

Interactive comment on “Evaluation of Southern Ocean cloud in the HadGEM3 general circulation model and MERRA-2 reanalysis using ship-based observations” by Peter Kuma et al.

Anonymous Referee #1

Received and published: 8 May 2019

Review of “Evaluation of Southern Ocean cloud in the HadGEM3 general circulation model and MERRA-2 reanalysis using ship-based observations” by Kuma et al. (acp-2019-201)

Summary:

The paper investigates cloud cover over the Southern Ocean through comparisons between numerous ship-based measurements and model outputs (including reanalysis). They demonstrate underestimation of low-level cloud cover in the HadGEM3 model and in the MERRA2 reanalysis. They investigate the link between boundary layer thermodynamics and low-level cloud cover and cloud biases. They show that the

C1

TOA SW biases are mainly related to places where the coldest near-surface airmasses are (near or below zero). They conclude on the subgrid-scale parameterisations being responsible for misrepresentation of clouds in model rather than boundary layer thermodynamics.

Relevance of the paper and overall comment:

The paper presents and describe a very valuable dataset of ship-based measurements of low-level cloud over the Southern Ocean, where observations are badly needed to understand the near-surface processes affecting cloud formation and responsible for the cloud/radiative biases in climate models over the SO. To this respect the paper addresses relevant science questions in the scope of ACP. However, it seems to me that more work is needed to achieve ACP standards, in the way the science is presented and discussed (major revision). The dataset deserves better scientific discussion and less vague or speculative comments in several parts of the paper. Figure 5, 7,8 and 10 are very interesting but the analysis and discussion should be better handled. I first list some major comments, and then line by line comments.

Major comments: _____

1)

The use of different time-periods needs to be much better introduced, justified, and discussed. I don't understand why the author use GCM simulations for the 1980-1989 period in a free-running mode, and then a nudged simulation for the year 2007 (only), while MERRA is used only for the 2015-2018 (the years where ship-based measurements took place). The reader needs much better justification for the choice the authors make to compare different periods. And a discussion on the shortcomings of doing so should appear in the paper. P4-Line 26, the authors say “Limited data availability meant that no nudged runs were available for the period 2015-2018”. Is this really the case? And if this is the case, why not having a free-running simulation for this period then? And why is the nudged run over 2007 only? Also, MERRA2 could

C2

be used for the 1980 period. MERRA2 is available for ≥ 1980 . MERRA could help bridge between the period 1980's/2007's of the GCM outputs and the period of the ship-based measurements (2015-2018). At least using MERRA2 also for the 1980's and 2007 + explaining/discussing the choice for the time periods of the GCM runs would be needed. The best case would be to have GCM runs over 2015-2018. How using different periods for GCM/MERRA2 would affect Figure 5 for instance? And what about Figure 1 and the TOA biases where only the year 2007 is shown? The authors speak about the years 2016-2018 that had unusually low sea ice extent (p15-Line16): how does this impact the comparisons with other years where sea ice was different? Having said that it is possible that the paper could be improved by giving up Figure 1 or 2, while focusing more on the novelty of this work, which is the ship-based measurements of clouds (+thermodynamics), and drop the comparisons to GCMs (mainly because they are different periods...) and keep the comparisons to MERRA2, and add the ERAI reanalysis. To this respect ERAI, which is widely used, could also be presented here. How is ERAI doing vs. MERRA2. Say, if ERAI has a contrasting behavior to MERRA2 (ie more like the GCMs) the authors could consider presenting the ship-based measurements + MERRA2 + ERAI over similar periods, and drop the GCM runs, which deal with other periods.

2)

The discussion of ship-based measurements should be better related to the TOA SW bias over all periods where these field measurements were available. Since the authors try to understand what the models are doing wrong, the discussion should make the most of the different measurements period. Because maximum insolation occurs in January, a focus is made on this month but Figure 2 clearly shows that March – for instance – can also and still show substantial biases. Since Ship-based measurements are also available in autumn and November-December, it would be very welcome to have also biases like the ones shown in Fig. 2 for the autumn and other summer months. Is the TOA SW bias spatial pattern (and the related comparisons between

C3

models) the same during these other months? I suggest Figure 3 to show only biases (the subplots m-p) for summer and autumn. The other maps (a-l) are difficult to read with the blue-shaded colourscale and I am not convinced they need to remain. Better discussing the cloud cover results in Summer/Fall (Figure 5) in relation to radiation biases in Summer/Fall would improve the overall discussion and conclusions of the paper. This would allow to make the most of the ship-based observations. Figure 5 is a great one and it would deserve better discussion in light of the motivations (i.e. the TOA SW biases over the SO and why these biases are present). The authors note that GA7.1 reduces the SO SW radiation bias (e.g. in the abstract p1-L9). Figure 5 does not show any cloud from GA7.1 (only GA7.0). Why GA7.1 is performing better? Is this really because of better cloud representation (but we cannot see it from Figure 5)? And if not, what does it say about cloud being the main/only reason of SW radiation bias? Related to 1), what ERAI would give in Figure 3? More like the GCMs or like MERRA2? Perhaps ERAI brings this contrasting behavior that the authors highlight between GCMs and MERRA2, and this would allow to have both observations and reanalyses (only) used over the same period (2015-2018)

3)

The authors tend too often to rely on previous conclusions from previous papers (e.g. the Bodas-Salcedo et al. ones) to comment on what they find, rather than more thoroughly commenting/discussing their own novel results. The discussion part for instance gives much room to results of previous published study and/or to speculative comments about why GA7.1 is doing better than GA7.0 etc. and how MERRA2 is overcompensating for the low cloud-cover etc. Several sentences using “we cannot conclude...”, “we cannot make the same conclusion... but it seems plausible...” “we cannot make substantial conclusions...” considerably weaken the discussion (section 4) from the beginning, and hence the paper, while it seems that all the ship-based measurements bring very valuable results (Figure 5, Figure 7, Figure 8) and interesting comparisons to MERRA2, and GCM runs (but cf. my point 1. on the time periods used).

C4

4)

The discussion on the effect of sea ice is overlooked while it seems that some discussion could be made from Figure 8 (q,r and w,x). Also, while it seems that 8w is still showing some correlation, Figure 8x shows very different behavior and no attempt is made to comment on this. Given that a lot of the soundings you use (65) were made in 100% sea ice regions, and many of your CBH observations as well (I see the number of points present in your Figure 8x compared to the other similar subplots), I would expect to see more in-depth study of these observations, and this is really missing the present version of the paper in my opinion. For instance, the recent study by Jolly et al. (2018) that you cite showed the influence of different regimes on cloud cover: can the observation in Figure 8x be explained by particular synoptic-scale regimes or just by the sea ice being 100%? And why? Also, that other recent study by Listowski et al. (2018) that you also cite showed that not all low-level clouds anticorrelate with sea ice fraction but only the liquid-bearing ones. Can the behaviour you see in Figure 8x also be explained by clouds being of different sort/phase? The absence of correlation between CBH and min(SLL and LCL) may lead to think that you could be observing clouds advected from other places not related to local atmospheric conditions, that may be different in nature/phase from clouds over open water (you mention some hints towards the detection of supercooled liquid water with one of your instrument, can't you improve the Figure 8 by adding information on the phase, notably for Figure 8x?) In other words, can what you are observing from regions with 100% sea ice be explained by changing synoptic scale regimes, or cloud phase, or other things? Speaking of the sea ice regions, in Figure 7 the very low CBH are identified as being due to fog/very low clouds (p13-L20) and these are the points we also see in Figure 8x. Could this be blowing snow since we are in a 100% sea ice-covered region where snow can accumulate? In relation to cloud phase, in Figure 9 you compare LWP and IWP for GA and MERRA only for a specific year and month (jan. 2007). This does not seem satisfying to conclude for the longer time scales/other periods. Here again the use of different time-periods in the paper is not very welcome (see my point 1.). Do you really need

C5

Figure 9? If you really want to go into the cloud phase, using the lidar observations to assess the nature of cloud phase would be welcome. Or, as suggested in 1), perhaps only using MERRA2, ERAI (to contrast with MERRA2?) over 2015-2018 (only) would be a better option rather than using 2007, i.e almost a decade before the 2015-2018 ship-based observations...

5)

Finally, the authors say that subgrid-scale processes should likely be responsible for the cloud misrepresentation in models rather than the boundary layer thermodynamics but it is never said and commented on what these subgrid-scale processes are. Do you mean the microphysics? Other processes? A discussion of what is used in the models regarding these processes would perhaps help to understand what should be improved in priority in the models and why the models are wrong. Using the contrasting behaviors of models to try to pin down the cause of cloud misrepresentation is an interesting method but the authors should provide with more clues in the discussion about what those subgrid-scale processes are and try to spot the main differences in the way the models implement these processes.

_____ Line by line comments: _____

_____ Abstract _____

P1-L9 By how much GA7.1 reduces the bias? P1-L17 The analysis you mention is referring to your Figure 9 and the related comments. They only refer to the period January 2007... as mentioned in my major comment 4) this is not satisfying I think. When one reads the abstract it seems that you compare modelling and MERRA2 over the same period as the ship-based measurements, which is not the case. This is misleading.

_____ 1.Introduction _____

P3 – L12 : “It was also more...” : what does “it” refer to exactly ? P3 – L14 : “more

C6

likely to have intermediate cloud fraction” This is not clear. What is meant here by “intermediate cloud fraction”? P3 – L19-20 Please double-check and be more precise here (what “tuning” do you mean?). Kay et al. (2016) changed the threshold temperature below which detrained condensates are ice crystals and not liquid any more. They lower this threshold, allowing for more condensates to remain in the supercooled liquid phase when being detrained. The way the sentence is written suggests that ice crystals only are detrained. P3 – L25 The reference to Jakob (2003) is a bit short or can be removed unless you specify what you mean by “cloud evaluation” regarding this specific study. P3 – L27-35 Please make a new paragraph and give section numbers to help the reader.

————— 2. Method —————

General comment: I would suggest a section 2. Datasets and 3. Method (lidar simulator). As it stands, it seems that this section combines too many different information about the data/methods used in the paper.

P4 – L2 As mentioned in my major comments, adding ERAI would be very interesting since this is a widely used reanalysis by the community, and would allow to contrast MERRA2 on same time-periods than ship-based obs. P4 – L9 I wonder whether a small appendix summarizing the main aspects of the lidar simulator would not be needed here, since the reference put is a paper in prep.

P4 – L16 What is the difference between GA7.0 and GA7.1 ? This would help understand and discuss the better performance of the latter in terms of SW bias (as stated in the abstract).

P4 – L 18 As said in the major comment it should be explained why these runs are used. 1980-1989, and then 2007. Why not having runs over more recent periods (as the ship-based measurements).

P4 24 – “Can only be compared statistically” What do you mean? Please clarify.

C7

P4 – L26 “Limited data availability. . .” What do you mean? See my major comment 1)

It does not seem that you are saying over which period you analyse MERRA2. This should appear in this section.

P5 – L15 “downsampled” from what initial resolution?

P6 – L5 “appears largely zonally symmetric” I don’t think we can say the pattern of the bias is symmetric, even zonally, but rather that the bias is present across all longitudes, but its magnitude does change zonally.

P6 – L6 “with a notable exception. . .” Precisions not needed in the section presenting the ship measurements. . .

P6 – L8 “Figure 1. . .” This Figure is already mentioned before P5 – L21.

P6 – L20 I am not sure what is meant by “directly reveal the cloud liquid. . .”. The strength of using a simulator is to compare the observables and not to rely on all the hypotheses used by inversion routines to retrieve IWC and LWC from lidar observations.

P7 L24 – Please clarify the title e.g. “Geographical areas/domains investigated” or “Domains used for the analysis”

Also, having 2.1 as “Datasets” then 2.2 as “Domains” then 2.3 “COSP simulator” is not ideal I think, and I would first present all datasets and tools, and then the domains.

P8 – L12 The title of this subsection is misleading since you are not using COSP simulator in the end, but your own simulator. Please change the title accordingly.

It seems to me you don’t need a section 2.3 and you could have everything put in current 2.4.3 where you could at once explain the modeling of the lidar signal along with its processing.

P9 L15 What is this known value of LR? Where does it come from?

C8

P9 L21-22 Citing Kotthaus et al. at the end of the paragraph falls a bit short and I am wondering if it should not appear earlier in the paragraph with some more explanation about why you refer to this study. Are you using their method? Then please say it.

P10 L7-9 Why do you need to do this? How are these random samples used then?

To shorten this section 2. I would not define SLL here, rather when it is used for the first time. Also SSL is neither a dataset, nor a tool, rather a variable defined to help with the analysis.

Also, is there any past reference using this definition? If yes, please cite relevant paper.

3. Results

P10-L26 to P11 L-16 There are too many statements dealing with observations made on Figure 3 that are actually difficult to see, whereas Figure 4, introduced after, is more helpful to confirm statements made by the authors. Also, as suggested in major comments, I would tend to simplify Figure 3 by showing only the biases and remove all the blue-shaded figures where the biases are difficult to read, especially regarding the statements made by the authors in the main text.

P10 – L28 “Lower” than what? And what biases? Please clarify.

P10 – L29-30 I would remove the sentence about the “predominantly zonally symmetric pattern” and the “more variable patterns in the tropics”, which is not very clear to me.

P11 – L1 “upwelling and downwelling” what?

P11 – L3 “large differences” between what? I would drop the mentions to the Peninsula and what is happening to the east of it as it is not clear why one would give so much importance to this since the ships did not get there anyway.

P11 – L1 I don’t understand the footnote. Also, I am not convinced there is a need to highlight a particular day in the present paper.

C9

P11 – L4 One cannot really see this “greater reflectivity”.

P11 – L13 “With some individual cloud systems being too bright”. I am not sure this should remain in the text. Again, I think all the consideration about the blue-shaded maps in Figure 3 (but biases maps should be kept) should be removed and Figure 4 should be used instead.

P11 – L21 “cyclical”. Rather “seasonal”?

P11 – L20 What is meant by “likely a secondary modulating factor”. Please be more explicit. A modulating factor for what?

P11- L26 “These panels also justify why...” Not needed.

P12 L4-6: The two sentences fall a bit short. Also, they would be in better place in the discussion part, with more explanations. “. . .in the GA7.1 model”: so what?

P12 L2 Figure 5 is very interesting and rich, and more analysis should be provided also regarding similarities or differences between summer results and autumns results. (Please consider adding letter to designate specific subplots of Figure 5). Also it seems that obs and model agree more where the statistics is larger (more days), can’t you say something about that? Isn’t it possible that at other time/places the larger disagreement between model/obs is partly due to smaller statistics of observations? This relates to my major comments that more analysis and discussion are really needed on this plot.

P12-L10 What period is used for MERRA2 here?

P12 L12. As mentioned in the major comment. Why can the authors trust comparisons between simulation of the 1980s period and the 2015-2018. This should be much better introduced/justified.

P12 L19-20 how much higher?

P12 L20-21 “Due to the zonal...of the whole SO” Could suit the discussion part. Not needed here.

C10

P12 L27-34 I would drop Figure 6 and give only numbers. It saves a Figure. Also what bothers me is that GA7.1 is said to be better from nudged simulations but, in the end, only GA7.0 is presented here, because of the decadal run being only available with GA7.0. This is again a shortcoming of accepting to work with so many different time periods for different simulations.

P13 – L1-5 Have the authors consider to use satellite data, or to rely on previous publications to try to assess how the comparison to models is biased by extinction of the ceilometer signal into the lowest thick clouds? At least this should be discussed in the discussion part. This is not the case now.

P13 L10-11 Is the extraction made above the lat/lon of the balloon launch or does it follow the radiosonde trajectory? I guess it is the latter but you may want to clarify this in the text.

P13L11 Can you make a subplot for each of the dataset? It is difficult like this to spot differing behaviours between coloured markers.

P13 – L14 What relationship?

P13 – L19 How large?

P13- L23 how weaker?

P13 – L25-27 The fact that LTS is not a good indicator should be discussed in the discussion part and I don't think it is the case for now. This relates to my major comment 3) where I suggest that more emphasis should be given in the discussion to all results obtained from these novel ship-based measurements.

P13 – L28-34. Figure 8 is introduced, but then some general statement are made about synoptic scale forcing. It would be much better, for the reader, to stick to the Figure.

P14 – L1 “As can be seen...where there is no sea ice”. What can be said about Figure8a and b where there is no cloud but at the same time GA7.0 is not in agreement

C11

at all with observations? Also what is the unit in the x-axis of subplots Figure 8a-f and Figure 8m-r?

In Figure 8g-l and s-x, you are not showing the modelled dots, only observations. I would have expected to see the model outputs as well. Or is it not useful here?

P14L3-4 “There is no substantial difference between...” This is not true for Figure 8a and b... which present non-sea ice cases. This should be discussed. “Plausible effect”? What do you mean?

P14 L8 – What is meant by subgrid-scale processes? Please be more specific.

P14 - L2 I am not sure about this subsection. I struggle with having it only focusing on January 2007. Since the novelty of the paper is the ship-based measurements I am not sure having this part here is relevant, especially that it is only about comparing Jan 2007 for two models. Plus, the GA7.0 one is not the one used in Figure 5, but the nudged one, and it is not clear what period is used for MERRA2. Why not showing also GA7.1 since it is spotted as reducing the SW biases (because of the modelling of larger supercooled LWP?)

P14L26-30 “We should note...” These are comments for the discussion part, but even so, these considerations are also and already mentioned in other places of the paper and remain very general and a bit speculative. I am not sure these zonal plots deserve a separate section, also because of these time period issues mentioned above.

4. Discussion

In general the discussion should be more focused on your results at least in the beginning and spend less time on explaining previous works. Figure 10 (which is interesting indeed) should come earlier in the discussion. Also, you don't seem to do discuss Figure 10b, but only Figure 10a.

Sentences like (P15-L18-21) “Combined...” are a bit speculative and more room should be rather given to discussing the results obtained from ship-based measure-

C12

ments, ie. Figure 5, Figure 7 and Figure 8, and 10. And then make the link to the TOA SW bias issue and relate it, possibly, to the LWP as modelled (cf. Figure 9 – if still considered relevant in a revised version).

P15-L34 to P16 L4. This is too much about other study, not enough discussing your results. Figure 10 comes after that and this is not appropriate. Also – as an example of additional discussion element – are the ship-based observations, which show larger discrepancies from MERRA2, in places where the near-surface temperature is the coldest? In other words, can you relate Figure 10 with your cloud results, instead of only speaking of the SW bias? Also, why is Figure 10 only showing the year 2007? Why not showing the decadal simulation, and the MERRA2 outputs as well (during the ship-based measurements)? What do they say? How is it consistent or not with the cloud simulations in these models? These sorts of analysis/discussions are really missing in the paper, in my opinion.

P17 L9 “Because sea ice is an important factor. . .”: What is meant by “secondary effect on cloud cover”? It seems to me you have the opportunity to say something about the effect of sea ice on very low clouds (and specifically the ones missed by satellites) – e.g. your Figure 8x – but you are not exploring this in the paper. This goes along with my major comments that not enough efforts are made to discuss the very interesting observations you have from ship over three years and in sea-ice free/covered regions.

5. Conclusion

In the conclusion only you speak again about the sugrid-scale processes without specifying them. This should be a paragraph on its own in the discussion part, trying at least to understand how the various models are doing different in parameterising these processes. This would give more perspective to the present work I think.

Figures.

Figure 2 If you still want to keep all the model results (provided you better justify your

C13

method – see my major comments) then you should add the time-periods for the simulations you use, and for the observations, so that one immediately knows you are using different times for comparisons (and that this is then discussed in the text).

Figure 3 As I said before, one struggles to see features with a single colour-shaded scale. As suggested I would keep only the plots showing the biases, and for summer and autumn (as these are seasons investigated with ship measurements).

Figure 4 The horizontal line indicates the “0” value for the bias (red curve). Please make it red (and thicker, or dashed).

Figure 5 What period is used for MERRA2? Why not also showing the nudged runs with the better (according to what you say) version GA7.1U.

Figure 6 Not sure this figure is needed. See my comment in the relevant section.

Figure 7 This would be better to separate the dataset in different subplots to see the different behaviours.

Figure 8 What are the x-axis units in the subplots a-f and m-r? The markers in the g-l and s-x subplots are quite small. Can you either make them larger or increase the size of the subplots.

Figure9 What are the contour values?

Interactive comment on Atmos. Chem. Phys. Discuss., <https://doi.org/10.5194/acp-2019-201>, 2019.

C14