Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2016-295-RC2, 2017

Geoscientific
Model Development
Discussions
Open Access

Interactive comment on "A multi-diagnostic approach to cloud evaluation" *by* Keith D. Williams and Alejandro Bodas-Salcedo

Anonymous Referee #2

Received and published: 7 March 2017

General comments:

This paper illustrates a model evaluation against a variety of different observational datasets of cloud and radiative properties. A variety of approaches are used to limit analysis to specific cloud regimes and to specific time-scales.

The paper mainly serves as an example of how model evaluation against available satellite and ground-based observations of cloud properties might be performed, including the use of techniques to account for uncertainties or biases in satellite retrievals (using simulators), techniques to isolate specific cloud or dynamical regimes (using compositing), and techniques to isolate the climatological or systematic biases in the model from short-timescale processes (using hindcasts). While this is a useful contribution, the paper leaves much to be desired in terms of physical interpretation,

attribution, and discussion of identified biases, and instead focuses primarily on listing the identified biases.

Additionally, there is little (if any) discussion of uncertainties in the observational products used, or of the uncertainties in the comparisons between the model fields and those observations.

In light of these shortcomings, I would recommend major revisions to the paper, in particular to dive somewhat deeper into identifying physical processes responsible for the identified biases in the model in terms of the model formulation.

Specific comments:

140: A definition of low, mid, and high cloud categories should be provided here (i.e., what are the altitude bounds for each category?). A short description of how these histograms are produced would also be useful to the reader here, in addition to providing the reference provided (i.e., cloud occurrence in each category is defined as that which exceeds a minimum backscatter ratio of ??).

153: A brief explanation of the approach for each simulator would be helpful here (i.e., the ISCCP simulator emulates the way the retrieval infers cloud top pressure by estimating brightness temperature...).

156-165: The addition of this diagnostic the combines the CALIPSO and CloudSat hydrometeor occurrence is fantastic, but this description and discussion of the implementation is not nearly sufficient. A much more thorough description of the algorithm should be provided. The rationale for the choice of threholds used seems somewhat incomplete as well, and it would be nice to see the comparison between GOCCP and RL-GEOPROF referred to on line 159. On line 160 it is suggested that the cloud detection algorithms differ between that used in COSP and that in RL-GEOPROF, but the nature of this difference is not explicitly stated and probably should be. Overall, some discussion of the uncertainties and sensitivities to the formulation of this new diagnostic should probably be provided to justify its use in the model evaluation. This could potentially be a significant contribution of this paper.

212-215: This is a nice result, and it would be worth expanding on the cause for the difference in cirrus between GA6 and GA7. In particular, some justification for the claim that the largest difference is due to the reduction in the rate of cirrus spreading could be shown, such as a figure showing the cirrus amount in GA7 with and without the adjusted cirrus spreading parameterization. I do not think the formulation of the cirrus spreading parameterization, or the changes made to improve the simulation, have been documented well enough in the manuscript. This result showing the decrease in cirrus and better agreement with both CALIPSO and CloudSat is a nice validation of the improvement in the simulation due to these changes, and would go nicely with a more thorough explanation of what is going on here.

221-222: How do we know that the revised numerics are responsible for the improvement in GA7? What specifically changed in the formulation of the model?

230: How is the "grid-box cloud fraction" being calculated? I am somewhat confused as to how this is produced alongside the profiles of reflectivity shown in the top panel. Is cloud fraction simply being aggregated onto a coarser grid from the reflectivity, calculated as the fraction within the coarser bins above some reflectivity threshold?

232-236: What does this imply about the model formulation (the cloud parameterizations)?

242: Add a note here that the drizzle rates cited are not shown here.

247-250: This is a nice demonstration of the impact of the new microphysics package, but this is lacking a discussion of the mechanisms for the improvement, and should be accompanied by a description of the changes.

258: Could the increase in cirrus here be explained by excessive advection of the cirrus outflow, or again maybe something to do with the cirrus spreading parameterization

referred to earlier? What is responsible for the improvement in GA7?

261-270: This discussion does not contain much substance, and inclusion of the ISCCP comparison seems to almost be an afterthought. This either needs a more complete treatment of the sources of differences, or consider cutting from the manuscript to make room for some of the more fleshed out analysis, such as the discussion of improvements in thin cirrus.

278-279: This statement could use evidence or a citation to back it up.

281-286: This could be better tied in with the discussion of cirrus above. In general though the results from this figure are not very compelling and do not seem to add much to the discussion. It is also not clear to me from Figure 5 that cirrus is overestimated in GA6. The most apparent biases in this figure are the altitude bias in the location of the cirrus maximum in GA6, and an overall underestimation of cirrus in GA7.

287-290: These conclusions are difficult to draw from Figure 5 as shown due to the scales of the axes used. If boundary layer cloud is the focus of this figure, it would be better to show just the boundary layer for the lower panel (SST composites), and on a cloud fraction scale that allows the reader to actually see the differences between the different curves.

340: I realize this is explained in the cited manuscript, but at least a simple explanation of the equation tested should be given here.

352: "Reasonably good" is awkward language to use here. I would suggest replacing with something like "while the cloud simulation was in reasonable agreement with observations".

356: Again "reasonably good" is awkward here.

368-369: Elaborate on how these biases are consistent with the radiation errors.

385-389: This is an excellent example of the utility of using multiple observations in the evaluation strategy. This would be a good point to emphasize, and perhaps use as a jumping off point for a more elaborate investigation of the source of these differences (multi-layered cloud vs excess precipitation) than is given in the sentences to follow.

401: Why is SYNOP data the most reliable here?

403-405: Need evidence or references to back this up.

410: How is an okta defined in the context of the model?

439: What caused the reduction in the cold bias in GA7?

447-450: I am not sure I entirely agree with these conclusions. The reflected shortwave biases around the subtropical cumulus transitions seem to have reversed in sign between HadGEM2 and GA7, but the magnitudes do not seem to be universally reduced. Perhaps I am looking at the wrong part of the figure though, so maybe a box or symbol on the figure indicating the region where the improvement is evident would be appropriate. The underestimate in reflective shortwave over the Southern Ocean also does not appear to be significantly reduced.

482-485: This seems to really be a key point of the paper: to demonstrate that the multi-diagnostic approach used reduces the possibility of drawing the wrong conclusions. This is hinted to at points in the paper, but I think this could be drawn together a little better here, perhaps by recounting the points in the preceeding analysis that illustrate this (such as the contrast in the comparisons between CloudSat and CALIPSO that demonstrate errors due specifically to thin cirrus, or to excess precipitation as opposed to cloud errors).

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-295, 2017.