Replies to Referee #2 on gmd-2022-162
The reviewer's comments are shown in bold font.

This manuscript describes the implementation of a stochastic, multi-plume Eddy Diffusivity Mass Flux (EDMF) boundary layer parameterization in the SCREAM GCM, using the Simplified Higher Order Closure (SHOC) to calculate eddy diffusivity and cloud properties. Single column model experiments are used to evaluate the scheme against large eddy simulations for marine shallow cumulus and continental convection cases. Improvements are shown relative to experiments using SHOC alone.

The topic is certainly of scientific interest, as boundary layer clouds remain one of the largest sources of uncertainty in future climate projections, and boundary layer parameterization is a topic of active research. The paper is well written and logically organized. However, a few details are missing that are needed for readers to interpret and reproduce results, as noted in comments below. I believe these can be addressed with minor revisions.

We would like to thank the reviewer for the constructive comments that helped to substantially improve the manuscript. Essentially all of the reviewer's comments were addressed by modifications and additions in the manuscript. As a result of the reviewer's comments, the manuscript was revised. The main changes are:

1. Description of SHOC's turbulence length scale in section 2.4.
2. Description of the coupling of SHOC and MF for the cloud macrophysical properties.
3. Figure 7 was updated following reviewer's comment #1.

Detailed replies to the reviewer's comments are listed below.

1) There are some discrepancies in cloud top height in the continental convection case between this study, Bogenschutz and Krueger (2013) describing SHOC, and Brown et. al. (2002) documenting the continental case. The two previous papers show LES cloud top heights around 2800 m, while in Fig. 7 here the deepest cloud tops are close to 2400 m. I am wondering if there are differences in case specification that might explain the LES difference, or if it may be due to the use of different LES codes and grid spacings? This has implications for the conclusion that SHOC-MF matches the LES, while SHOC by itself produces too-deep clouds. The original SHOC (Bogenschutz and Krueger, 2013) appeared to match LES in this case fairly well, so the different behavior here warrants some discussion.

Thank you for bringing this to our attention. In our Figure 7a, the LES cloud top appears to be lower than that in Brown et al. 2002 and Bogenschutz and Krueger (2013; BK13 hereafter) because of the threshold that we applied when producing the figure. In other words, in Figure 7a, we masked CF < 0.01 to eliminate the very small values that may be considered "noise" yielding to what seems to be a considerably lower cloud top. To illustrate this, please see the figure below for the LES cloud condensate (top panel) and cloud fraction (middle and bottom panels):

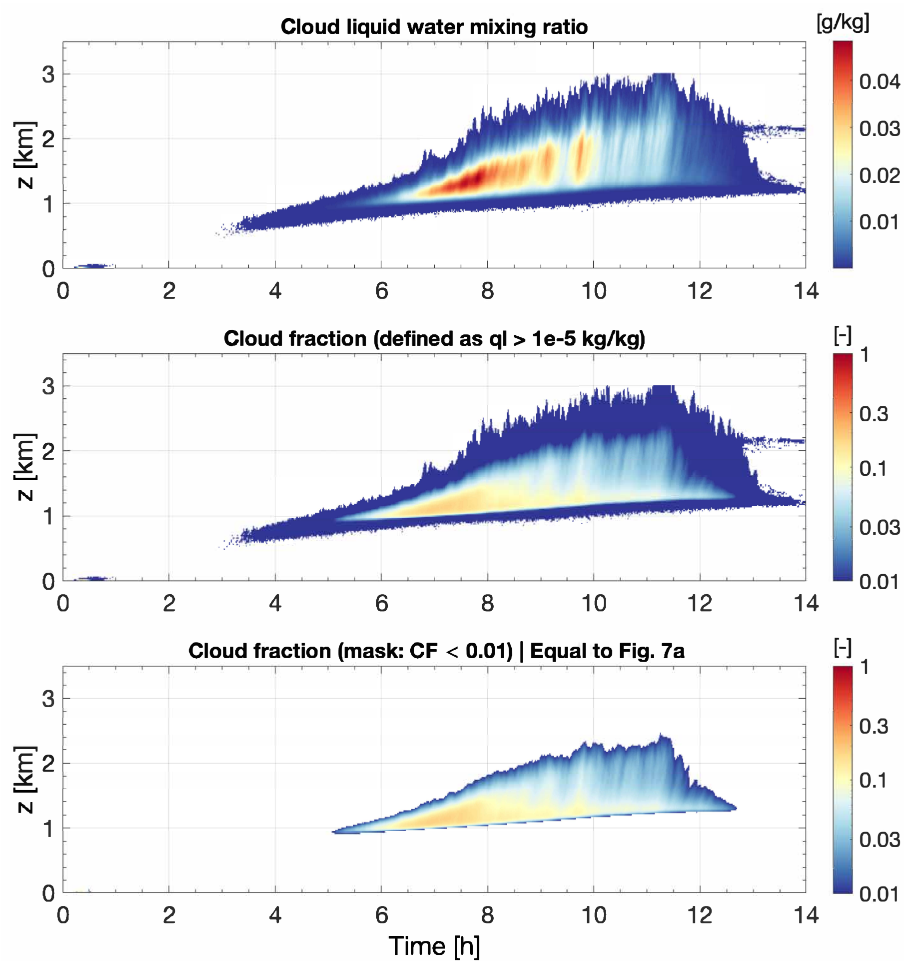

Figure 1 ARM shallow cumulus case: Time-height plot of LES (a) cloud liquid water mixing ratio, (b) cloud fraction which is defined as $q_l > 1 \times 10^{-5}$ *kg/kg* without any threshold/mask applied, and (c) cloud fraction as in (b) but only the values larger than 0.01 are displayed.

The cloud fraction defined as grid points where $q_l > 1 \times 10^{-5}$ kg/kg is shown in the middle panel, and the cloud fraction field also defined as $q_l > 1 \times 10^{-5}$ kg/kg but masked for CF < 0.01 is shown in the bottom panel—this is the one displayed in Figure 7a.

While our LES cloud-top height is not in disagreement with Figure 5d of Brown et al 2002 where it ranges from 2.4 and 3 km, it is in slight disagreement with Figure 9 of BK13. As the reviewer points out, the LES models and the vertical grid spacing used in both simulations are indeed different—the vertical grid spacing used in BK13 was 40 m, whereas we used 20 m. While we could only be sure that this is the reason for the discrepancy between these LES results by running both LES models with 20 m and 40 m (which is unfeasible), it is fair to conjecture that our cloud-top height is lower due to the higher vertical grid spacing. This is a well-known behavior of LES in which the utilization of higher grid resolutions generally yields slightly shallower boundary layers (e.g., Matheou and Chung, 2014; *Large-Eddy Simulation of Stratified Turbulence. Part II: Application of the Stretched-Vortex Model to the Atmospheric Boundary Layer*).

In conclusion, we replaced figure 7 with a similar figure but using a lower threshold of 0.001 instead of 0.01, and removed "… the cloud top is too deep" in L307.

Lastly, the caption of Figure 7 now reads (new text in blue):
"Figure 7: Time-height plot of (a) LES, (b) SHOC, (c) SHOC+MF cloud fraction for the ARM shallow cumulus case. The LES cloud fraction corresponds to the cloud sampling definition (i.e., $q_l > 1 \times 10^{-5}$ kg kg^{-1}). Cloud fraction values < 0.001 are masked and not plotted here."

2) Additional details are needed regarding the SHOC length scale and how it has been changed since Bogenschutz and Krueger (2013). The length scale formulation can have a major impact on PBL behavior, and should be described in the paper.

We added SHOC's length scale mathematical description in section 2.4.

3) It is not entirely clear how convective clouds are treated in this study. Is there any special treatment of cumulus cloud detrained from the mass flux? Or do the updrafts impact clouds only indirectly, through the mean state? This could be noted in Section 2.

This is indeed an important point. The first thing to note, however, is that for these particular shallow convection case-studies that we are analyzing, the longwave and shortwave radiation parameterizations are off, both in the LES as well as in the SCREAM SCM model. There is a simplified radiation forcing of the LES experiments that is replicated for the SCREAM SCM so as to have the SCM and the LES models forced in the same manner. The key physical reason for why the cloud-radiation interaction is off is related to the fact that with the low values of cloud cover typical of these shallow cumulus case-studies, the expectation is that the cloud-radiation interactions are of secondary importance. Large scale (i.e., in the environment surrounding the updrafts) cloud microphysics is also off, again because the expectation is that large scale microphysics plays a secondary role in these particular case-studies.

In this context, and although important, this topic is of less relevance for these particular shallow cumulus SCM simulations than it would be for other regimes. In any case, and in order to clarify the specific questions from the reviewer: For this version, although there is no explicit detrainment term that is a source of cloud water from the updrafts, the cloud fraction and cloud liquid water that are diagnosed (and shown in the figures) are simple linear combinations of the SHOC (PDF-based) cloud fraction and water with the contributions from the different updrafts.

This is summarized in the paper as follows in section 2.3:

"Condensation in each updraft takes place if the updraft water vapor reaches saturation. The MF contribution to the total cloud fraction corresponds to the sum of the area fraction of the updrafts that condense, and the MF contribution to the total cloud water is defined as the area-average of the cloud water of all moist updrafts."

And in section 2.5:
"The shortwave and longwave radiation, and the large scale cloud microphysics parameterizations are switched off for these experiments—basically because these processes are believed to be of secondary importance for these shallow convection case-studies and as such

are also off for the LES experiments. However, cloud fraction and water are calculated at every model level for diagnostic purposes. In SCREAM, the cloud macrophysical properties cloud fraction and liquid water mixing ratio are estimated using the SHOC pdf. Here, the combination of the grid-mean cloud properties is done as a simple weighted sum of the SHOC and MF contributions:

$$CF = CF_{SHOC} + \sum_{n=1}^{N} a_n(q_{l_n} > 0), \quad (12)$$

$$\overline{q_l} = a_e q_{l_{SHOC}} + \sum_{n=1}^{N} a_n q_{l_n}, \quad (13)$$

where CF_{SHOC} and $q_{l_{SHOC}}$ are diagnosed from SHOC's assumed double Gaussian distribution, and a_n and q_{l_n} are the fractional area and condensate loading of the *n*th updraft. Note that in practice, although the algorithm also imposes $CF \leq 1$, this value is not reached in these simulations because of the low values typical of these shallow convection case-studies. These low cloud values, the overall cloud vertical structure of these shallow convection regimes, and the fact that the radiation parameterization is off, are the key reasons for not using more complex cloud overlap algorithms in these estimates."

4) The experiments here used N=40 updrafts, while previous papers (e.g., Suselj et al 2013) have typically used a smaller number (N=10). Was N=40 chosen to reduce the effects of stochasticity, and do the results show any sensitivity to the value of N?

Yes, we choose *N* = 40 instead of e.g., 10 to reduce the intermittency of the results associated with the stochastic entrainment. Suselj et al. 2019 (*On the Factors Controlling the Development of Shallow Convection in Eddy-Diffusivity/Mass-Flux Models*) showed that using a smaller number of updrafts generally produces noisier results but results converge when a larger number of updrafts is used (please see their section 3a for further details). We tested the impact of *N* in our SCM experiments and saw reduced sensitivity to the results for $N > 10$ updrafts; please see the figure below with the BOMEX SCM results for *N* = {10, 40, 50, 100} updrafts.

We updated the text in the paper to reflect this discussion. In section 2.3, around L150, it now reads:

"Here, we use *N* = 40 updrafts. The number of updrafts was chosen based on a sensitivity analysis of SHOC+MF to its value in which SHOC+MF showed weak sensitivity to *N* > 30 updrafts (not shown). Note that a small number of updrafts can produce noisier results due to the lateral entrainment's stochasticity (Suselj et al. 2019a)."

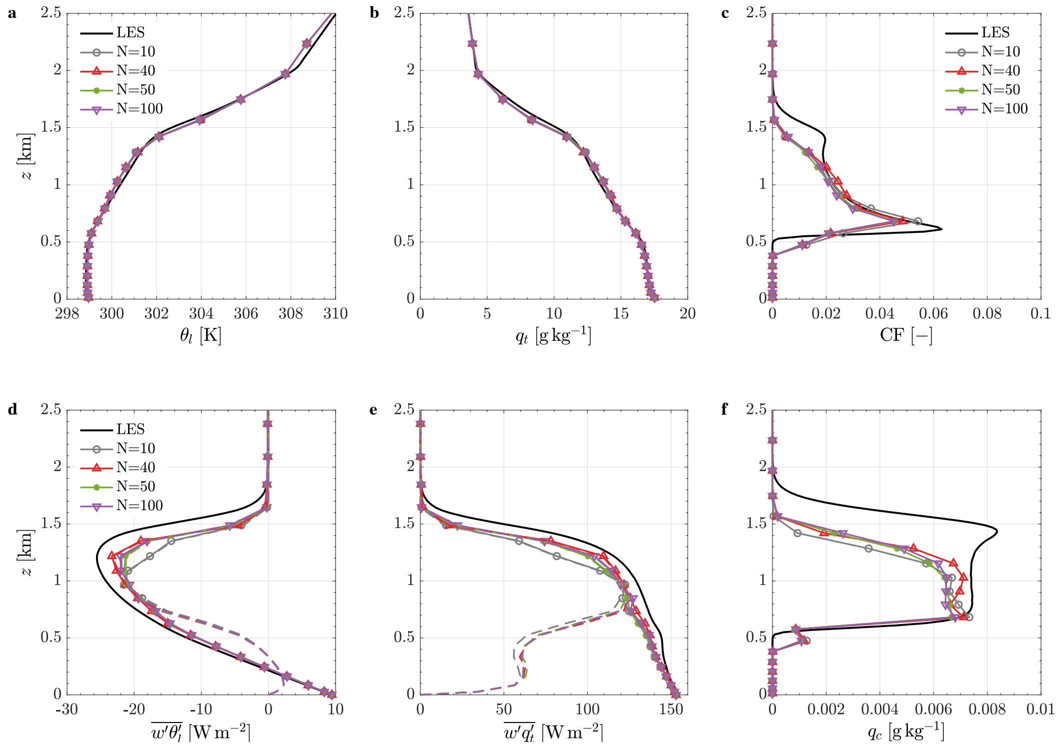

Figure 2 BOMEX: Sensitivity to number of updrafts *N*: vertical profiles of (a) liquid water potential temperature, (b) total water mixing ratio, (c) cloud fraction, (d) turbulent heat flux, (e) turbulent moisture flux, and (f) cloud water mixing ratio, for LES (black solid line), SHOC+MF (solid lines) and MF (dashed lines) for the BOMEX case.

5) Are the simple updraft microphysics mentioned on lines 61-64 included in this implementation? If so, do they have a non-negligible impact? This is relevant to the comparison with LES, for which precipitation was disabled.

No. At this point, we don't have the updrafts coupled to a microphysical scheme in our current SHOC+MF parameterization. Our current implementation reflects Suselj et al. 2019 (*On the Factors Controlling the Development of Shallow Convection in Eddy-Diffusivity/Mass-Flux Models*) as mentioned in L127 but without the downdrafts following the justification in L133-139.

6) On Line 247, which constant in SHOC was increased to reduce the mixing length, and by how much?

We increased the tunable constant l_c from 0.5 to 1 in equation 10 (section 2.4) in the SHOC+MF experiments. We updated this sentence to:

"Note that in SHOC+MF, we reduced SHOC's turbulence mixing length scale relative to SHOC alone, to prevent excessive mixing. This was done by increasing the tunable length scale factor l_c from 0.5 to 1 in equation 10. Thus, $l_c = 0.5$ in the SHOC experiments, and $l_c = 1$ in the SHOC+MF experiments."

L113: Typo: Eqn 1 is missing a "partial" symbol.

Thank you so much for catching this. It has been fixed.