Response to Reviewer 2

This paper used a parcel model to explore the parameter space and optimize the performance of the widely used Abdul-Razzak & Ghan aerosol activation scheme (ARG2000). The research question is well-defined and the comparison with other schemes and parcel model results is comprehensive. This paper addresses an important topic in aerosol-cloud interaction. However, the evaluation within the global model context could benefit from additional rigor. I recommend publication in GMD after addressing the comments below.

We sincerely thank the reviewer for their thoughtful and constructive comments on the manuscript. We have carefully considered each comment and have addressed them in detail below, making the necessary revisions and providing additional clarifications where appropriate. Reviewer comments are in **Blue**, our responses are in **Black**, and manuscript updates are in **Red**.

General Comment on Nd Comparison: While the comparison between model output Nd and MODIS Nd is interesting, it may be more complex due to the various differences between the model and satellite data. These differences make it challenging to determine whether the bias in the model with the old scheme is 'underestimated' or 'overestimated,' and consequently, whether the reported 'improvements' or 'degradations' are fully reliable. I believe the comparison using the MODIS simulator results might be more appropriate than using the model's Nd output.

We thank the reviewer for their suggestions. We agree that model-observation comparisons are inherently challenging, and it is difficult to definitively determine whether the updated parameterization performs better or worse due to the large uncertainties present in both the model and the observations. We have explored several approaches to address these shortcomings in the revised manuscript. First, we provide a better description of the cloud top N_d which attempts to mimic the way a satellite would see the cloud top. Second, based on Reviewer 1's suggestions, we calculate the error and report the magnitude to better evaluate the model performance. Finally, We also attempt to crudely match the Grosvenor et al. (2018) filtering criteria. Kindly see replies to the comments below.

The current study focuses mainly on updating the ARG parameterization, with the global model-observation comparison only intended to add context to the results. Although comparison with a MODIS satellite simulator could provide valuable insights, it is unclear whether it would be an improvement. The model simulates N_d. To use a satellite simulator such as COSP, we would need to use the effective radius and optical depth predicted by the satellite simulator to recalculate N_d using a similar algorithm to the satellite data. These variables are derived from the simulated N_d and liquid water content, and are therefore affected by assumptions about the droplet size distribution in the model's radiation scheme which do not impact the N_d we currently show. N_d predicted in this way is therefore less directly related to the ARG algorithm, and involves more modeling assumptions. It might be more closely related to the cloud radiative effect, but it would likely not be the same as the N_d predicted by the ARG algorithm.

In addition, to avoid excessive disk space usage we write out monthly mean N_d from our model, but if we wrote out monthly mean effective radius and optical depth, then calculated N_d, the N_d would be biased because it depends non-linearly on effective radius and optical depth. It would still be feasible to write these out and calculate N_d (say) every three hours in principle, and we agree that doing the comparison in multiple different ways would be interesting, even if there is no clear right answer.

We also think based on other work published and in progress that many of the biases in N_d we can identify are unrelated to the activation parameterization, but are instead due to aerosol concentration or updraft speed biases. Therefore a more careful comparison of N_d with observations would be of limited value without a corresponding evaluation of the aerosols and updraft speeds. All in all this would be a major undertaking we prefer to leave for a dedicated future study. As we comment in the text, "The biases likely originate mainly in aerosol modeling unrelated to the activation parameterization, or in the sub-grid updraft speed. The satellite retrievals used for the evaluation are also uncertain, and the comparison is imprecise due to the representativeness uncertainty". That said, we still feel that it is of value to set the improvements we present to the activation algorithm in the context of likely model biases and to draw the readers' attention here to the probable scale of the problems that must be addressed in this future work.

L210: It would be helpful to provide more details about how the model's N_d is calculated. Specifically, how is the liquid cloud-top defined in the model, and what methodology is used to derive Nd from this definition?

We thank the reviewer for their comment. We agree that it is important to describe in detail how cloud top droplet concentrations (N_d) are calculated in the model. The model diagnostic we use (which has existed in the model for many years but we are not aware of a reference) attempts to mimic the way a satellite would see the cloud top. In the revised manuscript, on line 231 we add the following:

"The cloud top is defined as the highest model level with non-zero liquid cloud fraction. If the cloud fraction is 100%, then N_d at that level is taken as the cloud top N_d. If not, the model-level-weighted N_d is calculated by summing over contributions from different model levels here indexed by i, as:

$$N_d = \frac{\sum_{i=1}^{n} N_d(i) \cdot cf(i) \cdot Pr(i)}{\sum_{i=1}^{n} cf(i) \cdot Pr(i)} \tag{1}$$

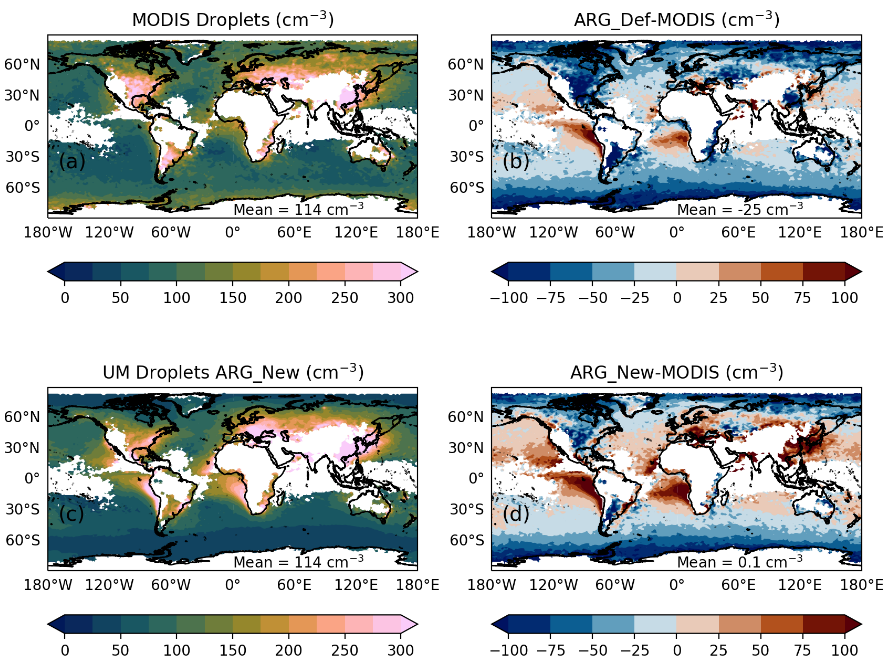
Here $N_d(i)$ is the droplet number concentration at the i-th model level. $cf(i)$ is the liquid cloud fraction at the i-th model level. $Pr(i)$ is the probability that a photon leaving the i-th layer can escape to space without encountering a cloud in the layers above. This is calculated based on the cloud fraction in the layer above the i-th layer. n is the total number of model levels. This equation is calculated only for the daytime. The cloud-top N_d is calculated in each timestep, and the monthly average is written out as an output diagnostic."

L215: The MODIS Nd estimates from Grosvenor et al. (2018) involve several thresholds to sample clouds. I would recommend clarifying whether similar thresholds are applied to your model results. If not, it would be valuable to discuss how the sampling differences due to post-processing thresholds and simulator assumptions may impact the comparison presented in Figure 3. Addressing these differences will enhance the clarity and reliability of the evaluation.

We thank the reviewer for their comment. We agree that the Grosvenor et al. (2018) dataset (GW2018) has several assumptions and thresholds to calculate N_d. We are able to apply most (but not all) of the thresholds while calculating the modeled N_d. The GW2018 dataset has a daily frequency, while the model provides monthly averages. First, we calculate the monthly mean GW2018 N_d using all available data points. Then we regrid it to the model gridpoints and filter out the model data (only liquid clouds) where we do not have any valid GW2018 retrievals. The GW2018 dataset includes retrievals only where the $1^\circ \times 1^\circ$ cloud fraction is at least 80%. Applying this threshold to the model output in Figure 3 is challenging due to the use of monthly-averaged cloud fraction.

In response to the comment, we perform a new sensitivity study. Since achieving precise filtering would require 3-hourly model outputs, we adopt an approximate approach to exclude gridboxes with low cloud fraction. Specifically, for each month, we filter out the bottom 25th percentile of weights (the denominator in the daytime cloud-top N_d equation) in the simulations and mask the corresponding MODIS data in those gridboxes. This ensures that we sample only gridboxes with relatively higher cloud fractions. In Supplement Figure S11, we show the model-observation comparison with this filtering technique. We also report the Normalized Mean Bias (NMB), Root Mean Square Error (RMSE), and Normalized Mean Absolute Error Factor (NMAEF). In the revised manuscript, in line 268 we add the following:

"The Grosvenor et al. (2018) dataset has several assumptions and thresholds to calculate N_d. An important assumption is a minimum of 80% cloud cover in a $1^\circ \times 1^\circ$ gridbox. Our model diagnostics are written out as monthly averages, to avoid the high cost in disk space associated with output at higher time resolution, and hence it is difficult to apply exactly the same threshold. We conduct a sensitivity study by filtering out the bottom 25th percentile of weights (the denominator in Equation 6) and their corresponding simulated N_d values, and mask out the same grid cells from the monthly-averaged MODIS dataset. This filtering ensures that we select gridboxes which has relative higher cloud fractions on average in the model. In Figure S8, we show the simulated and observed N_d with this filter in place. We find that the global NMB change with the updated ARG is similar to the unfiltered case (a change from -22% to 0.1% compared to the previous change from -21% to 2.8%). The changes to the RMSE and NMAEF with the filter are also small (RMSE increases from 59 cm^{-3} to 65 cm^{-3} compared to the previous change from 64 cm^{-3} to 72 cm^{-3}). While our investigation remains approximate, there is no strong evidence from this test that our results would be substantially changed if we were able to match the simulations to satellite data more precisely."



Section 3.4: The conclusion drawn in this section is somewhat unclear. It would be helpful to either explicitly state the implications of the findings or, if they are not central to the paper's argument, consider removing or rewording this part. This would enhance the clarity and focus of the discussion.

We thank the reviewer for the suggestion. We agree that the motivation for Section 3.4 was not very clear. Hence, in the revised manuscript, on line 329 we add the following:

"The alternative f and g we test here may be useful to modelers who wish to mitigate the biases in the default ARG parameterization while still favoring underestimating rather than overestimating the activation fraction predicted by the parcel model."

L85: To improve clarity, it would be helpful to show the formula of non-dimensional variables η and ζ. This would allow readers to better understand the discussion around kinetic limitations and how these variables related to Nd and w.

We thank the reviewer for this suggestion. We agree that a better description of η and ζ would help the readers better understand our updates to the ARG scheme. Hence, following the suggestions of the reviewer, we added the expression for η and ζ, and in the revised manuscript, in line 84 we have added a brief description.

"ζ and η_i are given by: :

$$\zeta = \frac{2A}{3}\left(\frac{\alpha w}{G}\right)^{\frac{1}{2}}$$
$$\eta_i = \frac{2\left(\frac{\alpha w}{G}\right)^{\frac{3}{2}}}{\gamma^* N_i}.$$

Here A is the Kelvin coefficient (a function of temperature), w is the updraft velocity, α is a thermodynamic term (function of temperature) that relates the updrafts to the tendency for water vapor to condense as it cools, N_i is the aerosol number concentration in mode i, γ^* follows from the thermodynamics of rising moist air with assumptions listed in Pruppacher and Klett (2010), and G is the growth coefficient which depends on the diffusivity of water vapor in air and on the thermal conductivity of the air. The ratio of ζ and η_i is proportional to the ratio of N_i and w. Since none of these parameters depend of the mode width, Abdul-Razzak and Ghan (2000) parameterized the dependence of S_{max} on σ by using two additional parameters, f and g."

L130: Could you clarify the nudging relaxation time scale?

We thank the reviewer for this comment. We agree that the choice of relaxation parameter is important. In this work, we set the relaxation parameter to 1/6 h^{-1}. In the revised manuscript, on line 141 we add the following:

"The choice of the nudging relaxation parameter is important; too small a value makes nudging ineffective, while too large a value can destabilize the model. Following Telford et al. (2008), we use the relaxation parameter of 1/6 h^{-1}, corresponding to the time spacing of the ERA5 data."

L174: 'AAF' should be changed into 'TAF'.

Changed (see line 189)

L202: the modification to the constant 'p' does not fix the underlying problem but alleviates it. It would be insightful to discuss why the improvement appears to be limited and consider other possible explanations or avenues for further improvement.

We thank the reviewer for their suggestion. We agree that more discussion is needed to better understand the limitations of the fix we proposed in the manuscript. Some activation parameterizations are able to treat kinetic limitations in a more principled way (Nenes and Seinfeld, 2003; Morales Betancourt and Nenes, 2014; Ming et al., 2006). However, some empirical relationships are still involved, for example to define the regime where kinetic limitations become dominant (Nenes and Seinfeld, 2003). Our approach maintains the simplicity of the ARG algorithm, in which no iterations or bisections are needed, but its disadvantage is that the kinetic growth is not treated in a physically motivated way: this is the reason for the persistence of poor performance in extreme cases.

If we tried to add more physics we would probably end up with something similar to the Nenes and Seinfeld (2003); Morales Betancourt and Nenes (2014); Ming et al. (2006) algorithms. Alternatively, of course, one could also guarantee fidelity to the parcel model using machine learning-based emulators, for example those of Silva et al. (2021). This would require larger changes to the code base of a climate model to implement, and may reduce the ease with which the behavior of the scheme can be interpreted.

More sophistication, for example considering hydration of aerosols and accounting for giant CCN using expressions based on first principles, could also improve model performance.

In the revised manuscript, on line 219 we add the following:

"Unlike our modification, the approaches of Nenes and Seinfeld (2003); Ming et al. (2006) have physically motivated treatments of kinetic droplet growth. However, introducing additional physics into the ARG scheme would likely lead to a formulation similar to those algorithms, compromising the simplicity. Alternatively, of course, one could also guarantee fidelity to the parcel model using machine learning-based emulators, for example, those of Silva et al. (2021). This would require larger changes to the code base of a climate model to implement and may reduce the ease with which the behavior of the scheme can be interpreted."

L241: 'Figure S12' should be changed into 'Figure 4'.

Changed (see line 300)

References

Abdul-Razzak, H. and Ghan, S. (2000). A parameterization of aerosol activation: 2. multiple aerosol types. *Journal of Geophysical Research*, 105:6837–6844.

Grosvenor, D. P., Sourdeval, O., Zuidema, P., Ackerman, A., Alexandrov, M. D., Bennartz, R., Boers, R., Cairns, B., Chiu, J. C., Christensen, M., Deneke, H., Diamond, M., Feingold, G., Fridlind, A., Hünerbein, A., Knist, C., Kollias, P., Marshak, A., McCoy, D., Merk, D., Painemal, D., Rausch, J., Rosenfeld, D., Russchenberg, H., Seifert, P., Sinclair, K., Stier, P., van Diedenhoven, B., Wendisch, M., Werner, F., Wood, R., Zhang, Z., and Quaas, J. (2018). Remote sensing of droplet number concentration in warm clouds: A review of the current state of knowledge and perspectives. *Reviews of Geophysics*, 56(2):409–453.

Ming, Y., Ramaswamy, V., Donner, L. J., and Phillips, V. T. J. (2006). A new parameterization of cloud droplet activation applicable to general circulation models. *Journal of the Atmospheric Sciences*, 63(4):1348 – 1356.

Morales Betancourt, R. and Nenes, A. (2014). Droplet activation parameterization: the population-splitting concept revisited. *Geoscientific Model Development*, 7(5):2345–2357.

Nenes, A. and Seinfeld, J. H. (2003). Parameterization of cloud droplet formation in global climate models. *Journal of Geophysical Research: Atmospheres*, 108(D14).

Pruppacher, H. and Klett, J. (2010). *Microphysics of Clouds and Precipitation*, volume 18.

Silva, S. J., Ma, P.-L., Hardin, J. C., and Rothenberg, D. (2021). Physically regularized machine learning emulators of aerosol activation. *Geoscientific Model Development*, 14(5):3067–3077.

Telford, P. J., Braesicke, P., Morgenstern, O., and Pyle, J. A. (2008). Technical note: Description and assessment of a nudged version of the new dynamics unified model. *Atmospheric Chemistry and Physics*, 8(6):1701–1712.