

The authors would like to thank the referee for thoroughly reviewing the manuscript. We have re-run some of the simulations, re-made almost all the plots, and modified the manuscript significantly, according to the suggestions and comments from the reviewers. Our fundamental findings remain unchanged. Please see below our response to each of the referee's comments starting with "**Response:**" and "**Changes Made:**". The line numbers ("L") referred here is based on the revised manuscript, not the revised manuscript with tracked changes.

Response to General Comments

This manuscript describes a study investigating aerosol-cloud interactions (ACI) in shallow cumulus and our ability to meaningfully represent them in LES. The authors conducted a "clustering analysis" to identify 4 distinct aerosol populations which were present during the TRACER field campaign, and then based on these populations they ran LES of a single shallow cumulus scene under various canonical aerosol conditions. Additionally, the study compared simulations of the control-case for aerosol conditions between LES using a bulk microphysics scheme and LES using the superdroplet method (SDM).

The study is clearly a substantial contribution to the atmospheric physics community, contributing to our understanding of how to characterise aerosols (in this particular case over Greater Houston), and how to interpret, through physical mechanisms, the results of ACI studies (here in the case of shallow cumulus). Their results are especially important for showing that the use of bulk schemes for studying ACI may be confounded by the unrealistic formation of precipitation as compared to SDM, which is somewhat corroborated by observations.

I found this manuscript quite exceptional both in terms of its scientific quality and presentation. The manuscript is well written and its arguments are compelling. As a result of thoroughly enjoying reading the manuscript I may have been a bit over-enthusiastic with my comments. This should be viewed as a sign of the manuscript's high quality, both in terms of readability and scientific content, which made it possible for me to engage with. Many of my comments could be viewed as suggestions / minor-revisions; the only points where I suggest substantial revision is in the description of the clustering method and the explanation of the clusters' relation to sea breezes, as well as in refining one of the key conclusions of the paper regarding the improvements required of bulk schemes. Notwithstanding all of my comments (somewhat over-eagerly given!), I recommend this paper for publication in ACP journal.

I look forward to the response of the authors to my comments, and also their forthcoming work on aerosol recirculation in developing cumulus.

Response: Thank you for your review comments. Below you will find our responses to more specific comments in greater detail.

Response to Specific Comments

Response to Major Comments

- **Section 2.1:** The description of the aerosol clustering would greatly benefit from a short description and schematic of the Atwood et al. (2017, 2019) methodology. I get the impression that the current state of Section 2.1 is useful for readers already familiar with this analysis method and may be useful for researchers who wish to follow-up on the work presented here, however, it is extremely difficult for readers with no previous knowledge of this method to understand, even on a surface level. This also makes it hard as a reviewer to comment on what errors/uncertainties need to be considered when deeming the identification of 4 clusters as appropriate (see below). In order to cater to a wider audience than just those familiar with Atwood et al. (2017, 2019), it is necessary to add an overview/description of the methodology in this section, or in a short appendix.

Response: We agree that the original version of Section 2.1 included minimal information on K-means clustering, which is an essential methodology for this work. In the revised version of this section, we now explain the basis of K-means clustering. As schematics of K-means clustering are not included in Atwood et al. (2019), we did our best to describe in words the fundamentals of K-means clustering and its performance evaluation as follows.

Changes made: Added “*K-means clustering is an unsupervised learning algorithm where input data are split into K groups or “clusters”, based on the similarity of input variables and their combinations. As this is unsupervised learning, there is no correct answers or “ground truth” given to the learning process, which fits well with our data where the correct group/cluster where each data point belongs to is unknown. In this algorithm, K number of cluster centroids are randomly chosen at the beginning, and each data point is assigned to the cluster with the nearest centroid. Then, the cluster centroids are recalculated, based on all the data points that now belong to that cluster. This recalculation is followed by the cluster reassignment, and recalculation of the cluster centroids, and so forth. By repeating this process, the clustering process converges (i.e., centroids stop moving much or remain unchanged) and K groups of data with similar features can be identified. The optimal value of K is determined through multiple testing, such as the Elbow Method (Thorndike, 1953) and the Silhouette Coefficient (Rousseeuw, 1987),—as explained in greater detail in Section 3.1. In the elbow method, the sum of squared errors (SSE) is first calculated when K is set to a value between 1 and 10. SSE is expected to decrease as K increases, since the number of cluster centers increases and therefore the distance between each point and its cluster center must decrease. When this reduction of SSE with K slows down, the K shows a deflection point or an “elbow” where a further increase of cluster number only marginally improves the clustering results. The Silhouette coefficient is a measure of how well each data point fits into its assigned cluster, ranging from -1 (closer to another cluster center) to 1 (well in the cluster), the higher the better.*” to Section 2.1.

- **L130-134:** As eluded to above, I cannot grasp what errors or uncertainties may be introduced by the remapping of the observational data to align more closely with Atwood et al. (2019). Superficially the coarsening of the data appears alarming and I question how the clusters would differ with different choices in this methodology, i.e. how different choices would translate into error bars/shading in Figure 2, and especially Figure 2d (hence also Figure 5). Some tests of the clustering results to different choices in the clustering methodology would alleviate my concern that the identified clusters are not robust, or else an explanation of why this concern is not valid.

Response: It is valid to wonder if any errors/uncertainties are introduced to the clustering results through the remapping of the aerosol size distribution data. Remapping was done both to reduce the number of input variables and to smooth the size distributions as raw data with fine bin sizes can be very noisy. Since K-means clustering is unsupervised learning (i.e., there is no “ground truth”), the “error” cannot be assessed due to the lack of answers to compare the clustering results with. Once we change the setting of the K-means clustering, such as the number of input variables or weight values, it is unlikely that the exact same 4 clusters will be identified and each data point gets classified into the same cluster (this is more possible with supervised clustering where the ground truth is known, and the training is based on it). In the course of identifying the presented results, we have tested about 20 different sets of input variables (this includes different remapping of aerosol size bins), weight values, and the number K through trial and error. The results varied, but similar 3-4 clusters were identified in many of the tests. This “robustness” may sound somewhat looser than what can be expected from supervised clustering or unsupervised clustering with known number of clusters, but this is due to the complexity of the real atmosphere and aerosols in the region. We have added a few sentences in the main text to emphasize our trial-and-error attempts and the reason for remapping the aerosol size distributions as follows.

Changes made: Changed to “*The~~is~~ re-mapping of the aerosol size bins is done to reduce the number of size-distribution variables for clustering to a similar value as that used in Atwood et al. (2019), which was 20, and to smooth out the noise from a large number of bins with fine bin widths.*” in L154. Added “*It should be noted here that, in the course of this study, we tested multiple sets of input variables and weight values that are different from those in Table 1. The clustering results varied, but similar clusters appeared repeatedly and the variables and weight values shown in Table 1 seem to provide us with results that are representative of outcomes from our trial and error.*” to L172.

- **L131:** Why do the droplet size bins range only until 478.3 nm when, if I understand correctly, the original data ranges until 1 micron? Is there an appreciable mass (even if not number) of aerosol in higher droplet sizes and, if so, how would the results of this study differ if that had been accounted for in the simulations (i.e. in Figure 5), especially

regarding precipitation? It would therefore be beneficial to have an explanation for the truncation of the DSDs and a conjecture of whether or not larger aerosols would have been significant.

Response: This is a valid point and needs to be clarified. The reason for using a subset of data for the clustering analysis, rather than utilizing the full instrument range, is because we wanted to identify aerosol clusters that differ over the size range between ~10 nm and ~500 nm, which is the range that often matters the most to aerosol activation into cloud droplets. If we included aerosols < 10 nm, they would dominate the relative concentrations to the total number and the clustering results would largely be based on their numbers, which may have resulted in clusters that differ greatly in the number of small aerosols but not so much in the range that matters to aerosol activation (i.e., 10-500 nm). As for the larger end (> 500 nm) that the reviewer pointed out, their contribution to the total *number* would be very low, and it is the *number* that is used as an input for this clustering (Table 1), and hence we prioritized the reduction of input features (and the computational cost) over the inclusion of larger particles that have very low numbers and therefore differ only slightly among input data. This point has been clarified in the revised manuscript as follows.

Changes made: Added “*Since we aim to identify aerosol clusters that differ over the primary size range for aerosol activation (i.e., roughly between 10 nm and 500 nm), we use the subset of the SMPS data for the clustering rather than its full size range.*” to L152.

- **L269-277:** Based on Figure 2e, I am not fully convinced by this explanation, please be more precise / clarify what is meant in the following:
 - How does Cluster 4 fit with the sea breeze argument? I interpret Cluster 4 in Figure 2e as a night-time phenomenon but this appears to contradict with the statements that “A continental air mass (i.e., Cluster 1) is more likely to reach the AMF1 site during nighttime/morning when a land breeze strengthens” and that “sea-breeze air mass (i.e., Clusters 2 and 4) in the afternoon/evening when a sea breeze prevails.”. It is also unclear to me how the Park et al., 2020 relates to the frequency of occurrence of Clusters 2 and 4. Also later, in L280 and L461, Cluster 4 is referred to as occurring in morning and evening (as opposed to night-time).

Response: We agree that the original description of the diurnal cycle, along with the labeling of Cluster 4, was confusing. We have significantly modified the paragraph and the labeling of Cluster 4 throughout the manuscript. Now we refer to Cluster 4 as “*Marine+Rural*” rather than “*Sea Breeze + Rural*”, since the timing of sea breeze should be late afternoon rather than nighttime as the reviewer pointed out. Please see the following changes.

Changes made: Added “... *For example, Cluster 2 or the Extreme Pollution cluster is most frequently observed midday (i.e., ~10am-2pm LT) when the industrial activities contribute to*

large aerosol emissions while the effects of convective dispersion and sea breeze are still limited. For Cluster 1 (Continental), its peak frequency is seen in the early morning (i.e., ~6-9am LT), which may be partly a result of land breeze or relative weakening of sea breeze. As for Cluster 3 (SB+Pollution), its frequency peaks in the late afternoon (i.e., ~2-10pm LT), which is consistent with the prevalence of classic sea breeze in the late afternoon as well as the continued emission of small particles from industrial activities. The interpretation of the diurnal cycle for Cluster 4 (Marine+Rural) is somewhat complex; the highest frequency is seen at nighttime (i.e., ~10pm-6am LT), which is in between the peak times for Clusters 3 (SB+Pollution; 6pm) and 1 (Continental; ~8am). There also seems to be a small second peak at 10am LT, which is in between the peak times for Clusters 1 (Continental: ~8am) and 2 (Extreme Pollution; ~noon). Given these timings, combined with the high frequency of Cluster 4 (i.e., ~45 %), we speculate that Cluster 4 includes all cases with southerly/southwesterly flows that are (1) in between the other three clear clusters, (2) sea-breeze cases that did not fall under the classic sea-breeze group of Cluster 3 due to somewhat different winds and aerosols, and (3) associated with strong synoptic patterns. Regardless of the exact origin of the air mass, the aerosol size distribution of this cluster shown in Fig. 2d strongly indicates the presence of large marine particles as mentioned above, due to the southerly wind components (Fig. 3). This observation justifies our labelling of Marine+Rural for this cluster.” to L306 and deleted “~~These observations align well with our assigned cluster labels. A continental air mass (i.e., Cluster 1) is more likely to reach the AMF1 site during nighttime/morning when a land breeze strengthens, a local polluted air (i.e., Cluster 3) may be observed at the AMF1 site midday when industrial activities are ongoing and before a sea breeze arrives and brings in a non-local air mass, and a sea breeze air mass (i.e., Clusters 2 and 4) in the afternoon/evening when a sea breeze prevails. As discussed in Park et al. (2020), sea breezes propagate faster during nighttime due to less turbulence associated with surface heating, which is consistent with the prevalence of Clusters 2 and 4 in the late afternoon and evening. The smooth size distribution suggests that aerosols in Cluster 1 may be aged and/or from many different sources.~~”.

- It’s ambiguous if statements in this paragraph are referring to more/less frequent as compared to each Cluster’s own average daily occurrence, or as compared between clusters, e.g. “Cluster 2 is more prevalent in the afternoon, whereas Cluster 4 is more frequently observed in the evening and morning.”. From Figure 2e, Cluster 4 is almost always more frequent than all the other clusters, so the reader can assume it is meant as more/less frequent relative to each Cluster’s own occurrence, but this should be clarified. And perhaps comparing occurrence between Clusters makes no sense because there are observational biases in how often different clusters are detected??

Response: In the revised manuscript, we clarified this point by adding “... *it is interesting that each cluster has a distinct diurnal pattern that is evident when individual frequency peak is*

examined.”. As the reviewer mentioned, comparing frequencies among different clusters simply tells us how often that particular cluster was observed in the absolute sense but not much about the diurnal cycle, which is more relative within a cluster.

Changes made: Added “... *it is interesting that each cluster has a distinct diurnal pattern that is evident when individual frequency peak is examined.*” to L305.

- **L426-429:** I find these statements quite important conclusions of this manuscript, however I think they could be more clearly differentiated / refined. To me points (1) and (2) are encompassing two fundamental problems of bulk schemes: (A) issues with DSD representation in bulk schemes (of aerosol and of droplets), and (B) issues with microphysical process representations in bulk schemes. However the current separation in (1) and (2) conflates them, making the separation instead between the activation and growth of so-called “cloud” droplets and then the growth of so-called “rain” droplets. This division may be how bulk schemes are constructed, but it is not really how the fundamental problem arises, nor how nature behaves (of course we cannot observe “autoconversion” — it is a modelling construct!). Perhaps this point is rather academic, but please consider re-dividing the problems contained in (1) and (2) along more natural/fundamental lines, as opposed to bulk scheme ones. For example, (1) “a deficiency in the representation of droplet activation rates” includes both the parametrisation of aerosol activation (point B) as well as the issue of bulk schemes prescribing constant aerosol distributions (point A). The same comment applies to L442-448 and in similar ways also to the conclusion in L468-471.

Response: The separation of the two points was, as the reviewer mentioned, between aerosol activation or CDNC and droplet-to-raindrop conversion in the original manuscript. The reviewer suggests the separation to be instead between aerosol/droplet size distribution (point A) and parameterization of microphysical processes (e.g., activation and droplet-to-raindrop conversion) within the microphysics scheme (point B). This separation does makes sense, and interestingly, we did use exactly this point-A/B separation when we first drafted the manuscript. However, we realized that the parameterizations of autoconversion/collision-coalescence and accretion used in this bulk Morrison scheme from Khairoutdinov and Kogan (2000) *do not* depend on droplet size distribution (DSD), but only on total cloud mass (q_c) and CDNC. Although DSDs indirectly contribute to q_c and CDNC (as an accurate DSD *would* lead to accurate q_c and CDNC), we intentionally did not point to aerosol size distributions and DSDs as a source of error with the awareness that some bulk schemes do not even include those size distributions into calculations. One can argue that this is a deficiency and a source of error for bulk microphysics schemes, but others can also argue that process rates *can* be accurately calculated without accurate DSDs if q_c and CDNC are right. Based only on our simulation results, we do not know how the lack or misrepresentation of DSD would lead to erroneous process rates, as the parameterizations in

question (i.e., collision-coalescence and accretion) do not depend on DSD in this study. We therefore retain the current separation of the two points. However, we have added the following sentences to the end of this paragraph to clarify this point above on aerosol size distributions and DSDs.

Changes made: Added “*Misrepresentations of aerosol and droplet size distributions can also contribute to erroneous process rates in bulk schemes. In our bulk scheme, however, neither the autoconversion parameterization nor the accretion parameterization depend directly on droplet size distributions (i.e., only on Q_c and $CDNC$), and therefore, we cannot point to aerosol/droplet size distributions as a reason for the overestimated rain production in our bulk simulations.*” to the end of the paragraph in L476.

Response to Minor Comments

- **L75-78:** In discussing the advantages of SDM, it would be a good idea to mention SDM’s explicit/fundamental representation of (de-/re-)activation, droplet evaporation and condensation (including droplet nucleation) according to Köhler theory, as opposed to “the growth process of wet aerosols into cloud droplets”. I think the current phrasing neglects droplets shrinking and de-activation and re-activation, all of which are likely to be relevant to this study.

Response: We have modified the sentence accordingly. Please see the change below.

Changes made: Changed the sentence to “... the major advantages of Lagrangian schemes include the capability to explicitly track the growth (e.g., activation, condensation) and shrinkage (e.g., deactivation, evaporation) ~~process of wet aerosols into cloud droplets~~ following the Köhler theory (Köhler, 1936),...” in L92.

- **L77-78:** I find “the elimination of artificial size thresholds”, a bit imprecise and “size thresholds” doesn’t generalise well when we talk about the cold-phase too, or more complex aerosols and/or aerosol chemistry. Just focussing on the warm-phase it may be clearer to state explicitly that it’s the elimination of categories which distinguish between cloud and rain droplets. In general it’s the elimination of the artificial hydrometeor categorisation, which in the warm-phase distinguishes between cloud and rain droplets. (At least I think that’s what you mean — please correct me if I misunderstood what was the intended meaning behind “size thresholds”!)

Response: We have modified the sentence as follows to clarify the meaning.

Changes made: Added “... and the elimination of artificial categorization of drops into cloud droplets or raindrops using a size thresholds for certain cloud processes.” to L95.

- **L100:** “LES scale” is too vague, please be more specific about the scale of interest, preferably in words not numbers.

Response: We have added somewhat longer description of an LES scale to its first appearance in L69 as follows.

Changes made: Modified to “... *on the large-eddy simulation (LES) scale that can capture fine variability of boundary layer structure, cloud structure, and ~~microphysical processes~~ turbulent flows associated with large eddies (i.e., often ≥ 10 -100 m), ...*” in L69.

- **L169-171:** This sentence is a bit hard to read and contains many points; consider separating the definition of multiplicity into a separate sentence and clarifying what is meant by “the Lagrangian framework” — such terminology for readers not familiar with SDM may invoke parcel trajectory modelling or something even further afield. Furthermore, clarifying “the Lagrangian framework” would also be a good opportunity to emphasise how well SDM’s representation of cloud microphysics (that is, hydrometeors and cloud microphysical processes,) matches our fundamental particle-based theory of cloud microphysics — in several places this point could be “read between the lines”, but this may be a good opportunity to make the point directly, especially for the benefit of readers less familiar with SDM.

Response: We have modified the sentence and greatly expanded on the Lagrangian approach in L197 as follows.

Changes made: Modified to “*In SDM, microphysical processes of individual particles or “super-droplets” (SDs) are calculated, which enables realistic representation of individual particle growth and decay in simulations. Each SD, ~~each of which~~ carries a weight factor called ‘multiplicity’ that is (i.e., equivalent to the number of physical particles each SD represents in a given volume of air location.), ~~are calculated in~~ Since the evolution of each SD is tracked, the microphysical calculations of SDM are ~~the a~~ Lagrangian framework approach, in contrast to the fixed-in-space Eulerian framework used by the host CM1 model for the dynamical flow field computations and by the bulk Morrison scheme for microphysical processes.*”

- **L193:** from Figure 1 is it hard to understand why the simulation was cutoff after 8 hours; consider extending Figure 1 beyond (and possibly before) the LES simulation period to show how the cumuli continue to develop that day, and please also justify why a longer simulation was not necessary.

Response: We cut the simulations at 8 hours due primarily to the computational cost of the SDM run. Another reason is the potential for higher departure of the simulation results from observations if the simulations continue longer, given that this is an idealized simulation and the observational data is directly ingested only at the beginning. We agree that the simulation could continue until the clouds are completely dissipated, for instance up to 12 hours. However, adding 4 extra hours to the current simulations would not change our conclusions, given that the rainfall starts well before the simulations hit 8 hours in the bulk runs and the SDM run likely would not have any rainfall even if the simulation is extended by 4 hours. Given the high computational cost in comparison to low value of the extra 4 hours, we will not run additional simulations, but we have added the following sentence in L228.

Changes made: *“To simulate these clouds with CM1, we run the simulations for eight hours, starting at 7am LT. Within eight hours, most of the cloud evolution is captured, which allows us to assess the discrepancy in cloud morphology and microphysics among the simulations while minimizing the computational cost for the SDM run.”*

- **L196-197:** Why is the scaling of superdroplets per grid-box done in this way? Is it to keep their multiplicities similar? Some information on how superdroplets are initialised would also be beneficial, especially whether or not there are any super-droplet sources during the simulations (I assume there are none but I am not certain).

Response: The original sentence/description here was confusing and somewhat misleading. We modified the sentence as follows to clarify the meaning here in L232. There is no additional superdroplet (SD) sources in the simulations, which is mentioned in L234 too. In L202, we now mention how the initial multiplicities of SDs are determined.

Changes made: Modified to *“In SDM, the ~~average number of SDs at the beginning of the simulation is 128 per grid at surface, and this number~~ multiplicities of SDs ~~are~~ is also vertically scaled by the same scaling factor from TRACER-MIP. The average number of SDs at the beginning of the simulation is 128 per grid at surface. No additional SDs are introduced to the simulation during the run.”* in L232. Added *“Initial multiplicities of SDs are determined by aerosol size distribution and dry particle sizes assigned to SDs.”* in L202.

- **L193-206:** Some plots showing these comparisons between the LES thermodynamics and the observations would be insightful, even if only in the supplement.

Response: These are observational data used as either an initial or boundary condition of the simulations, and therefore, the comparison of simulation inputs/outputs will show a match to the observations. We have included a supplementary figure to show these dynamical conditions for reference.

Changes made: Added Figure S1 and added “*These dynamical inputs and boundary conditions are presented in Fig. S1.*” in L244.

- **Figure 2:** I would highly recommend also including a plot of the mass size distribution, either in Figure 2 or in Figure 5. Also what is meant here by “cluster-median”? If there is a spread in the distributions of the DSDs in each cluster then it is important to represent this in these results, or at least in the supplement if it makes the plot too chaotic here.

Response: The aerosol/SMPS data does not include mass size distributions; we can calculate volume size distributions without assuming a particle density, but we are unsure if this plot would help readers better understand the paper. The size distributions shown in Figs. 2d and 5 were obtained by calculating the median concentrations for all the size bins. The authors think that the spread is large and showing the large spread here would make the figure more confusing. To clarify that these are median size distributions, we have modified the caption of Figure 2 as follows.

Changes made: Changed to “... (d) ~~cluster~~-median SMPS size distributions [$\text{cm}^{-3}(\text{dln}D)^{-1}$] composed of median values from all size bins, and (e) hourly counts of cluster occurrences. The numbers in the brackets within the legend in (b) and (c) are the ~~cluster~~-median values within the clusters.” in L272.

- **L312:** “cloud-base height increases with time from 500 m to 1.5-2 km” I dispute that the cloud base height in the observations reaches 1.5km, from Figure 1d it looks more like 1.25km. If so, then it should be pointed out that the cloud-base height is actually often overestimated, i.e. it rises too fast in the LES. Otherwise please explain how the cloud-base height in the observations is actually higher than it seems from Figure 1d.

Response: We have checked the exact value of the minimum height for non-zero values and it is 1.3 km at 1:30 pm LT, and hence we address this point by changing the sentence as follows.

Changes made: Modified to “*Although the cloud-base and ~~maximum cloud~~-cloud-top heights ~~are~~ both overestimated by 500 m in the simulations (i.e., $\sim 1.8\text{km}$ in comparison to $\sim 1.3\text{ km}$ as observed for cloud base, $\sim 3.5\text{ km}$ in comparison to $\sim 3\text{ km}$ as observed for cloud top), ...” in L349.*

- **L376-378:** Please be more specific about “macrophysical properties”, this study seems to only look at cloud fraction. I would expect that individual cloud lifetimes (within the scene) are different depending on the microphysics, and their precipitation certainly is different, both of those things I am inclined to view as “macrophysical” too.

Response: We agree that precipitation is also one of macrophysical features, and therefore, we have removed the word “macrophysical” throughout the manuscript and modified the original sentence as follows.

Changes made: Changed to “..., which signifies the primary control of dynamical conditions on the cloud development, evolution, and cloud-base and cloud-top~~macrophysical~~ features for the simulated shallow cumulus clouds, as opposed to the strong influence from microphysics. While these ~~macrophysical~~ features of the clouds are relatively unimpacted by the choice of microphysics schemes, ...” in L420.

- **L379-382 and L385-387:** These lines suggest there is something “counterintuitive” in the results for cloud mass and CDNC, possibly attributable to numerical diffusion. To me these results seem to be explainable by the fact that the bulk schemes rain out, whereas SDM does not. Have I missed something here? If so, please explain. I also do not fully understand how the supply of aerosols could play a role here because I don’t see signs of significant aerosol depletion in Figure 8e, i.e. it looks to me as though SDM doesn’t suffer from an aerosol supply deficiency until maybe very late in the simulation (circa. 1pm) — which may perhaps be rationalised by the fact these are very polluted simulations. Although perhaps I have misunderstood something here too. If so, please explain this too.

Response: Both points are very valid, and we have modified the sentences as follows.

Changes made: Modified to “In-cloud-mean cloud ~~mass~~liquid water content, shown in Fig. 6i-j, is 30 % ~~high~~larger on average in the SDM run (Fig. 6j)throughout the simulation, which suggests “thicker” clouds therein. Although higher condensation is expected for the bulk simulations upon initial aerosol activation due to ~~and is somewhat counterintuitive given the use of saturation adjustments (i.e., all supersaturated water vapor converted into liquid) in the bulk scheme, this higher cloud mass in the SDM run can be partly explained by the smaller conversion rate of clouds to rain as shown below, and/or that convert all supersaturated water vapor into liquid. This tendency of lower cloud mass in the bulk run, alternatively, can be a result of numerical diffusion.~~” in L423. Also, deleted “~~...is another counterintuitive discrepancy between the two simulations given that the bulk simulation gets an infinite supply of aerosols while the availability of aerosols is limited in the SDM run (i.e., Sect. 2.2).~~”.

- **L389-390:** Please rephrase “In the SDM run, non-zero rain mass is present in the atmosphere, ...”, I think this is extremely misleading. From my understanding of Figure 8, the maximum rain mass in SDM is at least 8 orders of magnitude smaller than in the bulk schemes — if SDM was a person, a bulk scheme would be about 10x the population of Texas, or the entire population of Japan — I think it’s safe to say the SDM rain mass is insignificant. Please also justify/explain or remove the following statement that “These

bulk properties of the clouds in the SDM run still technically follow the pattern of “more aerosols, less rain” found in the bulk simulations ...”

Response: We have modified the first sentence accordingly as follows. Also, we noticed the error in the second sentence and removed it.

Changes made: Modified to “*In the SDM run, ~~non-zero~~ rain mass is negligible present in the atmosphere, yet as its amount is orders of magnitude less than that in the bulk and the rain does not reach the ground (Fig. 9).*” in L431 and we removed “*These bulk properties of the clouds in the SDM run still technically follow the pattern of “more aerosols, less rain” found in the bulk simulations (Sect. 3.1), ...*”.

- **L403:** Please elaborate on “either by evaporation or conversion into raindrops”. From Fig. S4 the evaporation rate in SDM appears to be considerably larger and so I do not fully understand how evaporation can play a role here, perhaps it is really rather just raindrop conversion? Or maybe plotting evaporation rate also outside the cloud would help clarify this point.

Response: Thank you for mentioning this important point. We have modified the sentences as follows to reflect our findings regarding droplet activation rates and CDNC, based on an additional test simulation and our updated simulation results.

Changes made: Modified to “*Based on our analysis, the lower CDNC in the bulk scheme, despite the slightly higher activation rate, seems to mainly stem from the constraint on the slope parameter of the droplet size distribution therein; due to the minimum and maximum values of the slope parameter set in the scheme, CDNC cannot increase (or decrease) freely in the bulk scheme with the background aerosol loading.*” in L441.

- **L410:** “lower aerosol concentrations in CM1-SDM near the cloud base ($< \sim 1.5$ km)” really only applies to Figure 10a. Consider alternatively that it is lower aerosol concentrations in the sub-cloud layer and particularly towards the surface. It would also help to plot in Fig. 10 the cloud base of the bulk and SDM schemes (as in Figure 6d-e) to show this.

Response: We have modified the sentence accordingly and added the simulated cloud-base heights in the figure as a reference.

Changes made: Modified to “*The figure clearly presents the lower aerosol concentrations in CM1-SDM ~~near the cloud base ($< \sim 1.5$ km)~~ in the sub-cloud layer particularly towards the surface compared to the bulk run.*” in L451 and Fig. 10.

- **L442-448:** I am of the opinion there is already quite an extensive literature on some of these subjects, and so these points need to be refined. As an example, some of the works Adele Igel has co-authored (e.g. <https://doi.org/10.1029/2023MS004068>, <https://doi.org/10.1029/2022MS003039>, <https://doi.org/10.1175/JAS-D-15-0383.1>, <https://doi.org/10.5194/acp-17-4599-2017>) evaluate the assumptions of bulk and bin models, albeit perhaps not in comparison with observations. Notwithstanding this, other comparisons of bulk microphysics schemes to observations are abundant.

Response: We appreciate the suggestion. Here, our focus is specifically on the use of observationally constrained Lagrangian microphysics schemes to evaluate bulk scheme, rather than on the evaluation of bulk/bin schemes against observations. We emphasize this point in the revised manuscript as follows. However, we mention two of the papers the referee suggested in response to another comment below.

Changes made: Changed to “*It is therefore critical to assess the validity of bulk and bin simulation results and identify the reason for their discrepancy from ~~the~~ observationally constrained (Sect. 2.2) particle-based Lagrangian simulations.*” in L485.

- **Section 3.2 or Section 4:** I think a discussion is required of what this study misses by keeping the LES setup entirely the same except for changing the aerosol distributions in order to investigate ACI. Consider that Cluster 4 and Cluster 3 occur at almost diametrically opposite times of the day, and the implications this has for the radiative forcing on clouds, and potentially even the cloud-type which the aerosols play a role in. How might this change the results? In other words, the setup of this study prevents an analysis of how concomitant aerosol and large-scale forcing cycles/changes could complement or negate one-another, and this should be discussed.

Response: We agree that this needs to be mentioned. We have therefore added the following paragraph at the end of the Conclusion section in summarizing what this study did not cover.

Changes made: Added “*On the other hand, some aspects of ACI in shallow cumulus clouds in the region were not covered in this study. These points include (1) horizontal variability of aerosol concentration and size distribution; ~~Furthermore, using a horizontally variable aerosol composition to representing~~ spatially inhomogeneous aerosol sources ~~in future work~~ may reveal more complex ~~aerosol-cloud interactions~~ ACI in the region, (2) variability of chemical composition or hygroscopicity κ among the clusters; inclusion of chemical composition data (e.g., CCN, INPs) into clustering may enable the classifications of aerosols directly based on their potential impacts on clouds, and (3) differences in dynamical conditions and diurnal patterns that are particular to each cluster; all the clusters are observed under a particular wind pattern and frequently at a specific time of the day, which indicates their association with a particular atmospheric structure, flow patterns, and potential for cloud-radiation interaction. Addressing these points in future work*”

might lead to classification of aerosols into a larger number of (sub-)categories that have very distinct microphysical and radiative impacts on shallow cumulus clouds in the region.” to L525.

- **L76:** “exclusion of assumptions for the hydrometeor DSDs” may be better phrased positively, e.g. the complete freedom of the evolution of hydrometeor DSDs. A more profound/fundamental description of this advantage of SDM would perhaps instead point out that DSDs in SDM are emergent/prognostic rather than diagnostic — SDM model particles and only for analysis purposes are particles then turned into DSDs.

Response: We changed the sentence as follows to reflect the reviewer’s comment here.

Changes made: Changed to “... *complete freedom of the evolution*~~exclusion of assumptions for hydrometeor~~ particle size distributions *(i.e., including aerosols, droplets, and raindrops), ...*” in L94.

- **L69-81 and L82-92:** For the flow of the introduction it could make more sense to swap the order of these two paragraphs, i.e. introduce SDM after introducing the need for modelling ACI of shallow cumulus.

Response: We swapped the two paragraphs here. To align with this change, the following sentence has been modified (i.e., Points 3 and 4 were swapped).

Changes made: Changed to “... (3) *the relative lack of studies on the sensitivity of shallow cumulus clouds to aerosols in spite of their high prevalence over the region, and (4) significance of comparing simulations with a bulk/bin scheme and a Lagrangian scheme to assess the potential error in ACI estimated by operational forecasting and global models;*~~and (4) the relative lack of studies on the sensitivity of shallow cumulus clouds to aerosols in spite of their high prevalence over the region, ...~~” in L101.

- **L434-439:** This could be a good place to point out that historically we have attempted to make bulk schemes “bin-like” — perhaps the tuning we have done to bulk models to make them look like bin schemes can explain some of the discrepancy here between bulk and Lagrangian simulations??

Response: Since this is under *Conclusions* and our study did not use a bin scheme, it may be more appropriate to mention bin schemes in *Introduction*. We therefore included a sentence below, along with two of the papers the reviewer referenced above, in L87 under *Introduction*.

Changes made: Added “*Bin schemes can often mitigate erroneous calculations arising from artificial assumptions in bulk schemes (e.g., Igel and van den Heever, 2017; Hu and Igel, 2024), but ~~b~~Both bulk and bin schemes suffer from numerical diffusion ...*” to L87.

- **L92:** I encourage the authors to make this statement bolder! Consider replacing “highly desired” with “required”.

Response: We have modified the words accordingly.

Changes made: Changed to “*Given the prevalence of shallow cumulus and their role in setting the deep convection stage, more studies on those clouds, especially with respect to their sensitivity to aerosol perturbations, are ~~highly desired~~required.*” In L80.

- **circa. L174-175:** it may also be helpful to state which terminal velocity parametrisation was used too.

Response: We added the following sentence about terminal velocity calculations in our simulations to L208.

Changes made: Added “*Terminal velocities of SDs are calculated based on Beard (1976) and the sizes of SDs.*” to L208.

- **L266-268:** The identification of four clusters is compelling, but their current presentation seems to underappreciate the similarities between Clusters 2 and 4 and hinder readability. If Cluster 2 was relabelled as Cluster 3 (and vice versa) it would be easier to compare and contrast the clusters which have almost-purely land-aerosol sources (Clusters 1 and new-2) and those which have sea-breeze + another contribution (Clusters new-3 and 4). I hope this does not sound too pedantic, I think this simple re-labelling could really aid the comprehension of the results, especially when it comes to following the discussion from L279-286. Also adding labels Continental, Extreme Pollution, SB+Pollution, and SB+Rural to the sub-figures or figure caption would be good.

Response: The original order of presentation simply followed the raw outcome from the clustering. We agree that swapping the orders of original Cluster 2 (Sea breeze + Pollution) and original Cluster 3 (Extreme Pollution) facilitates the comparison of original Clusters 2 and 4 (Marine + Rural), and hence we swapped them throughout the manuscript, table, and figures. As for the labeling in Figs. 2-4 (i.e., labeling them as “Continental” etc. rather than “Cluster 1” etc.), the naming of the clusters takes place *after* these figures are discussed in L302. The authors therefore think that it is better to refer to the data presented in Figs. 2-4 as “Cluster 1” etc. until the official naming is done in L302.

Changes made: Figures 2-9

- **L347:** “all simulations (Fig. S4)” is misleading, please rephrase as “all bulk-model simulations” or alike — evaporation rates in the SDM simulations appear up to 100 times higher in some regions.

Response: The sentence has been removed from the manuscript.

- **L430:** I think the phrase “we cannot obtain confirmation of (1), (2), or (3) from observations” needs tempering, perhaps we cannot *directly* obtain ... from *in-field* observations?

Response: We modified the sentence accordingly as follows.

Changes made: Changed to “... we cannot *directly* obtain confirmation of (1), (2), or (3) from *in-situ* observations.” in L473.

Response to Technical Corrections

- **L12:** “such datasets” doesn’t refer to any previous dataset mentioned, please reword.

Response: The sentence has been modified accordingly as follows.

Changes made: Modified to “*The intensive operational period (IOP) during the U.S. Department of Energy (DOE) Tracking Aerosol Convection Interactions Experiment (TRACER) field campaign provides ~~such~~ datasets that are ideal for studies on aerosol-cloud interactions.*” In L11.

- **L53:** “the Gulf” is potentially misleading out of context because the same terminology is used in the Middle East, please refer to “the Gulf of Mexico” as is internationally recognised.

Response: The change has been made throughout the manuscript.

Changes made: Modified to “*the Gulf of Mexico*” in L55 and L289.

- **L77:** “stochastic drop collision-coalescence based on the kernel function” is rather unusual terminology. A suggested rephrasing may perhaps be along the lines of e.g. stochastic droplet/particle collision-coalescence for a given collision kernel.

Response: The sentence has been modified accordingly as follows.

Changes made: Changed to “... *the use of stochastic ~~drop~~-particle collision-coalescence based on the for a given collision kernel-function*, ...” in L95.

- **L93:** something doesn’t read quite right about the numbered statement “In response to the above-mentioned points of (1) needs for ...”, please reconsider the wording and/or grammar?

Response: The sentence has been modified as follows.

Changes made: Modified to “*In response to the above-mentioned points including of (1) the needs for aerosol characterization...*” in L100.

- **L529:** In the references, the date for the citation of Atwood et al., 2017 is 2019, please correct.

Response: It has been corrected.

Changes made: Changed to “... *Observations and source classification, Atmos. Chem. Phys., 17, 1105–1123. <https://doi.org/10.5194/acp-17-1105-2017>, 2017*” in L582.

- **L170:** “representsin” is missing a space; also “location” invokes a fixed (often infinitesimal) point, consider instead “volume of space”.

Response: The modifications have been made as follows.

Changes made: Changed to “... ‘multiplicity’ that is (i.e., equivalent to the number of physical particles each SD represents in a given ~~location~~ volume of air.” in L198.

- **L178:** remove comma in “TRACER-IOP, based on our analysis”.

Response: The comma has been removed.

Changes made: “... *during TRACER-IOP, based on our analysis of ...*” in L213.

- **L180:** is a word missing in “and showed macrophysical features ... of shallow cumulus clouds in the region.”, perhaps “typical of shallow cumulus”?

Response: Thanks for pointing this out. The change has been made as follows.

Changes made: “... developed in a prevalent southeasterly wind and showed ~~macrophysical~~ typical bulk features (cloud-base/cloud-top heights and lifecycle) of shallow cumulus clouds ...” in L214.

- **L194:** for consistency, place UTC time in parenthesis and local time in main text.

Response: The corresponding has been made as follows.

Changes made: “The initial atmospheric profile comes from the variational analysis data at 7am LT~~1200 UTC~~ (i.e., ~~7am LT~~1200 UTC).” in L230.

- **L213-224:** consider moving this paragraph into the methods Section 2.1, structurally it seems out of place here.

Response: We moved the first few sentences to Section 2.1 that are more appropriate in Section 2.1 as the reviewer suggested. We kept the last few sentences that are more appropriate under the *Results* section, while the wording has been changed as follows.

Changes made: “~~Based on these two analyses~~ elbow method and the Silhouette coefficient introduced in Section 2.1 (see Fig. S1 for further details), we found that the data should be split into either two or four major clusters (see Fig. S2 for further details).” in L251.

- **L231:** “The second most “polluted” Cluster 1 has” is missing “cluster, ” after the word polluted.

Response: The corresponding change has been made as follows.

Changes made: “The second most “polluted” cluster, Cluster 1, has ...” in L262.

- **L272:** missing space in “labels.A”.

Response: This sentence has been removed from the manuscript during this revision.

- **Figure 5:** the faint lines in Figure 5 do not reproduce well, please make darker and/or thicker.

Response: The grey line in Fig. 5 has been thickened.

Changes made: The grey line in Fig. 5.

- **Figure 6, 7 and 8:** Consider swapping the order of some of the sub-panels to aid the reader to follow the comparisons made in the text. In particular Fig. 6(f-j) would be better placed with Fig. 7(a-e) and/or Fig.8 (a-d), and Figure 7(f-j) complements well with Figure 6(a-e).

Response: The authors had an extensive discussion on the order of these panel plots before the submission, especially in terms of the combination of the panels as the reviewer pointed out. As somewhat explained in the manuscript (L359 & L428), the manuscript first discusses the *bulk properties* of clouds such as cloud mass, CDNC, and rain mass. And subsequently *process rates* are presented, so that the paper can explain the detailed reason(s) for the aforementioned bulk properties. Because of this structure that is used both in Sections 3.2 and 3.3, we retain the current format without changing the order of the plots.

- **Supplement:** please relabel figures with “Observed 7/20” to “Total observed (bulk)” to match the main text.

Response: The corresponding change has been made in the supplementary figures.

Changes made: Figure labels for the panel (d) and (e) in Figs. S3-4 to be consistent with the main text.

Thank you for reviewing the manuscript. We have re-run some of the simulations, re-made almost all the plots, and modified the manuscript significantly, according to the suggestions and comments from the reviewers. Our fundamental findings remain unchanged. Please see below our response to each of the referee's comments starting with "**Response:**" and "**Changes Made:**". The line numbers ("L") referred here is based on the revised manuscript, not the revised manuscript with tracked changes.

Response to General Comments

This study uses aerosol observations from the TRACER intensive operational period in La Porte, Texas, to identify four dominant aerosol regimes through K-means clustering. Representative aerosol size distributions from these clusters are then used in large-eddy simulations of shallow cumulus clouds with CM1, comparing a bulk microphysics scheme with the super-droplet method. The bulk scheme produces precipitation too early and too strongly relative to both observations and the super-droplet simulations for the considered intercomparison case, which the authors attribute to underestimated cloud-droplet number concentrations which then produce excessive parameterized collision-coalescence rates.

The observational-modeling framework is potentially valuable, but the interpretation of the simulation results would be strengthened by a clearer and more complete discussion of the assumptions underlying the bulk microphysics scheme, together with additional analysis of the cloud-water mass and number budgets and/or one or more targeted numerical experiments to identify the processes responsible for the differences between the simulations.

Response: Thank you for your review comments. We would like to also emphasize that one of our key findings is the clustering result itself, in addition to the modeling results. Below you will find our responses to more specific comments in greater detail.

Response to Specific Comments

Response to Major Comments

1. lines 326-8: "The CDNC variation [between simulations] can be simply explained by the variation in aerosol number concentration; the more available aerosol particles, the higher the domain-mean CDNC." The key here is "available aerosol." For a specific supersaturation, only the subset of the aerosol particles with critical supersaturations less than the supersaturation will activate. That fraction appears to be very small for clusters 2 and 4, which have total aerosol concentrations over 2000 cm^{-3} yet have CDNCs of about 50 cm^{-3} towards the end of the simulations. That means the "available aerosol" for

activation is less than 2.5 %. Even the largest aerosol mode, with GMD of 170 nm has a concentration of 312 cm^{-3} , so only a fraction of this mode is activated. This statement needs to be clarified by stating exactly what is meant by “available aerosol.”

Response: We would like to first mention that the bulk Morrison scheme uses the saturation adjustments, and therefore, no supersaturation is realized in the bulk simulations. Secondly, the in-cloud averages in Fig. 7a-d include data from thin or stagnant/downfalling parts of the clouds too, as long as cloud mass exceeds 0.01 gm^{-3} (which is generally a low threshold). And most importantly, we re-ran the bulk simulations based on the reviewers’ suggestions, and the updated simulations have higher droplet number concentrations (there was some error in the bulk microphysics scheme that kept droplet concentrations somewhat lower in the previous simulations, along with some parametric constraints that are mentioned below). Based on these updated results and the patterns we see, we have modified the sentence as follows.

Changes made: Modified to “*The CDNC variation can be simply explained by the variation of large (i.e., > 100 nm) aerosol number concentration; the more **large** particles are available (**Fig. 5; unchanged throughout each simulation**), the higher the domain-mean CDNC...*” in Line 363.

2. lines 341-51: Your discussion based on Figures 7 and 8 might be able to provide more robust conclusions if you analyzed all terms of the CDNC budget as well as of that of cloud droplet mass budget. As it stands, you can only speculate. The results and discussion presented in Section 3.3 also strongly motivate analyzing the budgets of cloud number concentration and cloud water mass.

As noted by the authors, the low bulk CDNC may reflect not only activation physics and cloud-droplet removal by autoconversion and accretion, but also limited aerosol availability for continued activation aloft. The authors could provide a quantitative cloud-droplet number budget together. Without separating aerosol transport and availability, activation, and subsequent cloud-droplet number loss, the cause of the bulk–SDM differences in CDNC and rain production may be difficult establish.

As an alternative to performing such analyses, which could be time consuming, you may be able to learn a great deal about the reasons for the differences between bulk and SDM for this case by doing at least one additional simulation, as described in the next comment.

Response: Thanks for the suggestion for an additional simulation. We have run the suggested simulation and found that autoconversion was not the major reason. We therefore modified the manuscript as follows. We also mention here that there’s no change in aerosol availability (or their transport) in the bulk simulations since there’s no aerosol model coupled with the bulk microphysics scheme. The concentrations and availability are fixed, regardless of the location (in-cloud/out of cloud).

Changes made: Changed to “Figure 8e shows slightly lower in-cloud droplet nucleation rates in the SDM run in comparison to the bulk run (Fig. 8d), which can be explained by the limited supply of aerosols in the SDM run as opposed to the unlimited aerosols in the bulk scheme. ~~It signifies that the activated droplets continuously get “lost” in the bulk runs after getting activated by evaporation and/or conversion into raindrops, which is why the activation rate in the bulk simulation is slightly higher than that in the SDM run but CDNC remains lower (Fig. 7d).~~ Based on our analysis, the low CDNC in the bulk scheme, despite the slightly higher activation rate, seems to mainly stem from the constraint on the slope parameter of the droplet size distribution therein; due to the minimum and maximum values of the slope parameter set in the scheme, CDNC cannot increase (or decrease) freely in the bulk scheme with the background aerosol loading.” in L439.

3. This reviewer believes that a particularly useful sensitivity experiment would be to repeat the Total Observed bulk simulation with the Khairoutdinov–Kogan autoconversion process disabled, including both the cloud-water mass transfer and its associated cloud-droplet number sink. In the corresponding SDM simulation, the reported droplet-to-raindrop conversion is essentially zero, whereas the bulk scheme produces substantial conversion. Thus, suppressing bulk autoconversion would remove a process that is negligible in the SDM result and would help isolate the origin of the large CDNC difference. If the bulk CDNC then approaches the SDM value, this would demonstrate that autoconversion and the resulting collision–coalescence sinks largely maintain the anomalously low bulk CDNC. If it remains low, the discrepancy would instead point primarily to aerosol availability, activation, or other aspects of the bulk droplet-number treatment. An intermediate response would quantify the contributions of both mechanisms. This experiment would complement, rather than replace, the requested cloud-droplet number budget.

Response: We have conducted this experiment where the autoconversion (droplet + droplet to raindrop) and accretion (droplet + raindrop to rain) processes were turned off in the bulk simulation with the Marine+Rural aerosol configuration. In this additional simulation, there was a reduction of rain mass, but not in CDNC. We have conducted further analysis with additional simulations to find the following two points; (1) the bulk scheme does not allow CDNC to exceed the number of background aerosols. While this setting itself is not an issue, this part of the code had some issue (i.e., the maximum CDNC was set to a lower number), and therefore, we have re-run all the bulk simulations with correct maximum CDNC. This correction has increased CDNC in all the bulk simulations. (2) Since a gamma size distribution is imposed for droplets in the bulk scheme, a maximum and minimum slope parameters are set for these distributions so that droplets remain within a reasonable size range even when CDNC is calculated to be very high or low. In the version of the bulk scheme used, the minimum and maximum *mean* droplet sizes are 1 μm and 60 μm , respectively. In other words, when CDNC is very high or very low, this limiter forces the droplet

size distribution to have mean sizes of 1 μm or 60 μm , which in turn modifies the original droplet number concentration to be not too high or not too low. This is another reason why CDNC remains relatively low in the bulk scheme. The second point is mentioned in the revised manuscript as follows.

Changes made: Changed to “*Based on our analysis, the low CDNC in the bulk scheme, despite the slightly higher activation rate, seems to mainly stem from the constraint on the slope parameter of the droplet size distribution therein; due to the minimum and maximum values of the slope parameter set in the scheme, CDNC cannot increase (or decrease) freely in the bulk scheme with the background aerosol loading.*” in L441.

4. lines 166-68: The bulk scheme uses the parcel-based Abdul-Razzak–Ghan activation parameterization, whereas the SDM represents activation through the supersaturation histories of individual superdroplets. The manuscript should specify the vertical velocity supplied to the activation parameterization, any imposed subgrid or minimum-updraft treatment, and where (cloud base or above cloud base) activation is calculated.

Response: For the bulk scheme, the input vertical velocity comes from the dynamical calculations by CM1, and no subgrid-scale fluctuation of updrafts is considered. The minimum updraft of 0.001 ms^{-1} was required for an activation to occur. For SDM, vertical velocity also comes from the dynamical calculations within the CM1 model but the subgrid-scale fluctuation of the vertical velocity is considered, as mentioned in L194. Activation could occur anywhere as long as the size of a superdroplet exceeds the critical radius (i.e., no minimum updraft speed was imposed). These points were clarified in Section 2.2.

Changes made: Added “... aerosol activation is based on Abdul-Razzak and Ghan (2000) *where a minimum updraft speed of 0.01 ms^{-1} was needed for activation and no consideration of subgrid-scale vertical velocity was made, and*” to L194 and “Aerosol particles go through wet growth based on the environmental humidity and get activated into cloud droplets when the environmental supersaturation exceeds aerosols’ critical supersaturation (i.e., no minimum updraft speed for activation).” in L205.

5. lines 380-81 and lines 166-68: The manuscript should clarify how supersaturation and condensational growth are treated in the bulk microphysics scheme. If saturation adjustment is used, excess water vapor is instantaneously condensed to restore saturation, so supersaturation is not prognosed and condensation is not explicitly controlled by droplet number or surface area. This treatment may be acceptable for some bulk cloud simulations, but it is potentially problematic in an LES comparison with the SDM, where supersaturation fluctuations and finite-rate condensational growth are explicitly represented. The authors should therefore discuss whether differences between the two

simulations may partly reflect their fundamentally different treatments of supersaturation, rather than differences in aerosol activation or collision-coalescence alone. A prognostic-supersaturation bulk formulation, such as that of Morrison and Grabowski (2008), could provide a more physically consistent basis for comparison.

Morrison, H., and W. W. Grabowski, 2008: Modeling supersaturation and subgrid-scale mixing with two-moment bulk warm microphysics. *Journal of the Atmospheric Sciences*, 65, 792–812, <https://doi.org/10.1175/2007JAS2374>.

Response: The use of saturation adjustment in the bulk scheme was mentioned in the original manuscript (L361 and L425 in the revised version). In the revised manuscript, we mention it additionally in Section 2.2 as follows. However, to our knowledge, majority of bulk microphysics schemes, especially those used in operational forecasting models and climate models, still use saturation adjustment in consideration of computational costs. Therefore, the comparison of the bulk and SDM runs presented in this study still provides readers with meaningful understanding of the bulk-SDM differences.

Changes made: Added “*In this scheme, saturation adjustment is used, aerosol activation is ...*” to L193.

6. lines 386-88: “This difference is another counterintuitive discrepancy between the two simulations given that the bulk simulation gets an infinite supply of aerosols while the availability of aerosols is limited in the SDM run (i.e., Sect. 2.2).” As noted in the comment for lines 326-8, the key here is “availability of aerosols.” For example, one could have a very large concentration of very small and/or hydrophobic aerosols, yet no activation in otherwise identical conditions. Or even have exactly the same size distributions of aerosols, yet encounter different supersaturations, so that many aerosols are activated (become “available” to become cloud droplets) in one case compared to the other. This statement should be revised to reflect other factors that control the “availability” of aerosols to activate.

Response: The original sentence including “*This difference is...*” pointed out here has been removed from the manuscript. The availability of aerosols (number, size distribution, and hygroscopicity) does not change in the bulk simulations due to the lack of an aerosol model, whereas the processes of superdroplets/aerosols are explicitly calculated in SDM. Because of this difference, it is generally true that the bulk scheme has the “infinite supply” of aerosols, whereas SDM has a limited number of aerosols in the domain (as in reality).

Changes made: Removed “~~*This difference is another counterintuitive discrepancy between the two simulations given that the bulk simulation gets an infinite supply of aerosols while the availability of aerosols is limited in the SDM run (i.e., Sect. 2.2).*~~”.

7. lines 399-407: The explanations for the differences between the bulk and SDM runs in this paragraph are not convincing. There are many differences between the simulations with the bulk scheme and the SDM scheme. These include not just that aerosols are processed and advected in the SDM run but not in the bulk model which leads to differences in the aerosol profiles in the later parts of the simulations, but also that supersaturation is prognosed in the SDM scheme and implicitly diagnosed by the ARG scheme in the bulk model, the collision-coalescence process is explicit in SDM but parameterized in the bulk scheme, the hydrometeor mixing ratios are advected using an Eulerian method in the bulk scheme but all particles are transported using a Lagrangian scheme in the SDM model. I am not sure what to recommend, other than that this paragraph needs to be completely revised based on more extensive analyses and/or numerical experiments, or simply omitted.

Response: The authors agree that there are many fundamental differences between the bulk and SDM simulations. Because of that, the presented results seem somewhat speculative without fully identifying the origin of the differences between the two simulations. However, we argue that the main point of this paper is (1) identifying the aerosol clusters in the region, (2) testing the sensitivity of the bulk scheme to different clusters of aerosols, and (3) examining the differences between the bulk scheme and the SDM, as mentioned in *Introduction*. As this study does not aim to correct/revise the bulk scheme based on the SDM results, understanding how shallow cumulus clouds are simulated differently and where the differences likely stem from does fulfill our aim (3). In the revised manuscript, we mention the fundamental differences between the two schemes as follows.

Changes made: Added “*It should be noted that the two microphysics schemes compared here have the fundamental differences in aerosol treatment, Eulerian/Lagrangian framework, and microphysical parameterizations. Therefore, it is challenging to identify exactly where the presented differences come from, but we show below the most likely cause of the discrepancy between the two, which is informative for studies and operational work using the bulk scheme for shallow cumulus cloud simulations.*” L414.

Response to Minor Comments

1. line 15: “particles sized around 180 nm” – state whether diameter or radius.

Response: It’s diameter and we have added the information as follows.

Changes made: Added “*sized around 180 nm in diameter*” in both L16 and L288.

2. line 18: “these aerosol clusters” – how many clusters? Two or four?

Response: It's four, and we have modified the sentence as follows.

Changes made: Changed to "... *the **fourse** aerosol clusters*" in L19.

3. line 60-61: As your study and many others show, numerous factors in addition to aerosol abundance play roles in convection and precipitation. Therefore, references to some key review articles might be helpful here, especially in regards to the roles of aerosols as well as meteorology in determining cloud properties.

Response: Some review articles on aerosol-cloud interaction are mentioned in *Introduction* (e.g., L28-29). Because it is difficult to isolate the dynamical effects from aerosol effects, their reviews on the effects of dynamical conditions are not as explicit. Nevertheless, we have added a short subclause and references to the studies mentioning the dynamical effects on convection as follows.

Changes made: Changed to "*On the weather scale, ~~many~~ **some** observational studies have identified strong controls of aerosols on cloud characteristics for a range of cloud types (e.g., Fan et al., 2016; Li et al., 2025), in addition to the control of dynamical factors (e.g., Grabowski, 2018; Wu et al., 2025).*" in L28.

4. line 64: Please state which cloud properties Huang et al. studied.

Response: They focused on liquid water content, droplet number concentrations ($< 50 \mu\text{m}$ and $> 50 \mu\text{m}$), and effective diameter. We have modified the sentence as follows.

Changes made: Changed to "... *to be the most important factors to **convective** cloud properties, such as liquid water content and droplet number concentrations, using a machine-learning approach*" in L65.

5. line 144: Change "weighed" to "weighted."

Response: This has been fixed.

Changes made: Changed to "*weighted*" in L168.

6. line 154: Change "Based on the results from the aerosol clustering above" to "Based on the results from the aerosol clustering described in section 3.1."

Response: The suggested change has been made as follows.

Changes made: Changed to "*Based on the results from the aerosol clustering described **in Sect. 3.1** above, we run ...*" in L181.

7. line 155: Change “double periodic” to “doubly periodic”

Response: The word has been modified accordingly as follows.

Changes made: Changed to “*doubly periodic*” in L183.

8. line 156: Change “50-m horizontal resolution” to “50-m horizontal grid size” because resolution and grid size are not the same. Same for “vertical resolution is 30 m.”

Response: The corresponding changes have been made as follows.

Changes made: Changed to “... with a 50-m horizontal ~~resolution~~ *grid size* (i.e., 256 points in both x and y directions). The vertical ~~resolution~~ *grid spacing* is 30 m ...” in L183.

9. line 174-75: You wrote “The collision kernel of Hall (1980) is chosen for this study.” More accurately, Hall (1980) provides/tabulates collision efficiencies for the hydrodynamic-gravitational collision kernel. For completeness, you could also mention this applies to laminar flows.

Response: The suggested change has been made as follows.

Changes made: Changed to “*For collision-coalescence, the collision efficiencies for hydrodynamic gravitational collision kernel (i.e., laminar flows) of-in Hall (1980) is-chosen forwere used-this study.*” in L206.

10. line 177: Please clarify “Subgrid-scale velocity fluctuation is accounted for in SDM but subgrid-scale supersaturation fluctuation is not.” Accounted for in terms of which processes? And for context, please mention how the same processes are accounted for in the Abdul-Razzak and Ghan (2000) aerosol activation scheme.

Response: In SDM, the subgrid-scale velocity fluctuation is accounted for in the calculation of superdroplets’ motion and supersaturation fluctuation for condensation/evaporation. These points have been clarified in the revised manuscript. Also, the revised manuscript now mentions the equivalent setting in the bulk scheme; no vertical velocity fluctuation is considered.

Changes made: Added “... aerosol activation is based on Abdul-Razzak and Ghan (2000) *where a minimum updraft speed of 0.01 ms^{-1} was needed for activation and no consideration of subgrid-scale vertical velocity was made, and*” to L194. Changed to “*In our SDM simulation, subgrid-scale velocity fluctuation is accounted for in the calculation of SDs’ motion in SDM but*”

subgrid-scale supersaturation fluctuation is not in the calculation of condensation /evaporation.”
in L210.

11. line 205-6: “The large-scale advections were modified slightly to better match observations.” Observations of what? Please state.

Response: We have modified this sentence as below to clarify the meaning.

Changes made: Changed to “*The large-scale advections were modified slightly to better match observations so that the simulated cloud evolution and cloud-base height better match the observation in Fig. 1d.*” in L243.

12. line 218: Should “deflection point” be “elbow point”?

Response: Either should work, and we have made the following modification.

Changes made: Changed to “... a deflection point or an “elbow” where ...” in L143.

13. Figure 7: I assume that observed radar reflectivity is available. This could be used for comparison to the simulations.

Response: Even though the bulk simulations (Fig. 7f-i) show rain mass, the surface rain gauge data show no sign of precipitation on the ground (Fig. 9). For these shallow cumulus clouds where cloud-top heights are low and no precipitation (or rain mass) is seen, precipitation radar data cannot provide cloud properties that are comparable to the simulated results.

14. Figure 9, lines 396-8: Point measurements of convective rainfall are not useful for comparison to domain means of simulations due to large spatial and temporal variability of convective rainfall. Observed radar reflectivity if from a scanning radar, would be much more useful. Conclusions drawn from the observed point measurements are almost certainly not reliable or robust and should be omitted from the text.

Response: We argue that the point measurements of precipitation data are still useful, given that the same point measurement of cloud cover, shown in Fig. 1d, clearly shows the presence of the shallow cumulus clouds above the observation site continuously for hours. Coarse horizontal resolution of precipitation radar might make it challenging to capture the small amount of rain from these small clouds whose cloud-top height is low.

15. line 379: “In-cloud-mean cloud mass is 29% higher on average in the SDM run throughout the simulation.” Cloud mass needs to be defined: is it liquid water path? or liquid water content at some level?

Response: It is liquid water content as shown in Fig. 6i-j. The percentage difference was calculated by computing the mean difference between Fig. 6i and 6j. As described in the figure caption, “in cloud” means grids with cloud mass $\geq 0.01 \text{ gm}^{-3}$. The sentence in L423 has been modified to clarify this.

Changes made: Changed to “*In-cloud-mean cloud ~~mass~~ liquid water content, shown in Fig. 6i-j, is ~~29~~³⁰ % ~~larger~~ higher on average in the SDM run (Fig. 6j) ~~throughout the simulation, ...~~” in L423.*

16. line 381-82: “This tendency of lower cloud mass in the bulk run, alternatively, can be a result of numerical diffusion.” You should state what field(s) you are referring to. If you are referring to the bulk hydrometeor fields, then yes, numerical diffusion could be a factor, but there are many other possible factors as well. It seems you might be referring to numerical diffusion in a bin scheme (lines 434-7). However the bulk scheme is not a bin scheme, so this specific type of numerical diffusion would not apply to the bulk model.

Response: The wording for this sentence has been modified as below, which should make it clearer that the field referred to here is *cloud mass*. We have also added more explanation on numerical diffusion in *Introduction*; numerical diffusion can arise both from the Eulerian grid framework itself (e.g., Hinneburg and Knoth, 2005) as there are fixed points of calculations in contrast to the Lagrangian framework with flexible points of calculations, and also from the use of particle size “bins” as in bin schemes, which is discussed more in previous studies (e.g., Grabowski et al., 2019). Here, the former was originally referred to, but it is difficult to assess the impact of this numerical diffusion in the bulk scheme. Therefore, the part of “*can be a result of numerical diffusion*” has been removed from this sentence.

Changes made: Changed to “... ~~b~~*Both bulk and bin schemes suffer from numerical diffusion that arises from fixed-point calculations of microphysical processes in the Eulerian framework (e.g., Hinneburg and Knoth, 2005) and/or from the use of particle size bins in bin schemes (Grabowski et al., 2019), which*~~that~~ potentially leads to erroneous calculations of microphysical processes (Grabowski et al., 2019), ~~which~~*but can be mitigated by using a Lagrangian scheme.*” in L88 and to “... *this larger cloud mass in the SDM run can be partly explained by the smaller conversion rate of clouds to rain as shown below*~~that convert all supersaturated water vapor into liquid. This tendency of lower cloud mass in the bulk run, alternatively, can be a result of numerical diffusion.~~” in L426.