Reply to the referee's comments

We thank the referee for the useful comments, which helped us to improve the quality of our manuscript.

In the following, the referee's comments are given in black.

Our point-to-point replies are marked by "R" and are in blue.

Changes to the manuscript text are in green.

Comments by referee #1

General comments:

This manuscript presents a valuable multi-instrument evaluation of Saharan dust transport simulated by ICON-ART and contains potentially important results concerning the respective roles of meteorological initialization, horizontal resolution, and aerosol size representation. In particular, the combination of vertically resolved lidar measurements with in-situ particle-size observations provides a useful opportunity to distinguish transport-related errors from aerosol-process errors. However, I do not think that several of the manuscript's main conclusions are yet fully supported by the current validation framework. In particular, the added value attributed to convection-permitting resolution appears minor and is largely demonstrated for a single event; the meteorological fields responsible for the claimed improvement after re-initialization are not directly evaluated; and some attribution of model biases to specific physical processes remains speculative. The manuscript would also benefit substantially from restructuring and comprehensive language editing. I therefore recommend major revision.

R: We thank the reviewer for the careful assessment and constructive recommendations. We appreciate the recognition of the value of combining vertically resolved lidar measurements with in situ particle-size observations. We agree that the conclusions should more clearly distinguish directly evaluated improvements from possible physical explanations. In response, we have revised the manuscript in the following main respects:

1. **Added direct meteorological validation.** We compared simulated temperature, wind speed, and wind direction profiles with radiosonde observations at eight radionsonde stations. The revised manuscript includes vertical bias profiles with interstation standard deviations and time series of station-mean full-column correlation, RMSE, and bias. These comparisons show that daily reinitialization generally reduces temperature and wind-profile errors, providing observational support for improved meteorology as a contributor to the better dust predictions.

2. **Qualified the conclusions concerning horizontal resolution.** We now explicitly restrict the conclusions from the systematic resolution experiments to the February 2021 event and emphasize that daily reinitialization produces a larger improvement than increasing resolution alone. The additional benefits of the kilometer-scale simulations are described as comparatively modest and localized, including improvements in selected lower-tropospheric profiles and reduced excessive backscatter in elevated layers under RI-EXP-1d. We also explicitly acknowledge that finer resolution does not consistently improve meteorological diagnostics.

3. **Reorganized the Results and strengthened the quantitative presentation.** We substantially rewrote the event descriptions to follow a clearer progression from observed evolution and model evaluation to the principal discrepancy, diagnostic evidence, and interpretation. Revised paragraph openings highlight the main findings, and the distinction between meteorological/transport errors and aerosol-representation effects is introduced earlier in the Results. Quantitative comparisons, including correlations, regression slopes, RMSE, bias, and variability across stations or height levels, replace or qualify several broad statements of agreement.

4. **Revised the language and presentation.** We corrected grammatical problems, removed duplicated instrument descriptions, and improved the wording and transitions, particularly in the case-study Results. The Discussion and Conclusions have been revised to reflect the scope of the sensitivity experiments and the strength of the available evidence.

These revisions strengthen the observational basis for the meteorological improvements, clarify the conditional benefits of finer resolution, and make the distinction between demonstrated results and physical interpretations more explicit. Detailed responses to each comment are provided below.

Scientific concerns:

1. Meteorological validation is surprisingly limited given the central role of meteorology in dust emission, transport and deposition. For example, a major conclusion is that daily meteorological re-initialization substantially improves dust prediction, increasing near-surface PM correlations from approximately 0.65–0.69 to 0.86–0.88. However, if the proposed mechanism is improved meteorology, then the meteorology itself should be validated. For example, please compare EXP-5d and RI-EXP-1d against observations/reanalysis for quantities directly relevant to vertical dust transport, e.g. 10-m wind / wind profiles, boundary-layer height, temperature, humidity, precipitation, cloud properties, and ideally vertical velocity or synoptic circulation. Without this step, the paper demonstrates that re-initialization improves dust, but does not fully demonstrate why. Statements such as errors in wind field, boundary-layer depth, or vertical mixing causing the differences are physically plausible, but presently remain interpretations rather than directly validated mechanisms. This additional validation would substantially strengthen the central scientific message.

R: We thank the reviewer for highlighting the need to validate the meteorological fields underlying the improved dust predictions. In the revised manuscript, we have added comparisons of simulated temperature, wind speed, and wind direction profiles against radiosonde observations at eight meteorological stations. Figure 10 presents station-mean bias profiles and their interstation standard deviations for the 13 and 1 km simulations on 22 and 25 February. A supplementary figure (Figure S6) shows the evolution of full-column RMSE, correlation, and bias from 21 to 26 February, calculated separately at each station and then averaged with equal station weights. The results show that daily reinitialization generally improves agreement with the radiosonde observations, with larger differences between the experiments emerging later in the simulation. For example, at 12:00 UTC on 25 February, reinitialization reduces the 13 km station-mean temperature RMSE from 1.89 to 1.07 °C, wind-speed RMSE from 4.61 to 2.18 $\mathrm{m\,s^{-1}}$, and wind-direction RMSE from 53.2° to 32.5°. The corresponding mean wind-speed and circular wind-direction correlations increase from 0.61 to 0.84 and from 0.43 to 0.75, respectively. These comparisons provide direct observational support for improved meteorology as a contributor to the better dust predictions. We also find that finer resolution does not consistently improve meteorological performance and discuss adjustments associated with nesting as a possible explanation, rather than an established cause. We have added following text in manuscript.

The EXP-5d and RI-EXP-1d simulations were conducted to investigate the effects of meteorological reinitialization and grid resolution on Saharan dust predictions. To evaluate the meteorological fields underlying these predictions, simulated temperature, wind speed, and wind direction profiles were compared with radiosonde observations at eight meteorological stations. Temperature profiles characterize the thermal structure relevant to atmospheric stability, while wind profiles constrain the horizontal transport of dust at different altitudes. Figure 10 shows station-mean model-minus-observation biases between 1000 and 300 hPa for the 26 and 1 km simulations at 12:00 UTC on 22 and 25 February 2021. Errors were calculated separately at each station before averaging, and the error bars indicate ±1 standard deviation across valid stations. Complementary full-column statistics were calculated individually for each station and then averaged with equal station weights to obtain the time series in figure S6. The two experiments show similar performance early in the simulation, whereas larger differences emerge later. By 25 February, EXP-5d exhibits pronounced wind-direction biases near 500 hPa and in parts of the lower troposphere, together with larger wind-speed biases at several levels; these departures are generally reduced in RI-EXP-1d (Fig. 10). The full-column statistics support this improvement. At 12:00 UTC on 25 February, daily reinitialization reduces the station-mean temperature RMSE from 1.89 to 1.07 °C and wind-speed RMSE from 4.61 to 2.18 $\mathrm{m\,s^{-1}}$ at 13 km, while the mean wind-speed correlation increases from 0.61 to 0.84. The mean wind-direction RMSE decreases from 53.2° to 32.5°, accompanied by an increase in the mean circular correlation from 0.43 to 0.75. Although RI-EXP-1d yields higher correlation coefficients, the near-unity full-column temperature correlations primarily reflect the shared vertical temperature structure and should therefore be interpreted alongside RMSE and bias. Increasing horizontal resolution does not consistently improve agreement with the radiosonde observations. The 26 and 13 km simulations generally perform better than the 2 and 1 km simulations in these station-averaged diagnostics, particularly for RI-EXP-1d (Figure S6). One possible explanation is that transferring meteorological fields from the parent domains to the finer nests introduces interpolation errors or adjustments as the flow adapts to the finer grids and their representation of topography. Such effects could contribute to the temperature and wind discrepancies in the nested simulations, although their contribution has not been isolated here. These meteorological errors could, in turn, affect predicted dust transport pathways and concentrations, potentially limiting the benefits of increased horizontal resolution. Nevertheless, the improved temperature and wind profiles in RI-EXP-1d provide observational support for improved meteorology as a contributor to its better dust predictions.
The lidar-observed and ICON-ART-simulated evolution of Saharan dust over KIT-CN for the EXP-5d and RI-EXP-1d experiments, using the 26 km nest of the global configuration and the 1 km nest of the limited-area

model (LAM) configuration, is shown in Figure 11. Results from the 13 and 2 km simulations are shown in Figure S7. It has to be noted that the effective resolution of resolved processes by the model is still around 5-6 horizontal resolution. At the synoptic scale, both experiments reproduce the timing of the Saharan dust outbreak and its large-scale transport pathway from North Africa toward Central Europe, whereas more pronounced differences emerge in the vertical dust distribution and near-surface concentrations. Compared with the continuous 5-day forecast (EXP-5d), the daily re-initialized simulations (RI-EXP-1d) provide a closer representation of the observed dust-layer heights, vertical backscatter structure, and near-surface PM_{10} variability. In particular, across all model resolutions, the correlation between simulated and observed near-surface PM_{10} concentrations increases from $R = 0.65$–0.69 in EXP-5d to $R = 0.86$–0.88 in RI-EXP-1d, while the RMSE decreases by approximately one third.

The improvement in RI-EXP-1d is consistent with a reduction in accumulated forecast errors through daily re-initialization as shown in figure 10 and figure S6, although the mechanism differs between the global and limited-area simulations. In the global simulations, re-initialization updates both the meteorological state and the dust fields, which can introduce discontinuities in the simulated dust distribution at the re-initialization times. In contrast, the limited-area model (LAM) simulations are not directly re-initialized internally and are affected primarily through the updated lateral boundary conditions. Consequently, the improvement in the LAM simulations is more directly associated with the repeated refresh of the meteorological forcing and boundary conditions. By constraining the large-scale circulation, frontal evolution, and associated vertical motions more frequently, daily re-initialization limits the accumulation of meteorological forecast errors and thereby improves the simulated transport and near-surface evolution of the dust plume. Despite the improved meteorological forcing, differences between the simulations and observations remain. This is expected because the re-initialization procedure only constrains meteorological variables, while no corresponding assimilation is available for mineral dust tracers. Consequently, uncertainties associated with dust emission, deposition, and aerosol transport processes continue to evolve freely within the model. These uncertainties can accumulate during transport from the Saharan source regions and may contribute to discrepancies in both dust concentration and vertical distribution, even when the meteorological conditions are well represented.

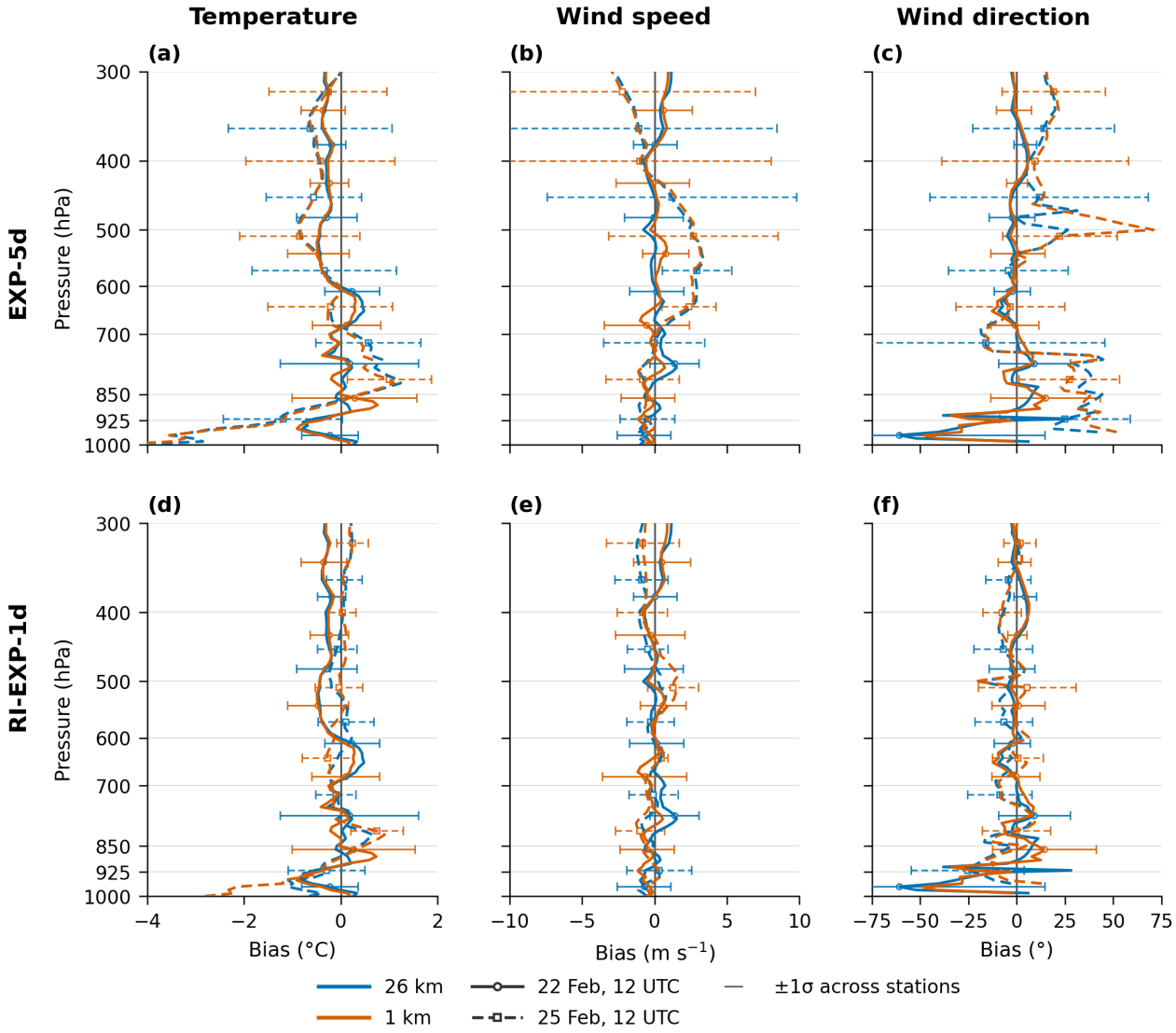


Figure 10: Vertical profiles of model-minus-radiosonde biases in temperature (a, d), wind speed (b, e), and wind direction (c, f) for EXP-5d (top row) and RI-EXP-1d (bottom row). Blue and orange indicate the 26 and 1 km simulations, respectively; solid and dashed lines indicate 22 and 25 February 2021 at 12:00 UTC. Errors are calculated separately at each station before averaging across the valid stations among the eight selected sites.

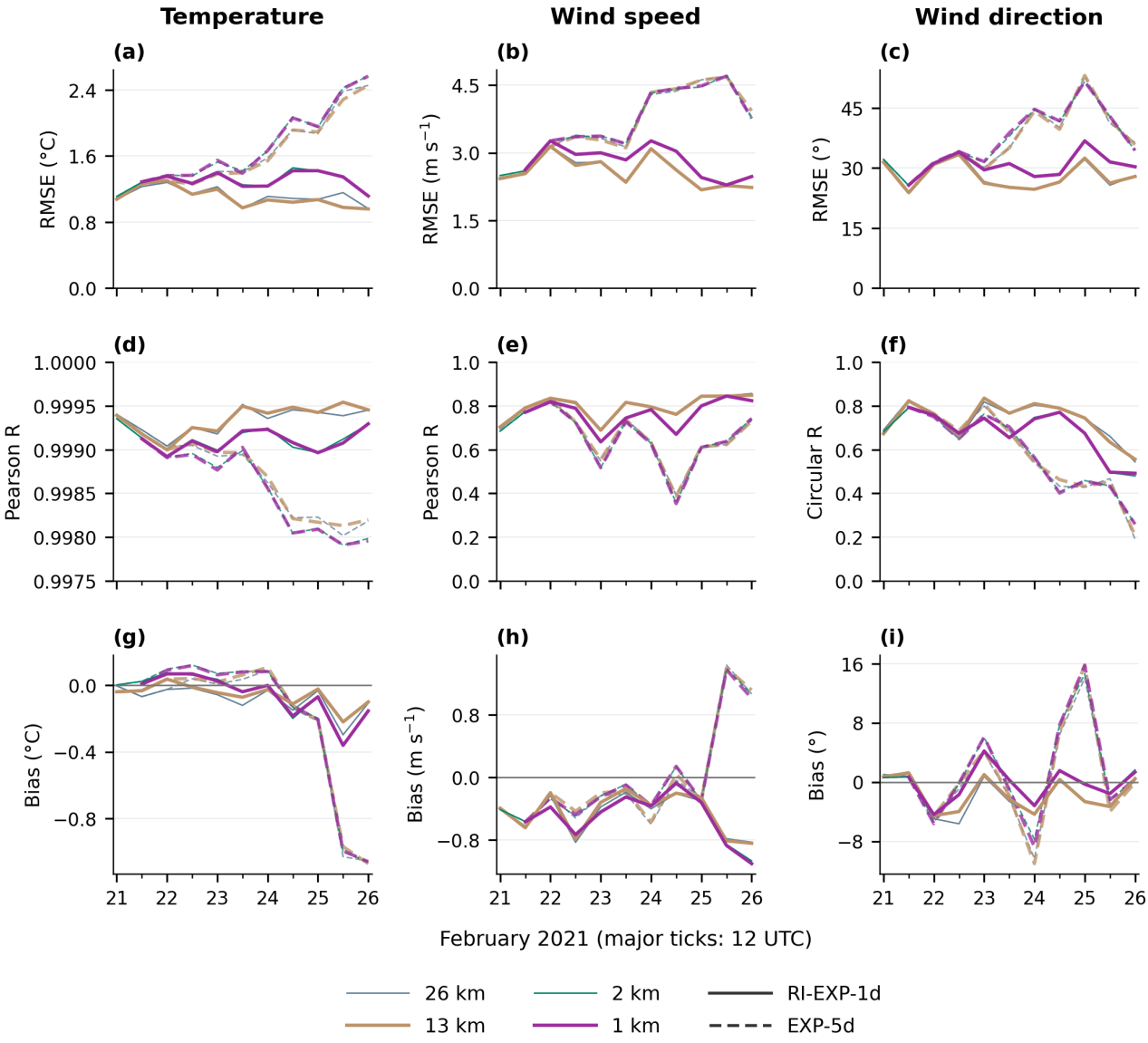


Figure S6: Time series of station-mean full-column RMSE (a–c), correlation coefficient R (d–f), and bias (g–i) for temperature (left column), wind speed (middle column), and wind direction (right column), from 21 February 2021 at 12:00 UTC to 26 February 2021 at 12:00 UTC.

2. The evidence for the added value of convection-permitting resolution is not yet fully convincing. The paper concludes that convection-permitting simulations better represent layer stratification, vertical gradients, and free-troposphere–boundary-layer exchange. However, the systematic resolution experiments appear to have been conducted only for the February 2021 event, whereas the four-event evaluation mostly uses the coarser operational configuration. And the changes from low- to high-resolution runs appear very minor to me (Figs. 11, 12). Furthermore, the manuscript itself shows that changing meteorological initialization produces a larger improvement in some metrics than changing horizontal resolution. For AERONET AOD, for example, statistics are described as relatively insensitive to resolution from 26 to 1 km.

R: We thank the reviewer for this comment. We agree that the added value of convection-permitting resolution should be distinguished from the larger improvement associated with daily reinitialization. We also acknowledge that the systematic resolution experiments are restricted to the February 2021 event; the four-event evaluation does not establish a general advantage of convection-permitting resolution across all events. Our interpretation of the resolution sensitivity therefore applies specifically to the February 2021 case.
The manuscript explicitly identifies meteorological reinitialization as the dominant source of improvement in several places. In Section 3.3, the discussion of Figure 11 reports that daily reinitialization increases the near-surface PM_{10} correlations from $R = 0.65$–0.69 to $R = 0.86$–0.88 across all resolutions, while reducing RMSE by approximately one third (p. 23, lines 441–450). We subsequently state that "the influence of horizontal resolution on the overall dust evolution is relatively modest" (pp. 23–24, lines 459–475) and that "forecast strategy has a stronger influence on the simulated dust evolution than horizontal resolution alone" (p. 25, lines 476–481). Thus, we agree with the reviewer that the broadly similar time–height structures in Figure 11 do not demonstrate a major improvement in large-scale dust transport from resolution alone.
The additional benefit of the finer grids concerns particular features of the vertical dust distribution, rather than a uniform improvement in all diagnostics. As described on p. 24, lines 465–467, the 2 and 1 km RI-EXP-1d simulations better reproduce the observed distribution below approximately 800 m in the selected profiles and reduce the excessive backscatter in the elevated layers at 6–8 km (Figure 12). A related improvement is described for the CALIPSO comparison: the kilometer-scale RI-EXP-1d simulations reduce the excessive backscatter at 7–8 km relative to the coarser simulations, whereas this resolution dependence is not apparent in EXP-5d (p. 26, lines 493–498; Figure S9). These results support a conditional and vertically localized benefit of finer resolution when combined with updated boundary forcing.
The AERONET results in Figure 15 are consistent with this qualified interpretation. The manuscript explicitly states that the statistics depend more strongly on reinitialization than on horizontal resolution and remain relatively insensitive to grid spacing within each experiment (pp. 28–29, lines 515–529). Nevertheless, within RI-EXP-1d, the 1 km simulation provides a modest additional improvement: its RMSE is 0.128, compared with 0.138–0.139 at the other resolutions, and its mean bias is 0.027, compared with 0.036–0.042. Its correlation is also slightly higher ($r = 0.656$), and its regression slope is closer to unity (0.972).
Taken together, the results support daily reinitialization as the principal improvement in this case, with finer resolution providing additional benefits for selected aspects of the vertical dust structure. We acknowledge that the statement about improved free-troposphere–boundary-layer exchange is stronger than can be established directly from the available diagnostics: the figures evaluate dust distributions, rather than vertical exchange rates. A more precise interpretation is that the improved vertical structure is consistent with changes in transport and redistribution, while the underlying exchange mechanisms have not been isolated.

3. Writing and presentation: The manuscript contains interesting results, but the English currently makes the science harder to follow than necessary. I recommend a thorough language edit. There are numerous grammatical problems even in the Introduction and Methods. Examples include phrases such as "Saharan dust outbreaks has impact," "A scanning aerosol lidar systems," "Case 1 was detailed discussed," and duplicated sentences describing the CPC instruments. More importantly, I would recommend a structural rewrite, not just grammatical correction. Much of Section 3 follows a repetitive pattern: observation shows X → model shows X → possible reason → next variable → observation shows Y → model shows Y. This makes the central findings difficult to extract. A clearer structure would be: Observed event evolution → model evaluation → identified discrepancy → diagnostic evidence → interpretation. Please always summarize the main idea in the opening sentence of each paragraph. The Discussion is stronger than the earlier Results because it begins synthesizing the errors into meteorological/transport versus aerosol-process components. I would bring this conceptual framework forward and organize the Results around it. I would also reduce repeated qualitative terms such as "good agreement," "relatively good," "consistent," "very well," and "reasonable agreement." Whenever possible, replace them with quantitative statements. For example, rather than saying "ICON-ART predicts dust layer heights very well," report the bias, MAE/RMSE, correlation, and uncertainty.

R: We thank the reviewer for these constructive suggestions. We agree that clearer language and a more diagnostic presentation are needed to make the central findings easier to follow. We have revised the language and presentation, with particular attention to the event descriptions in Section 3.1, and removed the duplicated description of the CPC instruments in Section 2.2. We have also corrected specific grammatical problems, including the singular form of "a scanning aerosol lidar system" and the description of Case 1 as having been

previously examined by Zhang et al. (2022).
Beyond grammatical corrections, we have substantially rewritten the case-study descriptions to distinguish the observed event evolution, model performance, identified discrepancies, diagnostic evidence, and interpretation. The revised opening of the February 2021 case, for example, identifies the persistent plume, its descent toward the surface, and the predominantly cloud-free conditions as the defining features of the event. The subsequent model-evaluation paragraph states the principal finding at the outset: ICON-ART reproduces the timing and broad vertical evolution, but larger discrepancies emerge as dust becomes mixed into the lower atmosphere. Similarly, the revised opening of the June 2021 case identifies clouds and precipitation as the main controls on its evolution and explains the resulting limitations of the lidar retrievals.
We have also brought the distinction between meteorological/transport errors and aerosol-representation effects into the Results. In the February case, the apparent discrepancy in lower-atmospheric backscatter is examined using the separation of dust and non-dust contributions, rather than being attributed immediately to a transport error. The comparison shows that the KASCAL total-aerosol backscatter exceeds the simulated dust backscatter by a factor of 1.94 ± 0.65. After isolating the observed dust contribution, the regression slope is 0.974 ± 0.01, although the correlation remains moderate at $R = 0.543$. The revised text therefore distinguishes agreement in backscatter magnitude from agreement in its detailed temporal and vertical variability. Section 3.3 additionally evaluates temperature and wind profiles against radiosondes before interpreting the effects of reinitialization and resolution on the dust simulations.
To reduce reliance on qualitative assessments, the revised Results provide explicit numerical comparisons where available. These include regression slopes and correlations for the dust-backscatter comparison, RMSE and bias for the meteorological and satellite comparisons, and correlation, regression slope, RMSE, and mean bias for AERONET AOD. We also distinguish the implications of these metrics: for example, the AOD analysis explicitly notes that improved correlation and a slope closer to unity do not necessarily imply a reduction in RMSE. These changes make the evidence supporting each interpretation more explicit and help connect the event-specific Results to the synthesis of meteorological/transport and aerosol-process limitations in the Discussion.

Minor suggestions:

1. The authors have referred Saharan dust to SAL throughout the manuscript. Yet, my impression is that when people refer to SAL they mostly mean Saharan Air Layer over the Atlantic Ocean. Please double check if it's appropriate to call it SAL when the dust is transported to Europe. Meanwhile, please write out the full name of SAL when it firstly appears in the manuscript.

R: Thank you for pointing this out. We have reviewed the manuscript and replaced "SAL" with "dust layer" as appropriate. We have also ensured that all abbreviations are defined at their first occurrence.

2. The color scheme in Figure 1 is not very informative.

R: We have changed Figure 1 as following to better demonstrate the Frequency of trajectories.

3. Please indicate the sample size in Figure 3.

R:We have added the sample size in Figure 3.

4. Meteorological fields are missing in Figure 6 but discussed in the text:

R: Thank you for pointing out it. We have added the meteorological field in Figure 6 as following.

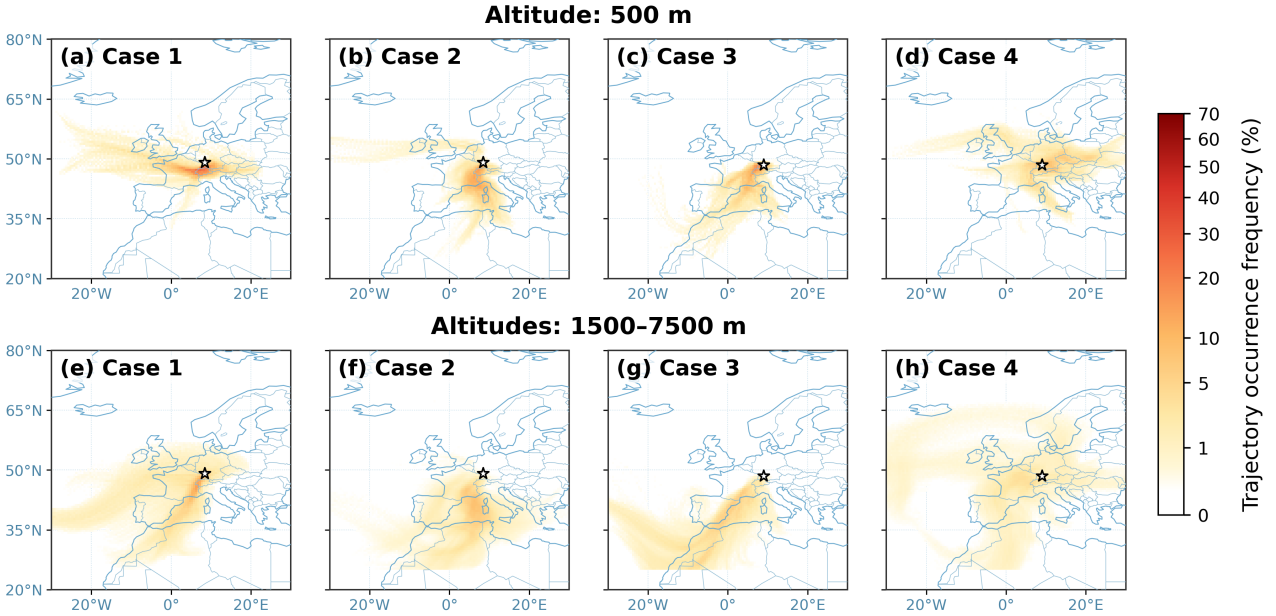


Figure 1: Frequency of trajectories at two altitude levels (500 m and 1500 - 7500 m) during four Saharan dust case periods based on back trajectories calculated by the HYbrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model. (The case IDs are labeled in Table 1)

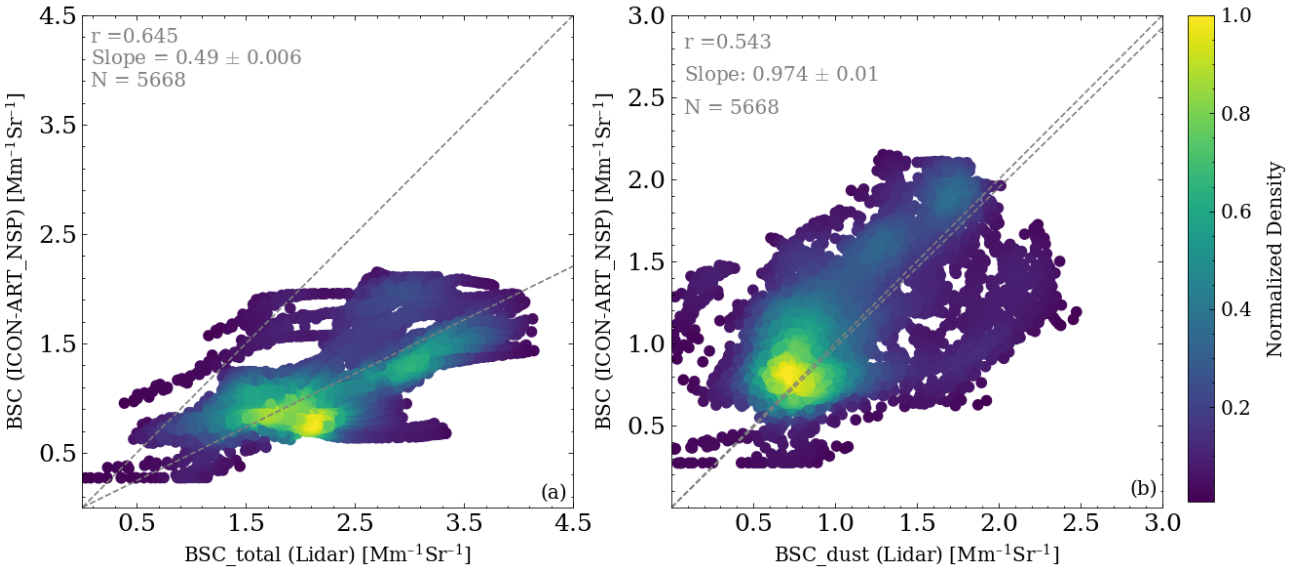


Figure 3: Correlation of backscatter coefficients predicted by ICON-ART and those retrieved from KASCAL for both total aerosol particles (a) and dust particles only (b) (from 20:00 on 24^{th} to 12:00 to 26^{th} February 2021 for Case 2).

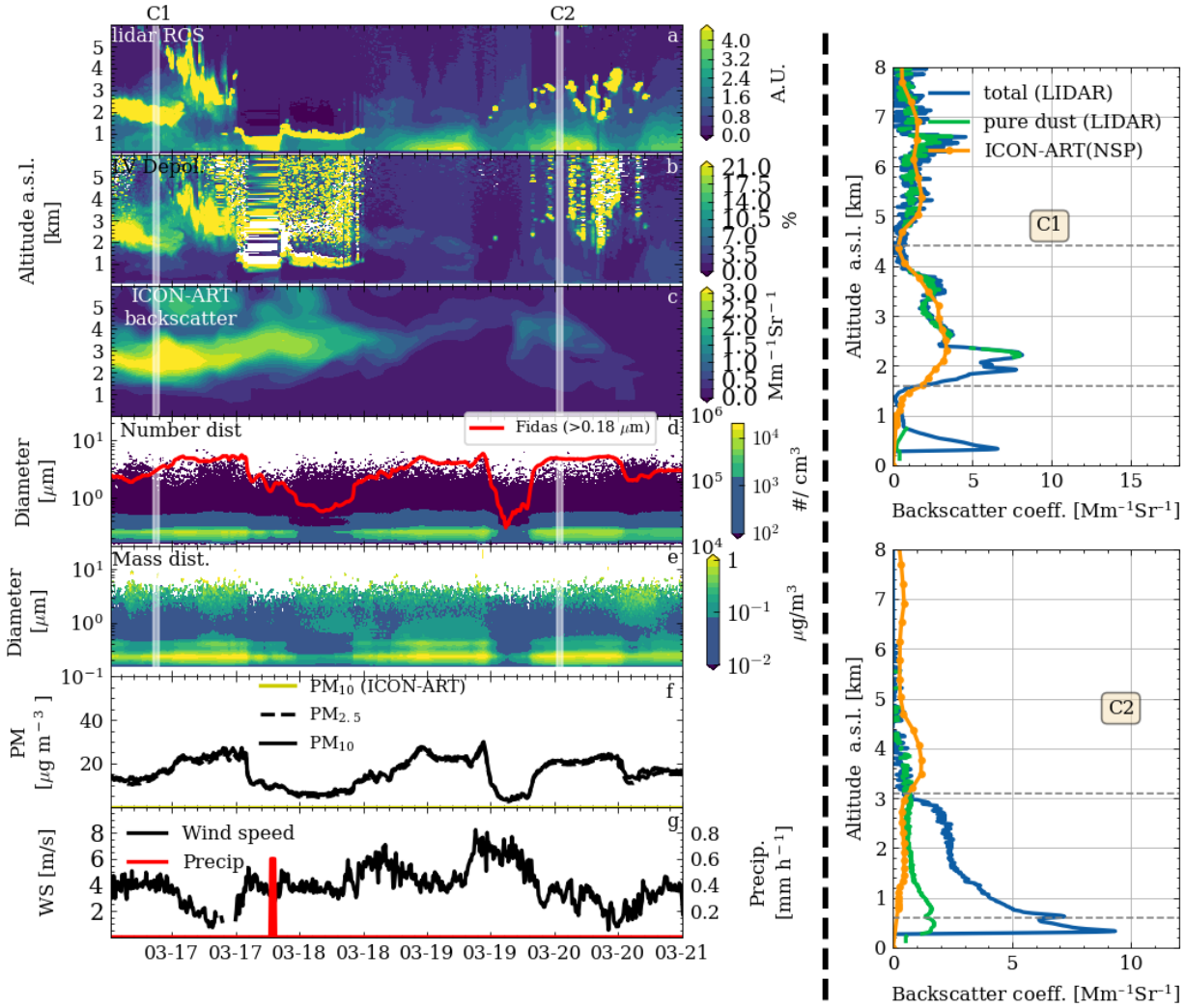


Figure 6: Time series of range corrected lidar signal for cross-polarized channel (a), and Linear Volume Depolarization Ratio (LVDR) (b) from KASCAL measurements, backscatter coefficients predicted by ICON-ART (c), particle number concentration (d), particle mass size distribution from Fidas 200 (e), particulate matter concentrations $PM_{2.5}$ & PM_{10} from the Fidas 200 and the lowest model level (150 m) of the ICON-ART simulation (f), and ground level meteorological parameters (e.g. wind, precipitation, WS700) (g) from 16^{th} to 21^{th} March 2022 for case 4. The profiles of backscatter coefficients measured by the two lidars and predicted by ICON-ART model for C1 case (Averaged profiles from 20:30 to 21:30 for lidar measurement and profile at 21:00 for ICON-ART prediction on 16^{th} March 2022) and for C2 case (Averaged profiles from 00:30 to 01:30 for lidar measurement and profile at 01:00 for ICON-ART prediction on 20 th March 2022) are shown on the right side of this figure. C1 and C2 represent time periods used for a more detailed data analysis as shown in the right side of this figure.