Dear editor and referees:

Thank you very much for the advice to improve the quality of this manuscript. We carefully addressed all the comments and made corresponding changes to the manuscript. In this "response to comments" document, we provided detailed responses in blue bold as below.

Referee #1: Peter May, peter.may@monash.edu

I am broadly happy with the paper and responses, except for the following.

1. Line 48: "Our goal is to provide a comprehensive evaluation of both horizontal pattern and vertical structure of cloud and precipitation." – How can you do this if you do not discriminate between convective, stratiform and large scale stratiform processes given the fundamentally different processes, heating and drying profiles. Radar simulations on such large grid sizes are already problematic, but given a key part of a convective paramaterisation is the fraction of the grid, surely the model and interpretation should could be improved. This is clearly beyond the current work, but should be acknowledged and planned for future work.

We have added the discussion about future work. "Future studies can also focus on separately evaluating properties in convective and stratiform regions, since the thermodynamic and reflectivity profiles are fundamentally different between the two regions."

2. CFADS without considering this are limited in value in my opinion. Processing the radar data for convective and stratiform fractions in observations is straightforward (e.g. Steiner's algorithm) and readily compared with convective fraction in cumulus paramaterisations and would provide a clear test for the models.

Convective cloud fraction is not parameterized in mass flux-based convection schemes including the ZM scheme. It is assumed to be <<1 for typical GCM resolutions such as at 1-degree grid spacing or coarser. Since radiative transfer calculations need it, convective cloud fraction is separately diagnosed in these schemes. In the ZM scheme, it is fitted to be a function of cloud mass flux, thus should be viewed as a tuning parameter for cloud radiation calculation. As such, it is not very meaningful scientifically to evaluate it in the current ZM scheme. However, in the future if it becomes an independent variable in a convection scheme (for instance, for grey-zone resolutions convective cloud fraction will be needed in parameterizations), then evaluating it will be meaningful.

3. L154. "In EAMv1, 50 sub-columns are used for calculating the mean radar reflectivity for a model grid box. There are 625 pixels inside each 1° grid for NEXRAD data to provide a probability density function (PDF) of observed reflectivity within the box. F. " – how can this be reliably done if you do not know what fraction is convective given radically different reflectivity profiles and magnitudes – even for similar rain rates?

We didn't consider the convective fraction when calculate the grid-mean reflectivity. All the subgrid reflectivity values and NEXRAD pixels within each 1° grid box are linearly average with no discrimination. We understand the reflectivity profiles are significantly different between stratiform and convective, but the COSP calculates the subgrid reflectivity independent of what E3SM uses. More importantly, the convective cloud fraction is not parameterized in mass flux-

based convection schemes including the ZM scheme (assumed to be <<1 for typical GCM resolutions such as at 1-degree grid spacing or coarser), therefore its evaluation is not very meaningful. The clarification has been added, "in addition, the subgrid distribution results from COSP are calculated based on the assumption about the distribution of cloud and precipitation among the 50 subcolumns, which is independent of what E3SM uses. Therefore, a higher-order consistency between the COSP and the host model is warranted in future studies. In this following analysis, we focus on the evaluation of the simulated 3D radar reflectivity field at the model's native grid, which is 1-degree, since the subgrid information from COSP does not directly reflect how E3SM does it. Also, the convective cloud fraction is not parameterized in mass flux-based ZM scheme and is diagnosed from cloud mass flux for cloud radiation calculation, which is treated as a tunable parameter, whose evaluation is not very meaningful unless it becomes an independent variable, for instance, for grey-zone resolutions."

4. L211 –" the reflectivity below 4 km is consistent", but in the CFAD's the model has an odd double peak in the reflectivity and low numbers of low reflectivity compared with the observations. Some of this could be associated with the downscaling, but could also be due to the lack of proper classifications when downscaling the model reflectivity. I think these differences are significant and potentially important. I would also note identifying issues such as this is exactly the point of papers such as this.

We have discussed the discrepancy of CFADs' overall shape between the model and observations and related this to the lack of separation between convective and stratiform. However, we think averaging to 1° should have more impact on the observation than the model as mentioned by Referee 4's minor comment 7 since the double peak shown in the model is only related to the choice of display interval. As a result, the following discussion has been added, "Regarding the overall shape of CFADs, the model follows the well-known pattern where the reflectivity value range of high frequency zone (> 3.2%) increases from cloud top to the freezing level, and then slowly decreases or remains constant below the freezing level. The cores of maximum frequency (> 5%) are located in the centres of the high frequency zones. However, these characteristics are not presented in the observations, whose high frequency zones are greatly skewed to the lower reflectivity values. The characteristics of NEXRAD's CFADs could be due to averaging from fine resolution (4 km) to coarse resolution (1°), as well as averaging of convective and stratiform components because the two components produce significantly different reflectivity profiles and magnitudes."

5. Figure 6 showing the diurnal cycle looks very odd. In the model there is essentially none despite the well known biases in global models for initial deep convection too early in the day and the observations show two very narrow peaks that certainly are not consistent with previous observations such as the cited Carbone and Tuttle paper. As the authors say, it may represent issues in triggering convection in the model, but without information on convective fraction it is difficult to know. Note that Carbone and Tuttle showed that the timing of diurnal maxima was longitude dependent with propagating modes that would smear the diurnal cycle. However, what you have plotted, the reflectivity maxima in the model may not be expected to vary much as long as there is some precipitation given the profile is paramaterised from the existence of rainfall. Frankly, I do not know what to make of the peaks in reflectivity in the observations. These seem very narrow. On what spatial scale are these reflectivity maxima and how has this been averaged

in time? If it is on the 4 km grid they may be too low, depending on how the temporal sampling is done. For the moment, I am not sure what this figure adds and would delete it but keep some of the text noting the diurnal variations (that are discussed in your previous paper?).

The figure of diurnal cycle has been removed, but we keep some discussion as suggested. "As evaluated in Zheng et al. (2019), E3SM v1 failed to simulate the diurnal variation of precipitation over the central United States, where the observed nocturnal peak is greatly underestimated. Xie et al. (2019) improved the diurnal cycle of convection in E3SM v1 recently by modifying convective trigger function in the ZM scheme. It will be interesting to see if the 3D radar reflectivity fields can be better simulated using the updated ZM scheme".

Referee #3: Anonymous

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.

We thank the referee for providing detailed line-by-line comments, which greatly helped to improve the readability of the manuscript.

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.

All NEXRAD data are averaged from 625 samples to 50 samples to match the simulation, then the PDF are generated accordingly. Figure 3 is also updated to account for the problems noted by Referee 4's minor comment 5. We have added "After averaging the NEXRAD pixels at subgrid scale to 50 samples to match the COSP's subcolumns, Fig. 3 compares the simulated subgrid reflectivity distribution to the NEXRAD distribution based on all the GridRad samples combined for the 3-year period at each individual level, where the interval of reflectivity bins is 1 dBZ."

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 upwardpointing 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.
The viewing geometry used in the COSP is not like an upwardpointing radar. As stated in the Section 2.3, the viewing geometry of the COSP mimics that of satellite, which is from space to the ground. Clarifications has been added to "the COSP mimics the satellite view from space to the ground, thus the impact of topography is not an issue as ground-based radars (Smalley et al., 2017). With the downward viewing geometry, the layer below 1-km altitude is most vulnerable to the possible attenuation caused by large precipitation particles, which has been excluded from the comparison."

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.
As clarified in the comment above, the viewing geometry of the COSP is from space to the ground. This said, the attenuation caused by near-surface convective core has no impact on the high-altitude reflectivities. Moreover, we stated in the text "the layer below 1-km altitude is most vulnerable to the possible attenuation caused by large precipitation particles, which has been excluded from the comparison."

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.
The reviewer might be misunderstanding what we did here. What we did was to make the microphysics assumptions used in the COSP to be consistent with those in the cloud microphysics scheme of the host model, which was just about fixing a problem. Fixing the problem indeed helped but the model is still significantly biased in the subgrid distribution, which might contribute by other issues. We have added discussion related to this. "Although the simulated subgrid reflectivity distribution is improved by setting the microphysics assumptions used in COSP consistent with the MG2, the model is still significantly biased. In addition to the intrinsic model-observation differences in the number concentrations and mixing ratios of hydrometeors, there are other possible error sources related to the reflectivity calculation as mentioned in Section 2.2. For example, (1) the mixing ratios are not directly passed from the host model to COSP, instead are converted from the model's precipitation fluxes, (2) the spectral parameters for defining a Gamma

distribution are not consistent from MG2, and (3) the assumptions of subgrid distribution and hydrometeor vertical overlap are simple and not consistent with other parts of the host model. In addition, the subgrid distribution results from COSP are calculated based on the assumption about the distribution of cloud and precipitation among the 50 subcolumns, which is independent of what E3SM uses. Therefore, a higher-order consistency between the COSP and the host model is warranted in future studies."

For the concern of the 13.6 GHz vs. 3 GHz, we performed a series of offline tests of COSP simulation using the frequency of 3 GHz, 13.6 GHz, and 94 GHz. The comparisons of their corresponding reflectivities are shown in Fig. 1. As shown, the reflectivities values with 3 GHz are very similar to those with 13.6 GHz, indicating the Rayleigh scattering is satisfied for both frequencies in this application. Note the particle size simulated by global model at 1-deg scale is known much smaller than the reality, with diameters far smaller than 2.2 cm which leads to the similar reflectivity simulation for any frequency lower than 13.6 GHz.

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.
The discussion of viewing geometry has been added to the Section 2.3.

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."
We agree that large ice particles could definitely occur in the warm season. The text has been modified as suggested.

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.
All the default values are added, and the table is modified as suggested.

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.
The width means the shape parameter in Gamma distribution for describing the dispersion of the distribution. This is clarified as a footnote of the table. A fixed value is used in two-moment microphysics schemes, so here we made it to be consistent with MG2.

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.
The model output is at hourly frequency, but the model time step is 30 min. Therefore the hourly radar reflectivity field represent the average state of the past 30 min, based on which the GridRad data are smoothed to gap the model-observation temporal mismatch.

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)."
Modification has been made as suggested.

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.
Modification has been made accordingly.

P2L31/32: please insert wavelength or frequency after S-band here. The earlier the better.
Modification has been made accordingly.

P2L49: "over the CONUS for the three years (2014-2016)" parentheses are jarring, please remove.
The parentheses has been removed as suggested.

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”.
Modifications has been made as suggested.

P3L76—77: “pressure-based terrain following coordinate” – you could optionally also insert “hybrid sigma” descriptor here to introduce it earlier than Section 2.
Modification has been made accordingly.

P3L89: “…spaceborne satellites…” typo: “satellites” is an adjective here so should be singular.
Correction has been made.

P3L92: “…direct measurements form 3D scanning radars…” typo: I think you mean “…from 3D…”
Correction has been made.

P4L95: “…pseudoobservations using forward calculation” typo: missing article or pluralisation (e.g. “forward calculations”).
Correction has been made.

P5L130: “we nevertheless perform the Gaussian smoothing” typo: I don’t think you need “the” before “Gaussian smoothing”.
Correction has been made.

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”.
It is “variability” that we intended to use. The typo has been corrected.

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…”
Modifications has been made as suggested.

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.
The missing “the” is added. Similar corrections have been made throughout the entire manuscript with careful examination.

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.
Modifications has been made as suggested.

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.
Thank you.

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?
Yes, a “to” has been added.

P8L250: “Different from the model evaluation of”. This seems grammatically weird to me, perhaps “Differently from…”, although I’d probably pick “Compared with…”
We have changed “different” to “differently”, because the two types of evaluation are truly different.

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”.
Thank you for the suggestion, modifications have been made accordingly.

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…”
It is a selected subset. Modifications have been made accordingly.

P9L278—279: “(i.e., those resolved by model resolution).” Is repetitive, how about “those resolved by the model”?
Modifications have been made accordingly.

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”.
Modifications have been made accordingly.

P9L283: “Attempts of accelerating” I think should be “attempts at accelerating”.
Modification has been made accordingly.

P9L287: “only gains 500-800 m increment” missing “a” before “500—800 m”
Modification has been made accordingly.

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.
“The” has been added as suggested.

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”.

Good suggestion. Modifications have been made accordingly.

Referee #4: Anonymous

Summary of paper:
The manuscript presents results of an evaluation of the E3SM model against NEXRAD radar observations for the summer periods during 2014-2016. The authors used the COSP forward simulator package to generate radar reflectivity values from their model cloud fields and averaged both the sub-grid COSP output and the NEXRAD observations to a 1-degree horizontal grid and 1-km vertical grid for like-with-like comparison. The model average reflectivity exceeds the observed value slightly at 2-km height, but at heights above 4 km the model generally does not produce enough cloud above the threshold reflectivity value. Sensitivity testing considering the convection and cumulus parameterisations does not improve this model bias.

Review summary:
This is generally a well-written paper with good quality figures. The evaluation of NWP models against 3D cloud and precipitation observations is of great importance and the present evaluation against NEXRAD is novel. The methodology is incomplete or insufficiently justified in places, which leads to serious concerns about the results. Nevertheless, these concerns might be overcome with appropriate clarifications or revisions and as such publication may be considered after major corrections.
We thank the referee for the accurate summary of our study and the detailed suggestions and comments.

Major comment 1:

The study is hindered by its original objective to evaluated the model against GPM and hence the implementation of the 13.6 GHz frequency in COSP. There are two issues at stake, namely (a) whether the comparison of 13.6 GHz simulated reflectivity against S-band (3-GHz) is appropriate and (b) whether the implementation has been done appropriately.

(a) The authors justify the 13.6 GHz versus 3 GHz comparison by citing their Wang et al. (2019b) study. While that is a nice paper, it is not a sufficiently comprehensive evaluation of the 13.6 GHz reflectivity against the 3 GHz reflectivity to convince the reader that these two are interchangeable. The key figure in that paper (Figure 2) uses normalisation, which removes any excess (or deficit) in cloud detection, which is of importance for this study. The normalisation within cloud, also performed in that Figure, masks the reduction in reflectivity values obtained with GPM (13.6 GHz) both due to attenuation and due to Mie scattering.

Beyond this general unease with the comparison, there are various studies that suggest a necessary conversion from Ku (13.6 GHz) to S band, with different equations used for ice and liquid phases. In particular, recent studies using the GPM radar to calibrate ground-based radars use such conversions:

Warren, R. A., A. Protat, S. T. Siems, H. A. Ramsay, V. Louf, M. J. Manton, and T. A. Kane, 2018: Calibrating Ground-Based Radars against TRMM and GPM. J. Atmos. Oceanic Technol., 35, 323–346, https://doi.org/10.1175/JTECH-D-17-0128.1.

To answer the first question "whether the comparison of 13.6 GHz simulated reflectivity against S-band (3-GHz) is appropriate?", we performed a series of offline tests of COSP simulation using the frequency of 3 GHz, 13.6 GHz, and 94 GHz. The comparisons of their corresponding reflectivities are shown in Fig. 1. As shown, the reflectivity values with 3 GHz are very similar to those with 13.6 GHz, indicating the Rayleigh scattering is satisfied for both frequencies in this application. Note the particle size simulated by global models at 1° scale is known much smaller than the reality, with diameters far smaller than 2.2 cm which leads to the similar reflectivity simulation for any frequency lower than 13.6 GHz.

Apparently, reflectivities at a frequency much higher than 13.6 GHz (such as 94 GHz, red cycles) would be a concern as mentioned by the reviewer.

For the attenuation, it would not cause a significant concern here as well, because 1) the model does simulate large particles that are enough to cause Mie scattering, and 2) the viewing geometry in COSP (from space to the ground) greatly relieve the attenuation from precipitation.

We agree with the reviewer that in real observation the conversion from Ku to S band is necessary. The reference listed is cited. In this application, it does not affect our results.

Please referrer to our response to the major comment 3 on the concern of "cloud detection".

We have rewritten the paragraph explaining the choice of 13.6 GHz, i.e., "The GPM radar frequency is higher than the NEXRAD (13.6 GHz vs. 3 GHz). Previous studies have shown conversions from Ku (13.6 GHz) to S band (3 GHz) are necessary when using GPM Ku band radar to calibrate the ground-based radars (Warren et al., 2018). Based on our previous study that quantitatively evaluated the coincident observations from NEXRAD and GPM over the CONUS, we found the 3D radar reflectivity fields obtained from the two independent platforms are highly consistent with each other after proper smoothing of GPM data in the vertical (Wang et al., 2019b). We performed a series of offline tests of COSP simulation using the frequency of 3 GHz (NEXRAD), 13.6 GHz (GPM Ku band), and 94 GHz (the cloud profiling radar onboard of the CloudSat satellite). Their corresponding reflectivities are compared in Fig. 1. As shown, the reflectivity values with 3 GHz are very similar to those with 13.6 GHz, indicating the Rayleigh scattering is satisfied for both frequencies in this application. Note the particle size simulated by global models at 1° scale is known much smaller than the reality (Marchand et al., 2009) with diameters far less than 2.2 cm (the wavelength of 13.6 GHz), which leads to the similar reflectivity simulation for any frequency lower than 13.6 GHz. To examine if the COSP can correctly handle the Mie scattering calculation, the frequency of 94 GHz used by the CloudSat is also tested, whose products have been widely used for the evaluation of coarse-resolution models (Zhang et al., 2010). As shown in Fig. 1, the reflectivities simulated with 94 GHz significantly deviate from those simulated with 3 GHz and 13.6 GHz when reflectivities > 10 dBZ, which reveals that the COSP simulator is capable of handling both Rayleigh and Mie scattering calculations. However, there is no difference using Ku band or S band in the COSP simulator in this study, because the simulated particles are too small to cause Mie scattering at these radar frequencies. An attenuation correction has been applied in case of existence of any large particles although they are extremely unlikely to occur in this application. Since the COSP mimics the satellite view from space to the ground, the layer below 1-km altitude is most vulnerable to the possible attenuation caused by large precipitation particles, which has been excluded from the comparison."

(b) It is not obvious that the implementation of 13.6 GHz in COSP is straightforward. The authors state that the simulator automatically uses Rayleigh scattering, but that cannot be appropriate under all circumstances, particularly if the focus is on convection. Attenuation will not only be significant below 1-km altitude: convective towers can cause attenuation in the ice phase as well. Similarly, the large hydrometeors found aloft may lead to Mie scattering. That the Rayleigh scattering assumption is inappropriate for the GPM PR has long been established in the literature, e.g.:
L'Ecuyer, T. S., and G. L. Stephens, 2002: An Estimation-Based Precipitation Retrieval Algorithm for Attenuating Radars. J. Appl. Meteor., 41, 272–285, https://doi.org/10.1175/1520-0450(2002)041<0272:AEBPRA>2.0.CO;2.
We agree that in reality, convective towers can easily bring large ice particles aloft that lead to Mie scattering. However, in the coarse-resolution climate model, no such large particles are simulated.

Having said that, it is entirely possible that the 13.6 GHz is a red herring here. If Rayleigh scattering is assumed, and no Mie scattering is included for large particles, the COSP calculation might as well be considered as if it were a 3 GHz radar. In that case, it is worth checking the COSP calculations for whether the frequency/wavelength matters.
The 3 GHz and 13.6 GHz simulations have been checked, which are consistent.

The authors have at least two options here. Either the authors provide corrected calculations following (for example) the papers above, for instance by applying such corrections to the COSP-simulated reflectivity. Alternatively, the authors develop a standalone forward simulator. If the latter, it is reasonable to assume Rayleigh reflectivity at 3 GHz (S-band) for comparison against the NEXRAD observations. Given the model microphysics assumptions (as listed in Table 1) it is relatively straightforward to calculate the Rayleigh reflectivity from the model ice and liquid water contents. This would be the most appropriate way to compare the model to the NEXRAD observations, but obviously requires some additional data processing, which may be difficult if the original cloud 3D fields were not included in the output.
Since the two frequencies are proven to give the almost identical results in the model, we do not need to go with either of the options here.

Major comment 2:

The lack of sufficiently high radar reflectivity aloft is concerning and while this could be a model bias, it would be helpful for the reader to have more information regarding the COSP calculations. In particular, in Table 1 the authors specify the density of ice and the distribution width. Following Morrison and Gettelman (2008), the remaining size distribution parameters lambda and N0 should be calculated from the mixing ratios directly. The COSP calculation will require the (consant) density of ice and distribution width as well as the (variable) lambda and N0, unless COSP has the appropriate information to calculate lambda and N0 itself from the mixing ratio. If COSP is not provided with the correct information, a constant lambda and N0 may be assumed by COSP, leading to erroneous calculations.
Thanks for the good point. The lambda for hydrometeor size distribution in COSP is derived only from the mass mixing ratio, but the intercept parameter N0 is fixed in COSP. This is not consistent with MG2 yet. We have clarified this in the section describing COSP and also added discussion about the uncertainties from the simulator to the end of Section 3.1, that is, "although the simulated

subgrid reflectivity distribution is improved by setting the microphysics assumptions used in COSP consistent with the MG2, the model is still significantly biased. In addition to the intrinsic model-observation differences in the number concentrations and mixing ratios of hydrometeors, there are other possible error sources related to the reflectivity calculation as mentioned in Section 2.2. For example, (1) the mixing ratios are not directly passed from the host model to COSP, instead are converted from the model's precipitation fluxes, (2) the spectral parameters for defining a Gamma distribution are not consistent from MG2, and (3) the assumptions of subgrid distribution and hydrometeor vertical overlap are simple and not consistent with other parts of the host model. In addition, the subgrid distribution results from COSP are calculated based on the assumption about the distribution of cloud and precipitation among the 50 subcolumns, which is independent of what E3SM uses. Therefore, a higher-order consistency between the COSP and the host model is warranted in future studies."

On a related note, if the E3SM has been evaluated against CloudSat and/or CALIPSO, that would provide helpful context to include about its ability to produce high-level cloud.
While not compared with CloudSat/CALIPSO, the E3SM's high cloud fraction has been thoroughly evaluated using MODIS data product, where the model with the same configuration as this study agree well with the observation.
Clarification has been added. "Differently from the model evaluation of cloud top height and high cloud fraction, where EAMv1 has shown good agreements with satellite observations over the CONUS, evaluation of radar echo top height indicates whether the processes internal to the cloud are producing precipitation correctly."

Major comment 3:

It is not clearly justified why the authors averaged their data to the 1-degree grid scale, when most of the information on the sub-grid scale is available to them. Averaging to 1-degree comes with its own problems (e.g. how to treat "cloud-free" regions) that may end up masking model deficiencies and it may have led to the disappearance of the characteristic CFAD shape in the NEXRAD analysis. Perhaps in Section 3.1, once the authors have performed their analysis using the sub-grid information, the authors could include some justification as to why the following analysis is done on the 1-degree averaged data.
We agree that averaging to 1-deg comes with its own problems. However, 1 degree is the model's native grid spacing and it is important to evaluate how model performs at its resolution. In addition, the subcolumns are not what the model used in the simulations. It is just something that the simulator COSP independently assumes, which does not reflect how model performs at the subgrid scale.
We have added the justification at the end of section 3.1, "in addition, the subgrid distribution results from COSP are calculated based on the assumption about the distribution of cloud and precipitation among the 50 subcolumns, which is independent of what E3SM uses. Therefore, a higher-order consistency between the COSP and the host model is warranted in future studies. In this following analysis, we focus on the evaluation of the simulated 3D radar reflectivity field at the model's native grid, which is 1-degree, since the subgrid information from COSP does not directly reflect how E3SM does it. Also, the convective cloud fraction in the ZM cumulus parameterization used in E3SM is fitted to be a function of cloud mass flux and should be viewed as a tunable

parameter, whose evaluation is not very meaningful unless it becomes an independent variable, for instance, for grey-zone resolutions."

Minor comments:

1. Line 52-53: It is important to acknowledge in the introduction that these "convective processes" are not resolved by the model and that some sub-grid representation is needed. The evaluation can then be performed either on coarsened observations (as this study does) or on the sub-grid sampled model, as COSP does. Please include such a clarification in the introduction.
The information that convective processes are parameterized is already described in the E3SM model description section. The model does not include a sub-column sampler for subgrid clouds so it is difficult to evaluate the subgrid clouds of the model. COSP is just a diagnostic package and the subgrid information assumed in the COSP does not reflect how E3SM does it.

2. Line 53-55: It is not obvious that these scattering effects are such an issue at S-band. Iguchi et al. (2018) consider the GPM-DPR which are smaller wavelength. At S-band, Rayleigh scattering could potentially be assumed which would make forward simulation much easier (easier than considering the sub-grid sampling for convection). Please rephrase this statement and consider studies using S-band radars specifically. [NB It should be noted that these lines are likely a result of the authors' original intent to evaluate the model against GPM PR observations.]
See our response to the major comment 1.

3. Line 69-72 and Line 98-99: It should be made clear to the reader that there is no difference in microphysics parameters between convective and stratiform (referring to Table 1). Some further clarification is then required regarding the sub-grid partition. Presumably, the model diagnoses a convective cloud fraction and a stratiform cloud fraction, with their respective water contents. These water contents may differ and therefore could lead to different simulated radar reflectivity. This is important information to include, so perhaps in Line 69-72 explain the convective-stratiform partition and in Line 98-99 clarify the typical differences between convective and stratiform water contents (noting that the microphysics parameters are the same).
The clarification of stratiform-convective partition in EAMv1 is added, "regarding the stratiform-convection partition, the MG2 stratiform cloud microphysics and CLUBB higher-order turbulence parameterization explicitly provide values for condensate mass and number, as well as an estimate of stratiform cloud fraction, whereas the convective cloud fraction is not parameterized in the ZM scheme, and is diagnosed from cloud mass flux for cloud radiation calculation, which is treated as a tunable parameter."
However, those mixing ratios of different hydrometeor types at sub-grid scale are not directly input into the COSP. Instead, the COSP converts model-simulated precipitation fluxes to mixing ratios at sub-grid scale, which is inconsistent with the host model and could lead to different radar reflectivity simulated. The clarification has been added, "note the COSP does not use the hydrometeor mixing ratios from the host model to construct the particle size distribution (PSD) and then to calculate the radar reflectivity. Instead, it converts the model-simulated precipitation fluxes into mixing ratios before calling the radar simulation."

4. Line 106-108: The adoption of model-specific parameters is not unique and is a widely used approach when implementing COSP (or when developing their own forward simulator). Perhaps rephrase: "Following general usage of COSP, we modified the microphysics assumptions…".

The Swales et al. (2018) paper explicitly mentions the need to "maintain consistency between COSP1 and the host model."
Modifications have been made as suggested.

5. Section 3.1: As stated above, the "out of the box" configuration of COSP is not advised and general use should always assume the model parameters. As such, it is recommended to remove the left-hand panels in Figure 2, as well as the standalone Figure 3, and focus instead on the differences between the model and observations from the right-hand panels. More specifically, in Figure 2: (1) Why does the x-axis start at 14 dBZ, when an 8 dBZ minimum reflectivity is considered? (2) What are the units of density? Per 2 dB (i.e. 2 dB bins)? (3) Why show these PDFs normalised? It should be important to note the absence of "cloud" above 8 dBZ as well. The authors should include a separate Figure showing the fraction of occurrence of Z>8 dBZ with height to compare this between model and observations.
We prefer to keep the left-hand panels in Figure 3, as well as Figure 4. Although Swales et al. (2018) explicitly mentioned the need to maintain the consistency between the COSP and the host model, the impact hasn't been quantified. Figures 3 and 4 give a clear demonstration of the consequence. In addition, for the E3SM model community, it is good to show the problem.

For Figure 3, the minimum reflectivity should be 8 dBZ. The 14 dBZ was for the comparison with the GPM, which has been corrected. The figure has been modified and the interval of reflectivity bins is 1 dBZ.

Please note this is the subgrid distribution within 1° model grid elements, we use this comparison to explore the how the simulated subgrid reflectivity distribution from COSP differs from the observation. Following the referee's suggestion, we have added a figure showing the fraction of occurrence of Z ≥ 8 dBZ with height (Fig. 6).

The related discussion has been added, "in addition to the mean values, the histograms of observed and simulated radar reflectivities are compared in for different altitudes, where the interval of reflectivity bins is 2 dBZ (Fig. 6). By comparing the occurrence of Z ≥ 8 dBZ between model and observations, the model apparently has narrower distribution than the observations, and the model-observation deviation in maximum values increases with height. At 8 km and below, the model generally overestimates the sample sizes of smaller reflectivity values but lacks extreme high reflectivity values. However, at 11-km altitude, the model greatly underestimates the sample sizes of the entire reflectivity spectrum compared to the observation, causing the severe underestimation in the mean value."

6. Section 3.2 and Figure 4: How is the mean calculated? In Section 2, we learn that for the instantaneous observation/simulated output, the mean is calculated in linear Z units (so that cloud-free areas are 0) and then converted to dBZ, with an 8 dBZ threshold. But how are values below 8 dBZ considered when calculating these long-term means? Or are the means (and standard deviation and 95th percentile) in-cloud only? In either case, it is useful to understand the occurrence of Z>8dBZ, so please add this to Table 2 and as a separate set of maps to complement Figure 4. The occurrence could help explain the difference in the mean, as the model could compensate for missing higher values by having a higher "cloud" occurrence.
We discarded all the instantaneous grid boxes with Z < 8 dBZ. So, yes, the means, standard deviations, and 95th percentile values are all in-cloud samples only.

The sample numbers have been added to Table 2 as suggested, and we have added an additional figure for the sample numbers (Figure 6) as mentioned above.

7. Section 3.3 and Figure 5: Again, normalization occurs in-cloud, so information is lost on the frequency of occurrence of cloud more generally. Could the authors comment on the difference in characteristic shape of the CFAD between NEXRAD and the model? The model follows the well-known shape with a maximum occurrence at dBZ that increases from cloud top to freezing level, and then slowly decreases or stays constant below the freezing level. That shape can be reproduced with NEXRAD, but it seems to have disappeared in the authors' analysis – is that solely due to the averaging to 1 degree? Perhaps the choice of Z for cloud-free regions is important here?

As shown in Figure 6, the frequency of occurrence of cloud with Z ≥ 8 dBZ are compared between the model and the observations.

We have added comment on the difference in characteristic shape of CFAD. "Regarding the overall shape of CFADs, the model follows the well-known pattern where the reflectivity value range of high frequency zone (> 3.2%) increases from cloud top to the freezing level, and then slowly decreases or remains constant below the freezing level. The cores of maximum frequency (> 5%) are located in the centres of the high frequency zones. However, these characteristics are not presented in the observations, whose high frequency zones are greatly skewed to the lower reflectivity values. The characteristics of NEXRAD's CFADs could be due to averaging from fine resolution (4 km) to coarse resolution (1°), as well as averaging of convective and stratiform components because the two components produce significantly different reflectivity profiles and magnitudes."

The choice of the threshold does not affect the shape of CFADs. In previous round of discussion, we have enlarged the sample size by using lower threshold of 0 dBZ, where the presentation of CFADs is not affected as shown in Figure R1.

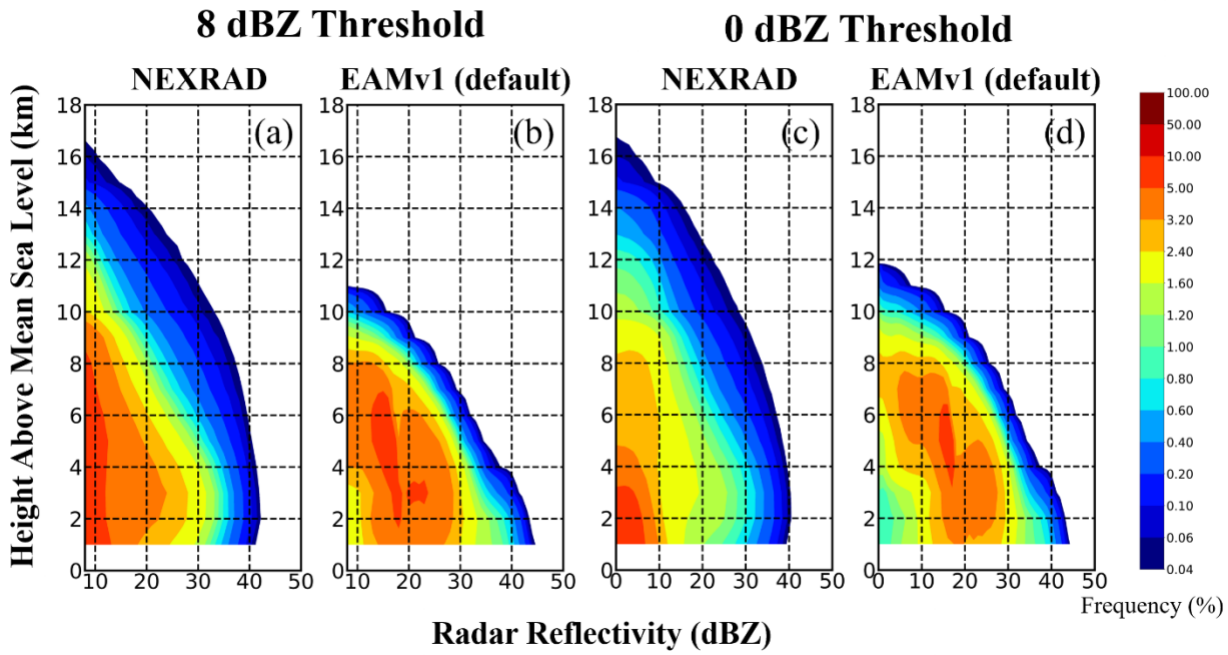

Figure R1. The sensitivity test of changing minimum reflectivity threshold from 8 dBZ (a, b) to 0 dBZ (c, d).

8. Line 219: "Above 11km, the model completely fails to simulate any reflectivity". There is some nuance here, as the authors use the 8-dBZ threshold. So: "Above 11km, the model fails to generate average reflectivity above 8 dBZ." Assuming that the authors have access to the data, it would be useful to report the typical reflectivity values that are generated by the model at these altitudes, even if below 8 dBZ.

Modifications have been made as suggested. The reflectivity values simulated at 12 km is shown below. All the model data and observational data less than 0 dBZ were truncated during data processing. The typical reflectivity value would be 0-2 dBZ. We have added this information to the manuscript, "above 11 km, the model fails to generate average reflectivity above 8 dBZ, and the typical reflectivity value is between 0 and 2 dBZ at 12 km".

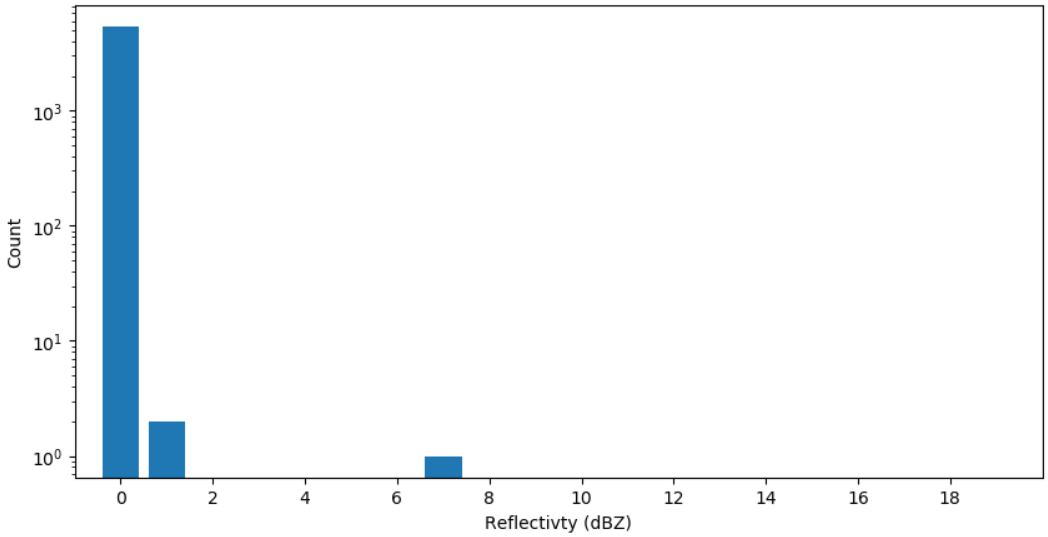

Figure R2. The histogram of simulated radar reflectivity at 12-km altitude.

9. Line 238-240 and Figure 6: It is unclear what is actually being considered here. Is column-maximum reflectivity the maximum in a column on the 1-degree grid? What is then the "radar reflectivity" at a given "local time" in Figure 6? Is this an average over the entire CONUS, but only for grid boxes with this value above 8 dBZ? Or is it the maximum over the entire CONUS? The way this is calculated might partly explain the signals that appear, so all this needs to be clarified in the text.

This figure has been deleted.

10. Section 3.4 and Figure 7: As above, a general understanding of frequency of occurrence of Z>8dBZ would be useful in addition to these (normalised) diagrams.

The sample sizes have been added to Table 2.

11. Line 301-305: This conclusion should be removed, as should be the related Figures and discussion, as it is widely established that the microphysics assumptions of the forward simulator should be consistent with those in the host model (e.g. Swales et al., 2018).

The figures and associated discussion help quantify the impact of inconsistent microphysics between the COSP and the host model. We prefer to keep these results.

Using Radar Observations to Evaluate 3D Radar Echo Structure Simulated by the Global Model E3SM Version 1

Jingyu Wang[1], Jiwen Fan[1,*], Robert A. Houze Jr[2], Stella R. Brodzik[2], Kai Zhang[1], Guang J. Zhang[3], and Po-Lun Ma[1]

[1]Pacific Northwest National Laboratory, Richland, WA 99354, USA
[2]University of Washington, Seattle, WA 98195, USA
[3]Scripps Institution of Oceanography, La Jolla, CA 92093, USA

Correspondence to: Jiwen Fan (jiwen.fan@pnnl.gov)

Abstract. The Energy Exascale Earth System Model (E3SM) developed by the Department of Energy has a goal of addressing challenges in understanding the global water cycle. Success depends on correct simulation of cloud and precipitation elements. However, lack of appropriate evaluation metrics has hindered the accurate representation of these elements in general circulation models. We derive metrics from the three-dimensional data of the ground-based Next generation radar (NEXRAD) network over the U.S. to evaluate both horizontal and vertical structures of precipitation elements. We coarsened the resolution of the radar observations to be consistent with the model resolution and improved the coupling of the Cloud Feedback Model Intercomparison Project Observation Simulator Package (COSP) and E3SM Atmospheric Model Version 1 (EAMv1) to obtain the best possible model output for comparison with the observations. Three warm seasons (2014-2016) of EAMv1 simulations of 3D radar reflectivity features at an hourly scale are evaluated. A general agreement in domain-mean radar reflectivity intensity is found between EAMv1 and NEXRAD below 4 km altitude; however, the model underestimates reflectivity over the central United States, which suggests that the model does not capture the mesoscale convective systems that produce much of precipitation in that region. The shape of the model estimated histogram of subgrid scale reflectivity is improved by correcting the microphysical assumptions in COSP. The model severely underestimates radar reflectivity at upper levels—the simulated echo top height is about 5 km lower than in observations—and this result is not changed by tuning any single physics parameter.

1 Introduction

Clouds and precipitation play a major role in Earth's budgets of energy, water, and momentum. However, the correct simulation of 3D structures of clouds and precipitation has been challenging in general circulation models (GCMs) (Trenberth et al., 2007; Randall et al., 2007; Eden and Widmann, 2012), partially because model grid spacings generally do not adequately resolve the cloud-structure details important to these budgets. In addition, the lack of appropriate evaluation metrics also hinders the evaluation of GCMs. Over the contiguous U.S. (CONUS), the detailed 3D radar reflectivity field (indicating the

3D distribution of precipitation particles) is observed by the ground-based Next Generation Radar (NEXRAD) network of S-band weather radars (3 GHz; Zhang et al., 2011 and 2015). In this study, we use the mosaic of NEXRAD observations called Gridded Radar Data (GridRad) developed by Homeyer and Bowman (2017), which have a horizontal resolution of 0.02° (regridded to 4 km in this study), vertical resolution of 1 km (24 levels), and an update cycle of 1 hour. In order to compare these data appropriately with output of the global model used here, we further coarsen the horizontal resolution, as described in Section 2.

The Energy Exascale Earth System Model (E3SM) is an ongoing effort of the Department of Energy (DOE) to advance the next-generation of climate modeling (Bader et al., 2014). Version 1 of E3SM Atmosphere Model (EAMv1) is a descendent of the National Center for Atmospheric Research (NCAR) Community Atmosphere Model version 5.3 (CAM5.3; Neale et al., 2012). However, it has evolved substantially in coding, performance, resolution, physical processes, testing and development procedures (Rasch et al., 2019). Previous model evaluation has focused on the long-term climatological properties of certain cloud fields, surface precipitation, and water conservation on the global scale (e.g., Qian et al., 2018; Xie et al., 2018; Zhang et al., 2018; Lin et al., 2019). Evaluations of the vertical structures of cloud and precipitation elements have used vertically pointing radar observations obtained during field campaigns (Zhang et al., 2018; Zhang et al., 2019). However, these tests lacked evaluation of fully 3D cloud and precipitation structure over large regions of the globe and over long time periods.

For this study, we have built data processing techniques to evaluate EAMv1 simulation of the 3D radar reflectivity field at its default setting of 1° grid spacing and 72 vertical layers at an hourly time scale. Our goal is to provide a comprehensive evaluation of both horizontal pattern and vertical structure of cloud and precipitation. We use radar observations obtained from the NEXRAD over the CONUS for the three years 2014-2016. In order to directly compare the model results with NEXRAD, we have implemented and improved the Cloud Feedback Model Intercomparison Project (CFMIP) Observation Simulator Package (COSP) (Bodas-Salcedo, et al., 2011) into EAMv1. We restrict the evaluation to the warm season (i.e., April to September). Over the CONUS, warm season precipitation is dominated by convective processes, which are very different from the more widespread frontal cloud systems of cold-season precipitation. As discussed by Iguchi et al. (2018), precipitating ice particles have 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.

This paper is organized as follows: Section 2 describes the model, the GridRad dataset, the COSP simulator, and the step-by-step methodology of data processing to account for differences between the modelled and observed datasets, specifically (1) horizontal and vertical resolutions of EAMv1 (1°, 72 vertical levels) and NEXRAD (4 km horizontally, 1 km vertically) and (2) minimum detectable limits between the model and NEXRAD. Section 3 presents the model evaluation results and tests of the sensitivity to physics parameters. Section 4 provides synthesis and conclusions.

2 Methodology

2.1 EAMv1 Description and Configuration

EAMv1's dynamics core and physics parameterizations are described in detail by Rasch et al. (2019). The continuous Galerkin spectral finite element method solves the primitive equations on a cubed-sphere grid (Dennis et al., 2012; Taylor & Fournier, 2010). Tracer transport on the cubed sphere is handled using a variant of the semi-Lagrangian vertical coordinate system of Lin (2004). The method locally conserves air mass, trace constituent mass, and moist total energy (Taylor, 2011). Turbulence, shallow cumulus clouds, and cloud macrophysics are parameterized with the Cloud Layers Unified By Binormals (CLUBB) parameterization (Golaz et al., 2002; Larson, 2017). Deep convection is based upon the formulation originally described in Zhang and McFarlane (1995, hereafter ZM), with modifications by Neale et al. (2008) and Richter and Rasch (2008). Stratiform clouds are represented with the "Morrison and Gettelman version 2" (MG2) two-moment bulk microphysics parameterization (Gettelman and Morrison, 2015). Aerosol microphysics and interactions with stratiform clouds are treated with an updated and improved version of the four-mode version of the Modal Aerosol Module (MAM4; Liu et al., 2016). Regarding the stratiform-convection partition, the MG2 stratiform cloud microphysics and CLUBB higher-order turbulence parameterization explicitly provide values for condensate mass and number, as well as an estimate of stratiform cloud fraction, whereas the convective cloud fraction is not parameterized in mass flux-based ZM scheme (assumed to be <<1 for typical GCM resolutions such as at 1-degree grid spacing or coarser), and is diagnosed from cloud mass flux for cloud radiation calculation, which is treated as a tunable parameter.

The EAMv1 used in this study has 30 spectral elements (ne30), which corresponds to approximately 1° horizontal grid spacing, and the total number of grid columns is 48,602. Vertically, there are 72 layers using a traditional hybridized sigma pressure coordinate. The simulation is run for the time period from 1 January 2014 to 1 October 2016. We use a dynamic timestep of 5 min and a cloud microphysics timestep of 30 min. The large-scale circulation in the simulation is constrained using the nudging technique (Zhang et al., 2014; Ma et al., 2015; Lin et al., 2016), so that the model simulations can be constrained by realistic large-scale forcing. Specifically, horizontal winds (U, V components) are nudged towards the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA2) reanalysis data (Gelaro, et al., 2017) with a relaxation time scale of 6 hours. Nudging is applied to all grid boxes at each time step, with the nudging tendency calculated using the model state and the linearly-interpolated MERRA2 data (Sun et al., 2019).

To facilitate the comparison with observations, model outputs are regridded to the geographic coordinate system with a horizontal grid spacing of 100 km, and the vertical coordinate is converted to the above mean surface level height in meters. By default, all the regridding processes in this study are based on the Earth System Modeling Framework (ESMF) Python Regridding Interface (https://www.earthsystemcog.org/projects/esmpy/) using bilinear interpolation.

2.2 COSP Radar Simulator

The retrieved spaceborne satellite and ground-based radar products such as cloud water content, and effective particle size (e.g., Randel et al., 1996; Wang et al., 2015; Tian et al., 2016; Um et al., 2018) are often treated as the ground-truth for model evaluation (e.g., Fan et al., 2017; Han et al., 2019). However, the retrieved products often have large uncertainty (Stephens and Kummerow, 2007). To allow the comparison of model results with direct measurements from 3D scanning radars (ground-based or satellite-borne), the CFMIP Observation Simulator Package (COSP) was developed for use in GCMs (Bodas-Salcedo et al., 2011). Instead of using retrieved products to evaluate the model simulation, COSP converts model output into pseudo-observations using forward calculations (Bodas-Salcedo et al., 2011; Swales et al., 2018; Zhang et al., 2010).

The COSP consists of three steps, as detailed in Zhang et al. (2010). The first step is to generate a subgrid-scale distribution of cloud and precipitation, which is done by using the Subgrid Cloud Overlap Profile Sampler (SCOPS; Klein and Jakob, 1999; Webb et al., 2001) and SCOPS for precipitation (SCOPS_PREC), respectively. Each GCM grid box is divided into 50 subcolumns in this study. Detailed description of SCOPS and SCOPS_PREC can be found in Zhang et al. (2010). Then, the radar signals are calculated by the QuickBeam code (Haynes and Stephens, 2007) using the column distribution of cloud and precipitation. Note the COSP does not use the hydrometeor mixing ratios from the host model to construct the particle size distribution (PSD) and then to calculate the radar reflectivity. Instead, it converts the model-simulated precipitation fluxes into mixing ratios before calling the radar simulation. Finally, the grid box mean radar reflectivity is calculated through the method of linear averaging (i.e., the reflectivity values [in dBZ] are converted to the Z values [$mm^6\ m^{-3}$] to calculate the mean Z, then mean Z is converted back to the dBZ). In addition to averaging, all the processing of radar reflectivity data from model and NEXRAD in this study utilizes the linearized Z values, including horizontal averaging, vertical interpolation, calculation and comparison of mean values, etc.

The COSP version 1.4 used in this study has no scientific difference from version 2.0 (Song et al., 2018, Swales et al., 2018). Following the general usage of COSP, we modified the microphysics assumptions used for the radar reflectivity calculation regarding hydrometeor density, size distribution, etc., making those assumptions consistent with those used in the MG2 cloud microphysics scheme that is used in E3SM. The detailed documentation of those changes is in Table 1. Note that, although we tried to make the COSP use the same hydrometeor size distribution functions as MG2, the three parameters (slope, intercept, and shape parameters) are still separately defined in COSP. We use horizontally homogeneous cloud condensate distribution within the model grid element, and maximum-random overlapping scheme for cloud occurrence (Hillman et al., 2018).

2.3 NEXRAD Observations

The NEXRAD network consists of 159 S-band (3 GHz) Doppler radars, which form a dense observational network nearly covering the CONUS. We use the GridRad mosaic product of Homeyer and Bowman (2017), which combines all NEXRAD radar data covering the region 155°W – 69°W, 25°N – 49°N. To compare the GridRad data to the E3SM model fields, the radar frequency in the COSP was set to 13.6 GHz, consistent with the Global Precipitation Measurement (GPM) Ku-band

radar, since we originally aimed at evaluating the E3SM simulation with GPM data. However, due to the high detectable threshold of 13 dBZ, low sampling frequency (4-7 overpasses over CONUS per day), and the narrow swath width (245 km) for each overpass, GPM data within the three-year period (2014-2016) have a significant under-sampling issue. That is, the GPM sample sizes over 1° model grid boxes are generally too small to robustly represent the grid element mean value. Therefore, we decided not to use GPM data in this study. As GPM operates over the whole earth and is anticipated to run for a long-time period, it will likely be a very useful dataset to evaluate the coarse-resolution global model in the future.

The GPM radar frequency is higher than the NEXRAD (13.6 GHz vs. 3 GHz). Previous studies have shown conversions from Ku (13.6 GHz) to S band (3 GHz) are necessary when using GPM Ku band radar to calibrate the ground-based radars (Warren et al., 2018). Based on our previous study that quantitatively evaluated the coincident observations from NEXRAD and GPM over the CONUS, we found the 3D radar reflectivity fields obtained from the two independent platforms are highly consistent with each other after proper smoothing of GPM data in the vertical (Wang et al., 2019b). We performed a series of offline tests of COSP simulation using the frequency of 3 GHz (NEXRAD), 13.6 GHz (GPM Ku band), and 94 GHz (the cloud profiling radar onboard of the CloudSat satellite). Their corresponding reflectivities are compared in Fig. 1. As shown, the reflectivity values with 3 GHz are very similar to those with 13.6 GHz, indicating the Rayleigh scattering is satisfied for both frequencies in this application. Note the particle size simulated by global models at 1° scale is known much smaller than the reality (Marchand et al., 2009) with diameters far less than 2.2 cm (the wavelength of 13.6 GHz), which leads to the similar reflectivity simulation for any frequency lower than 13.6 GHz. To examine if the COSP can correctly handle the Mie scattering calculation, the frequency of 94 GHz used by the CloudSat is also tested, whose products have been widely used for the evaluation of coarse-resolution models (Zhang et al., 2010). As shown in Fig. 1, the reflectivities simulated with 94 GHz significantly deviate from those simulated with 3 GHz and 13.6 GHz when reflectivities > 10 dBZ, which reveals that the COSP simulator is capable of handling both Rayleigh and Mie scattering calculations. However, there is no difference using Ku band or S band in the COSP simulator in this study, because the simulated particles are too small to cause Mie scattering at these radar frequencies. An attenuation correction has been applied in case of existence of any large particles although they are extremely unlikely to occur in this application. Since the COSP mimics the satellite view from space to the ground, the layer below 1-km altitude is most vulnerable to the possible attenuation caused by large precipitation particles, which has been excluded from the comparison. In this study, biases caused by the temporal mismatch are minimal at the horizontal resolution of 1° (~100 km), we nevertheless perform Gaussian smoothing of GridRad data to match the model time step (30 min) in the comparison.

2.4 Mapping the Radar Observations to the Model Grid

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. 2a). However, the model grid-mean reflectivity can be as low as -100 dBZ. Because our focus is on significantly precipitating clouds, the minimum threshold of reflectivity at 1° grid scale is set to be 8 dBZ (corresponding to rain rate $\geq$ 0.1 mm hr^{-1}). 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. 2b) and in the simulation (Fig. 2c). In the vertical direction, the EAMv1-simulated radar reflectivity field (72 vertical levels, hybrid coordinate) is interpolated to the levels of GridRad (vertical resolution of 1 km). The simulation data are saved hourly, consistent with the hourly GridRad data.

3 Results

After the horizontal averaging, vertical interpolation, and truncation at the identified minimum threshold of 8 dBZ, the 3D radar reflectivity fields obtained from GridRad and the model simulation become comparable. The EAMv1 simulated reflectivity is evaluated from the perspectives of subgrid distribution, horizontal pattern, and vertical distribution.

3.1 Comparison on Subgrid Distribution of Reflectivity

The horizontal resolution difference between GCMs (~100 km) and NEXRAD observations (4 km) presents a challenge for testing the model simulated radar reflectivity. To mimic the observations, COSP divides the grid-mean cloud and precipitation properties into subcolumns (Pincus et al., 2006) that statistically downscale the data in a way that should be consistent with observations. The way this is done in COSP is discussed by Zhang et al. (2010) and Hillman et al. (2018). In this section we examine whether the subgrid reflectivity distribution generated by COSP is consistent with the observed subgrid reflectivity distribution shown by the NEXRAD observations.

In EAMv1, 50 subcolumns are used for calculating the mean radar reflectivity for a model grid box. There are 625 pixels inside each 1° grid for NEXRAD data to provide a probability density function (PDF) of observed reflectivity within the box. After averaging the NEXRAD pixels at subgrid scale to 50 samples to match the COSP's subcolumns, Fig. 3 compares the simulated subgrid reflectivity PDF to the NEXRAD PDF based on all the GridRad samples combined for the 3-year period at each individual level, where the interval of reflectivity bins is 1 dBZ. The results for the default microphysics assumptions in COSP, which are for a single-moment scheme, produce a bi-modal distribution at and below 8-km altitudes (blue histograms in the left-hand column of Fig. 3). The bimodality is significantly different from the observed PDF, which forms a smooth gamma distribution. Song et al. (2018) also found bimodal distributions when the COSP was implemented in the CAM with the original microphysics assumptions, which are clearly unlike real observed radar reflectivity distributions.

Our modification of the microphysical assumptions in COSP (right-hand column of Fig. 3) greatly reduces the bimodality. In addition, the modified microphysical assumptions produce higher values of reflectivity, in better agreement with observations, and the grid-mean radar reflectivities increase by ~4 dBZ (Fig. 4) mainly for values less than 25 dBZ. The improvement in the subgrid distribution and grid-mean reflectivity brought by the change of microphysics assumptions indicates the necessity of microphysical consistency between the COSP and the host model. It should be noted that the simulated radar reflectivity and its subgrid distribution are sensitive to the overlap assumption and the distribution function of condensates that are set in COSP

(Hillman et al., 2018). Our results are from the default setup of these aspects of COSP. It is not the purpose of this study to test those assumptions.

Although the simulated subgrid reflectivity distribution is improved by setting the microphysics assumptions used in COSP consistent with the MG2, the model is still significantly biased. In addition to the intrinsic model-observation differences in the number concentrations and mixing ratios of hydrometeors, there are other possible error sources related to the reflectivity calculation as mentioned in Section 2.2. For example, (1) the mixing ratios are not directly passed from the host model to COSP, instead are converted from the model's precipitation fluxes, (2) the spectral parameters for defining a Gamma distribution are not consistent from MG2, and (3) the assumptions of subgrid distribution and hydrometeor vertical overlap are simple and not consistent with other parts of the host model. In addition, the subgrid distribution results from COSP are calculated based on the assumption about the distribution of cloud and precipitation among the 50 subcolumns, which is independent of what E3SM uses. Therefore, a higher-order consistency between the COSP and the host model is warranted in future studies.

In this following analysis, we focus on the evaluation of the simulated 3D radar reflectivity field at the model's native grid, which is 1°, since the subgrid information from COSP does not directly reflect how E3SM does it. Also, the convective cloud fraction is not parameterized in mass flux-based ZM scheme and is diagnosed from cloud mass flux for cloud radiation calculation, which is treated as a tunable parameter, whose evaluation is not very meaningful unless it becomes an independent variable, for instance, for grey-zone resolutions.

3.2 Comparison of Horizontal Patterns

Now we compare the temporal mean reflectivity through the entire study period between the NEXRAD observation (Figs. 5a, d, g and j) and EAMv1 simulation (Figs. 5b, e, h, and k) with the consistent microphysical assumptions between COSP and the host model at the vertical levels of 2, 4, 8, and 11 km. The mean, standard deviation, 95th percentile values, and valid sample numbers between the model and NEXRAD are compared in Table 2. At 2-km altitude, the EAMv1 estimates higher reflectivity than the NEXRAD observations (Figs. 5a-b) except over the central United States. The overall mean value is 28.7 dBZ for EAMv1 and 25.1 dBZ for NEXRAD. The negative bias for the model is in the region between the Rocky Mountains and Mississippi basin (Fig. 5c), where precipitation is heavily contributed by Mesoscale Convective Systems (MCSs). Those MCSs propagate eastward from their initiation over or just east of the Rocky Mountains, go through upscale growth, and finally dissipate in the eastern part of the Mississippi Basin (Yang et al. 2017; Feng et al., 2018, 2019). The standard deviations of the two individual datasets are quite similar, and EAMv1 generates a higher 95th percentile value than the observation, indicating the model overestimates the extreme high values at lower troposphere. In addition, those simulated extreme values are evenly distributed across the entire domain, which fail to mimic the spatial footprint of MCSs as depicted by the NEXRAD data.

At 4-km altitude (Figs. 5d-e), the model's underestimation over central U.S. becomes larger compared to the 2-km altitude and the overestimation at the foothills of Rocky Mountains also becomes larger. The model also overestimates reflectivity in

the east region of the domain. These results indicate that the E3SM simulation fails to capture the observed spatial variability. The domain mean value between the model and observations is the same (24.0 dBZ) as a consequence of the offset between the negative and positive biases in different areas. The standard deviation and 95th percentile values are comparable with the observations as well. At 8 km, underestimation of the reflectivity by the model occurs over almost the entire domain (Fig. 5i), with a domain mean of 15.0 dBZ, much lower than 19.2 dBZ in the NEXRAD data. Meanwhile, the modelled standard deviation and the extreme values are smaller, indicating the model has a difficulty capturing the observed variability.

At 11-km altitude, the EAMv1 severely underestimates the reflectivity values compared to NEXRAD (Figs. 5j-k), with a mean value of 9.8 dBZ for EAMv1 while 16.6 dBZ for NEXRAD. The negative bias is generally more than 7.5 dBZ in the central United States (Fig. 5l), and the model severely underestimates the standard deviation and extreme reflectivity. Moreover, EAMv1's sample size is 50 time lower than that of the NEXRAD, indicating the lower occurrence of reflectivity values ≥ 8 dBZ.

Clearly, above 4 km, the model's negative biases increase with height as shown from Figs. 5f, i, and l, manifested in the central United States. There is no valid reflectivity value simulated by EAMv1 above 12-km altitude, where NEXRAD still shows reflectivity values up to 15.7 dBZ, indicating that the simulated deep convection in the warm season is not deep enough, a problem that is further examined in the following section.

In addition to the mean values, the histograms of observed and simulated radar reflectivities are compared for different altitudes, where the interval of reflectivity bins is 2 dBZ (Fig. 6). By comparing the occurrence of Z ≥ 8 dBZ between model and observations, the model apparently has narrower distribution than the observations, and the model-observation deviation in maximum values increases with height. At 8 km and below, the model generally overestimates the sample sizes of smaller reflectivity values but lacks extreme high reflectivity values. However, at 11-km altitude, the model greatly underestimates the sample sizes of the entire reflectivity spectrum compared to the observation, causing the severe underestimation in the mean value.

3.3 Comparison of Vertical Distribution of Radar Reflectivity

To quantitatively examine the simulated vertical distribution of radar reflectivity, contoured frequency by altitude diagrams (CFADs, Yuter and Houze 1995) are generated from NEXRAD and EAMv1 and compared in Fig. 7. The CFADs represent the frequency of occurrence of reflectivity in a coordinate system having reflectivity bins (interval of 1 dBZ) on the x-axis and altitude bins (interval of 1 km) on the y-axis. The frequency within each bin box is calculated as the number of valid samples it contains divided by the total sample number of all reflectivity bins at all levels, meaning that the integrated value of all frequencies in each plot is 100%.

Fig. 7 shows the CFADs for both NEXRAD observations (Figs. 7a, d, g, j, m, and p) and the EAMv1 simulation (Figs. 7b, e, h, k, n, and q) for each month from April to September combined over 2014-2016. The most distinct difference between the model and observations is the simulated echo top height. The echo top height in the simulation generally is at 11 km, at least 5 km lower than the 16 km top seen in the observations. At levels below 4 km, the NEXRAD data show a high frequency zone

(> 3.2%) concentrated between 8-25 dBZ, whereas the simulated high frequency zone is at 13-28 dBZ. For reflectivity > 35 dBZ, the simulation has higher probability of occurrence than the NEXRAD observations.

Regarding the overall shape of CFADs, the model follows the well-known pattern where the reflectivity value range of high frequency zone (> 3.2%) increases from cloud top to the freezing level, and then slowly decreases or remains constant below the freezing level. The cores of maximum frequency (> 5%) are located in the centres of the high frequency zones. However, these characteristics are not presented in the observations, whose high frequency zones are greatly skewed to the lower reflectivity values. The characteristics of NEXRAD's CFADs could be due to averaging from fine resolution (4 km) to coarse resolution (1°), as well as averaging of convective and stratiform components because the two components produce significantly different reflectivity profiles and magnitudes.

The box-whisker plots (Figs. 7c, f, i, l, o, and r) represent the same results in a different way, where the normalization is conducted at each level rather than against the entire dataset at all levels. Below 4 km, the percentile values are consistent between the model and observations except for the 1-km altitude where the model overestimates the reflectivity. The simulated 25-75th percentiles are located at the reflectivity values of 15-27 dBZ at 1-km altitude, which is higher than the NEXRAD observation (12 - 28 dBZ). As noted in the discussion of Fig. 5, the consistency at low-levels (e.g., 2 km) between the model and observations is mainly due to the offset of negative and positive biases at different regions of the domain. Moreover, EAMv1 underestimates the frequency of echoes ≤ 15 dBZ and overestimate it for echoes between 15 and 30 dBZ, which causes the higher median values in the model. From 4 km upward, the model-observation differences become much larger than at low levels, consistent with the result shown in Fig. 5. The underestimation of 95th percentile value increases from 10 dBZ at 7 km to more than 20 dBZ at 11 km. Above 11 km, the model fails to generate average reflectivity above 8 dBZ, and the typical reflectivity value is between 0 and 2 dBZ at 12 km.

From Fig. 7 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 the CFADs is found in the model, but the echo top height of the observations is not changed much. As a result, the choice of threshold does not change the conclusion of severe model underestimation in echo top height.

The CFADs of NEXRAD observations vary from month to month. For example, the echo top height is at 15 km in April, which increases to 16 km in May, then reaches 17 km in June and July, and finally decreases to 15 km in September. Similarly, the 0.6%-0.8% contour level in the observations stops at 9-km altitude in April, but extends to 10 km in May and reaches 11 km in June. It increases to the highest at 11.5 km in July and August, then decreases to 11 km in September. This seasonality follows the seasonal variation of intensity of convection (Wang et al., 2019a), which is not captured in the EAMv1 simulation (Figs. 7b, e, h, k, n, and q).

The severe underestimation of the echo top height by EAMv1 has been reported for simulation of tropical convection with the Community Atmosphere Model version 5 (CAM5) in a recent study (Wang and Zhang, 2019). Although EAMv1 is different from CAM5 in many aspects such as vertical resolution and dynamical core, they share the same Zhang-McFarlane (ZM)

cumulus parameterization (Zhang and McFarlane, 1995) for representing deep convection. Wang and Zhang (2019) found the cloud top height of tropical convection is underestimated by more than 2 km, which can be alleviated by the adjustment of the ZM scheme. We have performed a series of sensitivity tests by changing physical parameters in ZM and cloud microphysics schemes to explore the possibility of model improvement in echo top height. These tests are detailed in Section 3.4.

As evaluated in Zheng et al. (2019), E3SM v1 failed to simulate the diurnal variation of precipitation over the central United States, where the observed nocturnal peak is greatly underestimated. Xie et al. (2019) improved the diurnal cycle of convection in E3SM v1 recently by modifying convective trigger function in the ZM scheme. It will be interesting to see if the 3D radar reflectivity fields can be better simulated using the updated ZM scheme.

3.4 Sensitivity of Simulated Echo Top Height to Tunable Parameters of the Global Model

Differently from the model evaluation of cloud top height and high cloud fraction (e.g., Xie et al., 2018), where EAMv1 has shown good agreements with satellite observations over the CONUS, evaluation of radar echo top height indicates whether the processes internal to the cloud are producing precipitation correctly. To examine if any model parameters in the ZM cumulus parameterization scheme and/or MG2 microphysics parameterization scheme can significantly influence the echo top height, we conducted a series of sensitivity tests for the tunable parameters as listed in Table 3. In each test a single parameter is changed, and all other parameters retain their default values.

Wang and Zhang (2018) suggested that the restriction of neutral buoyancy level (NBL) from the dilute CAPE calculation (Neale et al. 2008) can limit the depth of deep convection in ZM. When the convective plume reaches the NBL, all mass flux is detrained even if the updraft is still positively buoyant from the cloud model calculation (Zhang, 2009). To allow deep convection to grow deeper, we performed a sensitivity test following Wang and Zhang (2018), where the NBL determined in the dilute CAPE calculation is removed, and the upper limit of the integrals of mass flux, moist static energy, and other cloud properties is set to be very high (70 hPa in this study). After the modification, the convective cloud top height increases as shown in Wang and Zhang (2018), however there is no change in the radar echo top height, i.e., the maximum altitude at which precipitation-sized particles occur. A possible reason for the limited effect on echo top height is that the cloud ice content is too low in midlatitude continental convection without convective microphysics parameterization (Song et al., 2012), which cannot be improved by merely increasing the NBL.

Other parameters that we tested in the ZM cumulus parameterization with the dilute CAPE calculation include convective entrainment rate (zmconv_dmpdz), the convection adjustment time scale (zmconv_tau), the coefficient of autoconversion rate (zmconv_c0_lnd), ice particle size (clubb_ice_deep), convective fraction (cldfrc_dp), and number of layers allowed for negative CAPE (zmconv_cape_cin). The overall conclusion is that separately tuning any of these parameters does not improve the simulation of echo top height. For the convective entrainment rate (zmconv_dmpdz), we decreased its value from -0.7×10^{-3} to -1.0×10^{-5}, which means that the entrainment in convection is almost turned off, similar to the undiluted CAPE assumption. Results show the simulated echo top height is increased by 500-800 m in the EAMv1-test simulation, and the reflectivity span in the lower troposphere is narrowed by 1-2 dBZ, which is closer to the observations (Fig. 8). This result is consistent with the

previous studies that tested the undiluted CAPE assumption as well (Neale et al., 2008; Hannah and Maloney, 2014). However, that assumption is unrealistic given the fact that the undiluted CAPE-based closure strongly deviated from observations (Zhang, 2009). In summary, changing any of our selected parameters individually in the ZM scheme does not improve the simulation of echo top height.

The MG2 cloud microphysics parameterization in E3SM determines only large-scale cloud and precipitation (i.e., those resolved by the model). Changes in the MG2 cloud microphysics parameterization could affect the parameterized cumulus cloud and precipitation by changing the large-scale forcing which feeds into the cumulus cloud calculations. By decreasing the MG2 autoconversion rate (prc_coef1), ideally the depletion of moisture within the atmospheric column is slowed down and more water vapor can be supplied to cumulus convection. Results show, however, that the echo top height is not affected by changing the MG2 assumptions. Attempts at accelerating the Wegener–Bergeron–Findeisen process in MG2 to increase the conversion of liquid to snow/ice, as well as using lower size threshold for the ice-to-snow conversion have also proven to be unimportant to the simulation of echo top height.

Thus, echo top height proves to be insensitive to the available tunable parameters. Setting the value of convective entrainment rate to be unrealistically low only gains a 500-800 m increment in echo top height. Given that the model underestimation is more than 5 km, the increment is insufficient to solve the discrepancy. Note that each individual tunable parameter was changed without retuning the model to keep the top-of-atmosphere radiative energy budget balanced and the model performance optimized. Thus, some expected improvement in echo top height can be subsequently offset by other untuned processes. Instead of providing quantification of how the model responds to the changes of parameters, we emphasize the trend of change in echo top height, in which the simulation of the echo top height cannot be significantly improved by tuning only one of those physical parameters. Further investigation of combinations of two and more parameters is a topic for a future study.

4 Conclusions and Discussion

We have evaluated the model performance of E3SM EAMv1 in simulating the warm-season 3D radar reflectivity at an hourly scale over the North American sector of the globe by comparing the model results to the 3D distribution of radar reflectivity observed by NEXRAD radars over the CONUS during April-September of 2014-2016. The evaluation is achieved by improving the COSP radar simulator and employing special data processing techniques to ensure fair comparison between model and observations that are different in sampling frequency, horizontal-vertical resolutions, and minimum detection limit. We find that:

1. With the default microphysics assumptions in COSP, the simulated subgrid reflectivity PDF is bimodal, in disagreement with radar observations which show that the subgrid reflectivity follows a gamma distribution. Changing the microphysics assumptions in COSP to be consistent with the MG2 microphysics parameterization used in E3SM, the bimodality of the subgrid distribution is nearly eliminated. It is therefore important to maintain consistency of microphysics assumptions between the host model and radar-echo simulator attached to the model.

2. Below the 4-km altitude, the simulated domain-mean reflectivities by EAMv1 agree with NEXRAD observations in the magnitude, but the simulation fails to capture the spatial variability. The model underestimates the reflectivity in central U.S. between the Rocky Mountains and Mississippi River. This pattern suggests that the model is not adequately representing the mesoscale convective systems that dominate warm season rainfall in that region. The model overestimates the reflectivity outside this region.
3. Above 4-km altitude, EAMv1 shows a severe underestimation of the domain-mean reflectivity, and the negative bias increases with altitude up to 11 km, above which model fails to simulate any valid reflectivity at all, whereas NEXRAD observations show strong radar echoes up to 16 km.
4. EAMv1 is able to simulate the variability and extreme value of reflectivity at the lower troposphere but significantly underestimate them at high levels.

The NEXRAD observations used in this study reveal that EAMv1 fails to simulate the occurrence of large ice-phase particles at high levels in deep convective clouds. In addition, the conclusion of "simulated deep convection is not deep enough" also echoes the dry bias seen in GCMs as manifested in underestimations of total precipitation and individually large rain rates over the CONUS (e.g., Zheng et al., 2019). We have now shown that this model deficiency cannot be significantly improved by tuning a single value of the physical parameters in the ZM cumulus and MG2 cloud microphysics schemes. Note the large-scale circulation is nudged towards observations for the simulations in this study, so our results represent the best-case model performance. Compared to the nudged simulations, free running of EAMv1 has shown nonnegligible biases in the regional circulation (Sun et al., 2019). With the nudged simulations, the large biases in circulation can be excluded so that the performances of physics parameterizations in simulating convective systems can be more insightfully understood.

The data processing techniques and metrics we have developed in this study can be used globally for model evaluation when satellite-based radars provide global 3D radar observations. The GPM radar observations will eventually be able to provide global radar echo coverage (Houze et al., 2019), whose data have been proven consistent with NEXRAD (Wang et al., 2019b). However, as discussed in Section 2, the sampling by GPM at 1° model grid elements for only three years of GPM data is insufficient for obtaining robust grid-mean values to compare with the EAMv1 simulation. In addition to the restriction in the availability of observational data, the high computation cost with the incorporation of COSP simulator in simulation and the demand of large data space (14,000 core hours and 1.2 TB data per simulation month at hourly output frequency) have hindered the modelling for an extended period. When GPM has run for a much longer time period and more powerful computational resources become available, it will be a very useful study to evaluate the long-term model simulations at the global scale. 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. Future studies can also focus on separately evaluating properties in convective and stratiform regions, since the thermodynamic and reflectivity profiles are fundamentally different between the two regions.

Code Availability

The source code in this study is based on the Department of Energy (DOE) Energy Exascale Earth System Model (E3SM) Project version 1 at revision 9a86ab9 whose code can be acquired from the E3SM repository (https://github.com/E3SM-Project/E3SM/tree/kaizhangpnl/atm/cm20170220), which is also permanently archived in the National Energy Research Scientific Computing Canter (NERSC) High Performance Storage System (HPSS) at https://portal.nersc.gov/archive/home/w/wang406/www/Publication/Wang2020GMD.

Data Availability

The observational data is available through National Center for Atmospheric Research (NCAR) Research Data Archive (https://doi.org/10.5065/D6NK3CR7). Model results can be accessed from https://portal.nersc.gov/archive/home/w/wang406/www/Publication/Wang2020GMD.

Author Contributions

Jingyu Wang performed the simulations and conducted the analyses. Jiwen Fan and Robert A. Houze Jr developed the idea of this research. Kai Zhang helped on the model configuration and Po-Lun Ma implemented the radar simulator. Guang J. Zhang provided feedback and helped shape the research. All authors discussed the results and contributed to the final manuscript.

Acknowledgement

We acknowledge the support of the Climate Model Development and Validation (CMDV) project at PNNL. The effort of J. Wang, J. Fan, Kai Zhang, and Po-Lun Ma was supported by CMDV. Robert A. Houze was supported by NASA Award NNX16AD75G and by master agreement 243766 between the University of Washington and PNNL. Stella R. Brodzik was supported by NASA Award NNX16AD75G and subcontracts from the CMDV and Water Cycle and Climate Extreme Modeling (WACCEM) projects of PNNL. Guang J. Zhang was supported by the DOE Biological and Environmental Research Program (BER) Award DE-SC0019373. PNNL is operated for the US Department of Energy (DOE) by Battelle Memorial Institute under Contract DE-AC05-76RL01830. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under contract DE-AC02-05CH11231. The GridRad radar dataset is obtained at the Research Data Archive of the National Center for Atmospheric Research (NCAR) (https://rda.ucar.edu/datasets/ds841.0/).

References

Arakawa, A., and W. H.Schubert: Interaction of a cumulus cloud ensemble with the large-scale environment, part I. J. Atmos. Sci., 31, 674–701, doi:10.1175/1520-0469(1974)031<0674:IOACCE>2.0.CO;2, 1974.

Bader, D., Collins, W., Jacob, R., Jones, P., Rasch, P., Taylor, M., et al: Accelerated Climate Modeling for Energy. U. S. Department of Energy. Retrieved from https://climatemodeling.science.energy.gov/sites/default/files/publications/acme-project-strategy-plan.pdf, 2014.

Bodas-Salcedo, A., Webb, M. J., Bony, S., Chepfer, H., Dufresne, J.-L., Klein, S. A., et al.: COSP: Satellite simulation software for model assessment, Bulletin of the American Meteorological Society, 92(8), 1023–1043, doi:10.1175/2011BAMS2856.1, 2011.

Carbone, R. E., and J. D. Tuttle: Rainfall Occurrence in the U.S. Warm Season: The Diurnal Cycle. J. Climate, 21, 4132–4146, doi:10.1175/2008JCLI2275.1, 2008.

Dennis, J., Edwards, K., Evans, J., Guba, O., Lauritzen, P. H., Mirin, A. A., St-Cyr, A., Taylor, M. A., & Worley, P. H.: CAM-SE: A scalable spectral element dynamical core for the Community Atmosphere Model, International Journal of High Performance Computing Applications, 26(1), 74–89, 2012.

Eden, J.M. and M. Widmann: Downscaling of GCM-Simulated Precipitation Using Model Output Statistics, J. Climate, 27, 312–324, doi:10.1175/JCLI-D-13-00063.1, 2014.

Fan, J., Han, B., Varble, A., Morrison, H., North, K., Kollias, P., Chen, B., Dong, X., Giangrande, S. E., Khain, A., Lin, Y., Mansell, E., Milbrandt, J. A., Stenz, R., Thompson, G., & Wang, Y.: Cloud-resolving model intercomparison of an MC3E squall line case: Part I—Convective updrafts, Journal of Geophysical Research: Atmospheres, 122, 9351–9378, doi:10.1002/2017JD026622, 2017.

Feng, Z., Leung, L. R., Houze, R. A., Jr., Hagos, S., Hardin, J., Yang, Q., et al.: Structure and evolution of mesoscale convective systems: Sensitivity to cloud microphysics in convection-permitting simulations over the United States, J. Adv. Model. Earth Syst, 10, 1470–1494, doi:10.1029/2018MS001305, 2018.

Feng, Z., R. A. Houze, L. R. Leung, F. Song, J. C. Hardin, J. Wang, W. I. Gustafson, and C. R. Homeyer: Spatiotemporal Characteristics and Large-Scale Environments of Mesoscale Convective Systems East of the Rocky Mountains, J. Climate, 32, 7303–7328, doi:10.1175/JCLI-D-19-0137.1, 2019.

Feng, Z., X. Dong, B. Xi, S. A. McFarlane, A. Kennedy, B. Lin, and P. Minnis: Life cycle of midlatitude deep convective systems in a Lagrangian framework, J. Geophys. Res., 117, D23201, doi:10.1029/2012JD018362, 2012.

Gelaro, R., and Coauthors: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). J. Climate, 30, 5419–5454, doi:10.1175/JCLI-D-16-0758.1, 2017.

Gettelman, A., and H. Morrison: Advanced two-moment bulk microphysics for global models. Part I: Off-line tests and comparison with other schemes, J. Climate, 28, 1268–1287, doi:10.1175/JCLI-D-14-00102.1, 2015.

Golaz, J.-C., Larson, V. E., & Cotton, W. R. (2002). A PDF-based model for boundary layer clouds. Part I: Method and model description, Journal of the Atmospheric Sciences, 59(24), 3540–3551, doi:10.1175/1520-0469(2002)059<3540:APBMFB>2.0.CO;2, 2002.

Han, B., Fan, J., Varble, A., Morrison, H., Williams, C. R., Chen, B., et al.: Cloud-resolving model intercomparison of an MC3E squall line case: Part II. Stratiform precipitation properties, Journal of Geophysical Research: Atmospheres, 124, 1090–1117, doi:10.1029/2018JD029596, 2019.

Hannah, W. M., and Maloney, E. D.: The moist static energy budget in NCAR CAM5 hindcasts during DYNAMO, J. Adv. Model, Earth Syst., 6, 420– 440, doi:10.1002/2013MS000272, 2014.

Haynes, J. M., and G. L. Stephens, Tropical oceanic cloudiness and the incidence of precipitation: Early results from CloudSat, Geophys, Res. Lett., L09811, doi:10.1029/2007GL029335, 2007.

He, F. and D.J. Posselt, Impact of Parameterized Physical Processes on Simulated Tropical Cyclone Characteristics in the Community Atmosphere Model, J. Climate, 28, 9857–9872, doi:10.1175/JCLI-D-15-0255.1, 2015.

Hillman, B. R., Marchand, R. T., and Ackerman, T. P.: Sensitivities of simulated satellite views of clouds to subgrid-scale overlap and condensate heterogeneity, Journal of Geophysical Research: Atmospheres, 123, 7506–7529. doi:10.1029/2017JD027680, 2018.

Homeyer, C. R., and K. P. Bowman: Algorithm Description Document for Version 3.1 of the Three-Dimensional Gridded NEXRAD WSR-88D Radar (GridRad) Dataset, Technical Report. [Available online at http://gridrad.org/pdf/GridRad-v3.1-Algorithm-Description.pdf], 2017.

Houze, R. A., Wang, J., Fan, J., Brodzik, S., & Feng, Z.: Extreme convective storms over high-latitude continental areas where maximum warming is occurring, Geophysical Research Letters, 46, 4059–4065, doi:10.1029/2019GL082414, 2019.

Houze, R. A., Wilton, D. C. and Smull, B. F.: Monsoon convection in the Himalayan region as seen by the TRMM Precipitation Radar, Q. J. R. Meteorol. Soc., 133: 1389-1411, doi:10.1002/qj.106, 2007.

Iguchi, T., N. Kawamoto, and R. Oki, Detection of Intense Ice Precipitation with GPM/DPR. J. Atmos. Oceanic Technol., 35, 491–502, doi:10.1175/JTECH-D-17-0120.1, 2018.

Jensen, M. P., W. A. Petersen, A. Bansemer, N. Bharadwaj, L. D. Carey, D. J. Cecil, S. M. Collis, et al.: The Midlatitude Continental Convective Clouds Experiment (MC3E), Bull. Amer. Meteorol. Soc., 97, no. 9, 1667-1686, doi:10.1175/BAMS-D-14-00228.1, 2016.

Klein, S.A. and C. Jakob: Validation and Sensitivities of Frontal Clouds Simulated by the ECMWF Model, Mon. Wea. Rev., 127, 2514–2531, doi:10.1175/1520-0493(1999)127<2514:VASOFC>2.0.CO;2, 1999.

Larson, V. E.: CLUBB-SILHS: A parameterization of subgrid variability in the atmosphere, arXiv:1711.03675, 2017.

L'Ecuyer, T. S., and G. L. Stephens: An Estimation-Based Precipitation Retrieval Algorithm for Attenuating Radars. J. Appl. Meteor., 41, 272–285, doi:10.1175/1520-0450(2002)041<0272:AEBPRA>2.0.CO;2, 2002.

Lim, K.-S. S., Fan, J., Leung, L. R., Ma, P.-L., Singh, B., Zhao, C., Zhang, Y., Zhang, G., and Song, X.: Investigation of aerosol indirect effects using a cumulus microphysics parameterization in a regional climate model, J. Geophys. Res. Atmos., 119, 906– 926, doi:10.1002/2013JD020958, 2014.

Lin, G., Wan, H., Zhang, K., Qian, Y., and Ghan, S. J.: Can nudging be used to quantify model sensitivities in precipitation and cloud forcing? J. Adv. Model. Earth Syst., 8, 1073– 1091, doi:10.1002/2016MS000659, 2016.

Lin, G., Fan, J., Feng, Z., Gustafson, W. I., Ma, P.-L., & Zhang, K.: Can the multiscale modeling framework (mmf) simulate the mcs-associated precipitation over the Central United States? Journal of Advances in Modeling Earth Systems, 11, doi:10.1029/2019MS001849, 2019.

Lin, S.-J.: A "Vertically Lagrangian" Finite-Volume Dynamical Core for Global Models, Monthly Weather Review, 132(10), 2293–2307, doi10.1175/1520-0493(2004)132<2293:AVLFDC>2.0.CO;2, 2004.

Liu, X., Ma, P.-L., Wang, H., Tilmes, S., Singh, B., Easter, R. C., Ghan, S. J., & Rasch, P. J.: Description and evaluation of a new 4-mode version of Modal Aerosol Module (MAM4) within version 5.3 of the Community Atmosphere Model, Geoscientific Model Development, 9, 505–522. doi:10.5194/gmd-9-505-2016, 2016.

Ma, P.-L., Rasch, P. J., Fast, J. D., Easter, R. C., Gustafson Jr., W. I., Liu, X., Ghan, S. J., and Singh, B.: Assessing the CAM5 physics suite in the WRF-Chem model: implementation, resolution sensitivity, and a first evaluation for a regional case study, Geosci. Model Dev., 7, 755–778, doi:10.5194/gmd-7-755-2014., 2014.

Marchand, R., Haynes, J., Mace, G. G., Ackerman, T., and Stephens: A comparison of simulated cloud radar output from the multiscale modeling framework global climate model with CloudSat cloud radar observations, J. Geophys. Res., 114, D00A20, doi:10.1029/2008JD009790, 2009.

Neale, R. B., Gettelman, A., Park, S., Conley, A. J., Kinnison, D., Marsh, D., et al.: Description of the NCAR Community Atmosphere Model (CAM 5.0), tech. Note NCAR/TN-486+STR, Natl. Cent. For Atmos (pp. 2009–038451) [Available online at http://www.cesm.ucar.edu/models/ccsm4.0/cam/docs/description/cam4_desc.pdf], 2012.

Neale, R. B., Richter, J. H., & Jochum, M.: The Impact of Convection on ENSO: From a Delayed Oscillator to a Series of Events, Journal of Climate, 21(22), 5904–5924. https://doi.org/10.1175/2008JCLI2244.1, 2008.

Pincus, R, Richard S Hemler, and Stephen A Klein: Using Stochastically Generated Subcolumns to Represent Cloud Structure in a Large-Scale Model, Monthly Weather Review, 134, doi:10.1175/MWR3257.1, 2006.

Qian, Y., Wan, H., Yang, B., Golaz, J.-C., Harrop, B., Hou, Z., et al.: Parametric sensitivity and uncertainty quantification in the version 1 of E3SM atmosphere model based on short perturbed parameter ensemble simulations, Journal of Geophysical Research: Atmospheres, 123, 13,046–13,073. https://doi.org/10.1029/2018JD028927, 2018.

Randall, D. A., et al.: Climate models and their evaluation. Climate Change 2007: The Physical Science Basis, S. Solomon et al., Eds., Cambridge University Press, 589–662, 2007.

Randel, D. L., T. H. Vonder Haar, M. A. Ringerud, G. L. Stephens, T. J. Greenwald, and C. L. Combs: A new global water vapor dataset. Bull. Amer. Meteor. Soc., 77, 1233–1246, 1996.

Rasch, P. J., Xie, S., Ma, P.-L., Lin, W., Wang, H., Tang, Q., Burrows, S. M., Caldwell, P., Zhang, K., Easter, R. C., et al.: An Overview of the Atmospheric Component of the Energy Exascale Earth System Model, J. Adv. Model. Earth Syst., 11, 2377– 2411, doi:10.1029/2019MS001629, 2019.

Richter, J. H., & Rasch, P. J.: Effects of convective momentum transport on the atmospheric circulation in the Community Atmosphere Model, Version 3, Journal of Climate, 21(7), 1487–1499, doi:10.1175/2007JCLI1789.1, 2008.

Smalley, M., P. Kirstetter, and T. L'Ecuyer: How Frequent is Precipitation over the Contiguous United States? Perspectives from Ground-Based and Spaceborne Radars. J. Hydrometeor., 18, 1657–1672, https://doi.org/10.1175/JHM-D-16-0242.1, 2017.

Song, H., Z. Zhang, P.-L. Ma, S. Ghan and M. Wang: The importance of considering sub-grid cloud variability when using satellite observations to evaluate the cloud and precipitation simulations in climate models, Geosci. Model Dev., 11, 3147-3158, doi:10.5194/gmd-11-3147-2018, 2018.

Song, F., Z. Feng, L.R. Leung, R.A. Houze Jr, J. Wang, J. Hardin, and C.R. Homeyer: Contrasting Spring and Summer Large-Scale Environments Associated with Mesoscale Convective Systems over the U.S. Great Plains, J. Climate, 32, 6749–6767, doi:10.1175/JCLI-D-18-0839.1, 2019.

Stephens, G. L., and C. D. Kummerow: The remote sensing of clouds and precipitation from space: A review. J. Atmos. Sci., 64, 3742–3765, 2007.

Sun, J., Zhang, K., Wan, H., Ma, P.-L., Tang, Q., Zhang, S.: Impact of nudging strategy on the climate representativeness and hindcast skill of constrained EAMv1 simulations, Journal of Advances in Modeling Earth Systems, doi: 10.1029/2019MS001831, 2019.

Swales, D. J., Pincus, R., and Bodas-Salcedo, A.: The Cloud Feedback Model Intercomparison Project Observational Simulator Package: Version 2, Geosci. Model Dev., 11, 77-81, doi:10.5194/gmd-11-77-2018, 2018.

Taylor, M. A., and Fournier, A.: A compatible and conservative spectral element method on unstructured grids. Journal of Computational Physics, 229(17), 5879–5895, doi:10.1016/j.jcp.2010.04.008, 2010.

Taylor, M. A. (2011). Conservation of mass and energy for the moist atmospheric primitive equations on unstructured grids. In P. H. Lauritzen, et al. (Eds.), Numerical techniques for global atmospheric models, Lecture Notes Comput. Sci. Eng. (Vol. 80, pp. 357–380). Heidelberg, Germany: Springer, doi:10.1007/978-3-642-11640-7_12, 2011.

Tian, J., Dong, X., Xi, B., Wang, J., Homeyer, C. R., McFarquhar, G. M., and Fan, J.: Retrievals of ice cloud microphysical properties of deep convective systems using radar measurements, J. Geophys. Res. Atmos., 121, 10,820– 10,839, doi:10.1002/2015JD024686, 2016.

Trenberth, K. E., et al.: Observations: Surface and atmospheric climate change. Climate Change 2007: The Physical Science Basis, S. Solomon et al., Eds., Cambridge University Press, 235–336, 2007.

Um, J., McFarquhar, G. M., Stith, J. L., Jung, C. H., Lee, S. S., Lee, J. Y., Shin, Y., Lee, Y. G., Yang, Y. I., Yum, S. S., Kim, B.-G., Cha, J. W., and Ko, A.-R: Microphysical characteristics of frozen droplet aggregates from deep convective clouds, Atmos. Chem. Phys., 18, 16915-16930, doi:10.5194/acp-18-16915-2018, 2018.

Wang, H., Easter, R. C., Zhang, R., Ma, P.-L., Singh, B., Zhang, K., et al.: Aerosols in the E3SM Version 1: New developments and their impacts on radiative forcing. Journal of Advances in Modeling Earth Systems, 12, e2019MS001851, doi:10.1029/2019MS001851, 2020.

Wang, J, Dong, X, and Xi, B: Investigation of ice cloud microphysical properties of DCSs using aircraft in situ measurements during MC3E over the ARM SGP site, J. Geophys. Res. Atmos., 120, 3533– 3552. doi: 10.1002/2014JD022795, 2015.

Wang, J., Dong, X., Xi, B., and Heymsfield, A. J.: Investigation of liquid cloud microphysical properties of deep convective systems: 1. Parameterization of raindrop size distribution and its application for stratiform rain estimation, J. Geophys. Res. Atmos., 121, 10,739– 10,760, doi:10.1002/2016JD024941, 2016.

Wang, J., Dong, X., & Xi, B.: Investigation of liquid cloud microphysical properties of deep convective systems: 2. Parameterization of raindrop size distribution and its application for convective rain estimation. Journal of Geophysical Research: Atmospheres, 123, 11,637– 11,651, doi:10.1029/2018JD028727, 2018.

Wang, J., X. Dong, A. Kennedy, B. Hagenhoff, and B. Xi: A Regime-Based Evaluation of Southern and Northern Great Plains Warm-Season Precipitation Events in WRF, Wea. Forecasting, 34, 805–831, doi:10.1175/WAF-D-19-0025.1, 2019a.

Wang, J., R. A. Houze, Jr., J. Fan, S. R. Brodzik, Z. Feng, and J. C. Hardin: The detection of mesoscale convective systems by the GPM Ku-band spaceborne radar, J. Meteor. Soc. Japan, 97, Special Edition on Global Precipitation Measurement (GPM): 5th Anniversary, doi:10.2151/jmsj.2019-058, 2019b.

Wang, M. and G.J. Zhang: Improving the Simulation of Tropical Convective Cloud-Top Heights in CAM5 with CloudSat Observations. J. Climate, 31, 5189–5204, doi:10.1175/JCLI-D-18-0027.1, 2018.

Warren, R. A., A. Protat, S. T. Siems, H. A. Ramsay, V. Louf, M. J. Manton, and T. A. Kane: Calibrating Ground-Based Radars against TRMM and GPM. J. Atmos. Oceanic Technol., 35, 323–346, doi:10.1175/JTECH-D-17-0128.1, 2018.

Webb, M., C. Senior, S. Bony, and J. J. Morcrette: Combining ERBE and ISCCP data to assess clouds in the Hadley Centre, ECMWF and LMD atmospheric climate models, Clim Dyn, 17, 905–922, doi:10.1007/s003820100157, 2001.

Xie, S., Lin, W., Rasch, P. J., Ma, P.-L., Neale, R., Larson, V. E., et al.: Understanding cloud and convective characteristics in version 1 of the E3SM atmosphere model, Journal of Advances in Modeling Earth Systems, 10, 2618–2644, doi:10.1029/ 2018MS001350, 2018.

Xie, S., Wang, Y.-C., Lin, W., Ma, H.-Y., Tang, Q., Tang, S., et al.: Improved diurnal cycle of precipitation in E3SM with a revised convective triggering function. Journal of Advances in Modeling Earth Systems, 11, 2290–2310. doi.org/10.1029/2019MS001702, 2019.

Yang, Q., R. A. Houze, Jr., L. R. Leung, and Z. Feng, 2017: Environments of long-lived mesoscale convective systems over the central United States in convection permitting climate simulations. J. Geophys. Res. Atmos., 122, 13,288–13,307, doi:10.1002/2017JD027033, 2017.

Yuter, S. E., and R. A. Houze, Jr.: Three-dimensional kinematic and microphysical evolution of Florida cumulonimbus, Part II: Frequency distribution of vertical velocity, reflectivity, and differential reflectivity, Mon. Wea. Rev., 123, 1941-1963, 1995.

Zhang, G. J.: Effects of entrainment on convective available potential energy and closure assumptions in convection parameterization, J. Geophys. Res., 114, D07109, doi:10.1029/2008JD010976, 2009.

Zhang, G. J. and N. A. McFarlane: Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian climate centre general circulation model, Atmosphere-Ocean, 33:3, 407-446, doi: 10.1080/07055900.1995.9649539, 1995.

Zhang, J., K. Howard, C. Langston, B. Kaney, Y. Qi, L. Tang, H. Grams, Y. Wang, S. Cocks, S. Martinaitis, A. Arthur, K. Cooper, J. Brogden, and D. Kitzmiller (2016: Multi-Radar Multi-Sensor (MRMS) Quantitative Precipitation Estimation: Initial Operating Capabilities. Bull. Amer. Meteor. Soc., 97, 621–638, doi:10.1175/BAMS-D-14-00174.1, 2016.

Zhang, J., K. Howard, C. Langston, S. Vasiloff, B. Kaney, A. Arthur, S. Van Cooten, K. Kelleher, D. Kitzmiller, F. Ding, D. Seo, E. Wells, and C. Dempsey: National Mosaic and Multi-Sensor QPE (NMQ) System: Description, Results, and Future Plans. Bull. Amer. Meteor. Soc., 92, 1321–1338, doi:10.1175/2011BAMS-D-11-00047.1, 2011.

Zhang, K., H. Wan, X. Liu, S. J. Ghan, G. J. Kooperman, P. L. Ma, P. J. Rasch, D. Neubauer, and U. Lohmann: Technical Note: On the use of nudging for aerosol-climate model intercomparison studies, Atmos. Chem. Phys., 14, 8631–8645, doi:10.5194/acp-14-8631-2014, 2014.

Zhang, K., Rasch, P. J., Taylor, M. A., Wan, H., Leung, R., Ma, P.-L., Golaz, J. C., et al.: Impact of numerical choices on water conservation in the E3SM Atmosphere Model version 1 (EAMv1), Geoscientific Model Development, 11(5), 1971–1988. https://doi.org/ 10.5194/gmd-11-1971-2018, 2018.

Zhang, Y., Klein, S. A., Boyle, J., and Mace, G. G.: Evaluation of tropical cloud and precipitation statistics of Community Atmosphere Model version 3 using CloudSat and CALIPSO data, J. Geophys. Res., 115, D12205, doi:10.1029/2009JD012006, 2010.

Zhang, Y., S. Xie, S. A. Klein, R. Marchand, P. Kollias, E. E. Clothiaux, W. Lin, K. Johnson, D. Swales, A. Bodas-Salcedo, S. Tang, J. M. Haynes, S. Collis, M. Jensen, N. Bharadwaj, J. Hardin, and B. Isom: The ARM Cloud Radar Simulator for Global Climate Models: Bridging Field Data and Climate Models. Bull. Amer. Meteor. Soc., 99, 21–26, doi:10.1175/BAMS-D-16-0258.1, 2018.

Zhang, Y., Xie, S., Lin, W., Klein, S. A., Zelinka, M., Ma, P.-L., et al.: Evaluation of clouds in version 1 of the E3SM atmosphere model with satellite simulators, Journal of Advances in Modeling Earth Systems, 11, 1253– 1268, doi:10.1029/2018MS001562, 2019.

Zheng, X., Golaz, J.-C., Xie, S., Tang, Q., Lin, W., Zhang, M., et al.: The summertime precipitation bias in E3SM Atmosphere Model version 1 over the Central United States. Journal of Geophysical Research: Atmospheres, 124, 8935–8952, doi:10.1029/2019JD030662, 2019.

Table List

Table 1. Modification of the hydrometeor assumptions used in COSP.

Hydrometeor Type[1]	Distribution Type		Density (kg m^{-3})		Particle Mean Diameter (μm)		Distribution Width[2] (Unitless)	
	Default	Modified	Default	Modified	Default	Modified	Default	Modified
LSL	Lognormal	Gamma	$524\times D^3$	-	6	12	0.3	0
CVL	Lognormal	Gamma	$524\times D^3$	-	6	12	0.3	0
LSI	Gamma	-	$110.8\times D^{2.91}$	500	4	-	2	0
CVI	Gamma	-	$110.8\times D^{2.91}$	500	4	-	2	0
LSS	Exponential	-	100	250	N/A	-	N/A	-
CVS	Exponential	-	100	250	N/A	-	N/A	-

[1]LS: Large-Scale; CV: Convective; L: Cloud Liquid; I: Cloud Ice; S: Snow.

[2]Distribution width: ν in $N(D) = N_0 D^{(\nu-1)} e^{-\lambda D}$, which is a shape parameter in Gamma distribution describing the dispersion of the distribution.

Table 2. The statistical comparison of radar reflectivity between NEXRAD and EAMv1

Altitude	NEXRAD				EAMv1			
	Mean (dBZ)	Standard Deviation (dBZ)	95th Percentile (dBZ)	Sample Numbers	Mean (dBZ)	Standard Deviation (dBZ)	95th Percentile (dBZ)	Sample Numbers
2 km	25.1	7.7	32.1	1.7×10^6	28.7	7.4	35.8	4.1×10^6
4 km	24.0	7.2	31.6	1.6×10^6	24.0	6.4	30.2	4.2×10^6
8 km	19.2	5.2	24.4	7.9×10^5	15.0	3.9	21.0	1.5×10^6
11 km	16.6	4.4	21.8	2.2×10^5	9.8	1.6	12.9	4.1×10^3

655

Table 3. Changes of the tunable parameters in the sensitivity tests for echo top height.

	Parameter	Physics Meaning	Default	Changed Values	Impact
Cumulus parameterization	NBL restriction	The upper limit level of the integral of the mass flux, moist static energy etc. in ZM	Calculated NBL	200 hPa, 70 hPa	No
	zmconv_dmpdz	ZM entrainment rate in CAPE calculation	-0.7e-3	-1.0e-3, -1.0e-5	Yes
	zmconv_tau	Convection adjustment time scale	1 hr	15min, 6 hr	No
	zmconv_c0_lnd	Coefficient of autoconversion rate in ZM	0.007	0.01, 0.002	No
	zmconv_cape_cin	Number of layers allowed for negative CAPE	1	5, 10	No
	clubb_ice_deep	Assumed ice condensate radius detrained from ZM	16e-6	32e-6, 8e-6	No
	cldfrc_dp1	Convective fraction	0.045	0.01, 0.2	No
Microphysics parameterization	prc_coef1	Coefficient of autoconversion rate in MG2	30500	10000, 675	No
	berg_eff_factor	Efficiency factor for the Wegener–Bergeron–Findeisen process	0.1	0.2, 0.7	No
	thres_ice_snow	Autoconversion size threshold from cloud ice to snow	Temperature dependent	Maximize at 175e-6	No

Figure List

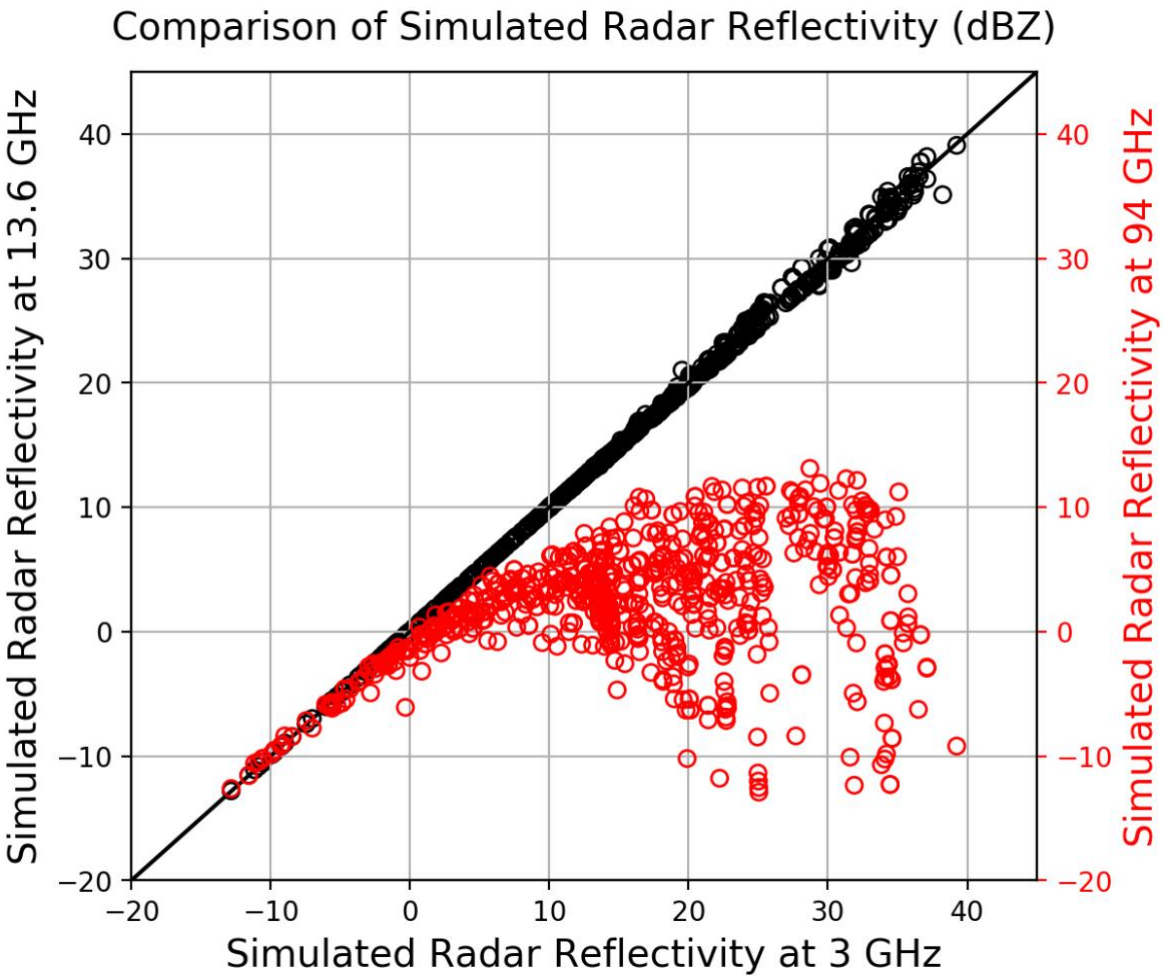

Figure 1: Scatter plots between radar reflectivity values simulated by the COSP simulator at 3 GHz (x-axis) versus those simulated at 13.6 GHz (left y-axis) and 94 GHz (right y-axis).

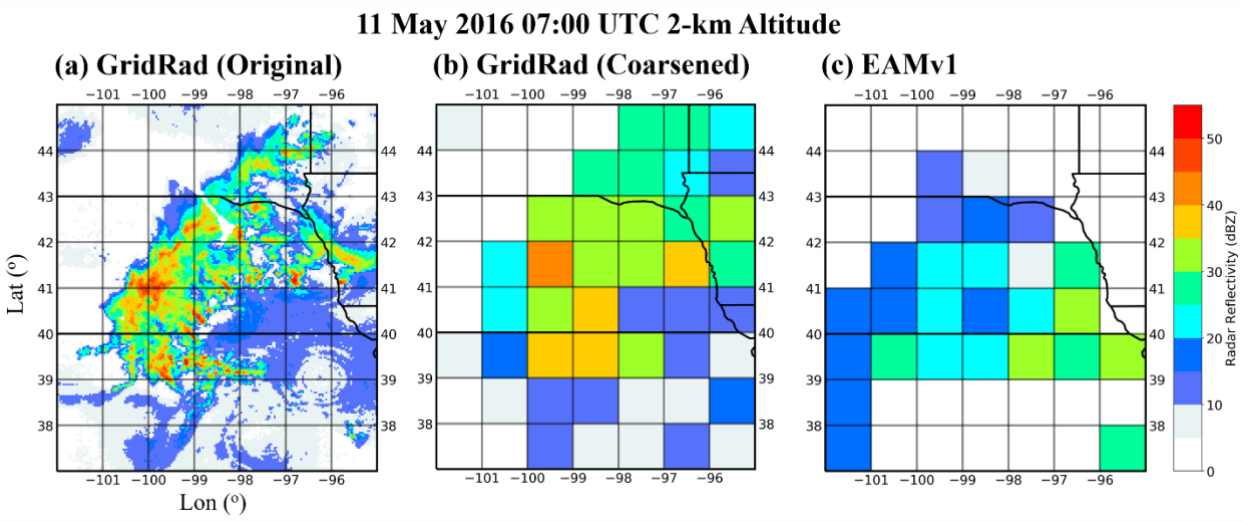

Figure 2: Examples of (a) original GridRad observation, (b) GridRad mapped over the E3SM model grid, and (c) the concurrent model simulation on 2016 May 11, 07:00 UTC, at the 2-km altitude.

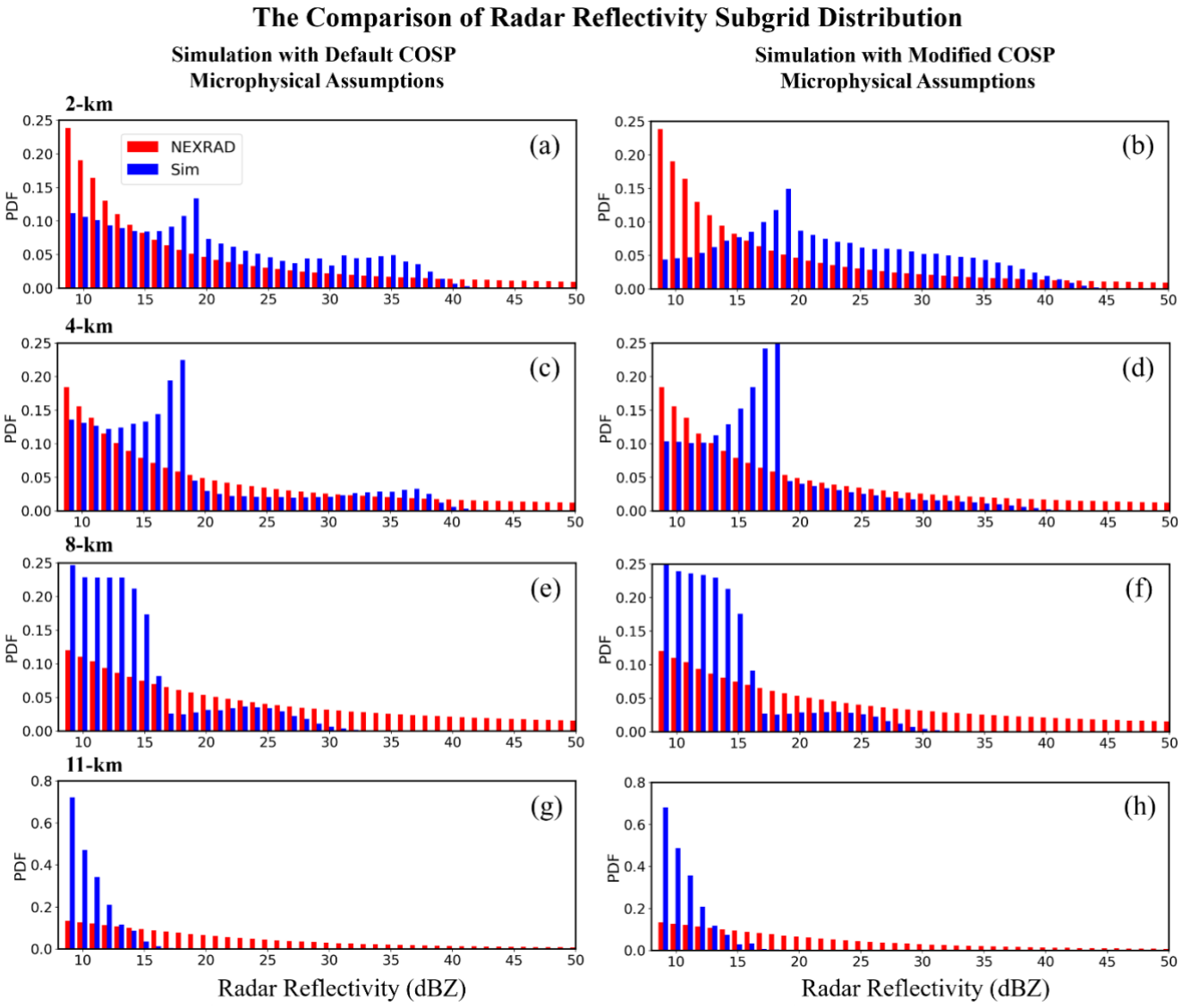

Figure 3: Comparison of radar reflectivity subgrid distribution between NEXRAD observations (red bars) and the simulations (blue bars) at the vertical levels of 2 km, 4 km, 8 km, and 11 km. Simulation results in the left and right columns are from the default microphysics assumptions in COSP and modified COSP microphysics assumptions, respectively.

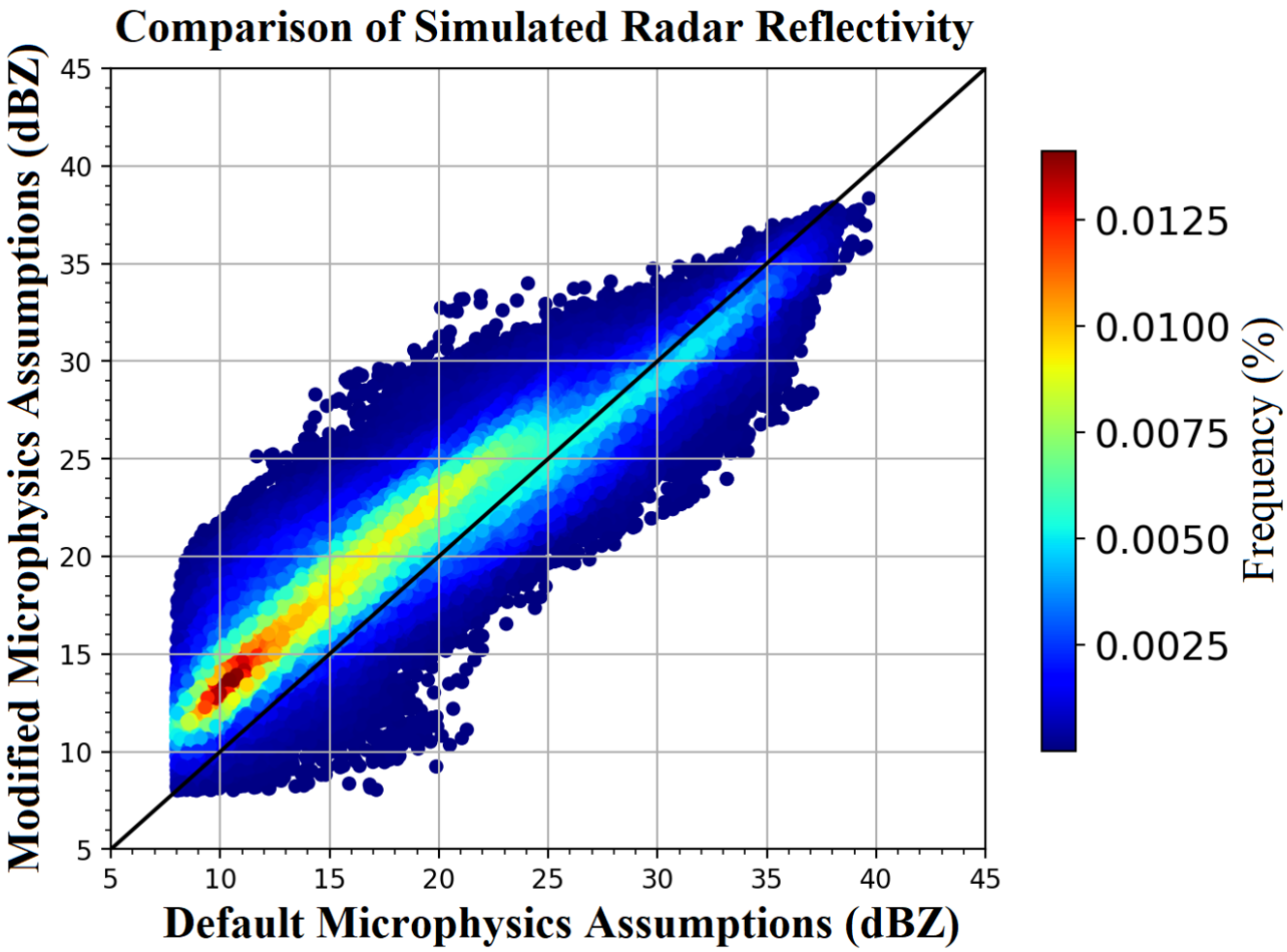

Figure 4: Scatter density plot between radar reflectivity values from the simulation with the modified microphysics assumptions (y-axis) versus those with the default microphysics assumptions (x-axis). The data shown are for April 2014. The dots are color labelled with their frequency of occurrence.

695

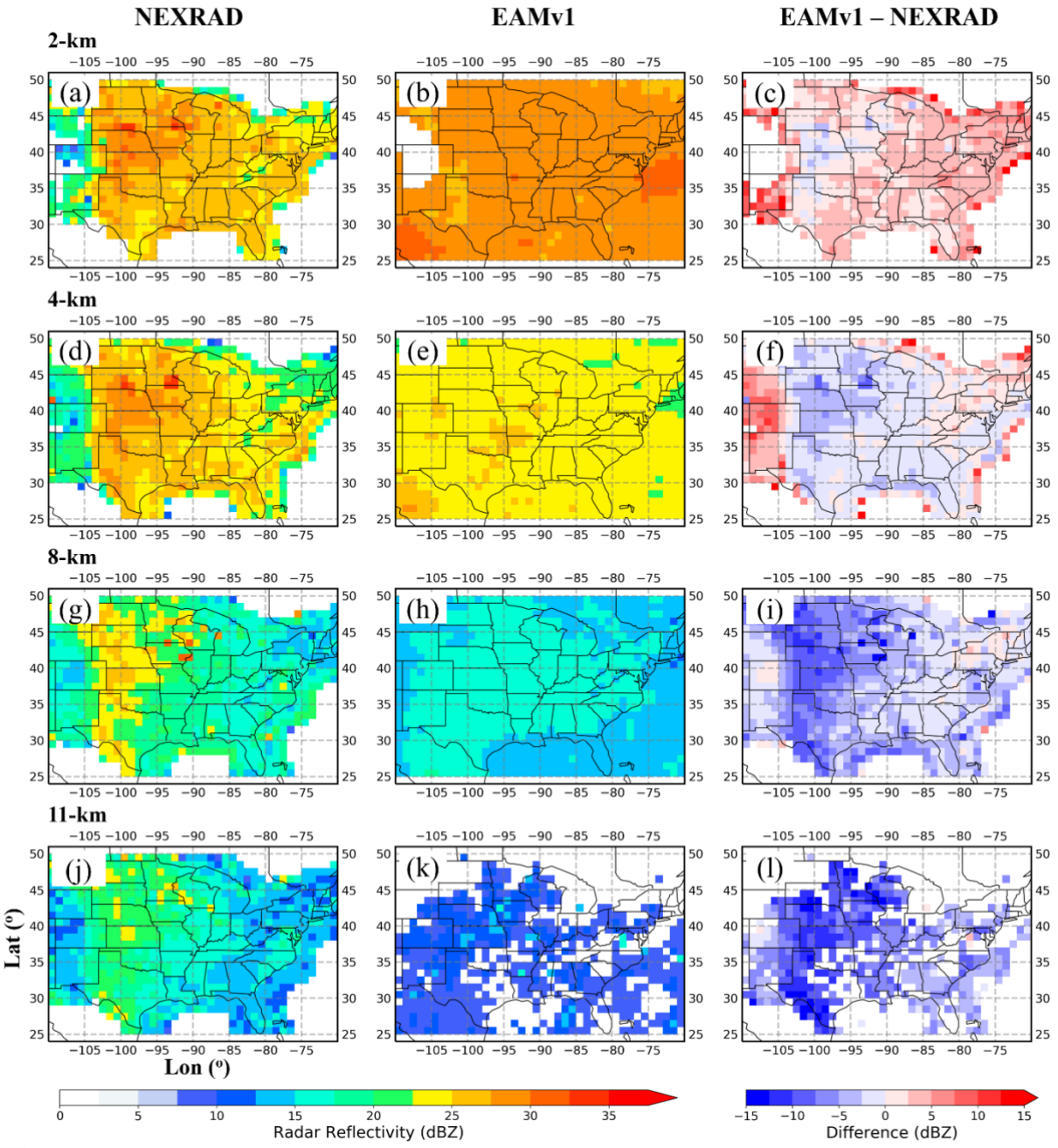

Figure 5: Plan view of radar reflectivity averaged from NEXRAD observations (a, d, g, j), EAMv1 simulation with the modified microphysics assumptions in COSP (b, e, h, k), as well as their absolute differences (c, f, i, l) at the level of 2-km, 4-km, 8-km, and

11-km altitude. The NEXRAD data are spatially averaged from native resolution to the model grid over 2014-2016 April-September period, and the simulation are vertically interpolated to the NEXRAD levels.

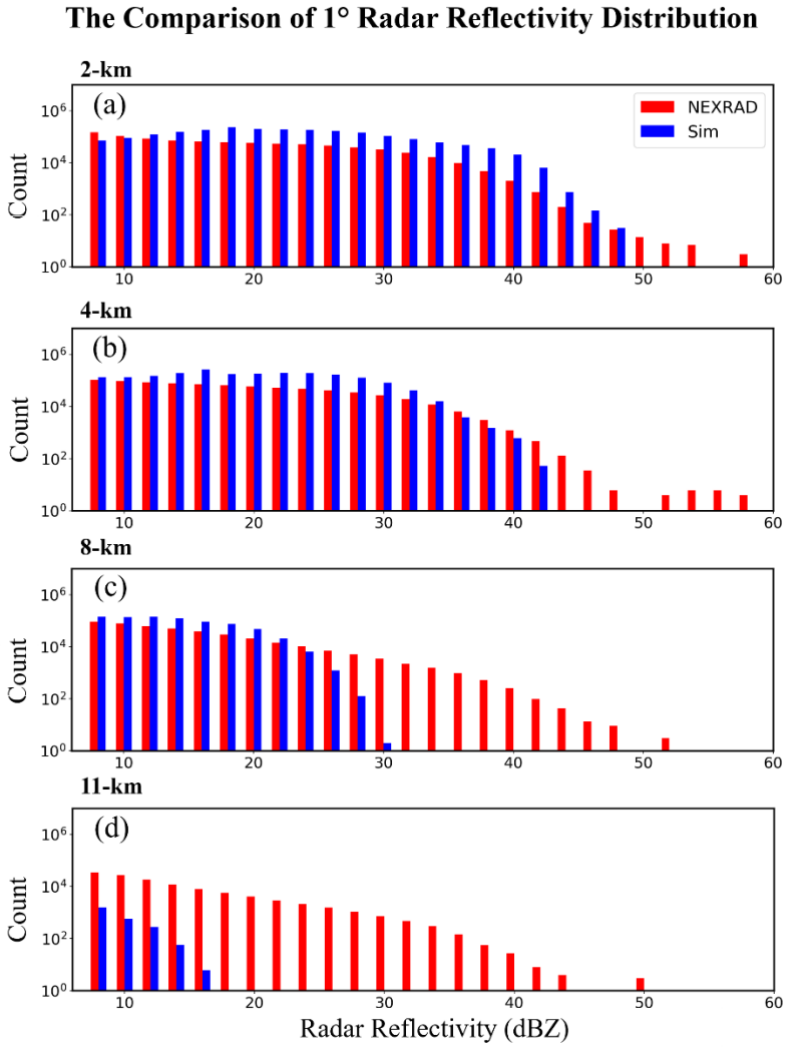

Figure 6: Comparison of radar reflectivity histograms at 1° scale between NEXRAD observations (red bars) and the simulations (blue bars) at the vertical levels of 2 km, 4 km, 8 km, and 11 km.

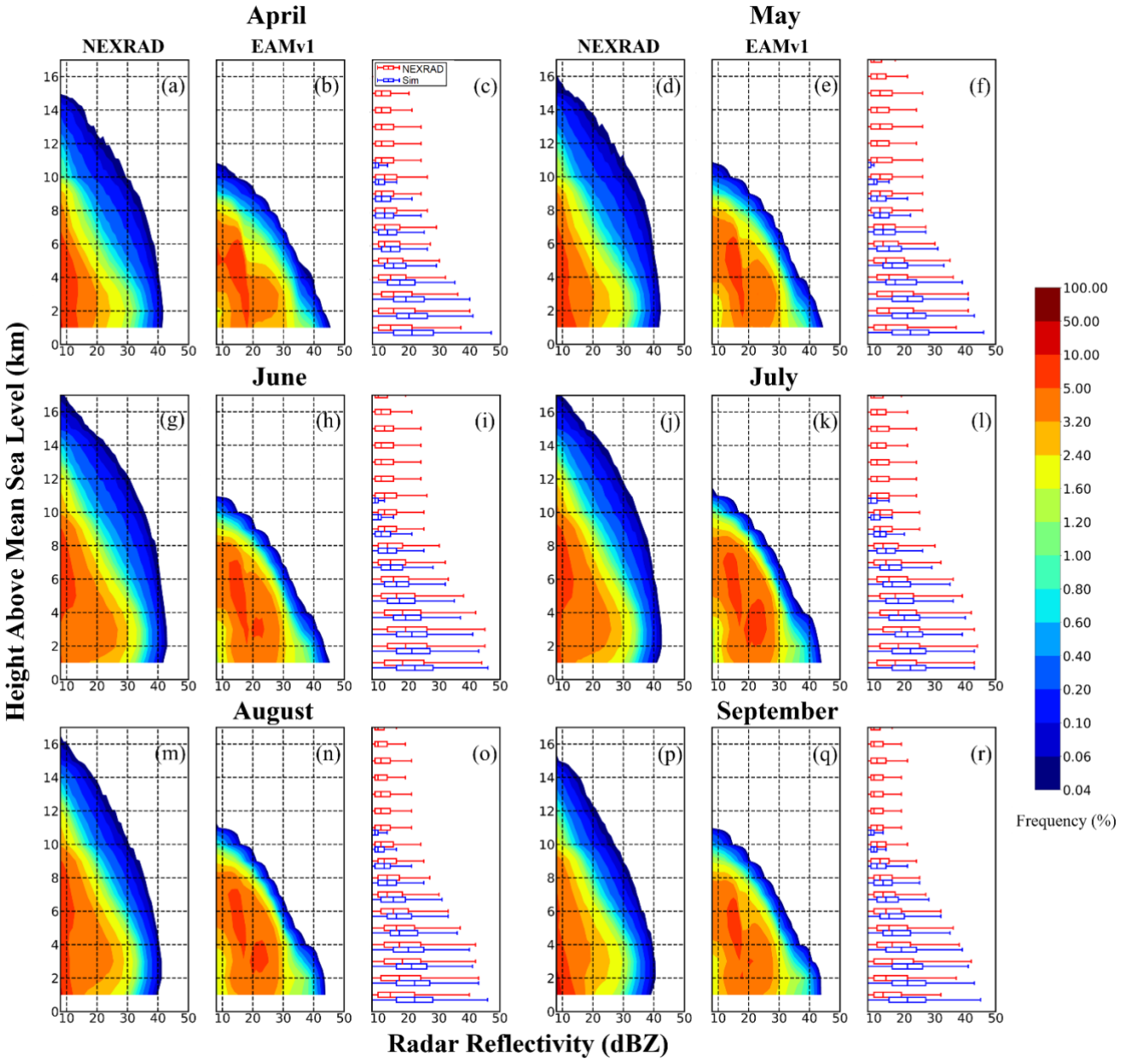

Figure 7: Contoured-Frequency-by-Altitude-Diagrams (CFADs) normalized by the total number of samples at all altitude levels for NEXRAD (a, d, g, j, m, p) and EAMv1 simulation with the modified microphysics assumptions in COSP (b, e, h, k, n, q) for the months from April to September averaged over 2014-2016 period. The box-whisker plots (c, f, i, l, o, r) for NEXRAD (red) and EAMv1(blue) are calculated using normalization at each individual level, where the center of the box represents the 50th percentile value, and the 25th and 75th percentiles are represented by the left and right boundary of the box, respectively. Whiskers correspond to the 5% and 95% values.

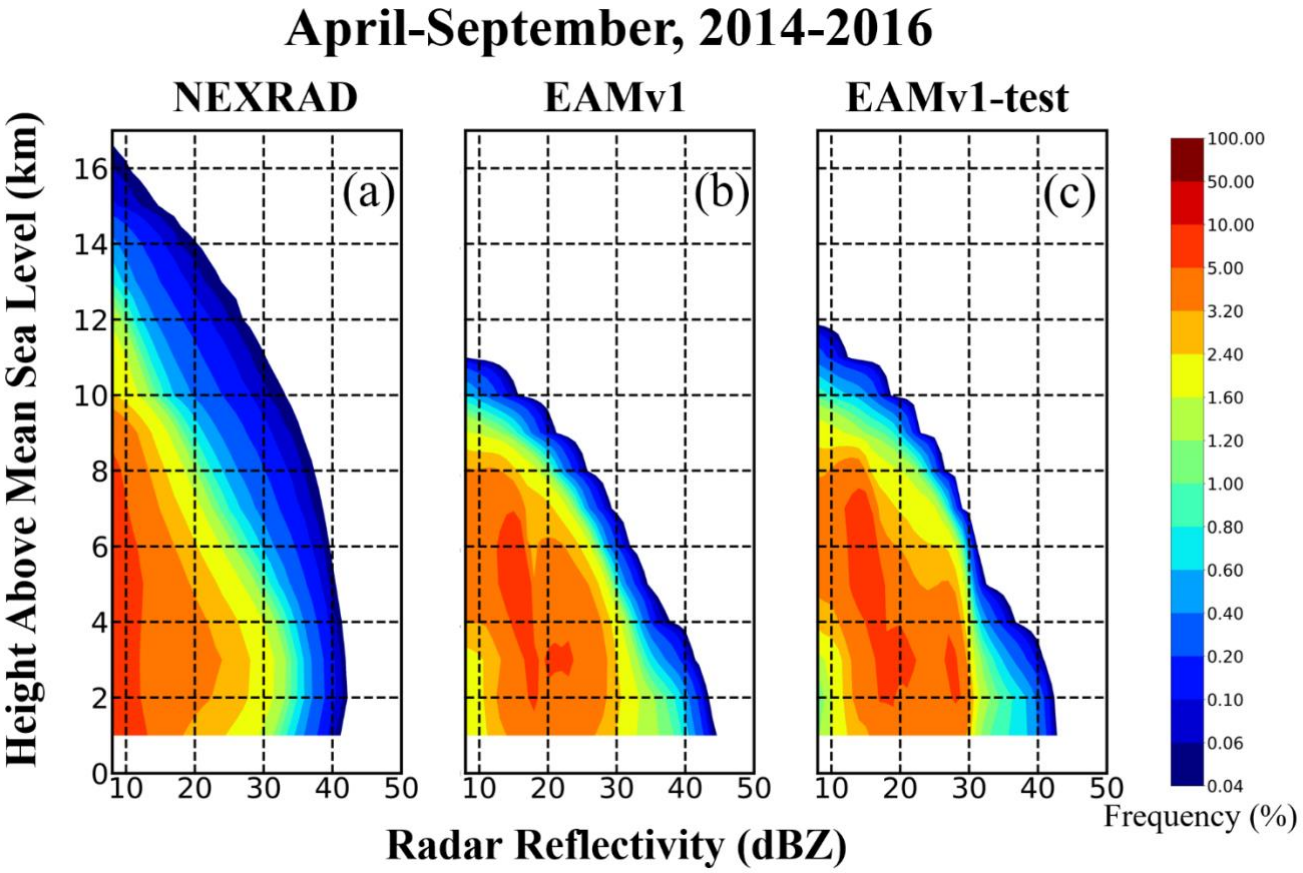

Figure 8: Comparison of Contoured-Frequency-by-Altitude-Diagrams (CFADs) for the warm seasons over 2014-2016 between (a) NEXRAD, (b) EAMv1 simulation, and (c) the EAMv1-test simulation with reduced convective entrainment rate.