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 #2

General comments:

The authors present 4 Saharan dust events observed over Southern Germany with lidar, sun photometer and ground-based in situ instrumentation. They compare these observations to the results of the dust transport model ICON-ART and found periods of good and bad agreement between model and observations. Generally, it is important to check the performance of aerosol transport models with observations. After reading the manuscript, I asked myself: What do I know now which I haven't known before? This questions already outlines the lack of novelty or the author's ability to convey a clear message to the reader. That is why I would refrain from publishing the manuscript in ACP. Below, I list some detailed points for my decision.

R: We thank the reviewer for this thoughtful comment. We agree that the scientific contribution of the original manuscript was not sufficiently emphasized and that the key messages were not clearly communicated. In the revised manuscript, we have substantially expanded both the analysis and the discussion to better highlight the scientific value of the study. While model validation is an important component of the work, the revised manuscript goes beyond a simple model–observation comparison. The study now provides several new insights into the long-range transport of Saharan dust over Central Europe.

First, we introduce a multi-resolution evaluation of ICON-ART using simulations at 26 km, 13 km, 2 km, and 1 km resolution. The results demonstrate that kilometer-scale simulations provide added value in representing dust-layer stratification and vertical gradients, which cannot be fully captured by the coarser simulations. To our knowledge, such a detailed evaluation of the resolution dependence of Saharan dust transport using this combination of observations and ICON-ART simulations has not previously been presented. (section 3.3)

Second, we introduced a comparison between two forecast strategies: a continuous 5-day forecast experiment (EXP-5d) and a daily re-initialized experiment (RI-EXP-1d). The new analysis demonstrates that daily re-initialization improves the representation of the observed dust-layer evolution by reducing the accumulation of meteorological forecast errors. This improvement is particularly evident in the convection-permitting simulations (2 km and 1 km), which produce less dust in the elevated layer and a more realistic vertical dust distribution. The convection-permitting simulations exhibit stronger resolved downdrafts, suggesting enhanced downward transport of dust from the free troposphere into the boundary layer. These results highlight the combined importance of accurate meteorological initialization and convection-permitting resolution for reproducing the vertical evolution of long-range transported Saharan dust. (Section 3.3)

Third, the combination of ground-based lidar observations and CALIPSO measurements provides complementary constraints on the vertical structure and optical properties of the Saharan dust plume. While the ground-based lidar at 355 nm shows that ICON-ART generally overestimates the dust backscatter coefficient, the CALIPSO observations at 532 nm show substantially better agreement with the simulations. This wavelength-dependent model performance, together with the consistency between the two independent observing systems, suggests that the remaining discrepancies are more likely related to uncertainties in the simulated dust particle size distribution and optical properties than to observational limitations. In particular, the underestimation of coarse particles and overestimation of intermediate-sized particles in ICON-ART may lead to an excessively strong wavelength dependence of the simulated backscatter coefficient. (Section 3.3, Figure 12 and 13)

Fourth, We have added a new evaluation of the simulated dust AOD against AERONET coarse-mode AOD using observations from 17 stations within the innermost model domain. The comparison covers both EXP-5d and RI-EXP-1d at horizontal grid spacings of 26, 13, 2, and 1 km. The results show that daily re-initialization improves the correlation and brings the regression slope closer to unity, whereas the AOD statistics vary only slightly with grid spacing. This suggests that the remaining discrepancies are primarily related to uncertainties in coarse-particle mass, size distribution, and associated optical properties rather than to horizontal resolution alone. (Section 3.3, Figure 14)

Fifth, we added a detailed analysis of aerosol mass size distributions measured by the Fidas 200 optical particle spectrometer and compared them with the simulated dust size distributions. The comparison with aerosol particle size measurements reveals a systematic bias in the simulated size distributions. ICON-ART consistently overestimates the mass concentration of particles with diameters between approximately 0.5 and 2 μm while underestimating particles larger than 2 μm. The persistence of this bias in both the EXP-5d and RI-EXP-1d experiments suggests that it is primarily related to the representation of size-dependent dust transport and removal processes rather than meteorological initialization. In particular, the underestimation of coarse particles indicates that large dust particles are removed too efficiently during long-range transport, likely because of excessive sedimentation and deposition. A similar conclusion was reached by Ratcliffe et al. (2025), who showed that additional processes counteracting the rapid sedimentation of coarse dust particles are required for climate models to reproduce observations. These size-dependent biases are also likely to contribute to the remaining discrepancies in the simulated optical properties and vertical dust structure. (Section 3.3, Figure 15 and S9)

We therefore believe that the revised manuscript contributes not only to model evaluation but also to a better understanding of the processes controlling Saharan dust transport, vertical redistribution, and observational detectability over Europe. We have revised the abstract, introduction, discussion, and conclusions to better emphasize these scientific findings and improve the clarity of the manuscript's main messages.

Specific comments:

1. Sometimes the model agrees, sometimes not. This is not a surprise, but besides speculations about possible reasons, the authors have not made the attempt to test different model parameterizations. Applying different dust emissions schemes, deposition parameterizations etc., one could detect reasons for agreement or disagreement. Alternatively, one could use different dust transport models to see which features lead to good/bad agreement in the different models (e.g., the mentioned model resolution). However, the authors have just used a single model setting of ICON-ART. In that sense, the paper remains merely descriptive and leads to the general statement that the model sometimes agrees and sometimes not.

R: We thank the reviewer for this important comment. We agree that sensitivity experiments using different dust emission schemes, deposition parameterizations, or alternative transport models would provide additional insight into the sources of model uncertainty. However, these investigations are beyond the scope of the present study and will be reported in a forthcoming publication. We would like to emphasize that the revised manuscript now extends beyond a simple description of model agreement and disagreement. In addition to the original simulations, we have incorporated two complementary sensitivity analyses that allow us to identify important sources of uncertainty in the model representation of Saharan dust.

First, we compare simulations at four different horizontal resolutions (26 km, 13 km, 2 km, and 1 km). This analysis demonstrates that model resolution has a measurable impact on the representation of dust-layer stratification and vertical gradients, with the kilometer-scale simulations providing a more realistic depiction of the observed dust structures.

Second, we compare a continuous 5-day forecast experiment (EXP-5d) with a daily re-initialized experiment (RI-EXP-1d). This comparison reveals that meteorological initialization strongly affects the simulated dust evolution and that accumulated forecast errors can significantly influence the vertical distribution of transported dust.

Furthermore, the combination of ground-based lidar observations, Fidas 200 size-distribution measurements, and independent CALIPSO observations enables us to distinguish between discrepancies related to meteorological forcing, model resolution, observational limitations, and dust process representation. For example, the persistence of size-distribution biases in both EXP-5d and RI-EXP-1d suggests that these discrepancies are more likely associated with dust emission, sedimentation, and deposition processes than with meteorological initialization.

We therefore agree that additional parameterization sensitivity experiments would be valuable future work. Nevertheless, we believe that the revised manuscript provides new process-level insights into the roles of meteorological initialization, model resolution, and dust process representation in determining the accuracy of Saharan dust simulations, rather than merely documenting instances of model agreement and disagreement.

2. The manuscript lacks to my opinion the focus. Is it a description of Saharan dust intrusions to Central/West Europe? Or is it the comparison to ICON-ART? The mere description of Saharan dust episodes in Europe is not new. The literature from the last 3 decades is full of such events (please have a look) – so what do we learn from these events presented here? What is the benefit for the community to read about these 4 events in ACP? If the focus is on the model comparison, then there are more sophisticated approaches already published – also here a deeper look into existing literature would be beneficial to sharpen the focus of a possible re-submission.

R: We agree that the focus of the original manuscript was not sufficiently clear and that the scientific message was not adequately emphasized. In the revised manuscript, we have substantially revised the abstract, introduction, discussion, and conclusions to better define the scope and objectives of the study. The primary purpose of this work is not to provide a descriptive account of four Saharan dust events over Europe. We agree that numerous studies have documented Saharan dust intrusions into Europe over the past decades. Rather, the focus of the revised manuscript is to use a unique combination of ground-based lidar observations, sun photometer measurements, aerosol size distributions (Fidas 200), CALIPSO satellite observations, and ICON-ART simulations to investigate the processes controlling the representation of Saharan dust in atmospheric transport models. The scientific contribution of the revised manuscript lies in three aspects. First, we investigate the influence of model resolution using ICON-ART simulations at 26 km, 13 km, 2 km, and 1 km resolution. This analysis demonstrates the added value of kilometer-scale simulations for reproducing observed dust-layer stratification and vertical gradients. Second, we assess the impact of meteorological initialization through a comparison between a continuous 5-day forecast experiment (EXP-5d) and a daily re-initialized experiment (RI-EXP-1d), providing new insight into the role of meteorological forcing in long-range dust transport simulations. Third, the combined use of ground-based lidar and CALIPSO observations allows us to identify cases where apparent model–observation discrepancies are related to observational limitations, such as signal attenuation in the ground-based lidar, rather than deficiencies in the model itself. Furthermore, the inclusion of aerosol size-distribution measurements enables us to identify systematic biases in the simulated particle-size spectra, suggesting that uncertainties in dust emission, sedimentation, and deposition processes remain an important source of model error. These findings provide process-based information that extends beyond a simple assessment of model agreement or disagreement.
To better reflect these objectives, we have revised the manuscript to emphasize the roles of meteorological initialization, model resolution, observational constraints, and dust process representation. We believe that the revised manuscript now provides a clearer scientific focus and contributes to a better understanding of the factors controlling the accuracy of Saharan dust simulations over Europe.

1. We have changed our title as following:

Saharan dust transport and boundary layer entrainment over Western Europe: multi-instrument observations and convection-permitting model evaluation

2. We have changed our abstract as following:

Long-range Saharan dust transport affects radiation, clouds, and air quality, yet its vertical structure and entrainment into the boundary layer remain challenging to simulate. We investigate four contrasting dust events over southwestern Germany using ground-based lidar, sun-photometer and in situ aerosol measurements, satellite observations, backward trajectories, and ICON-ART simulations. The events represent diverse atmospheric conditions, including a six-day cloud-free episode, a cloudy episode involving dust–cloud interactions, and an intense outbreak with aerosol optical depth approaching 1.4. A depolarization-based lidar method separated dust backscatter from contributions by other aerosols, enabling a like-for-like evaluation of the dust represented by ICON-ART and providing a useful approach for evaluating dust models in mixed aerosol layers. ICON-ART reproduced the large-scale transport and arrival of the dust layers, with mean differences of approximately 20 min in arrival time and 50 m in layer height. Daily re-initialization of the global simulations improved correlations with observed near-surface particulate matter concentrations from 0.65–0.69 to 0.86–0.88 and better reproduced the timing and persistence of boundary-layer dust. The high resolution runs (2, 1 km) better resolved dust-layer stratification, vertical gradients, and exchange between the free troposphere and boundary layer more realistically. However, ICON-ART consistently overestimated particles of 0.5–2 µm and underestimated particles larger than 2 µm. These results demonstrate the complementary benefits of accurate meteorological initialization and convection-permitting resolution while identifying size-dependent emission, sedimentation, and deposition as priorities for model development. By distinguishing errors in dust transport from those arising from aerosol size representation, this study provides an observation-constrained framework for evaluating and improving convection-permitting dust forecasts.

3. We have clarifyed our objectives in introduction as follows:

The objectives of this study are to (1) characterize the transport pathways, vertical structure, optical properties, and size distributions of Saharan dust using a unique combination of ground-based, in-situ, and satellite observations; (2) evaluate the ability of ICON-ART to reproduce these observed dust characteristics; and (3) investigate the influence of model resolution and meteorological initialization on the representation of long-range transported Saharan dust over Europe.

4. We have revised our discussion as follows:

The comparison of observations with ICON-ART simulations for four Saharan dust events and two sensitivity experiments provides insight into the main sources of model-observation discrepancies. Three main findings emerge. First, the large-scale transport pathway and plume arrival time are reproduced consistently across all model resolutions and meteorological initialization strategies, with a mean arrival-time difference of approximately 20 min even at 20 km horizontal resolution. Second, the aerosol loading and vertical distribution of dust are strongly influenced by both horizontal resolution and meteorological initialization, with the best agreement achieved only when both are improved simultaneously. Third, the particle size distribution and some systematic features of the optical properties remain largely unchanged across all experiments, suggesting that these persistent discrepancies originate from the aerosol representation rather than from the transport itself.
The improvement in the simulated vertical dust structure is one of the main findings of this study. Compared with the continuous 5-day forecast (EXP-5d), the daily re-initialized forecast (RI-EXP-1d) increases the correlation between simulated and observed near-surface PM_{10} concentrations from $R = 0.65$–0.69 to $R = 0.86$–0.88 across all model resolutions, while reducing the RMSE by approximately one third. The RI-EXP-1d also increases the correlation between simulated and observed AOD from $R = 0.29$-0.31 to $R = 0.41$–0.42 (Table S1). These results show that daily re-initialized forecast provides better representation of both surface dust concentration and column-integrated aerosol concentration. This sensitivity to meteorological forcing agrees with previous studies showing that Saharan dust outbreaks contribute to short-range forecast uncertainty over Europe (Hermes et al., 2024) and that updated meteorological analyses improve aerosol forecast skill in operational forecasting systems (Benedetti et al., 2009). The improvement is not solely attributable to increased horizontal resolution. Similar improvements from EXP-5d to RI-EXP-1d are found at both coarse (26 and 13 km) and convection-permitting (2 and 1 km) resolutions, indicating that reducing accumulated meteorological forecast errors substantially improves the timing and duration of near-surface dust events. Higher resolution provides an additional benefit by resolving circulations and enhancing vertical exchange between the free troposphere and the boundary layer, resulting in sharper dust-layer boundaries and a reduced overestimation of elevated dust concentrations (Figure 10 and 12). These findings are consistent with previous studies demonstrating that convection-permitting simulations improve the representation of dust-related vertical transport processes associated with convective downdrafts and cold pools (Marsham et al., 2011; Heinold et al., 2013). However, higher resolution alone does not compensate for deficiencies in the meteorological forcing. In the EXP-5d simulations, the 1 and 2 km experiments underestimate the duration of the near-surface dust event more strongly than the coarser-resolution simulations because accumulated meteorological errors are more directly reflected in the better-resolved boundary-layer transport. These results demonstrate that meteorological initialization and horizontal resolution provide complementary improvements, with the best performance obtained when both are combined.
At first glance, the backscatter-profile statistics appear inconsistent with the improvements obtained for near-surface PM_{10}. The normalized mean bias of the lidar-model comparison is smaller for EXP-5d (5.9–6.5 %) than for RI-EXP-1d (38–50 %), whereas RMSE and correlation remain similar in both experiments ($R \approx 0.37$–0.41) (Table S1). This behaviour is most likely explained by compensating errors rather than improved model performance. The comparison with CALIPSO observations (Figure 13 and S8) shows that all simulations, including the 1 km RI-EXP-1d experiment, systematically overestimate backscatter above approximately 2.6 km. A continuous forecast may therefore produce a smaller overall bias if underestimation at one altitude compensates for overestimation elsewhere, without improving the vertical profile itself. This interpretation is supported by the nearly unchanged RMSE between EXP-5d and RI-EXP-1d. Similar limitations of profile-integrated bias metrics have been reported previously, and profile-resolved statistics are generally recommended for evaluating vertically resolved aerosol simulations (Binietoglou et al., 2015). Consequently, the backscatter results do not contradict the improvements in vertical dust transport but instead indicate that improved meteorological forcing and resolution improve the vertical redistribution of dust without correcting the persistent overestimation of elevated aerosol layers.
The overestimation of elevated dust layers forms part of a broader pattern that is largely unaffected by either model resolution or meteorological initialization. ICON-ART overestimates the backscatter coefficient at 355 nm (KASCAL) but show agreement with the backscatter coefficient at 532 nm (CALIPSO). This wavelength-dependent behaviour is consistent with ICON-ART underestimating the abundance of coarse particles while overestimating intermediate-sized particles. This interpretation is independently supported by the particle mass size distributions measured by the Fidas 200 (Figure 15 and S8). ICON-ART consistently overestimates particles between approximately 0.5 and 2 μm while underestimating particles larger than 2 μm, and these biases remain nearly unchanged in both EXP-5d and RI-EXP-1d. Since meteorological re-initialization substantially improves the timing and duration of surface dust events without affecting the particle size distribution, the remaining discrepancy likely reflects limitations in the representation of the emitted size distribution and the size-dependent removal of coarse particles, potentially through sedimentation and dry deposition during long-

range transport. This interpretation is consistent with previous studies. Ratcliffe et al. (2025) showed that additional processes are required to maintain coarse dust during atmospheric transport, while Adebiyi and Kok (2020) demonstrated from a global synthesis of in-situ observations that current models remove coarse dust too efficiently. The present study extends these findings to long-range Saharan dust transport over Europe using both remote sensing and in-situ observations.

Wet deposition represents another potential source of uncertainty, particularly during Case 3. Observations show that moderate precipitation efficiently removed dust from the boundary layer. However, the present analysis does not allow a robust assessment of the wet deposition parameterization because the model evaluation point is located approximately 40 km from the observation site and precipitation was spatially heterogeneous. Consequently, differences between simulated and observed dust concentrations may partly reflect spatial displacement of precipitation rather than deficiencies in the wet removal scheme itself. Wet deposition should therefore be regarded as an important physical process that requires further investigation using more closely collocated precipitation observations.

Overall, the results indicate that further improvements in the simulation of long-range Saharan dust transport should focus on two largely independent aspects. Meteorological initialization together with convection-permitting resolution substantially improves the timing, vertical distribution, and surface concentrations of transported dust. In contrast, the persistent coarse-particle deficit and associated optical discrepancies remain largely insensitive to these improvements and instead reflect limitations in the current aerosol representation. Future model development should therefore focus on improving the the representation of the emitted coarse-particle size distribution and size-dependent sedimentation and deposition processes and on extending the aerosol representation to include ageing and mixing processes that modify aerosol optical properties.

3. It is valuable to include ground-based in situ data. However, they are not appearing in your discussion (Sect. 3.5, 3.6) any more. Where is your focus?

R: We agree that the role of the ground-based in-situ measurements was not sufficiently emphasized in the original Discussion section. To address this, we have expanded Sections 3.5 and 3.6 and explicitly incorporated the Fidas 200 aerosol size-distribution measurements into the discussion. In addition to the temporal evolution of surface dust concentrations, we now compare the observed and simulated aerosol/dust mass size distributions during different stages of the dust events. The Fidas 200 observations provide information that cannot be obtained from the lidar measurements alone. While the lidar observations primarily constrain the vertical structure and optical properties of the dust plumes, the Fidas measurements allow us to evaluate the particle-size characteristics of the transported dust at the surface. The comparison reveals a systematic bias in ICON-ART, which tends to overestimate the mass concentration of particles between approximately 0.5 and 2 μm while underestimating particles larger than 2 μm. A similar conclusion was reached by **?**, who showed that additional processes counteracting the rapid sedimentation of coarse dust particles are required for climate models to reproduce observations. Importantly, this bias is present in both the EXP-5d and RI-EXP-1d simulations, suggesting that it is primarily related to the representation of dust emission, sedimentation, and deposition processes rather than meteorological initialization. The inclusion of the aerosol particle size measurements therefore extends the model evaluation beyond plume timing and vertical structure and provides additional constraints on the simulated dust size distributions and associated physical processes. We have revised the Discussion section accordingly to better highlight the contribution of the in-situ observations to the overall conclusions of the study.

The particle mass size distributions measured by the Fidas 200 reveal a similar behaviour. ICON-ART consistently overestimates particles between approximately 0.5 and 2 μm while underestimating particles larger than 2 μm, and these biases remain nearly unchanged in both EXP-5d and RI-EXP-5d. Since meteorological re-initialization substantially improves the timing and duration of surface dust events but does not affect the particle size distribution, the remaining discrepancy most likely originates from the representation of aerosol microphysical processes, particularly sedimentation and dry deposition of coarse particles during long-range transport. This interpretation agrees with previous studies. Ratcliffe et al. (2025) showed that additional processes are required to maintain coarse dust during atmospheric transport, while Adebiyi and Kok (2020) demonstrated from a global synthesis of in-situ observations that current models remove coarse dust too efficiently. The present study extends these findings to long-range Saharan dust transport over Europe using both remote sensing and in-situ observations.

4. The general findings of the agreement in dust layer height and backscatter coefficient appear suddenly in the conclusion and before only in the abstract (L7/8, L488/489) and it remains unclear throughout the manuscript how they were derived. E.g., the mean backscatter in Tab. 2 differs much stronger between model and observations.

R:Thank you for pointing this out. We agree that the presentation of these metrics was not sufficiently clear in the original manuscript. We have therefore revised the Abstract and Conclusions. The revised abstract is as follows:

Abstract:

Long-range Saharan dust transport affects radiation, clouds, and air quality, yet its vertical structure and entrainment into the boundary layer remain challenging to simulate. We investigate four contrasting dust events over southwestern Germany using ground-based lidar, sun-photometer and in situ aerosol measurements, satellite observations, backward trajectories, and ICON-ART simulations. The events represent diverse atmospheric conditions, including a six-day cloud-free episode, a cloudy episode involving dust–cloud interactions, and an intense outbreak with aerosol optical depth approaching 1.4. A depolarization-based lidar method separated dust backscatter from contributions by other aerosols, enabling a like-for-like evaluation of the dust represented by ICON-ART and providing a useful approach for evaluating dust models in mixed aerosol layers. ICON-ART reproduced the large-scale transport and arrival of the dust layers, with mean differences of approximately 20 min in arrival time and 50 m in layer height. Daily re-initialization of the global simulations improved correlations with observed near-surface particulate matter concentrations from 0.65–0.69 to 0.86–0.88 and better reproduced the timing and persistence of boundary-layer dust. The high resolution runs (2, 1 km) better resolved dust-layer stratification, vertical gradients, and exchange between the free troposphere and boundary layer more realistically. However, ICON-ART consistently overestimated particles of 0.5–2 µm and underestimated particles larger than 2 µm. These results demonstrate the complementary benefits of accurate meteorological initialization and convection-permitting resolution while identifying size-dependent emission, sedimentation, and deposition as priorities for model development. By distinguishing errors in dust transport from those arising from aerosol size representation, this study provides an observation-constrained framework for evaluating and improving convection-permitting dust forecasts.

Conclusion:

This study combined vertically resolved lidar observations, sun-photometer and in situ measurements, satellite products, backward trajectories, and ICON-ART simulations to investigate four contrasting Saharan dust events over southwestern Germany. The selected cases covered prolonged clear-sky transport, dust–cloud interactions, and an intense dust outbreak, providing a range of observational constraints on Saharan dust transport, vertical structure, optical properties, and boundary-layer entrainment.

ICON-ART reproduced the regional transport pathways and arrival of the Saharan air layers well, demonstrating that the model captures the synoptic-scale processes controlling long-range dust transport towards Western Europe. Separating the lidar-derived backscatter into dust and non-dust components enabled a consistent comparison with the modelled dust field and avoided attributing the contribution of other aerosols to model error. This separation provides an important constraint for evaluating dust transport in mixed aerosol layers.

The sensitivity experiments show that meteorological forcing and horizontal resolution play complementary roles. Daily re-initialization of the global simulation provided more accurate lateral boundary conditions and improved the timing and persistence of dust within the boundary layer. Convection-permitting simulations better represented layer stratification, vertical gradients, and exchange between the free troposphere and boundary layer, but their added value was realized only when the large-scale meteorological state and lateral boundary conditions were sufficiently accurate. Thus, higher resolution alone does not necessarily translate into a better dust forecast.

In contrast, persistent biases in particle size and associated optical properties were largely insensitive to meteorological initialization and grid spacing. ICON-ART shifted the transported dust towards smaller particle sizes, overestimating particles between approximately 0.5 and 2 µm while underestimating particles larger than 2 µm. The results therefore indicate that transport-related and aerosol-process errors are partly separable: improved meteorological forcing and resolution can improve plume timing and vertical redistribution, whereas the persistent coarse-particle deficit points to limitations in the representation of coarse-dust emissions and size-dependent removal processes.

Overall, the combination of vertically resolved and size-resolved observations provides a means of distinguishing dynamical from aerosol-process limitations in long-range dust simulations. Future model development should therefore prioritize the joint improvement of lateral meteorological forcing and coarse-dust process representation. This will require targeted evaluation of coarse-particle emissions and removal processes together with more closely collocated observations of aerosol size, composition, clouds, and precipitation.

5. One has to be careful what to compare. In the case of the lidar observations, the authors have derived the dust-related backscatter coefficient which could be directly compared to the model output. However, in the case of the photometer observations, they compare total AOD to the dust AOD from the model. If the model contains only the dust as aerosol component, one has to find a good way to compare the results to atmospheric observations. A comparison of the extinction coefficient between observations and model would be a great addition. Furthermore, a temporal offset between model and observations could be investigated (e.g., in case 3) as possible reason for disagreement. Dust transport patterns and origins are only briefly discussed.

R: Thank you for this important comment. We fully agree that observational and modeled quantities must be defined consistently when evaluating a dust-only model configuration. Following the reviewer's suggestion,

we carefully revised the manuscript and now consistently compare dust-related observational products with the dust-only ICON-ART simulations throughout the study. For the lidar observations, Eq. (1) is used to derive the dust-related backscatter coefficient, which is directly compared with the modeled dust backscatter. Similarly, for the sun photometer observations, we now use the coarse model aerosol optical depth rather than the total aerosol optical depth when comparing with the ICON-ART simulations. We recognize that this distinction was not sufficiently clear in the original manuscript and may have led to confusion. Therefore, the methodology, results, and discussion sections have been revised to explicitly state that all lidar, photometer, and model comparisons are based on dust-related quantities. This ensures consistency between the observations and the dust-only model configuration.
In addition, we have expanded the discussion of dust transport pathways and source attribution by moving the backward-trajectory analyses to the main text and by providing supporting MODIS AOD observations. These additions strengthen the interpretation of the transport patterns and the Saharan origin of the observed dust plumes.

6. The argumentation has several deficits:
6.1. In Line 364, you state that the layer represents biomass burning aerosol. Why? Could you compare the optical properties to those observed in biomass burning aerosol layers from literature? Or could you indicate trajectories or fire locations from which the biomass burning aerosol originates? You postulate the origin of the layer without providing evidence or at least indications for the origin.

R: Thank you for this valuable comment. We agree that the original manuscript did not provide sufficient evidence to conclusively identify the observed layer as biomass-burning aerosol. In addition to the optical properties, we revisited the backward-trajectory analysis for this event. The trajectories associated with the low-depolarization layer indicate transport within the lower troposphere over continental Europe and suggest a significant contribution from regional and local aerosol sources. Therefore, we agree that the available observations do not allow an unambiguous attribution of the layer solely to biomass-burning aerosol. In the revised manuscript, we have adopted a more cautious interpretation and now describe this layer as **a low-depolarization continental aerosol layer** that may contain contributions from anthropogenic pollution, aged biomass-burning aerosol, and other regional aerosol types. The aerosol classification is based on both the observed optical properties and the trajectory analysis. We have revised the text accordingly and clarified the limitations of source attribution based solely on lidar observations.

Case 4 reveals the presence of Saharan dust, continental aerosol layers with low depolarization characteristics, and mixtures of these aerosol types. The backward-trajectory analysis indicates that the low-depolarization aerosol layers were transported within the lower troposphere over continental Europe, suggesting contributions from regional and local aerosol sources in addition to long-range transported Saharan dust. On 16^{th} March 2022, a dense dust layer was observed, characterized by a high depolarization ratio (0.32 ± 0.06) and a lidar ratio of 56 ± 11 sr. On 18^{th} March, two distinct layers were identified: an upper dust layer (LPDR: 0.32 ± 0.04, lidar ratio: 36 ± 7 sr) and a lower low-depolarization aerosol layer (continental aerosol) (LPDR: 0.02 ± 0.003, lidar ratio: 38 ± 8 sr). On March 19, a mixed layer of Saharan dust and continental aerosols (LPDR: 0.11 ± 0.019, lidar ratio: 27 ± 7 sr) was observed below a continental aerosol layer (LPDR: 0.05 ± 0.004, lidar ratio: 49 ± 3 sr). By 20^{th} March, the layers had mixed more uniformly, with LPDR and lidar ratio values of 0.1 ± 0.02 and 33 ± 13 sr, respectively.The temporal evolution of LPDR and lidar ratio indicates a progressive transition from a dust-dominated aerosol column to a more complex aerosol mixture. While the optical properties of the low-depolarization layers are consistent with aerosol types commonly associated with anthropogenic pollution and aged biomass-burning particles, the trajectory analysis suggests a substantial contribution from regional and local aerosol sources. The gradual decrease in LPDR and the emergence of intermediate optical properties are consistent with increasing mixing between Saharan dust and regional aerosols during transport. This mixing results in intermediate depolarization values and variable lidar ratios, depending on the relative abundance of dust and low-depolarization layers in the column. These findings underscore the complexity of aerosol interactions and their evolving optical properties during long-range transport events involving multiple sources. The observed aerosol layering illustrates that realistic representation of aerosol mixing is essential for accurately simulating optical properties during long-range transport.

6.2. In line 323/324, you speculate that the dust particles have formed ice crystals in time window C1. Have you checked the temperature in that layer? Do you observe structures of ice crystals in your lidar observations? Should the dust particles not form first a liquid cloud during summertime conditions?

R: Thank you for this valuable comment. We agree that our original interpretation was not sufficiently supported. Following the reviewer's suggestion, we examined the ERA5 temperature profiles corresponding to the dust layer observed during period C1. The analysis shows that temperatures within the layer were above 0 °C. Under such thermodynamic conditions, the formation of liquid cloud droplets is much more likely than the direct formation of ice crystals. We therefore agree that the original statement referring to ice-crystal formation was inappropriate. In the revised manuscript, we have replaced this interpretation with a discussion of possible liquid cloud formation associated with the dust layer. We have also revised the corresponding text to avoid unsupported speculation regarding cloud phase. We have modifed the text as follows:

One is that ICON-ART underestimated the dust loading in this case and the a second possible explanation is that the Saharan dust particles interacted with cloud droplets within the layer. ERA5 temperature data indicate that temperatures were above 0 °C at the corresponding altitude, suggesting that liquid cloud droplets were more likely to form. The presence of cloud droplets can increase particle size and enhance the observed lidar backscatter, contributing to the elevated signal within the dust layer.

6.3. And even more severe, Fig. 1a, c and d show structures of a possible dust layer up to around 6 km height. But somehow this layer is not captured in your profile C1. The time series agree much better with the model than your profile comparison. Please check carefully your profiles again (and provide profiles of the particle depolarization ratio additionally in Fig 2+7).

R: Thank you for carefully examining the figure and identifying this inconsistency. We revisited the profile extraction used for the C1 comparison and found that the discrepancy originated from an incorrect selection of the reference height range used for the profile analysis. As a result, the original C1 profile did not fully represent the elevated aerosol layer visible in the time–height cross section. After correcting the profile extraction procedure, the revised C1 profile is now consistent with the corresponding lidar observations shown in figure 2 and captures the elevated aerosol layer extending to approximately 6 km altitude. The figure and associated discussion have been updated accordingly. In addition, figures 2 and 7 have been removed from the manuscript to better focus on the main objectives of this study.

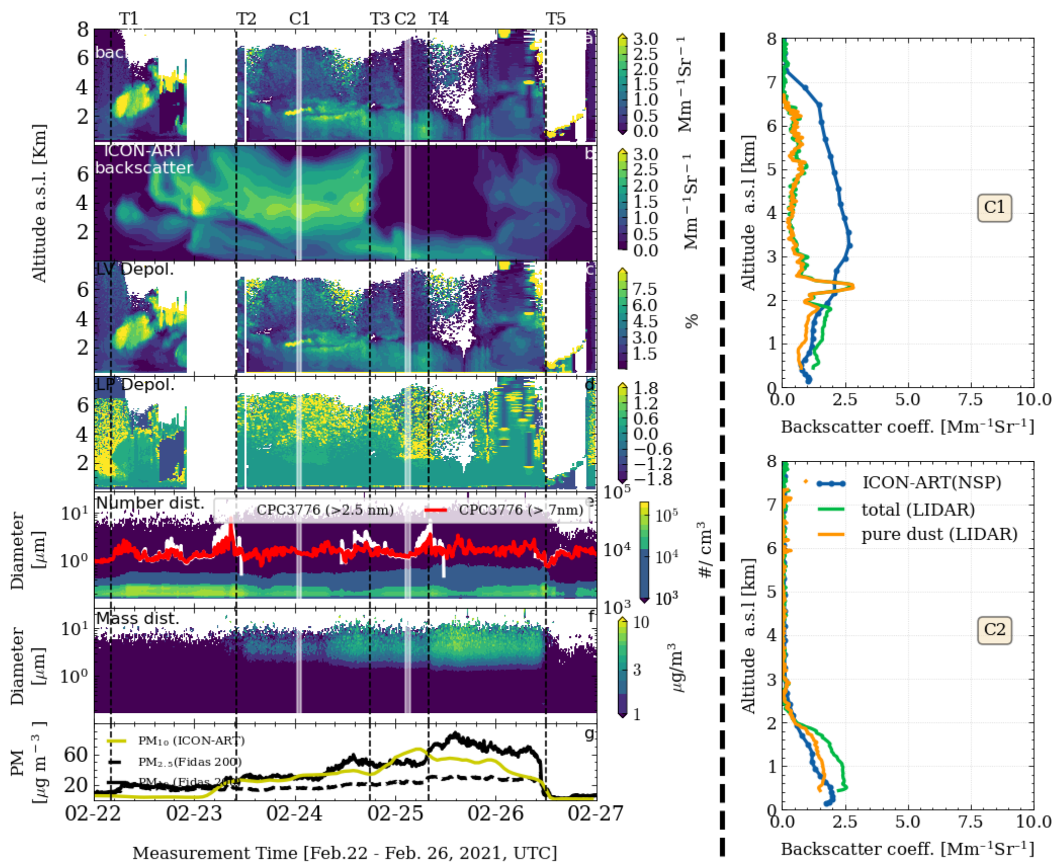

Figure 2: Time series of backscatter coefficients from KASCAL measurements (a), predicted backscatter coefficients by ICON-ART (b), Linear Volume Depolarization Ratio (LVDR) (c) Linear Particle Depolarization Ratio (LPDR) (d) from KASCAL measurements, particle number size distribution from Fidas 200 and particle number concentration from two CPCs (e), particle mass size distribution from Fidas 200 (f) and particulate matter concentrations (PM_{10} & $PM_{2.5}$) from the Fidas 200 and the lowest model level (150 m) of the ICON-ART simulation (g) from 22^{nd} to 26^{th} February 2021 for case 2. The profiles of backscatter coefficients measured by the lidar and predicted by ICON-ART model for case C1 (Averaged profiles from 00:30 to 1:30 for lidar measurement and profile at 01:00 for ICON-ART prediction on 24^{th} February 2021) and for a selected time C2 (Averaged profiles from 02:30 to 03:30 for lidar measurement and profile at 03:00 for ICON-ART prediction on 25^{th} February 2021) are shown on the right side of this figure. The vertical dashed lines in the contour plots indicate dust arrival (T1), dust touched to the ground (T2), an increase of PM concentration at ground level (T3 & T4) and a low-level cloud appeared (T5). C1 and C2 represent time periods used for a more detailed data analysis as shown in the right side of this figure.

7. The presentation shows deficits as well:

7.1. Please use a common date format throughout the entire manuscript.

R: Thank you for the comment. We have carefully reviewed the manuscript and revised all date expressions to a consistent format throughout the text, tables, and figure captions.

7.2. Please provide the units for the lidar ratio.

R: Thank you for the comment. The units of the lidar ratio (sr) have been added in the revised manuscript.

7.3. Fig 1,6+8 are much too small and hard to read. Furthermore, I don't see an added value of providing the lidar profiles at two angles.

R: Thank you for this valuable comment. we have increased the size of these figures to improve readability and facilitate a clearer comparison between the observations and model results. We also remove slant angle profiles and the revised plot is shown in Figure 2, 5, 6. The profile comparisons now focus on the total backscatter coefficient and the retrieved pure-dust backscatter coefficient. This modification provides a more direct comparison between the observations and the dust-only ICON-ART simulations and better highlights the contribution of mineral dust to the observed aerosol layers.

7.4. Fig 9: The description in the text for the red squares does not fit to the figure, because you mention a lidar ratio of 38 sr for the layer with the high depolarization ratio and 36 sr for the one with the low depolarization ratio. However, the red squares indicate the opposite.

R: Thank you for carefully checking the figure and identifying this inconsistency. The descriptions of the two aerosol layers in the text were inadvertently reversed. The layer with the higher depolarization ratio corresponds to a lidar ratio of approximately 36 sr, whereas the layer with the lower depolarization ratio corresponds to a lidar ratio of approximately 38 sr, consistent with the red squares shown in Fig. 9. The text has been corrected in the revised manuscript. We thank the reviewer for bringing this error to our attention.

On 18^{th} March, two distinct layers were identified: an upper dust layer (LPDR: 0.32 ± 0.04, lidar ratio: 36 ± 7 sr) and a lower low-depolarization aerosol layer (continental aerosol) (LPDR: 0.02 ± 0.003, lidar ratio: 38 ± 8 sr).

7.5. L400/401 Schuster et al., 2012 does not report lidar ratios at 355 nm. Freudenthaler et al., 2009 also not. Cao et al., 2014 also not. I am bit puzzled about your literature study, because then you continue with "This agreement suggests ..."

R: Thank you for carefully checking the cited references and identifying this issue. The reviewer is correct that Schuster et al. (2012), Freudenthaler et al. (2009), and Cao et al. (2014) do not directly provide lidar-ratio values at 355 nm and therefore do not fully support the original statement. To address this concern, we have revised the literature discussion and included additional references that report 355 nm Raman lidar observations of Saharan dust and dust mixtures, including GROß et al. (2011); Tesche et al. (2009). These studies provide lidar-ratio and depolarization-ratio values that are directly comparable to our measurements. We thank the reviewer for identifying this issue, which has helped us improve the accuracy and clarity of the literature comparison.

Overall, the retrieved optical properties exhibit values that are consistent with well-established characteristics of transported Saharan dust. In particular, the lidar ratios (LR), depolarization ratios, and SSA values fall within the range of previously reported measurements over North Africa and downwind regions (Freudenthaler et al., 2009; Schuster et al., 2012; Cao et al., 2014; GROß et al., 2011; Tesche et al., 2009).

7.6. The text describing Fig. 12 suggests that the highest AOD was observed in case 4 with a value of 0.30 (L432), but the figure shows higher mean values for case 3. Furthermore, you state that the AOD of ICON-ART is systematically lower (L434), but the mean in case 2 and the median in case 1 are higher than the observed one. You have to be more careful in describing your observations.

R: Thank you for carefully examining figure 12 and identifying these ambiguities. We agree that the original wording was not sufficiently precise. The statement regarding Case 4 referred to the highest peak AOD observed during the four events, whereas figure 12 primarily presents the statistical distribution of AOD values. We also agree that the phrase "ICON-ART systematically underestimated the observed AOD" was too strong and not fully supported. As pointed out by the reviewer, the model exhibits both underestimation and overestimation depending on the event. In order to get robust conclusion from comparison of AOD from Sunphotometer and ICON-ART model, We have added a new evaluation of the simulated dust AOD against AERONET coarse-mode AOD using observations from 17 stations within the innermost model domain. The comparison covers both EXP and RI-EXP at horizontal grid spacings of 26, 13, 2, and 1 km. The results show that daily re-initialization improves the correlation and brings the regression slope closer to unity, whereas the AOD statistics vary only slightly with grid spacing. This suggests that the remaining discrepancies are primarily related to uncertainties in coarse-particle mass, size distribution, and associated optical properties rather than to horizontal resolution alone.

Figure 14 compares the ICON-ART dust AOD with the AERONET coarse-mode AOD for the EXP and RI-EXP simulations at four horizontal grid spacings. ICON-ART captures the variability in AERONET coarse-mode AOD reasonably well across all four horizontal grid spacings, although the agreement depends more strongly on the experimental configuration than on grid resolution. For a given configuration, the correlation coefficients and other statistical metrics remain largely insensitive to changes in horizontal grid spacing from 24 to 1 km. The standard EXP simulations yield correlations of r=0.584–0.588 and regression slopes of 0.790–0.811. The slopes of approximately 0.8 indicate that the model increasingly underestimates AOD as the observed AOD increases, despite the small mean biases (-0.009 to 0.009). RI-EXP increases the correlations to r=0.642–0.656 and brings the regression slopes closer to unity (0.972–1.045), indicating an improved representation of the observed AOD magnitude and variability. However, this improvement is accompanied by positive mean biases of 0.027–0.042 and does not lead to a systematic reduction in RMSE. Among the RI-EXP simulations, the 1 km simulation shows the closest agreement, with r=0.656, a slope of 0.972, an RMSE of 0.128, and a mean bias of 0.027. These optical comparisons should, however, be interpreted in light of uncertainties in the par-

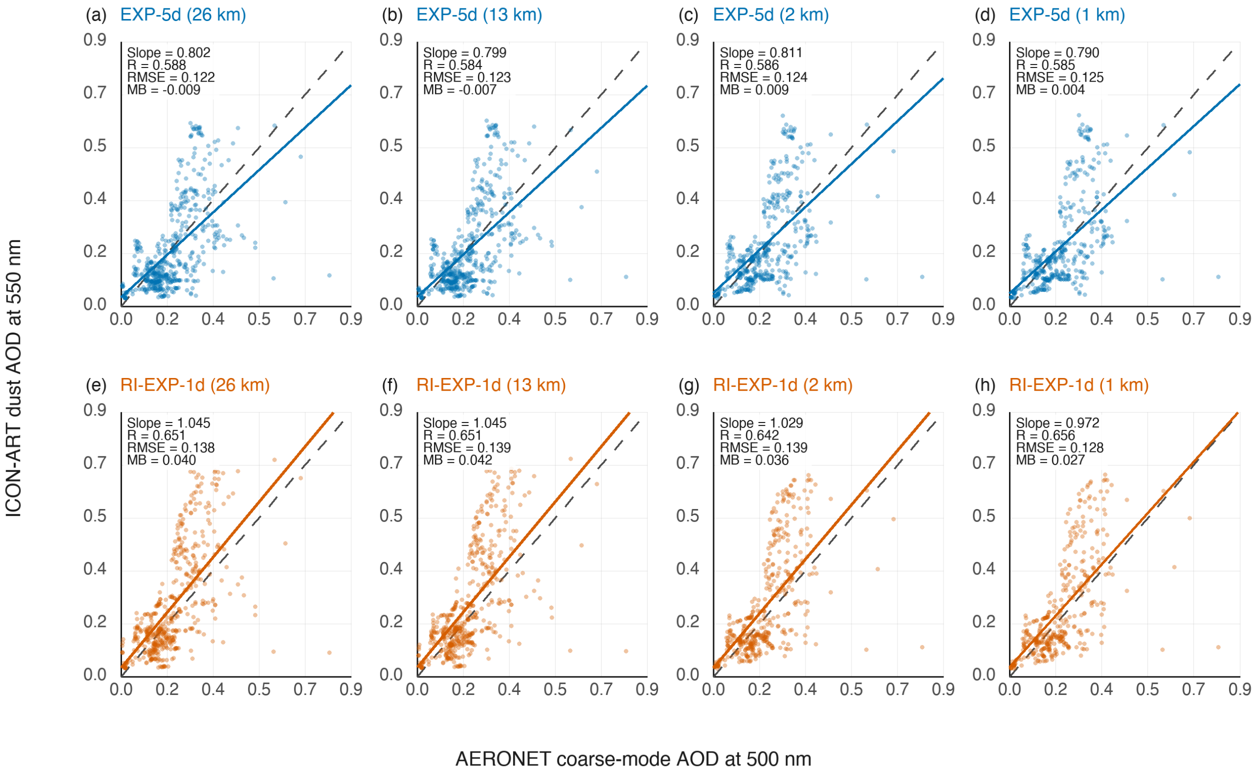

Figure 14: Comparison of AERONET coarse-mode AOD and ICON-ART-simulated dust AOD for the EXP (a–d) and RI-EXP (e–h) simulations at horizontal grid spacings of 24, 13, 2, and 1 km. The analysis is restricted to the innermost model domain, which contains 17 AERONET sites. Solid lines represent the least-squares linear regressions, and dashed lines indicate the 1:1 relationship. The reported statistics are the regression slope, Pearson correlation coefficient (r), root-mean-square error (RMSE), and mean bias (MB; model minus observation).

ticle size distribution (PSD) assumed in ICON-ART, as the PSD directly affects the conversion of simulated aerosol mass into optical properties. This issue is particularly relevant to the coarse mode: ICON-ART substantially underestimates both PM_{10} and coarse-particle mass and underrepresents the pronounced contribution of particles larger than approximately 5µm evident in the observations, as shown in figure 14 and reported in previous studies (Hoshyaripour et al., 2019). Consequently, part of the remaining AOD discrepancy may arise from uncertainties in the assumed PSD and the associated optical properties, in addition to the underestimated coarse-aerosol burden. The persistence of similar AOD statistics across resolutions therefore suggests that the dominant discrepancies are related to aerosol representation, including the coarse-particle size distribution and mass, and/or model–observation differences, rather than spatial resolution alone.

Concluding comment:
I feel a bit sorry for this critical review. However, the manuscript is a mere description of the detailed observations during 3 Saharan dust intrusions to Europe and the corresponding model runs with ICON-ART. In that sense, it does not represent a scientific advancement to merit publication in ACP.

General Comment: We thank the reviewer for the time and effort invested in reviewing our manuscript and for the many constructive comments provided throughout the review process. We understand the concern that the original manuscript may have appeared primarily descriptive and that its scientific focus was not sufficiently clear.
In response to the reviewer's comments, we have substantially revised the manuscript and expanded the scientific analysis. The revised manuscript is no longer limited to a description of four Saharan dust events and a comparison with a single model simulation. Instead, the four events are used as a testbed to investigate several factors controlling the representation of long-range transported Saharan dust in ICON-ART.
Specifically, we have added:

- A multi-resolution evaluation of ICON-ART (24 km, 13 km, 2 km, and 1 km), allowing assessment of the added value of kilometer-scale simulations for representing dust-layer stratification and vertical structure.
- A comparison between continuous forecasts (EXP-5d) and daily re-initialized forecasts (RI-EXP-1d), providing insight into the influence of meteorological initialization on dust transport simulations.
- Independent validation using CALIPSO observations, which helps distinguish genuine model deficiencies

from observational limitations associated with ground-based lidar measurements.

- A detailed evaluation of aerosol size distributions using Fidas 200 observations, allowing identification of systematic biases in the simulated particle-size spectra and providing constraints on dust emission, sedimentation, and deposition processes.
- A new evaluation of the simulated dust AOD against AERONET coarse-mode AOD using observations from 17 stations within the innermost model domain was added into manuscript.
- Expanded trajectory analyses and source attribution to better establish the transport pathways from North Africa to Europe.

These additions allow us to move beyond a simple model-versus-observation comparison and address process-oriented questions regarding the roles of model resolution, meteorological forcing, observational constraints, and dust process representation. We have revised the abstract, introduction, discussion, and conclusions accordingly to better communicate these scientific findings.
We appreciate the reviewer's critical assessment, which has significantly improved the clarity, focus, and scientific depth of the manuscript.

References

Adebiyi, A. A. and Kok, J. F.: Climate models miss most of the coarse dust in the atmosphere, Science Advances, 6, eaaz9507, https://doi.org/10.1126/sciadv.aaz9507, 2020.

Cao, C., Zheng, S., and Singh, R. P.: Characteristics of aerosol optical properties and meteorological parameters during three major dust events (2005–2010) over Beijing, China, Atmospheric Research, 150, 129–142, https://doi.org/10.1016/j.atmosres.2014.07.022, 2014.

Freudenthaler, V., Esselborn, M., Wiegner, M., Heese, B., Tesche, M., Ansmann, A., Müller, D., Althausen, D., Wirth, M., Fix, A., Ehret, G., Knippertz, P., Toledano, C., Gasteiger, J., Garhammer, M., and Seefeldner, M.: Depolarization ratio profiling at several wavelengths in pure Saharan dust during SAMUM 2006, Tellus B, 61, 165–179, https://doi.org/10.1111/j.1600-0889.2008.00396.x, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1600-0889.2008.00396.x, 2009.

GROß, S., TESCHE, M., FREUDENTHALER, V., TOLEDANO, C., WIEGNER, M., ANSMANN, A., ALTHAUSEN, D., and SEEFELDNER, M.: Characterization of Saharan dust, marine aerosols and mixtures of biomass-burning aerosols and dust by means of multi-wavelength depolarization and Raman lidar measurements during SAMUM 2, Tellus B, 63, 706–724, https://doi.org/https://doi.org/10.1111/j.1600-0889.2011.00556.x, 2011.

Hoshyaripour, G. A., Bachmann, V., Förstner, J., Steiner, A., Vogel, H., Wagner, F., Walter, C., and Vogel, B.: Effects of Particle Nonsphericity on Dust Optical Properties in a Forecast System: Implications for Model-Observation Comparison, Journal of Geophysical Research: Atmospheres, 124, 7164–7178, https://doi.org/10.1029/2018JD030228, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2018JD030228, 2019.

Ratcliffe, N. G., Ryder, C. L., Bellouin, N., Jones, A., Johnson, B., Woodward, S., Wieland, L.-M., Gasteiger, J., and Weinzierl, B.: New processes to counteract sedimentation of coarse dust particles are required for climate models to agree with aircraft observations, EGUsphere, 2025, 1–33, https://doi.org/10.5194/egusphere-2025-5143, 2025.

Schuster, G. L., Vaughan, M., MacDonnell, D., Su, W., Winker, D., Dubovik, O., Lapyonok, T., and Trepte, C.: Comparison of CALIPSO aerosol optical depth retrievals to AERONET measurements, and a climatology for the lidar ratio of dust, Atmospheric Chemistry and Physics, 12, 7431–7452, https://doi.org/10.5194/acp-12-7431-2012, publisher: Copernicus GmbH, 2012.

Tesche, M., Ansmann, A., Müller, D., Althausen, D., Engelmann, R., Freudenthaler, V., and Groß, S.: Vertically resolved separation of dust and smoke over Cape Verde using multiwavelength Raman and polarization lidars during Saharan Mineral Dust Experiment 2008, Journal of Geophysical Research: Atmospheres, 114, https://doi.org/https://doi.org/10.1029/2009JD011862, 2009.