Replies to the comments from Anonymous Referee #1

We would like to thank Anonymous Referee #1 for their constructive and insightful review. Their suggestions regarding the need for clarifications the evaluation metrics and the methodology have significantly enhanced the quality of the manuscript and are greatly appreciated. We have copied the comments into this document; the comments are in Times New Roman blue font while our answers are in Calibri black font. Line numbers refer to the revised version of the manuscript without track changes.

The article addresses a topic of interest for the dust and climate communities. It presents the improvements included in the COSMO-MUSCAT model to better characterize the size distribution of airborne minerals from arid regions and assesses their impact as compared to previous model versions. This will inform the scientific community on the ability of state-of-the-art models to reproduce a critical characteristic that shapes dust – Earth system interactions, its mineralogy. The article also presents two recently conducted campaigns that provide mineral fractions and elemental composition data, which constitute valuable assets for model evaluation. However, the mineral composition derived from one of the campaigns shows large uncertainty ranges for some minerals and sizes. While this aspect is acknowledged, it is not thoroughly considered in the model-observation comparison, hence weakening the validity of the conclusions. Additionally, the suggested reasons for model-observation discrepancies neglect in some cases relevant aspects (e.g. the assumptions made to derive the elemental composition, or the ability of the model to represent the dust cycle processes). Finally, some methodological aspects are not clear, which makes it difficult to properly interpret the results.

In my view, the article presents valuable information, but it still requires major revisions prior to its publication. I summarize below some of the suggested changes.

We thank the reviewer for this comment. The methodology has been substantially revised including incorporating iron content into the smectite conversion, which has affected the results and their analysis accordingly. Measurement uncertainties are now addressed more rigorously throughout the manuscript and are explicitly reported in the linear fit equations presented in Tables 5, 6, and 7. Additionally, the reasons for the model-observation discrepancies have been expanded.

***** General comments *****

1. Clarification of methodological aspects:
The method followed to calculate and compare modelled and observed mineral and elemental fractions needs clarification, particularly regarding:

The mass fraction calculation: How do the authors compute the modelled mass fraction of different minerals to compare with observations? Is this the mass fraction over total dust? How do the authors handle scenarios where certain minerals are measured in the field but not simulated in the model?

This is now explicitly described in the methodology Section 4.1 in the paragraph beginning at line 476.

The size range included: For elemental composition, BIN26 is explicitly mentioned as the upper size range for evaluation. Is this also the case for the mineralogy comparison? If not, how is the unspecified mineral fraction in BIN80 treated when calculating mineral mass fractions?

BIN26 is also the upper size range used for the mineralogy comparison. The unspecified mineral fraction in BIN80 is therefore not included in the analysis and does not require treatment in the calculation of mineral mass fractions.

The assumptions behind the elemental - mineral conversions. The authors apply different values for the elemental composition of dust minerals when deriving mineral composition from the SEM analyses (Table A) versus determining elemental abundances from the modelled minerals (Table 3). These choices impact the evaluation results, and the underlying reasoning must be clearly justified.

The differences between the two approaches are now explicitly stated and justified in Section 4.3.2 (line 599). In addition, the elemental percentages for smectite and feldspar have been updated to better reflect the underlying assumptions, and the reasoning for these choices is discussed therein.

2. Mineralogy from JATAC 2002 and treatment of observational uncertainty
The mineral fractions obtained from the JATAC 2002 campaign hold for some minerals and sizes large uncertainties. How is this information incorporated into the analysis?

All measurements are retained in the analysis; however, no conclusions are drawn in cases where uncertainties are too large. Furthermore, uncertainty-weighted linear fits are now reported in the results tables and applied throughout the subsequent analysis.

For instance, several measurements in the JATAC campaign show a BIN26 clay fraction of nearly 100%. Are these values physically realistic? What is the corresponding amount of other minerals, e.g., quartz, at these specific measurement points?

For illite, the coarser bins show amounts up to 70%. In addition, the variations of illite fraction with size seem small (Figure 6), and they seem to tend toward higher mass fractions in larger particles, which is unexpected for a typically clay-sized mineral. The authors recognize challenges to identify illite in the SEM-EDX measurements (L662-664), however, the associated measurement uncertainty is not considered in the evaluation beyond this comment. Do the estimated mineral fractions add to 100% in each observational point? Does this limitation in illite detection affect the mass fractions of other minerals?

All mineral fractions are normalized to sum to 100% at each observational point, which can be verified in the zenodo data repository. This normalization means that elevated illite contents inevitably affect the fractions of other minerals at those specific measurement points.
The elevated illite fractions observed in coarser size bins may be explained by the platelet formation and stable aggregation of illite in emitted dust. In soil analyzes, the

substrate is frequently treated mechanically (e.g. by sieving) and/or disaggregated in water or alcohol, which could have a higher tendency of breaking up the platelets and aggregates than the natural wind-driven disaggregation has. SEM-EDX images of such particles are provided in Appendix C.

3. Evaluation procedure and interpretation

Why is the comparison with the mineral measurement compilation restricted only to the winter period (Section 4.2)? According to Table 2, both model versions were run for both winter and summer, and then the reason given in lines 488–491 does not hold.

The simulation period was retained as is. Although some measurements in the compilation were taken during Northern Hemispheric summer, these were matched to simulated dust events originating from the same source regions and following similar transport pathways, as described in the "dust event analysis" in Section 4.2. This event-based matching approach ensures comparability between observations and simulations, making an additional simulation unnecessary.

Comparing mean biases and correlation values across experiments that rely on different samples does not seem fair (Table 4, L609-620). I would recommend comparing metrics that are calculated over the same measurement points.

We agree with this comment. The measurements haven been consequently grouped by similarity in surrounding soil and transport pathways, as described in Section 4.2 (line 556). Within each group, metrics are calculated over the same set of measurements, ensuring a fair comparison. The grouped results are presented in Table 4.

In some sections, potential causes for discrepancies between model and observations are given, but these point many times to the soil mineralogy or the particle size distribution. However, the actual ability of the model to represent the specific regions where the dust plumes originated, their transport and deposition processes influence the modelled minerals (and elements) at the time-and-location of the measurements.

We agree that the model's ability to accurately represent dust source regions, transport pathways, and deposition processes constitutes and additional potential source of discrepancy, and this has now been acknowledged in the relevant sections. To provide context for the model's performance in this regard, we also note that the simulations have been validated against dust aerosol optical thickness (AOT) from AERONET stations in previous studies, lending confidence to the representation of dust transport and loading along the modelled pathways.

***** Detailed comments *****

Some of the sentences are vague and require further clarification. Examples are provided below, but I would recommend reviewing carefully the text, if possible, with the objective of synthetising it more and being more specific.

We thank the reviewer for the suggestion. We have carefully revised the manuscript with the aim of making the text more concise and specific throughout, paying particular attention to the examples highlighted below.

L14–15 (Abstract): "The elemental validation approach provides complementary constraints that expose discrepancies in internal mixing assumptions and reveal limitations invisible to mineral-only comparisons". This is very difficult to follow unless the reader has already gone through the paper. I would recommend rephrasing for clarity.

We agree this sentence needed more clarity, and therefore it has been revised (line 13).

L74-76: "The performance of the two schemes is assessed against a multi-level observational dataset. We first utilized a compilation of regional North African mineralogical measurements of in-situ aerosol samples to evaluate broad mineralogical patterns and enable a direct comparison between the implemented schemes. This is followed by a high-resolution comparison with concurrent in-situ observations from two campaigns in Cabo Verde [...]" The authors use three different datasets, why are these qualified as multi-level?. What do "broad mineralogical patterns" refer to? Are these geographical patterns? Does this dataset allow direct comparison between the two schemes but not the data from the campaigns? What does the "high-resolution" refer to? High temporal frequency? High spatial resolution?

We agree that the term "multi-level" was unclear and the description of the differences between observational dataset was confusing. The text has been revised to explicitly clarify the nature of each dataset, the revised text can be found starting at line 73.

L84–85: The sentence "The validation results are presented through a tiered approach categorized by analytical scale" is vague and only becomes clear later in the text.

We agree that the original sentence was too vague. It has been replaced with a more explicit description of the structure of the results section (line 83): "Following this, the validation results are presented in two steps: Section 5.1 addresses mineralogical composition, beginning with the regional North African compilation and followed by a detailed assessment of the JATAC 2022 results; Section 5.2 then evaluates elemental composition against both JATAC 2022 and DUSTRISK 2022 campaigns."

L388-391: "In this study, one of the main objectives is to accurately represent the aerosol dust mineralogy in future modeling studies using measurement data. Hence, we adopt a classification scheme for mineral dust aerosols representative of the soil database as used in the model." This is difficult to follow. How do the authors plan to use the measurements to accurately represent dust mineralogy? What does the classification scheme refer to? Please, rephrase and clarify.

The revised text (line 388) clarifies that the measurement classification was adapted to match the mineral groups used in GMINER, ensuring consistency between the observational and modeled mineral categories and enabling a direct comparison between the two.

L631-635: The authors describe measurements taken close to dust sources as having an "observational bias." If the model's resolution is too coarse to capture localized features, this is a model limitation, not a measurement bias.

We agree with the reviewer that this was incorrectly framed. The text has been revised accordingly (line 662).

L48–49 defines 2.5 µm as the size boundary between clay and silt, but in L254 this is set at to 2 µm. Please remain consistent.

Thank you for pointing this out. The line 254 now changed to the 2.5 micrometer boundary.

L208–210: Dc needs to be explicitly defined within the main text.

We have explicitly defined it in line 209.

L256: Typo; change "thse" to "these".

Changed, thank you (line 256).

Section 2.4: This section contains redundant details regarding soil texture and soil mineral data sources that were already detailed in Section 2.3.

We agree, the redundant details have been removed and the section condensed accordingly (lines 356-358).

L509: The text mentions Panta et al. (2023), but the compilation described up to this point has only mentioned data from Perlwitz et al. (2015), which did not include Panta's results. I suggest reorganizing this section by moving the details currently in L528–530 to an earlier position.

The section has been reorganized so that the full measurement compilation, including Panta et al. (2023) data, is described together from the outset (line 521), in line with the reviewer's suggestion.

L611–612: The comment regarding the scope of feldspar fails to clarify what is actually occurring in Figure 5, which depicts mass fractions without size resolution. If the evaluation is done collocating model and observations in size, please, detail it in the methods.

The feldspar analysis has been revised following the introduction of the measurement grouping, and an updated explanation and analysis are provided in the paragraph beginning at line 646. Regarding the size co-location, this is indeed performed in the evaluation and is now more explicitly detailed in the methods section (line 546).

L615: Does the underlying soil mineralogy distribution actually change here? This information should theoretically come from the soil mineralogy map, which is supposed to be identical for both modeling approaches. Please clarify.

We agree. The underlying soil mineralogy map, GMINER, does not change for both modelling approaches. This statement has been removed.

L662–664: What is meant by a "localized observational trend"?

This refers to local dust emission events that could be captured in the observations but cannot be resolved by the model due to its coarse spatial resolution. The phrasing has been revised throughout to convey this more clearly.

L675: "This suggests that shifts in the mineral PSD do not necessarily align with the composition of the parent soil." The statement here is unclear. Why would be the changes in the PSD aligned with the composition of the parent soil? Please elaborate or rephrase.

We agree the original statement was unclear. The revised text (line 733) clarifies that while phyllosilicate abundance is based on the parent soil distribution from GMINER, this does not necessarily translate directly to coarser aggregates at the same location. Additional potential sources for the low correlations are also discussed, including differences between modeled and measured particles properties and the possibility that the model does not capture the exact dust event responsible for the observations.

L670–680 What do the correlation values represent? The value of r could be very high even if the bias or the slope of the linear fit is profoundly off. The low correlation could stem from source soil mineralogy, but it could also reflect the model's inability to capture specific transport and deposition patterns. How well does the model simulate dust itself? While these transport factors might average out in a climatological comparison, they gain relevance here since the authors are attempting to simulate hourly variations in mineral composition during an airborne campaign.

We agree with these observations. Uncertainty-weighted linear fits are now reported for all JATAC 2022 and DUSTRISK 2022 comparisons, providing a more complete picture of model performance.
Regarding the role of transport and deposition, we acknowledge that these factors are particularly relevant in the context of the JATAC 2022 campaign, where the goal is to capture hourly variations on mineral composition. The model has been previously validated against AERONET AOT for both simulation periods and has been shown to represent dust load reasonably well, providing confidence in the simulated transport. This has now been explicitly referenced in the text (line 745). Nevertheless, we acknowledge that inaccuracies in the model dust transport may contribute to the observed discrepancies for individual measurements and this has been noted in the analysis.

L735: Consider filtering the data to remove contamination from non-dust sources.

The data has been filtered to remove contamination from non-dust sources. However, some residual contamination cannot be entirely excluded.

L763–764: "This suggests that while the mass balance is improved, the inherently weak correlation, likely driven by the influence of non-simulated minerals, is further exacerbated by the proposed mass redistribution." This sentence is difficult to understand. Please rewrite.

We agree the original sentence was unclear. The revised text (line 846) clarifies that while the mass redistribution improves the overall bias, it does not improve the spatial correlation. We suggest this may be due to contributions from non-simulated Al-bearing minerals, whose are not including in the model and therefore cannot be corrected through mass redistribution.

L765-773: This discussion is sensitive to the chosen assumptions regarding the iron content within the simulated minerals.

We agree that the results are sensitive to the assumed iron content of the simulated minerals. Following the revision of the iron composition in smectite, the comparison has improved, as now reflected in the updated analysis. Additionally, a sensitivity study in which smectite is fully replaced by nontronite has been included (line 850).

L770: Note that not all iron present in minerals exists in the form of iron oxides. This distinction should be accurately reflected in the text.

We agree, and the text has been revised to acknowledge that iron content is present in minerals beyond iron oxides. The updated iron analyses for both campaigns are presented at line 850 for JATAC 2022 and line 931 for DUSTRISK 2022.

L877–L896: I think this discussion would fit best in the Conclusions section.

We agree and have move this discussion to the Conclusions section accordingly.