Author Responses

We thank both reviewers for their valuable comments and helpful suggestions, which we think helped improve the manuscript. Our responses and revisions are enumerated below.

Reviewer #1 (Comments to Author):

This manuscript documents a new observation simulator for model applications, focused on ground based radar and lidar at individual locations. It documents the simulator methodology, and illustrates a case study application to mixed phase clouds over Antarctica. The paper is well written, and should be publishable with minor revisions. Specifically, there is quite a bit of qualitative analysis in figures 4-5 before getting quantitative in figure 6. There is also discussion of, but no demonstration of the phase ratio. In particular, estimating the phase ratio from lira and radar properties when the 'truth' for the model exists is a bit strange, and doesn't seem to be the point of a simulator. But at least illustrate the method, and how it might differ or reproduce model 'truth'. I also think the description could be tightened up in the introduction. But it is basically a good paper, suitable for GMD with these minor revisions, echoed in specific comments below.

We understand the reviewer's interest in the 'truth' for the model. The main problem with using the 'truth' (e.g., the distribution of condensate mass and related metrics) is that even if it is known for a given model output, it introduces an inverse problem (e.g., retrieving an atmospheric state variable from the observations), and hence, cannot be consistently evaluated using observations. Instrument simulators convert this model evaluation problem into a forward problem (generating observed variables using model output fields), which provides a consistent definition between different models and between models and observations, thereby enabling direct comparisons to be made (see also the discussion in Hillman et al., 2018).

We already emphasize the importance of making direct comparisons using similar methodologies and metrics at the end of sect. 3.3 (after discussing fig. 8 in detail):

"Fig. 8 demonstrates the sensitivity of phase ratio statistics to the classification method, the viewing direction of the examined instrument, and the method by which "liquid" and "non-liquid" or "ice" classes are being counted. It shows that the use of forward simulators alone is not a guarantee for an "apples-to-apples" comparison, which requires matching processing steps to ensure its robustness."

Figs. 4 and 5 are indeed qualitative, but we think that it is, nonetheless, important to show them as they enable the reader to perceive the ability of EMC^2 to produce forward-calculated variables comparable to observations. While it is also true that fig. 6 is more quantitative (no color scale), the general impression concerning a model's output behavior observed through the simulator transfer function is valuable, especially for potential users who wish to have the option of making such quick qualitative evaluations.

Our understanding is that in "no demonstration of the phase ratio", the reviewer refers to the mass phase ratio calculated using the raw model output. To address this comment, we added the mass phase ratio from the raw ModelE3 SCM output and wrote a paragraph discussing it per our arguments above:

"We note that the treatment of the COSP emulator's "undefined" subcolumn bins as "ice" to produce phase ratio statistics leaves the impression that only ice hydrometeors exist below cloud top. However, a rather different impression of mostly liquid water dominance, though not as stark as in the radar-sounding method using the radiation approach, is perceived when the mass phase ratio calculated using the raw SCM output is examined (fig. 8, lower-left). Contrary to the COSP emulator, treating "undefined" bins as "ice" in the modified fixed lidar threshold method increases its apparent frequency phase ratio agreement with the mass phase ratio in multiple time-height bins. Phase classification depends on instrument measurement characteristics and limitations and hydrometeor properties such as their class, relative mixture with other hydrometeor classes, as well as their size distributions. Therefore, such an apparent agreement between different variables and phase occurrence metrics, as well as between the same variables and metrics based on different instruments and/or methodologies, should be taken with a grain of salt (cf. Cesana et al., 2021; see also Silber et al., 2021c)."

Specific Comments:

Page 1, L16: What is a cosp Lidar simulator emulator?

COSP is a package of multiple simulators, each imitating a different remote-sensing satellite instrument. One of these simulators is a lidar simulator imitating measurements made by the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) on-board the CALIPSO satellite. We didn't implement the actual COSP lidar simulator FORTRAN code, but instead added an emulator of that simulator; hence, the "COSP lidar simulator emulator".

Page 1, L17: So if Python it has to be run offline? Is it open source and publicly available? You never state that in the text. At least note it.

We note both explicitly and implicitly that EMC^2 is an open-source Python code in multiple locations in the text:

1. First, in the actual name of the package (Earth Model Column *Collaboratory*).
2. Abstract:
   "Here we present the Earth Model Column Collaboratory (EMC^2), *an open-source* ground-based lidar and radar instrument simulator and subcolumn generator…"
3. The final paragraph in the Introduction:
   "Here we present the Earth Model Column Collaboratory EMC^2, *an open-source* ground-based lidar and radar simulator and subcolumn generator."
4. Final sentence in the Summary section:
   "We invite the community to take advantage of the framework provided by EMC^2 and to contribute to its further development and applications."

We also mention that EMC^2 is currently an off-line simulator (sect. 2.4.2): "Noting that EMC^2 operates off-line, hydrometeor class fall velocities …"

Page 2, L19: Provide an example? Suzuki et al 2015 and/or Bodas-Salcedo et al 2013 maybe?

We added references for these two papers to the Introduction:

"Meaningful model evaluation benefits from a direct ("apples-to-apples") comparison with observations (e.g., Bodas-Salcedoet al., 2014; Suzuki et al., 2015)."

Page 2, L21: But retrieving microphysics is not the point. Microphysics in the model is used to simulate the observable (e.g. reflectivity pro backscatter or extinction).

We modified that sentence to make it clearer:

“However, model evaluation is challenging because of observational detectability constraints (e.g., signal extinction), and lack of retrievals or large uncertainties in some microphysical and atmospheric state quantities by these instruments, for example, hydrometeor number concentration or water content.”

Page 4, L1: Note #1 and #2 for the rad and micro approaches.

We removed that sentence from the text as part of our response to the reviewer comment on Page 21, Figure 6, so this comment is no longer relevant.

Page 8, L6: Can you list all the tuning parameters?

Unfortunately, no, because this is not a ModelE3 descriptive paper or a user manual. Model tuning is discussed in the GISS-E2.1 article describing ModelE3’s predecessor, but the tuning procedure and tuning parameters are quite different in the case of ModelE3. Several tens of ModelE3’s tuning parameters will be provided to readers as part of an upcoming manuscript describing ModelE3’s machine-learning tuning effort (Elsasser et al., in prep).

Page 11, L9: This is strange to simulate. The model knows explicitly what it’s hydrometers are. Why try to approximate them from the radar/Lidar? Applying this to observations would be useful, but the point of a simulator is to cast the model in observation space…I assume you will compare outputs to model hydrometers l(truth) later?

This comment was largely addressed in our response to the general reviewer comment. We added text following the sentence noted by the reviewer to motivate using the classification masks in accordance with our response above:

“Once the total lidar and/or radar variables are calculated, EMC^2 can be used to classify the subcolumn simulator output. Classification masks can serve as tools for direct comparisons between the simulator output and observational data utilizing similar classification methodologies, some of which can be used to calculate water phase ratios…”

Page 12, L18: But what is the emulator? How is trained? Unclear what this COSP emulator is. Please add a few sentence description.

We added a few sentences describing the COSP emulator:

“The emulator of the COSP lidar simulator follows the same equations and logic of the on-line lidar simulator (Cesana and Chepfer, 2013) implemented in numerous climate models. In short, the attenuated total backscatter (ATB) calculated in the COSP emulator routine while assuming $\eta = 0.7$ is used to calculate the lidar scattering ratio (the ratio of total to molecular attenuated backscatter) for the detection of hydrometeors in subcolumns by selecting scattering ratio values larger than 5. Calculated cross-polar ATB as a function of the total ATB is then used to classify the detected hydrometeors into liquid or ice, based on an empirical phase discrimination line. As the last step of this classification method, hydrometeors below (top-down lidar view) or above (bottom-up lidar view) a subcolumn bin with scattering ratio larger than 30 are classified as "undefined".”

Note that an emulator, by its definition in the Oxford Advanced Learner’s Dictionary, does not require training. We now clarify that at the first instance of the word emulator in the main text (sect. 2.5):

"… and the COSP lidar simulator emulator (henceforth referred to as the COSP emulator; 'emulator' is used here in its generic sense rather than a machine-learning context, and there is no training involved)."

Page 12, L30: Where is the model class with data in the figure?

We do not understand the question here. The model output data are loaded and stored in the Model class object as already shown in fig. 1's flowchart and noted in the text:

"Once loaded through the Model class internal methods, model output data are stored within the Model object …"

Page 14, L5: Would these model specific things be in the instrument class or model class? Seems like they should be in the model class? Please clarify.

That depends on whether the specifications are for instrument and measurement characteristics (Instrument class) or model output field names and cloud scheme logic (Model class). All of this information is already specified in the text above the sentence noted by the reviewer, for example:

"The Model class contains model output field namelists and default hydrometeor parameters (Table 1)."

"… the Instrument class contains relevant information about the instrument being simulated (some of which is listed in Table 2) as well as the single-particle and bulk scattering calculation LUTs …"

Page 17, L10: This is a good motivational sentence that I'm not sure was well reflected in the introduction, either to the whole paper or just section 4. Suggest stating this succinctly earlier.

That is an excellent suggestion. We revised the final paragraph of the Introduction:

"Here we present the Earth Model Column Collaboratory (EMC^2), an open-source ground-based lidar and radar simulator and subcolumn generator, which is designed to operate over large-scale model output while being faithful to the physics implemented in models' microphysics or radiation schemes but can also be applied to high-resolution model output. EMC^2 enables detailed evaluation of atmospheric thermodynamic profile and cloud properties extracted from local, regional, and global simulation outputs against long-term ground-based, air- or space-borne datasets."

Page 17, L22: Please state why you know it's the autoconversion parameterization at least.

We ran LES using the same model configuration except for the autoconversion parameterization, which was switched between Khairoutdinov and Kogan (2000) and Seifert and Beheng (2001). That is the reason we state "(not shown") at the end of the previous sentence:

"This reduced amount of rainwater in the SCM simulation relative to rainwater dominance in the simulations of Silber et al. (2019a) is largely the result of the different autoconversion parameterization schemes implemented in ModelE3 (Seifert and Beheng, 2001) relative to that implemented by default in the DHARMA LES (Khairoutdinov and Kogan, 2000), which produces significantly smaller rain mass mixing ratios in this case (not shown), …"

Page 18, L3: Good agreement is not quantitative. See comment on figure 4: skip to figure 6 and use mean profiles to make the statement quantitative please.

We modified this sentence to tone it down a bit by using the adjective "apparent":

"The processed LES output exhibits good apparent agreement with the observations, …"

Page 19, Figure 4: what do the white lines represent in the EMC2output?

We suspect that the reviewer refers to the apparent white vertical line in each of the microphysics approach panels. Because we set the number of subcolumns to 100 and some of the hydrometeor class fractions are quite consistently closer to 0.99 (fig. 3), only 99 subcolumns are to be allocated with hydrometeors. Because EMC^2 uses the maximum-random overlap approach in allocating hydrometeors to subcolumn bins (i.e., hydrometeor clusters are extended vertically), the hydrometeor-free subcolumn in the resultant subcolumn snapshot shown in fig. 4 appears as if there is a white line in the plot.

Figure 6 is much better than Figure 4. Maybe showing one row to orient the reader. But then show Figure 6 and make discussion more quantitative.

We discussed the motivation to have both figs. 4 and 5 in addition to fig. 6 in our response to the general comment. We think that having both figs. 4 and 6 condensed into a single figure would make it much more difficult for readers to orient themselves in the condensed figure, and hence, we prefer keeping these two figures as is.

Page 19, L2: Again: reasonable is not quantitative.

We agree that the word reasonable is not quantitative. We reworded that sentence in a way similar to our response to the previous comment on Page 18, L3:

"Evaluation of heights with full cloud cover, indicated by the large β_{p_tot} and α_{p_tot} values (see also fig. 3, right), suggests that the SCM has an apparent reasonable agreement with the observations there."

Page 20, Figure 5: Here having time mean profiles and temporal standard deviations would be more effective at discerning differences than the color scales you have chosen. Maybe just skip this figure and clean up figure 6?

We addressed this comment in our previous responses above.

Page 20, L7: appear to agree. Again, not quantitative.

That is correct, it is not quantitative but qualitative yet not definite, and therefore, we do not see a reason to further modify the text here.

Page 21, Figure 6 basically has all the information of Figure 4 and 5. Maybe show only one figure or one row of fugue 4 and discuss this figure instead. Add empirical approach? Would be nice to compare the model approaches on the same plot. Since only one line is different between top and bottom rows, I suggest you combine the plots to one for each variable, with all EMC2 methods on the same plot.

The empirical approach was originally implemented in EMC^2 to serve as a reference for evaluating the implementation of the radiation and microphysics approaches (whether the simulator output is "in the ballpark"; in other words, bug fixes). However, the empirical approach is deficient in its consideration of a given model physics as well as the limited flexibility of empirical parametrizations, which are generally derived for specific instrument operating wavelengths and certain geographical regions and/or conditions; hence, using this approach could impact comparisons as the simulator output is confounded by the utilized parameterizations. Therefore, after contemplating the reviewer's comment and some of the comments received from the second reviewer, we eventually decided to omit the empirical approach from this article (that is, from the text and figures 4 and 5) and focus on the two main approaches faithful to model physics. We note that the empirical approach is turned off in EMC^2 by default, and hence, most likely won't be used by an end-user.

Page 22, L7: Which method and result is more correct relative to the observations?

We think that an answer to this question would be somewhat subjective, as it depends on which metric does one harness to evaluate the methods in use. Moreover, we need to be careful here because stating which method is "better" would be misleading, as the choice of which method to use should be based on research objectives and associated assumptions.

Page 22, L15: Why is empirical approach not in figure 6? Remind the reader if there is an already stated reason.

We addressed this comment as well in our response to the comment on Page 21, Figure 6.

Page 24, L12: What is the COSP emulator exactly? What is the method? It's not the same code as the on line COSP simulator, so how is COSP emulated? This is not explained.

We addressed this comment as well in our responses to the reviewer comments on Page 1, L16 and Page 12, L18.

Page 25, Figure 8: Why is the vertical resolution of the on-line COSP simulator different than the COSP emulator? Where do you get more vertical resolution?

The COSP simulator has a coarser vertical resolution than ModelE3's vertical spacing within the troposphere. We already note that in the text when describing the COSP simulator emulator in sect. 2.5:

"Note that, unlike the on-line COSP lidar simulator, this emulator operates using the model vertical levels and does not interpolate the model output onto an evenly-spaced vertical grid."

Page 25, L1: I'm not seeing phase ratio statistics in figure 8 as mentioned here. Please clarify what you mean. See general comment: showing the phase ratio using model truth and EMC2 processed data would be good.

We addressed this comment in our response to the general reviewer comment. We wish to emphasize that phase ratio is a rather general name for various metrics describing the occurrence of liquid or ice hydrometeors relative to one another (e.g., frequency phase ratio, mass phase ratio). The fact that different metrics are occasionally treated as comparable suggests that cross-evaluation of different phase ratio results from the literature should be made with caution (see also Cesana et al., 2021, and the discussion in Silber et al., 2021).

Appendix B is never really referred to in the text (looking at different model configurations) and should be moved to the main text with the figure (should be B1, not A2) added to the main text (maybe remove figure 4 & 5).

We thank the reviewer for noticing that the figures for Appendix B were tagged as if they belonged to Appendix A. This error is corrected now. To keep the case study example concise, we prefer to keep the discussion on the ModelE3 configurations in Appendix B.

We note that Appendix B is already referred to in the second sentence of sect. 3.3:

"The SCM using this configuration … produced the best agreement with the observations and the LES relative to the other three configurations (see Appendix B)."