Reviewer Comment in response to:

What surface radiative fluxes reveal about Arctic cloud modelling
Accuracy

Ya.l Le Gars, Jean-Christophe Raut, and Louis Marelle

Overview:

This study evaluates how well the WRF mesoscale numerical weather prediction model simulates downwelling shortwave and longwave radiation under cloudy sky conditions compared to measurements collected on the N-ICE2015 ship-based research campaign in the Atlantic sector of the Arctic Ocean between January and June 2015. The dominant controls on downwelling longwave biases are assessed using a simplified two-layer model, sensitivity tests are implemented to assess the response of the model to changes in sea ice representation and cloud droplet number concentration, and the influence of boundary layer structure is assessed by separating the cloudy events into 'surface-coupled' and 'surface-decoupled' regimes. The results reveal a nuanced picture of the model performance and response to different forcings, in particular highlighting the sensitivity of surface radiative components to simulated liquid water path, albedo/ sea ice cover, and boundary layer structure. A key result is that important radiative biases in the model are not easily removed by adjusting a single parameter – as these tend to introduce compensating biases. Instead a consistent physical representation of all these processes are important to generate accurate simulations of the surface energy balance in the Arctic.

General comments:

This study is generally well written, and the figures are clear. I congratulate the authors especially on essentially no grammatical or spelling mistakes!

The problem of accurately simulating Arctic clouds to understand the Arctic surface energy budget, how it is changing, and what it means for the future, is urgent and important. In that sense studies such as this that dive deeply into where model biases are coming from are really important for understanding what needs to be done to make forward progress. This is not an easy thing to do due to the variety of interdependent processes, the complexity of the system, the assumptions involved in both the model, the observations and the methodology. In this sense I think this study has done a good job in keeping the analysis concise and physically motivated.

I do think that there are several major areas where this study could be substantially improved, and because it is a good and thorough study, I think it would be a missed opportunity for them not to be addressed. My major points are as follows:

1) I am struggling to understand why the authors chose to limit their analysis to comparisons with data from the N-ICE2015 campaign when there have now been sever other observational campaigns collecting radiation data in the Arctic sea ice that span multiple years, seasons, and some of which have more comprehensive datasets than N-ICE (see MOSAiC, SHEBA, AO2018, ARTofMELT). Some of these campaigns have collected measurements of liquid water path – which would go along way towards helping to answer some of the outstanding questions in this study, as well as co-located measurements of cloud base height and more frequent radiosonde launches. The data are all published and easy to access, the analysis would be same, but the statistical results would be much more compelling. Perhaps there is a resource limitation with model runs, but this ought to be less of an issue now than it used to be - I don't think there is a good excuse to not include more data in this analysis.

2) Although I found this study very interesting, I am finding it difficult to see how this drives the community forwards towards improving the representation of the modelled cloud radiative effect at the surface in the Arctic. Many of the results support things which we already know (the high sensitivity of cloud radiative properties to small changes in liquid water path when liquid water path is low, the importance of albedo and thermodynamic profiles, the seasonal variations in physical drivers). It's not completely clear to me (a) why the authors have targeted a WRF-NICE2015 evaluation specifically in order to glean more information that can be used to improve Arctic cloud and energy budget simulations, (b) how these results can now be used to help address the hard problem of improving the simulations, and (c) how these results generalise more broadly to some of the models discussed in the introduction (GCMs, regional models). I think these deficiencies can be addressed simply by adding a more thorough literature review of what previous model evaluations studies have found, adding some specific motivation for why they have selected the model and analysis they have and how it fits into the bigger picture, and adding some further discussion about what the next steps forward should be now. I think this would greatly improve the study.

3) The organisation of the text makes it a bit hard to follow at times. Specifically, the results section (section 3) is very long, and many of the subsections here contain background, methodological description, and discussion in addition to results. I think it needs reorganising so that all of the methods are in the methods section, and discussion in the discussion section ect – this would help make it more readable and also more useful. I don' think that the introduction and the data and methods section accurately prepare the reader for what to expect currently. This should be easy to fix with some re-organisation.

4) I think the data and methods section is deficient. There is very little information on why the specific methods have been selected, there is no information or discussion of measurement uncertainties and how they might impact the analysis. There are a lot of methodological choices made in the results section that are not introduced in the data / methods sections – these should be outlined in advanced to help the reader make sense of the direction and purpose of the paper.

5) There are several occasions where sentences are ambiguous or lacking in specific detail that leaves the results open to different possible interpretations, these should be addressed and hopefully I've mentioned them all in specific comments below. Please check that every sentence, especially when stating results, is clear and unambiguous – in particular by clearly stating what you are referring to in each sentence (models / observations / wrf runs / variables) rather than relying on the reader interpreting context from previous sentences. Adding specific numbers rather than words such as 'strong' or 'less accurately' would also help.

Specific comments:

- Since the cloud fraction comparison is discussed in detail, and included in the abstract and conclusion, I think some data for this needs to be included – I would prefer to see a figure in the main text. But there should at least be one in the supplement.
- Line 35: A formal definition of the cloud radiative effect at the surface is warranted here.
- Line 51: What do you mean by "derive model cloud evaluation"? Can you be more specific about what you're trying to achieve here?
- Line 55: Why WRF? Why NICE?
- Figure 1: Please describe what the sea ice cover concentration map is showing in the caption and include the reference. I think it would also be good to have mean sea ice cover during N-ICE plotted on the background of panel B?
- Section 2.1: A short discussion of the uncertainties in the radiation measurements are warranted here, to contextualise the magnitude of the bias results.

- Section 2.2: Can you justify why you have selected these particular parameterisation schemes? And what influence you expect these decisions to have on the results of this study and how they might be interpreted?
- Figure 2:
  - Why are air temperature and RH shown with an hourly resolution but the radiation shown with daily resolution. All quantities show daily cycles, so it seems odd to not have them at the same resolution. I think the hourly resolution would be more informative for the radiation data.
  - Are all the data gaps during the ship repositioning? Or are there other reasons for the missing data?
  - Why does the simulated RH drop to zero briefly sometimes? It would be nice to start the y-axis at 40% RH so you could see more of the main data variability.
  - I think this figure could be improved by highlighting the three periods described in the text, and indicating where the storms are that are discussed for P1. Is missing data completely due to ship movement? Or are there other periods of missing data in there?
- Are the three periods you discuss the same as other studies have used?
- Clear sky thresholding (line 162) – where these numbers determined by statistically comparing the distributions? Or just eyeballed? Also I don't think you need to use 'centered around' – just give the actual specific value..
- Line 166: I would like to see some more about this in the discussion section – optically thin clouds can have an outsized and non-linear impact on the surface radiation components as you demonstrate. Can you speculate what impact underestimating the occurrence of optically thin clouds might have on the results of this analysis and their interpretation?
- Line170: "The latter results from an underestimation of the upwelling component." Please exapand on this, how do you know this? Why is this the case?
- Line 174: Please describe how you are determining whether or not the simulated and observed distributions "agree well" and why? The uncertainties of the measurements and model are important for determining whether they actually agree well or not.
- Line 180: It would be nice to see how sensitive the overall study results are to this thresholding on both datasets
- Line 185: "Primarily" – replace with likely? I don't think you've actually shown this is true or have the data to do so?
- Line 189: A plot of how the cloud occurrence varies with the threshold would be helpful here, even in the supplement.
- Line 189: "and shows weak temporal overlap (50 %) " – what does this mean? How do you calculate a percentage temporal overlap? Please show this if you're going to draw conclusions from it.
- Line 189: Change to: indicating that "the observed" opaque clouds are rarely simulated
- Line 193: What does slight to moderate mean and how have you determined this? Please show these results in the supplementary.
- Section 3.2.1: The discussion of cloud occurrence in models versus obs needs supporting with data / plots if discussed in this much detail, but I understand it's perhaps not the main focus, so could move the discussion to the supplement. I prefer the former though because I think a lot of readers might be interested in this.

- Throughout, there is not much discussion of the fact that cloud height and vertical temperature profile errors could be another reason for errors in downwelling longwave variation in the model compared to observations. I understand that this is difficult to validate with the data you are currently using, but I think you need to caveat that this might still be the case in the abstract, discussion, and conclusion. You are effectively adding any errors in cloud height and temperature profiles into errors in cloud emissivity by assuming that the T cloud base = T surface, and therefore possibly overestimating the sensitivity to cloud emissivity. This is a really important point to discuss, since it is one of your main conclusions.

- Line 199: “The distribution shape is retrieved” – what does this mean? Please be specific.
- Line 199: Please can you give the specific value instead of “about” ?
- Line 206: P3 radiative analysis. Low values of downwelling longwave are overestimated and high values are underestimated. But for the shortwave, low values are overestimated while high values are underestimated. This suggests that in the model, you have more clouds with a high optical depth (less shortwave transmission) but less longwave emission (i.e. colder clouds) – could this be errors in cloud height? Too many high clouds? Or not enough liquid water? Maybe add some of this to the discussion if it’s not already covered.
- Line 209: “Likely reflect biases in cloud optical depth” –why likely? what else could they be? What about multiple scattering and cloud edge enhancement?
  Does surface albedo impact downwelling shortwave by a significant amount?
- Please make sure you clearly define what you mean by CRE somewhere – there can be several different definitions of this.
- Line 222: The IAOOS dataset should be introduced and described in the datasets / methods section.
- Line 225: I think the assumption that the cloud base temperature = surface temperature in the observations is awkward and problematic in this analysis. Some specific comments to address:
  a) How does the Maillard identification of cloud presence line up with the observed cloud presence from the radiometers (are they seeing the same clouds?). I ask this because the 808 nm lidar will be biased towards seeing low clouds anyway as it will suffer from attenuation higher up.
  b) “We do not have any direct measurements of cloud base temperature” – is sounds like you actually do, at least twice a day from radiosondes. At the very least you should validate your assumption using these data.
  c) There are several other campaigns that have more frequent atmospheric temperature profiles as well as better measurements of cloud base height, like MOSAIC and ARTOFMELT for example.
  d) Another way of testing if you’re assumption is reasonable: If you’re assuming all the clouds have the same temperature as the surface, you’re assuming that all the net longwave variation in the cloudy mode of figure 3 is due to changes differences in cloud and surface emissivity – can you demonstrate whether or not that assumption is reasonable? How much do you expect these to change?
  e) By making the assumption above, perhaps it’s not surprising that you find changes in cloud emissivity to be of such high importance? Can you comment on this in the discussion and what it might mean for your results?
  f) Even if all the cloud bases were located below 90 m, that is not a good argument for assuming that the cloud base temperature and 2 m temperature are the same. Low level temperature inversions are common in the Arctic, you should be able to have a look how relevant these are for NICE by examining the radiosonde profiles, and show some data for that in the supplement.
- Line 233: Please can you explain why you have chosen these thresholds? Are they based on literature from elsewhere? Are they a conservative threshold for cloud presence compared to the LW emission threshold from the observations or not? Perhaps most importantly, are the results sensitive to this choice?
- Line 237: Is this from both the observations and the model derivations?
- Table 2: Please make the caption more descriptive / specific. You’ve defined this in the text but make it easier for your reader to interpret your results by being explicit in the captions. Something like “Relative contribution of the four variables from the two-layer atmospheric emission model to the variance in the difference between the modelled and observed downwelling longwave radiation during cloudy scenes ($\Delta LW^{\downarrow}$) “
- Line 240: Do you have a reference for this exact analysis? If not please describe it in more detail with equations (ideally in the methods section).

- Line 244: The properties you list are specifically cloud microphysics as opposed to cloud properties more generally which might include cloud height, depth, temperature, so change “cloud properties” → “cloud microphysics”
- Line 245: The values in table 2 are reflecting your assumption that T2m=Tcb in the observations as well as biases or problems in the model, how can you tell how much of this just comes from the former or the latter? I think some further analysis is warranted here.
- Line 246: You go from discussing the contribution of different variables to the variance in the model-obs bias (I think), and then you talk about “errors in Tcb” – are you now talking about the contribution of Tcb to that variance – or are you talking about something else? The difference between Tcb in the model and the assumed Tcb in the observations perhaps? Please be very specific and consistent with your language here. If you’re taking about the latter, you probably need a new paragraph, and it would also help to see some of this data so the reader can interpret what you mean by substantial (or at least some numbers).
- Line 247: The “temperature profile misrepresentation” could presumably either be a misrepresentation in the model, or an incorrect assumption that T2m == Tcb in the observations – please add that detail here.
- Line 248: Change “well captured” to “well captured by the model” – it’s important to be specific. Also change “with 95% within the first model level” to “with 95% of observed cloud base heights falling within the first model level” (if that is really what you mean).
- Line 249: “Thus, $T_{cb} \approx T_{2m}$ during P1” - do you mean in the model here? Please be specific!
- Line 249: Do you mean less accurately captured compared to the IAOOS observations? If so, please show some data, and also introduces the IAOOS as a dataset that you’re using for analysis in the data and methods section. In what way are they less accurately captured? Is the IAOOS data showing more low cloud bases and the model showing higher? Or the other way around?
- Line 251: Versus what value for the observations? Please show some of this data / comparison if you’re using it to derive results.
- Line 256: Why do you choose modelled cloud water specifically here? When you have also stated that radiative errors could be driven by particle size? I think that this is the right thing to focus on because the impact of cloud water on the LW emissions is so large – but guide a reader who might be less familiar with the relative radiative effects of changes in cloud microphysics.
- Line 259: this could easily be solved by using one of the many Arctic field campaigns with microwave radiometer data…
- Table 3 caption: add ‘relative to observations’ and define which way you’ve calculated the bias (model minus obs?)
- Table 3: Please explain in the methods section why you have chosen to focus the evaluation on the statistics MB, NMB and NMAE
- Line 268: Please be specific about what you mean by low.
- Line 267: Instead of using the word strong, can you put some numbers to this? The word ‘strong’ is quite subjective without relative context.
- Line 269: What do you mean by mirroring?
- Figure 5: Can you use the same colour bars and y-axis for the all the lw plots and the same colour bars for all the shortwave plots so that the reader can compare the magnitude of the errors between the periods?
- Line 286: I believe this could also come from too much ice? I.e. clouds that are too deep generally?
- Line 290: Is this still the case when you only consider cases when ISWP is > 0 ? Certainly, radiative biases in the liquid only cloud are not related to ISWP misrepresentation, but I could imagine that downwelling SW biases for deep mixed phase clouds could be related to misrepresentations of cloud depth and total water content (liquid +ice), can you address that?
- Section 3.3, I am a bit confused about the ordering of this. WRF-SEAICE clearly does a much better job of representing the surface albedo – surely it would make sense to show this first, and then present and discuss the radiative bias results from the WRF-SEAICE run in the main paper rather than the WRF-CTRL run? With the WRF-CTRL results in the supplement?

- 319: Please reference figure 6 here. It took me a while to make sure I was understanding this result correctly due to the lack of specificity in the sentence, ‘reduced SW underestimation’ is a double negative and the result is unclear (is it now an over estimation? – I see now from fig 6 that it is not), and a ‘systematic negative bias persists for LWP’ – the immediate interpretation of this is that there as a bias in LWP – but I think you mean ‘systematic negative bias in downwelling shortwave persists for LWP values greater than..’ . Please be very specific about what you are saying both here and elsewhere.
- 320: Later in the text you also say this could be due to phase partitioning?
- Figure 7 – the caption says WRF-CTRL but the figure says WRF-SEAICE?
- Line 335: You haven’t shown or mentioned the LW changes in the CDNC runs – maybe add to the supplement? It would be interesting to know how the magnitude of the CDNC forced changes in downwelling LW compares to the biases.
- Line 337: Maybe for the discussion section – what do we know about the variability of this on an event basis? The reference value of 100 cm-3 might be good for a seasonal mean, but the radiative biases are event specific, and I suspect you can get very low aerosol events and very high aerosol events in all seasons. Can you comment on whether getting the event scale CDNC closer to reality might help improve these biases?
- Line 340: change “too efficiently” to “too efficiently in the model”
- 3.3.5: this section starts by some background / literature references, and then some methodology. This paper would be more readable if you could move the background / literature from each subsection into the introduction, and the methods from each subsection of section 3 into the data and methods in section 2.
- Figure 8: I find it very difficult to distinguish between the purple circles and blue squares or to see any real difference in these distributions from these plots, so I was left interpreting differences from the NMB calculations. I think you can at least make the two sets of points more contrasting colours and shapes. Perhaps you could also add a centroid like you have on the previous plots?
- Line 370, can you do a statistical test between the two different distributions to confirm whether these differences are statistically significant? I suspect that they are, but that would help support the point.
- Line 387: Presumably also CDNC would be different between coupled and decoupled clouds – it would be interesting to so how varying the CDNC impacts the radiative bias separately for coupled versus decoupled clouds.
- Line 399: Be clear on what is a novel result of this study and what is not – I think the dominance analysis identified that cloud emissivity was the leading driver of the LW errors – but the fact that this is primarily driven by LWP is not a direct result of this study, but just based on physics / past studies? If I’m correct, then please rephrase this sentence.
- Line 405: add “when averaged over the season” (i.e. there is still an overestimation of SW at low LWP in all seasons).
- Line 410: I think you need to look at total cloud water content before you can state this as a conclusion – there may be no systematic relationship when you just look at ice and snow water paths because you’re considering both ice and mixed-phase clouds, but there could still be a relationship with total cloud water content (ice + liquid), i.e. it might be the ice in the mixed-phase that is important for the cloud radiation errors in the high LWP cases. These will be the deepest clouds – if the model is not capturing the ice phase the clouds might be too optically thin even if there are capturing the liquid phase correctly.
- Line 415: I wasn’t convinced that this was a ‘strong’ relationship – some statistical testing on the differences would help. As would more data from a wider range of campaigns.
- Code data availability: Where can we find the IOASS data?