Review of "Using Radar Observations to Evaluate 3D Radar Echo Structure Simulated by the Global Model E3SM Version 1" by Jingyu Wang et al.

I've been drafted in and did not review the original version of this paper, but I agree with the reviewers' original compliments regarding its readability and usefulness. I judge that the authors address the original reviews and the addition of the 0 dBZ threshold sensitivity test is particularly helpful for interpretation.

The authors reject reviewer 1's suggestion to analyse uncoupled runs and I side fully with the authors' response. The submitted paper explores issues with physical parameterisations and COSP within the model, this alone is a large undertaking that would be hindered by additional model-observation discrepancies introduced via natural variability in free-running simulations. The new version of the text makes this argument succinctly.

I have one technical comment regarding Figure 2 that I believe must be addressed for accuracy, and request that the authors cover some additional ground in the discussion to really tie this paper up. I also put a set of "minor" and "very minor" points where I suggest grammar or text to reduce ambiguity.

I am requesting only minor changes and I don't think the conclusions will be strongly affected. I think the authors have done a nice job with this paper and would support publication after these changes.

Main points
   1. **Figure 2**

My understanding is that Figure 2 shows "sub-grid" statistics, i.e. there are N grid cells included, and the observational histograms contain $625N$ entries while the model histograms contain $50N$ entries. Isn't this like comparing the histograms of properties at different spatial resolutions? The histogram from the smaller-grid-cell sample (i.e. observerations) would generally be broader anyway, so the two cannot be directly compared. Instead, the observations should be averaged somehow in order to provide 50 per grid cell.

If I am wrong about this then please clarify in the text.

Fortunately, it doesn't look like your overall conclusions would be strongly affected.

   2. **Discussion**

Some potential limitations of the analysis are not covered, primarily related to the model-observation comparison.

I haven't used COSP for surface measurements but I presume your assumed viewing geometry is for an upward-pointing radar. Please provide the assumed viewing geometry in Section 2.2 and in the discussion you should cover whether this is somehow addressed or if it may introduce model-observation discrepancies.

For example, are there regions in which topography may affect the NEXTRAD data like it can do for precipitation frequency (e.g. Smalley et al., 2017: https://doi.org/10.1175/JHM-D-16-0242.1)? Probably not for your big central MCS regions, but maybe in others.

Is attenuation a problem? If your simulations are looking up and you compare with a NEXRAD radar that's looking side on to a convective storm, then perhaps model pulse gets attenuated by the dense convective core so you see lower high-altitude dBZ than you would get from the NEXRAD radar. It's my understanding that even S-band upper-level dBZ can be affected in extreme cases (e.g. lots of big particles in the melting layer), and these extreme cases might contribute a substantial fraction of the very high altitude/high dBZ results.

Finally, the discussion might also benefit from laying out the primary other factors which could contribute to your model-observation differences. For example; you modified the COSP particle size distribution and see changes, but who's to say it's now realistic? What features of the model-observation discrepancy could perhaps be explained by these factors; how far can you exclude them; and are there obvious tests for future studies to address and rule out such factors? I remain concerned about the 13.6 GHz to 3 GHz change too, and this should be mentioned again in the Discussion section to remind readers. Putting these things all together will make it easier for the community to contextualise and use your results.

I insist on commentary about the viewing geometry and its expected implications for the results, the rest of the potential expanded discussion I leave to the authors' discretion.

Minor points
P2L54—56:
"As discussed by Iguchi et al. (2018), precipitating ice particles have a large variation in habits and scattering properties, and the effect of non-Rayleigh scattering and multiple scattering by large precipitating ice particles could introduce large uncertainty into simulating the cold-season radar reflectivity field. To avoid this uncertainty, we examine only the warm season of the three years from 2014 to 2016."

I interpret that as saying that it is only in the cold season that you see (i) ice particle scattering, (ii) non-Rayleigh scattering and (iii) multiple scattering. Specifically, you "avoid" it (i.e. it is zero) during the warm season. I think it happens sometimes in warm season storms. Suggestion:

"As discussed by Iguchi et al. (2018), precipitating ice particles have a large variation in habits and scattering properties, and the effect of non-Rayleigh scattering and multiple scattering by large precipitating ice particles could introduce large uncertainty into simulating the radar reflectivity field. To reduce uncertainty due to these factors, we examine only the warm season of the three years from 2014 to 2016."

P4L108:
"…The detailed documentation of those changes is in Table 1…" this table is appreciated and efficiently carries important information. I would like to see all the Default values, where there are no changes you could insert "-" in the Modified columns. This would provide complete information and visually guide the reader to identify where changes have been made.

The top two rows report Gamma distributions with "width" of "0". What is this width? Since you're already reporting the mean then I think it would be consistent to insert the variance in the final column and then add text to the caption to explain this is what you've done.

If I have misunderstood, then please expand the caption to avoids such misunderstandings.

P5L130/131 and PL140:
"…we nevertheless perform the Gaussian smoothing of GridRad data to match the model time step (30 min) in the comparison." AND "The simulation data are saved hourly, consistent with the hourly GridRad data."

I can't understand this – these sentences appear contradictory. Please clarify. For interpreting the results I assumed the second description applied.

P5L136:
"We also did the test with 0 dBZ to look at the sensitivity of our key results to the choice of the threshold value. Thus, after coarsening the 4-km GridRad data to a model grid element, only the grid elements with a mean value larger than 8 dBZ are taken into account in both observations (Fig. 1b) and simulation (Fig. 1c)."

This is the first time you mention the sensitivity test and I think it could be clearer. Example suggestion:

"We also tested with a threshold of 0 dBZ and report later on how it only has minor effects on our conclusions. For our main results, after coarsening the 4-km GridRad data to a model grid element, only the grid elements with a mean value larger than 8 dBZ are taken into account in both observations (Fig. 1b) and in the simulation (Fig. 1c)."

Very minor points
P1L30: "Over the continental U.S." would be a good point to introduce the "CONUS" acronym which is used later without expansion.

P2L31/32: please insert wavelength or frequency after S-band here. The earlier the better.

P2L49: "over the CONUS for the three years (2014-2016)" parentheses are jarring, please remove.

P2L52: "Over the CONUS, warm-season is dominated by convective processes". The hyphen makes me think "warm-season" is intended as a compound modifier so I guess you're just missing the word "precipitation".

P3L76—77: "pressure-based terrain following coordinate" – you could optionally also insert "hybrid sigma" descriptor here to introduce it earlier than Section 2.

P3L89: "…spaceborne satellites…" typo: "satellites" is an adjective here so should be singular.

P3L92: "…direct measurements form 3D scanning radars…" typo: I think you mean "…from 3D…"

P4L95: "…pseudoobservations using forward calculation" typo: missing article or pluralisation (e.g. "forward calculations").

P5L130: "we nevertheless perform the Gaussian smoothing" typo: I don't think you need "the" before "Gaussian smoothing".

P6L189: "Meanwhile, the modeled standard deviation and the extreme values are smaller, indicating the model has a difficulty to capture the observed verifiability." – I don't understand "observed verifiability" and I'd replace "to capture" with "capturing".

P7L208/209: "For the reflectivity >35 dBZ, simulation has a higher probability". Looks like "the" is in the wrong place, I think it should be: "For reflectivity >35 dBZ, the simulation has a higher probability…"

P7L211: "the percentile values are consistent between model and observations". Looks like missing "the" before "model". There are some other cases like this, just keep an eye out for that when you skim through.

P7L222: "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". I think "the" is needed before CFAD (it is not a proper noun, there are many CFADs) and "observation" should be pluralised here.

P8L246—248 "Xie et al. (2019) improved the diurnal cycle of precipitation in E3SM v1 recently by modifying the convective trigger function in the ZM scheme. It will be interesting to see if it can simulate the double-peaks in observed column-maximum reflectivity in the future.". This is interesting and useful context, good inclusion.

P8L249: "3.4 Sensitivity of Simulated Echo Top Height Tunable Parameters of the Global Model" I had trouble parsing this. Do you mean sensitivity of simulated echo top hight *to* tunable parameters?

P8L250: "Different from the model evaluation of". This seems grammatically weird to me, perhaps "Differently from…", although I'd probably pick "Compared with…"

P8L254—255: "tunable parameters as listed in Table 3. Each test is based on the default setup for all other parameters.". This sentence makes sense but thanks to the structure I re-read it a couple of times to be sure I'd understood it. I suggest something more explicit, like "In each test a single parameter is changed, and all other parameters retain their default values".

P9L276—277: "In summary, changing any single parameter alone in the ZM scheme does not improve the simulation of echo top height." Did you change all the parameters, or is this a select subset? If a select subset then I think you should specify here: "changing any of our selected parameters individually in the…"

P9L278—279: "(i.e., those resolved by model resolution)." Is repetitive, how about "those resolved by the model"?

P9L280: "and precipitation by changing the large-scale forcing on which cumulus clouds are calculated" it sounds unnatural to me that clouds are calculated "on" a forcing. How about: "…the large-scale forcing which feeds into the cumulus cloud calculations".

P9L283: "Attempts of accelerating" I think should be "attempts at accelerating".

P9L287: "only gains 500-800 m increment" missing "a" before "500—800 m"

P10L301: "With default microphysics assumptions" I think this would make more sense as "the default microphysics assumptions", since you're referring to the individual set of assumptions in this model, rather than assumptions in general.

P11L1—2: "circulation is nudged towards observations for the simulations in this study, which represents the upper bound of model performance." Again the phrasing is ambiguous here to me, because it's not clear what the "which" refers to among all the nouns in the earlier part of the sentence (nudging? Circulation? Simulations?). My first choice would be to remove everything after the comma because the next two sentences explain it, but if you really want to keep that bit then how about "…for the simulations in this study, so our results represent the best-case model performance".