

We thank the reviewers for their thoughtful and constructive comments. We made substantial changes to improve the overall structure of the paper and clarify the study's objective and novelty compared with previous studies. Detailed responses to the reviewers' comments are provided below.

Reviewer #1

This study investigates the dominant factors contributing to the spatial (regional and aridity-level dependent) and temporal (interannual) variability of dust emission, drawing upon Earth System Model (ESM) results for the present day provided to CMIP6. Similar prior works (not cited in the manuscript) have previously reached comparable conclusions. So the work lacks originality. A key distinction is that previous studies relied on observations, allowing for absolute comparisons. This study, however, does not utilize global long-term observations of dust properties, such as satellite-based Dust Optical Depth (DOD). Consequently, its findings are limited to a model inter-comparison, without the ability to assess potential strengths or weaknesses.

Response: Unlike previous studies, our work provides a more focused and in-depth analysis on dust emission. We evaluate how the simulated interannual variability of dust emission differs across CMIP models, which has not yet been addressed in the literature. The results of this work thus provide critical insights into how dust emission projections would differ across these models under future climate change. A novel aspect of this study is that dust emission is viewed as an unobservable, model-specific quantity, instead of a true, observable geophysical quantity, reflecting the lack of direct observational constraints on the dust emission magnitude. Multiple linear regression, which was used in past studies, cannot disentangle the influence of various dust emission drivers due to multicollinearity. Here we used the dominance analysis technique to quantify the joint and relative importance of correlated drivers. Finally, we include CESM and E3SM models with newly updated dust schemes, which remove the use of dust source functions and incorporate more physically based parameterizations, compared to models participating in the CMIP experiments. By comparing results from these new models, we demonstrate, for the first time, how the updated emission scheme simulates dust variability and sensitivities to the physical drivers. All these represent the originality of this work compared to past studies.

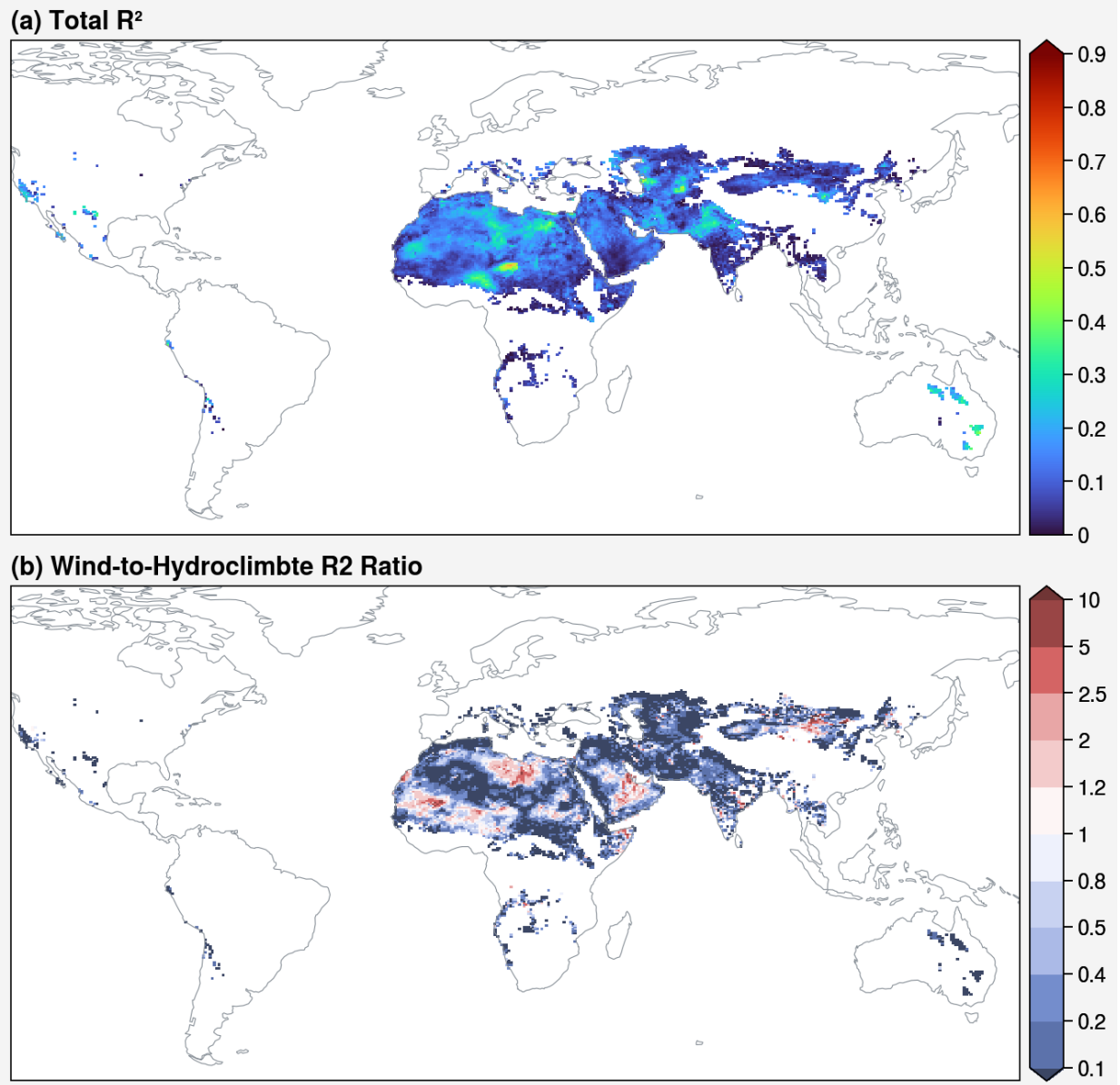
In response to reviewer's suggestions, we have attempted to use MODIS dust AOD (DAOD) for model evaluation. Unfortunately, the results are not encouraging, presumably because MODIS aerosol observations are not a good proxy for dust emission fluxes. See details below.

I would suggest to the authors to build upon the work of Zender and Kwon (2005) and Kim et al. (2017), perhaps focusing on long-term variability, given that MODIS offers over 20 years of daily global DOD data. Surface concentration at Barbados has been observed daily since 1965, providing 60 years of data—ample for studying inter-annual variability. Prospero and Lamb (2003) demonstrated that hydroclimate factors control dust long-term variability. By examining the ESM results that best reproduce the interannual variability observed at Barbados and/or in satellite data, we could discern the strengths and weaknesses of their controlling factor(s).

Response: This study's objective to evaluate the representation of interannual variability in dust emission within CMIP models. The surface concentration at Barbados is influenced by numerous factors beyond dust emission, such as those affecting deposition rates and transport pathways. Consequently, this variable cannot yield a definitive conclusion regarding dust emission itself. Nevertheless, we performed a new analysis by using MODIS deep blue aerosol products to evaluate model representations of the relative influence of dust emission drivers. The analysis is conducted in a consistent manner as for the models. Specifically, we performed dominance analysis using MODIS monthly DAOD (2001-2024) as target variable, and MERRA2 wind speed, precipitation, soil moisture, specific humidity, air temperature, and LAI as predictors. The MODIS DAOD is derived as follows: we first derive daily DAOD using the approach of Li and Ginoux (2025) for both Terra and Aqua MODIS level-2 products. The Terra and Aqua DAODs are averaged for the same day and regridded to $0.65^{\circ} \times 0.5^{\circ}$ resolution (same as MERRA2) and monthly levels. To isolate dust source regions, we focus on grid cells with at least 14 days of DAOD larger than 0.2, based on the frequency-of-occurrence method in Ginoux et al. (2012).

In Fig. (a) below, the total R^2 or variance explained by the six predictors is very low (<0.3) globally, substantially lower than R^2 values (>0.8) in MERRA2 dust emissions as shown in Fig. 7f from the paper. This indicates weak coupling between DAOD and the reanalyzed wind and hydroclimate variables, suggesting that MODIS DAOD is not a good proxy for dust emission fluxes. Furthermore, Fig. (b) below shows the ratio of wind speed-associated R^2 to hydroclimate-associated R^2 , calculated the same way as in Fig. 8 from the paper. We find that North Africa and Arabian Peninsula exhibit extensive areas of R^2 Ratio <0.2 , which suggests that dust emissions from these areas are predominantly controlled by hydroclimate variables, rather than wind speed. This result does not align with the models, and the fact that hyperarid regions are dominated by wind speed due to the permanent dry, barren surfaces and low precipitation variability.

Given the above results, we conclude that MODIS AOD is not a good proxy for dust emission fluxes and thus decide to not include it into the paper. There are several possible explanations. While dust AOD offers some information on dust source strength, it essentially measures the column dust burden, which mostly comprises fine particles and is influenced by atmospheric transport and removal in addition to emission. Compared to emission, the column burden is more sensitive to factors that affect the atmospheric lifetime (e.g., updraft velocity, precipitation). There are also other limitations, such as MODIS algorithm biases over bright desert surfaces, limited revisit frequency (only once a day), and cloud interference (i.e., leading to missing observations), all of which can affect its ability to represent dust source strength.



However, when analyzing interannual variations, certain factors influencing these variations should also be considered. These include land-use, which significantly contributes to dust emission (Ginoux et al., 2012; Stanelle et al., 2014), and fires (Yu and Ginoux, 2022; Wagner and Schepanski, 2025).

Response: We included the leaf area index (LAI), an indicator of green vegetation structure and density (which indirectly reflects the land use state), as one of the predictors to account for the effect of vegetation and land-use. We did not aim to evaluate the model performance in capturing the LAI effect, but whether the models have consistent representations of the LAI effect.

Some fundamental information regarding the ESMs, crucial for understanding their differences in dust emission, is either incorrect or inadequately explained. Furthermore, the analysis exhibits a bias towards CESM, with minimal or no discussion of other models.

The citations for MPI-ESM-1.2 and GFDL-ESM4 are erroneous. Tegen et al. (2019) described ECHAM6.3-HAM2.3 with constant roughness and vegetation mask. Is Mauritsen et al. (2019) not the correct reference for MPI-ESM-1.2? MPI-ESM-1.2 utilizes MAC-v1 prescribed aerosol distribution (Kinne et al., 2013). For GFDL-ESM4, it would be appropriate to refer to Shevliakova et al. (2024) instead of Evans et al. (2016). The authors are advised to consult these references or contact the lead authors to ensure an accurate description of their models.

Response: We have updated the descriptions of ESMs and the references for dust schemes and models in Table 1 and the body text.

Table 1 lacks critical information, such as LAI. Is it calculated online? Is it static or dynamic? Does it incorporate land-use? Is brown vegetation included? For 10-meter wind-speed derived from the first model level, the robustness of the derivation diminishes with increasing altitude of this level. Horizontal resolution is paramount for all fields. How can models be compared without knowledge of their spatial resolution?

Response: We added a new column for horizontal resolutions. To not make Table 1 too big, we removed the details of soil moisture and vegetation treatments, but added new columns for LAI treatment and dust scheme references. Some models (e.g., NorESM2, CanESM5.1) use dynamic vegetation simulations, while others (e.g., GISS-E2.1, HadGEM3-GC31-MM) use prescribed LAI. The different treatments in vegetation, land-use, winds, and other factors are the primary reasons why dust emission flux should be considered as a model-specific quantity.

Given these significant issues, I cannot recommend the publication of the present manuscript.

Reviewer #2

This article analyzed the inter-model discrepancies of the spatial distributions and interannual variability over dust source regions and climate zones, and quantified the total and relative explainability of wind speed and 5 hydroclimate parameters with the dominance analysis. The analyses of this study could help better understand the uncertainties of dust emissions among various Earth system models. The analyses and methods are reasonable. However, it would benefit from several improvements in terms of the size distribution and size cut of emitted dust, model resolution, model screening, and comparisons of dust burden and physical parameters against observations.

Response: We would like to emphasize that, unlike previous multi-model studies, our analysis provides a more focused and in-depth assessment of dust emissions rather than the entire dust life cycle. Specifically, we quantify how the simulated interannual variability of dust emission differs across CMIP models - an aspect that, to our knowledge, has not been systematically

examined before. This focus enables us to identify the physical drivers of dust emission variability and provide new insights into how dust emission projections may diverge among ESMs in response to future climate changes.

Because dust emission is an unobservable quantity, we explicitly treat it as such rather than as an observable geophysical field, a distinction that is essential given the absence of direct, large-scale observational constraints. We acknowledge that there are multiple high-quality dust observations (e.g., AOD, surface concentration) other than dust emission fluxes. However, we cannot draw solid, clear conclusions on dust emission itself, particularly its temporal variability, considering that these observations are influenced by numerous factors beyond emissions, such as those affecting dust deposition rates, transport pathways, and atmospheric lifetime.

Please see below our detailed response to other comments.

General Comments

1. As the simulated size distribution and size cut would largely affect the total dust emissions, it is recommended to better reconcile the total dust emissions for a defined size cut for a fair comparison and attributing the physical drivers. As shown in Table 1 in the study, there are large variations in the maximum diameter simulated in different Earth system models. In addition, different size distributions used would induce different total dust emissions even with the same model and dust emission scheme, but there is no information shown in this study about the size distribution of emitted dust.

Response: The models typically calculate a total emission flux (mass per area per time) using a vertical dust flux equation as a function of wind speed, soil moisture, and other soil erodibility factors. The total emission flux is then partitioned into discrete size bins based on assumed emission size distributions. That is, the emitted dust size distribution is not dynamically predicted in the models. The assumed size distributions and upper cutoffs can lead to large differences in the total mass flux, but have limited effects on the interannual variability of the simulated dust emission fluxes, especially how they respond to different drivers represented in the CMIP models (the focus of this study).

We have added below to the revised manuscript (L185).

We use the monthly total dust emission flux as target variable and consider six predictors, including 10-m wind speed, total precipitation (including liquid and solid phases), water content in the uppermost soil layer (hereafter as soil moisture), 2-m specific humidity, 2-m air temperature, and leaf area index (LAI). Here the total dust emission flux is a bulk quantity that represents the source strength. Although ESMs differ in the representation of emitted dust size distribution—a key factor influencing the transport and atmospheric lifetime of dust particles—we expect the particle size partitioning to have limited impact on diagnosing the emission process, particularly its dependence on wind and land surface conditions. The primary drivers of emission variability operate upstream of the size partitioning of the mobilized soil particles.

2. As suggested in this study and various prior studies, different model resolutions could induce substantial and even orders of magnitude differences in total dust emissions. It is recommended to add the information of the model resolution used and potentially reconcile or better separate the differences induced by the model resolution.

Response: We have added model resolutions in Table 1. The models differ in many aspects, such as model physics, dust schemes, parameter uncertainties, and configurations (i.e., horizontal resolution, vertical levels). Together, these differences contribute to inconsistent simulations of dust interannual variability and predictor influence. Separating the effect of model resolution would require experiments using the same model but with different resolutions, which is beyond the scope of this study.

3. It was suggested several times that some models may have intrinsic unreasonable representation in terms of dust emissions, for example in lines 218-223 and 368-370. Why are these models included in the inter-model comparisons for dust in this study? Would it make more sense to exclude them if they showed unreasonable representation of dust emissions?

Response: We aim to include as many models as possible in order to provide a comprehensive assessment of state-of-the-art Earth system models. Because there is no 'ground truth' or validation data for dust emission fluxes, it's almost impossible to determine which model performs the best. Through a multi-model comparison, however, we may be able to identify so-called outliers, for example, models which overestimate hydroclimate influences in hyperarid areas. The intrinsic unreasonable representations mentioned by the reviewer was already known in the community, including the model developers. Here, we just kept it as a sidenote.

4. What are the model performances in terms of total dust burden compared to ground-based and satellite-based observations? Why are the 5 hydroclimate parameters specifically chosen?

Response: We only compared total dust emission fluxes in this study. Evaluating model performance against dust concentration or AOD observations would require separate efforts and are not the focus in this work (please refer to our response above the general comments). We choose the five hydroclimate parameters, because they are either directly used in dust flux calculations or strongly correlated with dust emissions. The choice of predictors is described at L190 of the revised paper.

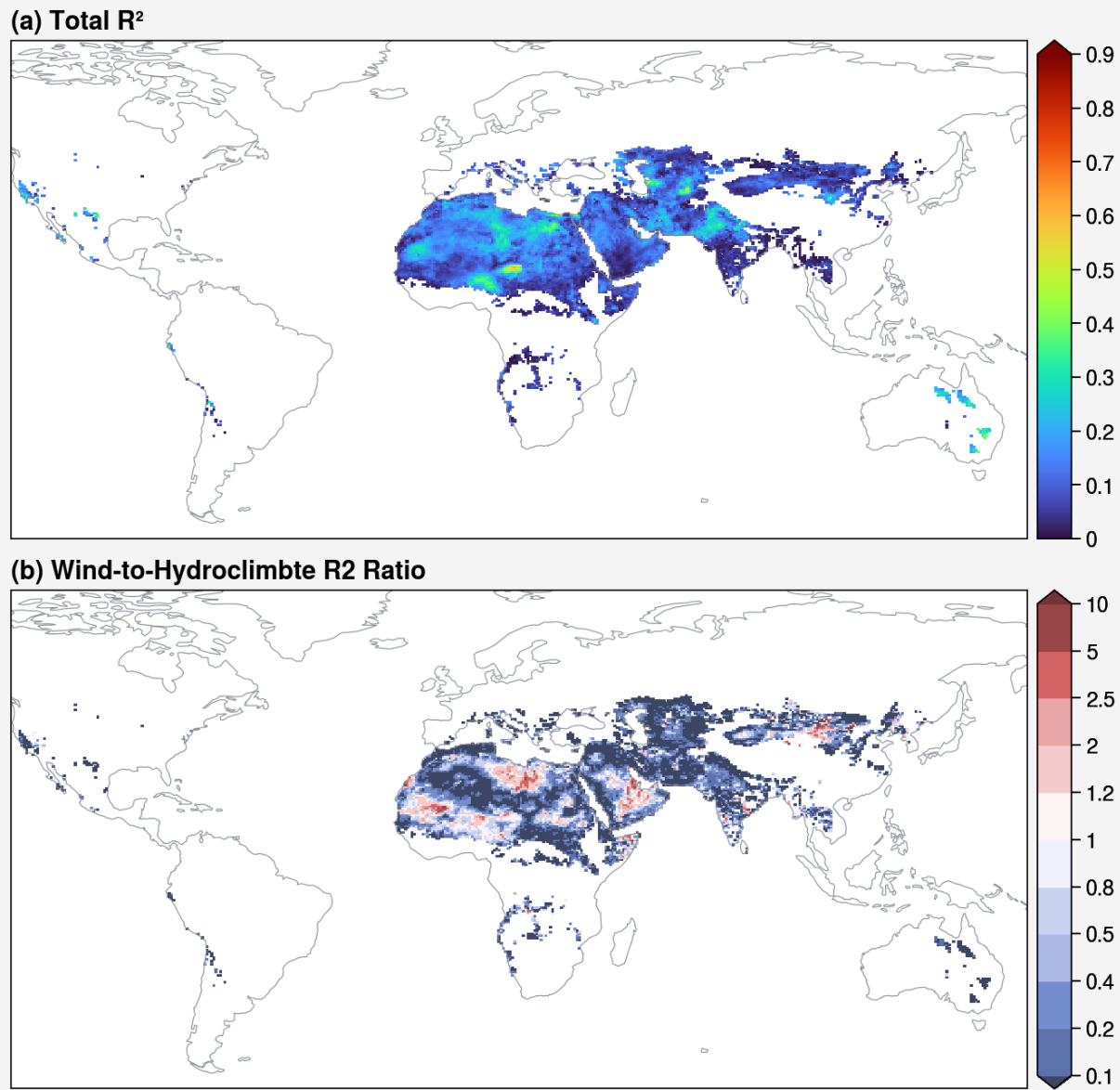
In response to Reviewer #1, we made an attempt to use MODIS deep blue aerosol products to evaluate model representations of dust emissions. However, the results are not encouraging and we decide to not include them in the revision.

We performed dominance analysis using MODIS monthly DAOD (2001-2024) as target variable, and MERRA2 wind speed, precipitation, soil moisture, specific humidity, air temperature, and LAI as predictors. The MODIS DAOD is derived as follows: we first derive daily DAOD using the approach of Li and Ginoux (2025) for both Terra and Aqua MODIS level-2 products. The Terra and Aqua DAODs are averaged for the same day and regridded to 0.65°x0.5° resolution (same as MERRA2) and monthly levels. To isolate the dust sources, we focus on grid cells with at least

14 days of DAOD larger than 0.2, based on the frequency-of-occurrence method in Ginoux et al. (2012).

In Fig. (a) below, the total R^2 or variance explained by the six predictors is very low (<0.3) globally, substantially lower than the R^2 values (>0.8) in MERR2 dust emissions as shown in Fig. 7t from the paper. This indicates weak coupling between DAOD and the reanalyzed wind and hydroclimate variables, and suggests that MODIS DAOD is not a good proxy for dust emission. Furthermore, Fig. (b) below shows the ratio of wind speed-associated R^2 to hydroclimate-associated R^2 , calculated the same way as in Fig. 8 from the paper. We find that North Africa and Arabian Peninsula exhibit extensive areas of R^2 Ratio <0.2 , which suggests that dust emissions from these areas are predominantly controlled by hydroclimate variables, rather than wind speed. This result does not align with the model analysis, and the fact that hyperarid regions are dominated by wind speed due to the permanent dry, barren surfaces and low precipitation variability.

Given the results in Fig. (a) and (b), we conclude that MODIS AOD is not a good proxy for dust emission fluxes and thus decide to not include it into the paper. There are several possible reasons. While dust AOD offers some information on dust source strength, it essentially measures the column dust burden, which mostly comprises fine particles and is influenced by atmospheric transport and removal in addition to emission. There are also other limitations, such as MODIS algorithm biases over bright desert surfaces, limited revisit frequency (only once a day), and cloud interference (i.e., resulting in missing observations), all of which can affect its ability to represent dust source strength.



5. Out of curiosity, what are the annual trends of dust emissions and driving factors for different regions and climate zones? Are regions showing insignificant trend of total dust emissions the regions with larger interannual variability among models?

Response: We did not look at dust trends but focused on dust emission's relationships with physical drivers. The recent study by Kok et al. (2023) found that CMIP6 models produced no significant trends in global dust burden from 1850 to 2000, and suggested that the models failed to reproduce the increased dust burden since preindustrial times, likely because of the model failure to capture historical climate and land-use drivers and/or the dust sensitivity to the drivers. This study addresses the latter issue, that is, the sensitivity of dust emissions to the driving factors.

Specific Comments

1. For Table 1, what are the size distributions of emitted dust, and horizontal resolutions for different models?

Response: We have added model resolutions in Table 1. Because our analysis uses total dust emission fluxes, we only show the size upper cutoff in the table.

We have added below to the revised manuscript (L185).

“the total dust emission flux is a bulk quantity that represents the source strength. Although ESMs differ in the representation of emitted dust size distribution---a key factor influencing the transport and atmospheric lifetime of dust particles---we expect the particle size partitioning to have limited impact on diagnosing the emission process, particularly its dependence on wind and land surface conditions. The primary drivers of emission variability operate upstream of the size partitioning of mobilized soil particles.”

2. Line 111-115: what the rationality of including these additional models or generally what is the criteria for selecting certain models for comparison?

Response: We did not have specific selection criteria, but tried to include as many models as possible in the inter-comparison to provide a comprehensive assessment of state-of-the-art Earth system models. This would help identify outliers or models with different behaviors from the majority. We also incorporated the newly updated CESM and E3SM models because they have implemented revised dust emission schemes that eliminate prescribed source functions and represent emission processes more physically. By comparing paired versions of these models, we wanted to demonstrate, for the first time, how the updated schemes modify dust emission responses to wind speed and hydroclimate drivers.

3. Line 125-127: why are these two reanalysis products chosen? Are they treated as the reference? What are their model performance compared to observations?

Response: JRAero and MERRA2 are both aerosol reanalysis products. We choose them because they provide data on monthly total dust emission fluxes, same as CMIP6 models. The reanalyzed dust emission fluxes are not treated as ‘truth’ or reference, because they are produced by global aerosol models, similar to CMIP6 models. The key difference is that the meteorological, land surface processes, and total aerosol information (e.g., aerosol optical depth) in the reanalysis are constrained by observational data assimilations. But still, dust emissions are simulated with prescribed dust sources and parameterizations. JRAero and MERRA2 also assimilate satellite-derived AOD, which constrains the dust column burden but has limited influences on dust emissions, which has no direct observational constraints.

4. Line 139-141: what is the data source of the aridity index? Are they climatological mean from a certain Earth system model? What are the annual trends of the aridity over these climate zones? Would semiarid zones shift to arid zones over the climatological range considered in this study?

Response: The aridity index is calculated from the 1970–2000 climatological mean precipitation and potential evapotranspiration provided by Zomer et al., (2022). We choose this approach to ensure that climate zones are consistent for all models, allowing a direct comparison of the same dust-emitting regions. We expect the climate aridity to have slight changes from year to year due to precipitation variations. But the aridity index is normally defined using a multi-decadal climatology. We use the climatological aridity index to keep the geographic coverage unchanged while comparing dust emissions from each climate zone.

5. Lin 173: are the target grid cells the same among different models? Do they vary as different models have different grid cells with nonzero dust emissions? If they vary, would the different number and location of target grid cells affect the intercomparisons?

Response: Yes, the number of target grid cells differs among the models, due to different resolutions and dust source area coverage (as shown in Fig. 2). In the inter-comparison, we used the median value of grid cells from a climate zone (Fig. 9) or the statistical distributions of the grid cells (Fig. 10).

6. Line187-190: why do these models use a truncated version of dust source function? How do they compare against observations for dust optical depth? If they show worse results, should they be excluded for the comparisons? As E3SM2-Zender uses the original dust source function and CESM2-Zender uses the truncated dust source function, this would be misleading for the conclusion about the effects of different dust emission schemes shown in line 481-487 as they are different Zender schemes.

Response: Dust models use different dust source functions to reflect the geomorphic, topographic or hydrologic constraints on the spatial pattern of dust emissions. The original Zender scheme uses a continuous dust source function (Zender et al., 2003). When the scheme was implemented in CESM2, the dust source function was truncated for values below 0.1, resulting in discontinuous emission patterns. In contrast, E3SM2-Zender uses the original, continuous source function. While comparing the models, e.g., between CESM2-CAM-Zender and CESM2-CAM-Kok, we used two strategies: all grid cells from each model, and using only common dust-emitting grid cells. We found that the results are very similar.

7. Line 234-236: how about the size distribution in other models?

Response: We have added below to the revised manuscript (L185).

“the total dust emission flux is a bulk quantity that represents the source strength. Although ESMs differ in the representation of emitted dust size distribution---a key factor influencing the transport and atmospheric lifetime of dust particles---we expect the particle size partitioning to have limited impact on diagnosing the emission process, particularly its dependence on wind and land surface conditions. The primary drivers of emission variability operate upstream of the size partitioning of mobilized soil particles.”

8. Line 247-249: how do the winds in MIROS-ES2L and MIROC6 compared to observations?

Response: We did not compare the modeled winds with observations. We compared the MIROC-ES2L and MIROC6 winds in order to understand the causes of their dust emission differences.

9. Line 263: what are their original resolutions?

Response: We have added model resolutions to Table 1.

References

Ginoux, P., Prospero, J. M., Gill, T. E., Hsu, N. C., & Zhao, M. (2012). Global-scale attribution of anthropogenic and natural dust sources and their emission rates based on MODIS Deep Blue aerosol products. *Reviews of Geophysics*, 50(3).

Kok, J. F., Storelvmo, T., Karydis, V. A., Adebisi, A. A., Mahowald, N. M., Evan, A. T., ... & Leung, D. M. (2023). Mineral dust aerosol impacts on global climate and climate change. *Nature Reviews Earth & Environment*, 4(2), 71-86.

Li, X., & Ginoux, P. (2025). An empirical parameterization to separate coarse and fine mode aerosol optical depth over land. *Geophysical Research Letters*, 52(6), e2024GL114397.

Zender, C. S., Newman, D., & Torres, O. (2003). Spatial heterogeneity in aeolian erodibility: Uniform, topographic, geomorphic, and hydrologic hypotheses. *Journal of Geophysical Research: Atmospheres*, 108(D17).

Zhao, A., Ryder, C. L., & Wilcox, L. J. (2022). How well do the CMIP6 models simulate dust aerosols?. *Atmospheric Chemistry and Physics*, 22(3), 2095-2119.

Zomer, R. J., Xu, J., & Trabucco, A. (2022). Version 3 of the global aridity index and potential evapotranspiration database. *Scientific Data*, 9(1), 409.