I am grateful to the editor for his comments that pointed out the shortcomings of this manuscript and gave me a second chance to improve it.

Requests for clarification or additional discussion raised by reviewers often indicate a lack of clarity of the manuscript or missing information, making it difficult for readers to follow the paper or argumentation. Therefore, additional explanations and clarifying statements should not only be provided in the author's response (e.g., RC1 comment on gravitational settling), but also be included in the revised version, except the authors consider a referee comment as inappropriate. In the latter case, a clear argument should be given in the author's response. Furthermore, a rephrasing of the initial text instead of a simple repetition often helps to increase clarity (e.g., RC1 comment on Fig. 6 and related discussion).

For RC1 comment on gravitational settling, the following discussion has been added to the manuscript.

- This scheme not only calculates the gravity settlement from the upper layer of the model to the lower layer of the model, but also calculates the gravity settlement from the bottom layer of the model to the ground. Meanwhile, the GOCART dry deposition protocol [Chin et al., 2000] was used for GEFS-Aerosols. Dry deposition in GEFS-aerosol was calculated by dry deposition velocity based on aerodynamic resistance, sublayer resistance and surface resistance. Therefore, aerosol gravity deposition and aerosol dry deposition are completely separated in GEFS-Aerosols.

For RC1 comment on Fig 6, this paragraph has been rewritten as:

- "To better understand the model error shown in Fig. 6, the GEFS-Aerosols output frequency was changed from every 3 hours (orange line in Fig. 7) to every hour (blue line in Fig. 7). Fig. 7 shows that the hourly variation in model error for dust simulations (in a 5-day simulation) is actually similar to that at 3-hour intervals, but the magnitude of peaks is reduced by about 60%, suggesting that model error is sensitive to model output frequency. "
- The linear assumption of aerosol deposition and emissions when testing the mass balance equation (Eq. 1) are the main cause of the model error. The deposition and emissions output by the GEFS-Aerosols diagnostic system are instantaneous values rather than cumulative values. Therefore, to calculate the cumulative amount of aerosol deposition or emissions over a model output time interval (e.g., three hours), simply multiply this value by three based on the linearity assumption. This treatment only affects deposition calculations for BC and OC (The daily emissions of BC and OC are constant.), but for dust and sea salt it affects not only deposition but also emissions calculations. At the same time, the wind threshold velocity makes the dust emissions more nonlinear than the source or sink terms of the other aerosol types. Therefore, the model errors for dust and sea salt are higher than those for BC and OC, while the model errors for dust are the highest. In general, when aerosol deposition or emissions increase, the error in calculating them in the analysis also increases due to linearity assumptions. For example, when aerosol emissions increase, in addition to BC and OC, the error in calculating aerosol emissions in the mass balance equation also increases. Correspondingly, the error in deposition calculations will also increase because an increase in ambient aerosol concentration will lead to an increase in the amount of deposition. This is why the model error for dust has a higher correlation with aerosol emissions than for BC and OC. However, for sea salt this correlation does not exist because sea salt emissions shown in Figure 6 are relatively stable. "

RC1 raised the point that the mass balance equation and associated processes did not mention the model's advection, diffusion, and physical processes. In your reply you wrote: "Although aerosols are affected by advection, diffusion, and physical processes, these processes are not specifically considered in the equations because they do not cause aerosols to leave/enter the system or change aerosol species." Unfortunately, this statement is not necessarily valid for a numerical model as the numerical schemes applied to solve advection, diffusion and other processes are not necessarily mass conserving, i.e., numerical artefacts might indeed lead to an artificial gain or loss of aerosol mass in the model, and this is exactly the referee's point. The same holds for the rather low model top. So to properly address this issue it needs to be shown and discussed how the model's advection, diffusion, physical processes and model top affect the aerosol mass balance, either by appropriate model experiments or by adding relevant references.

The following discussion has been added to the manuscript in response to comments from the reviewers and editor.

- "For a given system (such as the entire atmosphere), the amount of chemicals entering the system is equal to the amount of chemicals leaving the system."
- "Although aerosol mass is affected by advection, diffusion, and physical processes, these processes are not specifically considered in the equations because they do not cause aerosols to leave/enter the system or change the aerosol species. However, these processes do change the concentration of aerosols in the atmosphere, and this effect is ultimately included in the "Initial", "Reaction", "Removal" and "Final" terms of the mass balance equation."
- "Theoretically, aerosol mass in GEFS-Aerosols simulations should be conserved, which means that the model error should be zero if calculation accuracy is not taken into account. Possible reasons for the non-conservation of aerosol mass in GEFS-Aerosols as shown in Fig. 6 include: 1) The aerosol mass is not conserved in the advection, diffusion and physical processes of the model; 2) Aerosol leakage at the top of the model layer; 3) There are problems with calculating aerosol emissions and deposition in the mass balance equations; "
- "First, aerosol transport in GEFS-Aerosols is based on the FV3 dynamic core [Lin et al., 1994], which is also used in NASA-GOCART and GEOS-Chem. The mass conservation problem of this dynamical framework has been discussed by Lin and Rood [1996]. The physical processes of GEFS-aerosols are derived from the GFDL (Geophysical Fluid Dynamics Laboratory) cloud microphysics scheme [Lin et al., 1983], which strictly adheres to the conservation of moist energy during phase changes. Secondly, the pressure at the top of the model in GEFS-Aerosols is set to 200 Pa. Since the pressure at the top of the model is low enough and the layers of the model are dense enough near the top [Campbell et al., 2022], the aerosol concentration in GEFS-Aerosols is the background concentration (1×10^{-16} μg/kg) in these layers. There may be mass conservation issues at the top of the model, but their impact is minimal."

Through the editor's comments, I realized there was a misunderstanding about how the mass balance equations were used in this study. Therefore I also added this discussion in the manuscript.

- "Because the linearity assumption in the analysis can lead to model errors as mentioned above, the model error shown in Fig. 6 does not represent the true model simulation error, but rather

the calculation error in the mass conservation analysis. Therefore the main purpose of using the mass balance equation in this study is to verify the aerosol deposition and emissions calculated in the model budget analysis, rather than verifying whether aerosol mass is conserved in the GEFS-Aerosols."

Both referees mentioned the lack of comparison / verification of your model results with observational data. As response to this point you added a short paragraph (lines 40-48) to the revised manuscript, listing a number of other studies which compared GEFS-aerosols with observational data. However, this paragraph still lacks a detailed discussion /summary of the outcome of the cited studies. Overall, the presentation of your data and results needs a more comprehensive discussion including results of relevant studies in the field

The following discussion has been added to the manuscript to summarize the work of Zhang et al. and Bhattacharjee et al.

- "These assessments found that GEFS-Aerosols captures not only major wildfire plumes in southern Africa, Siberia, the central Amazon, and central South America, as well as agricultural fire plumes over India, but also high dust events in North Africa and the Arabian Peninsula; At the same time, GEFS-Aerosols has good performance in reproducing the seasonal variations at most surface observation sites dominated by dust and biomass plumes, as well as reproducing the vertical profiles of OC, BC, sulfate, dust and sea salt observed by ATOM. However, these findings are based on comparisons of AOD or aerosol concentrations and lack other assessments beyond AOD and concentration."

RC1 suggested to show observations for sea salt near sea surface instead of AOD (Fig. 2). Such a figure has been included in the author's response to RC1, although without any discussion, but not in the revised version of the manuscript. Why? Please explain. Furthermore, I would suggest to add a figure showing the difference between GEFS and MERRA2. This would clearly facilitate the comparison.

AOD represents aerosol column concentration and may better match the aerosol removal process. Because aerosol removal doesn't just come from the ground. AOD is derived from aerosol concentration. One possibility is that the AODs may be very similar, but the aerosol concentrations are very different. I guess this is why the first reviewer asked how AOD is calculated in GEFS-Aerosols.
A sea salt surface mass concentration plot has been added to Figure 2, and the following discussion is included in the manuscript:

- "Fig. 2 represents the monthly mean sea salt AOD (top) and surface mass concentration ($\mu g/m^3$) (bottom) simulated in October 2019 from GEFS-Aerosols (left) and Modern-Era Retrospective analysis for Research and Applications Version 2 (MERRA-2) (right) [Molod et al., 2015]. GEFS-Aerosols and MERRA-2 show very similar results in simulating sea salt AOD, as does the distribution pattern of sea salt surface mass concentration. This is due to the fact that in GEFS-Aerosols, AOD is calculated using look-up tables (LUTs) of aerosol optical properties in the NASA GOCART model, consistent with AOD calculations in MERRA2."

RC1 suggested to show separate profile over land and sea in Fig. 10 and add some more discussion. I do not see any of this in the revised version, but also no related statement in the authors' response. Furthermore, you mentioned that for the evaluation of Fig. 11 the HIPPO experiment and CALIOP observations were used. However, it remains unclear to me how this evaluation has been done as I do not see any observational data plotted in Fig. 10 or 11.

The following discussion has been added to the manusctipt

- "Fig. 11 is more representative of the vertical distribution of aerosols near the source (such as over land), while Fig. 10 is more representative of the vertical distribution of aerosols far away from the source (such as over the ocean)."

The difference in vertical distribution of aerosols over ocean and land has been discussed in the manuscript, for example Figure 10 "For aerosol species other than sea salt, aerosol mass peaks at pressure levels between 800 and 600 hPa."; for Figure 11, "The contribution of aerosol surface emissions to aerosol concentration decreases with increasing altitude".

The HIPPO experiments and CALIOP observations are only used for qualitative comparison with Figure 10. Because Figure 10 shows the simulation results in September 2019, later than the HIPPO experiment, and the CALIOP observations cited in our manuscript are earlier than our simulations.

The following discussion has been added to the manuscript:

- "Note that the HIPPO experiment and CALIOP observations are only used for qualitative comparisons in this study because they are of different timing than the GEFS-aerosol simulations."

In your response to RC1, last point related to L350-365, you provided the following paragraph, which is obviously copied from the model description section of your manuscript: "In GEFS-Aerosols: "The GOCART dry deposition protocol [Chin et al., 2000] was used for GEFS-Aerosols…." " How are these technical details related to the referee's comment? Please clarify and put your response into context. Furthermore, it would be beneficial to add a short statement to your paper that measurements of aerosol deposition fluxes are extremely challenging and therefore rather limited.

I would say that the removal protocols used in GEFS-Aerosols may differ from those used in GOCART, but they still come from a standard protocol that is now widely accepted by the aerosol modeling community. The differences mainly arise from the way these schemes are parameterized in the model. Therefore, it is not surprising to see large differences. The lack of observations may give us a freedom to adjust the model at will. However, that doesn't mean it's correct.

The following discussion has been added to manuscript.

- "The budget analysis in Table 2 again demonstrates that two models can have completely different sources and sinks but end up with very similar concentration predictions, while at the same time it is difficult to discover which model is more correct. Because few observational are available for verification, especially for aerosol removal processes or net mass fluxes at surfaces. Measuring aerosol deposition fluxes is extremely challenging [Farmer et al., 2021], so such observations are rare. For example, the in-cloud mass scavenging efficiency of BC [Yang et al., 2019], the number of studies in this area is small but the reported data vary greatly, making it difficult to use for model evaluation."