Response to reviews of 'Attribution of sea ice model errors enabled by new induced surface flux framework

This document is set out in the following way:

- Reviews reproduced, with inline response in red.
- Summary of changes to the document.
- Tracked changes version of document.

Review #1 (Francois Massonnet)

The authors have to be commended for the hard work in reshaping and revising the manuscript. I appreciate the work done to expand their framework to include subgrid ice thickness distribution information, to estimate the role of oceanic heat convergence in setting the biases, and the discussion on the validity of the approach (especially the new Fig. 5, Fig. 7, Fig. A1; these figures are very important for explaining the methodology and were clearly missing from the first version). It is fair to say that the manuscript has gained much in scientific quality. The structure of the text is still a bit dense, with sections 4 and 5 being hard to digest in one reading, with a lot of equations and formulas within the text.

We agree that the paper remained rather hard to understand, and have refocused and restructured the paper once again.

As part of the overhaul of Section 4 (method description):

- A fuller summary of the ISF framework is now given in the Introduction.
- The ISF framework motivations are described more clearly during the model evaluation in section 3.
- More detail is given early in Section 4 about why the ISF framework is needed and useful.
- Conversely, the detailed description and derivation of the surface flux formula has been moved to a reintroduced Appendix A. Responding to the reviewer's suggestion, many equations have been moved from the text to their own lines.
- Only two examples of the ISF framework are now given (as opposed to the previous 4).

Section 5 (results) has also been overhauled by the use of subsections. In 5.1, Arctic-wide ISF biases and internal variability are presented; in 5.2, spatial patterns discussed, in 5.3, feedbacks discussed (as suggested below); in 5.4, residuals and observational uncertainty compared. Section 6 has also been divided into subsections.

Also, adding subsections could help for orienting oneself in the manuscript. Finally, I would move the feedback discussion, currently in the "Discussion" section, as a sub section of Section 5.

We have added subsections in Section 5, and moved the feedback discussion as suggested.

An interesting finding, which I did not fully appreciate during my first review, is that observational inaccuracy is so large that it participates greatly in the estimated model bias (e.g., page 12, lines 34-36). This statement should be better emphasized, for example in the abstract, because it is of high

importance for future evaluation studies. Most researchers think that the "model minus obs" quantity reflects a genuine model bias, while it can be caused by many other factors including observational inaccuracies.

We have emphasised this finding in the abstract and in the Conclusions, as well as clarifying our reasoning for it in the new section 5.4.

I still have a couple of minor comments, which when addressed will give the green lights for publication. My comments are referenced with "p/l" where p is the page number and l the line number.

1/9 "allows the local dependence": I have the impression a verb is missing --> "allows emphasizing the local dependence" or "allows singling out the local dependence"

The verb is 'described' – the wording of this sentence has been altered.

1/13 after "variable", there is a ". ' ." --> remove

Change made as suggested.

1/35 "Massonnet" takes two n's.

Apologies – change made as suggested.

2/6 "Ice volume is, to first order, proportional to the heat required to melt the ice" is somewhat imprecise. "heat required to melt the ice" might be interpreted as "heat needed to trigger melt onset", while what the authors mean is "heat required to melt the ice out" or "heat required to entirely melt the ice".

We have altered this to the reviewer's first suggestion – 'heat required to melt the ice out'.

2/19 Calling HadGEM2-ES a "member" of CMIP5 is troubling because "member" is usually a term reserved to mean "individual realizations" within the same model.

We have changed this to simply calling HadGEM2-ES a 'CMIP5 model'.

2/19 "ensemblewhich"

Corrected.

3/33 Does the definition of "Arctic Ocean" come from somewhere, or did you make it up yourselves? How was the region defined? Can you elaborate in Fig. 1 caption?

The region was custom-designed – we have added elaboration to both the figure caption and reference.

4/5-8 The second part of the sentence "The approach..." seems to be missing something: "... and to use as reference datasets [WHAT?]"

The wording of this section has been amended.

4/9 Consider "concentration" instead of "fraction" to be consistent.

Change made as suggested.

4/11 If I'm not mistaken, PIOMAS also assimilates sea surface salinity. Check it out.

We have looked at the main PIOMAS references (e.g. Zhang and Rothrock, 2003; Schweiger et al, 2011) and cannot find a reference to this. If the reviewer can identify a reference we will be happy to add this statement.

5/5 "From 1980-1999" --> "From 1980 to 1999"

Change made as suggested.

5/21-26 I find your discussion on Fig. 3 interesting but it deserves more physical explanations. (This is somehow briefly discussed later at page 13, lines 11-22). Essentially, your results suggest (Fig. 3) that the October to April change in thickness is overestimated with respect to reference products, and you find that there is a "clear association between areas where modelled annual mean ice thickness is biased low, and areas where the modelled seasonal cycle is overamplified". The physical reason is - I think - that the negative ice-growth ice thickness feedback is much stronger for thin ice. When ice is biased thin, much more ice is grown in winter than if there was no bias. We have somehow shown this in our recent paper (Massonnet et al., 2018, https://www.nature.com/articles/s41558-018-0204-z) with the introduction of the "Ice Formation Efficiency" (IFE). The IFE is the regression between ice volume produced during winter on summer sea ice volume. It is negative in all CMIP5 models: a summer sea ice volume negative anomaly results in a higher-than-normal ice production; and it is strongly mean-state dependent, being more intense at low thickness (see Fig. 2a of our paper). It might be good to comment the results of Fig. 3 in light of this paper.

We have added a comment to the effect that the thickness-growth feedback explains this association, and noted that it is common across CMIP5 with a reference to Massonnet et al 2018.

6/26 See my comment 2/6

A similar alteration has been made.

7/4 By "surface flux", do you mean "net surface flux"? Be more specific.

Yes, net surface flux – this has been clarified as suggested. Indeed, when first introduced, we have expanded this to 'total downwards surface energy flux'.

7/9 When giving the analytical formula for estimating the induced surface flux bias, it might be worth already mentioning here that the partial derivative needs to be evaluated at a reference state (otherwise the equation at line 9 gives a functional dependence but is not of practical use).

Clarification made as suggested.

7/16 Use "ice concentration" instead of "ice area"

Change made as suggested.

7/23 The equation (1) features a term $\gamma^{cat}_{ice-REF}$ but in the text that term is referred to as γ^{cat}_{REF}

8/26 "... variables which have the required property to affect the surface flux on timescales shorter than on which they affect each other". (Also at page 7 line 5). Can you be a bit less vague here? I understand the general idea, but can you give rough time scales of how surface fluxes respond to these state variables and rough time scales of how these variables affect each other?

This has now been clarified: the point is that the independent variables all affect the surface flux instantaneously in the model (in the real world, there is a time lag associated with ice and snow thickness; we have added a paragraph regarding this in the Discussion).

9/23 Add white space between 0 and "m".

Change made as suggested.

9/24 Expanding the induced surface flux bias framwork to account for the ice thickness distribution is a good idea, but it readily opens new problems, e.g. as how to estimate biases at the subgrid scale. I somehow follow the authors reasoning here, and think there is just no best solution given that the problem is underconstrained (too much information to create). I think it would be useful for the general audience to sketch what is the geometric idea behind your redistribution of thickness biases among categories. I'm asking because the formula that you give at line 26 does not give a good idea of what is going on physically.

A figure along the lines you suggest has been included in Appendix A.

11/3 "anomalous" --> biased?

Change made as suggested.

11/4 "anomalous" --> biased?

Change made as suggested.

14/19 There are two consecutive dots.

Corrected.

14/24 A white space is missing between SW and bias.

Corrected.

14/24 A white space is missing between bias and "is"

Corrected.

15/3 A dot is missing between two sentences.

Corrected.

17/22 There is a white space before the "In"

Corrected.

17/32-35 I am not sure to understand the meaning of this paragraph: do you mean that the approach can also be used to understand the sensitivity of model results to the inclusion of new processes?

Figure 2 Readers might be puzzled that for panels b, c, d, the model (black) lines are different from panel to panel. I understand that the average thickness is computed for the corresponding observational region of coverage each time (this is acknowledged in the figure caption). However, does the time period for averaging change as well from panel to panel? It looks it does not (1980-1999 according to the caption). However, the ERS product runs from 1993 to 2000. Does that mean that averages are not done on consistent periods? Also, it would be good to use similar y-axis limits in panels b-d and to set the lower bound at zero.

The reviewer is correct that the ERS comparison was not consistent in terms of time period. This has now been corrected, and the associated model bias quoted in Section 4 amended (there is no qualitative change).

Figure 3 Font size appears to change from sentence to sentence in the caption.

Corrected.

Figure 4 Change "anomaly" in your difference plots to "bias".

Corrected.

Figure 4 The units of y-axes are W/m-2, they should be W/m2

Corrected.

Figure 5 In panel C, the units W/m2 are missing.

Corrected.

Reviewer 2 (Anonymous)

Dear Authors,

Dear Editor,

I have now completed my reading of the revised version of « Attribution of sea ice model biases to specific model errors enabled by new induced surface flux framework ».

I well appreciated the energetic and thorough effort to address both (sometimes contradictory) reviewer's comments.

However, the main comment from my first review has not been successfully addressed, I'm afraid: science looks a priori fine, but awkward writing and logics makes proper understanding of the paper hard.

The methods section is at the centre of this: too much time is spent on the details and too little is made to explain the very main points of the reasoning, the assumptions and the validity of the method.

The ultimate strengths and weaknesses of the methods, which are essential points, are hidden into the limbs of the paper.

I'm afraid this more or less applies to the whole paper. I'm still encouraging, though. I'm pretty sure all elements are there to turn the paper properly.

We apologise that the clarity of our presentation is still lacking. There was no deliberate intention to ignore the concerns raised in the first round of reviews, but in our efforts to address the scientific concerns this aspect was probably somewhat neglected.

We have taken time to consider what the main points of the paper should be, and have refocussed the document accordingly, with particular attention to Section 4 (methods).

Accordingly, the details of the surface flux formula, and their derivation, have been moved to a reintroduced Appendix A. In place of this, we have taken some more time to describe how we have arrived at the ISF framework, and why it is useful. In particular:

- In section 3 (model evaluation), the sea ice biases are described more systematically to motivate what follows. Firstly the extent and thickness biases, and then the mass balance bias, from which the surface flux evaluation is motivated. The surface melt onset, as a primary driver of sea ice albedo, is now also evaluated directly in Section 3, and a figure has been added to describe this.
- The model biases are then summarised, followed by the main ISF motivation: we have many model biases that probably affect the sea ice mass balance; we now want to quantify precisely how they affect the mass balance.
- At the start of Section 4, we now describe how the ISF framework arises quite naturally from this aim: the function we use to approximate surface flux is designed to encapsulate all the relationships by which the evaluated variables can affect sea ice mass balance.
- The chief advantages of the method (linearity of induced surface fluxes, quasi-independence of variables) are described in more detail.
- Now, only two examples of ISF calculation are given (melt onset and downwelling LW).

We have also refocussed some other parts of the study. In particular:

- There has been a general aim to clarify that we are trying to attribute model bias in **sea ice mass balance** (which is the principal driver of the sea ice state)
- Section 5 (ISF results) has been restructured and divided into subsections. 5.1 – Arctic aggregate results and internal variability; 5.2 – spatial variability; 5.3 – forcings and feedbacks discussion; 5.4 – ISF residuals and observational uncertainty.
- A fuller description of the ISF method has been given in the Introduction, for greater prominence.
- More prominence has also been given to the finding that observational uncertainty is the greatest cause of uncertainty in the ISF biases themselves.

- In turn, the reasoning behind this conclusion, in section 5.4, has been clarified. We evaluate the total ISF bias with respect to the net radiation biases and the ice heat uptake bias (themselves only proxies for the true surface flux bias which cannot be known). We compare the theoretical ISF errors (Appendix B) to the ISF spread amongst reference datasets, and find the latter to be far larger.
- Section 6 (Discussion) has also been split into subsections.
- Section 7 (Conclusion) has also been refocussed, to concentrate more on the ISF framework itself, and less on HadGEM2-ES in particular.

These are my major points.

1. Make sure the paper is fully understandable.

2. Correctly qualify the confidence in your methods. There have been progresses, but the situation is still blurred. To be clearer on how your method is reliable, you could show 3 things on the same seasonal cycle plot: the mean ISF; estimated minus directly retrieved surface flux with HadGEM; sum of all ISF contributors minus ISF.

We have tried to evaluate the ISF errors in a more systematic manner, in the new Section 5.4, quoting differences between total ISF and net surface flux bias proxies. We also show difference between estimated and directly retrieved surface flux in HadGEM2-ES in Appendix B. However, we do not completely understand the reviewer's specific request at the end of this paragraph. Our guess is that 'mean ISF' corresponds to 'total ISF bias' – but in the last suggestion, what does 'ISF' mean, if it is distinct from 'sum of all ISF contributors'?

We are also not sure that estimated minus directly retrieved surface flux, as an Arctic aggregate, would be a useful quantity to compare to the ISF residuals. Estimated surface flux is not used directly in the ISF calculation – it is only useful in so far as it gives information about error in the surface flux *dependence*, as calculated from the partial derivatives.

If the reviewer could give some more details about his/her reasoning here, we would be happy to consider another figure, although we note that there are already 11 in total.

3. The explanation on the early melt onset problem is not entirely convincing (page 11, lines 3-12). You claim ice melts too early in HadGEM (end of May). Based on the used datasets, melt onset is supposed to happen in mid-late June. However, in my experience, surface melt onset in the Arctic falls quite regularly near June 1st. It was true at SHEBA (29 May, Persson et al JGR 2002), it is true at Barrow I believe. It also seems generally true from passive microwave datasets (Markus et al., GRL2009, Fig. 3). Hence your claim that melt onset is too early sounds invalid, and the melt pond explanation as well. I would tend to think that melt onset looks fine. What could be wrong is that the initial drop in sea ice albedo, that could be happening too fast.

We thank the reviewer for drawing our attention to this issue. We have again checked the product we are using to determine melt onset (Snow Melt Onset Over Arctic Sea Ice from SMMR and SSM/I-SSMIS Brightness Temperatures, Version 3, https://nsidc.org/data/nsidc-0105, based on Drobot and Anderson 2001), and think that our description of the average 1980-1999 melt onset date as measured by this dataset was not satisfactory (Figure 1).

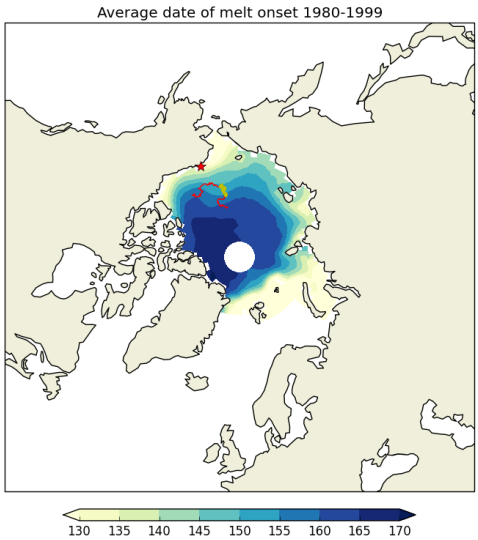


Figure 1. Average date of melt onset 1980-1999 as measured by the dataset used in the present study. Point Barrow, Alaska is indicated by the red star, the track of the SHEBA experiment by the red line, and the track covered during May-June 1998 by the thick yellow line.

As the day 165 marks the rough midpoint of June, it can be seen that melt onset occurs after this date, in 'late June', in only a small part of the Arctic Ocean, next to Greenland. However, it could be fairly said that melt onset occurs after 10th June in a large part of the Central Arctic – that roughly corresponding to the area where the principal ISF anomaly is measured (Figure 6 of our study). While our case was overstated, therefore, we believe that it remains qualitatively true. We will correct our wording here to 'mid-June'.

We note in particular the very steep latitudinal dependence of melt onset across the Beaufort Sea region, with average melt onset measured as occurring before 15th May on the Alaskan coast (Point Barrow is indicated by the red star on Figure 1), and around the turn of the month in the mid-Beaufort Sea, where the SHEBA experiment was deployed (the SHEBA track is also indicated in Figure 1, with the yellow section describing the position of the camp during May-June 1998). We also note that as overall ice thinning was observed at all sites during SHEBA (Perovich et al 2003), it is very likely that thermodynamic forcing on the ice was anomalously high during 1997-1998, and likely that the 1998 melt onset date was early relative to the 1980-1999 average. Hence it is not clear to us that our dataset conflicts with the in situ findings mentioned by the reviewer.

Similarly, looking at Figure 3 of Markus et al (2009) as prompted by the reviewer, a pronounced gradient in melt onset date between the Central Arctic and the Alaskan coast is apparent in all three products. It is hard to determine the exact date from the colour scale used, but it looks as if deep, verging on dark, red colours dominate the Central Arctic, signifying late June, as opposed to the orange colours that occur close to the Alaskan coast, signifying mid-to-late May. (This discussion refers to the 'early melt onset' figures, signifying the first date of snowmelt, which is analogous to the observational product and the surface temperature-derived model proxy used in the present study, rather than the 'continuous melt onset' figures).

In response to the reviewer's question, we will describe the spatial characteristics of the melt onset anomaly earlier, to clarify that it is a Central Arctic phenomenon, and change the description of the modelled melt onset time to 'mid-June' as described above.

4. There are important observational references on surface energy fluxes (Lindsay 98, Personn et al 2002) that could be of help to select which products are more plausible than others.

(Page 6, line 9-12). These two papers analyse the energy budget in the Arctic as derived from in situ measurements in the ~50 Russian polar drift stations, and from SHEBA. Regarding downwelling longwave fluxes these references suggest between 150 and 160 W/m2, and nothing above 170 W/m2. I guess this places your model at the low range and places ISCCP out of the game.

Lindsay et al (1998) is a very interesting and useful observational reference on many Arctic climate variables, and we are grateful to the reviewer for drawing our attention to it; our study would clearly be improved by consideration of Lindsay's findings in comparing the various different radiation datasets. Figure 6 of Lindsay et al shows in the sixth panel a seasonal cycle of boxplots of downwelling LW observed on drifting stations from 1947-1990. The median lines from December-February appear to fall close to 160 Wm-2, lower quartiles close to 140 Wm-2, and upper quartiles between 180-190 Wm-2. Figure 2 superimposes the Lindsay et al figure over the Arctic Ocean mean fluxes quoted in the present study for the relevant datasets; as the reviewer suggests, the HadGEM2-ES Arctic Ocean means are well within the interquartile ranges (close, in fact, to the median); the CERES and ERAI estimates towards the upper end of the interquartile range; and ISCCP-FD well outside the interquartile range, close to the upper end of the distribution. This appears at first sight evidence that the true model bias is less severe than suggested in the present study, or non-existent.

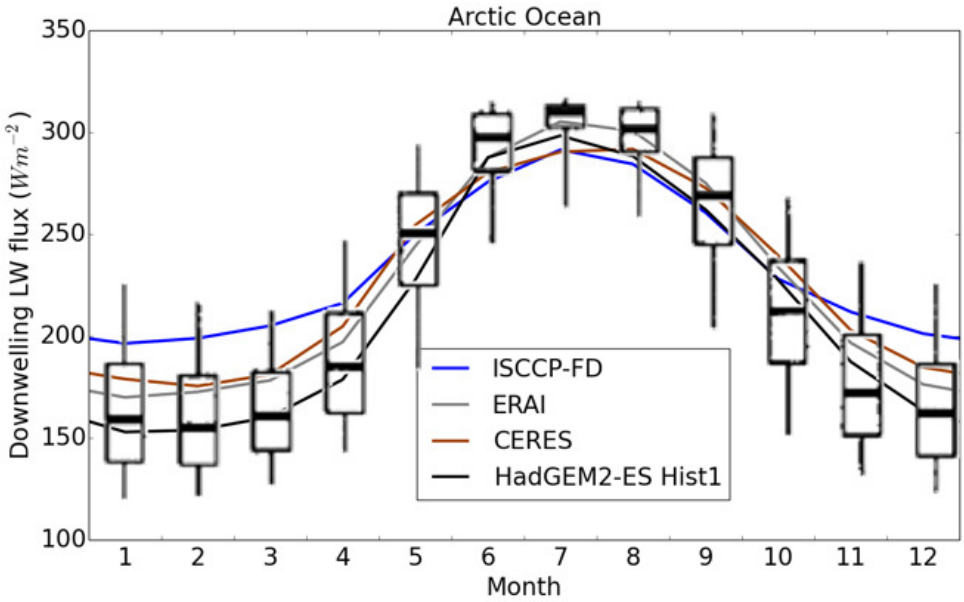


Figure 2. Comparing boxplots of downwelling LW as compiled by Lindsay et al (1998) to the Arctic Ocean mean values for ISCCP-FD 1983-1999, ERAI 1980-1999, CERES 2000-2013, and the first historical member of HadGEM2-ES 1980-1999.

However, as with the melt onset date issue, spatial variability needs to be taken into account. Figure 3 shows maps of Dec-Feb mean downwelling LW radiation for the three reference datasets across the Arctic Ocean region used in the present study, as well as the first historical member of HadGEM2-ES, with the tracks of the drifting stations used by Lindsay et al superimposed (derived from Figure 2 of their study). Though the datasets disagree on the area mean, they show broadly similar spatial patterns: a wide area of low values across the Central Arctic, rising steeply towards Alaska, the Bering Strait and the Atlantic Ocean ice edge. The drifting station tracks analysed by Lindsay et al lie almost entirely within the central cold region. The peripheral regions, which dominate the high end of the distribution, are barely sampled. This introduces a bias to the comparison in Figure 2.

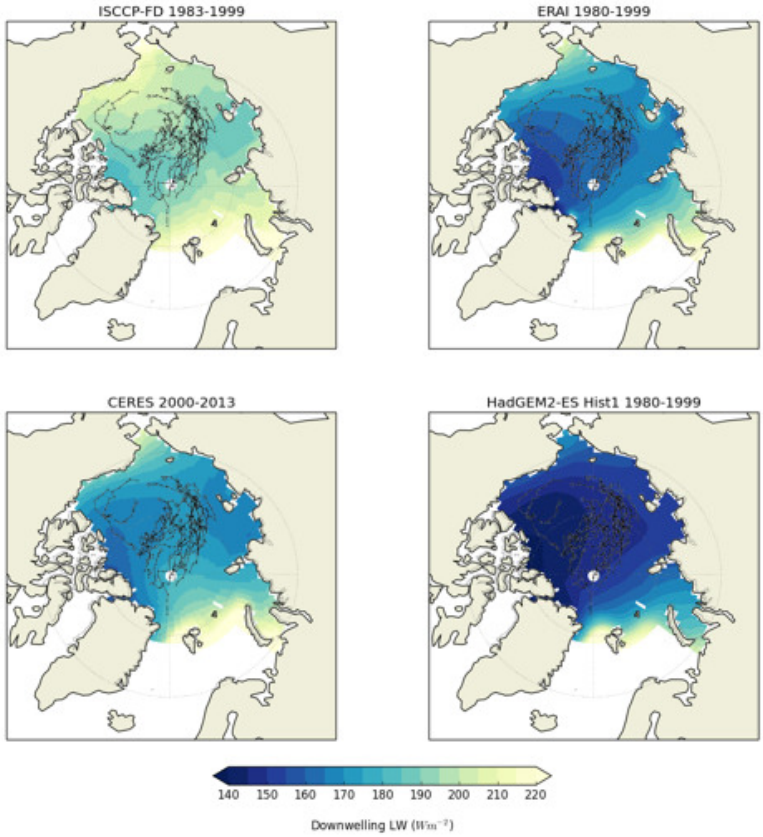


Figure 3. Maps of mean winter downwelling LW datasets for the reference datasets and for the first historical member of HadGEM2-ES, with drifting station tracks analysed by Lindsay et al overlaid.

To account for this bias, a new region 'Inner Arctic Ocean' was defined (Figure 4), which corresponds more closely to the region sampled by the drifting stations. Winter downwelling LW statistics were recomputed for this region for the three reference datasets, and for the first historical member of HadGEM2-ES. In Figure 5 these are compared to the Lindsay et al values. Using the Inner Arctic Ocean region shifts the mean downwelling LW by about -10 Wm-2, and the effect relative to the Lindsay distributions is decisive: roughly speaking, HadGEM2-ES now falls close to the lower quartiles, CERES and ERAI close to the medians, and ISCCP-FD close to the upper quartiles. This is consistent with our finding that HadGEM2-ES is likely to be biased low, but that ISCCP-FD is likely to be biased high (i.e. using ISCCP-FD as reference dataset overstates the model bias). We also note that is it likely that average Arctic downwelling LW from 1947-1990 was lower than that from 1980-1999, due to the underlying warming trend.

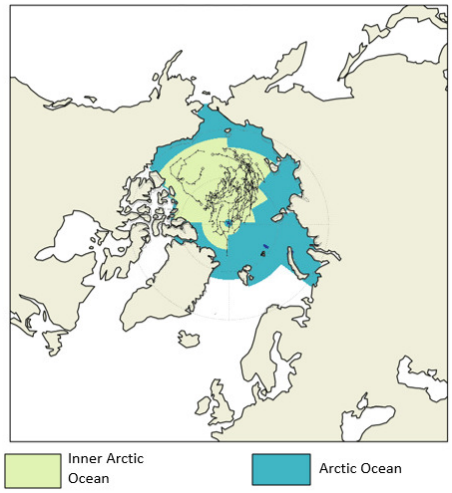


Figure 4. The Inner Arctic Ocean region defined for comparison with Lindsay et al (1998)

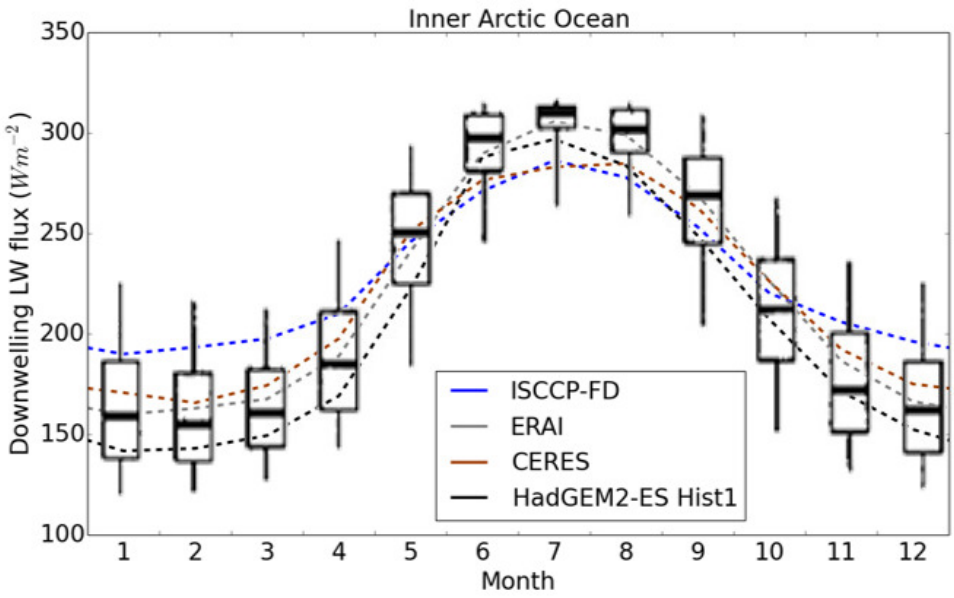


Figure 5. Comparing boxplots of downwelling LW as compiled by Lindsay et al (1998) to the Inner Arctic Ocean mean values for ISCCP-FD 1983-1999, ERAI 1980-1999, CERES 2000-2013, and the first historical member of HadGEM2-ES 1980-1999.

We are less sure that discrepancies with respect to the SHEBA fluxes (Persson et al, 2002, figure 22c) require explanation, as these represent only a single measurement from a single year. The SHEBA track lies very close to the edge of the 'cold' region in Figure 3. Hence precisely how spatially representative the SHEBA downwelling LW measurements were of the Arctic Ocean as a whole is sensitive to the dataset used, and probably also to the weather conditions of the year in question. At an earlier revision of the paper, the SHEBA fluxes were in fact compared directly to HadGEM2-ES

(restricted to a small region covering the drift track). Here, too, the modelled downwelling LW fluxes were found to be biased low by around 20Wm^{-2}.

The discussion of the relative accuracy of the various downwelling LW sources in section 3 has been expanded by comparing the Lindsay et al fluxes to the various data sources as above. We could not show any new figures because the number of figures was already rather high (11) but the qualitative finding that CERES and ERAI are likely to be quite accurate, ISCCP-FD biased high and HadGEM2-ES biased low, will be stated.

A few minor items:

- Make sure your figures are absolutely clear. Fig. 2, 3, 4 could be better labelled to warrant immediate understanding.

We have added labelling to these figures where we thought it would be helpful, particularly where we thought the reference datasets and variables being plotted were hard to identify.

- Explain how consistent time periods are between reference and model products in the construction of Fig. 2 and 3.

Reviewer 1 raised a similar issue. Time periods were not entirely consistent in Figure 2and this has been corrected. In each figure, time periods have now been clarified.

- Don't use "observation" to refer to PIOMAS. Use "reference data set" throughout.

We have made this change in all sentences where PIOMAS is among the datasets under discussion.

- Color scales and axes scales are to be carefully chosen. In particular, on Fig. A1 there are no means to know the order of magnitude of the differences. Use line contours ?

A mistake was made in the contour levels for the difference plots in Figure A1 which made this figure very hard to interpret. This has been corrected.

- Page 6, line 25. In the SM0L framework, "latent heat storage" and ice volume are equivalent not because sensible heat is small, but by construction.

Indeed, and hence ice volume and ice energy of melting are directly proportional in HadGEM2-ES. However, we are talking about model biases, and need to keep in mind how the real world operates – so it is important to state that in reality, although ice volume and ice energy of melting are not directly proportional, they are quite close to being so.

- I still find the use of « latent heat flux » for internal energy storage in sea ice confusing.

We agree this is confusing and have removed the use of the word ‘latent’ in this context in the revised document, using instead ‘energy uptake’ or, in full, ‘energy required to melt the ice out’.

- There are many small typos or poorly constructed sentences.

We have corrected all typos that could be identified, and have reworded and shortened many sentences, removing unnecessary conversational phrases or qualifying adjectives.

Summary of changes to document

In each of the responses above, an overview of the changes has been given, with particular attention to the points raised by each respective reviewer. Here we describe all changes in turn, divided by section. Page and line numbers refer to the tracked changes version of the document, which is appended below. Changes in punctuation, or simple sentence division, are not mentioned.

Title and abstract

Page 1, line 1: Title reworded and 'Arctic' added.

Page 1, line 9: 'Arctic' added to start of abstract.

Page 1, line 11: Reference to ice mass balance added.

Page 1, line 12: Sentence reworded for clarity.

Page 1, line 14: Unnecessary phrase removed.

Page 1, line 15: Sentence reworded for clarity.

Page 1, line 22: Reference to observational uncertainty finding appended to end of Abstract.

Section 1 (Introduction)

Page 1, line 26: 'most notably' removed (unnecessary qualifier)

Page 1, line 28: wording simplified

Page 1, line 29: many unnecessary qualifiers removed from this sentence.

Page 1, line 33: extraneous 'much' removed

Page 2, line 1: qualifier removed, and 'considerable' changed to 'large'

Page 2, line 3: extraneous 'clearly' removed

Page 2, line 4: 'Massonnet' corrected

Page 2, line 7: 'or internal variability' removed – it was felt that mentioning internal variability in this sentence (as opposed to the following) was technically incorrect.

Page 2, line 11: Physical reasoning for study refocussed on ice mass balance.

Page 2, line 15: 'melt the ice' changed to 'melt the ice out' (Reviewer 1 suggestion)

Page 2, line 18: 'surface flux' defined

Page 2, line 24: 'in any meaningful sense' removed (thought to be unnecessary). Summary sentence added at end motivating focus on separating causes of surface flux bias.

Page 2, line 30: Reference to HadGEM2-ES model, and 1980-1999 period, deleted (to be moved to a later point of the Introduction).

Page 2, line 35: This is the introductory paragraph to the ISF framework, and has been refocussed so as to describe the most fundamental aspects of the analysis.

Page 3, line 9: Two new paragraphs describing the ISF framework in more detail have been added; these include the points removed from the previous paragraph.

Page 3, line 23: Information about HadGEM2-ES model and 1980-1999 period re-inserted here.

Page 3, line 29: Paper structure description reworded as appropriate.

Section 2 (Model and reference datasets)

Page 4, line 4: 'Observational' changed to 'reference' (Reviewer 2 suggestion)

Page 4, line 13: Reference to Martin et al altered slightly as it was felt that previously a spurious reference to the following paragraph was implied.

Page 4, line 15: Reference to section 4 removed (seemed unnecessary)

Page 4, line 19: Model description re-ordered, so that most important aspects for the current study were stated first.

Page 4, line 35: Reasoning for using 1980-1999 period explained more accurately.

Page 5, line 3: Arctic Ocean region defined more clearly (Reviewer 1 suggestion).

Page 5, line 7: 'Observational' changed to 'reference' (R2)

Page 5, line 13: 'the reanalysis' removed (thought unnecessary)

Page 5, line 15: Sentence reworded and split for clarity.

Page 5, line 19: 'ice fraction' changed to 'ice concentration' (Reviewer 1 suggestion)

Page 5, line 21: Reference to PIOMAS corrected

Page 6, line 3 & 7: 'Barrow' changed to 'Point Barrow'

Page 6, line 6: 'Finally' removed (now spurious)

Page 6, line 9: Reference to Lindsay et al (1998) added, following from Reviewer 2 suggestion and response.

Page 6, line13: Reference to addition ice concentration datasets added (these are now quoted in section 5.4 to aid evaluation of ISF sensitivity to reference dataset).

Section 3 (Evaluating HadGEM2-ES)

Page 6, line 20: 'sea ice and surface radiation' removed from title to make it more general purpose

Page 6, line 21: 'bias' defined (previously this paragraph was at the start of section 4, but we thought it made more sense to have it here, as we talk about model biases in this section too)

Page 6, line 27: Paragraph re-ordered for clarity.

Page 6, line 37: Submarine comparison split into new sentence

Page 6, line 38: Ice thickness bias summarised and mass balance bias introduced.

Page 7, line 8: Mass balance bias described explicitly with respect to all datasets.

Page 7, line 16: 'Low bias' changed to 'thin bias' (thought less ambiguous)

Page 7, line 18: 'Even' removed (thought unnecessary)

Page 7, line 22: 'association' changed to 'correspondence' (thought a more appropriate word)

Page 7, line 24: Reference to thickness-growth feedback and Massonnet et al (2018) added.

Page 7, line 28: New paragraph added. The mass balance bias is used to motivate a surface flux focus (and hence the ensuing surface radiation evaluation). Reasons for neglecting oceanic heat convergence and ice export are presented (previously this argument was in section 4).

Page 8, line 4: Surface radiative flux evaluation introduced; downwards=positive framework clarified.

Page 8, line 7: Discussion of May downwelling SW observational errors removed (these are not important for the study's conclusions and so this sentence was probably just confusing matters)

Page 8, line 18: New paragraph added. The surface melt onset bias is evaluated, motivated by the net SW bias described in the previous paragraph. Following on from Reviewer 2's concerns, the observed Central Arctic melt onset is now described as being in 'mid-June'. The new Figure 6 is referenced.

Page 9, line 3: Model evaluation summary paragraph removed as thought inadequate; instead…

Page 9, line 8: …two new paragraphs have been added, summarising the model evaluation and describing why a more rigorous approach is helpful (i.e. properly quantifying the effect of the model biases.

Section 4 (The induced surface flux framework: a way to quantify the effect of each model bias on sea ice mass balance)

Page 9, line 29: Section title clarified

Page 9, line 32: This section has been completely rewritten. The way in which the ISF framework arises quite naturally from the considerations of the model evaluation is described; the approach is described in more detail; the advantages (linearity of the ISF biases, and quasi-independence of the variables) are set out with justification. Conversely, the details of the surface flux formula, and its derivation, have been moved to a (re-introduced) Appendix A. At the end of the section, ISF calculation examples are still given, but only the first two examples (melt onset and downwelling LW) have been retained.

Section 5 (Induced surface flux bias)

Page 16, line 5: Title simplified

Page 16, line 6: First subsection 'Aggregate ISF bias' inserted

Page 16, line 7: Unnecessary reference to previous section removed.

Page 16, line 8: Clarification added (we're evaluating these variables because they're the ones we have reference datasets for)

Page 16, line 11: Figure numbering amended; 'net radiative flux bias' clarified

Page 16, line 13: 'sea ice latent heat uptake' changed to 'sea ice energy uptake' (Reviewer 2 suggestion); 'ice thickness bias' changed to 'seasonal ice mass balance bias'

Page 16, line 15: New sentence added summarising ISF results.

Page 16, line 19: The two paragraphs describing the Arctic aggregate ISF results have been shortened, reworded, and worked into one paragraph.

Page 17, line 19: ISF implied mass balance bias summarised again.

Page 17, line 24: Large section of text on ISF residuals and observational uncertainty moved due to section restructuring.

Page 18, line 32: New spatial variability subsection introduced

Page 18, line 33: Spatial variability discussion restructured to describe each ISF bias in turn.

Page 19, line 29: Figure numbering amended

Page 19, line 31: Unnecessary qualifier removed

Page 19, line 33: 'near the Atlantic Ocean ice edge' changed to 'in the Atlantic sector' (thought a more accurate description)

Page 20, line 3: New 'forcings and feedbacks' subsection introduced.

Page 20, line 4: Forcings and feedbacks discussion, previously in section 6, introduced. This has been reworded to justify more carefully why the ice area and thickness ISF biases can be identified with the surface albedo and thickness-growth feedback effects on the ice mass balance.

Page 21, line 9: New 'residuals and observational uncertainty' subsection introduced.

Page 21, line 10: This section has also been completely rewritten. The total ISF bias is evaluated using four proxies for net surface flux bias (which cannot be exactly known). The direct effect of observational uncertainty on ISF uncertainty is evaluated, and compared to the estimated effect of theoretical errors (as evaluated in Appendix B).

Section 6 (Discussion)

Page 22, line 22: First subsection introduced (understanding sea ice state biases)

Page 22, line 23: Sea ice state analysis introduced with reference to the 'feedback' discussion in section 5.3.

Page 22, line 37: wording clarified.

Page 23, line 27: Second subsection introduced (looking beyond proximate drivers)

Page 23, line 38: Figure numbering amended

Page 24, line 22: Third subsection introduced (missing processes)

Page 24, line 24: 'missing process' discussion expanded, as there is now less discussion of this in section 5.4.

Page 25, line 10: New paragraph added, discussing the quasi-independence issue in more detail (ice and snow thickness are not truly independent of the other variables as they don't affect surface flux instantaneously)

Section 7 (Conclusions)

Page 25, line 22: This section has also been rewritten, to focus more on the method, its limitations, its applicability to other models and the impact of observational uncertainty.

Appendix A (Description and derivation of the surface flux formula used in the ISF calculation)

Page 27, line 28: This section contains much text formerly in Section 4, but expanded. The process by which the surface flux formula is derived from first principles is completely described. As requested by Reviewer 1, the process by which model category ice thickness biases is illustrated with another new figure.

Appendix B (Analysis of potential errors in ISF bias calculation)

Page 30, line 25: appendix labelling amended

Page 30, line 26: introductory paragraph rewritten to define the problem more clearly

Page 31, line 5: Subsections introduced for clarity.

Page 31, line 10: Error analysis re-ordered to be chronological (winter – May – summer)

Page 31, line 16: unnecessary qualifier removed

Page 31, line 29: error analysis re-ordered (see above)

Page 33, line 3: second subsection introduced

Page 33, line 13: new introductory paragraph, again to define the problem more clearly

References

Page 35, line 16: Reference to NSIDC ice concentration (Cavalieri et al, 1996) added

Page 36, line 16: Lindsay et al (1998) reference added

Page 37, line 4: Massonnet et al (2018) reference added

Page 38, line 24: Reference to HadISST.1 ice concentration (Titchner and Rayner, 2014) added

Figures

Figure 1: caption amended to define the Arctic Ocean more clearly

Figure 2: individual labelling provided for each reference dataset on each subplot; caption amended to clarify time periods

Figure 3: Reference dataset labelled on each subplot; caption amended to clarify time periods

Figure 4: Respective fluxes labelled on Y-axis of each subplot; caption expanded to clarify meaning of absolute and bias subplots.

Figure 5: New melt onset figure inserted.

Figure 6: This was formerly figure 5. The lower two rows have been removed.

Figure 7: This was formerly figure 6. The caption has been updated and clarified.

Figure 8: This was formerly figure 7, but is otherwise unchanged.

Figure 9: This was formerly figure 8, but is otherwise unchanged.

Figure A1: New figure demonstrating calculation of category ice thickness bias.

Figure B1: This was formerly figure A1. The contour levels on the surface flux bias figures (right-hand column) have been corrected.

A tracked changes version of the paper, to which the above page and line numbers refer, follows beneath.

Induced surface fluxes: A new framework for attributing Arctic sea ice mass balance biases to specific model errors

~~Attribution of sea ice model biases to specific model errors enabled by new induced surface flux framework~~

Alex West[1], Mat Collins[2], Ed Blockley[1], Jeff Ridley[1], Alejandro Bodas-Salcedo[1]

[1]Met Office Hadley Centre, FitzRoy Road, Exeter, EX1 3PB
[2]Centre for Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, EX4 4SB, UK
Correspondence to: Alex E. West (alex.west@metoffice.gov.uk)

Abstract. A new framework is presented for analysing the proximate causes of model Arctic sea ice biases, demonstrated with the CMIP5 model HadGEM2-ES. In this framework the sea ice volume is treated as a consequence of the integrated surface energy balance, via the mass balance. A system of simple models allows the local dependence of the surface flux on specific model variables to be described, as a function of time and space~~, on specific model variables (ice area, ice thickness, surface melt onset and downwelling longwave and shortwave radiation) to be described~~. When these are combined with reference datasets ~~of the variable in question~~, it is possible to estimate the surface flux bias induced by the model bias in each ~~that~~ variable~~.~~. The method allows ~~quantification of~~ the role played by the surface albedo and ice thickness-growth feedbacks in sea ice mass balance biases to be quantified, along with the roles of model bias in variables external to the sea ice state.~~causing anomalous sea ice melt and growth to the role played by other forcings which can be viewed as external to the sea ice state on short timescales~~. It shows biases in the HadGEM2-ES sea ice volume simulation to be due to a bias in spring surface melt onset date, partly countered by a bias in winter downwelling longwave radiation. The framework is applicable in principle to any model and has the potential to greatly improve understanding of the reasons for ensemble spread in modelled sea ice state. A secondary finding is that observational uncertainty is the largest cause of uncertainty in the induced surface flux bias calculation.

1. Introduction

The Arctic sea ice cover has witnessed rapid change during the past 30 years, ~~most notably~~ with a decline in September extent of 1.05 x 10^6 km^2 / decade from 1986 to 2015 (HadISST1.2, Rayner et al 2003). In association with the changes in extent, ~~evidence of declining~~ Arctic sea ice thinning ~~thickness loss~~ has been observed from submarine and satellite data (Rothrock et al, 2008, Lindsay and Schweiger, 2015). Arctic sea ice ~~is~~ has also ~~also thought to have~~ become younger on average as ~~reserves of~~ older ice has~~ve~~ been lost (Maslanik et al, 2011), the onset of summer melt has ~~been observed to~~ become earlier in the year (Markus et al, 2009) and the onset of winter freezing has ~~been observed to~~ become later (Stammerjohn et al, 2012).

The changes have focussed ~~much~~ interest on model projections of Arctic sea ice, the loss of which influences the climate directly through increased absorption of shortwave (SW) radiation during summer and through greater release of heat from the ocean to the atmosphere during winter (Stroeve et al, 2012b). However, substantial spread remains in model simulations of present-day Arctic sea ice~~,~~ and of the long-term rate of

decline under climate change (Stroeve et al, 2012a). The causes of this spread are ~~at present~~ poorly understood, resulting in large~~considerable~~ uncertainty in future projections of Arctic sea ice.

Evaluating sea ice extent or volume with respect to reference datasets shows that some models ~~clearly~~ reproduce present-day sea ice state more accurately than others (e.g. Wang and Overland, 2012; Massonnet et al, 2012; Shu et al, 2015). However, an accurate simulation of sea ice extent and volume under the present-day climate does not necessarily imply an accurate future projection of sea ice change, as a correct simulation can be obtained by accident due to cancelling model errors~~, or internal variability~~. Sea ice extent in particular is known to be a very unsuitable metric for diagnosing model performance due to its high internal variability (Notz, 2015; Swart et al, 2015). Hence there is a need to better understand the drivers which lead a model to simulate a given Arctic sea ice state.

This study presents a new framework (the induced surface flux, or ISF, framework) to improve understanding of sea ice model bias, by identifying proximate drivers of model bias in sea ice mass balance. The framework is motivated in the following way. Changes in sea ice volume are driven by the sea ice mass balance. ~~Ice volume arises from ice mass balance~~. ~~I~~Sea ice mass balance in turn arises from the surface and basal energy balance, as the ice mass is~~,~~ to first order~~,~~ proportional to the heat required to melt the ice out, and therefore acts to integrate the surface and basal energy balance. Basal melting in the interior ice pack has been shown to derive in the main from direct solar heating of the ocean (e.g Maykut and McPhee, 1995), while basal freezing derives principally from conduction of energy upward through the ice (Perovich and Elder, 2002).~~;~~ T~~t~~his implies that the total downwards surface energy flux (surface flux) ~~balance (SEB)~~ contains the principal sources and sinks of energy for the sea ice on an Arctic-wide scale. However, a complex two-way relationship exists between sea ice thickness and surface energy balance, via the surface temperature and surface albedo, giving rise to the thickness-growth feedback (Bitz and Roe, 2004) and the surface albedo feedback (Bitz, 2008), both of which exert first-order control on the sea ice state. Hence many components of the SEB cannot be viewed as independent of the sea ice state ~~in any meaningful sense~~, but are directly affected by it. Therefore, there is a need to separate surface flux biases that are caused by the sea ice state (representing feedbacks of the sea ice mass balance) from those that are not (representing forcings on the sea ice mass balance).~~.~~

~~This study, which presents a new framework to investigate the causes of modelled sea ice biases, is motivated by a desire to separate, to first order, many external drivers of the SEB (and hence the sea ice state) from the thickness-growth and albedo feedbacks, and thereby better understand the processes that result in a particular modelled sea ice state being simulated. The analysis uses as a case study the four members of the historical ensemble of the coupled CMIP5 model HadGEM2-ES, a member of the CMIP5 historical ensemblewhich simulates anomalously low annual minimum ice extent, and which simulates an ice volume which is both too low in the annual mean and too amplified in the seasonal cycle, a similar behaviour to that identified by Shu et al (2015) in the CMIP5 ensemble mean.~~

In the ISF framework ~~presented~~, the total downwards surface energy flux is expressed as an explicit function ~~in terms~~ of key Arctic climate variables, allowing ~~as functions of space and time using two simple models, for the freezing and melting seasons respectively, which are shown to capture well the large-scale spatial and seasonal variation of the surface flux. With the use of the simple models,~~ the ~~local~~ dependence of modelled surface flux

on key variables can be described. Hence, using observed or reference datasets for the climate variables, the model bias in surface flux induced by each climate variable can be estimated. In this way, biases in ice growth and melt over the course of the year are attributed via the surface flux to biases in specific model quantities. The method allows the contributions to model biases in ice growth and melt caused by the sea ice albedo feedback, the ice thickness-growth feedback, and various external factors, or 'forcings', to be separately quantified. In this way it can be seen how model biases in the external forcings ~~drivers~~ drive model bias in the sea ice mass balance~~are able to produce a particular sea ice state~~, offering a valuable tool for setting sea ice state biases in context, and for understanding spread in sea ice simulation within multi-model ensembles.

In more detail, the ISF framework works by expressing the total downward surface energy flux F_{sfc}, at each point in time t and space x, as an explicit function $g_{x,t}$ of quasi-independent climate variables v_i. The variables are quasi-independent in the sense that while they affect each other on timescales varying from days to months, they affect the surface flux instantaneously. Hence by taking partial derivatives, the dependence of the surface flux on each variable can be separately expressed at each point in time and space ($\left[\partial g_{x,t}/\partial v_i\right]^{MODEL}$). Given an estimate of the model bias in any variable via a reference dataset $\left(v_{i,x,t}^{MODEL} - v_{i,x,t}^{REFERENCE}\right)$, the field of surface flux dependence can be multiplied through to produce an estimate of the surface flux bias induced (instantaneously) by the model bias in that variable, as a function of space and time.

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

This method has two key strengths; firstly, that the fields of induced surface flux bias (ISF bias) can be averaged in time or space to determine the large-scale effects of particular model biases, bypassing nonlinearities in surface flux dependence. Secondly, due to the quasi-independence of the variables, the effects of each on the surface flux are separated, such that the sum of the ISF biases theoretically approaches the total, true, model surface flux bias. In this way, the instantaneous, or proximate, causes of the model surface flux bias, and hence sea ice mass balance bias, can be separated and quantified.

The analysis is applied to the four members of the historical ensemble of the coupled CMIP5 model HadGEM2-ES. This model simulates anomalously low annual minimum ice extent, as well as ice volume that~~which~~ is both too low in the annual mean and too amplified in the seasonal cycle, a similar behaviour to that identified by Shu et al (2015) in the CMIP5 ensemble mean. A variety of reference datasets are used to assess model biases, to demonstrate the large observational uncertainties present in the Arctic, and their effect on our ability to attribute sea ice mass balance bias with the ISF framework.

The paper is structured as follows. In Section 2, the HadGEM2-ES model and the reference~~observational~~ datasets used are described in turn. In Section 3 the sea ice and surface radiation simulations of HadGEM2-ES are evaluated. In Section 4, the ISF framework is described in more detail, and examples shown.~~induced surface flux method is introduced~~, and in Section 5 the ISF~~induced surface flux~~ analysis is applied to HadGEM2-ES, allowing ~~quantification of~~ the role of each model bias identified~~assessed~~ in Section 3 in causing sea ice mass balance bias to be quantified.~~played by biases in specific Arctic climate variables in causing anomalous ice growth and melt.~~ In Section 6 the implications of the results are discussed, in particular the mechanisms by

which the identified external drivers determine the modelled sea ice state, and the likely drivers behind the corresponding model biases. Conclusions are presented in Section 7.

2. Model and reference datasets~~observational data~~

2.1 The HadGEM2-ES model

HadGEM2-ES is a coupled climate model employing additional components to simulate terrestrial and oceanic ecosystems, and tropospheric chemistry (Collins et al, 2011). It is part of the 'HadGEM2' family, a collection of models that all use the HadGEM2-AO coupled atmosphere-ocean system. HadGEM2-AO is developed from HadGEM1 (Johns et al, 2006), a coupled atmosphere-ocean model whose sea ice extent simulation was recognised as being among the closest to observations out of the CMIP3 ensemble models (Wang and Overland, 2009). While the atmospheric and ocean components of HadGEM2-ES contain a large number of improvements relative to HadGEM1, many of these targeted at improving simulations of tropical weather, the sea ice component is very similar to that of HadGEM1 (except for three minor differences, summarised in ~~modifications (~~Martin et al, 2011, table A4).

A fundamental feature of the sea ice component of HadGEM2-ES is the~~, which is important for the analysis described in Section 4 below, is that it includes a~~ sub-gridscale sea ice thickness distribution (Thorndike et al, 1975). In this formulation, ice in each grid cell is separated into five thickness categories with boundaries at 0, 0.6m, 1.4m, 2.4m, 3.6m and 20m, each with its own area, thermodynamics and surface exchange calculations. Another key aspect of the sea ice model is the zero-layer thermodynamics scheme (appendix to Semtner, 1976), in which the sea ice surface temperature responds instantaneously to changes in forcing, conduction is uniform within the ice and snow column, and neither sea ice nor overlying snow have heat capacity (sensible heat storage is parameterised in the top 10cm of the snow-ice column during surface exchange calculations, to aid stability).

The HadGEM2-ES sea ice model also includes elastic-viscous-plastic sea ice dynamics (Hunke and Dukowicz, 1997) and incremental remapping (Lipscomb and Hunke, 2004). ~~It also includes elastic-viscous-plastic sea ice dynamics (Hunke and Dukowicz, 1997) and incremental remapping (Lipscomb and Hunke, 2004). The thermodynamic component is a zero-layer model, with no heat capacity, described in the appendix of Semtner (1976). The insulating effect of snow is modelled by means of a single layer with conductivity 0.33Wm-1K-1, also with no heat capacity (although sensible heat storage is parameterised in the top 10cm of the snow-ice column during surface exchange calculations, to aid stability).~~ Most processes are calculated in the ocean model, but the surface energy balance (SEB) calculations are carried out in the atmosphere model, which passes top melting flux and conductive heat flux to the ocean model as forcing for the remaining components. A more complete description of the sea ice component can be found in McLaren et al (2006).

This study uses the four ensemble members of the CMIP5 historical experiment of HadGEM2-ES, forced with observed solar, volcanic and anthropogenic forcing from 1860 to 2005. The period 1980-1999 is used~~chosen~~ for the model evaluation, chosen to be close to the end of the period of the historical experiments so as to be ~~as a period which predates much of the recent rapid Arctic sea ice loss and is hence at least partially independent of~~

~~the period normally used to evaluate sea ice trends. It has the added advantage of being~~ recent enough to allow the use of a reasonable range of observational data. All analysis is carried out with data restricted to the Arctic Ocean region, defined as the area enclosed by the Fram Strait, the northern boundary of the Barents Sea, the western boundary of the Kara Sea, the Bering Strait and the northern edge of the Canadian Arctic Archipelago,, shown in Figure 1.

~~2.2~~ Reference datasets~~Observational data~~

Uncertainty in observed variables tends to be higher in the Arctic than in many other parts of the world. There are severe practical difficulties with collecting in situ data on a large scale over regions of ice-covered ocean. While satellites have in many cases been able to produce Arctic-wide measurements of some characteristics, most notably sea ice concentration, the relative lack of in situ observations against which these can be calibrated means knowledge of the observational biases is limited. Reanalysis data over the Arctic is also more subject to ~~the reanalysis~~ model errors than in other regions, due to errors in atmospheric forcing, and the existence of fewer direct observations available for assimilation (Lindsay et al, 2014). The approach of this study is to use a wide range of observational data to evaluate modelled sea ice state and surface radiative fluxes, setting results in the context of in situ validation studies. The same datasets are then used~~, and to use~~ as reference datasets for the induced surface flux framework.~~, using the small number of in situ validation studies to set results in context as far as possible.~~

To evaluate modelled sea ice concentration~~fraction~~, we use the HadISST1.2 dataset (Rayner et al, 2003), derived from passive microwave observations. To evaluate modelled sea ice thickness Arctic-wide, we use the ice-ocean model PIOMAS (Zhang and Rothrock, 2003~~Schweiger et al, 2011~~), which is forced with the NCEP reanalysis and assimilates ice concentration data. Laxon et al (2013) and Wang et al (2016) found PIOMAS to estimate anomalously low winter ice thicknesses compared to satellite observations in some years. In particular, Wang et al (2016) found PIOMAS to have a mean bias of -0.31m relative to observations from the ICESat (Ice, Cloud and land Elevation Satellite) laser sensor. To set the PIOMAS comparison in context, we use two additional datasets to evaluate the model over smaller regions; measurements from radar altimetry aboard the ERS satellites from 1993-2000 (Laxon et al, 2003), limited to latitudes below 82°N; and estimates compiled by Rothrock et al (2008), derived from a multiple regression of submarine transects over the Central Arctic Ocean from 1975-2000, constrained to be seasonally symmetric.

To evaluate modelled surface radiative fluxes across the whole Arctic Ocean, three datasets are used. Firstly, we use the CERES-EBAF (Clouds and Earth's Radiant Energy Systems – Energy Balanced And Filled) Ed2.7 dataset (Loeb et al, 2009), based on direct measurements of top-of-atmosphere radiances from EOS sensors aboard NASA satellites, available from 2000 – present. Secondly, we use the ISCCP-FD (International Satellite Cloud Climatology Project FD-series) product (Zhang et al, 2004). Lastly, we use the ERA-Interim (ERAI) atmospheric reanalysis dataset, which provides gridded surface flux data from 1979-present using a reanalysis system driven by the ECMWF (European Centre for Medium-range Weather Forecasts) IFS forecast model and the 4D-Var data assimilation system (Dee et al, 2011).

Formatted: Font: (Default) Times New Roman, 10 pt

In-situ validation of these datasets in the Arctic has been limited, but Christensen et al (2016) found CERES to perform quite well relative to other products, albeit underestimating downwelling LW fluxes from November – February by 10-20 Wm^{-2} relative to in situ observations at Point Barrow (Alaska). Liu et al (2005) found ISCCP-FD to simulate SW radiative fluxes fairly accurately relative to observations from SHEBA, but to underestimate downwelling SW fluxes in spring by over 30 Wm^{-2}, also overestimating downwelling LW fluxes in winter by around 40 Wm^{-2} . ~~Finally,~~ Lindsay et al (2014) identified ERAI as producing a relatively accurate simulation of surface fluxes compared to in situ observations at Point Barrow (Alaska) and Ny-Ålesund (Svalbard), although tending to underestimate downwelling SW fluxes in the spring by up to 20 Wm^{-2} and overestimate downwelling LW fluxes in the winter by around 15 Wm^{-2}. Comparison of winter downwelling LW fluxes in all datasets to in situ measurements compiled by Lindsay et al (1998) suggests that while ISCCP-FD is likely to be biased high, ERAI and CERES may be relatively accurate.

In addition to the datasets above, in section 4 we make use of satellite estimates of date of melt onset over sea ice (Anderson et al, 2012), also derived from passive microwave sensors.~~;~~ In section 5, to evaluate the impact of observational uncertainty in ice area, we use the NSIDC 'Sea Ice Concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data, Version 1' (Cavalieri et al, 1996) and HadISST.2 (Titchner and Rayner, 2014). Finally, in section 6, ~~and in section 5,~~ the CERES-SYN dataset (Rutan et al, 2015), similar to CERES-EBAF but available at higher temporal resolution, is used to examine modelled surface radiation evolution during May in more detail.

3. Evaluating ~~sea ice and surface radiation in~~ HadGEM2-ES

In this section, and throughout the rest of the paper, a difference between a model simulation of a particular quantity~~variable~~, and any reference dataset for that quantity~~variable~~, is referred to as a 'bias'. In a similar way, the difference in model surface flux judged to arise from the difference in a particular quantity~~variable~~ relative to a reference dataset is referred to as an 'induced surface flux bias'. Attention is drawn to the fact that, due to observational inaccuracy, true model bias relative to the real world may be somewhat different from the biases described in this way.

Modelled September sea ice extent in HadGEM2-ES ~~F~~from 1980 to ~~-~~1999 is systematically lower than that observed (Figure 2a);~~,~~ the four members of the HadGEM2-ES historically-forced ensemble simulate a mean September sea ice extent of 5.78 x 10^6 km^2, with ensemble standard deviation of 0.24 x 10^6 km^2. By comparison, the mean observed September sea ice extent over this period was 6.88 x 10^6 km^2 according to the HadISST1.2 dataset. ~~Over the reference period, therefore, modelled September sea ice extent is systematically lower than that observed (Figure 2a).~~

Mean ice thickness is consistently lower than that estimated by PIOMAS for the Arctic Ocean region (Figure 2b), with the highest biases of -0.4m occurring in October, close to the minimum of the annual cycle, and a near-zero bias in May, close to the maximum. Modelled ice thickness is also biased low relative to the ERS satellite measurements (Figure 2c), with thickness biases ranging from -0.57m in November to -0.16m in April. Finally, modelled ice thickness is biased low~~, and~~ relative to the submarine data (Figure 2d), with thickness biases ranging from -1.5m in August to -0.8m in January and May. Hence there is clear evidence of a thin~~low~~

model bias in annual mean ice thickness, but there is also evidence of a model bias in the seasonal cycle of mass balance, as model biases at minimum ice thickness tend to be larger (i.e. more negative) than those at maximum ice thickness, implying excess ice melting and ice freezing in the model.

~~Hence it is very likely that ice thickness in HadGEM2-ES is biased low in the annual mean, with biases tending to be higher when ice thickness is lower. In other words, the ice thickness annual cycle of HadGEM2-ES is likely to be too amplified, with both anomalously high ice melt during the summer and ice growth during the winter.~~

Relative to PIOMAS, a mass balance bias of 38cm excess ice melt in summer and ice growth in winter is implied; relative to the ERS satellite measurements, a mass balance bias of 42cm. As the seasonal cycle of submarine ice thickness is constrained to be symmetric, a more useful measure of the mass balance bias here can be gained by comparing the maxima and minima of the submarine ice thickness seasonal cycle to that of HadGEM2-ES; in this case, a mass balance bias of 38cm excess melt in summer, and excess growth in winter, is implied. Hence HadGEM2-ES is likely to overestimate the magnitude of the ice thickness seasonal cycle by around 40cm across the Arctic Ocean on average. In particular, the large negative model bias in ice thickness in late summer is likely to be the principal cause of the low bias in September ice area.

Maps of the ice thickness bias in April and October (Figure 2b-d) show agreement that the thin~~low~~ ice thickness bias is smaller on the Pacific side of the Arctic than on the Atlantic side of the Arctic, becoming very small or ~~even~~ positive in the Beaufort Sea. There is also striking agreement in the spatial pattern of the amplification bias of the seasonal cycle, as diagnosed by April-October ice thickness difference (Figure 3). All three ice thickness datasets show the HadGEM2-ES ice thickness seasonal cycle to be too amplified across much of the Arctic, by up to 1m in the Siberian shelf seas; in addition, all show that in the Beaufort Sea, the amplification is nonexistent or even negative. There is clear correspondence~~association~~ between areas where modelled annual mean ice thickness is biased low, and areas where the modelled seasonal cycle is over-amplified, and vice versa. This is likely to be associated with the ice thickness-growth feedback, whereby a steeper temperature gradient induces stronger conduction and hence ice growth for thin ice. Indeed, negative correlations between summer sea ice and sea ice growth the following winter are a ubiquitous feature of CMIP5 models (Massonnet et al, 2018).

The bias in ice mass balance is associated with a bias in ice energy uptake. For example, the 40cm ice melt bias during summer would be associated with an energy uptake bias of around 1.5 x 10^8J, or 15Wm^{-2} over a 4-month melting season; the 40cm ice growth bias during winter would be associated with an energy uptake bias of -7.5Wm-2 over an 8-month freezing season. The ice energy uptake has three drivers: the surface energy balance, the oceanic heat convergence, and ice divergence. Sea ~~However,~~ ice divergence is generally recognised to be a small term (e.g. Serreze et al, 2007). Although Arctic Ocean heat convergence can be significant in size, across much of the Arctic the sea ice is insulated from the main source of heat energy from beneath, the warm Atlantic water layer, by fresh water derived mainly from river runoff (e.g. Serreze et al, 2006; Stroeve et al, 2012b). Because of this, in the Arctic Ocean interior direct solar heating of the ocean is a much larger contributor to sea ice basal melting than oceanic heat convergence, as observed by Maykut and McPhee, 1995, McPhee et al, 2003 and Perovich et al, 2008, and modelled by Steele et al (2010) and Bitz et al (2005). In particular, it has been

found that in HadGEM2-ES oceanic heat convergence is of negligible importance to the sea ice heat budget (Keen et al, 2018). Hence for the main purposes of this study, we concentrate on the surface energy balance, and neglect the other two terms (although the ocean heat convergence ~~y~~ is~~are~~ briefly discussed in Section 6).

Surface radiative fluxes are now evaluated. In the following discussion of radiative fluxes, and throughout this study, the convention is that positive numbers denote a downwards flux. Fluxes of downwelling SW radiation are higher in HadGEM2-ES than in all observational estimates during the spring (Figure 4a-c), with May biases of 22, 43 and 53 Wm^{-2} relative to CERES, ERAI, and ISCCP-FD respectively. ~~We note that as ERAI and ISCCP-FD have been found to underestimate downwelling SW during spring at specific locations, the true model bias is perhaps more likely to lie towards the lower end of these estimates.~~ During the summer, upwelling SW radiation is consistently lower in magnitude than in HadGEM2-ES, with June biases of 16, 37 and 44 Wm^{-2} with respect to ERAI, CERES and ISCCP-FD respectively (a positive bias in an upward flux demonstrates that the model is too low in magnitude). There is no consistent signal for a low bias in downwelling SW during the summer, suggesting a model surface albedo bias. The effect is that modelled net downward SW flux is too large with respect to all observational datasets in May and June, and with respect to some in July and August. Relative to CERES, the May downwelling SW bias displays no clear spatial differentiation over the Arctic Ocean (Figure 4a), but the June upwelling SW bias, and hence the net SW bias, tend to be somewhat higher in magnitude towards the central Arctic (Figure 4b-c).

As the June net SW bias is likely to result from a surface albedo bias we briefly digress from the radiation evaluation to discuss the parameters affecting surface albedo over sea ice in HadGEM2-ES: ice fraction, snow thickness and surface melt onset. Ice fraction has already been evaluated and for snow thickness no reference dataset is available; however, surface melt onset can be evaluated using satellite observations (Figure 6). We define the date of melt onset for any grid cell as the first day on which the surface temperature exceeds -1°C (varying this threshold by 0.5°C in either direction changes the date in only a small minority of grid cells). The average date of melt onset as estimated by this method (Figure 5a) is then compared to that measured by the satellite-derived dataset described in Section 3 (Figure 5b), with model bias shown in Figure 5c. Large spatial variability is evident in the observations. Melt onset occurs in early to mid-May around the Arctic Ocean coasts, but much later in the Central Arctic, around mid-June. In contrast, the HadGEM2-ES surface melt onset date is in mid- to late May across the Arctic Ocean, without the strong gradients seen in the observations. This would cause a surface albedo, and hence net SW, bias with a strong maximum in the Central Arctic, similar to that discussed above.

We now return to the radiation evaluation to evaluate LW fluxes. Fluxes of longwave (LW) radiation are lower in magnitude in HadGEM2-ES throughout the winter than in all observational datasets (Figure 4d-f). For downwelling LW, the mean model biases from December-April are -16, -22 and -40 Wm^{-2} for ERAI, CERES and ISCCP-FD respectively; for upwelling LW, the biases are 11, 16 and 18 Wm^{-2} for CERES, ERAI and ISCCP respectively. Because the downwelling LW biases vary more than the upwelling LW biases, there is uncertainty in inferring a model bias in net downwelling LW; ISCCP suggests a large model bias of -22 Wm^{-2}, CERES a smaller bias of -11 Wm^{-2}, while ERAI suggests a biasof only 1 Wm^{-2}. As in situ studies have shown both underestimation (by CERES) and overestimation (by ERAI and ISCCP-FD) of downwelling LW in winter, there is no clear indication as to where the true model bias in this quantity may lie. Maps of the downwelling

and net down LW bias relative to CERES in February (Figure 4d,f) show the bias tends to be somewhat higher towards the North American side of the Arctic, and lower on the Siberian side.

~~In summary, there is evidence of a low bias in net downward LW during the winter, and a high bias in net downward SW during the summer, each of order of magnitude ~10 Wm^{-2}. This is consistent with surface radiation fluxes being the likely first-order cause of the amplified sea ice thickness seasonal cycle. In the next section we describe the process by which surface radiation biases can be attributed to particular model processes by calculating induced surface flux biases.~~

The surface radiation evaluation provides clues as to the causes of the HadGEM2-ES ice mass balance bias, but also underlines why a more detailed analysis is required to properly quantify these causes. For example, i~~In the~~ winter~~,~~ the low downwelling LW bias provides a clear mechanism for the bias in ice freezing. However, ~~, but~~ the reference datasets also suggest a ~~counteracting~~ low bias in upwelling LW that at first sight would tend to counteract this. In fact, these biases are fully consistent: a low bias in downwelling LW would be expected to cause a low surface temperature bias, causing both a high bias in ice growth and a low bias in upwelling LW. The qualitative evaluation fails to capture the full causal relationship; for this, an analysis of exactly how the downwelling LW bias affects surface flux, including the upwelling LW response, is necessary. ~~In addition, the low bias in ice thickness present in early winter would be expected to cause a high bias in ice growth: thinner ice supports a steeper temperature gradient, higher conduction at the ice base, a more negative basal energy balance and hence stronger ice growth. This is also manifested in the surface flux via a warmer surface temperature and stronger upwelling LW.~~

In the summer, meanwhile, the surface radiation evaluation suggests that a bias in net SW radiation is responsible for the ice mass balance bias, and that this in turn is related to a surface albedo bias. However, at least two possible drivers of this have been identified: the surface melt onset bias, and the underlying ice area bias that is itself likely to be caused by the ice mass balance bias. Quantifying the extent to which each driver is important, and at which times of year, is likely to help in resolving this circular causal loop. In the next section, it is described how the effects of each model bias on the sea ice mass balance can be separated and quantified through their effect on the surface flux, in the ISF framework.

~~4. Calculating induced surface flux bias: Methods~~ The induced surface flux (ISF) framework: a way to quantify the effect of each model bias on sea ice mass balance

~~1.~~

Formatted: Outline numbered + Level: 1 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0.63 cm + Indent at: 1.27 cm

We are motivated by the observation that each of the model biases described above affects the sea ice mass balance by acting through the total downwards surface energy flux (referred to as the surface flux). An excess of downwelling radiation leads directly to a higher surface flux, higher sea ice energy uptake~~-~~ and a bias towards ice melting. A bias in ice area, or in surface melt onset, is associated with a bias in surface albedo, hence a bias in net SW, and in the total surface flux. Finally, a bias in ice thickness alters the thermodynamics of the entire

snow-ice column, and of the near-surface atmosphere layer. Flux continuity considerations show that in the freezing season, thinner ice is associated with a warmer surface temperature, and hence a higher upwelling LW flux and a smaller total downwards surface energy flux. Thinner (thicker) ice is also associated with a stronger (weaker) upwards conduction flux, and hence stronger (weaker) ice growth.

Each of these relationships can be quantified, in principle, at any point in model space and time. Specifically, the rate at which the surface flux depends on each variable alone, with others being held constant, can be estimated. To this end, we approximate the surface flux F_{sfc} ~~FSFC~~ at each point in model space x and time t by an explicit function $g_{x,t}$ of quasi-independent variables v_i. The functions $g_{x,t}$ are constructed in such a way as to capture each of the relationships described above in a manner that best represents both HadGEM2-ES, and also the conditions at the point x and time t. In addition, the function captures the indirect effect of any model bias on surface flux via surface temperature and upwelling LW, which will tend to counteract the direct effect to a degree. Hence the dependence of the surface flux on each of the independent variables at point x, time t can be approximated by $\left[\partial g_{x,t}/\partial v_i\right]^{MODEL}$. Given a model bias in variable v_i at (x,t) we can then estimate the surface flux bias induced by that model bias as $\left[\partial g_{x,t}/\partial v_i\right]^{MODEL} \partial g_{x,t}/\partial v_i \left(v_{i,x,t}^{MODEL} - v_{i,x,t}^{REFERENCE}\right)$.

The function $g_{x,t}$ and its derivation are described fully in Appendix A, but are summarised briefly here. The surface flux is expressed as a sum of separate radiative and turbulent components. Ice-covered and ice-free portions of a grid cell are treated separately; in ice-covered areas, dependence on surface temperature is linearised, and flux continuity at the surface and a uniform vertical conductive flux through the ice are assumed, allowing surface temperature to be eliminated, and the dependence of surface flux on upwelling LW to be captured. The surface albedo, upon which the upwelling SW component depends, is expressed in terms of ice fraction, snow fraction and melt onset occurrence, based on the albedo parameterisation of HadGEM2-ES. Latent heat flux over ice is neglected. In this way the surface flux is expressed as an explicit function of downwelling shortwave (SW) and longwave (LW) radiation, ice concentration, category ice thickness, snow thickness, sensible and latent heat fluxes over ice and ocean, and melt onset occurrence (a logical determining whether or not the snow surface is undergoing melting). Induced surface flux due to ice thickness bias is determined by summing each of the separate ISF biases by category.

The usefulness of this approach is that surface flux operates linearly on the sea ice mass balance, meaning that each of the ISF biases at (x,t) can be averaged over large regions of time and space to understand large-scale sea ice biases. Clearly, none of the driving model biases operate on the sea ice state in a linear sense. For example, given identical surface melt onset (and hence surface albedo) biases at different points in the Arctic, each could have very different implications for local sea ice mass balance, depending on the downwelling SW modelled at each point. Conversely, identical downwelling SW biases would have different implications for sea ice mass balance depending on the modelled surface albedo. Model bias in ice thickness behaves in a particularly nonlinear fashion, with bias in regions of thinner ice having far more influence than that in regions of thicker ice. The effect of estimating ISF bias at each point separately, and then averaging to determine large-scale effects, is to bypass all nonlinearities.

Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed

A second advantage of this approach lies in the quasi-independence of the variables: while each variable may affect the others over timescales varying from days to months, each affects the surface flux instantaneously (in HadGEM2-ES). Hence a model bias in any variable represents an effect on the surface flux that is separable from the effect of a model bias in any other. If the surface flux variation is completely described by the function $g_{x,t}$, therefore, the sum of the ISF biases, over all variables, must approach the true model surface flux bias (although it will be seen that this claim is impossible to evaluate precise~~direct~~ly due to observational uncertainty). In this way, large-scale model biases in surface flux, and hence sea ice mass balance, can be broken down into separate contributions from model biases in each of the independent variables.

~~In this section, and throughout the rest of the paper, a difference between a model simulation of a particular variable, and any reference dataset for that variable, is referred to as a 'bias'. In a similar way, the difference in model surface flux judged to arise from the difference in a particular variable relative to a reference dataset is referred to as an 'induced surface flux bias'. Attention is drawn to the fact that, d~~Due to observational ~~inaccuracy, true model bias relative to the real world may be somewhat different from the biases described in this way.~~

~~Because~~Due to the latent heat of sea ice is~~being an order of magnitude greater than the sensible heat required to raise the ice to the melting temperature, ice volume is very nearly proportional to the heat required to melt the ice out. Ice volume therefore acts to integrate the surface and basal energy balance, and is largely determined by the fluxes at these interfaces. Across much of the Arctic the sea ice is insulated from the main source of heat energy from beneath, the warm Atlantic water layer, by fresh water derived mainly from river runoff (e.g. Serreze et al, 2006; Stroeve et al, 2012b). Because of this, in the Arctic Ocean interior direct solar heating of the ocean is likely to be an order of magnitude higher in accounting for basal melting of the sea ice, as observed by Maykut and McPhee, 1995, McPhee et al, 2003 and Perovich et al, 2008, and modelled by Steele et al (2010) and Bitz et al (2005). In particular, it has been found that in HadGEM2-ES oceanic heat convergence is of negligible importance to the sea ice heat budget (Keen et al, 2018). Hence the surface energy balance in the Arctic Ocean is of primary importance in controlling the evolution of sea ice volume.~~

~~We use a system of well-understood simple models, similar to those used by~~in Thorndike (1992), ~~to estimate, for each model grid cell, and month within the period, the rate at which the surface flux would be expected to change with a particular model variable. For each model grid cell and month, we construct a function $F_{sfc} \approx g_{x,t}(v_1,...,v_n)$, F_{sfc} being net surface flux, where the v_i are climate variables that affect the surface flux on timescales shorter than that on which they affect each other, and can therefore be said to be independent for the purposes of this analysis. In this way, at each model grid cell and month the rate at which the surface flux depends on variable v_i can be approximated by $\partial g_{x,t}/\partial v_i$, labelled hereafter as $\left[\partial g_{x,t}/\partial v_i\right]^{MODEL}$. Given a reference dataset for variable v_i, it then becomes possible to estimate, for each point in time and space, the surface flux bias induced by the bias in v_i as $\left[\partial g_{x,t}/\partial v_i\right]^{MODEL} \partial g_{x,t}/\partial v_i \left(v_{i,x,t}^{MODEL} - v_{i,x,t}^{REFERENCE}\right)$. The chief~~ An advantage of this method is that the resulting fields of induced surface flux bias can then be averaged in time

Field Code Changed

Field Code Changed

or space to determine the large-scale effects of particular model biases, effectively bypassing nonlinearities in surface flux dependence.

The functions $g_{x,t}$ are constructed as follows. Firstly, a model grid cell in a particular month is classified as freezing or melting depending upon whether the monthly mean surface temperature is greater or lower than -2°C. If the grid cell is classified as freezing, the surface flux is approximated as

$$F_{sfc} \approx g^{w}_{x,t} = a_{ice}\left(F_{atmos-ice} + BT_{ocn}\right)\sum_{cat}\gamma^{cat}_{ice-REF}\left(1 - BR^{cat}_{ice}\right)^{-1} + \left(1 - a_{ice}\right)F_{atmos-ocean} \quad (1)$$

Here

$$F_{atmos-ice} = F_{LW\downarrow} - \varepsilon_{ice}\sigma T^{4}_{sfc-REF} + F_{sens-ice} + \left(1 - \alpha_{ice}\right)F_{SW\downarrow}$$

, where a_{ice} is ice concentrationarea, $F_{LW\downarrow}$, $F_{SW\downarrow}$ and $F_{sens-ice}$ downwelling longwave, shortwave and sensible heat flux respectively, the latter evaluated over only the ice-covered portion of the grid cell, ε_{ice} ice emissivity, $\sigma = 5.67\times10^{-8} Wm^{-2}K^{-4}$ the Stefan-Boltzmann constant, $T_{sfc-REF}$ is monthly mean surface temperature and α_{ice} is mean surface albedo over ice.

$F_{atmos-ocean} = F_{LW\downarrow} - \varepsilon_{ocn}\sigma T^{4}_{ocn} + F_{sens-ocn} + F_{lat-ocn} + \left(1 - \alpha_{ocn}\right)F_{SW\downarrow}$, where ε_{ocn} is ocean surface emissivity, T_{ocn} ocean surface temperature (assumed to be -1.8°C), and $F_{sens-ocn}$ and $F_{lat-ocn}$ sensible and latent heat flux respectively over the ice-free portions of the grid cell.

$B = 4\varepsilon_{ice}\sigma T^{3}_{sfc-REF}$ approximates the local rate of dependence of surface flux on surface temperature, T_b is ice base temperature, $\gamma^{cat}_{ice-REF}$ is the area of ice in ice thickness category *cat* as a fraction of total ice area, where *cat* ranges from 1 to 5, and $R^{cat}_{ice} = \dfrac{h^{cat}_{ice}}{k_I} + \dfrac{h_{snow}}{k_s}$ is the thermal insulance of the snow-ice column in category *cat*, where *cat* ranges from 1 to 5, k_I and k_s being ice and snow conductivity respectively, h^{cat}_{ice} and h_{snow} local ice and snow thickness.

If the grid cell is classified as melting, the surface flux is approximated as

$$F_{sfc} = g^{s}_{x,t} = F_{LW\downarrow} - \varepsilon_{ice}\sigma T^{4}_{f} + F_{sens} + \left(a_{ice}\alpha_{ice} + \left(1 - a_{ice}\right)\alpha_{ocn}\right)F_{SW\downarrow} \quad (2)$$

where $T_f = 0°C$.

The ice surface albedo α_{ice} is further expressed as

$$\alpha_{ice} = \left(\alpha_{melt\_ice} - \alpha_{sea}\right) + I_{snow}\left(\alpha_{melt\_snow} - \alpha_{melt\_ice}\right) + \left(1-\gamma_{melt}\right)\left(1-I_{snow}\right)\left(\alpha_{cold\_ice} - \alpha_{melt\_ice}\right) + \left(1-\gamma_{melt}\right)I_{snow}\left(\alpha_{cold\_snow} - \alpha_{melt\_snow}\right) \quad (3)$$

Here $\alpha_{sea} = 0.06$, $\alpha_{melt\_ice} = 0.535$, $\alpha_{melt\_snow} = 0.65$, $\alpha_{cold\_ice} = 0.61$ and $\alpha_{cold\_snow} = 0.8$ denote the parameterised albedos of open water, melting ice, melting snow, cold ice and cold snow respectively, and γ_{melt} denotes melting surface fraction as a fraction of ice area, while I_{snow} is an indicator for the presence of snow that is set to 1 or 0 depending on whether monthly mean snow thickness exceeds 1mm.

The derivation of the formulae is briefly described. The surface flux is composed of four radiative fluxes (downwelling and upwelling SW and LW), two turbulent fluxes (sensible and latent) and of an additional flux due to snowfall (which affects the surface flux as it represents a transfer of negative latent heat, since snow lying on ice changes the enthalpy of the snow–ice system). Hence

$F_{sfc} = \left(1-\alpha_{ice}F_{SW}\right) + F_{LW\downarrow} - \varepsilon_{ice}\sigma T_{sfc}^4 + F_{sens} + F_{lat} + F_{snowfall}$ is used as a starting point from which the derivation of (2) follows in the melting season, assuming a surface temperature of 0°C and neglecting the snowfall contribution. (3) is designed to mimic the calculation of ice albedo in HadGEM2-ES, which parameterises the effect of meltponds after Curry (2001), reducing albedo linearly as surface temperature rises from −1°C to 0°C. (1) is derived by considering separately the contributions to F_{sfc} from the area of the grid cell covered by each ice category (and by open water). For each ice category, the conductive flux through the ice is assumed to be uniform; the dependence of F_{sfc} on surface temperature is then linearised, using monthly mean surface temperature at each grid point, $T_{sfc-REF}$, as a reference about which to take the linearization. By setting the conductive flux equal to F_{sfc}, the variable T_{sfc} is eliminated. Finally, the contributions to F_{sfc} are multiplied by category ice area and summed. In deriving (1), the contributions of the snowfall flux and of the latent heat flux over ice are neglected.

In this way, using equations (1)–(3), we construct the functions $g_{x,t}$, which depend on downwelling LW, downwelling SW, sensible heat flux, category ice thickness, category ice area (freezing cells), total ice area (melting cells), snow thickness, snow area and surface melt onset, variables which have the required property of tending to affect the surface flux on timescales shorter than that on which they affect each other. Hence at each point in space and time the rate of dependence of surface flux on each variable can be approximated by $\partial g_{x,t}/\partial v_i$. We describe for the case of three variables how this process can be used to estimate the surface flux bias induced by biases in that variable, firstly for the variable of melting surface fraction (for simplicity, we describe only the process over grid cells judged to be melting). Model daily surface temperature fields are used to judge, for each month of the year, the average melting surface fraction in each grid cell. The satellite-derived observational estimates of surface melt onset described in Section 2.2 are used to produce a climatology of melting surface fraction for each month and grid cell, and this is subtracted to produce a model bias. This bias is

then multiplied by the partial derivative of equation (2) with respect to melting surface fraction, $-F_{SW\downarrow}\left((1-I_{snow})\left(\alpha_{cold\_ice}-\alpha_{melt\_ice}\right)+I_{snow}\left(\alpha_{cold\_snow}-\alpha_{melt\_snow}\right)\right)$, evaluated with monthly mean fields of $F_{SW\downarrow}$ and I_{snow}, to produce a monthly mean field of surface flux bias induced by the model bias in melting surface fraction.

By a similar method, the effect of downwelling LW radiation on surface flux can be estimated, illustrated here using CERES as a reference dataset (in section 4 below the analysis is performed using multiple datasets) to produce fields of model bias in downwelling LW radiation. For freezing grid cells, these are then multiplied by $\sum_{cat} a_{cat}\left(1-BR_{ice}^{cat}\right)^{-1} \Big/ \sum_{cat} a_{cat}$, the partial derivative of (1) with respect to downwelling LW, to produce fields of surface flux bias induced by model bias in downwelling LW. For melting grid cells, the induced surface flux bias is equal to the downwelling LW bias, as the surface temperature does not change in response to the bias.

The most complex variable to analyse in this way is the ice thickness. Ice thickness strongly affects the surface flux in the freezing season; thicker ice is associated with less conduction, a colder surface temperature and a weaker negative surface flux, and hence reduced ice growth. However, it appears in equation (1) only implicitly, in the form of the individual category mean thicknesses h_I^{cat}. To use this equation to estimate the effect of ice thickness biases on surface flux, a method of estimating the way biases are distributed amongst thickness categories is needed. Given an estimated model bias in mean thickness $\overline{h}_{ice}^{'}$, it can be argued that the least arbitrary approach is to estimate the model bias in each thickness category to be $\overline{h}_{ice}^{'}$ also (i.e. the thickness distribution is uniformly shifted to higher, or lower values). However, this leads to unphysical results at the low end of the distribution; in the case of a negative bias, it implicitly assumes the creation of sea ice of negative thickness; in the case of a positive bias, it assumes that no sea ice of thicknesses between 0_m and $\overline{h}_{ice}^{'}\_m$ exists.

Hence we use a slightly modified approach. The model bias in the lowest thickness category is estimated to be $\overline{h}_{ice}^{'}/2$, equivalent to translating the top end of the category by $\overline{h}_{ice}^{'}$ but allowing the lower end to remain at 0. The model biases in the other four categories are then estimated to be $\overline{h}_{ice}^{'}\dfrac{a_{ice}-a_1/2}{a_{ice}-a_1}$, i.e. the translation is increased to ensure that the mean ice thickness bias remains correct. Following this, we iterate through the categories, identifying grid cells where the bias is such that a negative category sea ice thickness in the reference dataset is implied; in these cells, the bias is reduced such that the reference thickness in that category becomes 0, and the bias in the remaining categories is increased proportionally to ensure the mean sea ice thickness bias remains correct.

Hence we create, for each category, fields of sea ice thickness bias. These are multiplied by the partial derivative of equation (1) with respect to category ice thickness, $\left(A+BT_b\right)a_{cat}\left(1-BR_{cat}^{ice}\right)^{-2}\left(\sum_{cat} a_{cat}\right)^{-1}$, to create

~~fields of induced surface flux bias for each category. These are then summed to obtain the total induced surface flux bias due to ice thickness bias.~~

~~The process of calculating induced surface flux bias is illustrated in Figure 5 for example months for these three variables. Figure 5a-c illustrates the melt onset analysis. Figure 5a shows the HadGEM2-ES bias in melting surface fraction for the month of June 1980, relative to the NSIDC climatology; the bias is generally positive, reflecting melt onset modelled earlier than observed during this month. Figure 5b shows the field of rate of change of surface flux with respect to melt onset occurrence (effectively downwelling SW multiplied by the difference in parameterised albedos); this tends to be higher in the Central Arctic, reflecting a greater tendency to clear skies here. Finally Figure 5c shows the product of these two fields, the modelled surface flux bias induced by the model bias in melt onset. This is also generally positive, by up to 25 Wm^{-2} in the central Arctic, reflecting the greater absorption of SW radiation induced by the early melt onset.~~

~~Figures 5d-f demonstrate the same process for the downwelling LW radiation in January 1980, using CERES as reference dataset. Modelled downwelling LW radiation is seen to be considerably lower in magnitude than that observed by CERES for the 2000-2013 period, by up to 30 Wm^{-2} in many parts of the Central Arctic (Figure 2d). The rate of change of surface flux with respect to downwelling LW is shown to be higher (closer to 1) in regions of thinner ice (Figure 2e). This has the result that the induced surface flux bias is greatly reduced relative to the downwelling LW bias in regions of thicker ice (Figure 2f), reflecting the lower efficiency of ice creation in regions of thicker ice; the bias is below 10Wm^{-2} over much of the Arctic, only approaching 20 Wm^{-2} in the Barents and Kara seas.~~

~~Figures 5g-i demonstrate the process of calculating surface flux bias induced by the bias in ice thickness in model category 1 (0-0.6m) for the month of January 1980, using PIOMAS as reference dataset. Modelled ice thickness tends to be thinner than estimated by PIOMAS over much of the Arctic for this month, except for an area on the Pacific side of the Arctic; as described above, the bias in category 1 is assumed to be half the total bias. Figure 5h shows the rate of change of surface flux with respect to category 1 ice thickness, which tends to be high in regions where category 1 ice covers higher fractions of the grid cell, generally near the ice edge.~~

The ISF calculation process is now illustrated for two processes in turn. The model bias in melting surface fraction for the month of June 1980 is positive over most of the Arctic, although only weakly so towards the coasts (Figure 6a), reflecting melt onset modelled earlier than observed during this month. The reference dataset is derived from SSMI microwave observations. The rate of change of surface flux with respect to melt onset occurrence tends to be higher in the Central Arctic (Figure 6b). This reflects a greater tendency to clear skies in the Central Arctic, as this field is effectively downwelling SW multiplied by the difference in parameterised albedos. The product of these two fields represents the modelled surface flux bias induced by the model bias in melt onset; this is also positive over most of the Arctic Ocean, by up to 25 Wm^{-2} in the central Arctic, reflecting the greater absorption of SW radiation induced by the early melt onset (Figure 6c)

The model bias in downwelling LW radiation in February 1980 is predominantly negative (Figure 6d); here CERES-EBAF is the reference dataset. The rate of dependence of surface flux on downwelling LW (Figure 6e) is everywhere between 0 and 1, tending to be lower in regions of thicker ice, associated with a greater tendency

for downwelling LW biases to be counteracted by upwelling LW biases in these regions. The resulting ISF bias (Figure 6f) is negative almost everywhere, but is lower in magnitude than the driving downwelling LW bias.

~~2.~~5. I~~Calculating induced~~ surface flux biases~~: Results~~

5.1 Aggregate ISF biases

We~~Using the methods described in Section 4 we~~ calculate surface flux biases induced by model biases in downwelling SW, downwelling LW, ice area, local ice thickness and surface melt occurrence (the variables for which reference datasets are available). The resulting fields are averaged over the model period and over the Arctic Ocean region, to produce for each variable a seasonal cycle of surface flux bias induced by the bias in that variable. The induced surface flux (ISF) biases are displayed in Figure 7~~6~~, together with total ISF bias, net radiative flux bias~~es~~ estimated by the direct radiation evaluation relative to ISCCP-FD, CERES and ERAI, and also sea ice energy~~latent heat~~ uptake biases implied by the seasonal ice mass balance~~thickness~~ bias~~es~~ relative to PIOMAS. The ISF biases are also shown in Table 1, using CERES as reference dataset for the radiative terms. During winter, ISF biases generally sum to negative values, indicating that model biases in this season induce net additional surface energy loss and ice growth. During summer, ISF biases generally sum to positive values, indicating additional net surface energy gain and ice melt. In both seasons, these results are consistent with the radiation and ice mass balance evaluation.

Major ~~ISF biases tend to sum to negative values during the winter (indicating a model bias towards additional anomalous modelled energy loss and ice growth) and to positive values during the summer (indicating a model bias towards additional~~anomalous modelled ~~energy gain and ice melt), consistent with the radiation and ice thickness evaluation. Major~~ roles are identified for particular processes in certain months. ~~Firstly, in June a bias in surface melt onset induces a surface flux bias of -13.6 Wm^{-2}, equivalent roughly to an extra 11cm of melt. This is associated with the meltpond parameterisation of HadGEM2-ES lowering the surface albedo at the end of May as the surface reaches the melting point, in contrast to SSMI observations which show surface melting to commence on average in mid to late June in the 1980-1999 period. Secondly, in August a bias in ice fraction induces a surface flux bias of 9.6 Wm^{-2}, equivalent to an extra 8cm of melt. This is associated with the overly fast retreat of sea ice in HadGEM2-ES, and the low extents in late summer, as noted in Section 3.~~ Major roles are identified for particular processes in certain months. Firstly, in June the surface melt onset bias, through its effect on the surface albedo, induces a surface flux bias of -13.6 Wm^{-2}, equivalent roughly to an extra 11cm of melt. Secondly, in August a bias in ice fraction induces a surface flux bias of 9.6 Wm^{-2}, equivalent to an extra 8cm of melt. This is associated with the overly fast retreat of sea ice in HadGEM2-ES, and the low extents in late summer, as noted in Section 3. Thirdly, from October-March the downwelling LW biases induce substantial surface flux biases ranging from -6.5 to -3.8 Wm^{-2} depending on choice of reference dataset, equivalent to additional sea ice growth of 20-33cm. Finally, from November-March the negative ice thickness bias induces substantial surface flux biases reducing from -8.3 Wm^{-2} to -2.0 Wm^{-2} as the freezing season progresses, equivalent to total additional ice growth of 24cm.

Formatted: Tab stops: 10.66 cm,

Formatted: Outline numbered + Level: 1 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0.63 cm + Indent at: 1.27 cm

Formatted: Font: (Default) Times New Roman, 10 pt, Bold

Formatted: Normal, Indent: Left: 0.63 cm, No bullets or numbering

~~Thirdly, the large model biases in downwelling LW present throughout the freezing season induce substantial surface flux biases, ranging from -6.5 to -3.8 Wm^{-2} from October-March (the surface flux biases are considerably lower than the original downwelling LW biases because of the increasing inefficiency by which surface heat loss is converted to sea ice growth as ice thickens). Throughout this period, the total extra heat loss estimated by this process is roughly equivalent ice growth ranging from 20-33cm. Fourthly, the negative biases in ice thickness present at the end of summer also induce substantial surface flux biases which tend to decrease throughout the freezing season as the thickness biases decrease, with an induced surface flux bias of -8.3 Wm^{-2} in November reducing to -2.0 Wm^{-2} in March. This effect is roughly equivalent to an extra 24cm of ice growth. It is noted that while large ISF biases due to downwelling SW and LW are evident during summer, there is very large spread in these values between observational datasets, to the extent that the sign of the biases are uncertain. It is concluded that it is not possible to determine the net effect of downwelling radiative biases on surface flux during the summer with current observational data.~~

Internal variability in the ISF biases is measured by taking the standard deviation of the whole-Arctic ISF bias for each process and month across all 20 years in the model period, and all four ensemble members used. Variability is highest in the ice area term, reaching 4.0 Wm^{-2} in July. Variability reaches considerable size in some other terms in some months, for example 1.1 Wm-2 for surface melt onset in June, 1.9 Wm^{-2} for ice thickness in November, but is otherwise mainly under 1 Wm^{-2} in magnitude. In each case, therefore, the ISF biases noted above are persistent features of the model; surface melt onset and ice fraction biases induce additional ice melt in summer, while downwelling LW and ice thickness biases induce additional ice growth in winter. The total summer mass balance bias accounted for by the analysed processes is of the order 20cm, while the total winter mass balance bias is of the order 45-55cm depending on radiation reference dataset.

~~.~~

~~Residuals between the total ISF bias and the directly evaluated radiative flux biases (demonstrated using CERES as radiation reference dataset in Table 1) are comparable in magnitude to the differences between the three different evaluations of the radiative flux biases, indicating that observational uncertainty is likely to dominate uncertainty in the ISF biases themselves. For example, the residual between total ISF bias and net radiation bias varies from -15.4 Wm^{-2} in June to 8.1 Wm^{-2} in November, while the difference between net radiation bias as evaluated by CERES and ERAI respectively varies from -16.9 Wm^{-2} in July to -2.4 Wm^{-2} in September. As discussed in Section 3 above, evidence from in situ validation studies is inconclusive as to the true size of the modelled downwelling LW bias, and hence as to the magnitude of the surface flux bias induced by downwelling LW. On the other hand, the evidence of PIOMAS underestimating winter sea ice thickness suggests that the magnitude of this bias, and the assosicated ISF bias, may be underestimated. It is also noted that there is a high uncertainty of the order ±10 Wm^{-2} in the ice area contribution during the winter. This is because the rate of dependence of surface flux on ice area is very high in freezing grid cells (generally 100-200 Wm^{-2}) due to the large differences between turbulent fluxes over sea ice and open water.~~

In Appendix A potential errors in the ISF analysis are discussed and are found to be quite small in magnitude relative to the difference between observational datasets. Firstly, due to sub-monthly variation in the component variables, the winter downwelling LW component may be underestimated in magnitude by around 0.6 Wm^{-2} on average, and the ice area component in August may be overestimated by around 1.6Wm^{-2}. Secondly, due to a separate effect by which the ISF biases do not exactly sum to the total surface flux bias, the total bias in October is likely to be overestimated in magnitude by 3.6 Wm^{-2}. Thirdly, due to nonlinearities in the surface flux dependence on ice thickness, the ice thickness component is overestimated in magnitude by 0.7 Wm^{-2} on average from October-April, with a maximum overestimation in November of 1.9 Wm^{-2}. We note that it is possible that in some months the sum of the ISF biases may be a truer representation of the actual surface flux bias than any of the individual evaluations, as the method combines observational estimates with physical relationships between the various flux components. For example, in the satellite datasets observational errors in the different components are not constrained to correlate in a physically realistic sense.

The most obvious discrepancy between the total ISF bias and the net radiation bias occurs in July, when the sum of the induced surface flux biases is small and of indeterminate sign, while a large positive bias is implied by the sea ice thickness and surface radiation simulations. This may be due to the 'missing process' of surface albedo bias due to the presence of snow on sea ice. Early surface melt onset, and sea ice fraction loss, as modelled by HadGEM2-ES, would be expected to be associated also with early loss of snow on sea ice, with an associated surface albedo bias, with this process reaching its maximum influence at a time between that of the surface melt onset (June) and that of the sea ice fraction loss (August). We note also that the direct effect of thinning ice on ice albedo could induce an additional flux bias relative to the real world, despite the fact that this effect is not modelled in HadGEM2-ES.

An annual mean total ISF bias of -3.6 (CERES) and -4.5 Wm^{-2} (ERAI) is present when the satellite datasets are used as reference (the annual mean total ISF bias for ERAI is -0.1 Wm^{-2}). It is noted that given a negligible contribution of oceanic heat convergence to the sea ice heat budget in HadGEM2-ES or in the real world, as is argued in Section 4, the annual mean surface flux bias would be expected to be substantially smaller than these figures, as a surface flux bias of -4.5 Wm^{-2} is equivalent to a relative thickening of the model sea ice cover by 9m over the 1980-1999 period. Analysis of potential sources of error in the ISF calculations in Appendix A does not produce evidence of a systematic bias that could explain these large annual mean negative biases, although the early-winter errors in the ice thickness component could explain a small portion (0.4 Wm^{-2}). Given the large discrepancy amongst observational datasets, therefore, it is likely that observational inaccuracy plays a significant part in introducing this annual mean bias.

5.2 Spatial variability

Formatted: Font: Bold

Spatial patterns in the ISF biases are now discussed. Consistent with the pattern of net SW bias identified in section 3, the spatial pattern of surface flux bias induced by melt onset occurrence is characterised by a weak maximum in the central Arctic, with values falling away towards the coast. A more sharply-defined pattern is produced by the ice fraction bias in August, with high values across the shelf seas and the Atlantic side of the Arctic falling to low or negative values in the Beaufort Sea; the pattern displayed by the ice thickness-induced bias in November is almost a mirror image. Finally, the surface flux bias induced by downwelling LW in

~~February displays slightly higher values on the Siberian side of the Arctic than the North American side, the reverse pattern to that displayed by the downwelling LW itself in Figure 4d. The contrast is due to the role the effective ice thickness scale factor plays in determining the induced surface flux bias; thicker ice, such as that which tends to be found on the American side of the Arctic in both model and observations, tends to greatly reduce the flux bias. This represents the thickness-growth feedback, the reality that thicker ice will grow less quickly than thin ice under the same atmospheric conditions.~~

Examining first the June melt onset ISF bias, the spatial pattern of the bias is characterised by a maximum in the central Arctic, with values falling away towards the coast; this is directly related to the spatial pattern of the melt onset bias itself shown in Figure 6. It is very similar to the spatial pattern of the directly evaluated net SW bias in this month, providing additional evidence that the melt onset bias is the principal cause of this. This implie~~nfer~~s that the additional ice thinning induced by this bias is greatest in the Central Arctic and least at the coasts.

The August ice concentration ISF bias displays a sharply defined pattern, with high values across the shelf seas and the Atlantic side of the Arctic falling to low or negative values in the Beaufort Sea, again very similar to the pattern of the ice concentration bias itself. The implication is that the model bias towards ice thinning in August is largely based in areas where ice concentration is already biased low.

The November-March ice thickness ISF bias displays a pattern which is almost identical to that of the August ice fraction ISF bias, but with the opposite sign, with high negative values on the Atlantic side of the Arctic rising to near-zero values in the Beaufort Sea. Hence this model bias has the reverse effect to that of the ice concentration bias, reducing existing ice thickness biases by promoting additional ice growth in these areas.

Finally, the winter downwelling LW ISF bias is much more spatially uniform, but displays slightly higher values on the Pacific side of the Arctic than the Atlantic side, a different pattern to that displayed by the downwelling LW itself in Figure 4d. The contrast is due to the role the effective ice thickness scale factor plays in determining the induced surface flux bias; the thicker ice present on the American side of the Arctic, tends to greatly reduce the flux bias. This represents the thickness-growth feedback; thicker ice will grow less quickly than thin ice under the same atmospheric conditions. The downwelling LW bias tends to increase ice growth Arctic-wide, but less so in regions where ice is already thick.

Formatted: Font:

The spatial patterns of total ISF bias shows many similarities to total net radiation bias evaluated by CERES in most months of the year (Figure 8~~7~~), notably a tendency in July and August for positive surface flux biases to be concentrated on the Atlantic side of the Arctic, and a tendency throughout the freezing season for negative surface flux biases to be least pronounced in the Beaufort Sea, where the ice thickness biases are ~~likely to be~~ lowest. We note that the spatial pattern of amplification of the ice thickness seasonal cycle displayed in Figure 3 is very similar, with amplification most pronounced in the Atlantic sector~~near the Atlantic Ocean ice edge~~, and least pronounced in the Beaufort Sea.The surface flux biases produced by ice fraction biases in August, and ice thickness biases in November, provide reasons for the spatial variation in amplification of the ice thickness seasonal cycle seen in Figure 4, as well as the close resemblance of this pattern to the model biases in annual mean ice thickness. Ice which is thinner in the annual mean will tend to melt faster in summer, due to the net

SW biases associated with greater creation of open water (the ice albedo feedback), and to freeze faster in winter, due to greater conduction of energy through the ice (the ice thickness-growth feedback).

5.3 Forcings and feedbacks

~~**3. Discussion**~~

~~The calculation of the surface radiative flux biases induced by various key processes in the Arctic Ocean produces results qualitatively consistent with the surface radiation evaluation, and with the surface flux biases implied by the sea ice simulation. Melt onset occurrence and sea ice fraction biases tend to cause anomalous surface warming, and sea ice melt, during the summer, in the HadGEM2-ES historical simulation; downwelling LW and ice thickness biases tend to cause anomalous surface cooling, and hence sea ice growth, during the winter.~~ It is helpful to divide the processes examined into feedbacks (surface flux biases induced by biases in the sea ice state itself) and forcings (those induced by downwelling radiative fluxes and melt onset occurrence). In this sense, a 'forcing' refers to a variable which is independent of the sea ice volume on short timescales, rather than being used in the traditional sense of a radiative forcing. Of the variables examined, downwelling SW and LW radiation, as well as the surface melt onset, have this property, and hence their corresponding ISF biases can each be regarded as a 'forcing' on the sea ice state. However, ice thickness and area do not have this property, and their corresponding ISF biases should be regarded instead as intrinsic feedbacks of the sea ice state.

~~The surface flux bias induced by biases in ice fraction during the melting season can be identified with the effect of the surface albedo feedback on the sea ice state. This is because during the melting season the ice area affects the estimated surface flux only through the surface albedo, and the surface flux biases induced in this way cause associated biases in ice melt.~~

The ice concentration ISF flux bias (specifically during the melting season) can be identified with the effect of the surface albedo feedback on the sea ice state. During the melting season the ice area affects the estimated surface flux only through the surface albedo, and the surface flux biases induced in this way cause associated biases in ice melt.

On the other hand, the ice thickness ISF bias (specifically ~~surface flux bias induced by biases in ice thickness~~ during the freezing season) can be identified with the effect of the thickness-growth feedback on the sea ice state. This is perhaps less obvious, as the ice thickness affects the estimated surface flux via the surface temperature and upwelling LW radiation, while the thickness-growth feedback is usually understood to result from differences in conduction. However, the assumption of flux continuity at the surface in constructing the estimated surface flux means that the cooler surface temperatures, and shallower temperatures gradients occurring for thicker ice categories are manifestations of the same process. Slower ice growth at higher ice thicknesses has a manifestation in a smaller negative surface flux, and the surface temperature is the mechanism by which this is demonstrated. Hence the effect of the thickness-growth feedback is described by the ice thickness-induced component of the surface flux bias.

The~~In this way, the~~ ISF analysis allows the effect of the surface albedo and thickness-growth feedbacks on the sea ice state to be quantified, and compared to the effect of other drivers. Arctic-wide, the surface albedo feedback, diagnosed as the ice area-induced component of the surface flux bias, contributes an average of 5.2

Formatted: Indent: Left: 0 cm

Formatted: Outline numbered + Level: 1 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0.63 cm + Indent at: 1.27 cm

Wm^{-2} to the surface flux bias over the summer months, equivalent to an extra 13cm of ice melt; this is very similar to the effect of the surface melt onset-induced component, which contributes an average of 5.3 Wm^{-2}, equivalent also to an extra 13cm of ice melt. In the freezing season, meanwhile, the thickness-growth feedback, diagnosed as the ice thickness-induced component of the surface flux bias, contributes an average of -4.4 Wm^{-2} to the surface flux bias from October-April, equivalent to an extra 26cm of ice freezing, while the downwelling LW-induced component (using CERES as reference dataset) contributes an average of -4.9 Wm^{-2}, equivalent to an extra 29cm of freezing over this period.

5.4 ISF residuals and observational uncertainty

The ISF biases, summed over all independent variables, should approach the true total surface flux bias. However, this is difficult to evaluate as the true surface flux bias is not known. Hence it is necessary to use proxy quantities to evaluate the total ISF bias: directly evaluated surface net radiation bias (relative to ISCCP-FD, ERAI and CERES respectively); and ice energy uptake bias, derived from ice mass balance bias relative to PIOMAS.

For most months of the year, all estimates of total ISF bias fall within the spread of these four datasets (Figure 6), the exceptions being June and July when total ISF bias is smaller than all surface flux proxies. However, the spread is extremely large. For example, in the month of January the estimates of total ISF bias are -12.3, -8.2 and -6.1 Wm^{-2} (with ISCCP-FD, CERES and ERAI used as downwelling radiation datasets respectively), while the estimates of net radiation bias are -18.2, -11.6 and 0.6 Wm^{-2} from ISCCP-FD, CERES and ERAI respectively, and ice heat uptake bias is estimated as -10.1 Wm^{-2}. Hence it is difficult to evaluate the total ISF bias within current observational constraints, and at best it can be said that the total ISF bias is qualitatively consistent, over the year as a whole, with the surface flux bias proxies. A possible cause of the lower total ISF bias in June and July is the 'missing process' of snow on ice, which cannot be evaluated here due to the lack of a reference dataset. The early surface melt onset, and sea ice fraction loss, as modelled by HadGEM2-ES, would be associated with an early loss of snow on ice, with an additional surface albedo bias ~~that is not accounted for here due to the lack of a reference dataset.~~ and hence an additional ISF bias.

On the other hand, the annual mean ice heat uptake bias (0.0 Wm^{-2}) provides a strong constraint on the annual mean surface heat flux bias, in the absence of a significant oceanic heat convergence contribution. For example, the annual mean total ISF biases are -3.6 Wm^{-2} and -4.5 Wm^{-2} when CERES and ISCCP-FD are used as reference datasets respectively; these would imply sea ice thickening of 7m and 9m over the 1980-1999 in HadGEM2-ES, which does not occur. Hence in the annual mean, the total ISF bias is too low. This annual mean bias is related to the tendency for the ISF analysis to account for a greater bias towards ice growth in winter (45-55cm), than that towards ice melt in summer (20cm). It is likely to derive, at least in part, from the use of multiple reference datasets whose errors are not constrained to correlate in a physically realistic sense, but may also be related to the missing processes in June and July.

Formatted: Superscript

Formatted: Superscript

Observational error is one potential cause of error in the ISF biases. An idea as to the potential magnitude of this can be seen from the large spread in SW and LW ISF bias (across different datasets) during summer (Figure 6).

For example, in July the model downwelling LW bias with respect to ERAI produces an aggregated ISF bias of -7.0 Wm^{-2}, but that with respect to CERES produces an aggregate ISF bias of 8.0 Wm^{-2}. Calculation of ice area ISF biases using NSIDC and HadISST.2 as reference, described in section 2.2 and not shown here, showed a similar magnitude of uncertainty in the ice area term (±10Wm^{-2} in summer, ±2Wm^{-2} in winter). We note that the evidence from in situ validation studies suggests that the winter downwelling LW estimates of ERAI and CERES are more likely to be accurate than that of ISCCP-FD. Hence the downwelling LW ISF bias is likely to be estimated more accurately when ERAI or CERES are reference dataset, and the bias towards ice growth is likely to lie closer to the lower end of the range (20cm).

In Appendix B, inherent theoretical errors in the ISF analysis are discussed and are found to be small relative to the sensitivity to use of observational datasets. The largest errors are listed here: firstly, due to sub-monthly variation in the component variables, the winter downwelling LW component may be underestimated in magnitude by around 0.6 Wm^{-2} on average, and the ice area component in August may be overestimated by around 1.6Wm^{-2}. Secondly, due to the evaluation of surface flux dependence at a model state which is itself biased, the total ISF bias in October is overestimated in magnitude by around 3.6 Wm^{-2}. Thirdly, due to nonlinearities in the surface flux dependence on ice thickness, the ice thickness component is overestimated in magnitude by 0.7 Wm^{-2} on average from October-April, with a maximum overestimation in November of 1.9 Wm^{-2}. As these biases are, in the main, considerably smaller than the differences between ISF biases when different reference datasets are used, it is concluded that observational errors are the more important contribution to error in the ISF biases.

6. Discussion

6.1 Using the ISF framework to understand the HadGEM2-ES sea ice state

The HadGEM2-ES ISF biases are qualitatively consistent with the direct net radiation evaluation and with the sea ice simulation, both in terms of seasonal and spatial variation, and allow the effect of the surface albedo feedback and thickness-growth feedback on the sea ice mass balance to be separated. This allows the HadGEM2-ES sea ice biases to be

~~The biases of the HadGEM2-ES sea ice state can be~~ understood by considering in turn the separate ISF components, their magnitudes, and the times of year when they are important. ~~.~~The anomalous summer sea ice melt is initiated by the early melt onset occurrence, and maintained by the surface albedo feedback, which acts preferentially in areas of thinner ice; the anomalous winter ice growth is maintained both by the thickness-growth feedback (occurring mainly in areas of thinner ice, of greater importance in early winter) and by the downwelling LW bias (more spatially uniform, in late winter). It is unclear that any significant role is played by the downwelling SW bias, as at the only time of year when the radiation datasets agree that this bias is of significant value (May), the induced surface flux biasis more than balanced by that induced by downwelling LW. However this may have a role in causing the later melt onset bias, as discussed below.

The means by which the external forcings – anomalous LW winter cooling, and early late spring melt onset – cause an amplified seasonal cycle in sea ice thickness are clear. It can also be seen~~It is also possible to~~

Formatted: Superscript
Formatted: Superscript
Formatted: Superscript
Formatted: Superscript
Formatted: Numbered + Level: 1 + Numbering Style: 1, 2, 3, … + Start at: 6 + Alignment: Left + Aligned at: 1.27 cm + Indent at: 1.9 cm
Formatted: Font: (Default) Times New Roman, 10 pt, Bold
Formatted: Normal, No bullets or numbering

~~understand~~ how, in the absence of other forcings, these combine to create an annual mean sea ice thickness which is biased low, as seen in Section 3. The melt onset forcing, by inducing additional ice melting through its effect on the ice albedo, acts to greatly enhance subsequent sea ice melt through the surface albedo feedback. The downwelling LW, on the other hand, by inducing ice freezing, acts to attenuate subsequent sea ice freezing through the thickness-growth feedback. The effect is that surface flux biasesinduced by melt onset occurrence are enhanced, while those induced by downwelling LW are diminished.

Acting together, the ice thickness-growth feedback and surface albedo feedback create a strong association between lower ice thicknesses and amplified seasonal cycles, because ice which tends to be thinner will both grow faster during the winter, and melt faster during the summer. Hence the melt onset bias, acting alone, would induce a seasonal cycle of sea ice thickness lower in the annual mean, but also more amplified, than that observed, because the surface albedo and thickness-growth feedbacks act to translate lower ice thicknesses into faster melt and growth. For similar reasons, the downwelling LW bias, acting alone, would induce a seasonal cycle of sea ice thickness higher in the annual mean, and also less amplified, than that observed. The bias seen in HadGEM2-ES is a result of the melt onset bias 'winning out' over the downwelling LW, due to its occurring at a time of year when the intrinsic sea ice feedbacks render the ice far more sensitive to surface radiation. The anomalously low ice cover in September arises as a consequence of the low annual mean ice thickness, and in particular of the anomalously severe summer ice melt. The finding that the low annual mean ice thickness is driven by surface albedo biases is consistent with the finding by Holland et al (2010) that variance in mean sea ice volume in the CMIP3 ensemble was mostly explained by variation in summer absorbed SW radiation.

The feedbacks of the sea ice state explain the association between spatial patterns of annual mean ice thickness bias and ice thickness seasonal cycle amplification. However, the external forcings (melt onset and downwelling LW bias) cannot entirely explain the spatial patterns in the mean sea ice state biases, because on a regional scale effects of sea ice convergence, and hence dynamics, become more important. The annual mean ice thickness bias seen in HadGEM2-ES is associated with a thickness maximum on the Pacific side of the Arctic, at variance with observations which show a similar maximum on the Atlantic side. It was shown by Tsamados et al (2013) that such a bias could be reduced by introducing a more realistic sea ice rheology.

6.2 Looking beyond proximate drivers

As noted in the Introduction, the ISF framework can only identify the proximate causes of sea ice biases. Here, we briefly discuss ~~The study would be incomplete without a discussion of~~ possible causes of the two external drivers identified ~~by this analysis~~ as causing sea ice model biases. Underestimation of wintertime downwelling LW fluxes in the Arctic is known to be a widespread model bias in the CMIP5 ensemble (e.g. Boeke and Taylor, 2016). Pithan et al (2014) showed that this bias was likely to be a result of insufficient liquid water content of clouds forming in subzero air masses, resulting in a failure to simulate a particular mode of Arctic winter climate over sea ice; the 'mild mode', characterised by mild surface temperatures and weak inversions, whose key diagnostic is observed to be a net LW flux of close to 0 Wm^{-2} (Stramler et al, 2011; Raddatz et al, 2015 amongst others). HadGEM2-ES was not one of the models assessed by Pithan et al (2014), but its winter climate simulation displays many of the characteristic biases displayed by these, notably a tendency to model very low cloud liquid water fractions during winter compared to MODIS observations (Figure 9~~8~~a) and a failure to

simulate the milder mode of Arctic winter climate as demonstrated in SHEBA observations, diagnosed by 6-hourly fluxes of net LW (Figure 9~~8~~b). Here we conclude that a similar mechanism is likely to be at work in HadGEM2-ES, and that insufficient cloud liquid water is the principal driver of the anomalously low downwelling LW fluxes.

The causes of the early melt onset bias of HadGEM2-ES are harder to determine. For most of the spring, comparison of daily upwelling LW fields of HadGEM2-ES to CERES-SYN observations (not shown) shows the Arctic surface to be anomalously cold in the model, as during the winter. During May, however, upwelling LW values rise much more steeply in the model, and surface melt onset commences during mid-to-late May, far earlier than in the satellite observations. A possible cause of the overly rapid surface warming during May is the zero-layer thermodynamics approximation used by HadGEM2-ES, in which the ice heat capacity is ignored. Comparing fields of surface temperature in HadGEM2-ES between the beginning and the end of May shows a 'missing' ice sensible heat uptake flux of 10-30 Wm^{-2} over much of the central Arctic, which would in turn be associated with a reduction of flux into the upper ice surface of 5-15 Wm^{-2}. Examination of modelled and observed daily timeseries of downwelling LW and net SW fluxes in late May and early June suggests that a surface flux reduction of this magnitude could delay surface melt by up to 2 weeks, a substantial part of the modelled melt onset bias seen.

Another cause of the rapid warming may be the increasing relative magnitude of the downwelling SW response to cloud biases as May progresses (compared to the downwelling LW response). Comparison of 5-daily means of HadGEM2-ES radiative fluxes during May to those from the CERES-SYN product (not shown) support this hypothesis; a modelled bias in downwelling SW grows quickly during early May, from ~ $0Wm^{-2}$ to ~ $30Wm^{-2}$, while the modelled bias in downwelling LW remains roughly constant.

6.3 Missing processes in the ISF analysis

The ISF analysis, as presented, does not comprise an exhaustive list of processes affecting Arctic Ocean surface fluxes. The missing processes of the effect~~s~~ of snow fraction on surface albedo has already been noted, and its likely effect on the total June and July ISF bias. We note also that the direct effect of thinning ice on ice albedo could induce an additional flux bias relative to the real world, despite the fact that this effect is not represented in HadGEM2-ES. ~~and ice thickness bias on the surface albedo have already been noted; T~~the effect of snow thickness bias on winter conduction and surface temperature is another such process which cannot be included due to inadequate observations. Model biases in the turbulent fluxes may also be significant; while the process which is likely most important in determining these during the winter is captured (ice fraction in the freezing season), a more detailed treatment of turbulent fluxes would also examine the effect on these of the overlying atmospheric conditions. It is also noted that snowfall itself is a component of the surface flux which could in theory be evaluated directly given a sufficiently reliable observational reference.

~~Finally, it is noted that a~~ A complete treatment of model biases affecting the sea ice volume budget would also examine causes of bias in oceanic heat convergence. For the reasons discussed in Section 4 these are likely to be small in the Arctic Ocean interior in HadGEM2-ES and observations, but the model bias could nevertheless conceivably be of considerable size in the context of the surface flux biases shown in Figure 6. The total Arctic

Ocean heat convergence modelled by HadGEM2-ES for the period 1980-1999 is 4.4 Wm^{-2}, although this figure shows high sensitivity to the location of the boundary in the Atlantic sector, suggesting that most of this heat is released close to the Atlantic ice edge. This figure is slightly higher than the 3 Wm^{-2} found by Serreze et al (2007) in their analysis of the Arctic Ocean heat budget, but is broadly consistent with observational estimates of oceanic heat transport through the Fram Strait (likely to be the major contributor to Arctic Ocean heat convergence) from 1997 to 2000 by Schauer et al, 2004. This suggests that errors in oceanic heat convergence are unlikely to contribute significantly to sea ice volume biases in HadGEM2-ES. However, for a hypothetical model which simulated greater oceanic heat convergence in the Arctic Ocean interior, the surface flux analysis presented here would fail to adequately describe the model bias in the sea ice volume budget.

Finally, the assumptions underpinning the ISF framework include an instantaneous effect of the independent variables on the surface flux. In the real world, and in many models, there is a time lag associated with the effect of the ice and snow thickness, due to the thermal inertia of the snow and ice. In the time taken for a change of ice thickness to cause a change in surface flux, other variables such as downwelling radiation could in theory be affected, raising doubts as to whether ice and snow thickness are truly independent variables. However, it is likely that any mechanism by which ice thickness could affect another variable would act first through the surface flux, and hence that the timescale on which ice thickness affects surface flux is shorter than that on which it affects other variables. It is also seen in this study that the ISF biases are largest, by far, in the thinnest ice category, where the effect of ice thickness on surface flux would be near-instantaneous. ISF biases in the thickest ice category, where the time lag could be of significant size, tend to be negligible.

~~4.~~7. Conclusions

Formatted: Font: (Default) Times New Roman, 10 pt, Bold

Formatted: List Paragraph, Add space between paragraphs of the same style, Numbered + Level: 1 + Numbering Style: 1, 2, 3, … + Start at: 6 + Alignment: Left + Aligned at: 1.27 cm + Indent at: 1.9 cm

~~HadGEM2-ES simulates a sea ice cover which is not extensive enough at annual minimum. Comparison to various ice thickness datasets shows that it also has too low an annual mean ice thickness, and that its ice thickness seasonal cycle is likely to be overamplified. Evidence of a positive net SW bias during the ice melt season, and a negative net LW bias during the ice freezing season is apparent from evaluations using multiple radiation datasets.~~

~~An evaluation of processes influencing surface radiation, combined with simple models to estimate their effect, produces results consistent with the evaluation of the sea ice state and surface radiation; processes tend to cause anomalous ice melt during the melting season, and anomalous ice growth during the freezing season. Consequently model biases in sea ice growth and melt rate can be attributed in detail to different causes; in particular, the roles played by the sea ice albedo feedback, by the sea ice thickness-growth feedback, and by external forcings, can be quantified.~~

A framework has been designed (the ISF framework) that allows the proximate causes of biases in sea ice mass balance to be separated and quantified. Given reference datasets for independent variables, fields of induced surface flux bias can be calculated from the underlying model bias; these in theory sum to the total surface flux

bias. In practice, the total ISF bias matches both the net radiation bias, and the ice mass balance bias to first order: processes evaluated cause around 40cm additional ice growth during the ice freezing season, and 20cm additional ice melt in winter; a missing process for which we have no reference (snow thickness) is likely to account for at least some additional ice melt in summer. However, observational uncertainty in the evaluated terms prevents direct evaluation of the total ISF bias, and is the largest contribution to ISF uncertainty.

The ISF analysis enables model biases in sea ice growth and melt rate to be attributed in detail to different causes. In particular, the roles played by the sea ice albedo feedback, by the sea ice thickness-growth feedback, and by external forcings, can be quantified. The analysis reveals how the melt onset bias of HadGEM2-ES tends to make model ice thickness both low in the annual mean, and too amplified in the seasonal cycle, with the downwelling LW bias acting to mitigate both effects. The result is consistent with the prediction of DeWeaver et al (2008) that sea ice state is more sensitive to surface forcing during the ice melt season than during the ice freeze season. The analysis also suggests that through an indirect effect on surface albedo at a time when sea ice is particularly sensitive to surface radiation biases, the zero-layer approximation, which was until recently commonplace in coupled models, may be of first-order importance in the sea ice state bias of HadGEM2-ES.

The ISF analysis also allows more detailed analysis of the spatial patterns in sea ice mass balance simulation. In particular, the mechanisms behind the near-identical spatial pattern of biases in annual mean ice thickness (likely driven by ice dynamics) and that of biases in the ice mass balance are explicitly demonstrated. Where ice thickness is biased low in the annual mean, an enhanced seasonal cycle is apparent. This is due to the ice thickness ISF bias (in freezing season) and the ice area ISF bias (in melting season), corresponding to the thickness-growth and ice albedo feedbacks. The downwelling LW and melt onset biases, by contrast, are more spatially uniform, and do not contribute to the annual mean ice thickness control on the ice mass balance.

The finding that observational uncertainty is the most important cause of uncertainty in the ISF bias calculation itself suggests that if observational uncertainty could be reduced, the ISF analysis could become a very powerful tool for Arctic sea ice evaluation. In particular, large observational uncertainties for snow cover and summer surface radiation limit the overall accuracy of the methodology presented here. The addition of freezing season snow thickness, and melt season snow fraction, would represent useful extensions to the analysis presented. An additional caveat is that the ISF framework does not consider factors influencing turbulent fluxes (with the exception of the ice area, but this contribution is subject to particularly high uncertainty). It also does not consider the influence of oceanic heat convergence on sea ice state; in HadGEM2-ES the latter is small (~10%), but might be more significant in other models.

The ISF analysis as presented here is designed specifically to approximate HadGEM2-ES, but could in principle be generalised to other models, particularly by altering the surface albedo parameterisation used here, or by using different sea ice thickness categories. The zero-layer thermodynamic assumption used in the ISF analysis is likely to be appropriate for any model during the ice freezing season, as the largest ISF biases tend to arise from the thinnest ice categories, where the zero-layer approximation is closest to reality. However, there is a question as to whether the zero-layer approximation conceals significant surface flux bias relating to ice sensible heat uptake in the late spring.

Formatted: Font:

~~A clear link has been demonstrated between the spatial pattern of biases in annual mean ice thickness, likely driven by ice dynamics, and that of biases in the April-October thickness. Where ice thickness is biased low in the annual mean, an enhanced seasonal cycle is apparent. This is due to the thickness-growth and ice albedo feedbacks, initiated by melt early melt onset and downwelling LW bias, both of which are spatially uniform.~~

~~Large observational uncertainties for snow cover and summer surface radiation limit the overall accuracy of the methodology presented here. The addition of freezing season snow thickness, and melt season snow fraction, would represent useful extensions to the analysis presented. An additional caveat regarding this analysis is that it does not consider factors influencing turbulent fluxes (with the exception of the ice area, but this contribution is subject to particularly high uncertainty). It also does not consider the influence of oceanic heat convergence on sea ice state; in HadGEM2-ES the latter is small (~10%), but might be more significant in other models.~~

~~-~~In the case study presented here, the analysis provides mechanisms behind a model bias in sea ice simulation. However, the analysis could also be used to investigate a sea ice simulation that was ostensibly more consistent with observations, to determine whether or not the correct simulation was the consequence of model biases that cause opposite errors in the surface energy budget; a negative result would greatly increase confidence in the future projections of such a model. The analysis could be also used to investigate a whole model ensemble, to attribute spread in modelled sea ice state to spread in the underlying processes affecting the SEB, focussing attention on ways in which spread in modelled sea ice could be reduced. It is noteworthy that Shu et al (2015) found the CMIP5 ensemble mean Arctic sea ice volume to be biased low in the annual mean, and overamplified in the seasonal cycle, relative to PIOMAS (albeit over the entire Northern Hemisphere), suggesting that the behaviour exhibited by HadGEM2-ES may be quite common in this ensemble.

Finally, it is suggested that the ISF method, as well as being used to compare a model to observations, could also be used to understand the reasons for the biases of one model with respect to another. Such a comparison would avoid the issues of observational uncertainty discussed above, enabling the contributions of the different model variables to the surface flux biases to be evaluated more accurately. However, the choice as to which model parameters on which to base the ISF framework would be subjective.

Appendix A: Description and derivation of the surface flux formula used in the ISF calculation

The ISF framework depends upon the construction, at each point in model space and time, of an explicit function $g_{x,t}$ which approximates the surface flux as a function of quasi-independent variables v_i. The functions $g_{x,t}$ are constructed as follows. We start from the standard equation for surface flux:

$$F_{sfc} = \left(1-\alpha_{sfc}F_{SW}\right)+F_{LW\downarrow}-\varepsilon_{sfc}\sigma T_{sfc}^4+F_{sens}+F_{lat}+F_{snowfall} \quad \text{(A1)}$$

Field Code Changed

Field Code Changed

Field Code Changed

where the surface flux is expressed as the sum of separate radiative and turbulent components. In this equation, α_{sfc} represents surface albedo, $F_{SW\downarrow}$ downwelling SW flux, $F_{LW\downarrow}$ downwelling LW flux, $\varepsilon_{sfc} = 0.98$ ice emissivity (as parameterised in HadGEM2-ES), $\sigma = 5.67 \times 10^{-8} Wm^{-2} K^{-4}$ the Stefan-Boltzmann constant, T_{sfc} surface temperature, F_{sens} sensible heat flux, F_{lat} latent heat flux, and $F_{snowfall}$ heat flux represented by the transfer of negative enthalpy from the atmosphere to the ice associated with snowfall.

Given a model grid cell *x*, over a model month *t*, the cell is classified as freezing or melting depending upon whether the monthly mean surface temperature is greater or lower than -2°C. In the case that the cell is classified as melting, (A1) is simplified in the following way: $T_{sfc} = T_f$, where $T_f = 0°C$, and $F_{snowfall}$ and F_{lat} are both neglected. In addition, we expand

$$\alpha_{sfc} = a_{ice}\alpha_{ice} + (1 - a_{ice})\alpha_{ocn} \qquad \text{(A2)}$$

where a_{ice} is ice concentration, α_{ice} mean surface albedo over sea ice and $\alpha_{ocn} = 0.06$ the albedo of open water used by HadGEM2-ES. Finally the ice albedo is further expanded

$$\alpha_{ice} = (\alpha_{melt\_ice} - \alpha_{sea}) + I_{snow}(\alpha_{melt\_snow} - \alpha_{melt\_ice}) + (1 - \gamma_{melt})(1 - I_{snow})(\alpha_{cold\_ice} - \alpha_{melt\_ice}) + (1 - \gamma_{melt}) I_{snow}(\alpha_{cold\_snow} - \alpha_{melt\_snow}) \qquad \text{(A3)}$$

Here $\alpha_{sea} = 0.06$, $\alpha_{melt\_ice} = 0.535$, $\alpha_{melt\_snow} = 0.65$, $\alpha_{cold\_ice} = 0.61$ and $\alpha_{cold\_snow} = 0.8$ denote the parameterised albedos of open water, melting ice, melting snow, cold ice and cold snow respectively, and γ_{melt} denotes melting surface fraction as a fraction of ice area, while I_{snow} is an indicator for the presence of snow that is set to 1 or 0 depending on whether monthly mean snow thickness exceeds 1mm. Equation (A3) mimics the parameterisation of ice albedo in HadGEM2-ES, in which the albedo of both snow and ice is progressively reduced from the 'cold' to the 'melting' values as surface temperature rises from -1°C to 0°C.

If the grid cell is classified as freezing, we likewise ignore $F_{snowfall}$. We assume $F_{SW\downarrow}$ and $F_{LW\downarrow}$ to be constant across a grid cell, and parameterise α_{sfc} as above. The remaining terms ($-\varepsilon_{ice}\sigma T_{sfc}^4$, F_{sens} and F_{lat}) we expect to vary over the six different surface types present in a grid cell: open water, and the five different ice thickness categories. Hence we express each as a sum over the surface types:

$$-\varepsilon_{ice}\sigma T_{sfc}^4 = -(1 - a_{ice})\varepsilon_{ice}\sigma T_{sfc-water}^4 - \sum_{cat=1}^{5} a_{ice-cat}\varepsilon_{ice} T_{sfc-cat}^4 \qquad \text{(A4)}$$

Field Code Changed

$$F_{sens} = (1 - a_{ice}) F_{sens-water} + \sum_{cat=1}^{5} a_{ice-cat} F_{sens-cat} \quad \text{(A5)}$$

$$F_{lat} = (1 - a_{ice}) F_{lat-water} + \sum_{cat=1}^{5} a_{ice-cat} F_{lat-cat} \quad \text{(A6)}$$

We make the following further approximations: firstly, that $F_{lat-cat} = 0$ for all ice categories; secondly, that $F_{sens-cat} = F_{sens-ice}$ for all categories (i.e. that the sensible heat flux does not vary across categories); thirdly, that $T_{sfc-water} = 1.8°C$. Lastly, for each ice category we approximate $T^4_{sfc-cat} = A + B(T_{sfc-cat} - T_{sfc-REF})$, where $A = T^4_{sfc-REF}$ and $B = 4\varepsilon_{ice}\sigma T^3_{sfc-REF}$, $T_{sfc-REF}$ being the monthly mean surface average temperature of the grid cell *x*.

Flux continuity implies that over each category the surface flux is equal to $F_{condtop-cat}$, the downwards conductive flux from the ice surface, unless surface melting is taking place; as the freezing case is being discussed, melting is assumed to be zero. We also make the zero-layer approximation used in HadGEM2-ES, that the sea ice has no sensible heat capacity and that conduction is therefore uniform in the vertical for each category. This implies that

$$F_{condtop-cat} = (T_{sfc-cat} - T_{bot}) R_{ice-cat} \quad \text{(A7)}$$

where $T_{bot} = -1.8°C$ is ice base temperature and $R_{ice-cat} = \left(\frac{h_{ice-cat}}{k_{ice}} - \frac{h_{snow}}{k_{snow}} \right)$, k_{ice} and k_{snow} being ice and snow conductivity respectively, and $h_{ice-cat}$ and h_{snow} local ice and snow thickness (the latter of which is assumed to be uniform across ice categories). For each category, setting $F_{sfc-cat} = F_{condtop-cat}$ allows $T_{sfc-cat}$ to be eliminated. This results in the following equation for F_{sfc}:

$$F_{sfc} \approx g^{w}_{x,t} = a_{ice} (F_{atmos-ice} + BT_{ocn}) \sum_{cat} \gamma^{cat}_{ice-REF} (1 - BR^{cat}_{ice})^{-1} + (1 - a_{ice}) F_{atmos-ocean} \quad \text{(A8)}$$

where

$$F_{atmos-ice} = F_{LW\downarrow} - \varepsilon_{ice}\sigma T^4_{sfc-REF} + F_{sens-ice} + (1 - \alpha_{ice}) F_{SW\downarrow} \quad \text{(A9)}$$

and

$$F_{atmos-ocean} = F_{LW\downarrow} - \varepsilon_{ocn}\sigma T^4_{ocn} + F_{sens-water} + F_{lat-water} + (1 - \alpha_{ocn}) F_{SW\downarrow} \quad \text{(A10)}$$

Field Code Changed

Hence we approximate the surface flux as a function of ice area, category ice thickness, snow thickness, downwelling SW, downwelling LW, melting surface fraction, sensible heat fluxes over ice and open water, and latent heat flux over open water. In the ISF analysis, we analyse the resulting dependence of surface flux on ice area, ice thickness, melting surface fraction, and downwelling SW and LW, all of which in HadGEM2-ES affect the surface flux instantaneously, and can therefore be said to be quasi-independent.

Ice thickness does not appear in the surface flux formula directly; instead, the surface flux is expressed as a function of the individual category thicknesses $h_{ice-cat}$. To estimate the ISF bias due to ice thickness, it is hence necessary to sum over categories the ISF biases due to bias in each $h_{ice-cat}$. The estimation of model biases in $h_{ice-cat}$, therefore, requires some discussion.

Given an estimated model bias in mean thickness $\overline{h}'_{ice}$, it can be argued that the least arbitrary approach is to estimate the model bias in each thickness category to be $\overline{h}'_{ice}$ also (i.e. the thickness distribution is uniformly shifted to higher, or lower values). However, this leads to unphysical results at the low end of the distribution; in the case of a negative bias, it implicitly assumes the creation of sea ice of negative thickness; in the case of a positive bias, it assumes that no sea ice of thicknesses between *0 m* and $\overline{h}'_{ice}$ *m* exists.

Hence we use a slightly modified approach (Figure A1). The model bias in the lowest thickness category is estimated to be $\overline{h}'_{ice}/2$, equivalent to translating the top end of the category by $\overline{h}'_{ice}$ but allowing the lower end to remain at 0*m*. The model biases in the other four categories are then estimated to be $\overline{h}'_{ice}\,\frac{a_{ice}-a_1/2}{a_{ice}-a_1}$, i.e. the translation is increased to ensure that the mean ice thickness bias remains correct. Following this, we iterate through the categories, identifying grid cells where the bias is such that a negative category sea ice thickness in the reference dataset is implied; in these cells, the bias is reduced such that the reference thickness in that category becomes 0*m*, and the bias in the remaining categories is increased proportionally to ensure the mean sea ice thickness bias remains correct.

Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed

Appendix BA: Analysis of potential errors in ISF bias calculation

~~Due to observational uncertainty, it is difficult to directly evaluate the ISF bias calculations. Instead, we examine in turn the two principal sources of error in the method; firstly, error in correctly characterising the dependence of surface flux on a climate variable, and secondly, error in approximating the surface flux bias induced by this as the product of the surface flux dependence with the model bias in that variable.~~

The two principal sources of error in the ISF bias calculation method are examined in turn. Firstly, error in correctly characterising the dependence of surface flux on a climate variable is estimated; secondly, error in approximating the surface flux bias induced by this as the product of the surface flux dependence with the model bias in that variable is estimated.

B1 Error in calculating surface flux dependence

Formatted: Font: Bold

To understand error in calculating dependence of surface flux on model variables~~analyse the first source of error, we begin by comparing~~ fields of the approximated surface flux $g_{x,t}$ are compared to those of the real modelled surface flux F_{sfc}. The $g_{x,t}$ are found to capture well the large-scale seasonal and spatial variation in surface flux, but are prone to systematic errors which vary seasonally, indicated in Figure A1; firstly, a tendency to underestimate modelled negative surface flux in magnitude from October-April by 13% on average; secondly, during May, a underestimation varying from 5-20 Wm-2; thirdly, a tendency to overestimate modelled positive surface flux from June-August by up to 10 Wm-2.~~; thirdly, during May, a underestimation varying from 5-20 Wm-2.~~

Formatted: Left

Examining first the winter underestimation (demonstrated in Figure A1 a-c), it is found that for each model month the relationship between estimated and actual surface flux is ~~strongly~~ linear, with underestimation factors ranging from 6 ± 1% in December to 17 ± 2% in April. This suggests that the cause lies in systematic underestimation of the scale factor $\sum_{cat} \gamma_{ice-REF}^{cat} \left(1 - BR_{ice}^{cat}\right)^{-1}$. A possible cause is covariance in time between $\gamma_{ice-REF}^{cat}$ and R_{ice}^{cat} within each month, particularly in the first ice category; during the freezing season, occurrence of high fractions of ice in category 1, the thinnest category, would be expected to be associated with formation of new ice, and correspondingly lower mean thicknesses of ice in this category, lower values of R_{ice}^{cat} and higher values of $\left(1 - BR_{ice}^{cat}\right)^{-1}$. A calculation using daily values of $\gamma_{ice-REF}^{cat}$ ranging from 0.1 – 0.5, and daily values of h_I^{cat} ranging from 0.2 – 0.5m, predicts that this effect would in this case lead to an underestimation of 9% in the magnitude of the surface flux, sufficient to explain all of the underestimation in October, December and January, and most in November, February and March. This effect would produce a corresponding underestimation of the rate of dependence of surface flux on downwelling LW radiation and ice thickness throughout the freezing season. It was estimated that the downwelling LW component of the ISF bias is underestimated by 0.6 Wm^{-2} for the freezing season on average due to this effect.

Secondly, we examine the reasons for the underestimation of surface flux in May (Figure B1d-f), a pattern unique to this month which is seen to be small in the central Arctic but to approach 20 Wm^{-2} at the Arctic Ocean coasts. A likely cause of this inaccuracy is the classification of grid cells as 'freezing' or 'melting' for entire months. During May, as has been seen, most model grid cells in fact cross from one category to the other; however, virtually all Arctic Ocean grid cells are classified as freezing for the month as a whole. The difference field between estimated 'freezing surface flux' and 'melting surface flux' is similar in magnitude and in spatial

pattern to the underestimation field, being near-zero in the central Arctic but rising to 25 Wm^{-2} close to the Arctic Ocean coasts. It is concluded that the actual model mean surface flux is much higher than that estimated near the coast due to these grid cells experiencing melting conditions from relatively early in the month. Although this error is not directly relevant to the results of this paper, as no unequivocal ISF biases were identified for May, it would have the potential to lead to overestimation of the dependence of surface flux on ice thickness, and underestimation of dependence on all other variables, as the upwelling LW flux is unable to counteract changes in surface forcing once the surface has hit the melting point.

~~Second~~Thirdly, we examine the tendency to overestimate surface flux during the summer (Figure A1d-f), an effect that displays a spatially uniform bias rather than a spatially uniform ratio, ranging from 5-15 Wm^{-2} in July and August; the bias is smaller, and in the central Arctic negative, during June. A possible contributing factor to this bias is within-month covariance between ice area and downwelling SW; during July and August, both downwelling SW and surface albedo fall sharply, an effect that would tend cause the monthly mean surface flux to be overestimated. To estimate this effect, monthly trends in these variables were estimated by computing half the difference between modelled fields for the following and previous month. For July, an overestimation in surface flux of magnitude 5-15 Wm^{-2} was indeed predicted in the Siberian seas, as well as the southern Beaufort and Chukchi Seas; however, in the central Arctic no overestimation was predicted, due to near-zero trends in ice area in the summer months. It is possible that some covariance between ice area and downwelling SW is nevertheless present in these regions, due to enhanced evaporation and cloud cover in regions of reduced ice fraction.

However, this effect would have no direct impact on the ISF biases because these are computed from monthly means of the model bias in one variable by the model mean in the other; hence, it is covariance between bias and mean that would induce inaccuracy in this case. By similarly approximating the trend in monthly mean model bias as half the difference between model bias in the adjacent months, the error in downwelling SW and ice area contributions were evaluated. Error in the downwelling SW term was found to be significant early in the summer, with an error of -2.7 Wm^{-2} in June; error in the ice area term was found to be significant later in the summer, with errors of -1.7 Wm^{-2} and -1.6 Wm^{-2} in July and August respectively. However, the August error is small relative to the total ISF bias identified.

~~Thirdly, we examine the reasons for the underestimation of surface flux in May (Figure A1g-i), a pattern unique to this month which is seen to be small in the central Arctic but to approach 20 Wm^{-2} at the Arctic Ocean coasts. A likely cause of this inaccuracy is the classification of grid cells as 'freezing' or 'melting' for entire months. During May, as has been seen, most model grid cells in fact cross from one category to the other; however, virtually all Arctic Ocean grid cells are classified as freezing for the month as a whole. The difference field between estimated 'freezing surface flux' and 'melting surface flux' is similar in magnitude and in spatial pattern to the underestimation field, being near-zero in the central Arctic but rising to 25 Wm^{-2} close to the Arctic Ocean coasts. It is concluded that the actual model mean surface flux is much higher than that estimated near the coast due to these grid cells experiencing melting conditions from relatively early in the month. Although this error is not directly relevant to the results of this paper, as no unequivocal ISF biases were identified for May, it would have the potential to lead to overestimation of the dependence of surface flux on ice~~

~~thickness, and underestimation of dependence on all other variables, as the upwelling LW flux is unable to counteract changes in surface forcing once the surface has hit the melting point.~~

B2 Error in characterising induced surface flux bias

The surface flux dependencies, for each variable, are evaluated at a model state which is itself biased. This introduces an error in characterising the induced surface flux bias. For example, ~~Having examined potential causes of error in estimating dependence of surface flux on individual variables, the validity of estimating ISF biases as the product of these with model variable biases is now discussed. Even if the dependence of monthly mean surface flux on variable v_i at a model grid cell is perfectly described by $\partial g_{x,t}/\partial v_i$, that dependence changes as the realisation varies from the model state to the real-world state. As a simplified example,~~ a component of the surface flux, net SW, is equal to $F_{SW\downarrow}\left(1-\alpha_{sfc}\right)$, and induced surface flux biases due to model biases in $F_{SW\downarrow}$ and α_{sfc} would be calculated as $F'_{SW\downarrow}\left(1-\alpha^{\mathrm{mod}}_{sfc}\right)$ and $F^{\mathrm{mod}}_{SW\downarrow}\alpha'_{sfc}$ respectively. However, the sum of the two induced surface flux biases will not be exactly equal to the true surface flux bias, $F^{\mathrm{mod}}_{SW\downarrow}\left(1-\alpha^{\mathrm{mod}}_{sfc}\right)-F^{obs}_{SW\downarrow}\left(1-\alpha^{obs}_{sfc}\right)$, but will differ from it by $F'_{SW\downarrow}\alpha'_{sfc}$. ~~This is due to the dependencies being evaluated on model states which are themselves biased.~~

This apparent problem can be resolved only by viewing the ISF method as a way not simply of estimating model biases due to a particular variable, but of characterising them, i.e. by accepting that the quantity that we are trying to estimate is itself somewhat subjective. Instead of requiring the ISF method to be correct, it is required that it gives useful, physically realistic results. In the case given above, a sufficient condition is that $F'_{SW\downarrow}\alpha'_{sfc}$ is small relative to $F'_{SW\downarrow}\left(1-\alpha^{\mathrm{mod}}_{sfc}\right)$ and $F^{\mathrm{mod}}_{SW\downarrow}\alpha'_{sfc}$, i.e. that the model bias in both downwelling SW and in surface albedo is small relative to the absolute magnitudes of these variables.

More generally, the difference between the surface flux bias F'_{sfc} and the sum of the induced surface flux biases $\sum_i v'_i \partial g_{x,t}/v_i$ can be approximated by $\sum_{\substack{i,j \\ i\neq j}} v'_i v'_j \partial^2 g_{x,t}/\partial v_i \partial v_j$, a term that can be calculated relatively easily as many of the derivatives go to zero. Averaged over the Arctic Ocean this term was small in most months of the year, but of significant size in October (3.6 $\mathrm{Wm^{-2}}$), due to co-location of substantial negative biases in downwelling LW and category 1 ice thickness in this month, indicating that the true surface flux bias in this month may be substantially smaller (in absolute terms) than the -11.5 $\mathrm{Wm^{-2}}$ obtained from summing the ISF biases.

Finally, the induced surface flux calculation implicitly assumes a linear dependence of surface flux on each climate variable. However, this is not the case for the ice thickness, where higher-order derivatives do not go to zero, and in some regions of thinner ice actually diverge. It is possible to quantify the error introduced by the assumption of linearity by comparing the partial derivative $\left(A+BT_b\right)a_{cat}\left(1-BR^{ice}_{cat}\right)^{-2}\left(\sum_{cat}a_{cat}\right)^{-1}$ to the

quantity $(A+BT_b)a_{cat}\left(1-BR_{cat}^{ice}\right)^{-1}\left(1-BR_{cat}^{ice-REF}\right)^{-1}\left(\sum_{cat}a_{cat}\right)^{-1}$, where $R_{cat}^{ice-REF}=h_I^{OBS}/k_I+h_S/k_S$, h_I^{OBS} being climatological ice thickness in the reference dataset, in this case PIOMAS, and all other terms defined as in Section 4. It can be shown that multiplying this quantity by the model bias produces the exact bias in estimated surface flux that is being approximated by $\partial g_{x,t}/\partial h_I\left(h_I^{MODEL}-h_I^{OBS}\right)$. Hence the bias in the ice thickness component induced by the nonlinearity can be calculated directly. It is found that the nonlinearity causes the ice thickness component to be overestimated in magnitude by 0.7 Wm^{-2} on average from October-April, with a maximum overestimation of 1.9 Wm^{-2} in November.

Code availability

The code used to create the fields of induced surface flux bias is written in Python and is provided as a supplement (directory 'ISF'). The code used to create Figures 1-8, as well as Figure A1, is also provided(directory 'Figures'). In addition, the routines used to estimate errors in the ISF analysis are provided (directory 'Analysis'). Finally, the code used to create Table 1 is provided (directory 'Tables'). A set of auxiliary routines used by most of the above are also provided (directory 'Library').. Most routines make use of the open source Iris library, and several make use of the open source Cartopy library.

Data availability

Monthly mean ice thickness, ice fraction, snow thickness and surface radiation, as well as daily surface temperature and surface radiation, for the first historical member of HadGEM2-ES, is available from the CMIP5 archive at https://cmip.llnl.gov/cmip5/data_portal.html.

NSIDC ice concentration and melt onset data can be downloaded at http://nsidc.org/data/NSIDC-0051 and http://nsidc.org/data/NSIDC-0105 respectively.

PIOMAS ice thickness data can be downloaded at http://psc.apl.uw.edu/research/projects/arctic-sea-ice-volume-anomaly/data/.

ERAI surface radiation data can be downloaded at http://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/.

ISCCP-FD surface radiation data is available at https://isccp.giss.nasa.gov/projects/browse_fc.html.

CERES surface radiation data is available at https://climatedataguide.ucar.edu/climate-data/ceres-ebaf-clouds-and-earths-radiant-energy-systems-ceres-energy-balanced-and-filled.

Acknowledgements

This work was supported by the Joint UK BEIS/Defra Met Office Hadley Centre Climate Programme (GA01101), and the European Union's Horizon 2020 Research & Innovation programme through grant agreement No. 727862 APPLICATE. MC was supported by NE/N018486/1.

Thanks are due to Richard Wood and Ann Keen for helpful comments on multiple versions of this study.

References

Anderson, M., Bliss, A. and Drobot, S.: Snow Melt Onset Over Arctic Sea Ice from SMMR and SSM/I-SSMIS Brightness Temperatures, Version 3. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi: http://dx.doi.org/10.5067/22NFZL42RMUO [accessed October 2015], 2001, updated 2012

Bitz, C. M. and Roe, G. H.: A Mechanism for the High Rate of Sea Ice Thinning in the Arctic Ocean, J. Clim., 17, 18, 3623–3632, doi: https://doi.org/10.1175/1520-0442(2004)017, 2004

Bitz, C. M., Holland, M. M., Hunke, E. C., and Moritz, R. M.: Maintenance of the Sea-Ice Edge, J. Clim., 18, 15, 2903–2921, doi: https://doi.org/10.1175/JCLI3428.1, 2005

Bitz, C. M.: Some Aspects of Uncertainty in Predicting Sea Ice Thinning, in Arctic Sea Ice Decline: Observations, Projections, Mechanisms, and Implications (eds E. T. DeWeaver, C. M. Bitz and L.-B. Tremblay), American Geophysical Union, Washington, D.C.. doi: 10.1029/180GM06, 2008

Boeke, R. C. and Taylor, P. C.: Evaluation of the Arctic surface radiation budget in CMIP5 models, J. Geophys. Res. Atmos., 121, 8525-8548, doi:10.1002/2016JD025099, 2016

Cavalieri, D. J., C. L. Parkinson, P. Gloersen, and H. J. Zwally. 1996, updated yearly. Sea Ice Concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data, Version 1. [Indicate subset used]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi: https://doi.org/10.5067/8GQ8LZQVL0VL. Accessed 2016

Christensen, M. W., Behrangi, A., L'ecuyer, T. S., Wood, N. B., Lebsock, M. D. and Stephens, G. L.: Arctic Observation and Reanalysis Integrated System: A New Data Product for Validation and Climate Study, B. Am. Meteorol. Soc., 97, 6, 907–916, doi: https://doi.org/10.1175/BAMS-D-14-00273.1, 2016

Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Hughes, J., Jones, D., Joshi, M., Liddicoat, S., Martin, G., O'Connor, F., Rae, J., Senior, C., Sitch, S., Totterdell, I., Wiltshire, A. and Woodward, S.: Development and evaluation of an Earth-System model – HadGEM2. Geosci. Model Dev., 4, 1051-1075. doi:10.5194/gmd-4-1051-2011, 2011

Dee, D. P, Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Belsol, C., Dragani, R., Fuentes, M., Geer, A., J., Haimberger, L., Healy, S. B., Hersbach, H., Holm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N. and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Quarterly Journal of the RMS 137:553:597. doi: 10.1002/qj.828, 2011

DeWeaver, E. T., Hunke, E. C. and Holland, M. M.: Sensitivity of Arctic Sea Ice Thickness to Intermodel Variations in the Surface Energy Budget. AGU Geophysical Monograph 180: Arctic Sea Ice Decline: Observations, Projections, Mechanisms and Implications, 77-91. doi: 10.1029/180GM07, 2008

Gultepe, I., Isaac, G. A., Williams, A., Marcotte, D. and Strawbridge, K. B., Turbulent heat fluxes over leads and polynyas, and their effects on arctic clouds during FIRE.ACE: Aircraft observations for April 1998, Atmosphere-Ocean, 41:1, 15-34, DOI: 10.3137/ao.410102, 2003

Holland, M.M., Serreze, M.C. and Stroeve, J., The sea ice mass budget of the Arctic and its future change as simulated by coupled climate models, Clim Dyn (2010) 34: 185. doi:10.1007/s00382-008-0493-4

Hunke, E. C. and Dukowicz, J. K., An Elastic–Viscous–Plastic Model for Sea Ice Dynamics, J. Phys. Oceanogr., 27, 9, 1849–1867, doi: https://doi.org/10.1175/1520-0485(1997)027<1849:AEVPMF>2.0.CO;2, 1997.

Johns, T. C., Durman, C. F., Banks, H. T., Roberts, M. J., McLaren, A. J., Ridley, J. K., Senior, C. A., Williams, K. D., Jones, A., Rickard, G. J., Cusack, S., Ingram, W. J., Crucifix, M., Sexton, D. M. H., Joshi, M. M., Dong, B.-W., Spencer, H., Hill, R. S. R., Gregory, J. M., Keen, A. B., Pardaens, A. K., Lowe, J. A., Bodas-Salcedo, A., Stark, S. and Searl, Y.: The New Hadley Centre Climate Model (HadGEM1): Evaluation of Coupled Simulations. J Clim 19:1327-1353. doi: 10.1175/JCLI3712.1, 2006

Keen, A. B., Blockley, E., Investigating future changes in the volume budget of the Arctic sea ice in a coupled climate model, The Cryosphere, 12, 2855-2868, doi: 10.5194/tc-12-2855-2018, 2018.

Laxon, S., Peacock, N. and Smith, D.: High interannual variability of sea ice thickness in the Arctic region. Nature, 425, 947-950, doi:10.1038/nature02050, 2003

Laxon, S., Giles, K. A., Ridout, A., Wingham, D. J., Willatt, R., Cullen, R., Kwok, R., Schweiger, A., Zhang, J., Haas, C., Hendricks, S., Krishfield, R., Kurtz, N., Farrell, S. and Davidson, M.: Cryosat-2 estimates of Arctic sea ice thickness and volume. Geophys Res Lett 40:732-737. doi: 10.1002/grl.50193, 2013

Lindsay, R.: Temporal Variability of the Energy Balance of Thick Arctic Pack Ice, J. Clim., 11, 3, 313-333, doi: 10.1175/1520-0442(1998)011<0313:TVOTEB>2.0.CO;2, 1998.

Lindsay, R., Wensnahan, M., Schweiger, A. and Zhang, J.: Evaluation of Seven Difference Atmospheric Reanalysis Products in the Arctic, J. Clim., 27, 2588-2606, oi: 10.1175/JCLI-D-13-00014.1, 2013.

Lindsay, R. and Schweiger, A.: Arctic sea ice thickness loss determined using subsurface, aircraft, and satellite observations. The Cryosphere, 9, 269-283. doi: 10.5194/tc-9-269-2015, 2015

Lipscomb, W. H. and Hunke, E. C., Modeling Sea Ice Transport Using Incremental Remapping, Mon. Weather Rev., 132, 6, 1341–1354, doi: https://doi.org/10.1175/1520-0493(2004)132<1341:MSITUI>2.0.CO;2, 2004

Liu, J., J. A. Curry, Rossow, W. B., Key, J. R. and Wang, X.: Comparison of surface radiative flux data sets over the Arctic Ocean. J. Geophys. Res.: Oceans, 110, C2, doi: 10.1029/2004JC002381, 2005

Loeb, N. G., Wielicki, B. A., Doelling, D. R., Louis Smith, G., Keyes, D. F., Kato, S., Manalo-Smith, N. and Wong, T.: Toward Optimal Closure of the Earth's Top-of-Atmosphere Radiation Budget. J Cli, 22, 3, 748–766. doi: 10.1175/2008JCLI2637.1, 2009

Markus, T., Stroeve, J. C. and Miller, J.: Recent changes in Arctic sea ice melt onset, freezeup, and melt season length. J. Geophys. Res., 114, C12024. doi: 10.1029/2009JC005436, 2009

Martin, G. M., Bellouin, N., Collins, W. J., Culverwell, I. D., Halloran, P. R., Hardiman, S. C., Hinton, T. J., Jones, C. D., McDonald, R. E., McLaren, A. J., O'Connor, F. M., Roberts, M. J., Rodriguez, J. M., Woodward, S., Best, M. J., Brooks, M. E., Brown, A. R., Butchart, N., Dearden, C., Derbyshire, S. H., Dharssi, I., Doutriaux-Boucher, M., Edwards, J. M., Falloon, P. D., Gedney, N., Gray, L. J., Hewitt, H. T., Hobson, M., Huddleston, M. R., Hughes, J., Ineson, S., Ingram, W. J., James, P. M., Johns, T. C., Johnson, C. E., Jones, A., Jones, C. P., Joshi, M. M., Keen, A. B., Liddicoat, S., Lock, A. P., Maidens, A. V., Manners, J. C., Milton, S. F., Rae, J. G. L., Ridley, J. K., Sellar, A., Senior, C. A., Totterdell, I. J., Verhoef, A., Vidale, P. L. and Wiltshire, A., The HadGEM2 family of Met Office Unified Model climate configurations, Geosci. Model Dev., 4, 723-757, doi: https://doi.org/10.5194/gmd-4-723-2011, 2011

Maslanik, J., Stroeve, J. C., Fowler, C. and Emery, W.: Distribution and trends in Arctic sea ice age through spring 2011. Geophys. Res. Lett., 38, 47735. doi: 10.1029/2011GL047735, 2011

Formatted: Line spacing: 1.5 lines

Massonnet, F., Fichefet, T., Goosse, H., Bitz, C. M., Philippon-Berthier, G., Holland, M. M. and Barriat, P.-Y.: Constraining projections of summer Arctic sea ice, The Cryosphere, 6, 1383–1394, doi: 10.5194/tc-6-1383-2012, 2012

Massonnet, F., Vancoppenolle, M., Goosse, H., Docquier, D., Fichefet, T. and Blanchard-Wrigglesworth, E., Arctic sea-ice change tied to its mean state through thermodynamic processes, Nat. Clim. Ch., 8, 599-603, doi: 10.1038/s41558-018-0204-z, 2018

Maykut, G. A. and McPhee, M. G., Solar heating of the Arctic mixed layer, J. Geophys. Res., 100, C12, 24,691-24,703, doi: 10.1029/95JC02554, 1995

McLaren, A. J., Banks, H. T., Durman, C. F., Gregory, J. M., Johns, T. C., Keen, A. B., Ridley, J. K., Roberts, M. J., Lipscomb, W. H., Connolley, W. M. and Laxon, S. W.: Evaluation of the sea ice simulation in a new coupled atmosphere-ocean climate model (HadGEM1). J. Geophys. Res., 111, C12014. doi: 10.1029/2005JC003033, 2006

McPhee, M. G., Kikuchi, T., Morison, J. H. and Stanton, T. P.: Ocean-to-ice heat flux at the North Pole environmental observatory, Geophys. Res. Lett., 30, 2274, doi:10.1029/2003GL018580, 2003

Notz, D.: How well must climate models agree with observations?. Philos T Roy Soc A, 373, 2052. doi: 10.1098/rsta.2014.0164, 2015

Perovich, D. K. and Elder, B., Estimates of ocean heat flux at SHEBA, Geophys. Res. Lett., 299, 9, 58-1 – 51-4, doi: 10.1029/2001GL014171, 2002

Perovich, D. K., Richter-Menge, J. A., Jones, K. F. and Light, B.: Sunlight, water, and ice: Extreme Arctic sea ice melt during the summer of 2007. Geophys. Res. Lett., 35, 11. doi: 10.1029/2008GL034007, 2008

Pithan, F., Medeiros, B. and Mauritsen, T.: Mixed-phase clouds cause climate model biases in Arctic wintertime temperature inversions, Clim. Dyn., 43, 289-303, doi:10.1007/s00382-013-1964-9, 2014

Raddatz, R. L., Papakyriakou, T. N., Else, B. G., Asplin, M. G., Candlish, L. M., Galley, R. J. and Barber, D. G.: Downwelling longwave radiation and atmospheric winter states in the western maritime Arctic. Int. J. Clim., 35, 9, 2339-2351. doi: 10.1002/joc.4149, 2014

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C. and Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. J Geophys Res 108:4407. doi:10.1029/2002JD002670, 2003

Rothrock, D. A., Percival, D. B. and Wensnahan, M.: The decline in arctic sea ice thickness: Separating the spatial, annual, and interannual variability in a quarter century of submarine data. J. Geophys. Res., 113, C05003. doi: 10.1029/2007JC004252, 2008

Rutan, D. A., Kato, S., Doelling, D. R., Rose, F. G., Nguyen, L. T., Caldwell, T. E. and Loeb, N. G.: CERES Synoptic Product: Methodology and Validation of Surface Radiant Flux, J. Atmos. Ocean. Tech., 32(6), 1121-1143. http://dx.doi.org/10.1175/JTECH-D-14-00165.1, 2015

Schauer, U., Fahrbach, E., Osterhus, S. and Rohardt,, G.: Arctic warming through Fram Strait: Oceanic heat transport from 3 years of measurements, J. Geophys. Res. (Oceans), 109, C6, doi: 10.1029/2003JC001823, 2004.

Schweiger, A. J.., Lindsay, R. W., Zhang, J., Steele, M. and Stern, H.: Uncertainty in modeled arctic sea ice volume, J. Geophys. Res., doi:10.1029/2011JC007084, 2011

Semtner, A. J.: A Model for the Thermodynamic Growth of Sea Ice in Numerical Investigations of Climate. J. Phys. Oceanogr., 6, 3, 379–389. doi: 10.1175/1520-0485, 1976.

Serreze, M. C., Barrett, A. P., Slater, A. G., Woodgate, R. A., Aagaard, K., Lammers, R. B., Steele. M., Moritz, R., Meredith, M. and Lee, C. M.: The large-scale freshwater cycle of the Arctic, J. Geophys. Res., 111, C11010, doi:10.1029/2005JC003424, 2006

Shu, Q., Song, Z. and Qiao, F. (2015) Assessment of sea ice simulation in the CMIP5 models. The Cryosphere, 9, 399–409. doi: 10.5194/tc-9-399-2015

Stammerjohn, S., Massom, R., Rind, D. and Martinson, D: Regions of rapid sea ice change: An inter-hemispheric seasonal comparison, Geophys. Res. Lett., 39, 6. doi: 10.1029/2012GL050874, 212

Stramler, K., Del Genio, A. D. and Rossow, W. B.: Synoptically Driven Arctic Winter States. J. Clim., 24, 1747–1762. doi: 10.1175/2010JCLI3817.1, 2011

Steele, M. Zhang, J.; Ermold, W.: Mechanisms of summer Arctic Ocean warming, J. Geophys. Res. (Oceans), 115, C11, doi: 10.1029/2009JC005849 (2010)

Stroeve, J. C., Kattsov, V, Barrett, A., Serreze, M., Pavlova, T., Holland, M. M. and Meier, W. N.: Trends in Arctic sea ice extent from CMIP5, CMIP3 and observations. Geophys. Res. Lett., 39, doi: 10.1029/2012GL052676, 2012a.

Stroeve, J. C., Serreze, M. C., Holland, M. M., Kay, J. E., Maslanik, J., Barratt, A. P.: The Arctic's rapidly shrinking sea ice cover: a research synthesis, Clim. Ch., 110, 1005–1027, doi: https://doi.org/10.1007/s10584-011-0101-1, 2012b

Swart, N. C., Fyfe, J. C., Hawkins, E., Kay, J. E. and Jahn, A.: Influence of internal variability on Arctic sea ice trends, Nat. Clim. Ch., 5, 86–89, doi:10.1038/nclimate2483, 2015

Thorndike, A. S., Rothrock, D. A., Maykut, G. A. and Colony, R., The thickness distribution of sea ice, J. Geophys. Res., 80, 4501-4513, doi: 10.1029/JC080i033p04501, 1975

Titchner, H. A., and N. A. Rayner (2014), The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations, J. Geophys. Res. Atmos., 119, 2864-2889, doi: 10.1002/2013JD020316.

Tsamados, M., Feltham, D. L. and Wilchinsky, A. V., Impact of a new anisotropic rheology on simulations of Arctic sea ice, J. Geophys. Res. – Oceans, 118, 1, 91-107, doi: 10.1029/2012JC007990, 2013

Wang, M. and Overland, J. E.: A sea ice free summer Arctic within 30 years? Geophys. Res. Lett., ., 36, L07502, doi:10.1029/2009GL037820, 2009

Wang, X., Key, J., Kwok, R. and Zhang, J.: Comparison of Arctic Sea Ice Thickness from Satellites, Aircraft and PIOMAS data, Remote Sens., *8*(9), 713; doi:10.3390/rs8090713, 2016

Zhang, Y., Rossow, W. B., Lacis, A. A., Oinas, V., and Mishchenko, M. I.: Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: Refinements of the radiative transfer model and the input data, J. Geophys. Res., 109, D19105, doi:10.1029/2003JD004457, 2004

Month	Downwelling SW	Downwelling LW	Ice thickness	Ice area	Melt onset occurrence	Total induced surface flux bias	Radiative flux bias	ISF residual	CERES-ERAI net radiation difference
Jan	0.0	-6.5	-4.3	2.7	0.0	-8.1	-1.6	3.3	-12.2
Feb	0.0	-4.8	-2.8	2.3	0.0	-5.3	-10.4	4.6	-12.5
Mar	0.1	-3.8	-2.0	2.0	0.0	-3.7	-10.5	5.9	-12.0
Apr	0.4	-4.4	-1.4	1.4	0.2	-4.2	-12.2	7.6	-9.7
May	2.1	-4.8	-0.6	0.0	0.1	-3.2	-3.7	0.5	-3.5
Jun	-8.3	7.4	0.0	1.7	11.4	12.2	27.8	-15.4	-4.2
Jul	-13.6	8.0	0.0	3.7	3.3	1.4	5.1	-3.9	-16.9
Aug	0.5	-3.3	-0.1	9.6	1.9	8.5	8.6	-0.0	-12.9
Sep	3.3	-7.1	-0.7	-0.7	0.0	-5.2	-0.2	-4.8	-2.4
Oct	0.9	-5.7	-4.2	-1.8	0.0	-10.8	-14.4	3.4	-4.6
Nov	0.0	-5.4	-8.3	-0.3	0.0	-14.0	-21.6	8.1	-11.7
Dec	0.0	-6.4	-6.4	2.4	0.0	-10.4	-14.9	4.1	-12.2

Table 1. Surface flux biases induced by model bias in 5 different variables in HadGEM2-ES (Wm^{-2}), with CERES used as reference dataset for the radiative components. Total ISF bias and total net radiative flux bias relative to CERES are shown for comparison, as well as their residual; the difference between net radiative flux bias as evaluated by CERES and ERAI is also shown. A positive number denotes a downwards flux, and vice versa.

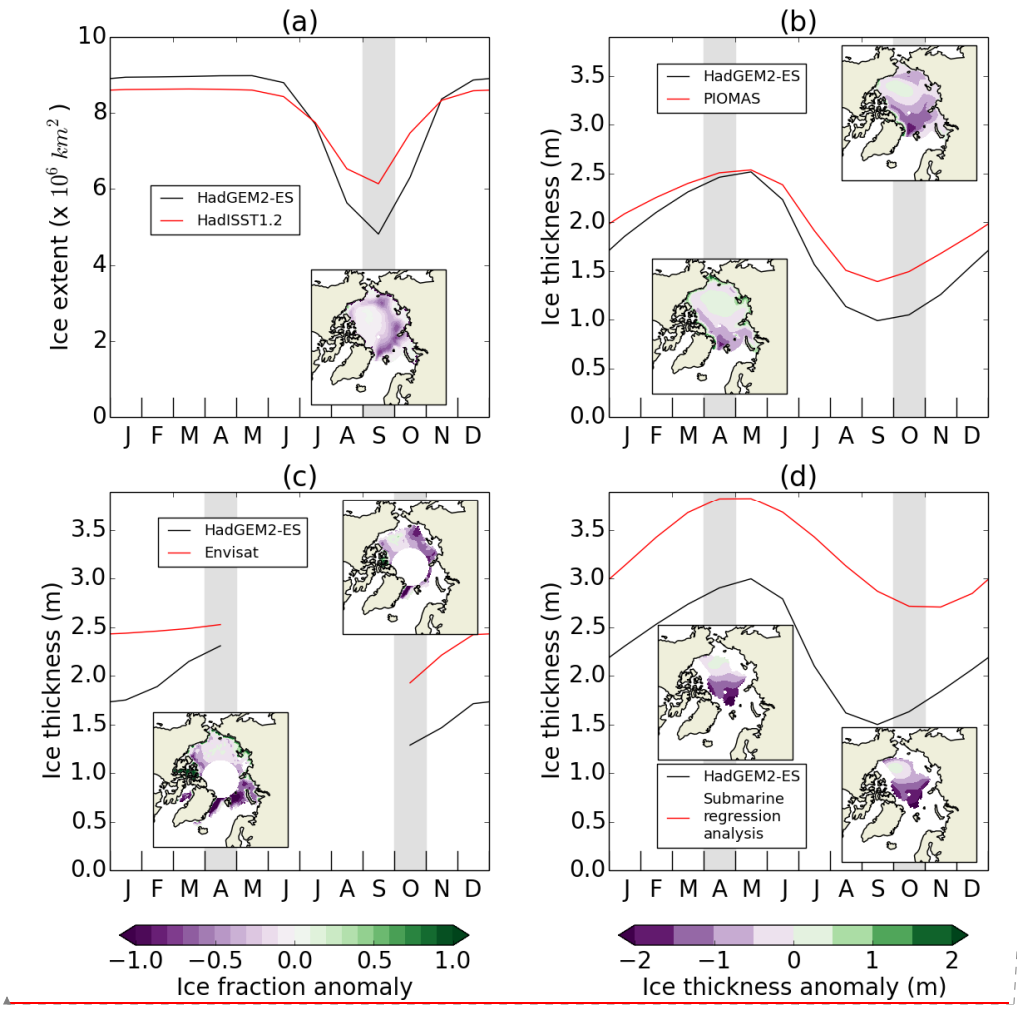
Formatted: Font: Bold

Figure 1. The Arctic Ocean region used in the analysis, defined as the area enclosed by the Fram Strait, northern boundary of the Barents Sea, western boundary of the Kara Sea, the Bering Strait and the northern boundary of the Canadian Archipelago.

Formatted: Font: 9 pt, Bold

Formatted: Font: 9 pt, Bold

1



1

Formatted: Font: (Default) Times New Roman, 9 pt, Bold

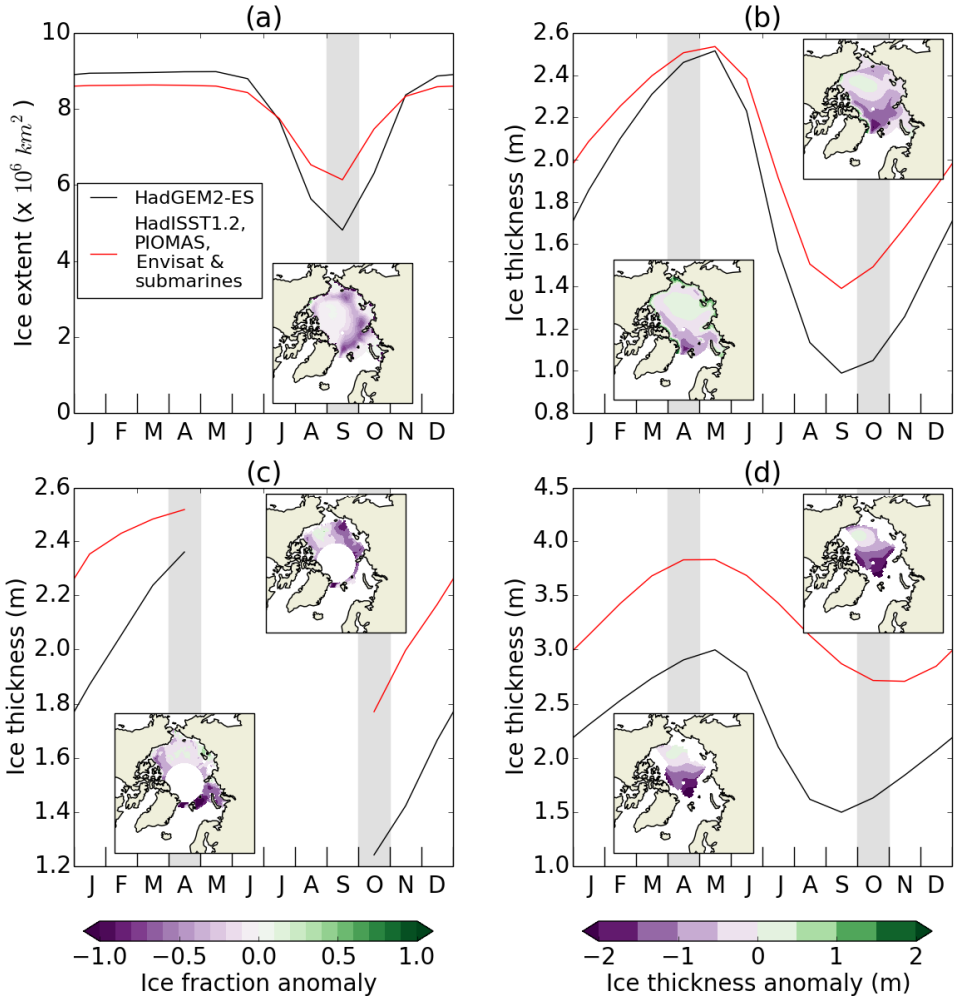
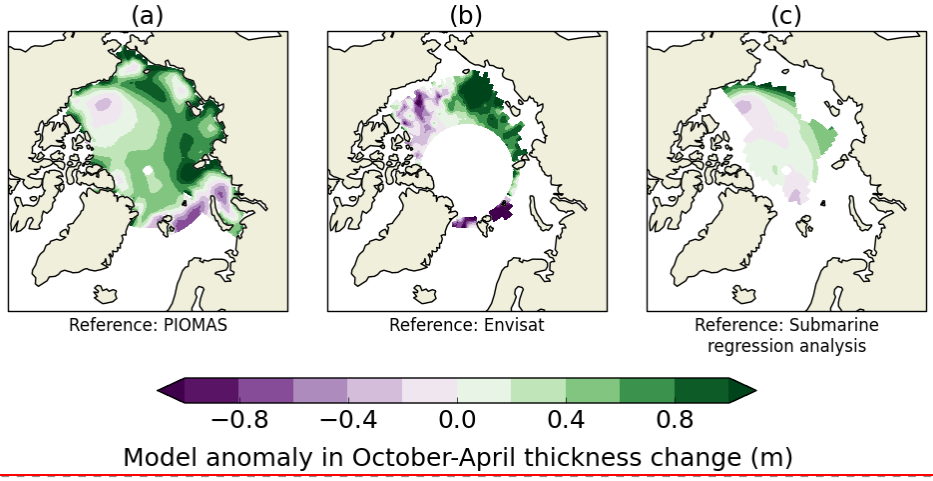


Figure 2. (a) HadGEM2-ES 1980-1999 mean Arctic Ocean ice extent, compared to HadISST1.2 1980-1999, with September ice fraction bias map; (b-d) HadGEM2-ES ~~1980-1999~~ ice thickness compared to (b) PIOMAS 1980-1999, (c) Envisat 1993-1999 and (d) submarine datasets from 1980-1999 over respective regions of coverage, with April and October ice thickness bias maps. For each seasonal cycle plot, the model is in black and observations in red. In (c), data is not plotted from May-September due to the region of coverage being very small.



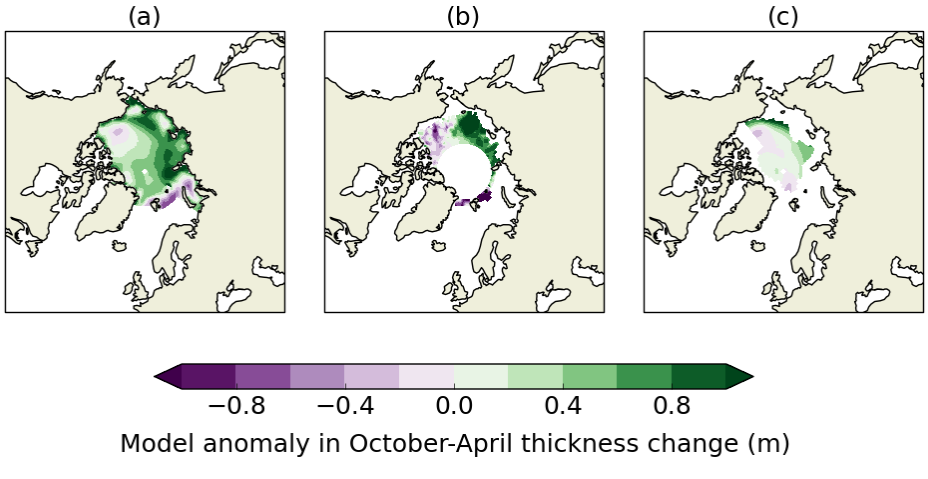
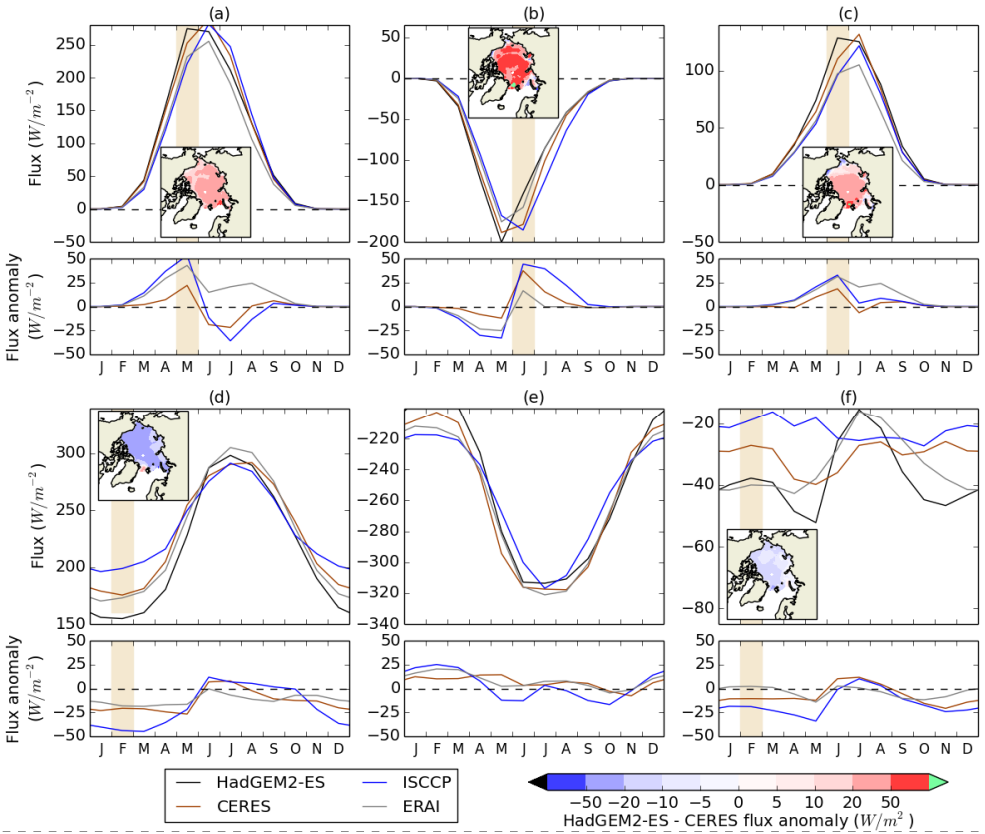


Figure 3. HadGEM2-ES 1980-1999 model bias in ice thickness change from October-April compared to (a) PIOMAS 1980-1999; (b) Envisat 1993-2000; (c) submarine regression analysis 1980-1999. Differences are taken as model-observation so that areas of green (purple) correspond to areas where the HadGEM2-ES model simulates too much (not enough) sea ice growth through the winter

Formatted: Font: (Default) Times New Roman, 9 pt, Bold

Formatted: Font: 9 pt



1

Formatted: Font: (Default) Times New Roman, 9 pt, Bold

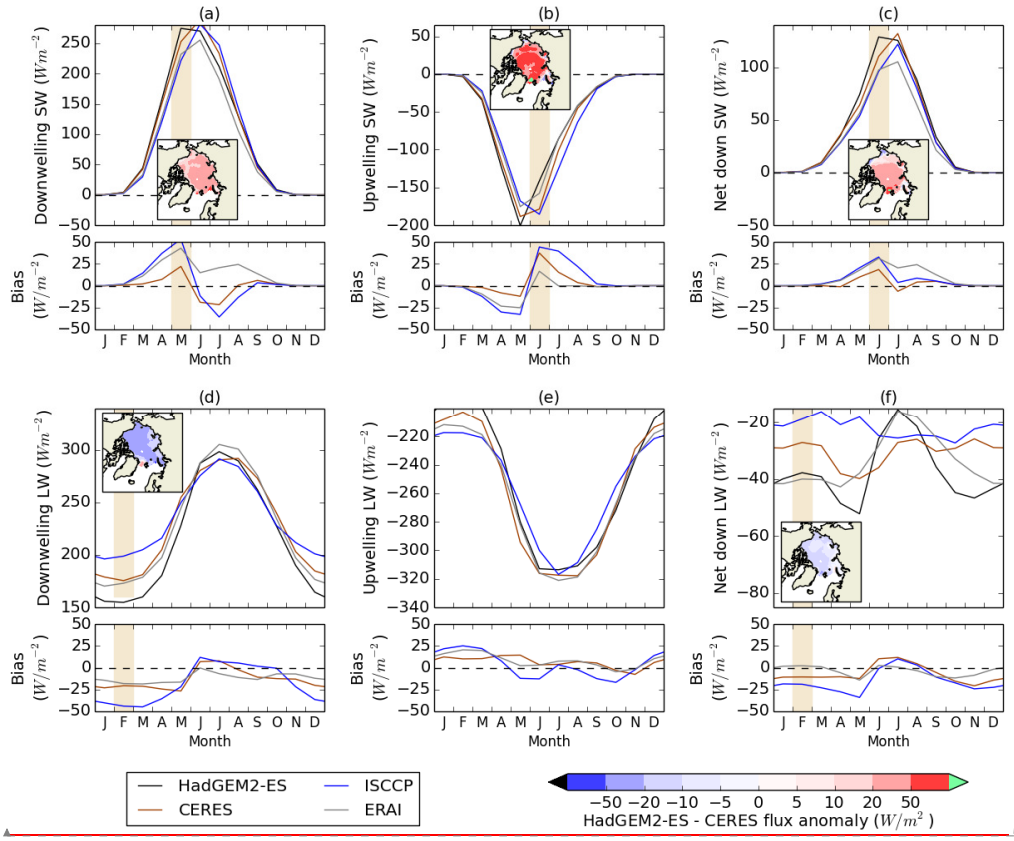
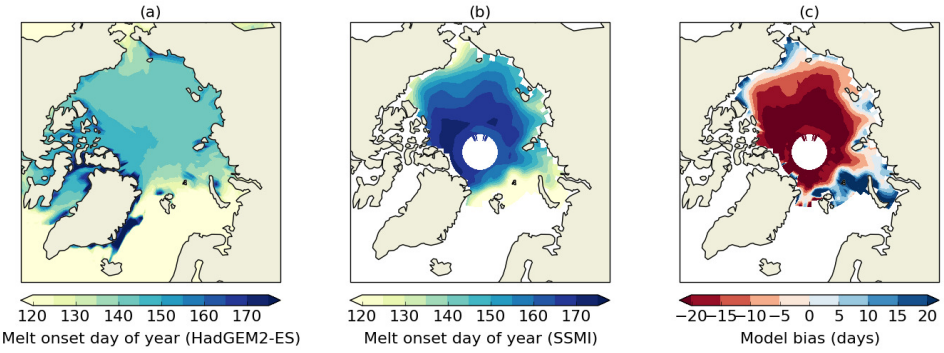


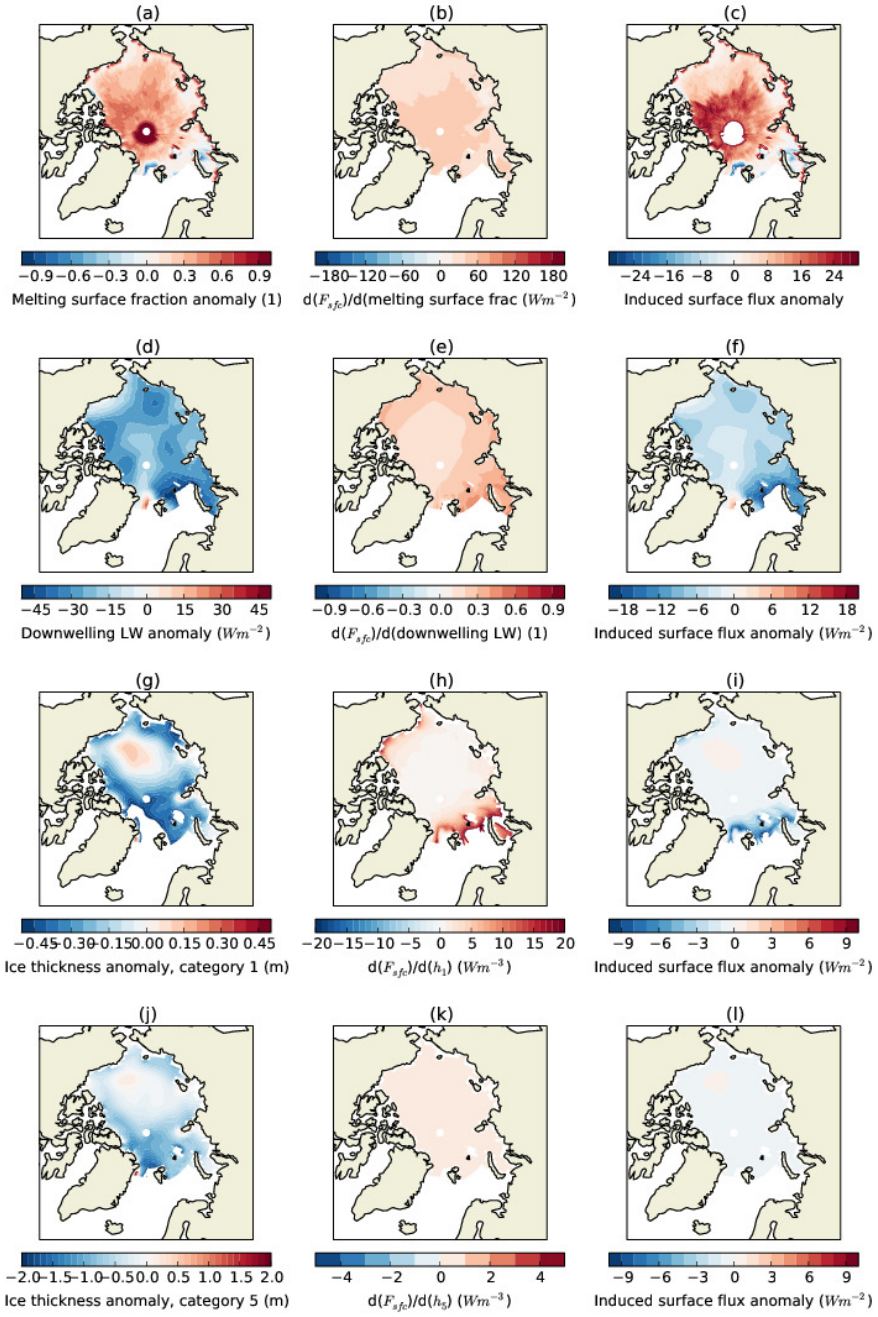
Figure 4. (a) Downwelling SW, (b) upwelling SW, (c) net down SW, (d) downwelling LW, (e) upwelling LW, (f) net down LW, for HadGEM2-ES 1980-1999 over the Arctic Ocean region, compared to CERES 2000-2013, ISCCP-D 1983-1999 and ERAI 1980-1999. Upper panels show absolute values; lower panels show model bias relative to each respective dataset. For all fluxes, a positive number denotes a downward flux and vice versa. Maps of flux bias relative to CERES are shown for downwelling SW in May, upwelling and net down SW in June, and downwelling and net down LW in February.

Formatted: Font: (Default) Times New Roman, 9 pt, Bold



Formatted: Font: (Default) Times New Roman, 10 pt

Figure 5. Average date of year of surface melt onset, 1980-1999, (a) as modelled by HadGEM2-ES, (b) as measured by SSMI observations. (c) shows model bias.



Formatted: Font: (Default) Times New Roman, 9 pt, Bold

1

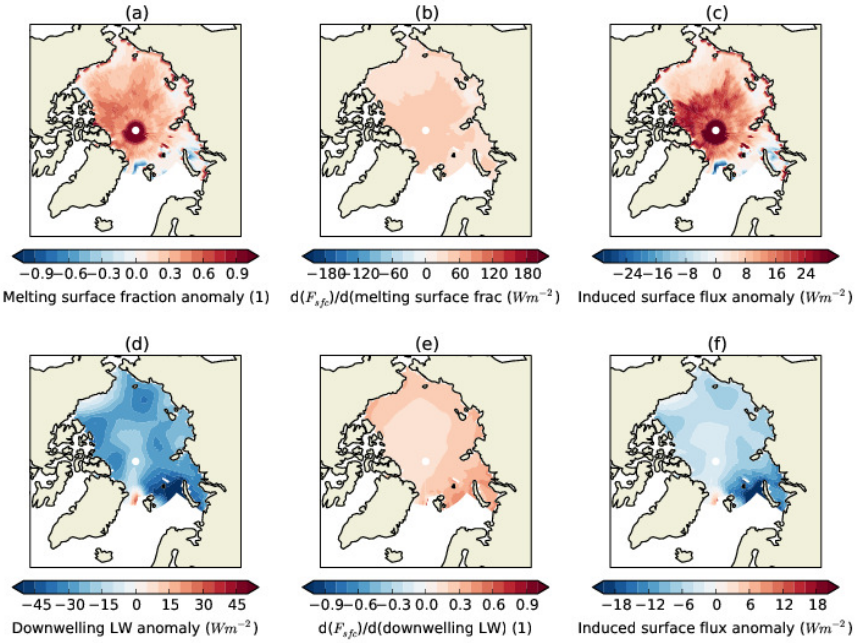


Figure 65. Demonstrating the calculation of fields of surface flux bias due to model bias in melting surface fraction (a-c), downwelling LW (d-f), category 1 ice thickness (g-i) and category 5 ice thickness (j-l). The left-hand column shows model bias in each variable; the middle column the local rate of dependence of surface flux on each variable as calculated above; the right column the induced surface flux bias, calculated as the product of these two fields.

Formatted: Font: (Default) Times New Roman, 9 pt, Bold

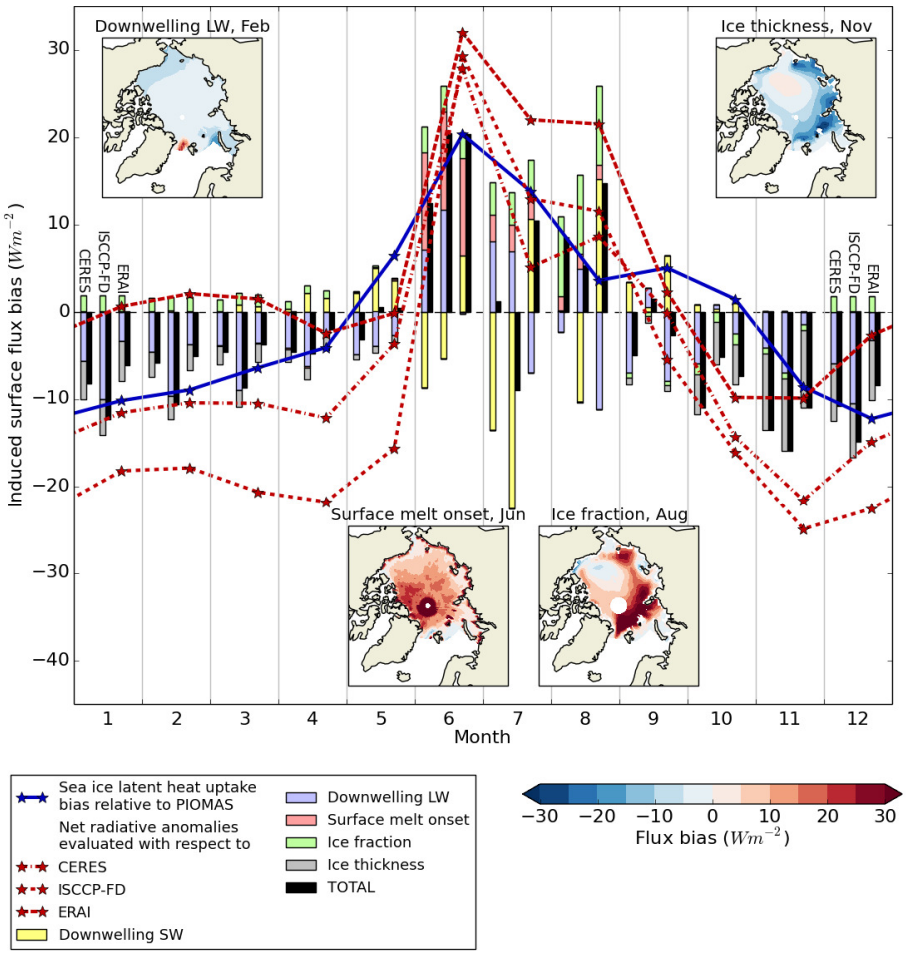


Figure 76. Surface flux bias induced by model biases in ice fraction, melt onset occurrence, ice thickness, downwelling SW and downwelling LW respectively, for the Arctic Ocean region in HadGEM2-ES, 1980-1999, as estimated by the simple models described in Section 2.3. Total ISF bias is indicated in black bars. For each month, induced surface flux biases are estimated using in turn CERES, ISCCP-FD and ERAI as radiation reference datasets, from left-right. Sea ice latent heat flux uptake bias relative to PIOMAS is indicated in black. Net radiative flux biases relative to CERES, ISCCP-FD and ERAI are indicated in brown. Spatial patterns of induced surface flux bias for four processes in key months, with CERES as reference dataset, are displayed beneath.

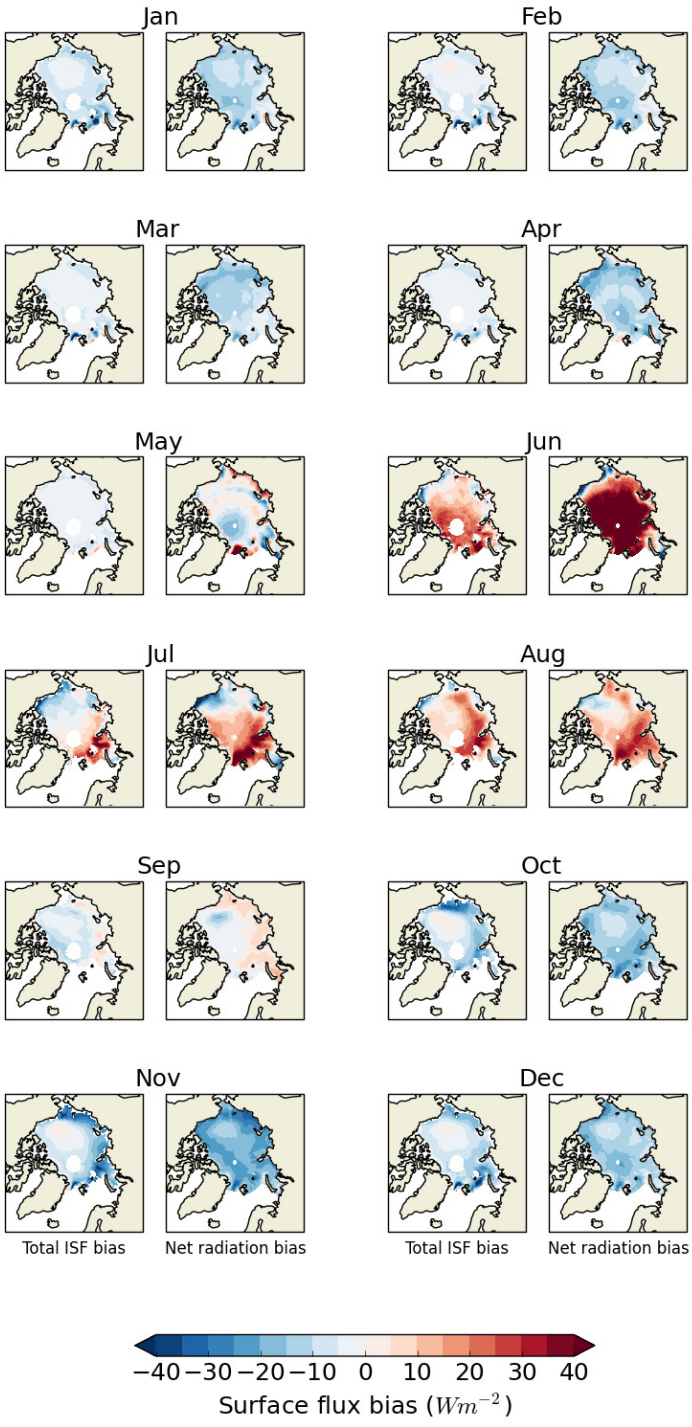


Figure 87. Comparing fields of total ISF bias to net radiation bias relative to CERES for each month of the year, for the four historical members of HadGEM2-ES, 1980-1999.

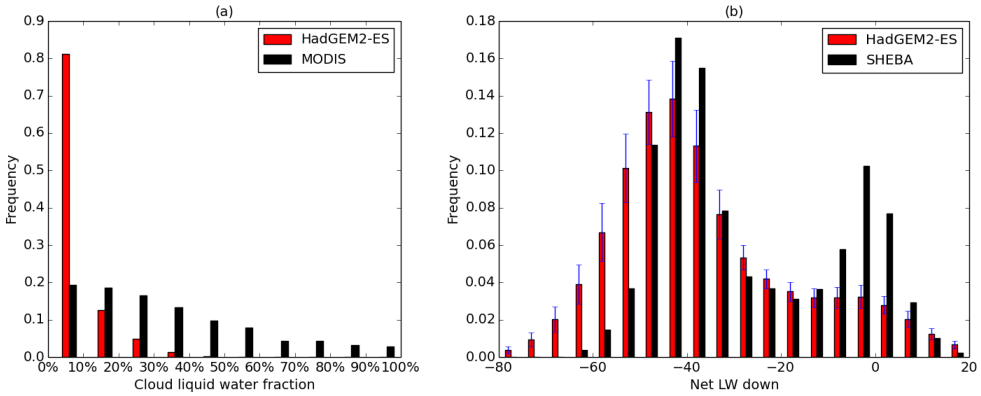


Figure 98. Frequency distributions of (a) October-April cloud liquid water percentage in HadGEM2-ES compared to MODIS observations, for the Arctic Ocean region; (b) December-February surface net downwelling LW in HadGEM2-ES in the SHEBA region, compared to the values observed at SHEBA.

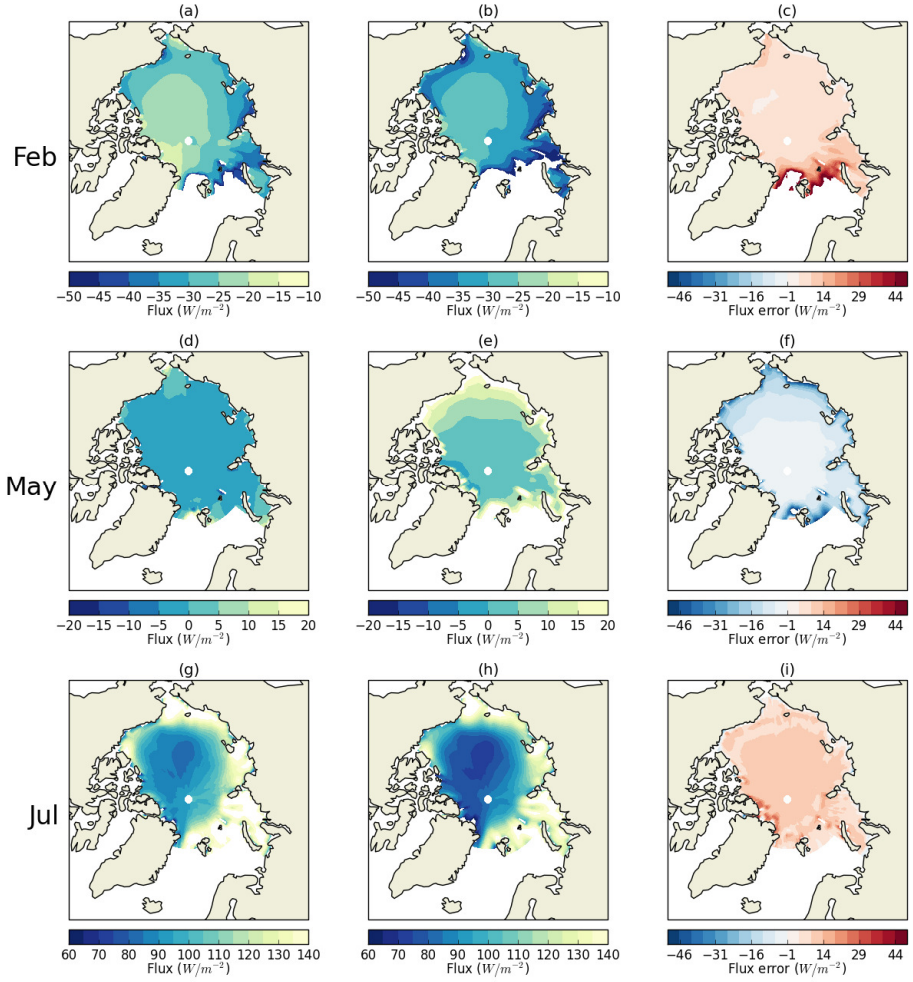


Figure A1. Illustrating approximated (left) and actual (centre) model net surface flux, as well as the approximation error (right), in (a-c) February; (d-f) May; (g-i) July, for the period 1980-1999 in the first historical run of HadGEM2-ES.