Dear Editor,

Please find there our responses to the reviewers' comments. Our responses are in blue.

Sincerely,

Zhipeng Qu

Referee #1 (Toshi Matsui)

Summary: This manuscript describes generation and validation of numerical atmosphere-surface test frames for upcoming EarthCARE satellites. Authors conducted multiple multi-scale modeling to generate high-resolution scenes of virtual atmosphere and surface across different regions. The manuscript, then, qualitatively validated generated virtual fields against the satellite observations. This is a very ambitious paper that intends to validate this many different scenes all together in a single manuscript. This is why the validation method lacks detailed quantitative analysis and non-evident arguments in ice microphysics. Other problems include lack of citations related to previous studies and lack of aerosol validation. Thus, this paper needs "major revision" to resolve aforementioned problems before publishing in AMT. Please read the major and minor comments below.

Thank you for your review and suggestions. We agree that more quantitative verification of the test frames would improve the quality of the manuscript, and some have been included in the revised version, but we also stress that this was not, at this point, the purpose of the manuscript. This is elaborated in our responses to subsequent points.

Major Comments:

1) Lack of citations and discussion in the previous studies.

This numerical satellite mission frame has emerged from the last decades. We have developed a benchmark of numerical test frames before the launch of Global Precipitation Measurements (GPM) Core satellite.

Matsui, T. T. Iguchi, X. Li, M. Han, W.-K. Tao, W. Petersen, T. L'Ecuyer, R. Meneghini, W. Olson, C. D. Kummerow, A. Y. Hou, M. R. Schwaller, E. F. Stocker, J. Kwiatkowski (2013), GPM satellite simulator over ground validation sites, Bull. Amer. Meteor. Soc., 94, 1653–1660. doi: http://dx.doi.org/10.1175/BAMS-D-12-00160.1

For this, we have validated numerical simulation against the various measurements from the GPM field campaigns.

Iguchi T., T. Matsui, J. J. Shi, W.-K. Tao, A. P. Khain, A. Hou, R. Cifelli, A. Heymsfield, and A. Tokay (2012), Numerical analysis using WRF-SBM for the cloud microphysical structures in the

C3VP field campaign: Impacts of supercooled droplets and resultant riming on snow microphysics, Journal of Geophysical Research, 117, D23206, doi:10.1029/2012JD018101.

Iguchi, T., T. Matsui, A. Tokay, P. Kollias, and W.-K. Tao (2012), Two distinct modes in one-day rainfall event during MC3E field campaign: Analyses of disdrometer observations and WRF-SBM simulation. Geophysical Research Letters, 39, L24805, doi:10.1029/2012GL053329.

Iguchi, T., T. Matsui, W. Tao, A. Khain, V. Phillips, C. Kidd, T. L'Ecuyer, S. Braun, and A. Hou, 2014: WRF-SBM simulations of melting layer structure in mixed-phase precipitation events observed during LPVEx. J. Appl. Meteor. Climatol. 53, 2710-2731, doi:10.1175/JAMC-D-13-0334.1.

There should be a lot more papers related to these topics. Please search manuscripts and discuss similarities and differences between your work and other previous work in the introduction section.

Thank you for these suggestions. New text and references have been added to the conclusion and perspective section. While we agree with the Reviewer that verification of the test frames using independent sources of information is a worthy and very demanding task, we emphasize, in the revised version, that the test frames, as described in this manuscript, were produced expressly for end-to-end assessment of EarthCARE's algorithms and its data-handling procedures. While the frames had to be "reasonably realistic" and sufficiently large and demanding for retrieval algorithms to be tested rigorously, they were never expected to be "perfect facsimiles" and the manuscript was never intended to focus on verification of the frames via use of independent observations. In fact, given the size of the test frames, extensive verification of them would be extremely challenging and in all likelihood unsatisfactory. We tried to give a cursory indication.

"Previous studies have simulated atmospheric conditions for purposes of satellite algorithm development and evalua-tion (e.g., MPB Technologies Inc. 2000; Voors et al. 2008; Tao et al. 2009), constraint of cloud microphysical schemes by observations (Matsui et al. 2013, Iguchi et al. 2012a, 2012b, 2014), and assimilation of retrieved aerosol properties into Numerical Weather Prediction (NWP) models (Zeng et al. 2020; Cornut et al. 2023), but the simula-tions done for this study had to serve several purposes simultaneously, and this put unique demands on them. Most notably their size, for they had to provide sufficient detail to meet several wide-ranging aspects of observation simulation and algorithm assessment, with enough areal extent to evaluate data processing and archiving proce-dures."

"As such, the overarching requirement placed on the time-sensitive production of these test frames was that they be deemed, by myriad mission researchers and managers, "sufficiently" realistic and "necessarily" expansive enough to provide adequate assessment of the numerous key steps that will be required to produce EarthCARE data. No doubt, this requirement compromised some aspects of both the quality of the simulations and their verification against independent sources of information. Moreover, efforts are being made to improve upon these test data. In particular, the two-moment bulk cloud microphysics scheme Predicted Particle Properties (P3) (Morrison and Mil-brandt 2015; Milbrandt and Morrison 2016; Cholette et al. 2019; Qu et al. 2022) is being used."

2) Lack of quantitative validation.

This paper intends to validate many different scenes that contain many different types of clouds and microphysics. All validation here is qualitative “eyeballs” validation. Thus, the results are stated “very good (line 209). At least, I suggest creating histograms or PDFs for each plot (Figs. 8, 9, 11, 12, 14) for more quantitative discussion in addition to existing qualitative validation. You can see Fig 1d in Matsui et al. 2014 for example.

Matsui, T., J. Santanello, J. J. Shi, W.-K. Tao, D. Wu, C. Peters-Lidard, E. Kemp, M. Chin, D. Starr, M. Sekiguchi, and F. Aires, (2014): Introducing multisensor satellite radiance-based evaluation for regional Earth System modeling, Journal of Geophysical Research, 119, 8450–8475, doi:10.1002/2013JD021424.

Thank you for this suggestion. We added PDFs to figure 8, 9, 11, 12, 14 and 15. Additional discussions are also added in the main text.

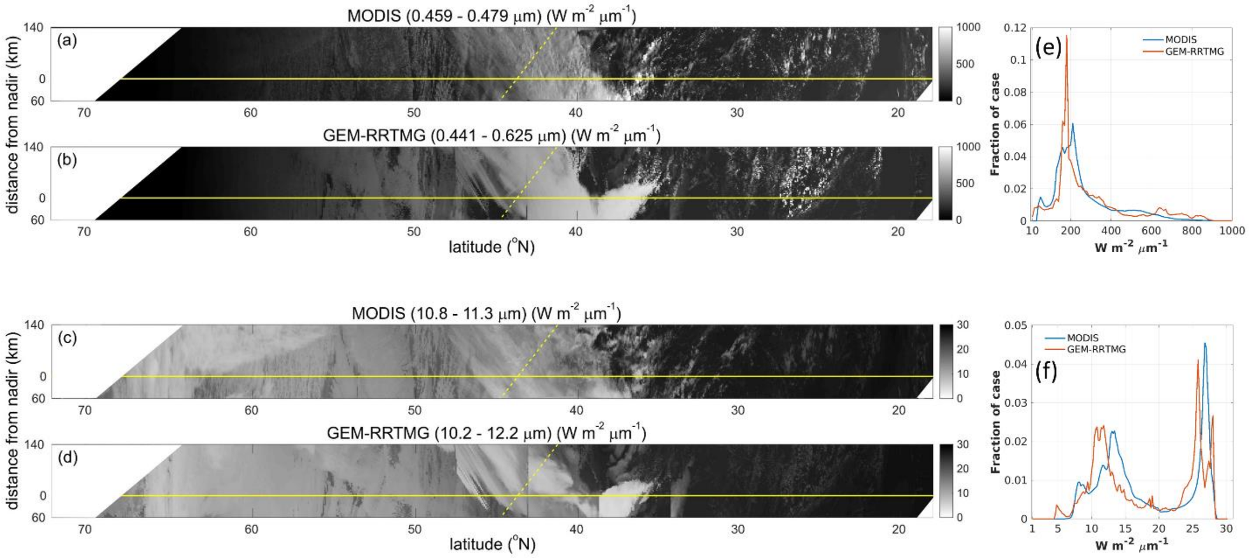

Figure 8: (a) MODIS TOA flux (approximated by radiance*□) for band 3 (459 - 479 nm) between 17h15 and 17h35 UTC on 2014-12-07. (b) RRTMG simulated upward TOA flux for 441.5 - 625 nm for GEM’s simulation of the Halifax frame. (c) as in (a) but for band 31 (10.8-11.3 □m). (d) in as but for wavelengths 10.2 - 12.2 □m. Solid and dashed yellow lines indicate EarthCARE’s and CloudSat’s nadir-tracks. Blank areas are outside MODIS’s field-of-view. (e) frequency distributions of fluxes for band 3 (bin size of 10 W m-2 μm-1). (f) as in (e) but this is for band 31 (bin size of 0.2 W m-2 μm-1).

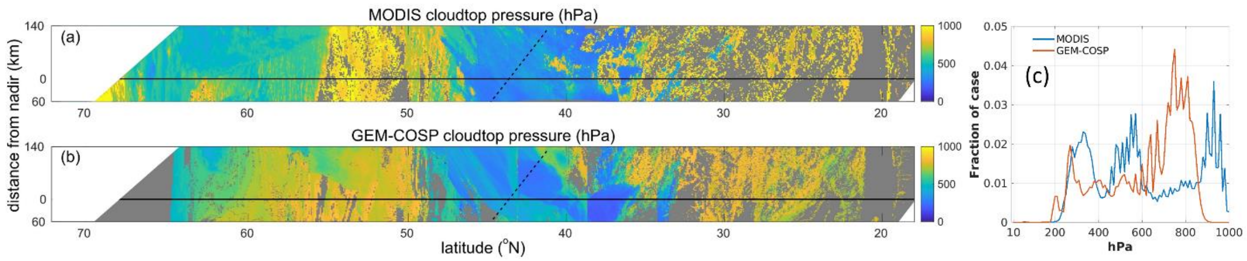

Figure 9: (a) MODIS cloudtop pressure (MYD06_L2 product) between 17h15 and 17h35 UTC on 2014-12-07. Blank areas are outside MODIS's field-of-view. (b) GEM's cloudtop pressure for the Halifax frame based on COSP's MODIS simulator. Grey area in the northern portion has $\theta_0 > 90°$ and so no COSP values. (c) frequency distributions of cloudtop pressure (bin size of 10 hPa).case with regard to cloudtop pressure (bin size of 10 hPa is used).

"Figure 8a and c show MODIS spectral fluxes (MYD02HKM product; MCST 2017a) for 0.459 - 0.479 µm and 10.8 - 11.3 µm for the Halifax frame. Key cloud-related features are a cold front between 40°N and 45°N, scattered clouds to its south, and mostly overcast conditions to its north. Figure 8b and d show TOA spectral fluxes for two wavebands, close to MODIS's bands, as simulated by the Rapid Radiative Transfer Model for GCMs (RRTMG - Mlawer et al. 1997; Iacono et al. 2000; 2008) using GEM data. At large-scales, GEM did well with respect to cloud occurrence. Figure 8e and f show distributions of visible and infrared spectral fluxes, respectively. While the distributions of fluxes derived from observations and models follow similar patterns, there are some notable differences in the imagery. For the GEM scenes, discontinuities, stemming from the stitching together of the semi-independent high-resolution inner-most domains, are clearly visible across the frontal system. They do not pose a serious problem for the task at hand."

"Near 38°N GEM's longwave fluxes are significantly less than MODIS's. This is because GEM simulated widespread convection in this area whereas MODIS only observed isolated convective cells. This is also evident in Figures 8e and f as GEM shows higher frequencies around 800 W m^{-2} μm^{-1} and 5 W m^{-2} µm-1, respectively. This is also apparent in Figure 9, which shows cloudtop altitudes both inferred from MODIS radiances (Platnick et al. 2015) and comput-ed by the MODIS simulator of the Cloud Feedback Model Intercomparison Project (CFMIP) Observation Simulator Package (Bodas-Salcedo et al. 2011; abbreviated as the COSP simulator)."

"GEM's cloudtop altitudes are too high for low clouds between latitude 20°N and 30°N and 50°N and 55°N; most are near 750 hPa, whereas MODIS's values are mostly near 920 hPa. This can also be inferred from Figure 8f in which higher frequencies of infrared spectral fluxes from GEM are found between 23 and 25.5 W m-2 µm-1 for the southern section and between 10 and 12 W m-2 µm-1 for the northern section. Additionally, Figure 9c shows that GEM under-estimates the amount of mid-level clouds between 500 and 600 hPa in the region between latitude 55°N and 62°N."

“Figure 11 compares MODIS TOA fluxes to those computed by RRTMG acting on GEM data for the Baja frame. As with the Halifax frame, agreement is generally good, though GEM’s fields exhibit some peculiarities. For instance, GEM’s fluxes associated with clouds are less variable than MODIS’s; especially between 40°N and 50°N. This could be due to both GEM’s clouds being simply too homogeneous due to missing mesoscale forcing (Stensrud and Gao, 2010) or RRTMG’s use of 1D radiative transfer models (Barker et al. 2017). Also, the thin high clouds near latitude 32°N, which are also evident in Figure 12 and positioned well in space, show an on-off pattern that is not seen in the observations. Furthermore, near latitude 55°N GEM failed to produce the very thin, but extensive, clouds below 800 hPa. This is most apparent in Figure 12. GEM’s overestimation of cloudtops close to 400 hPa near latitude 50°N is consistent with Figure 12c and Figure 11e and f which show significant overestimations of fluxes between 620 and 730 W m^{-2} μm^{-1} for visible band and between 7 and 10 W m^{-2} μm^{-1} for infrared band. Note too, that the discontinui-ties that stem from stitching together GEM’s innermost domains are less apparent for this frame than they are for the Halifax frame, though the discontinuity near 26°N is notably bad for it stands out in both visible and IR imagery.”

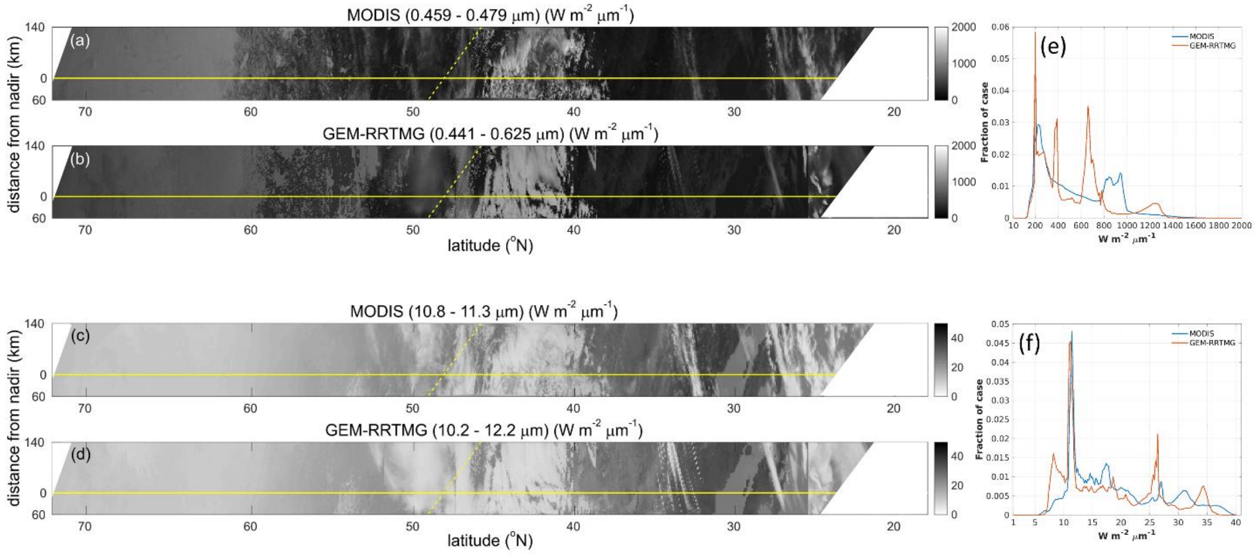

Figure 11: As in *Figure 8* except these are for the *Baja frame*. MODIS TOA fluxes are observed between 20h10 and 20h30 UTC on 2-Apr-2015

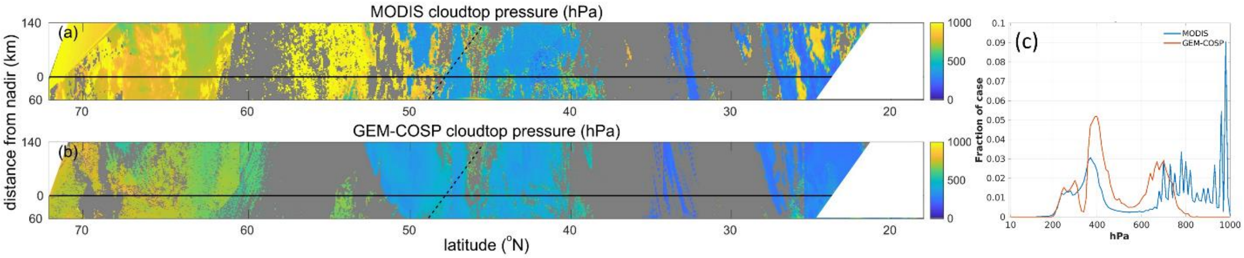

Figure 12: As in *Figure 9* except these are for the *Baja frame*. MODIS cloudtop pressure are retrieved between 20h10 and 20h30 UTC on 2-Apr-2015.

“Figure 14 shows that for the Hawaii frame, GEM’s positionings and approximate intensities of cloud systems near the Equator and ~25°S agree well with the MODIS observations. The harsh discontinuity in GEM’s string of inner-most domains near 2°S is due to a lack of high ice cloud, as seen in Figure 15, which likely stems from the lack of information, in the form of reduced outflow of high cirrus, coming into the sub-domain from the equatorial mesoscale system. Likewise, near 15°N the lack of upper-level cloud in GEM could be because this sub-domain was too disconnected from the mesoscale system to the south. The lack of high cloud in the simulation can be inferred from Figure 14e which shows an overabundance of fluxes by GEM near 200 W m^{-2} μm^{-1}; a value that resembles TOA visible fluxes from ocean surface. This is also seen in Figure 14f and Figure 15c. Again, however, the point of this section is to show the gross verisimilitude of the test frames and hence their suitability for EarthCARE algorithm assessments.”

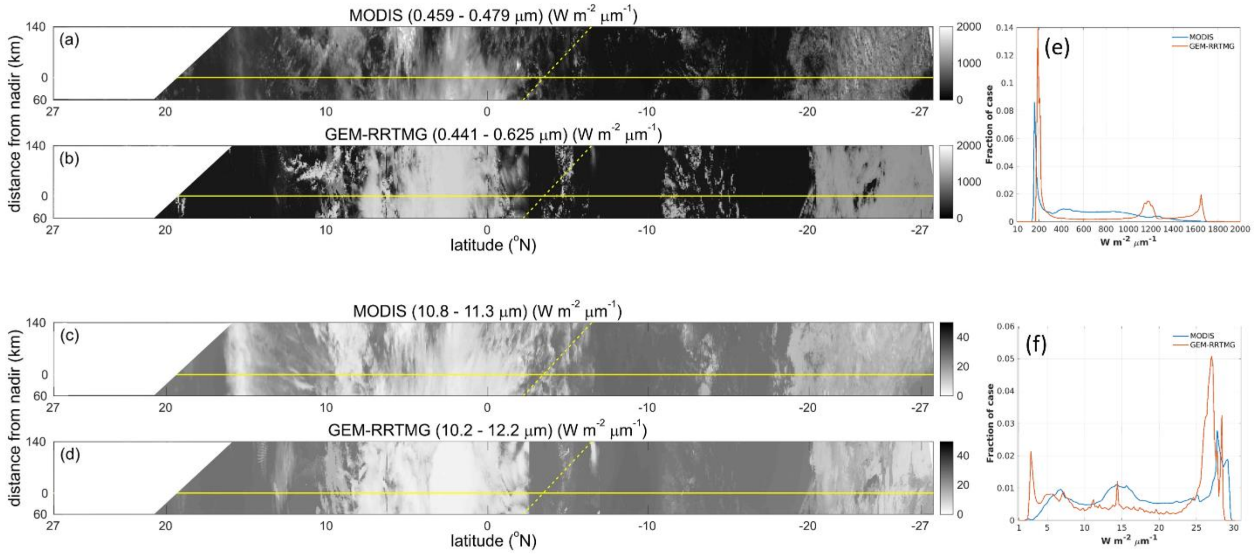

Figure 14: As in *Figure 8* except these are for the *Hawaii frame*. MODIS TOA fluxes are observed between 00h35 and 00h55 UTC on 24-Jun-2015

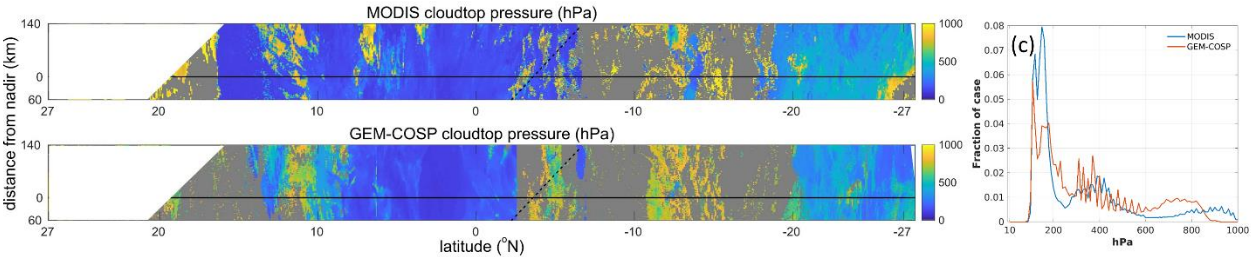

Figure 15: As in *Figure 9* except these are for the *Hawaii frame*. MODIS cloudtop pressure is retrieved between 00h35 and 00h55 UTC on 24-Jun-2015

I also suggest utilizing more detailed satellite simulators to generate observation-equivalent scenes. For example, RRTMG is broad-band RT. The COSP simulator does not account for details in size and phase of MY2. Detailed satellite simulator must closely follow assumptions in size distributions, phase, and shapes in cloud microphysics. You can read section 2 of above paper for more principles and radiance-based model evaluation.

The results presented in this manuscript are meant to provide an overview of the ability of the simulations to provide "reasonably realistic and large" test frames. Given this goal, we used output from the COSP simulator and RRTMG radiative fluxes. While they are not as complex or comprehensive as other systems used to simulate observations, they are one step toward the use of similar quantities to compare model output and observations. Other manuscripts in this special issue present results of applying very sophisticated instrument simulators (ATLID, CPR, MSI and BBR) to the test frames (the list of the manuscripts).

3) Non-evident argument of ice microphysics bias and improvement.

Section 7 argument suddenly starts with "Basically, GEM predicts too many overly small ice crystals….". Unfortunately, I don't see any such evidence in this manuscript or previous manuscript using MY2 microphysics. What is this argument based upon? You must show evidence using either observations or previous manuscript using MY2 in different cloud types. Then, discussion goes to "these alterations were found to …. Improvement of GEMS's simulated cloud properties…." Again, based on what?? No evidence. Essentially Figure 17 compares Reff-T distributions before and after modification.

There are potential pathways to provide this evidence.

1. Validate against Reff products of MODIS/VIIRS satellites. Although these products have their own assumptions, it is better than nothing.
2. Simulate CloudSat reflectivity before and after the change in ice size. In this case, you must use a detailed radar simulator that accounts for size, phase, and non-sphericity of ice crystals.

Reff is the inverse function of lambda in generalized gamma distributions, thus, change in Reff can significantly impact CloudSat radar reflectivity. You can construct contoured frequency of altitude diagrams (CFADs) or similar statistical composites. See examples Fig 10 & 11 of Shi et al. 2010 for example.

Shi, J. J., W.-K. Tao, T. Matsui, A. Hou, S. Lang, C. Peters-Lidard, G. Jackson, R. Cifelli, S. Rutledge, and W. Petersen (2010), Microphysical Properties of the January 20-22 2007 Snow Events over Canada: Comparison with in-situ and Satellite Observations. Journal of Applied Meteorology and Climatology. 49(11), 2246–2266.

Thank you for the suggestions. Assessment of adjustments to Reff and resulting improvements were addto the manuscript. The focus of the assessment is whether the adjustments improved the relationship between Reff and other variables (e.g., Reff_i vs. T; Reff_i vs. cloudtop P; and IWC vs. Reflectivity) instead of assessing whether the simulated Reff values agree well with observations.

"Impacts of these adjustments can be seen in Figure 19, which shows fractions of cases as functions of effective radius and cloudtop pressure. Figure 19a is for MODIS retrievals (MYD06_L2) and shows that for most cases with cloudtop pressure between 200 and 400 hPa effective radii are between 30 and 50 µm. Figure 19b shows that for COSP simulations based on GEM data, most ice clouds for the same cloudtop pressures have effective radius smaller than 15 µm. After applying the adjustments, however, COSP values improve significantly with most effec-tive radii between 30 and 50 µm. Though not shown, similar impacts exist for the Baja and Hawaii scenes."

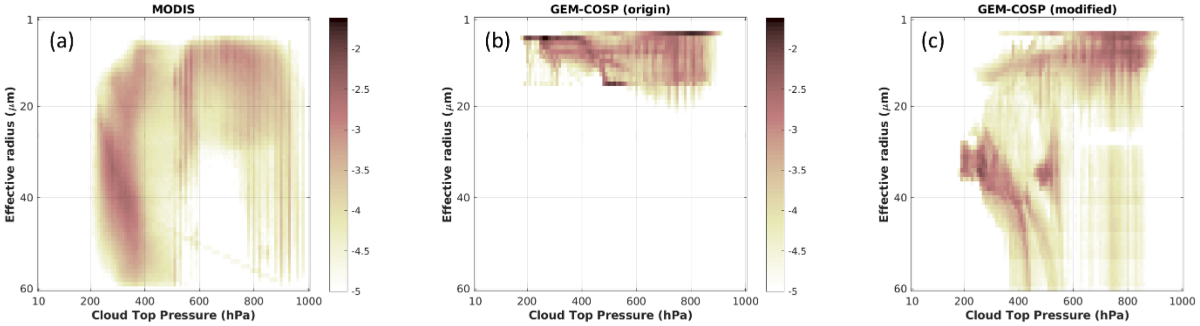

Figure 19: fraction of cases in logarithmic scale. Y-axis: effective radius with bin size of 1 µm. X-axis: cloudtop pressure with bin size of 10 hPa. (a) MODIS retrieval from MYD06_L2. (b) original GEM data simulated by MODIS simulator of COSP. (c) GEM data after the adjustments of effective radius simulated by COSP.

"In addition to these improvements in cloud optical properties, the same adjustments were found to improve the realism of cloud properties that are relevant to simulation of CPR observations. For example, after applying the adjustments the relationship between ice CWC and simulated radar reflectivity now falls in phase-space that agrees well with real observations (e.g., Matrosov and Heymsfield 2017; Heymsfield et al. 2005). Figure 20 shows the IWC vs. Ka-band reflectivity for the nadir Halifax scene path. It can be seen that after adjustment (Figure 20b) the best-fit line of GEM data compares well with the relationships shown in Fig 4 of Matrosov and Heymsfield (2017). The agreement is even more striking when distributions of data shown in Matrosov and Heymsfield (2017) are consid-ered instead of just best-fit lines. Various cross-sections of adjusted GEM+CAMS-derived fields can be found in the Supplementary Material."

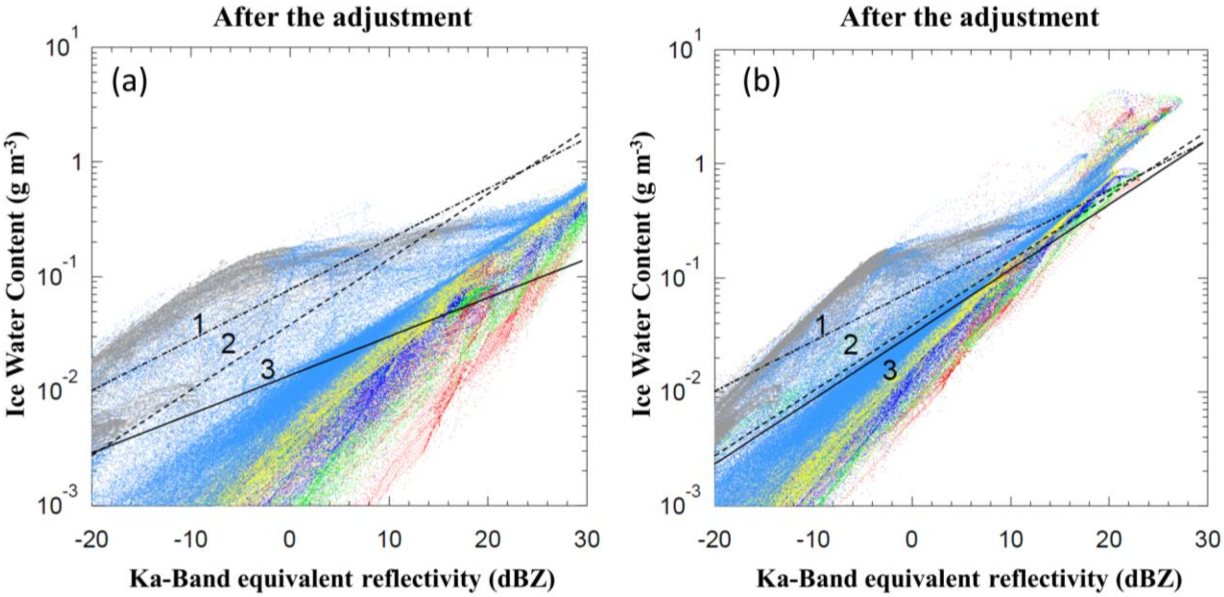

Figure 20: IWC vs Ka band equivalent reflectivity for the nadir track of the Halifax scene. Lines 1 and 2 are the best-fits line to GCPEX and CRYSTAL-FACE respectively data taken from Matrosov and Heymsfield et al. (2017). Line 3 is the best-fit data to the GEM results. The Red points are for temperatures between 0 and -5 C, Green:-10<T<-5 C, Dark-Blue:-15C<T-10C, Yellow: -20<T<-10C, Light-Blue: -20C<T-40C, Grey: T<-40C. (a) the relationship before the adjustment. (b) the relationship after the adjustment.

4) Lack of aerosol validation.

Although aerosols are part of the numerical testing frame, only cloud properties are validated. Aerosols are supposed to be one of the major components of EarthCARE satellites, right? Why not validate aerosols? I understand that it is incorporated from the CAMS field, and applied ECSIM scattering properties. You can simulate these quantities and present statistical quantities. Are these realistic against existing Lidar measurements, like CALIOP/CATS sensors? See example of how to validate aerosol total backscattering in Figs 16 of Choi et al. 2020.

Choi, Y., S.-H. Chen, C.-C. Huang, K. Earl, C.-Y. Chen, C. Schwartz, and T. Matsui (2020), Evaluating the impact of assimilating aerosol optical depth observations on dust forecasts over North Africa and the East Atlantic using different data assimilation methods, Journal of Advances in Modeling Earth Systems, 12, e2019MS001890. https://doi.org/10.1029/2019MS001890

Validation of aerosols is beyond the scope of this manuscript. As mentioned in the Supplementary Materials, our goal was limited to producing ‘realistic enough’ aerosol fields for the purposes of algorithm development and testing. Hence, a large degree of tolerance for ad-hoc choices and procedures was approved and accepted by all involved. Regarding general validation of CAMS’s aerosols properties, we added text and references to relevant CAMS papers (e.g., Flemming et al. 2017).

Flemming, J., Benedetti, A., Inness, A., Engelen, R. J., Jones, L., Huijnen, V., Remy, S., Parrington, M., Suttie, M., Bozzo, A., Peuch, V.-H., Akritidis, D., and Katragkou, E., 2017: The CAMS interim Reanalysis of Carbon Monoxide, Ozone and Aerosol for 2003–2015, Atmos. Chem. Phys., 17, 1945–1983, https://doi.org/10.5194/acp-17-1945-2017.

Minor Comments:

Title: Title including “Data Management”, but I don’t see particular discussion in DM. Either omit DM from title, or add more plentiful discussion of DM.

We prefer to keep “Data Management” in the title. We added the following description at the end of the second last paragraph of the Introduction:

“Use of high-resolution full-frame data in ECSIM not only allows assessment of the quality of EarthCARE’s retriev-als, it also facilitates meaningful estimation of required computational resources and processing times for each algorithm.”

Line 13: Write acronym NWP in abstract.

We changed the text to “Numerical Weather Prediction (NWP)”

Line 39: “ensemble” does not sound right. It probably means “diverse surface-atmosphere scene”

We changed the wording to “surface-atmosphere conditions”

Line 112-115: Is PBL scheme and shallow convection scheme applied to which domains (coarse grid only)?

We added a statement at the end of the paragraph:

“Both schemes are used in all domains.”

Elsewhere: “inner-domains” should be “inner-most domains”.

The changes were made.

Line 127: “Saved variables are listed in the Appendix” but I guess it’s also available in Table 5??

We agree this could be confusing, so we’ve renamed them as Table A1 and Table A2.

Line 138-142: Equations of spectral and white albedo can be omitted since they’re not very important for this paper. Just a reference is enough or say “white albedo is used for diffuse radiation”.

For the convenience of readers, we prefer to keep the equations in the manuscript so that users of the test frames can compute the albedos.

Line 149: Please put a citation for “another emissivity database”, although I understood it is more discussed in the companion paper.

The reference is added into the manuscript.

“Huang, X., Chen, X., Zhou, D. K., & Liu, X.: An observationally based global band-by-band surface emissivity dataset for climate and weather simulations. J. Atmos. Sci., 73(9), 3541–3555. https://doi.org/10.1175/jas-d-15-0355.1, 2016.”

Line 172: Omit sentence “While annoying to look at,”

The change was made.

Line 187 & Figure 10: “making a comparison to GEM useless” This statement is too strong.

We changed the wording to be,

“Unfortunately, CloudSat’s retrieval of liquid CWC is problematic (e.g., Li et al. 2018). The comparison of liquid CWC is therefore not shown.”

Line 195: 1D RT is not the major reason to create homogeneous brightness temperature. Mis-representation of cloud structures due to missing mesoscale forcing should be the main reason.

We change the phrase as:

“This could be due to GEM’s clouds being simply too homogeneous due to missing mesoscale forcing as well as RRTMG’s use of 1D radiative transfer models (see Barker et al. 2017).”

Line 209: “very good” is too qualitative and vague a statement (not scientific statement).

We changed the phrase as:

“Figure 14 shows that for the Hawaii frame, the location and intensity of cloud systems simulated by GEM near the Equator and ~25°S agrees well with MODIS observations.”

Line 230: “by algorithm development groups (….” should be “in the previous section.”

We changed the phrase as:

“While the macrophysical cloud properties simulated by GEM were deemed satisfactory in the previous section, it was clear that there were shortcomings with its predicted ice cloud microphysical properties (cf. Qu et al. 2018).”

Anonymous Referee #2

Summary: The paper describes the creation of the test frames for testing the EarthCare retrieval algorithms. As such it provides substantial utility, which is justification for publication in principle. Publication in practice is to be decided based on its ability to fulfill its objective, which I judged based on the clarity and completeness of the description. Overall it passes with flying colors, as it provides a clear, concise, and compelling description of what is done and what is available. The authors are to be congratulated.

Thank you very much for your review and compliment! Please find below our response to each point.

I only have minor editorial comments or suggestions that the authors may want to consider for their final revision.

1. Line 26: Perhaps say “early to mid 2024 or perhaps later”.

   Thank you for the suggestion. We change the phase to:

   “which is scheduled for launch in early- to mid-2024”

2. Line 42: “Lacks this luxury” is a rather conversational way to make the point which might take readers a few passes to digest.

   Agreed. We changed the phrase as:

   “One could stop here and assess performance by comparing retrieved geophysical quantities to their simulated counterparts (cf. Mason et al. 2023), but in the real mission this is impossible to do routinely.”

3. Line 63: I was a bit puzzled by the reference to the bin-resolved, as what a bin scheme can resolve is a non-parameteric distribution. Passing a parametric distribution to a bin scheme leads to a lack of resolution and seems simply a matter of practicality when interacting with the radiation, as such this strikes me as an unnecessary detail, elaboration, that is not necessary to understand the present paper.

   Agreed. We made the following change:

   “Bulk properties of atmospheric attenuators, such as 3D distributions of GEM’s cloud water contents (CWC), are used in conjunction with assumed aerosol/cloud size distributions in order for ECSIM to produce physically-consistent synthetic measurements for each of EarthCARE’s sensors.”

4. line 76: I don’t think I fully understood the rationale for not considering night scenes. The simplicity assumption would be that nocturnal situations don’t fundamentally sample a

different meteorology, which might be true, but it should be stated, rather than simply focusing on the effect on the instruments.

Agreed. We changed the phrase as:

“Assuming that night-time atmospheric conditions are not fundamentally different from day-time conditions, night retrievals can be approximated by neglecting MSI solar channels and solar back-ground for ATLID.”

5. Fig : I would have preferred a qualitative coloring of the frames, and a label of the colors

   We made changes for Figure 1:

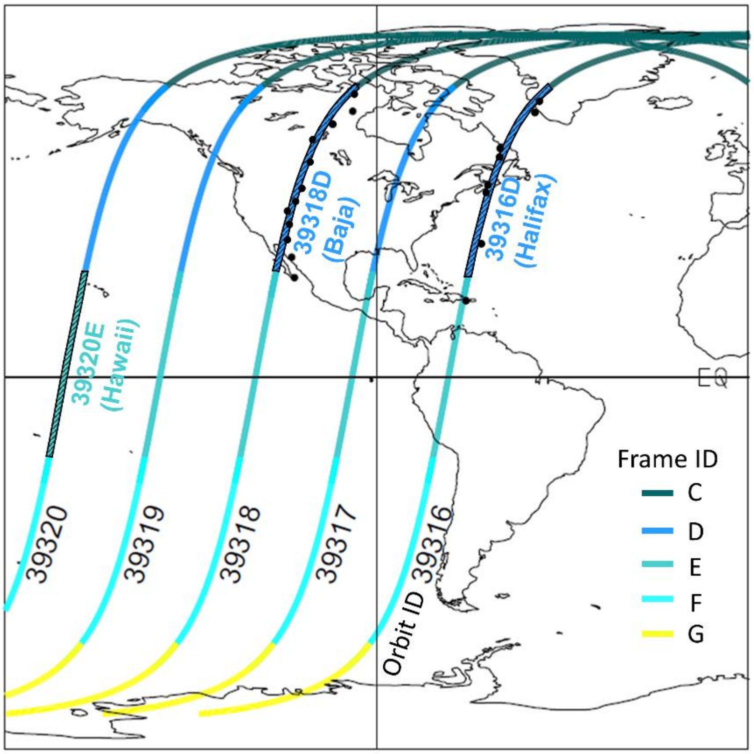

Figure 1: Examples of several successively numbered EarthCARE orbits as provided by ESA. Frames are colour-coded. The test frames are indicated by shaded areas. Frames 39316D, 39318D, and 39320E are referred to as “Halifax”, “Baja”, and “Hawaii”, respectively.

6. line 84: Why not use the ISO-8601 standard for date formatting.

   Thank you for this suggestion. The format of the dates is changed to ISO-8601 standard.

7. line 103: I know the phrase non-hydrostatic primitive equations is used, but I find it confusing because I think of the hydrostatic assumption one of the things that make the

primitive equations the primitive equations. I would prefer, the non-hydrostatic extension of the primitive equations.

Agreed. The change was made in the manuscript.

8. Fig 5: For domains 2 and 3, I inferred that they are implemented 13 times, for each of the instances of domain 4, but if this could be said more explicitly it would avoid confusion arising from Fig. 5 which shows just one instance.

   Thanks for this suggestion. The phrase is now changed to:

   "The downscaling transitional domains at Δx of 2.5 km and 1 km adapt themselves to the locations of the Δx=0.25 km domains (both domains at Δx of 2.5 km and 1 km are repeated 13 times). A common Δx=10 km domain was used for all 13 segments."

9. Fig 6 - wouldn't a binary color scale be more appropriate for what I infer to be a binary mask.

   The "water-land mask" and "ice fraction" variable are actually continuous values between 0 and 1 (e.g. 50% of grid is land), although most of cases are either 0 or 1. Given this we prefer to keep the continued color map. However, the variable name "water-land mask" is confusing and we have changed it to "water-land fraction". We also changed the description in the caption.

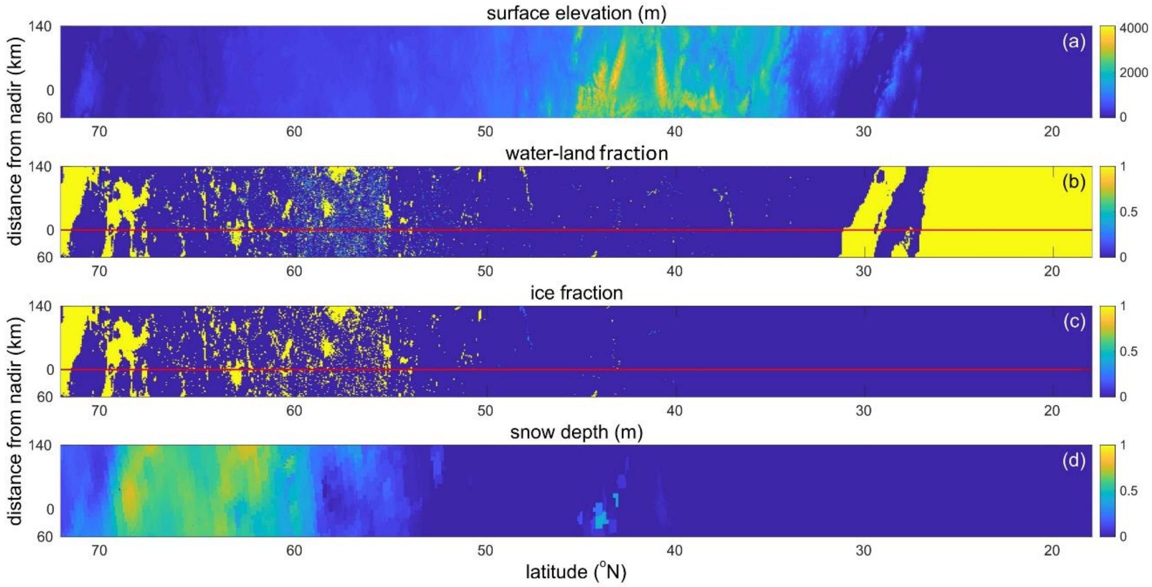

Figure 6: All panels are for the Baja frame (see Figure 1 and Figure 3) and each panel's title is self explanatory. For (b) and (c), blue (fraction of 0) corresponds to 100% land and yellow (fraction of 1) to either 100 % water or 100% ice.

10. line 134: April is not that late in spring, so I was surprised by how little snow there was in the Rockies, making me wonder if this was a bias, or just a false expectation on my part.

    From Figure 6d, we can still find some areas with snow depth of ~30 cm near 44°N. The snow information is based on NWP model outputs using a global surface analysis (relatively low resolution), hence there might be inaccuracies. However, since the primary purpose of the test frames is for end-to-end simulation, we, and other algorithm developers, considered inaccuracies and uncertainties such as these to be acceptable.

11. line 203: I thought the ‘quite good’ was a bit of an overstatement. I guess it depends on one’s expectations, and raises the question as to whether the qualitative judgments that are made in these sections are appropriate.

    We replaced the phrase with:

    “Despite these discrepancies, Figure 13 shows that in the vicinity of where the satellite tracks intersect, vertical reali-zations of clouds from both GEM simulations and CloudSat retrievals indicate smooth mid-level low density clouds, although GEM’s are more extensive. The altitudes of GEM’s clouds over the Rooky Mountains are also in fair agreement with CloudSat’s. Unlike the Halifax frame, the magnitudes of modelled and “observed” IWPs agree quite nicely, in general.”

    With regard to the qualitative judgement in the manuscript, we added PDF plots in Figs. 8, 9, 11, 12, 14. More discussions are also added in the revised manuscript. Please refer to the answers to reviewer #1 for more details.