissions, loading, and deposition but have discrepancies in representing seasonal cycles of dust over North China, North America, and South America and show a large diversity in simulating DOD. It's a timely paper providing an informative evaluation of the life cycle of dust in CMIP6 models. While I appreciate the authors' great work and detailed analysis, I found the uncertainties of the data and comparisons are not well addressed in the paper. My comments are listed as follows.

Major comments:

1. While aerosol reanalysis products, e.g., CAMS and MERRA-2, provide high spatial and temporal resolutions of data to study dust, it should be noted that variables such as dust emission, deposition, dust loading, and DOD, are not directly constrained by observations—only AOD is directly assimilated with satellite retrievals and ground observations. The uncertainties associated with reanalysis variables and how these may affect the comparison with CMIP6 model output need to be discussed.

I'm also not fully convinced by the argument that “we did not evaluate the models against ground-based measurements but instead perform a large-scale analysis focusing on the more spatio-temporally available fields from reanalyses that are in good agreement with ground-based observations where they exist (Wu et al., 2020a, b).” (lines 347-349). While MERRA-2 may generally have a good agreement with ground-based deposition data as noted by previous work, it does not necessarily imply that ground-based measurements can be substituted by reanalysis products. Not to mention ground-based deposition data also have uncertainties.

In short, a clarification and detailed discussion of the uncertainties of the data used here as references for CMIP6 model evaluation would be useful. For satellite products, e.g., MIDAS DOD and FMI AOD, providing error ranges, which can be found in the referred papers in the text, would be helpful, too.

2. How the timing of dust emissions affects the seasonal peak of DOD is briefly discussed for North Africa in Fig. 13 (lines 337-339). I wonder if it's possible to add more discussion about generally how the representation of dust emissions and deposition affect model performance in simulating DOD.

Minor points:

1. Line 138, how is bare soil fraction defined? Do you use the output of bare soil fraction from each model to perform the regression?
2. Section 2.4, a similar multiple linear regression was used by Pu and Ginoux (2018) to study drivers of global DOD.
3. Line 167, it would be better to add the definitions of regions (boxes) to Fig. 1 instead of keeping it in the supplement (Fig. S1).
4. Line 175, please consider adding a discussion about the uncertainties associated with dust emissions from CAMS and MERRA-2.
5. Line 194, have you compared dust emissions for CMIP6 models with and without dynamic vegetation schemes to see if the latter generally have larger uncertainties and emissions?
6. Line 197, “28-69”, is this a range from the 10th to 90th percentile or the minimum to maximum?
7. Line 245, cannot find any clear impact of soil bareness (brown shading) from CAMS in Fig. 4I.
8. Lines 274-275, very interesting. Any idea about why? e.g., do you have a figure similar to Fig. 7 to clarify over what regions the wet deposition are high in CESM2?
9. Lines 281-282, this adds to the need to discuss the uncertainties of evaluating CMIP6 model output against CAMS and MERRA-2.
10. Line 319, “Fig. 11a”, referred to a wrong figure?
11. Line 329, “black crosses”, refers to model mean?
12. Lines 363, “dust accumulations along the southern edge of the Himalayas” can you please highlight that part in the figure? And what about uncertainties associated with satellite retrievals of dust from snow surface?
13. Fig. 2 caption, “The global annual total dust emission budgets (E_g ; Pg yr^{-1})”, I think you refer to Fig. 2(d), right?
14. Fig. 6, consider adding boxes of the meridionally-averaged regions, Asia and Africa, to plot 6(a).
15. Fig. 8, maybe use a boxplot or add error bars (e.g., the 10th to 90th percentiles) for AMIP MEM instead of using a plus sign?

16. Fig. 9, why does the font size of the letters in (h) differ? What is spatial R^2 ? Is it pattern correlation?
17. Fig. 12(a)-(c), what is the x-axis? Is the y-axis frequency (%)?
18. Fig. 13, maybe add the lines of AMIP MEM dust emissions for a better comparison of the timing of emission and DOD maxima?