Review of "Using radar observations …" by Wang et al.

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.

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. But what do you learn with this about the deficiencies in the model, especially about the convective vertical velocities and convective mass flux assumptions. 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 heghts, which would be very interesting. 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).

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.
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).
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.
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.
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.
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.
7. Line 323: typo "Brodzik"
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.

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