Reviewer #1

This paper performs hindcast CAM simulations for data taken during the EURECA field campaign to attempt to improve the representation of the boundary layer in trade-wind regimes by testing experimental versions of the CLUBB parameterization that include upgradient fluxes for momentum. Overall, I found this to be an interesting paper that addresses an issue that typically receives little attention. Additionally, I found this manuscript to be exceptionally well written, which always makes the job of a reviewer much easier and should not go unnoticed. I do feel this is worthy of publication, after the authors consider some of the following points which I feel would improve the presentation of results and perhaps increase the impact.

Thank you for offering your time to review this manuscript and constructive feedback. Please see point-by-point responses below.

Unlike the Larson et al. (2019) paper, the authors here find that the boundary layer thermodynamic structure is improved with their experimental configurations. I feel it is important to show what impact this has on the simulation of clouds. If the impact is negligible then that could be displayed in a figure or two and would be a worthy reference point for other studies. If the change is more significant then perhaps this should be highlighted a bit more (but I understand this was never the intended focus of the paper) with a discussion of potential implications for climate length simulations and other modeling centers wishing to improve momentum transport.

We have added a figure (Fig. 11) that shows profiles of both the vertical cloud fraction and cloud liquid water content towards the end of the discussion of the experimental length scale. We also add a brief discussion in the text which is reproduced below. However, as the reviewer notes, we emphasize that this isn't a core motivation for this project (and likely merits significant further exploration) so we show it as descriptive and leave a more detailed interpretation for future work.

The following text has been added to the manuscript (in addition to the aforementioned Fig. 11).

"While the specific focus of this work is on the transport of momentum, we show vertical profiles of cloud liquid and cloud fraction in Fig. 11 since a key motivation for understanding boundary layer processes in this region is to improve the representation of low clouds in Earth system models (and their associated forcing on the climate system). When prognostic momentum is turned on (ED-O to PM-O) both cloud liquid and cloud fraction decrease. A decrease in the height of peak cloudiness also occurs. Both of these changes tend to represent a better agreement with the CM1 LES results, although we stress that we have not undertaken a rigorous comparison with observations from a cloud perspective. Nonetheless, we do note these results are qualitatively similar to those published in Narenpitak et al. (2021) and Schulz and Stevens (2023). Turning on the experimental length scale formulation (ED-X and PM-X) results in an increase in cloud liquid and a further reduction in the height of the peak cloudy layer. Both of these further improve the correspondence of the profile shape to the CM1 results, although both liquid and fraction are overestimated in magnitude relative to the LES run. Somewhat interestingly, going from ED-X to PM-X increases cloud liquid, which is counter to the same

change using the original length scale formulation (ED-O to PM-O). While this is just a cursory look at cloud fields, it would imply that changes in the treatment of momentum fluxes also feed back into cloud fields, but that the updated treatment of τ may play an equally or larger role. This is unsurprising given that τ appears in many equations throughout CLUBB, not just those associated with momentum (Golaz et al., 2002). These cloud responses to both momentum treatment and length scale formulation are complex and merit additional evaluation and calibration."

I am not a fan of the naming convention for the simulations (x001, x101, etc,). Why not refer to "x001" simply as "CNTL" and devise similarly brief but easily distinguishable names for the other sets of simulations?

We agree that the original naming convention wasn't overly clear – it was our internal organization strategy, and we should have essentially built a lookup table in our plotting scripts for these in the original submitted manuscript. We have renamed the experiments based on this suggestion. x001 is now "ED-O" (eddy diffusivity, original length scale), x101 is now "PM-O" (prognostic momentum, original length scale), x204 is now "PM-X" (prognostic momentum, experimental length scale), and x304 is now "ED-X" (eddy diffusivity, experimental length scale). This is described in the text and we have also added a table (Table 1) to help readers interpret the experiment IDs.

The authors present results for control CAM, CAM with prognostic momentum fluxes, and CAM with prognostic momentum fluxes but revised length scale definition. I feel it is also very important to present results from CAM with only the revised length scale modification (essentially the Guo et al. 2021 configuration). Having this data point is essential to help the reader tease out the relative contribution of improvement stemming from this modification alone.

In the revised manuscript, we have added another configuration (ED-X) that does this. We do not apply the updated prognostic momentum flux formulation but do update the length scale formulation. This now gives us four simulations (see below response re: the other sensitivity runs) spanning the full space of T/F for both the prognostic momentum and experimental length scale combinations.

I am confused why the authors chose to perform tuning experiments for the configurations that include the revised length scale definition. While I understand it is important to exploit model sensitivity to tunable parameters, it is odd that they chose to do it for this configuration but not experiment x101 (which also introduces new tunable parameters) and this makes the article feel a bit disjointed. In addition, presenting results for all four x20 simulations is a bit redundant and cumbersome. If the authors feel it pertinent to include these tuning simulations perhaps they could just show the average result (and summarize the sensitivity/spread in a short section with a figure or two)? In this instance I feel the authors would also need to add a strong justification why they felt it necessary to do a tuning suite for this particular configuration while neglecting to do so for the others (or perhaps add a tuning suite for the other configurations). Overall, though (unless justification is provided) the main purpose of this paper didn't seem to be about finding optimal tuning parameters nor exploiting CLUBB's sensitivity to them. Therefore, I'm inclined to suggest the authors just present results from one realization of x201 (with whatever the default parameters are from Guo et al. 2021) and briefly summarize that the sensitivity to tunable parameters was explored and highlight the most important points of that analysis.

In response to this (and the above comment) we have decided that having four versions of the experimental length scale indeed does muddy the picture a bit. Based on this feedback (and the above addition of ED-X), we choose to only keep one set of parameters for the experimental length scale (formerly x204, now PM-X). We feel that this tightens the manuscript and allows for focus on the two experimental changes explored here.