Replies to Referee #3 on gmd-2022-162
The reviewer's comments are shown in bold font.

General comments

This manuscript presents a nice discussion of the EDMF and PDF-based HOC PBL and shallow convection literature, describes a new combination of physics schemes to improve on SHOC, and presents relatively preliminary results of using this combination in a SCM to simulate two shallow cumulus cases. It contains excellent writing and grammar and the authors are able to convey their main points in a concise manner, appropriate for the chosen journal. My recommendation would be to accept the paper after minor revisions despite what I perceive as a few major weaknesses of the manuscript, described below. The reason for my recommendation is based on the demonstrated efficacy of the described approach and a recognition that it has potential to make a substantial improvement to future GCM/NWP models if further developed.

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.

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

Specific Comments

1. **I'm concerned that this approach is conceptually "double-counting" effects due to the largest, coherent turbulent eddies. My understanding of the SHOC scheme is that one of the Gaussian components of the underlying PDF is already supposed to account for the type of subgrid-scale PBL-spanning updrafts that the MF scheme is also trying to represent. If I remember correctly, although the complexity of SHOC is boiled down to a K-theory implementation (as evidenced by the first term on the RHS of Eq. 3, 4, 9, 10), several of its underlying assumptions, notably its length scale and the "updraft" portion of the trivariate binormal PDF, are formulated to try to represent the same physical phenomenon that the MF scheme is. I.e., in regions of the column where PBL-scale convective eddies are present, K is substantially increased in SHOC in order to try to represent the effects of those eddies. The need for an additional turbulent transport "boost" from a MF component potentially speaks to the inherent limits of SHOC's approach of maintaining first-order closure (neglecting the TKE component that gives it a 1.5-order closure) at its heart. The manuscript comes close to touching on this point in several places, but never explicitly discusses it. For example, on lines 83-84 where it mentions SHOC represents "local mixing", on lines 197-199 mentioning the down-gradient term from SHOC, and lines 246-248 that describes the modifications to SHOC to achieve satisfactory results. In my reading of this paper, this thought was a through-line that I feel needs to be addressed/discussed. The fact that lines 246-248 are put in the paper lead me to believe that the authors are aware of this issue, but the lack of detail (how was SHOC's length scale reduced, which constant was increased and why, sensitivities to these changes) elicits an impression of "sweeping this under the rug", so to speak. A discussion/explanation doesn't even have to be scientifically/phenomenologically-motivated necessarily. It may suffice just to say that this is a pragmatic approach to patching a "weakness" of the SHOC formulation, etc.**

In the introduction, L74-81, and later in L343-347, we briefly address the limitations of double-Gaussian PDF closures in representing the high skewness and kurtosis of the distributions of shallow convection and the need for a larger number of PDFs to properly represent the higher-order moments and cloud statistics. Firl and Randall (2015) illustrate this issue very well in their Figure 6 where four PDFs are needed to properly represent the cloud properties and second-order moments of BOMEX. Regarding this and the "double-counting" issue, Witte et al. 2022 also show in their Figure 3 (see below) the inability of CLUBB (black contours) in representing the extremes of the LES joint distribution (gray contours) and how the MF updrafts successfully capture these extremes in their CLUBB+MF parameterization. It is then reasonable to expect that, if the higher order closures, based on pdfs, such as CLUBB have difficulties in simulating the skewed part of the pdf where most of vertical transport takes place during moist convection, much more simplified pdf approaches such as SHOC will also suffer from (at least) similar problems.

Regarding SHOC, in section 2.4, we added the mathematical description of SHOC's turbulence length scale L (now equation 10) where l_c is a length scale factor that we increase from 0.5 to 1 and that reduces L. Please note that, apart from SHOC+MF, this reduction was necessary for the ARM case since SHOC mixes too much in its default configuration (Figures 3–6); however, the reduction of L required to improve SHOC's performance in ARM would degrade even further BOMEX where SHOC is not mixing enough (Figure 1). Overall, we believe the lack of a consistent representation of shallow convection by SHOC is partly due to the general inability of HOC-PDF schemes in representing the shallow convection variability (e.g., Firl and Randall 2015, Fitch 2019, and Witte et al. 2022), although the additional simplifications made in SHOC potentially make the matters worse.

To clarify this, we edited L77-81 and now reads (new text in blue):
"An alternative solution was recently proposed in Witte et al. (2022) where CLUBB is combined with multiple stochastic MF plumes leading to a modified CLUBB+MF parameterization where the plumes represent the extreme tail of the joint distribution which is not represented by CLUBB (see their figure 3). Furthermore, their results showed a large improvement of the higher-order moments for two benchmark shallow cumulus convection cases. Thus, the multiple MF plumes offer a physics-based and cost-effective solution by representing the extreme values of the joint distribution not well captured by the assumed PDF."

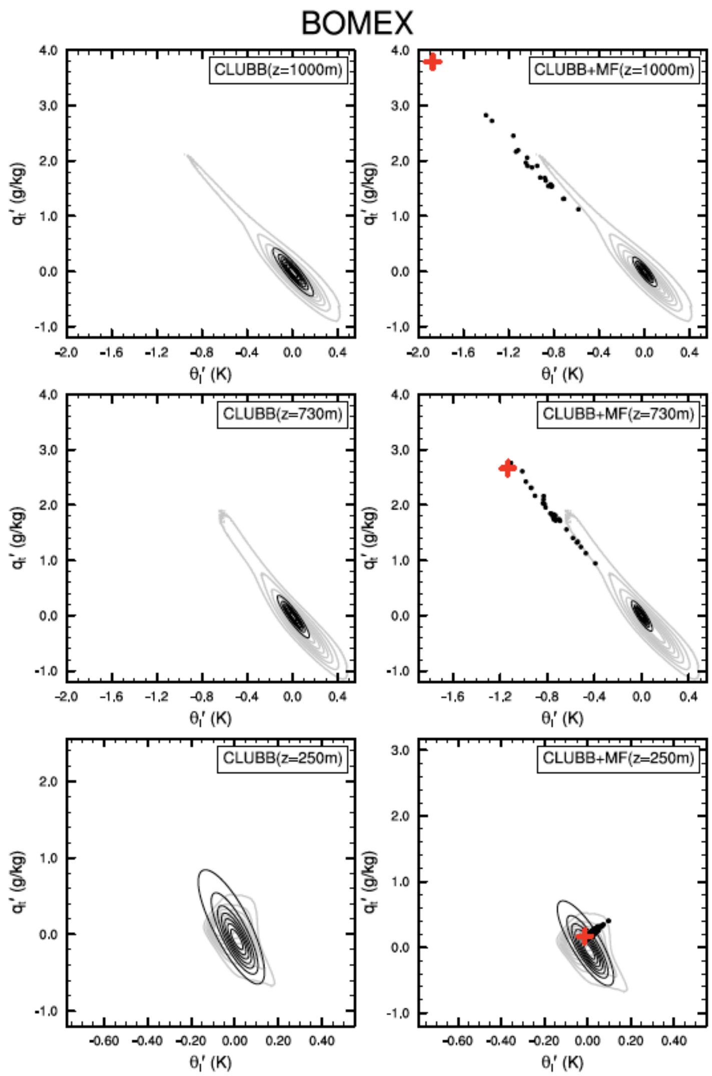

FIG. 3. Joint probabilities of perturbation q_t' and θ_l' relative to level mean $\overline{q_t}$ and $\overline{\theta_l}$ at $t = 6$ h for three levels: (bottom) $z = 250$ m, (middle) $z = 730$ m, and (top) $z = 1000$ m from the BOMEX simulations for LES (gray contours), CLUBB (black contours), and MF plumes (black dots). Red crosses denote mean surface properties relative to the mean environmental properties aloft. Probabilities are expressed as 1 minus a unitless cumulative probability, $C(\theta_l, q_t)$, such that the mean of the joint distributions has a value of 1, decreasing radially outward. The contours intervals are logarithmic and identical for both CLUBB and LES. The outermost contour represents a cumulative probability of 3.5% and the innermost contour represents a cumulative probability of 70%. The CLUBB contours refer to the (left) CLUBB-only and (right) CLUBB+MF simulations, with the black dots denoting the q_t' and θ_l' values of individual members of the MF plume ensemble.

2. **The applicability of the paper is limited by the chosen cases. The readers are definitely left "wanting more" in the sense that no indication is given for how the new scheme performs for stratocumulus, deep convection, frontal cloudiness, clear/dry convection, stable PBLs, mixed-phase cloudiness cases, etc. The manuscript as presented is fairly typical in its scope in this regard, and this criticism applies generally to other similar papers, but it is worth pointing out. I'm not saying that the authors need to expand the scope for the particular paper, just that it is much less exciting/convincing without more meteorological regimes (higher "N") to go on.**

This is a really important point. As mentioned by the reviewer, this is a fairly typical approach of several similar investigations and papers. To be clear, we are currently extending SHOC+MF to stratocumulus cases, the stratocumulus to cumulus transition, and the diurnal cycle of convection over land. Our initial focus on shallow convection is essentially because i) these are benchmark case-studies that any successful unified mixing parameterization needs to get right and ii) SHOC does not appear to be able to represent shallow convection cases such as these ones in a realistic manner.

3. **There are components of the scheme that seem arbitrary to the reader. For example, line 144 describing which part of the PDF is used for the MF model with [1.5σw – 3σw], or 6.65%. Can this be justified? E.g., why not use the entire part of the PDF with w > 0?**

We use the updraft model presented in Suselj et al. 2019a; this is mentioned in L127. For clarity, we now reiterate this in L142. We also edited L144 and now reads: "here defined as $1.5\sigma_w < w_n < 3\sigma_w$, where σ_w is the vertical velocity standard deviation (note that the interval $[1.5\sigma_w, 3\sigma_w]$ corresponds to a total updraft surface fraction area equal to 6.65% in agreement with the sensitivity analysis to the surface updraft area presented in Suselj et al. (2019a))." Please note that in the EDMF approach, the MF aims to represent the nonlocal turbulent transport by the strongest updrafts. Accordingly, the updrafts correspond to the tail of the joint distribution as shown for instance in figure 3 of Suselj et al. 2013.

Also, we added information about the number of updrafts ($N = 40$ updrafts) around L149 and 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)."

4. **I'd like a more physical explanation for the formulation of Lϵ, which I have interpreted as the "mean free path" between entrainment events. It seems like this should be related to the local strength of turbulence, e.g. TKE. I realize that h_{CBL} might be a proxy for TKE, but how it is used seems to be opposite to my intuition. I would think that high TKE (and high h_{CBL}) would lead to more frequent entrainment/mixing events, creating a lower Lϵ. This leads me to think that this formulation was "tuned" to get the desired scheme behavior/results rather than using physical reasoning. A more thorough explanation (perhaps it is in the Suselj 2019 paper) for Eq. 8 would be beneficial to the reader.**

Th reviewer is partly correct in that there is, in principle and from a physics perspective, a relation between entrainment and turbulence. However, that precise relation is not really well known (and lateral entrainment parameterizations often reflect this state of knowledge) and in practice the contradiction that the reviewer highlights does not really a play a particularly important role in the context of these specific simulations. While we utilize a stochastic entrainment approach, the foundations in which the parameterization is built are relatively simple and common practice. Equation 8 captures the fact that deeper convective clouds tend to be wider which protects them from the environment leading to small entrainment rates (e.g., Boing et al. 2012, Takahashi et al. 2021). In other words, if the updrafts entrain too often, here through a lower L_ε, they will get diluted and cease too quickly. We edited L170 to clarify this and now reads:
"In agreement with previous studies (e.g., Böing et al. 2012; Takahashi et al. 2021), the entrainment length scale L_ε is larger for deeper clouds (i.e., higher h_{CBL}) as these tend to be wider and thus better

protected from the environment leading to smaller entrainment rates. Note that diagnosing L_ε as the square-root of h_{CBL} allows for continuous adjustment of ε_n as a function of the CBL state, i.e., the entrainment rate is reduced for deeper CBLs allowing the updrafts to reach higher vertical levels and vice-versa for shallower CBLs, which is particularly important to represent the strong diurnal cycle over land while remaining insensitive to small oscillations of h_{CBL}."

5. **Lines 53-67 discuss EDMF schemes more broadly and the last couple sentences imply that the SHOC+MF approach is on par computationally to existing EDMF implementations. Is this claim accurate? It would be nice to have some numbers to back this up if so.**

This number is largely model-dependent and unfortunately, we don't have an EDMF parameterization, apart from SHOC+MF, available in SCREAM which precludes us from providing meaningful numbers regarding the computational efficiency of EDMF and SHOC+MF. Nonetheless, EDMF and SHOC+MF are significantly more efficient than for instance CLUBB where 9 prognostic equations of higher-order moments must be solved.

6. **Lines 98-99: Why is C++ code better than Fortran code in this case? Seems strange to make this claim in this paper.**

In this paragraph (L94-101), we provide a general description of SCREAM. SCREAM is a global convection-permitting model aiming to run global 3D simulations at a target horizontal resolution of 3 km. The code is being rewritten in C++ to enable faster simulations on GPUs and future architectures than what would be possible using Fortran. The SCREAM model is a fairly recent DOE effort and we expect most readers to be unfamiliar with it, thus a generic introduction seems appropriate.

7. **Line 115: What are "fluctuations" with respect to? Time, space, both? I'm assuming just space, e.g., fluctuations from the spatial, grid mean values.**

In our equation 1, the fluctuations are relative to spatial averages. We added the word "horizontally" to clarify this ("… prognostic horizontally averaged thermodynamic variable…"

8. **Line 169: It is perhaps awkward to have two different length scales. How does Lϵ related to L used in SHOC?**

The reviewer is broadly correct and this is a critical point. Indeed, it can be argued that these two length scales could/should be related. In practice, however, these length scales play different roles and are somewhat independent in SHOC+MF; and EDMF for this matter since most if not all EDMF-type parameterizations contain a length scale equivalent to SHOC's turbulence length scale (as examples see Suselj et al. 2013, 2019; Han and Bretherton 2019; Lopez-Gomez et al. 2020). Developing more realistic links between these length scales so as to better represent in a unified way the key mixing length scales of turbulence and convection, is one of our critical research tasks in moving forward.

9. **Line 213: I'm confused by this. Doesn't SHOC produce tendencies for u, v like other PBL schemes? Are these not applied too?**

That is correct, i.e., SHOC contains 6 prognostic variables ($\theta_l, q_t, TKE, \boldsymbol{u}, \boldsymbol{v}$, and tracer) in the 3D model. However, the single-column model (SCM) used here sets the winds to the values provided in the forcing file each time the forecast routine is called and unfortunately the forcing file of ARM only contains the initial profiles of *u* and *v*, i.e., it doesn't contain the time-varying vertical profiles of *u* and *v* even though the ARM case represents a diurnal cycle of shallow convection over land. To avoid the wind profiles being "reset" to the initial constant profiles every host model time step, we replaced the wind profiles in the forcing file with the ones from our LES reference run. To be clear, this issue is specific to the ARM case and it is because and citing Brown et al. 2002: "At the time when the case was set up, detailed estimates of the time and height variation of the geostrophic wind, and of any large-scale advective tendencies of the wind components, were not available."

This paragraph was rewritten and reads:
"It is worth noting that we modified the ARM case forcing file to run the model with a 30-minute time step (i.e., the ARM forcing file available in the E3SM SCM library contains values at every 20 minutes). Also, the SCM reads the wind information from the forcing file at every host model time step, however for the ARM case, the large-scale advective tendencies of the winds were not available when the case was setup (Brown et al. 2002), and consequently, the time-varying u profiles in the forcing file were set equal to the initial profiles which are constant with height. Resetting the u profile to the initial vertically constant profile at every host model time step interferes with the development of the TKE field through the shear production term. To circumvent this issue, we replaced the u profiles in the forcing file with the u profiles from our LES reference data; the meridional wind component v is zero in the ARM case. Note that this is specific to the SCM used here and to the ARM case as the large-scale advective tendencies of the winds were not available when the ARM case was setup (Brown et al. 2002)."

10. Line 217: Would it be correct to say that physics then is ONLY using SHOC + MF?

Yes, that is correct. But please note that (i) both BOMEX and ARM correspond to non-precipitating warm shallow convection, so cloud microphysics does not play a key role; (ii) the cloud cover is low so cloud-radiation interactions are not expected to play an important role and (iii) and that the clear-sky radiative forcing is included in the 'SCM forcing file'.

11. Figure 1: Aren't there observations for these case studies? Why are they not plotted alongside LES results?

Both shallow convection cases are idealized case studies based on observations in which the large-scale forcings represent a spatial average of the BOMEX square (500×500 km^2) and the ARM Southern Great Plains site (140 000 km^2), respectively, instead of local forcings. Accordingly, there are no direct observations to compare the modeling results. Further details on this are available in Siebesma and Cuijpers (1995) ("Evaluation of Parametric Assumptions for Shallow Cumulus Convection") and Brown et al (2002) ("Large-eddy simulation of the diurnal cycle of shallow cumulus convection over land"). Nevertheless, LES provides an explicit solution of the fundamental fluid flow equations given the prescribed initial and boundary conditions. As the reviewer likely knows well, SCM comparisons against LES output of benchmark cases are common practice in numerous studies. For example, the Siebesma et al. (2003) paper entitled "A large eddy simulation intercomparison study of shallow cumulus convection" has over 700 citations.

12. Figure 2: Doesn't SHOC have a binormal PDF? Could you also plots means from the "updraft" component of the PDF? Also, it's explained why there are no data points for SHOC + MF at z >=1.5km, but why wouldn't you plot all lines returning to 0? Shouldn't they all do that at some height?

Yes, SHOC has a binormal PDF. Please, see below the properties of SHOC's "updraft" Gaussian (dashed gray line with asterisks). The updraft contribution of SHOC's gaussian is very small and ceases around 800 m as expected based on Figure 1 where the MF contribution takes over in the cloud layer and becomes the only contribution to the total fluxes above 800 m (see the MF contribution in Figure 1d and e; dashed red line).

Regarding SHOC+MF, we thank the reviewer for pointing this out. The SHOC+MF data points stop at the cloud top where the updrafts cease. Please note that the hourly average vertical profiles in panels c and d correspond to $\theta_{l_u} - \theta_l$ and $q_{t_u} - q_t$, respectively; we applied a mask to the data so that grid points where there are no moist updrafts are not included in the hourly average and the difference calculation relative to the mean fields. Figure 2 in the paper has been updated.

To keep Figure 2 as simple and easy to interpret as possible, we did not add SHOC's gaussian contribution, instead we added the following explanation in L276:
"Note that the "updraft" (second Gaussian) moist properties of the SHOC's PDF are not shown because they are quite small (e.g., maximum $w_u \approx 0.3$ m/s) and vanish around 800 m in agreement with Figure 1d-e where the MF contribution makes up the total turbulent fluxes."

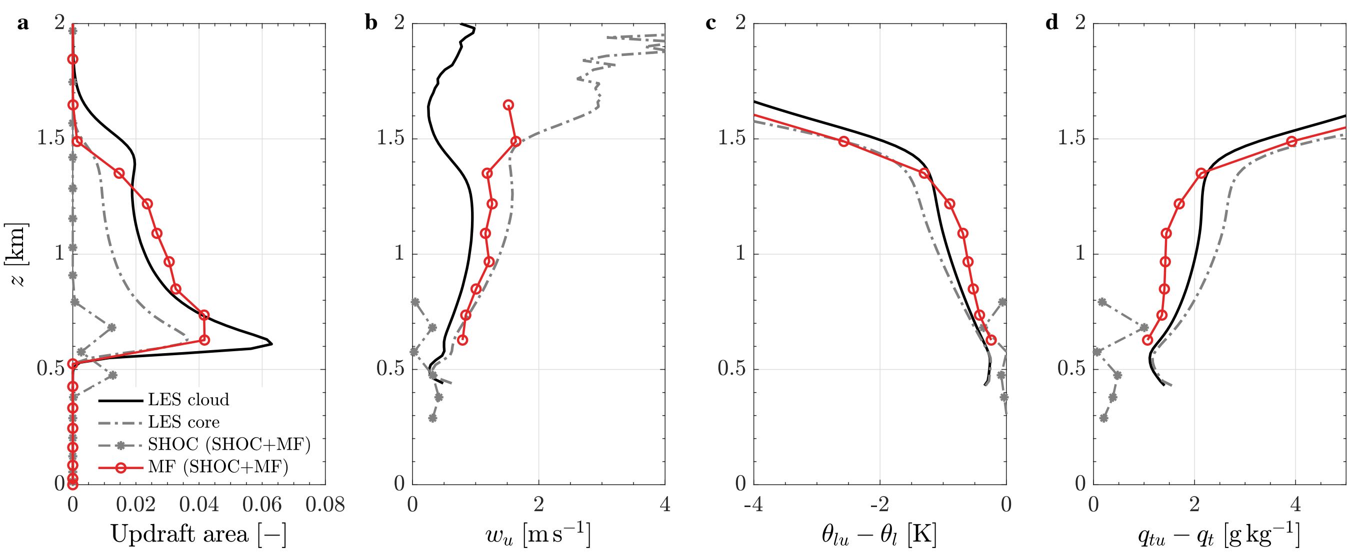

Figure: Vertical profiles of moist updraft properties for the BOMEX case. (a) Updraft area, (b) updraft vertical velocity, and excess relative to the grid-mean values of (c) liquid water potential temperature, $\theta_{lu} - \bar{\theta}_l$, and (d) total water mixing ratio, $q_{tu} - \bar{q}_t$. The solid and dashed black lines correspond to the LES cloud sampling and the cloud core sampling, the dashed gray lines to the SHOC's contribution of SHOC+MF (i.e., the properties from the "updraft" Gaussian of SHOC), and the solid red lines to the MF contribution of SHOC+MF. The profiles correspond to a time average over t = 4–6 h.

13. Figure 3: It's pretty hard to see the LES curves. It might be better to plot the bias (difference from LES) for clarity.

These biases (SCM−LES) are shown in figure 4, and same for figures 5 and 6 but as time-height curtain plots since ARM is a non-stationary case.

14. Figure 6a,b: There are discussions of oscillations like this in the literature. IIRC, Anning Cheng mentions cloud water oscillations related to IPHOC, so it might be worth mentioning that these oscillations are not unique. Also, why do these oscillations not show up in the means in figures 4a,b?

Thanks for bringing this paper to our attention. At a first glance, these oscillations seem to be unrelated to the ones described in Cheng 2004 where the "liquid water oscillations are caused by the interaction of the LWB and the mean gradient of $\bar{s}$ in the third-order equations" (LWB: liquid water buoyancy terms). Here, these oscillations appear to be related to the eddy turnover timescale τ used in SHOC's turbulence length scale (please see now equation 10 in section 2.4) when using its dynamic definition instead of a constant value (in the current development codebase of SCREAM, $\tau = 400$ s).

Lastly, these oscillations show up in figures 4a and b but they are only visible if we increase the colormap limits; however, we would prefer to have the same colormap limits on both the top and bottom panels for comparison purposes.

15. Figures 8,9: Same comment about potentially plotting biases for clarity since all lines are on top of each other.

We appreciate the reviewer's comment but we believe our message comes across as effectively through our current figures. Figures 8 and 9 are for BOMEX which contrarily to ARM (figures 3–4 and 5–6 where 4 and 6 show the biases over time of the profiles shown in figures 3 and 5) is a steady-state case. Ultimately, the lines being on top of each other is somewhat the goal here as it means that the differences between the reference LES and the SCM are negligible.

<u>**Technical Corrections**</u>

1. Line 77 Instead of "PDFs", it is more appropriate to say Gaussian components or modes. The PDF should mean the entirety (weighted sum of Gaussian components).

Thank you for pointing it out. We fixed this sentence and now reads:
"… higher-order moments and cloud statistics appear to only be properly represented when a larger number of Gaussians is used in the joint PDF"

2. Line 121: The second approximation is worded strangely in my opinion. You're talking about a small updraft area but the mentioned term doesn't even contain updraft area. I think it would be better to say that the environmental fraction is large such that we ≈ w, etc.

Thanks. We rephased this sentence and now reads:
"… the second term is neglected because the environmental and grid-mean values are approximately equal (i.e., $w_e \approx \overline{w}$) following the assumption of small updraft area (i.e., $a_u \ll 1$)".