Due to the recent U.S. government shutdown, Referee #1 was unable to submit their comments during the formal review period. They have now followed up with additional comments, I kindly request that you address them in a revised manuscript, so the published paper avoids potential misunderstandings with model developers and is strengthened for the community.

Response: Sorry about this unfortunate situation. We are very grateful for the editor's efforts in compiling the new comments and coordinating this new round of revision.

1. Incorporate or acknowledge available observational constraints used in dust-model intercomparisons
The reviewer reiterates that most dust–emission intercomparison studies incorporate some type of observational reference — whether satellite AOD, in-situ flux data, or campaign-based measurements (e.g. FRAGMENT field campaigns). They suggest the manuscript would benefit from acknowledging the existence of such datasets and briefly explaining why they were not used here. Since you previously explored MODIS AOD and found it unsuitable for your purpose, you may wish to expand the explanation and comment on the limitations or relevance of other observational datasets.

Response: During the first-round review, we used more than two decades' MODIS Level-2 deep blue aerosol records to compare with the ESMs and MERRA-2 reanalysis. We used the Li and Ginoux (2025) approach to isolate dust from total AOD and the Ginoux et al. (2012) approach to isolate regions frequently affected by persistent dust loadings. As seen in Fig. 1 below, the total R^2 or variance in MODIS dust AOD (or DAOD) explained by the six predictors is very low globally, compared to the R^2 values (>0.8) in MERRA-2 dust emissions (see Fig. 2 below). Our analysis shows that the MODIS DAOD record is not a good proxy for dust emission fluxes for temporal variability analysis, probably because DAOD is affected not only by emission, but also by atmospheric transport, dry/wet removal, and size partitioning processes.

Granted dust emission flux measurements are available from field campaigns, these measurements are from only a few locations and short time periods, making them unsuitable for comparison with multidecadal global model simulations considered in this study.

In the revised manuscript (Lines 92-97) we acknowledged the existence of satellite AOD and in situ data and explained why they are not used in this study:
"Although satellite-derived dust AOD and in-situ dust measurements provide valuable constraints on dust variability (e.g., Prospero and Lamb, 2003; Voss and Evan, 2020), they integrate the effects of emission, transport, and deposition, making it difficult to isolate the emission process itself. Also, due to lack of global validation data, we focus on diagnosing inter-model inconsistency in representing the dust emission variability

and its physical controls, rather than validating individual model performance against observations."

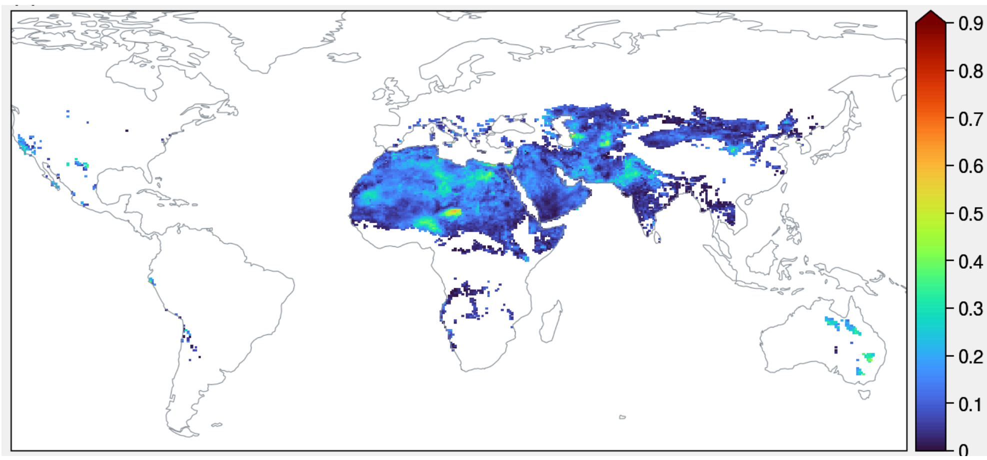

Figure 1. Total R^2 explained by MERRA-2 wind speed and hydroclimate drivers in the MODIS DAOD.

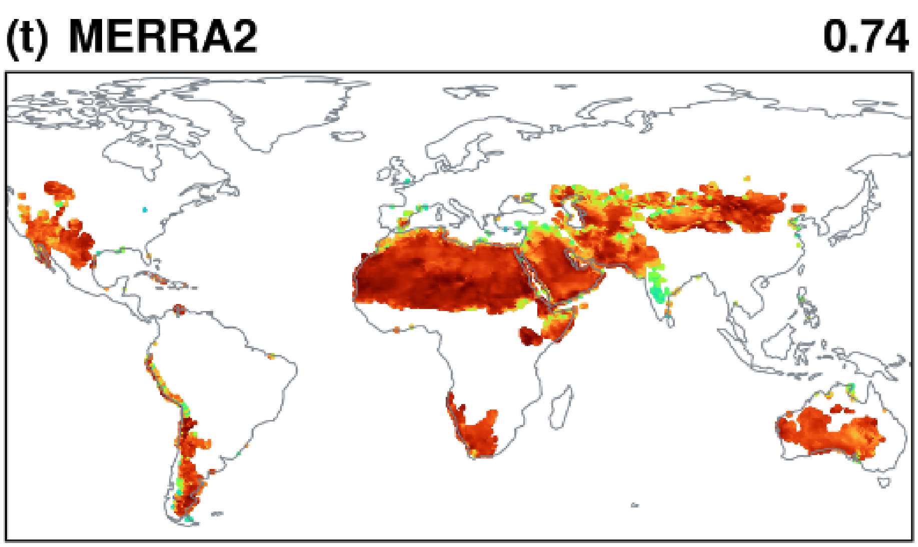

Figure 2. Total R^2 explained by MERRA-2 wind speed and hydroclimate drivers in the MERRA-2 dust emission fluxes. Copied from Fig. 7t in the manuscript.

2. Clarify/Remove the interpretation of GFDL-ESM4 dust-emission sensitivities
The reviewer emphasizes that several statements in the manuscript may mischaracterize dust-emission drivers in GFDL-ESM4. They point specifically to Section 2.5 of Shevliakova et al. (2024), which states that soil-moisture dependence is disabled in the CMIP6 version of GFDL-ESM4. This contradicts the interpretation in Line 470 that "anomalous sensitivities to soil

moisture" explain model behavior. They also stress that the GFDL scheme contains additional controlling factors beyond wind, soil moisture, and LAI, including both LAI (green vegetation) and SAI (brown vegetation), which can introduce low-frequency variability. For these reasons, the reviewer strongly recommends that all material related to GFDL-ESM4 be removed as doing so would be simpler than revising the manuscript in depth. If you decide not to fully remove the GFDL-ESM4 results, please provide sufficient justification. In my view, the bottom line is to remove any discussions and conclusions attributing the GFDL-ESM4 behavior to soil-moisture sensitivity.

Response: Thanks for providing the additional information about GFDL-ESM4. The dominance analysis is purely statistically driven, and independent from the physical linkages between dust emission and the considered drivers. In the case of GFDL-ESM4, since the soil moisture effect is disabled in the model, the strong hydroclimate influence from dominance analysis is very likely driven by spurious statistical covariations between the model-simulated dust emission fluxes and the topsoil water content. Indeed, Shevliakova et al. (2024) reported that GFDL-ESM4 overestimates the top-layer (0-5 cm) soil water content over global drylands compared with satellite estimates, with positive biases of up to 100% in regions such as central North Africa, South Africa, northwestern China, southwestern USA and Australia (see their Fig.9). This overestimated soil moisture may covary with the dust emission fluxes, even if soil moisture is not used in the dust parameterizations in GFDL-ESM4. To avoid misunderstandings and per the reviewer's request, we decided to remove GFDL-ESM4 from the paper. In the revision, we removed all content related to GFDL-ESM4, and added CanESM5.0, alongside CanESM5.1 which was already included.

In addition, we acknowledge that we did not consider all the physical drivers represented in the model, but only focus on six common physical drivers in order to ensure a fair comparison of all ESMs. This is now clarified in the revised paper as follows.

At Lines 171-174, we added:
"Note that we donot not include all the physical drivers represented in each model because of limited data availability in the CMIP6 online archive, and because some models incorporate additional drivers not used by others. Hence we focus on a common set of six readily available predictors to provide a consistent and fair comparison across the ESMs and MERRA-2 reanalysis."

At Lines 426-430 (Section 4 Conclusions), we added:
"Note that the physical drivers considered in this study may not fully represent all the dust emission driving factors for specific emission schemes; instead, we focus on a common set of drivers for all models to provide a fair comparison across the ESMs.

Therefore, the inferred relative importance from this analysis is limited to those common drivers considered and their influences on dust emissions in different models. Also, because of the statistical nature of dominance analysis, the predictor importance results shall be interpreted with caution when linking to model parameterizations."

3. Acknowledge other region-specific drivers
The reviewer notes that factors such as SAI, land-use change, and related vegetation dynamics play important roles in regions such as the Sahel, India, Australia, and the western United States. They recommend that the manuscript explicitly acknowledge these influences and clarify that the dominance analysis does not include all possible drivers in each Earth System Model.

Response: We acknowledge that our analysis does not consider all the physical drivers represented in every model, but focuses on six common physical drivers in order to provide a fair comparison of all ESMs. Also, not all the physical drivers (such as stem area index) are available in the CMIP6 online archive. This is now clarified in the revised paper as follows.

At Lines 171-174, we added:
"Note that we donot not include all the physical drivers represented in each model because of limited data availability in the CMIP6 online archive, and because some models incorporate additional drivers not used by others. Hence we focus on a common set of six readily available predictors to provide a consistent and fair comparison across the ESMs and MERRA-2 reanalysis."

At Lines 426-430 (Section 4 Conclusions), we added:
"Note that the physical drivers considered in this study may not fully represent all the dust emission driving factors for specific emission schemes; instead, we focus on a common set of drivers for all models to provide a fair comparison across the ESMs. Therefore, the inferred relative importance from this analysis is limited to those common drivers considered and their influences on dust emissions in different models. Also, because of the statistical nature of dominance analysis, the predictor importance results shall be interpreted with caution when linking to model parameterizations."

4. Concern about broader mischaracterization of multiple models
The reviewer mentions that GFDL-ESM4 is not the only ESM potentially mischaracterized in the manuscript. They fear this could create long-term confusion in the community about how dust emissions are represented in specific models. Because contacting ESM model developers at this stage may be too late, the reviewer urges careful removal or significant softening of model-specific conclusions unless supported directly by model documentation.
I appreciate your patience and understanding with this unusual late-review situation. Please feel free to reach out if you need clarification on any of the reviewer's points.

Response: We have removed all content related to GFDL-ESM4. We also carefully revised the paper by cross-referencing with original model documentations. In the revision, we also suggest caution in interpreting the statistically inferred predictor influences to model parameterizations.

At Lines 426-430 (Section 4 Conclusions), we added:
"Note that the physical drivers considered in this study may not fully represent all the dust emission driving factors for specific emission schemes; instead, we focus on a common set of drivers for all models to provide a fair comparison across the ESMs. Therefore, the inferred relative importance from this analysis is limited to those common drivers considered and their influences on dust emissions in different models. Also, because of the statistical nature of dominance analysis, the predictor importance results shall be interpreted with caution when linking to model parameterizations."

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).

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.