Review of "Employing airborne radiation and cloud microphysics observations to improve cloud representation in ICON at kilometer-scale resolution in the Arctic"

by Kretzschmar et al.

submitted to Atmos. Chem. Phys. Disc.

Summary

This paper presents an analysis of surface radiative biases in kilometre-scale ICON simulations as compared to field observations obtained during the ACLOUD campaign which was conducted May-June 2017 around the region of Svalbard. Measurements were obtained over partial-sea ice and full sea-ice covered surfaces. Biases in surface solar and terrestrial irradiance within standard ICON configurations are attributed in this study to a misrepresentation of the surface albedo in short-time-scale simulations and an overestimation in cloud transmissivity for low-level clouds above sea ice. The latter is further attributed to an underestimation in cloud-droplet number concentration of small to medium-sized cloud droplets (10-25 µm), which also leads to an underestimation of cloud water content. In sensitivity experiments the authors demonstrate that a more accurate description of the activation process in kilometre-scale ICON simulations, and an adjustment of the background CCN profiles to Arctic conditions, decreases cloud transmissivity and thus improves the simulated cloud-radiative effect (CRE). The authors also show that an active coupling of the two-moment microphysics scheme to the ICON RRTM radiation scheme does not yield a considerable improvement of the simulated net CRE in this case.

Recommendation

This study presents a comprehensive evaluation of the CRE of low-level Arctic clouds in ICON simulations above sea-ice covered surfaces. Low-level clouds and in particular mixed-phase clouds impact the surface radiative balance substantially in this region and are often miss-represented in climate models. This analysis is thus addressing one of the key concerns within the community and will be of interest to a wide readership.
The paper is really well written and very logically structured. Their results are presented clearly and concisely and I agree with their scientific conclusions. Occasionally their arguments could be strengthened, which I point out in my comments below. Overall, I think this is an excellent paper that deserves publication once these minor revisions are addressed.

General comments

1. I understand that the case descriptions etc. are given in other papers. Yet from this paper it is not clear for which conditions you have tested the TKE-based activation approach and its impact on net CRE and for which conditions you associate the largest biases. While detailed case descriptions are not necessary, context should be given for the reader in terms of the conditions of June2-June5th (for which the bias attribution and sensitivity analysis is done). In particular information with respect to temperature regime, integrated water vapour content, optical depth regime, precipitation characteristics and stability would be useful. Also are these predominantly stratiform or broken cloud-decks? I would also suggest to contextualise your findings in the discussion section in terms of how far you would be comfortable to extrapolate your findings beyond the optically thin (I assume), single-layer cloud regime that you explored here in greater detail.

2. You argue the utility and necessity to evaluate and improve kilometre-scale simulations. In this paper you provide a pathway to improve the simulated net CRE for “kilometre-scale” ICON simulations for Arctic low-level clouds, which may even yield to improvements to similar cloud regimes simulated in other regions of the globe. In order to use this approach more widely, it would be helpful to be aware of its potential limitations within the range of “kilometre-scale” grids. Here, you show results of a particular configuration of horizontal (1.2km resolution) and vertical resolution. Terms like “kilometre-scale”, “convection-permitting”, etc. are often used in the community for a range of resolutions ranging from, say, 1-5 km, which apply all kinds of vertical grid refinement within the boundary layer. How valid do you expect your conclusions to remain across the range of spatial resolutions that fall under the category “kilometre-scale”/”convection-permitting”? Would you expect your TKE fix to droplet activation to work equally well at a (say) 5km grid spacing, or when only half the vertical grid spacing is applied?

3. In section 3 during your evaluation of surface radiative quantities you argue that you can compensate for the temporal irregularity of your model output (every 3h) by the increased spatial coverage and thus increased sampling of spatial variability. This essentially assumes that spatial and temporal variability are equivalent. This is assumption is commonly made during simulation-observation comparisons. Can you demonstrate this to be valid though for radiative quantities subject to a diurnal cycle?

4. Your analysis of biases regarding net CRE is focused on the period of 2-5th of June. In L234 you state that you select this period because you largely are dealing with single-layered low-level clouds and have a high density of flights. I am assuming that the bias in CRE (Fig. 4) is also largest during this time period and for this particular cloud regime as well? Given the significance of the analysis that follows for the overall manuscript, I would include a couple more sentences on this selection for clarity.

5. I agree with your general sentiment conveyed in the introduction and conclusion sections of this manuscript that high-resolution LES simulations are quite limited in their spatial and temporal coverage and that coarser-resolution simulations allow longer-term evaluations over larger domains. Yet I wonder, if you are not subject to the same limitations in this particular application, since you are restricting this evaluation to the location of 15 linear flight tracks within a particular region (although admittedly you can afford to simulate more flight hours), and most of your analysis is focused on the period June 2nd-5th. The argumentation of the benefits and limitations of kilometre-scale versus LES simulations does not seem an essential part of your analysis. Thus I would consider to reduce the emphasis on this point, as this is not something you actually show.

6. Fig. 6 very clearly shows the bias in simulated cloud properties that are consistent with an overestimation in cloud transmissivity. From the observations you are under constrained and cannot (I presume) say with certainty whether this is a source or sink issue. In your analysis you show that the bias can be fixed by increasing the source in Nd. Can you provide an equally strong argument, that you could not obtain the same improvement, by adjusting the sink? I think this could be done in the context of a discussion of cloud-base or surface precipitation rates, or a couple of additional numerical experiments where you explicitly show that adjustments to the autoconversion rate by: either turning it off altogether – essentially shutting off warm rain – or reducing its efficiency, does not yield the same kind of improvement.

Specific comments

1. L44: This seems like a somewhat random selection of LES studies in the Arctic and by no means complete. I suggest to either include a comprehensive list of references, or to make it clear that this list of studies is merely exemplary.

2. L48ff: In addition to the representation of in-cloud turbulence and cloud-top inhomogeneity, LES setup also allows the study and evaluation of microphysical processes (e.g. Ovchinikov et al (2014), Solomon et al (2015), Fridlind et al (2017)) and aerosol-cloud interactions (e.g. Possner et al (2017), Solomon et al (2018), Eirund et al (2019)) at scales where the dynamics and thermodynamics are largely resolved. Since you identify the representation of CCN and the activation process itself as one of your primary sources of bias regarding net CRE. It seems fair to mention this here.

   Refs:
   1. Ovchinikov et al (2014): doi:10.1002/2013MS000282 (JAMES)
   2. Fridlind et al (2017): doi:10.1029/2007JD008646 (JGR)
   3. Solomon et al (2015): doi:10.5194/acp-15-10631-2015 (ACP)
   4. Possner et al (2017): doi:10.1002/2016GL071358 (GRL)
   5. Solomon et al (2018): https://doi.org/10.5194/acp-2018-714 (ACP)
   6. Eirund et al (2019): https://doi.org/10.5194/acp-19-9847-2019 (ACP)

3. L102: What is your reasoning for using the all or nothing cloud-cover scheme? Did it impact your results?

4. L132: “The daily averaged observed albedo is parameterized as a function of day of the year”. I did not follow this. Did you not simply prescribe the daily mean albedo value from the full sea-ice covered surface observations. So how is it a “function” of the day of year?

5. L177/178: Why did you not fix the sea ice fraction in a similar fashion as the sea ice albedo in your simulation setup to exclude the impact of biases from essentially prescribed surface properties?

6. L217: I agree with your conclusion that the underestimated cooling in the solar spectral range is likely due to an incorrect simulation of cloud transmissivity, rather than remaining biases in surface albedo. As this a central aspect to your overall argument, I was wondering if you could not show this explicitly. Do your conclusions remain the same if you restrict the phase space your analysis of the observations to surface albedo values < 0.8 such as to match the simulations?

7. L231: I personally would argue that cloud water content is to first order a thermodynamic variable and thus also a macrophysical variable that is adjusted by microphysical processes (i.e. the efficiency of autoconversion/accretion in warm-phase clouds anyway). Especially in a model with saturation adjustment I have a hard time referring to qc as purely microphysical, but can be convinced.

8. “it shows a slight underestimation”. Can you quantify this? What is the average/median cloud depth?

9. L260: I would suggest to be more specific/quantitative here, as this is a key argument in your assertion that the bias stems predominantly from biases in cloud water content and droplet concentration. For the typical cloud optical depth seen in your simulations or during the observation record, how large would a geometrical cloud depth bias have to be to affect transmissivity substantially? How does that quantity relate to your biases assessed?
   In that context, I am not sure Fig. 5 is best suited. I wonder if a PDF-based comparison is not more informative. Cloud transmissivity is strongly non-linearly related to geometrical cloud depth. Thus biases in the distribution of geometrical depth (although means may agree), could induce substantial biases in mean cloud transmissivity. To follow your line of assertion, the argument that cloud depth biases are unlikely to contribute signficantly should be strengthened quantitatively.

10. L270: "droplets plus ice crystals": essentially droplet concentration at >1cm-3. I would not expect to see any impact of ice in the shown range. It may be worth to state explicitly.

11. L300: Can you provide a quantitative estimate of rain rate? Do you have any constraint here from the observation?

12. Fig 2: In the range of surface albedos between 0.6-0.8 where the number of occurrence is highest, the simulations show a much considerably narrower range of Fnet,sol than the observations. Do you have any idea whether this is indicative of a model bias, or simply a sampling issue between the simulation and observation datasets?

13. Fig5: I personally find it hard to draw quantitative conclusions from this plot going beyond the overall range of values in observations and simulations. You can sort of see that geometrical cloud depth is likely underestimated, but its hard to tell due to the many overlapping points where the real density of points is. As suggested previously, I wonder if a PDF comparison would not be more informative

14. Fig. 6-8: Panel a): I am fairly sure the ICON and ACLOUD lines are swapped? Otherwise there would be a mismatch between your figure and the discussion and the results of sections 4.2ff. Panel b): The ACLOUD in Fig6 is slightly different to 7/8. Why?

15. Fig. 9: Ultimately the net cloud-radiative effect is of interest, but your argument primarily relates to CRE_sol. What is the impact of CRE_sol alone?

16. Table1: I personally would find a a total number of included flight hours as part of the caption helpful.

Edits

1. L165: suggested rephrase "to the previous comparison" to "to the previously used model setup".
2. L248: "the the"
3. L291: suggest rephrase "than observed the mean of" to "than the observed mean of"
4. L423: Typo? Underestimation of geometrical cloud depth, right?
5. L424: "represent" instead of "simulate" (since you do not really simulate it)