We thank the reviewer for a speeding and thoughtful review. We appreciate the opportunity to clarify the novelty and distinct contributions of our study.

We agree with the reviewer that our study is a model inter-comparison. However, our approach differs substantially from previous studies. As the reviewer correctly pointed out, prior model intercomparison studies often rely on observational datasets (e.g., satellite-derived AOD) to evaluate or correlate with model simulations. While dust AOD (or other observations) offers some insight into the source strength, it primarily reflects the total atmospheric dust burden, which is influenced by transport and removal processes. Its limitations—such as retrieval challenges over bright desert surfaces, limited revisit frequency, and cloud interference—can further affect its ability to represent dust source activity.

In contrast, our study focuses on dust emissions only. As discussed in the Introduction, dust emission fluxes simulated by Earth system models are inherently a model-specific quantity, governed by internal processes such as dust emission parameterization, PBL, land surface and other coupled processes. Fig. 4 to Fig. 6 illustrate the inconsistent variability in dust emission simulations, highlighting a key motivation for our study. A key originality of our analysis lies in the examination of internal model drivers, rather than relying on external observational datasets. This approach enables us to probe the internal causal mechanisms within each model.

Moreover, our study introduces a novel spatial framework by partitioning the global domain into three distinct climate regimes. This approach reveals how model consistency varies systematically with climatic aridity. Particularly, the near-surface wind speed has been shown as the primary controlling factor over hyperarid regions, such as the Sahara. One might expect Earth system models to reproduce this feature reliably, however, our analysis reveals some models do not. Thus, despite being a model intercomparison, our analysis exposes fundamental biases in the dust emission representations in some models.

Another unique contribution of our study is the inclusion of a broader set of physical drivers than typically considered, combined with the use of dominance analysis to disentangle the individual contributions of correlated predictors. This method accounts for multicollinearity among drivers and quantitatively attributes their relative importance in explaining dust variability.

Finally, our study incorporates a wider range of Earth system models than previous studies. In particular, we include E3SMv2 and E3SMv3, which have not been compared within the CMIP6 framework, and an updated version of CESM2 with revised dust emission scheme and mineralogy. The paired model experiments with Zender and Kok schemes in E3SM and CESM allow us to examine how the same dust emission scheme performs in different ESM, and how different dust emission schemes perform in the same ESM.

We hope this response clarifies the key distinctions and contributions of our study. Below, we provide specific replies to the reviewer's additional comments.

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.

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).

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).

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.

Response: The Introduction section aims to give a broad overview of inconsistent dust emission representations in ESMs, as a main motivation for our inter-comparison approach. We may have misinterpretations or miss important details .Can you please clarify what we may have missed?

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: Thanks for pointing out the correct references. When choosing the references, we try our best to use the paper with most details on dust emission representations. We will update the references based on your recommendations.

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: Thanks for pointing out the additional model differences. We will add more details in the Table to highlight the dust emission differences. As the reviewer pointed out, the models are different in many aspects. This is exactly the reason why we approach the inter-comparison by treating dust emission as a model-specific quantity, and focusing on the internal variability and predictors, instead of using external observations.

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