Responses to comments from Reviewer #1

We thank you for the helpful comments.

L55 (Table 1): What is the horizontal length scale used to define these weights? Is it something like a ~100 km box surrounding the SGP?

Following Bogenschutz et al. (2020), the E3SM SCM used in this study is configured with ne4np4 grid. The horizontal length scale for the grid cell is ~ 800 km. This detail has been added into the context now in L62.

L89: How big of an impact do you think using a homogeneous u could have? Please comment on this if possible*

The control run HOM is using a homogeneous u*, computed on the basis of grid-cell mean momentum. The impact generally can be considered as the differences arising from the HOM and HET runs as discussed in Section 3.

L115: Is the near-surface atmospheric state used for each patch the same? i.e. is the atmosphere treated as homogeneous?

The near-surface atmospheric states for each patch are different. E3SM land model has subgrid-scale treatment for patches. Though the atmospheric states passed from atmosphere to land are treated as homogeneous (grid-cell mean), E3SM land model applies Monin-Obukhov similarity theory to determine the atmospheric states for individual patches according to their unique characteristics. We have clarified this now in L122.

L130: You mention in figure 2 and in the conclusion that the fluxes are computed offline. Could you define/explain that in more detail here?

Done. We clarified it starting from L136 now as "In the HET configuration, we used the atmospheric state variables and surface fluxes for each patch in ELM saved at every model time step from the HOM configuration to compute the spatially heterogeneous characteristics (i.e., surface variances and covariance of potential temperature and specific water content) following Eqs. (7-9) in Section 2.2. These surface heterogeneous characteristics were then provided to CLUBB as the lower boundary condition." We also clarified "offline" in the caption of Fig. 2 as "The heterogeneous surface moments are computed offline by our implementation of the HET approach (white boxes on the right)."

Responses to comments from Reviewer #2

Summary

In this paper, the authors present a modeling study that aims to improve the representation of subgrid heterogeneity when coupling with the atmospheric boundary layer (ABL). A new coupling parameterization is implemented into the E3SM model which accounts for the variability from each individual sub-grid patch and inter-patches variability. The new parameterization is tested in a single-column model over the ARM site in Southern Great Plain and the results are analyzed for non-precipitating days, cloudy days, and precipitating days.

This research topic is very important in the land and atmosphere modeling communities and the authors have documented the new parameterization in detail in this manuscript. Please see the attached document for some places where the authors could further improve this manuscript. Please see the comments and questions below.

We appreciate all your comments.

General comments

There is no observation data shown in the manuscript. If this study is performed with the ground truth fraction of land patches in the ARM site, it makes sense to compare the HET and HOM simulation performance with site observation data, if there are any. Many of the analyses conducted in the paper, i.e. temperature, humidity, cloud fraction, precipitation, can be compared with observation data. This can add creditability to this study and readers can have a direct sense of the HET performance over HOM.

We have included analysis comparing the SCM simulations with observations in Fig. S7. When comparing the SCM results with the ARM measurements of near surface air temperature, humidity, total cloud fraction, and precipitation, we find that the differences between HOM and HET are negligible compared to their differences from the ARM measurements. This has been added in L151. We note that this is a sensitivity study aimed to assess the effects of the HET scheme on the simulations. In an on-going effort to assess the effects of the HET approach on global simulations, we perform model evaluation against various observations and results will be documented in a separate manuscript.

Another point would be that water and vegetated bare ground are two major contributors to interpatch variability, which is also a major component in HET different from HOM, while these two patches only account for a small portion of the study area. Generally, in a model sensitivity study, authors can have larger free space perturbing the different fractions for the eight patches. Would it be possible to demonstrate to what extents the findings in this paper are subject to fractions of land patches? A similar question would also apply to the atmospheric forcing too.

In this study, the initial condition and large-scale forcings are derived from the surface measurements and reanalysis for the ARM SGP site. Therefore, applying those forcings to an arbitrary grid cell that is very different from SGP may potentially introduce inconsistency between land and atmospheric conditions. We assessed the impacts of patch weights by performing additional sensitivity tests that (1) perturb the lake patch, and (2) increase the horizontal resolution to typical ne30, so that the patch weights were perturbed and yet still represented the land conditions over the SGP site. In the lake patch experiment, we perturb the lake patch by 20% of its original weight. In "Lake+," the lake portion is increased from 1% to 1.2% relative to grid cell while the other patches are reduced accordingly. In "Lake-," the lake patch weight is reduced from 1% to 0.8%. In the resolution sensitivity test, we increase the land model resolution from ne4np4 (~ 7.5°) to ne30np4 (~ 1°). As a result, the patch weight perturbation that we introduced on the premise of a realistic SGP forcing has negligible impact on our analysis and conclusion (Fig. S2-S6), even though the surface variances are moderately affected (the dashed lines in Fig. S1). We added this discussion now in L183.

Specific comments

1. *Introduction*

The authors briefly summarized the representation of subgrid heterogeneity in land and atmosphere components in ESMs, with a specific focus on E3SM version 1. It will benefit a larger group of readers if the authors can provide how the coupling process is handled between the land and atmosphere model in other ESMs or RCMs. If it is a common issue, the new parameterization and findings in this paper could be a valuable lesson for a large modeling community.

Yes, it is a common approach. This comment was addressed in L42.

2. *Methodology*

Table 1: It will be good to show or describe the spatial extent of the ARM SGP site, i.e. how large is the study area?

The study area is a 7.5° × 7.5° grid cell (~800 km) centered around the ARM SGP site. Description about the grid size was added into the text at L62.

L118 "A schematic diagram . . . ": should it be Fig. 1 or Fig. 2? Fig. 1 shows the potential temperature for eight different patches, indicating spatial heterogeneity. (I couldn't find the description of these different potential temperatures in the texts).

It is meant to be Fig. 1. Using Fig. 1 we illustrate the concept of inter-patch and sub-patch variability that we defined in Eqs 7-9. These potential temperatures are computed according to Monin-Obukhov similarity theory. This has been clarified now in L122.

L132 "A short-term hindcast approach . . . ": Can the authors provide more details on this configuration used in the SCM? What are the large-scale prescribed forcing and initialization data? A note to this run 48-hour hindcast is that it may not have enough time for land surface feedback to the atmosphere, so the land model may be strongly controlled by the atmospheric forcing.

The large-scale forcings are from ARM continuous forcing which is well documented by Tang et al. (2019) and Xie et al. (2004). Details and corresponding references were added into the text at L147.

The model is indeed constrained by the large-scale condition with the hindcast approach. We have added the text in L143: "This hindcast approach avoids mixing long-term simulation biases with fast physics errors, allowing us to focus on the effects of different land-atmosphere coupling schemes."

3. *Results*

Figure 3: the modified HET simulation is quite different from the default HOM at two time periods in a day, near 18:00 LT and 12:00 LT. While the authors explain the discontinuity between two consecutive days is expected due to the hindcast configuration, why does it affect more strongly on HET than HOM simulation? It is because HET simulation is more sensitive to the initialization at 18:00 LT? But the first 24 hours are not included in the analysis. It might be good to provide an explanation.

The HET variances largely differ from HOM at 1800 and 1200 LT because the atmospheric states over lake (patch #32) at that time are quite different from others (refer to the discussion about Figs. 4 & 5 in the manuscript). To examine the reviewer's hypothesis, we show the average surface temperature variance from simulated 0-48 h in Fig. 1. It reveals that the diurnal patterns for 0-24 h and 24-48 h are similar but of different magnitudes. The initialized HET temperature variance differs from the HOM variance because the HET variance is derived from the patch anomalies of the grid cell mean while the HOM variance is based on the grid cell mean flux. Hence, we consider the first day as spin up and apply the short-term hindcast approach in order to avoid mixing long-term simulation biases with fast physics errors. It has been clarified now in L142.

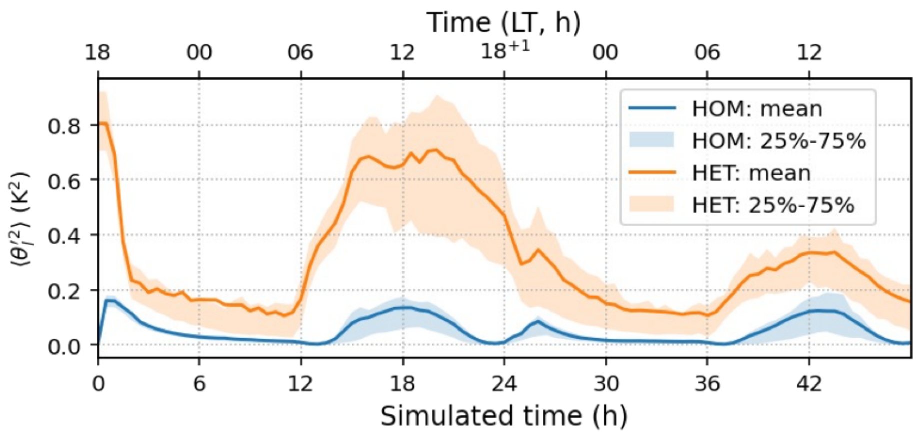

Figure 1: 0-48 h mean temperature variance averaged from June to August 2015.

L189: it is surprising that the sensible and latent heat flux between HET and HOM are not so different. From a modeling perspective, what does this finding suggest? Is it because the study area is by half occupied by the croplands while water and vegetated bare ground only make up a small portion? Can one expect larger differences in turbulent fluxes between HOM and HET with a higher fraction of water and bare ground fraction?

In Eq (6) at L112, if we let x=w and y=θ, the heat flux $< w'\theta' >$ will be equal to $< \overline{w^s\theta^s} >$ because $< \overline{w}'' >= 0$. Therefore, the sensible heat fluxes from HOM and HET are expected to be identical. Also, as can be seen in Figs. S3 and S5, the differences in turbulent fluxes remain small in the sensitivity runs with large lake fraction ((i.e., HOM_lake+ versus HET_lake+).

Fig. 9a: the cloud fractions for both HOM and HET seem very low for cloudy days.

Considering the clouds in this study are populated over land, we think the cloud fractions are fair. As shown in Fig. S7, the simulated total clouds from HOM and HET are even higher than observations.

Table 3: another possible thought that HOM and HET are very similar might be the radiative and turbulent fluxes are so strongly controlled by the atmospheric forcing, while land surface heterogeneity, under this configuration, does not provide sufficient feedback to the atmosphere. The authors may provide a more enriched discussion on this and readers will surely benefit from it. Similar comment to section 3.3 as well.

Though the near-surface atmospheric states are hardly influenced by the HET scheme because of the strongly constrained atmospheric forcing, we still see in Section 3.3 that the vertical profiles in particular of the skewness of vertical velocity (Sk_w) and CLUBB-diagnosed low clouds are in response to the surface boundary conditions provided by HET. This single column model configuration allows us to improve mechanistic understanding without being confounded by complex feedbacks in the atmosphere. Assessing the full effects of the HET approach on the global climate requires further investigation and will be documented in a separate manuscript. We have included these text in L310 now.

Fig. 12: it will also be good to show the precipitation observation in the same figure to demonstrate whether HET improves from HOM.

The precipitation observation is shown in the Supplement (Fig. S7).

L244-L253: the authors are trying to make connections between the vanishing of inter-patch variability with precipitation occurring in June. So, it might help to show also the timeseries of the inter-patch variability in Fig. 14. What's also helpful is the potential temperature and water content for different patches, as in Fig. 4 or 5, during precipitating days. This will help to make this argument stronger.

The timeseries of inter-patch variability are depicted in Fig. 5. We discussed the results in L275.

References

Bogenschutz, P. A., S. Tang, P. M. Caldwell, S. Xie, W. Lin, and Y.-S. Chen, 2020: The E3SM version 1 single-column model. *Geoscientific Model Development*, **13**, 4443–4458, https://doi.org/10.5194/gmd-13-4443-2020.

Tang, S., S. Xie, M. Zhang, Q. Tang, Y. Zhang, S. A. Klein, D. R. Cook, and R. C. Sullivan, 2019: Differences in eddy-correlation and energy-balance surface turbulent heat flux measurements and their impacts on the large-scale forcing fields at the ARM SGP site. *Journal of Geophysical Research: Atmospheres*, **124**, 3301–3318.

Xie, S., R. T. Cederwall, and M. Zhang, 2004: Developing long-term single-column model/cloud system–resolving model forcing data using numerical weather prediction products constrained by surface and top of the atmosphere observations. *Journal of Geophysical Research: Atmospheres*, **109**.