Reponses to Reviewer #2

This paper uses three years of post-processed NEXRAD radar data to assess the performance of a recent version of the new E3SM model, using reflectivity simulations. The results are interesting, and the presentation of the results is clear. I have a number of comments and suggestions to improve the scientific content of the paper, which is a bit on the light side. They are all somewhat minor comments for consideration by the authors, except one more major comment that should be fully addressed. Therefore, I recommend the paper be accepted with minor revisions provided that at least my major comment is addressed.

Thanks Alain, for your valuable comments to improve the paper. See our point-by-point responses as below.

Major comment:

My main comment is about the use of an 8 dBZ threshold and the implication for echo top height conclusions in the paper. There is nothing wrong with using such threshold, but it introduces in my opinion a possible misinterpretation of the results regarding the echo top height statistics to address the all-important question: does my model reproduces the vertical development of convection well, statistically? Comparing echo top heights between observations and models is very tricky, because using a threshold in reflectivity implicitly carries the assumption that the echo top height is not affected by the threshold when you draw conclusions. Let me give an example: say the model is underestimating reflectivities in ice phase by 5 dBZ (in your case it seems to be more than that). If you want to learn something about how good the model approximates cloud top height statistics (and indirectly your vertical air velocities in deep convection), you should actually use 3 dB echo top heights from the model compared to 8 dB echo top height in the observations to be fair to the model. In your case, you find a substantial underestimation of reflectivities in the upper levels, well, then it's not surprising that you are underestimating the 8 dB echo top height by a large amount.

Thanks for the comment. We agree with the reviewer on the possible caveats with the threshold of reflectivity. But this cannot be avoided when comparing with model values over a 100-km grid. We have used a lower threshold of 0 dBZ to see how the results are sensitive to the choice of the threshold. As shown in Figure R1, we do see an increment of ~1 km in the simulated echo top height, however the observation doesn't change much. As a result, switching to the lower threshold of 0 dBZ has a very limited impact on the main conclusion that the model severely underestimates the echo top height.

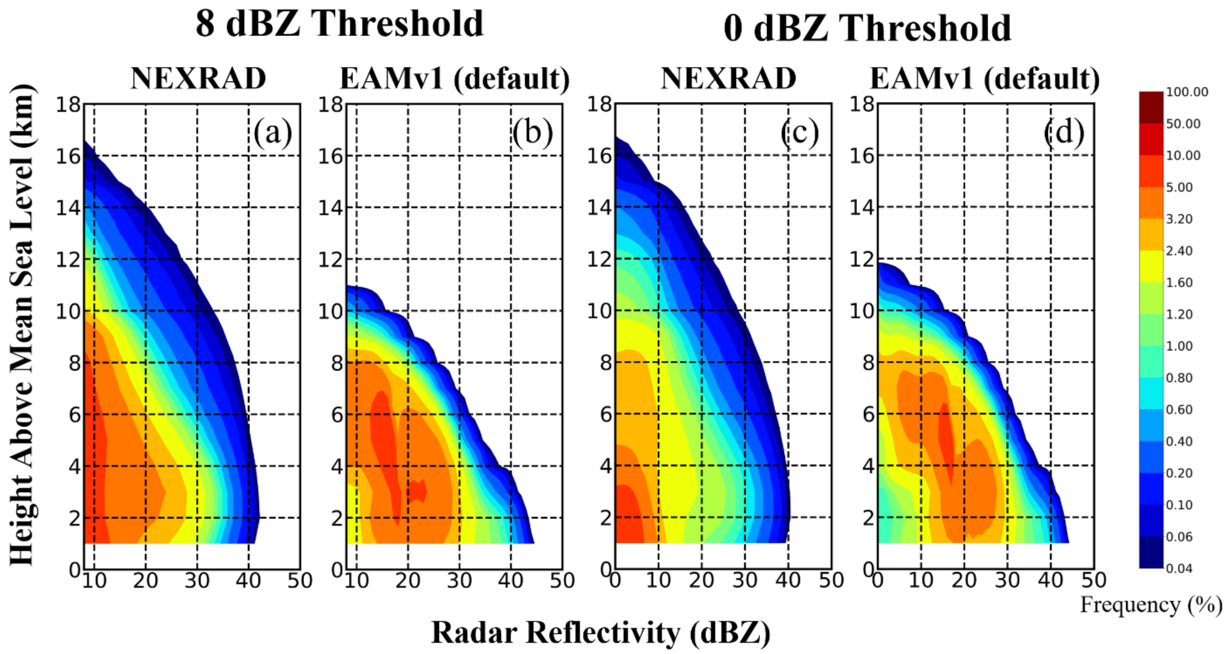

Figure R1. The sensitivity test of changing the minimum reflectivity threshold from 8 dBZ (a, b) to 0 dBZ (c, d).

We have added statements to Section 3.3, "From Fig. 5 it is clear that the model severely underestimates the echo top height by at least 5 km. To look at how this result is sensitive to the threshold reflectivity, we reprocessed the results with the 0 dBZ threshold. By lowering the threshold to 0 dBZ, an increment of ~1 km in the vertical extension of CFAD is found in the model, but the echo top height of the observation is not changed much. As a result, the choice of threshold does not change the conclusion of severe model underestimation in echo top height."

But what do you learn with this about the deficiencies in the model, especially about the convective vertical velocities and convective mass flux assumptions. You could check, maybe. There is a very good discussion on this in the Labbouz et al. (2018) paper. This paper also tackles similar issues but using different types of comparisons, so I believe it should be quoted in your paper (as well as references therein).

The literature recommended and references therein provided an in-depth discussion of how to improve the modeling of convective clouds in GCMs to better match radar retrievals. In this study, we are not aiming for that purpose. We conducted the direct comparison of reflectivity between model and radar to identify model biases, and we did some tests by tuning a series of parameters in the ZM cumulus scheme and cloud microphysics scheme to see if the large biases in the echo top height can be alleviated. The results of this study can provide metrics for evaluating the cumulus parameterizations or provide insights for improving the cumulus parameterizations, which would be nice follow-on work.

We have cited the paper and provided a discussion about the further work at the end of Section 4. "In addition, the results of this study can provide metrics for evaluating the cumulus parameterizations or provide insights for further improving the cumulus parameterizations like Labbouz et al. (2018), which can be a follow-on work."

On another hand if you take the model cloud top height, chances are that the NEXRAD radars won't have the sensitivity to detect it, and this time the radar statistics will show lower cloud top heights than the model (I think I have seen statistics somewhere showing a systematic ~ 2km difference between cloud top height and 0-dBZ echo top heights, but I can't find the reference just yet). If it does not though, it would mean that the model really underestimates cloud top heights, which would be very interesting.

We would like to clarify that we only focused on the echo top height and did not look at cloud top height. Evaluating echo top height allows us to know that the model failed to simulate the occurrence of large ice-phase particles at high levels in deep convective clouds. As the reviewer mentioned, NEXRAD won't have enough sensitivity to detect the cloud top height so we cannot compare the observed echo top height with the modeled cloud top height. To ensure a fair comparison, the same radar threshold has to be applied. We tested the sensitivity of our results with a different threshold (0 dBZ) as shown in Fig. R1, and the model still severely underestimates the echo top height.

Specific comments:

1. The 13.6 GHz versus 3 GHz difference: my only issue with this is that you should make sure that you have switched off any Ku-band attenuation correction in COSP. Have you? If so you should mention it in the paper.

We actually turned on the Ku-band attenuation correction in COSP, and we believe this is still a valid comparison. First, Rayleigh scattering is satisfied at 13.6 GHz frequency with respect to gases and most ice/liquid particles, thus the attenuation correction making no differences for those hydrometeor species. Secondly, it is extremely difficult for the global model to simulate ice particles with a size large enough to be comparable with the wavelength (~2 cm), which has been discussed in the CFAD comparison. Last and the most important, the COSP mimics the satellite view from space to the ground, therefore the layer most vulnerable to the attenuation caused by large precipitation droplets is close to the ground (i.e., 1 km), which has been excluded from the comparison. The clarification has been added, "In the COSP simulator, the 13.6 GHz frequency ensures the Rayleigh scattering calculation. Although an attenuation correction has been applied, because the COSP mimics the satellite view from space to the ground, the layer below 1-km altitude is most vulnerable to attenuation caused by large precipitation particles, which has been excluded from the comparison."

2. Line 139: it does not make sense to say that the minimum detectable reflectivity of NEXRAD radars is 0 dBZ. The MDR varies with range (MDR(range)=MDR(1km) + 20 log10(range_km)), so you either need to provide the sensitivity at 1 or 10 or 100km (whatever you prefer) or state that a hard threshold of 0 dBZ is applied somewhere in the NEXRAD processing. Also, I would be surprised if the NEXRAD radars can detect 0 dBZ at 250km range (but I don't have those numbers).

We didn't directly use the original NEXRAD scan data, but the gridded 3D mosaic data. The 0 dBZ should not be the threshold of the NEXRAD but the threshold of the dataset we use. A correction has been made in the text, "as shown in previous studies (e.g., Wang et al., 2015, 2016, 2018; Feng et al., 2012, 2019), the minimum reflectivity of the 3D mosaic NEXRAD dataset is 0 dBZ (Fig. 1a)."

3. Back to the 8-dBZ threshold: how sensitive are all your results to the use of 8-dB threshold? What happens if you take 0 dB (you say NEXRAD can detect 0dB), so why didn't you use 0 dB to get closer to the true cloud top? It would make more sense in my opinion.

See our response to your major comment above.

4. Line 180-181: again, you say here that your most distinct result is about echo top height. But I believe it does not teach you anything about potential model deficiencies, including the convective vertical velocities and mass flux issues highlighted in Labbouz et al. (2018) and others. Your main result is actually that the model is underestimating reflectivities, but may well be excellent at producing realistic cloud tops. The only thing is that they have lower reflectivities than in reality.

See our response to the major comment above.

5. Line 194: this statement is likely wrong. I would bet that it simulates reflectivities at heights greater than 11km, but they are under your 8 dB threshold due to ice microphysics deficiencies.

We used a lower threshold 0 dBZ and the conclusion is not affected, as seen from our response to your major comment.

6. Line 226: ~5dBZ: it is more like 4 dB and only for Z<25 dBZ, so maybe this statement should be modified to reflect this.

We have modified the statement to "In addition, the modified microphysics assumptions produce higher values of reflectivity, in better agreement with observations, and the grid-mean radar reflectivities increase by ~4 dBZ (Fig. 3) mainly for values less than 25 dBZ.

7. Line 323: typo "Brodzik"

The correction has been made.

8. Figure 2 and associated discussion in section 3: comparing mean reflectivities is interesting, but only one aspect of what you'd like to get right with a model. The two main other things I would personally try to assess is the standard deviation of reflectivity at 100km scale (to check if the model reproduces the observed variability even if it does not have the mean right) and the 95th percentile or even 99th percentile if you have enough samples (to check if the model has any skills in forecasting extremes). These two additional things would greatly enhance the scientific content and scope of this model evaluation exercise.

We have added a Table to the manuscript (Table 2) to include the standard deviation and the 95th percentile values. The discussion has been added correspondingly in Section 3.2 and the following sentence has been added to the Conclusion. "EAMv1 can simulate the variability and extreme value of reflectivity at the lower troposphere but significantly underestimate them at high levels."

References:

Labbouz, L., Z. Kipling, P. Stier, and A. Protat, 2018: How well can we represent the spectrum of convective clouds in a climate model? Journal of the Atmospheric Sciences, 75(5), 1509–1524.

Good luck with the review,

Alain Protat

Melbourne, 25/06/2020