Responses to Reviewer #1

This is a nice paper testing the performance of the NCAR climate model against observations from the US weather radar network. My view is this is important and that more climate models need to be tested against observations, both for current and recent climate but also studies such as this focusing on representation of key processes.

We thank the reviewer for recognizing our efforts and providing helpful suggestions and comments. Our point-by-point responses are provided as below.

Note the large scale circulation in this model is being nudged towards observations, so that the performance being discussed represents an upper bound. It would be interesting to also compare the output from an extended period of the model in free running mode as for CMIP and looking at the latter part of 20th Century and early 21st century runs so that the forcings are consistent with current observations. I suspect the key limitations outlined in the nudged runs will be at least as large and possibly greater.

Free-run simulations could cause a large bias in circulation, especially regional circulation, making the comparison with observation in convective systems difficult. We wanted to start from the nudged runs to exclude the large biases from circulation. We agree that further study can be extended to free runs. We have added this clarification in the conclusions: "Note the large-scale circulation is nudged towards observations for the simulations in this study, which represents the upper bound of model performance. Compared to the nudged simulations, the 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." We were not able to provide evaluation for a longer period, because "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 modeling for an extended period."

The analysis looks at a few metrics including the overall spatial distribution and the vertical profiles of reflectivity. This is OK as far as it goes, but I feel further and deeper analysis will yield more information on the process limitations in the model. For example, further insights would be gained by examining the mean diurnal cycle of convection and how that compares with observations over the great Plains. Does the model reproduce the night-time maxima over the eastern plains and as a difficult test are propagating modes observed modulating the diurnal convective activity (cf. Carbone and Tuttle, J Clim., 2008). Is the spatial distribution of time of peak convection at all captured or is it dominated by a morning maxima as convection triggers too early in daily heating as occurs in many simulations in the tropics. The diurnal cycle of convection is also important for the resulting cloud and radiation climatology of the model.

The precipitation including the diurnal cycle has been evaluated for EAMv1 (Zheng et al., 2019), which showed the model failed to simulate the diurnal variation of precipitation over the central United States. To avoid the redundancy, here we have added a plot for comparing the diurnal cycle of column-maximum reflectivity (Fig. 6), which can indicate the intensity of precipitation. The following text has been added to the last paragraph of Section 3.3, "As evaluated in Zheng et al. (2019), E3SM v1 failed to simulate the diurnal variation of precipitation over the central United States. Here we examine the diurnal cycle of column-maximum reflectivity (Fig. 6), which can

indicate the intensity of precipitation (Carbone and Tuttle, 2008). The observation shows two peaks, one in the early morning and the other in the late afternoon. This pattern differs from the observation of total precipitation, which only has one nocturnal peak with a smooth transition from the minimum at local noon. The difference between the two observed variables is expected, as the column-maximum reflectivity most likely represents convective (not stratiform) precipitation, which occurs significantly in the early morning and late afternoon. In contrast with the two peaks in observed column-maximum reflectivity, the EAMv1 simulation demonstrates a flat diurnal curve without any obvious peak, suggesting the model has a difficulty in simulating the convective precipitation.

In Sect 3.1, where there is a mean difference in reflectivity. Noting these are linear averages over a 100 km area, do you have a feel how much of this is associated with differing convective fractions within the grid points, differing fractions of precipitation or differences in the PDF of the reflectivities not associated with the convective/precipitating fraction? Have you compared convective fraction from the model paramaterisations with observations? Diagnostics looking at fractional cover can also aid interpretation and diagnose issues.

Thanks for the comment. We did not output the convective fraction for model grids, which prevents us from looking at the relationship of mean reflectivity with the convective fraction. Since the subgrid distributions of cloud and precipitation assumed in the COSP simulator have nothing to do with a convective fraction which is calculated by the ZM cumulus parameterization, we think the relationship of mean reflectivity with the convective fraction might not mean much.

In Section 3.3, comparing NEXRAD and subscale distributions – testing maybe could be earlier in the paper and is there is a degree of circularity in your argument since you are adjusting the sub-grid scale distributions with the observed NEXRAD data and so naturally there is increased agreement. Note that the bimodality in the original distributions shown in Fig 4 are not generally observed in nature.

We agree with the reviewer that the structure of the manuscript should be adjusted to avoid circularity. Section 3.3 has been moved to the beginning of the results.

We agree that the bimodality shown in original distribution in the left column of Figure 5 disagree with the general observation. However, this was the result from the COSP in the E3SM v1 in which cloud microphysical parameters are not aligned with the microphysics scheme used in the host model. After the correction, the distribution is more-like a Gamma distribution,

As a minor point, on line L130, using NEXRAD also simplifies the radar scattering calculations compared with GPM and TRMM with the 10 cm wavelength radar being close to Rayleigh scatter most of the time although the scattering calculations are still complex for ice habits.

We agree with the reviewer and have added a clarification at the end of Section 2.3, i.e., "For the NEXRAD observation, its 10 cm wavelength guarantees Rayleigh scattering for most situations."

Overall, I think this is a useful study, but would benefit from being taken further. It is clearly addressing important issues with climate models and as noted these kind of studies are sorely needed. The methods are clearly articulated.

Thanks.