Authors' response associated with first resubmission of 'Using Arctic ice mass balance buoys for evaluation of modelled ice energy fluxes

This response is in three parts:

a) An updated point-by-point response to the reviewers. This is in the main very similar to the original response, but in some cases the actual changes made differ from the changes originally suggested.
b) A summary of the changes to the manuscript.
c) A tracked changes version of the manuscript.

a) Point-by-point response to reviewers

Once again, we thank the reviewers for their effort in producing a helpful, constructive set of suggestions for improving the manuscript. We apologise for the length of time it has taken to produce a revision. This is mainly associated with an overhaul of the code (intended to allow the data associated with the publication to be made more widely available), problems encountered during this process, as described in the summary of changes below, and the need for an internal review.

1. Response to reviewer 1

Original reviewer comments are shown in italic font, our response in normal font.

General comments: *The authors present a thorough analysis of the well known dataset of ice mass buoys deployed covering the Central Arctic and Beaufort Gyre regions since 1993 to provide climatology seasonal estimates of the top and bottom ice conductive and melt and ocean fluxes over and under sea ice. The novelty of the method lies in the fine analysis of the data and in the physical processing applied to retrieve meaningful fluxes that can then be used to evaluate climate models such as the HadGEM2-ES Met Office model. I am supportive of this paper being accepted in this general as this dataset and methodology offers a useful tool to the modelling but also remote sensing and in-situ communities. Nevertheless one of the main strength of this paper is the (clean) dataset produced as well as the algorithm developed to analyse these dataset and I strongly encourage the authors to make these data available to the community. In addition,in this case, the nature of the product calls for more transparency and sharing the code and data will warrant easier reproducibility for the scientific community. Finally, a significant effort is needed to clarify the sensitivity of the analysis to the variousconstants and approximations made. I provide some detailed comments below on how this could be achieved.*

We thank the reviewer for their kind remarks. Although it was not possible to publish the code in time for the publication of the discussion paper, due to internal procedures, we are publishing a new version of the code with this revision of the paper.

As noted before, we strongly agreed with the reveiwer's suggestion that, in addition, a version of the processed data should be made available to the community. Following this suggestion, we overhauled the code to allow the data to be produced in netCDF4 format. In addition, we divided the code into two stages, each with an associated output dataset. In the first stage, IMB data is read, quality controlled, and mapped to a consistent set of data series at consistent time points. In the second stage, this processed data is used to calculate the datasets of monthly mean variables.

Because of this division into stages, there are two code repositories published with this paper (on GitHub) and two datasets (on Zenodo), one associated with each stage.

Specific comments:

Abstract

Introduction

P1L23: add reference to Kwok, 2018

This has been included in a rewritten Introduction.

Calculating monthly-mean energy fluxes from the IMBs

P3L26: would a more advanced optimal interpolation scheme improve the results?

Yes, probably – such a scheme could make use of information about other variables, for example. However we decided early on to use a very simple scheme for the present study to ease the processing of data. The results of the estimation scheme used were, individually, sensible. The estimation scheme is contained in a single function in the code, which could easily be replaced by a more advanced scheme in a future study. We have justified this choice with a sentence in our revision.

P3L33: define z_srf and explain a little more (maybe in appendix or with figure 4) how zsrf and zint are sufficient to estimate both changes of surface and bottom sea ice.

This point was not explained very clearly. z_sfc is now defined (surface elevation) and we have tried to improve the clarity of this paragraph.

P4L2: King et al, 2018 and Mercouriadi et al, 2018 have shown that such snow ice formation is prevalent in some regions of the Arctic. Discuss.

We have added a more detailed discussion of the snow-ice formation issue, noting the observations you mention (we have chosen to cite Rösel et al, 2018).

P4L13: couldn't you ask the data providers?

We asked the data providers, and now cite their response.

P4L19: a link to the code would be very valuable here.

The code is now linked.

P4L26: you can cite Alexandrov et al, 2010 for values of snow and ice density. Snow evolves throughout the season with values typically from ~200 to 350 (i.e. Tilling et al, 2017)

We have cited the paper the reviewer suggests. The new revision includes a more comprehensive uncertainty analysis, and as part of this we examine fluxes produced with snow densities of 274 – 374 kgm^{-3}, and ice densities of 917 – 944 $kg\ m^{-3}$; these seemed to us realistic boundary values given the conditions under which IMBs are deployed.

P4L34: not clear where this formula comes from and if it applies to the real snow on sea ice. At what depth?

As suggested in our initial response, we have rewritten this paragraph and provided a figure, and hope that the result is clearer.

P5L9: this fixed thickness (say L) is a parameter of your analysis. Discuss how sensitive your results are to this choice.

We have included this in our expanded uncertainty analysis.

P5L29: similarly how does the uncertainty on these constants impact your results? Discuss.

We have evaluated the impact of using a different scheme for estimating sea ice conductivity, that of Pringle et al (2006). As discussed above we also now evaluate uncertainty due to snow and ice density.

P6L1: These values come from where? Recent work Nandan et al 2018 show salinity at the snow ice interface to be larger than 1. There are more references buy Turner et al 2015 (model work on CICE) but also Notz etc…

We thank the reviewer for providing these references. Our view is that the subject of sea ice salinity is sufficiently complex that the only way of properly accounting for this in the present study, without seriously detracting from its main purpose, is to use uncertainty ranges that encompass all realistic salinity values. Hence in the revised paper, we use expanded uncertainty ranges in the next paper revision, allowing values up to 10 at both the top and lower surfaces of the ice.

P6L5: equation is no readable

This has been corrected.

P6L17-21 where is that shown. Perform proper sensitivity analysis to all these parameters inyour plots, discussion etc…

The reviewer has correctly identified that there are indeed additional sources of uncertainty not accounted for here, and we have tried to incorporate these into our expanded uncertainty evaluation section.

P7L1: interesting. How would you inform the S value.Is it measured? Explain. What problem are you referring to here.

Salinity is not measured, but the temperature and elevation data act to constrain the salinity ranges. For example, if the ice surface is at -0.1m, and the temperature at -0.2m is -0.1 deg C, this implies the melting temperature of the ice at -0.2m is greater than -0.1 deg C. Hence the salinity is lower than 1.9. The ‘problem’ as described in this paragraph is that occasionally the temperatures are in this way inconsistent with the assumed salinity ranges.

P7L5: Tsamados et al, 2015 has implemented the three equation boundary conditions and discussed false bottom impact on sea ice - ocean bottom fluxes

Instead of our stating that false bottom formation renders the computation of ocean heat flux impossible, it is probably more accurate to say that it greatly complicates its calculation – and given the number of data points affected, may be outside the scope of this study. We decided not to add a reference to Tsamados et al because it seemed to us that this study concerned the indirect impact of false bottoms on the ice energy budget, rather than the direct impact.

P7L13: Interesting. Can you see synoptic signal related to snow forcing (i.e. storms?). At what timescale are you solving these? Monthly? Should you pre-process such erroneous signals before monthly averaging? Explain -> share code!

In the overhauled code, we have switched to removing 'cold' top melt data prior to monthly averaging.

Deriving monthly-mean flux distributions from the IMBs

P7L27: why not two regions in the table

We did not want to include too much information in a single table, to improve ease of reading. In the revision, we have provided two additional tables, giving the fluxes by region.

P7L28: why don't you discuss changes between decades I.e. 90s vs 00s vs 10s?

We decided that analysis of interannual variability in the IMB fluxes would be a valuable addition to the paper; this is in a new subsection (3.2). However, this is difficult to perform satisfactorily because of a lack of data points. Only 7 buoys are available for the 1990s; 6 of these were from the SHEBA campaign (i.e. in the same year, 1997-1998, at the same location in the Beaufort Sea), and the remaining buoy, deployed in 1993, was also located in the Beaufort Sea. Hence there is not enough data from this decade to properly sample spatial or interannual variability in the Beaufort Sea region, and none at all in the North Pole region. We chose instead to use a series of irregular periods to compare, with roughly comparable numbers of data points. However, there is still not enough data, or variation, to unequivocally detect interannual variability, and we state this.

P9L1: explain a bit more how these errors on the individual monthly scatter points are obtained. Are you performing an error propagation or are these simply a standard deviation?

This point is superseded by the new uncertainty analysis in subsection 3.3. We do not use scatter points any more in the figure in question, instead demonstrating the IMB dataset with a series of boxplots.

4 ***Evaluating modelled sea ice using the. IMB-derived fluxes***

P9L21: not clear if you estimate the fluxes at the same location in time and space as the IMBs or average over the whole region for the whole month. You should both to test impact of IMB sampling on your results.

We average over the whole region. Model internal variability is such that we do not expect the model to exactly capture the conditions at each point in space and time, and therefore did not see any particular value in sampling the model only at identical points to the IMBs. As discussed in the appendix, we suspect that the largest impact of IMB sampling is through the ice thickness – this would not be solved by evaluating the model at the same points in space and time, as we would still be evaluating fluxes over the entire grid cell. It was not clear to us, therefore, that the impact of IMB sampling would be revealed by the methods that the reviewer suggests. We have included a paragraph to this effect in our revision.

P9L27: why didn't you perform your analysis on a more advanced model with more Ice thickness categories?

We chose to evaluate HadGEM2-ES because its sea ice simulation was already fairly well-understood. Confidence in the IMB-based evaluation could therefore be informed by how consistent this was with the sea ice and surface radiation evaluation. We are now evaluating the new UK CMIP6 models (HadGEM3-GC3.1 and UKESM1.0) in the same way, but this evaluation appears to be outside the scope of this study, which is intended only to demonstrate the new method. Note that although the new models are more advanced (multilayer thermodynamics and explicit meltponds), the number of thickness categories is the same.

P9L35: is it West2018 or 2019.

2019. This has been corrected.

P10L4: again not clear if you perform comparison like for line (i.e. forsame days and grid cells) or not.

No – the model distribution is calculated over the whole region, for reasons described above. Model internal variability means that we do not expect fluxes at the exact same pathways, at the same times, to better represent the conditions than the fluxes over the whole region and time period.

P10: here you list various fluxes but don't explain why you find these results. A bit too descriptive.

We have added some discussion of the results in this section.

P11L5: West 2018 or 2019?

2019 – again this hs been corrected.

P11: discuss role of melt ponds (summer) and snow cover (winter)

The role of snow cover in winter in the conductive flux biases can be investigated directly by comparing IMB-measured snow depths to those modelled by HadGEM2-ES. We have carried out such a comparison alongside a comparison of conductive flux and ice thickness which was requested by Reviewer 2 (see below), in a new subsection 4.3.

The role of melt ponds in summer in the top melt biases (which is what we assume the reviewer means) is more difficult to evaluate directly using the IMBs. However, the meltpond parameterization of HadGEM2-ES was strongly implicated in causing the net surface flux biases found by West et al (2019), and is hence probably implicated in the top melt biases found by this study. We have noted this.

P12: I think a lot of your analysis is missing the link to melt pond coverage

We apologise for omitting to respond to this point in our earlier response.. We have discussed the role of melt ponds in the HadGEM2-ES top melt bias, as indicated above. If Reviewer 1 has additional suggestions to improving this aspect of our analysis we will be very happy to consider how to implement them.

2. Response to reviewer 2

The authors use ice mass balance buoys (IMB) to estimate fluxes through the top and bottom of sea ice. The authors present this new method and then compare the observed fluxes in the North Pole and Beaufort Sea regions. The authors then compare the observed fluxes with modeled fluxes from the HadGEM2-ES climate model. The main findings are that there are biases in these fluxes in the model, which are likely due to the biased mean state of the model. Additionally, there are differences in the fluxes in the Beaufort Sea and North Pole regions. I havea fewm ajorand moderate concerns about the way the model and observations are compared and these need to be addressed before I can recommend publication.

Major concerns

1) Internal Climate Variability
The elephant in the room for a comparison between a climate model and observational data is the issue of internal climate variability (see references below), which you never mention, and leads me to have major concerns with your method. This is also relevant for when the Arctic will become ice free, which you mention in the introduction. Since HadGEM2-ES is a fully-coupled, freely-evolving climate model a single model experiment should not be expected to match the observed sea ice conditions. You do not mention using ensembles and where/how the observations fit in an ensemble spread.

We have expanded the analysis to include the full ensemble (albeit only 4 members) to estimate the internal variability in the fluxes, and compare these to the model biases. The differences are described below, in the summary of changes (briefly, no changes to the model means, but some difference to standard deviation in the top melt flux).

As before we note that the main purpose of the study is to demonstrate the value of the IMB-based evaluation, rather than to draw conclusions from the model biases demonstrated. To demonstrate the value of the evaluation method, it is in our view sufficient that a) the IMB-measured model biases are consistent with biases in the sea ice and surface radiation simulations, and b) they are larger in many months than the observational uncertainty in the IMB fluxes. Whether or not the biases result from internal variability, or another cause, is in our view of only secondary importance for this study. This would not be the case if we were trying to draw conclusions from the model biases, in which case the internal variability context would be vital (hence the many studies of internal variability in sea ice extent trends, some of which the reviewer quotes).

We have rewritten the abstract and introduction to try and clarify this purpose.

Indeed, it appears you are comparing the mean climate model state with the mean observations (Fig. 8). This is not a particularly useful comparison –we know the model is biased from your previous work and therefore we expect to see biases in these mean fluxes as a result!

But we would argue that to find biases in the mean fluxes that are physically consistent with biases in the sea ice simulation is in itself a significant result – because we are demonstrating an entirely new method of model evaluation. We have stated this more explicitly.

Instead, a more useful analysis would be to evaluate situations when the model does have similar thicknesses to those observed, do the fluxes match the observations? That would tell us more about the processes going on in the model and how well they compare to those *observed. You could do this by plotting the distribution of the conductive fluxes by thickness for the model and observations for a particular month or throughout the year.*

A comparison of conductive fluxes to ice thickness, for both the model and IMBs, has been added in a new subsection 4.3. In addition we compare conductive fluxes to snow depth, as suggested by Reviewer 1.

Two additional comments on the model: a) It would be useful to quantify other relevant model biases to the sea ice mass budget like SST or ocean heat transport.

Ocean heat transport has been quantified in our revision, in a paragraph following the evaluation of modelled ocean heat flux. We quantify OHC over the Arctic Ocean as a whole, and also over the North Pole and Beaufort Sea regions, and discuss the likely nature of the link between this and ocean-to-ice heat flux.

b) You compare different years from the model (1980-1999) and observations (1997-2016). I know you did analysis about how the periods are different (Pg.12, lines 13-24) but why not just use the same years that presumably have comparable radiative forcing?

We chose the period of 1980-1999 for consistency with the earlier study of HadGEM2-ES sea ice and surface radiation. The historical ensembles of HadGEM2-ES actually end in 2005, so it would be necessary to use a scenario experiment to get a comparable time period.

2) Additional sources of uncertainty

You mention uncertainty in salinity as one of the big uncertainties in the IMB flux calculations. I think you need to mention that there are also large ranges in the observed snow and ice densities (see refs below) that could cause uncertainty in the retrievals. The values you use are reasonable, but you need to at least acknowledge this and do some basic calculations about how big a difference these values make. Franz et al. 2019 (doi: 10.5194/tc-13-775-2019) Webster et al. 2018 (doi: 10.1038/s41558-018-0286-7)

We have included an analysis of uncertainty due to ice and snow densities in our revision, as part of an expanded uncertainty analysis (new subsection 3.3). We chose ranges of 274-374 kgm^{-3} (snow) and 917-944 kgm^{-3} (ice).

3) Significance

You spend much of section 3 describing differences in observations in two regions and sections 4.1/4.2 describing differences in the mean state of the model and observations. No significance tests were discussed to indicate whether themeans are really significantly different in Figs. 7/8. A simple t-test should suffice, but without this I have a hard time believing some of the conclusions (e.g. that September fluxes differ in the IMB between regions).

This appeared an excellent suggestion, and significance tests have been incorporated throughout sections 3 and 4. A Welch t-test appeared to us to be appropriate, as in most cases the sample sizes will not be the same and the variances of the distributions cannot be assumed to be equal. We have

Moderate concerns

1) Model thermodynamics

I am surprised that HadGEM2-ES, a CMIP class climate model, uses the very simple zero-layer thermodynamics and I think that this should be addressed. The Bitz and Lipscomb thermodynamics is more realistic than zero-layer, and even this has been superseded by the mushy layer thermodynamics of Turner and Hunke (see below). I realize you can't change the model at this point and this shouldn't prevent publication, but your own results show that the assumptions of the zero-layer scheme for conductive fluxes are bad (Fig. 7 and Table 1). Maybe one of the conclusions should be that the zero-layer cannot represent the observed processes so HadGEM2-ES might want to stop using it?

Bitz and Lipscomb 1999 (doi: 10.1029/1999JC900100) Turner and Hunke 2015 (doi: 10.1002/2014JC010358)

The new UK CMIP6 models, HadGEM3-GC3.1 and UKESM1.0, use the multilayer thermodynamics formulation of Bitz and Lipscomb (1999); a further paper is planned to compare these to HadGEM2-ES. But our view was that they are not relevant for the present study, which is intended to demonstrate a new method of evaluation rather than to compare models.

2) Figures

Individually these comments are fairly minor, but the sum of them is moderate.

• *Table 1 –Please add the # values per month and per flux to this table rather than just listing them in the text. Also, adding the units below the flux names (not just in the table description) would be helpful. I expect top melt to usually be in cm/day (or kg m-2s-1), not W/m2.*

Adding number of samples and units are both good suggestions to aid clarity, and have been carried out. For the reasons given in our original response however we have left the units unchanged.

Fig. 1 – It would be helpful to add the sign convention for fluxes here at the interfaces (show the flux direction for positive!). Either label or remove the red/yellow arrows.

The red and yellow arrows have been removed as it was considered they did not add useful information for a diagram that was purely for illustration of processes involved. We were not sure that a sign convention was necessary here for similar reasons.

Fig.3 and Fig.4 –you don't have units on the y-axis or in the labels.

Apologies – these have been added (now figures 4 and 5).

Fig. 4 – Please define surface_r and interface_r. Is snow depth the difference between these two lines? Please clarify on the figure and in text.

We have improved the labelling on this figure (now figure 5).

Fig. 5 –I think a diagram of an IMB would be helpful for modelers, which I think this one is, but it's poorly labeled.What do the dots represent (thermistors?) Is this why they go above the snow interface)? What does the L/R position of the dots mean (temperature?)? Why is there a green line over part of the dots?

We have added a diagram of an IMB (figure 2) and incorporated most of the information on the original figure 5 onto here.

Fig. 6 –the circle colors on this figure are very hard to make distinguish. Perhaps different shaped symbols in black would be better? Again, define surface_r, interface_r, and bottom_r. Are points where there is blue (aka surface_r) mean that the blue and green circles are overlapping? Adding arrows to indicate the transitions for the false bottom would also be helpful? This figure also makes me question why you don't linearly interpolate between the "correct" depth–it looks possible so why lose that data?

We have carried out the corrections to the figure the reviewer suggests, but the figure has been moved to supplementary material to make room for other figures.

Interpolating over the false bottom would probably work in this case. Such a step would need to be carried out at an earlier part of the analysis. However, interpolation might fail in other cases where

- false bottom formation was more gradual in onset
- the false bottom lasted for a longer time, concealing additional ice formation at the true ice base. In this case, analysis of the ice temperature might help, but it remains the case that false bottom formation greatly complicates the analysis.

Fig. 7 – It might be clearer to show the total spread in values with shading and then just a central dot for the mean since the individual points with their spread get hard to distinguish. Also, in text you list the means but they need to be shown to be significantly different.

We have replaced the scatter points with boxplots, and have added a subsection in which uncertainty due to various parameters is evaluated in detail.

Fig. 8 – What purpose does this figure provide other than the model bias, which we already expected from your previous work. Again, if you do show it, significance tests are important.I think a PDF of the flux by thickness would be more helpful to supplement this figure.

We agreed that significance tests, and an examination of the relationship between conductive flux and ice thickness, would both be valuable enhancements of the study, and have added these.

It is true that the figure 'only' shows the model bias – but that is the whole point of a model evaluation, and the purpose of this paper is to demonstrate the value of this method of evaluation. Showing that the method is able to demonstrate model biases despite observational uncertainty, and that these biases are consistent with previous information, is vital to this.

3) Code availability

The effort put in here by the authors to make the data available is lackluster. I understand the model code itself may not be available. However, the authors should make more effort to list how to get the buoy data (raw and processed) as well as the model data (if the code isn't available) since these should be public if it's part of the CMIP archive. See the guidance on the website: https://www.geoscientific-model-development.net/for_authors/code_and_data_policy.html

The code, and associated data, is published alongside this revision. We have overhauled the code to enable production of data in netCDF4, and have divided the code into two stages, each with associated output data. In the first stage, IMB data is read, quality controlled, and processed into a consistent set of data series on consistent time points. In the second stage, monthly mean energy fluxes are calculated from the resulting data. Each stage has an associated code repository on GitHub and an associated dataset repository on Zenodo.

Minor concerns

Shu et al. 2015 isn't in your references.

This has been added.

Pg.4, line 6-8 – do you mean the ice temperature or air temperature? What's going on to cause the non-physical temperatures?

We contacted the data providers to confirm this, and have added their response.

Pg.4 line 14 -How long do buoys last? Give a range here.

We have provided this.

On page 7 you state the convention of positive fluxes indicate downwards. It would be very helpful if you mention that much earlier on Page 4 line 21. And put this on a diagram (Fig.1 and/or Fig.5) too.

We have stated this in the places you suggest (for the reasons above we thought Figure 5 might be preferable to Figure 1 – this information may not be appropriate for an introductory schematic).

Pg.7 line 20 –how thin is "quite thin"? Be specific!

In fact we have removed this paragraph, and associated quality control check, as on reflection we did not think it was important.

Pg.7 line 26-28 –These means are over all available buoys and all years, right? It would be good to be explicit.

Yes, they are over all buoys and all years.

Pg.9 line 15 –change "7 to 143" W/m2 because negative fluxes are possible and the current wording is unclear.

This has been changed.

Pg.8 line 25 AND Pg.10 line 1, mention that those regions are defined in Fig.2.

This has been mentioned.

Pg.12, line 5 –what is the model grid(it hasn't been mentioned)?Why the huge range in grid cell size?

The model grid is a regular latitude-longitude grid (the 'HadGOM grid'), with width one degree latitude and longitude throughout the world except in the tropics, where the latitudinal resolution is somewhat increased. Hence the range in grid widths in km in the Arctic is quite large, falling from ~40km near 70N to ~2km near the North Pole. The grid height, meanwhile, is one degree throughout (~110km). This information has been provided in the model description.

Summary of changes to manuscript

1. Brief summary of changes

The most substantial change to the manuscript arises from the overhaul of the code used to produce the IMB fluxes, to enable production of output data in netCDF format (the code itself is also now published). In the course of this, some bugs and problems were identified and corrected that have resulted in changes to the IMB dataset, that are in the main small. There are also three substantial new areas of analysis: interannual variability in the IMB fluxes, more comprehensive evaluation of uncertainty in the IMB fluxes, and an investigation of the relationship between conductive fluxes, ice thickness and snow depth. In addition, the analysis of internal model variability is improved by including all 4 ensemble members of HadGEM2-ES.

These five major modifications are described in more detail in subsection 2 below. In subsection 3, a complete line-by-line description of all changes to the manuscript is presented.

2. Description of major changes to the study

2.1 Overhauled code, and resulting changes to the IMB dataset

Strong desires were expressed by both reviewers that code and data should be published with any revision of the study. While permission was obtained to publish the code soon after the initial submission, no consideration had previously been made of publishing the associated data, which had not been produced in a particularly commonly-used format. Hence it was decided to overhaul the code to enable production of the IMB data in netCDF4 format. In the course of this, the code was divided into two repositories, each representing different stages of the analysis. In the first stage, the IMB data is read, quality controlled, interpolated or averaged to temperature measurement points, and processed to a standard set of data series, before being written to netCDF4. In the second stage, the processed data is used to produce a set of monthly mean energy fluxes, using the processes described in section 2. The code of each stage is published in separate GitHub repositories, and the two resulting netCDF4 datasets are published in separate linked Zenodo repositories.

There are a number of small differences in the new IMB dataset relative to the old, resulting from a mixture of bugfixes and necessary changes. Specifically:

- All fluxes in the new dataset are produced from timeseries regularised to temperature measurement points. Before, some series were produced from timeseries regularised to daily time points (usually midnight GMT). In the case of the conductive fluxes, this produced a bias when a diurnal cycle was present in the temperature data. Use of the temperature measurement points, which tend to be bi-hourly, removes this bias. This causes large differences in particular in the top conductive flux, especially during the spring.
- The use of different time points also causes missing data points to be treated slightly differently, such that months which may have reported valid data in the old dataset

no longer do so (because temperature was not measured correctly for a significant portion of that month).

- It was found that salinity had not been properly taken into account when calculating latent heat of freezing of the basal surface of the ice; this has been remedied in the new dataset, and generally results in slightly higher ocean-to-ice heat fluxes.
- At the base of the ice, the salinity uncertainty range now takes ice temperature into account. While over most of the dataset, a salinity range of 0-10 is used, salinities which imply a melting temperature below the actual temperature of the ice are ruled out. This avoids the occurrence of singularities which previously had to be manually identified after processing and removed from the dataset.
- As suggested by Reviewer 1, removal of erroneous top melting data is now performed before monthly averaging, instead of afterwards. Instances of falling surface elevation are now judged due to top melting only when surface temperature exceeds a fixed threshold (-2°C), being otherwise assigned to a separate data series.

The cumulative effect of these changes does not change the qualitative properties of the IMB dataset discussed in Section 3, but it does lead to small changes in the numbers shown in Table 1. There is also no effect on the qualitative statements about model biases in Section 4, with the exception of the (small) ocean heat flux bias in the Beaufort Sea region which is eliminated by the change to latent heat calculation. This bias was only incidental to the paper's conclusions as it was small compared to the biases in top melting and conductive fluxes.

2.2 Interannual variability

As indicated in the original response to the reviews, there are not enough data points to permit a year-by-year comparison of the IMB data. Instead, the dataset was divided into three periods of roughly equal data points and distributions from each period compared. To accommodate this analysis, section 3 was divided into subsections. The previous evaluation of seasonal and spatial variability in the IMB fluxes is now in section 3.1, and the interannual variability analysis is in a new section 3.2.

2.3 Improved analysis of uncertainty

In the original version of the paper, we analysed uncertainty in the IMB fluxes due only to salinity, where salinity was assumed to be 1 ± 1 near the upper surface of the ice, and 4±4 at the base. The reviewers made a number of suggestions as to how this could be improved. Based on these suggestions, in the revision we have also examined the impacts of

- Using another widely-used scheme to estimate ice conductivity, based on Pringle (2007)
- Using different reference layers to calculate conductive fluxes and ocean heat flux
- Using different values for ice and snow density, based on Alexandrov et al (2010) and its sources Cox and Weeks (1982) and Romanov (1995)
- Using larger ranges for ice salinity (from 0 – 10).

The evaluation performed in sections 3.1 and 3.2 is that for the 'standard configuration': Maykut and Untersteiner conductivity scheme, a reference layer of .4m-.7m above the ice base to calculate conductive fluxes, and the original values for snow and ice density (330 and 917 kgm-3), but with error bars shown over all possible values of salinity (as before). In a new subsection, 3.3, we then explore the impact on these results of altering in turn the parameters above.

2.4 The relationship between conductive flux, sea ice thickness and snow depth

Both reviewers suggested that an analysis of the link between conductive flux and sea ice thickness would be of value, to judge whether the conductive flux biases of HadGEM2-ES were entirely caused by ice thickness biases. In addition, Reviewer 1 requested an analysis of the role of snow depth. Both are examined in a new subsection 4.3. IMB and model data points are now sorted into distributions based on intervals of (in turn) ice thickness and snow depth, and variation in conductive flux distributions with these variables is examined.

2.5 Internal variability

To more fully assess the impact of internal variability on the results (as requested by Reviewer 2), we evaluated the full HadGEM2-ES historical ensemble (4 members); in the original manuscript, only the first member was evaluated. The addition of the three other members did not significantly alter mean fluxes in any instance, (so the model biases were qualitatively unchanged). However, standard deviation of top melting flux did rise significantly in July and August in the North Pole region, with standard deviation twice as high in July (44.0 compared to 20.1 Wm-2) and nearly five times as high in August (53.1 compared to 11.2). No comparable effect is visible in any other variable and region, and given the positive-definite nature of the top melt flux it is likely that this effect is produced by a handful of model years (confined to ensemble members 2-4) with exceptionally high top melting in the North Pole region, thus producing a highly skewed overall distribution of fluxes.

--------------------	First ensemble member only		Whole ensemble	
Topmelt (Wm^{-2})	Mean	Std.	Mean	Std.
1	0.0	0.0	0.0	0.2
2	0.0	0.1	0.0	0.1
3	0.0	0.0	0.0	0.1
4	0.1	0.3	0.1	0.4
5	3.0	3.5	3.5	3.7
6	56.6	14	55.8	14.3
7	57.2	20.1	56.6	44.0
8	11.7	11.2	11.9	53.1
9	0.3	5.5	0.4	4.0
10	0.1	0.4	0.1	0.4
11	0.0	0.0	0.0	0.8
12	0.0	0.1	0.0	0.2

3. Complete description of changes to manuscript

In the following description, line and page numbers refer to the non-tracked changes version.

Abstract

P1, L8: The abstract has been rewritten, to make the purpose of the study clearer: to demonstrate a new method of sea ice model evaluation.

Introduction

P1, L25: The Introduction has been rewritten for similar reasons, with the citation suggested by Reviewer 1 (Kwok, 2018) added at P1, L33.

Calculating monthly mean energy fluxes from the IMBs

P3, L5: A new figure, with a diagram of the main components of an IMB, is referenced (Reviewer 2 suggestion).

P3, L12: The range of buoy lifetimes is described.

P3, L25: Justification of interpolation scheme added.

P3, L30: This paragraph (description of the process of estimating interface from surface) has been reworded for clarity, and zsfc is now defined in the text.

P4, L5: Reasoning for rarity of snow-ice formation in the IMB dataset is discussed in more detail.

P4, L15: Possible reasons for failure of temperature sensors are discussed, following communication with the data providers, as suggested by Reviewer 1.

P4, L20: We note here the sign convention, as suggested by Reviewer 2.

P4, L20: The paragraph discussing assumptions about the elevation of the top temperature measurement point has been deleted, as all such elevations are now known following communication with the data providers.

P4, L24: Incidence of surface elevation decrease accompanied by cold temperatures is now removed from the top melt dataset prior to monthly averaging, and that is made clear here.

P4, L26: The values of ice and snow density used have been moved from here to the thermodynamic parameters section, for consistency.

P4, L29: The top conductive flux section has been rewritten for clarity, with the additional of a new figure (Figure 6).

P5, L18: When the reference layer used for the basal conductive flux is defined, it is noted that we examine the sensitivity to this reference layer elsewhere in the study.

P5, L25: a reference to the new Figure 2 (IMB diagram) is added.

P6, L3: The new approach to evaluating uncertainty is outlined: the use of a standard set of parameters to calculate the main dataset, and calculation of the dataset with altered values of parameters.

P7, L5: Sources of uncertainty examined later in the study are summarised.

P7, L14: This paragraph (temperatures too warm for the assumed salinity) has been reworded for clarity. In addition, a slightly different approach is used in the revised code: when the data used to calculate a monthly flux contains temperatures too high for the assumed salinity, the highest salinity consistent with the observed temperatures is used instead.

P7, L21: figure reference changed to supplementary material

P7, L26: Sentence added discussing the possibility of interpolating over false bottoms.

Description of monthly mean fluxes from the IMBs

P7, L31: Title changed for greater contrast with the title of Section 2.

P7, L32: This section has been split into subsections, to allow for separate analysis of interannual variability, and a more comprehensive analysis of uncertainty due to parameters of the analysis.

P7, L36: In the dataset description section, there are many changes to the numbers (mostly small), related to the code changes discussed above. These are not individually listed here.

P8, L3: A reference to the strong positive skew in the top melting dataset is added.

P8, L33: New tables have been added (Tables 2 and 3) to describe the dataset in the North Pole and Beaufort Sea regions separately (Reviewer 1 suggestion).

P8, L33: The distributions are now visualised using boxplots, rather than a large number of individual points.

P8, L35: Throughout this subsection, references to the significance of differences between distributions, evaluated using Welch t-tests, are added.

P9. L36: New subsection added discussing interannual variability in the dataset.

P10, L33: New subsection added, describing the sensitivity of the derived monthly fluxes to parameters of the analysis: salinity, snow density, ice density, conductivity and reference layer chosen.

Evaluating modelled sea ice using the IMB-derived fluxes

P13, L5: Reference to the grid used by HadGEM2-ES added.

P13, L8: West et al (2018) changed to (2019), here and below.

P13, L17: Reference to significance test now used to compare modelled and observed distributions added.

P13, L21: The approach of using model cells from the entire region in the comparison is justified.

P13, L29: Discussion of significance of differences is added throughout this section.

P13, L37: Here, and in many other places, more discussion of the results is added, in recognition of Reviewer 1’s point that it was previously too descriptive. In particular, the likely role of the HadGEM2-ES meltpond scheme in the top melting biases is discussed, following the suggestion by Reviewer 1 that the role of meltponds should be examined in more detail.

P14, L34: After the evaluation of ocean heat flux, a paragraph examining the role of oceanic heat convergence is added (Reviewer 2 suggestion).

P15, L15: The model biases are compared to the dataset uncertainties evaluated in section 3.3.

P15, L33: Additional reference to the role of the HadGEM2-ES meltpond scheme inserted.

P16, L19: New subsection added examining the relationship between conductive flux, ice thickness and snow depth in model and IMB observations.

Representativity of the IMB fluxes

Conclusions

P18, L25: This section has been rewritten to include additional conclusions resulting from

a) Interannual variability analysis
b) More comprehensive evaluation of uncertainty in the IMB dataset
c) The comparison of conductive flux, ice thickness and snow depth

Appendix A

Code availability

P22, L29: References to the two GitHub repositories where the IMB analysis code is now stored have been added. However, code used in producing figures and tables for the paper is still provided as a .tar supplement (except Figures 1 and 2 which were produced in Powerpoint).

Data availability

P22, L37: References to the Zenodo repositories where the processed IMB data is now stored have been added. A clarification is made regarding the model data: the code can be made available upon request to the author.

References

New references in this revision include:

- For constants and equations used in the IMB flux calculation, particularly in the evaluation of uncertainty: Alexandrov et al 2010, Cox and Weeks 1983, Pringle et al 2006
- For links between oceanic heat convergence and ocean-to-ice heat flux, Bitz et al 2008, Keen and Blockley 2018, Schauer et al 2004, Steele et al 2010
- For recent observations of snow-ice formation, Rösel et al 2018
- For additional analysis of CMIP5 models, Shu et al 2015

Tables

Table 1 has been updated to include number of observations of each flux, as suggested by Reviewer 2. As noted above, most values have changed slightly due to the issues discussed in the code changes section.

Tables 2 and 3, detailing fluxes in the North Pole and Beaufort Sea region, are new in this revision.

Figures

Arrows have been removed from Figure 1, in line with the suggestion of Reviewer 2; we agree that these added no useful information.

Figure 2, a diagram of an IMB, has been added (Reviewer 1 suggestion). The information included in the previous Figure 5, indicating reference layers, has been added to this figure instead and is referenced from later in the study.

Figure 4 (previously Figure 3): units added

Figure 5 (previously Figure 4): units added, and line labelling improved.

A new figure 6 has been added, demonstrating the process of calculating top conductive flux in the event of the reference layer lying partially within snow and partially within ice.

The previous Figure 6 has been moved to supplementary material to help allow for 3 new figures. The edits suggested by Reviewer 2 have been implemented.

Figure 7: instead of plotting individual points, we are now showing boxplots.

Figure 8 is new for this revision, and shows interannual variability in the IMB fluxes (Section 3.2).

Figure 9 (previously Figure 8): For consistency with Figure 7, and because of high skew in some of the fluxes, boxplots are now used instead of mean and standard deviation bars.

Figure 10 is new for this revision, and shows the relationship between conductive flux, ice thickness and snow depth in IMBs and model.

The previous Figure A1 has also been moved to supplementary material.

c) Tracked changes version of manuscript

Using Arctic ice mass balance buoys for evaluation of modelled ice energy fluxes

Alex West[1], Mat Collins[2], Ed Blockley[1]

[1]Met Office Hadley Centre, FitzRoy Road, Exeter EX1 3PB
[2]Centre for Engineering, Mathematics and Physical Sciences, University of Exeter, Stocker Rd, Exeter EX4 4PY

Abstract. A new method of sea ice model evaluation is demonstrated. Data from the network of Arctic Ice Mass Balance buoys (IMBs) is used to estimate distributions of vertical energy fluxes over sea ice in two densely sampled regions – ~~(~~the North Pole and Beaufort Sea~~)~~. The resulting dataset captures ~~well~~ seasonal variability in sea ice energy fluxes well, and captures spatial variability to a lesser extent~~, well~~. The dataset is used to evaluate a coupled climate model, HadGEM2-ES, ~~in the North Pole and Beaufort Sea regions.~~ in the two regions. The evaluation shows HadGEM2-ES to simulate~~model~~ too much top melting in summer and too much basal conduction in winter. These results are consistent with a previous study of sea ice state and surface radiation in this model, increasing confidence in the IMB-based evaluation. In addition, the IMB-based evaluation suggests an additional important cause for excessive winter ice growth in HadGEM2-ES, ~~(~~lack of sea ice heat capacity, ~~)~~ which was not detectable in the earlier study.

Uncertainty in the IMB fluxes caused by imperfect knowledge of ice salinity, snow density, and other physical constants is quantified (as is inaccuracy due to imperfect sampling of ice thickness) and in most cases is found to be small relative to the model biases discussed. Hence the IMB-based evaluation is shown to be a valuable tool with which to analyse sea ice models, and by extension better understand the large spread in coupled model simulations of present-day ice state. Reducing this spread is a key task both in understanding the current rapid decline in Arctic sea ice, and in constraining projections of future Arctic sea ice change.

~~**Abstract.** Arctic sea ice has declined rapidly over recent decades. Models predict that the Arctic will be nearly ice-free by mid-century, but the spread in predictions of sea ice extent is currently large. The reasons for this spread are poorly understood, partly due to a lack of observations with which the processes by which Arctic atmospheric and oceanic forcing affect sea ice state can be examined. In this study, a method of estimating fluxes of top melt, top conduction, basal conduction and ocean heat flux from Arctic ice mass balance buoy elevation and temperature data is presented. The derived fluxes are used to evaluate modelled fluxes from the coupled climate model HadGEM2-ES in two densely sampled regions of the Arctic, the North Pole and Beaufort Sea. The evaluation shows the model to overestimate the magnitude of summer top melting fluxes, and winter conductive fluxes, results which are physically consistent with an independent sea ice and surface energy evaluation of the same model.~~

1. Introduction

Evaluation of sea ice simulations using metrics based on sea ice extent (e.g. Stroeve et al, 2012; Wang and Overland, 2012) is known to be an imperfect method of assessing models (Notz, 2015). This is partly because of the very high interannual variability of sea ice extent (Swart et al, 2015), but also because it does not address the accuracy of the many variables influencing sea ice extent, in which compensating errors may be present. Sea ice volume, evaluated ~~in~~for CMIP5 by Stroeve et al. (2014) and Shu et al. (2015), is less sensitive to internal variability (Olonscheck and Notz 2018) but is also driven by multiple complex processes and so is equally susceptible to compensating errors. These issues hinder understanding of the very large spread in modelled present-day sea ice simulation. In turn, this increases the uncertainty in future projections of Arctic sea ice, which has declined rapidly over the past 30 years both in extent and volume (Lindsay and Schweiger, 2015; Kwok, 2018). In this study, we present a new, complementary method of evaluating sea ice simulation, motivated by the following reasoning.

The ~~-~~proximate driver of sea ice volume is sea ice mass balance. In turn, sea ice mass balance is driven by energy balance at the upper and lower surfaces of the snow-ice column. Energy balance is driven partly by external factors in the atmosphere (radiative fluxes, upper air temperature) and ocean (ocean heat flux) but also by sea ice thermodynamics (temperature, albedo and conduction). Finally, the sea ice thermodynamics, in turn, are partially driven by the sea ice state itself (area and thickness), closing the two causal chains known as the thickness-growth feedback and the surface albedo feedback (Figure 1). Ideally, for any model, the entire sea ice causal chain would be evaluated, along with the external drivers, greatly increasing understanding of why a particular sea ice state i~~wa~~s modelled.

Large-scale evaluation of ice mass balance, ice thermodynamics, and energy balance at the lower surface of the ice, has not to date been performed for any model. This is largely because these quantities cannot yet be measured remotely~~,~~ but must instead be measured in-situ using systems of instruments frozen into the ice. In particular, a device called an Ice Mass Balance buoy (IMB) measures mass balance at the upper and lower surfaces of the snow-ice column, and temperature profiles within the ice, at simultaneous locations (Perovich and Richter-Menge, 2006). Data from individual IMBs has been used to estimate sea ice energy fluxes such as conduction and ocean heat flux in the past (e.g. Perovich et al, 2002; Lei et al, 2014; Lei et al, 2018). However, IMB data has not yet been used to directly evaluate a climate model, due to the large disparity in the relevant spatial and temporal scales.

In this study, we use data from the whole IMB network to perform a large-scale evaluation of sea ice mass balance and thermodynamics in a coupled climate model. Monthly mean fluxes of top melt, top conduction, basal conduction and ocean heat flux are calculated from temperature and elevation data obtained from 104 IMBs released between 1993 and 2015: the resulting observational dataset, and the code used in its production, are published alongside this study, with references given in the code availability and data availability sections below. This dataset is then used to evaluate the sea ice in a coupled climate model (HadGEM2-ES, part of the CMIP5 ensemble) in two densely-~~ ~~sampled regions of the Arctic, the North Pole and the Beaufort Sea. Modelled and IMB-measured fluxes are restricted to each region in turn: distributions of fluxes in each month are compared, and likely model biases identified. The results of the IMB-based evaluation are compared with a previous

Formatted: Font: (Default) Times New Roman, 10 pt
Formatted: Normal, No bullets or numbering
Formatted: Font: (Default) Times New Roman, 10 pt
Formatted: Font: (Default) Times New Roman, 10 pt
Formatted: Font: (Default) Times New Roman, 10 pt
Formatted: Font: (Default) Times New Roman, 10 pt
Formatted: Font: (Default) Times New Roman, 10 pt
Formatted: Font: (Default) Times New Roman, 10 pt
Formatted: Font: (Default) Times New Roman, 10 pt
Formatted: Font: (Default) Times New Roman, 10 pt
Formatted: Font: (Default) Times New Roman, 10 pt
Formatted: Font: (Default) Times New Roman, 10 pt
Formatted: Font: (Default) Times New Roman, 10 pt
Formatted: Font: (Default) Times New Roman, 10 pt
Formatted: Font: (Default) Times New Roman, 10 pt

evaluation of the sea ice state and surface radiation in the same model (West et al., 2019): the results are found both to be consistent with the results of this study, and to enhance understanding of the first evaluation.

The manuscript is structured as follows. In section 2, the IMB data, and the process by which vertical energy fluxes are calculated from the data, are described. In section 3, the IMB flux dataset is described: seasonal, spatial and interannual variability is discussed, and uncertainty in the IMB fluxes due to various parameters used in the analysis is examined. In section 4, the data are used to evaluate HadGEM2-ES, and the results are interpreted in the context of West et al. (2019). In section 5, the representativity of the IMB fluxes is discussed. In section 6, conclusions are presented.

~~The climate of the Arctic Ocean region is characterised by the presence of a semi-permanent sea ice cover that acts to reflect solar energy in summer and insulate ocean heat in winter. Arctic sea ice has changed rapidly over the past 30 years, with a decline in extent of 0.84 x 10^6 km^2 per decade observed during the month of September according to the HadISST1.2 dataset (Rayner et al., 2003), and evidence of concurrently declining ice volume (e.g. Rothrock et al., 2008, Lindsay and Schweiger, 2015). Model projections from CMIP5 suggest that the Arctic Ocean is likely to become nearly ice-free in September by mid-century (Massonnet et al, 2012), an event that would have serious implications for the climate (through decreased surface albedo and decreased insulation between the ocean and atmosphere) and for geopolitics. However, these model projections vary greatly in their simulation of the mean state of Arctic sea ice in the present day, as well as the speed of sea ice decline.~~

~~Evaluation of model simulations using metrics based on sea ice extent (e.g. Stroeve et al, 2012; Wang and Overland, 2012) is known to be an imperfect method of assessing projections (Notz, 2015). This is partly because of the very high inter-annual variability of sea ice extent (Swart et al, 2015), but also because it does not address the accuracy of the many variables influencing sea ice extent, in which compensating errors may be present. A more fundamental measure of the sea ice state is the sea ice volume, evaluated in CMIP5 by Stroeve et al, 2014 and Shu et al., 2015. However, there remain many variables influencing sea ice volume, and hence this metric is equally susceptible to the problem of compensating errors (although the effect of internal variability on sea ice volume is now quite well-understood, e.g. Olonscheck and Notz, 2018). Clearly, a logical next step would be to evaluate the primary driver of sea ice volume, i.e. sea ice mass balance. However, although detailed studies of processes driving Arctic sea ice volume have been undertaken (e.g. Lei et al, 2018), in general these are on too small a scale to directly evaluate model processes.~~

~~Many of the processes driving sea ice volume are, to first order, well understood; sea ice volume is the area integral of sea ice thickness, which is driven by the surface and basal mass budget, and hence energy budget. The surface energy budget is driven partly by variables external to the sea ice, such as fluxes of downwelling shortwave (SW) and longwave (LW) radiation, and air temperature. However it is also influenced by properties of the sea ice state, such as surface temperature, albedo and the top conductive flux. Similarly, the basal energy budget is driven partly by one particular external variable (the ocean heat flux), but also by a property of the sea ice (the basal conductive flux). In turn, these properties of the ice surface and base are determined in the main by ice thermodynamics, which is strongly influenced by ice thickness; the surface albedo is also affected by the ice thickness directly. This closed causal chain (Figure 1) gives rise to the thickness-growth feedback, whereby~~

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

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

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

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

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

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

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

~~thinner ice grows more quickly in winter, and the surface albedo feedback, whereby thinner ice melts more quickly in summer.~~

~~Sea ice state is determined by the response of the feedback loops to the external variables (downwelling radiation, air temperature, ocean heat flux), which are in turn influenced by the sea ice state over longer timescales. It is difficult to evaluate the simulation of these processes in coupled models due to a lack of large-scale observational datasets. CMIP5 surface radiative fluxes were evaluated by Boeke and Taylor (2018), but the evaluation was unable to separate out flux variability directly caused by sea ice variability. West et al (2019, in discussion) proposed a framework by which external drivers of sea ice mass balance could be separated from feedbacks of the sea ice state given available observations, but the applicability of this framework is limited by observational uncertainty. Hence there remain significant shortcomings in ability to evaluate modelled sea ice processes. In this study, a method is presented of evaluating sea ice mass balance, and sea ice thermodynamics, using data from Arctic ice mass balance buoys, devices that measure ice surface and base elevation and internal ice temperatures.~~

~~An observational dataset of sea ice vertical energy fluxes is created by estimating monthly mean fluxes of top melt, top conduction, basal conduction and ocean heat flux for the entire IMB network. The extent to which this dataset is representative of wider regions is investigated by comparing the ice thickness distribution sampled by the buoys to that measured by submarine sonar. Evaluation of the coupled model HadGEM2-ES (part of the CMIP5 ensemble), is then carried out using this dataset. This is done by restricting model data to the densely sampled regions of the North Pole and Beaufort Sea, transforming modelled fluxes to be over ice only, and comparing the distribution of model fluxes obtained for each region and month to that measured from the IMB network. By combining this evaluation with a previous evaluation of the sea ice state and surface radiation in the same model (West et al., 2019, in discussion), a detailed picture of the sea ice simulation emerges, with drivers of model biases easily identifiable.~~

~~The study is structured as follows. In section 2, the IMB data, and the process by which vertical energy fluxes are calculated from the data, are described. In section 3, the resulting dataset is presented, and seasonal and spatial variability is discussed. In section 4, the data is used to evaluate HadGEM2-ES, and the results are interpreted in the context of West et al (2019). In section 5, the representiveness of the IMB fluxes is discussed. In section 6, conclusions are presented.~~

2. Calculating monthly-mean energy fluxes from the IMBs

The ice mass balance buoy (IMB; Perovich and Richter-Menge, 2006) is a system of instruments frozen into a sea ice floe, allowing the simultaneous measurement of surface and base elevation, internal ice temperature (usually at 10cm resolution), and position; many also measure surface air pressure and temperature. A diagram of an IMB is shown in Figure 2. An IMB provides, by design, measurements of sea ice thickness, and of surface and basal mass balance, via the measurements of surface and base elevation. Fluxes of conduction can also be estimated from the ice temperature data (e.g. Perovich and Elder, 2002), although uncertainty is considerable due to lack of knowledge of ice salinity. In particular, the thermodynamics and basal elevation measurements can be combined to estimate ocean heat flux (Lei et al., 2014).

Data from the 104 IMBs deployed by the Cold Regions Research and Engineering Laboratory (CRREL) are stored in a series of comma-delimited CSV files at http://imb.erdc.dren.mil/buoysum.htm. The buoys were deployed between 1993-2017; spatial coverage is mainly in the North Pole and Beaufort Sea regions (Figure 3~~2~~). The buoys are identified by the year of deployment followed by a letter, for example '2012L'. Buoy lifetimes range from 4 days (2015C) to 20 months (2006C) with an interquartile range of 4-11 months. ~~The buoys are identified by the year of deployment followed by a letter, for example '2012L'.~~ All buoys report time series of ice base elevation, snow/ice surface elevation, latitude, longitude, as well as a collection of ice temperature time series taken at a number of vertical positions above, within and below the ice. In general, temperature profiles are reported at very high temporal resolution, hourly or bi-hourly, and tend to be noisy, with much high-frequency variability. From 2006 onwards, elevation data are reported at similarly high resolution, but before 2006 are reported much less frequently, with intervals of a week or more between measurements.

As most analysis of the data depends on the ability to perform arithmetic operations on different series, it was necessary to produce data series at consistent points in time for each buoy. To this end, modified elevation data series were produced at times coincident with the temperature measurements, using either interpolation (where there were fewer than 3 measurements in the 2-day period centred on the time in question) or a binomially-weighted mean (where there were 3 or more measurements in this period). This regularisation process is illustrated in Figure 4~~3~~. Although a more advanced optimal interpolation scheme would likely produce more accurate time series, inspection of individual data series shows~~ed~~ that the current scheme produces~~d~~ data that is sufficiently realistic for the purposes of this study. For example, linear interpolation produces unrealistic sharp changes in the time derivative of elevation, but the effect of these on monthly mean elevation change, the derived variable used in this study, is likely to be very small.~~.~~

The set of elevation measurements provided also varies between buoys, necessitating some processing before full regular time series of surface elevation, snow thickness, interface elevation, ice thickness and base elevation can be obtained. Some later buoys do not report surface elevation directly, but report snow-ice interface elevation and snow depth, which must be summed to obtain the surface elevation. A more difficult problem is presented by ~~T~~the earlier buoys, which tend to produce data of surface and base elevation only. Snow-ice interface elevation must therefore be deduced from surface and base elevation, by a process illustrated in Figure 5.~~From these, it is usually possible to deduce snow-ice interface elevation, as by design the snow-ice interface is always at position z=0 at the time of deployment. The process is illustrated in Figure 4:~~ ~~I~~iterating through the times of observation $t_1,\ldots,t_n$, the interface elevation $z_{\text{int}}(t_1)=0m$ by construction, as the thermistor string is always referenced to the snow-ice interface at the time of deployment. At time t_i, if $z_{\text{int}}(t_{i-1})\le z_{sfc}(t_i)$, where $z_{sfc}(t_i)$ represents surface elevation of the snow-ice column, we set $z_{\text{int}}(t_i)=z_{\text{int}}(t_{i-1})$; but if $z_{\text{int}}(t_{i-1})>z_{sfc}(t_i)$ we set $z_{\text{int}}(t_i)=z_{sfc}(t_i)$. In this way, the interface elevation changes only when top melting of ice is detected, i.e. when the surface elevation is judged to fall below the interface elevation estimated for the previous time of observation.

This method would fail in the presence of ice flooding and snow-ice formation (e.g. as documented by Provost et al, 2017). However, while snow-ice formation is known to occur in some areas sampled by the IMBs (particularly in the North Pole region, e.g. Rösel et al, 2018), it is almost certainly a rare event in the IMB dataset. This is because the snow layer is almost always sufficiently thin relative to the ice layer that snow-ice formation is unlikely from hydrostatic principles. There are four instances when snow depth becomes sufficiently large that snow-ice formation is a possibility, but these are always associated with failure of other sensors, such that the associated data does not reach the final dataset produced in this study.~~,~~

~~but there are no clear signs of such events taking place in any of the IMB records examined in this study. In addition some later buoys do not report surface elevation directly, but report snow-ice interface elevation and snow depth, which must be summed to obtain the surface elevation. Hence full regular time series of surface elevation, snow thickness, interface elevation, ice thickness and base elevation can be obtained for most buoys.~~

Processing the temperature data is also necessary. Instances of air, ice or ocean temperature data that are obviously wrong occur very frequently, usually characterised by sudden step changes in the temperature measurements at single, or multiple layers, that are inconsistent with simultaneous measurements in other layers, often to physically unrealistic values. The incorrect values can be caused by failure of the sensors or the datalogger, or by an inability to communicate data to the receiving satellite (Donald K. Perovich, personal communication). In most cases, wrong values occurred in large groups that were difficult to identify with automatic data processing, and therefore had to be identified by inspection and removed.

~~For the early buoys, the depths at which temperature measurements are taken are clearly labelled, tending to begin at 60cm or 70cm above the surface, occurring at resolutions of 5cm above the surface and 10cm below. For later buoys, however, the measurement depths are not labelled. It was inferred that for all such buoys, the measurements began at 60cm above the ice surface, descending at a resolution of 10cm. A curious effect occurs with the long-lived buoy 2006C, in which the position of the thermistor string appears to materially change a year after deployment, as the elevation data becomes abruptly dislocated from the temperature contours normally associated with the presence of sea ice and snow. The change occurred soon after the buoy had reported an exceptionally high basal melting flux during the summer of 2007, as reported by Perovich et al (2008). In this case, the instance of the change was estimated, and all temperature data after this point translated downwards by 70cm.~~

~~With~~From the processed temperature and elevation data, monthly mean fluxes of top melt, top conduction, basal conduction and ocean heat flux were produced in the following way. Throughout this study, the sign convention is that a positive value denotes a downwards flux, and vice versa.

Top melting of ice and/or snow. This flux, commonly reported by models, represents the total energy gain by sea ice (snow) in a grid cell over the course of a month associated with melting of ice (snow) at the upper surface. It is estimated from the IMBs using the surface elevation series. A change between two adjacent daily data points in surface elevation is judged due to top melting if and only if the change is negative, and the surface temperature is above a threshold value (-2°C). The energy gain associated with the melting is calculated~~.~~ by multiplying the elevation change by ice or snow density, depending on whether the snow depth is nonzero, and by specific latent heat of fusion of ice (all parameters are defined below)~~using ice density of 917 kgm-3, snow density of 330kgm-3 and latent heat of melting of 3.34 x 10^5 Jkg-1, the standard values used by the sea ice model CICE (Hunke et al, 2013)~~. The daily top melt estimates are then averaged to obtain monthly mean top melt.

Top conductive flux. This flux is defined as the conduction from the snow/ice surface into the ice interior. In this study it is calculated using temperatures in the top 50cm of the snow-ice column. Where this layer lies entirely within snow (ice) the conductive flux is calculated as the temperature gradient across the layer, determined by a linear fit, by snow (ice) conductivity: values of snow and ice conductivity used are defined below.

In many cases, however, the top 50cm is located partly within snow and partly within ice. Because snow conductivity tends to be much lower than ice conductivity, the snow-ice interface is usually associated with a sharp change in gradient that renders a linear fit meaningless. In these cases, the top conductive flux i~~wa~~s determined by a linear fit through the same layer, using an 'adjusted' temperature profile: ~~Its calculation from the IMB data is made difficult by large step changes in temperature gradient associated with the snow-ice interface, which is usually located within 50cm of the upper surface. To calculate top conductive flux, adjusted temperature profiles were calculated for each point in time, in which for each point z above the snow-ice interface, the adjusted temperature can be written as~~

$$\cancel{T_{adj} = \mu T + (1-\mu)T_{\mathrm{int}}}$$

$$T_{adj}(z) = \begin{cases} \mu T(z) + (1-\mu)T_{int-r} & z > z_{int} \\ T(z) & z \le z_{int} \end{cases}$$

w~~W~~here z_{int} is the elevation of the snow-ice interface, $T_{int-ref}$ ~~T_{int}~~ is temperature 5cm below the interface~~snow-ice interface temperature~~, and $\mu = k_{ice}/k_{snow}$ where k_{ice} and k_{snow} are ice and snow conductivity respectively~~; their calculation is described below~~. Physically, T_{adj} represents the temperature profile that the snow-ice column would have, if the snow was converted to ice, $T_{int-ref}$ remained the same, and the vertical conductive fluxes remained the same. The effect of the adjustment is to 'straighten' the profile by rotating the profile section located in the snow about $T_{int-ref}$, by a factor determined by the ratio of conductivities μ. A linear fit i~~wa~~s then taken through a layer 0-50cm below the snow surface, and multiplied by k_{ice} to produce estimates of instantaneous top conductive flux. These a~~we~~re then averaged to obtain monthly means. The process is illustrated in Figure 6.

Basal conductive flux. This flux is defined as the conduction from the ice base into the ice interior. As an important ~~it is a vital~~ component of the energy balance at the ice base it has frequently been estimated from individual buoys in ocean heat flux calculations. Typically, temperature gradients at the ice base are small due to higher salinities here (e.g. Schwarzacher, 1959), with correspondingly higher heat capacities and lower conductivities; hence previous studies have commonly used a reference layer of a fixed thickness above which the basal conduction is estimated. In this study we use the approach of Lei et al. (2014), and calculate the basal conduction by taking temperature gradients across a layer 40cm-70cm above the ice base, illustrated in Figure 2. In section 3.3 we examine the sensitivity of the derived fluxes to changes in the elevation of this reference layer, amongst other parameters. As above, the instantaneous values were averaged to a monthly mean.

Ocean heat flux. This flux is defined as the diffusive heat flux arriving at the ice base from the ocean beneath. In theory, it can be calculated as the residual of the basal conductive flux and the latent heat of melting/freezing at the ice base. However, using the basal conductive flux as defined above it is necessary also to take into account the sensible heat uptake of the intervening layer (the 'buffer zone'), 0-40cm above the ice base, illustrated in Figure 2~~5~~. The ocean heat flux can then be written as

$$F_{ocn} = F_{condbot} - F_{sens} - F_{lat}$$

as in Lei et al (2014).

The basal conductive flux $F_{condbot}$ is defined as above. Monthly mean F_{sens}, the sensible heat flux in the 0-40cm layer, is calculated as the average of daily heat uptake rates obtained by taking linear fits through all temperature points within 1 day of a given time instant for all vertical points in this layer, summing these (weighted according to layer thickness), and multiplying by ice density and heat capacity, defined below. Finally, monthly mean latent heat of melting at the ice base, F_{lat}, is calculated from the base elevation time series, by multiplying daily differences in elevation by specific latent heat of fusion.

The calculation of thermodynamic parameters is now described. In this study, we take the approach of using a 'standard' set of thermodynamic parameters to calculate the main dataset of energy fluxes, demonstrated in sections 3.1 and 3.2 below, and subsequently evaluate sensitivity to the values of these parameters in section 3.3. Ice density ρ_{ice}, snow density ρ_{snow} and latent heat of melting q_{fus} are set to 917 k~~K~~kgm^{-2}, 330 k~~K~~kgm^{-2} and 3.34 x 105 Jkg^{-1} respectively, the standard values used by the sea ice model CICE (Hunke et al, 2013).

Ice conductivity is defined after Maykut and Untersteiner (1971) as

$$k_{ice} = k_{fresh} + \frac{\beta S}{T}$$

where S and T are ice salinity and temperature respectively, $k_{fresh} = 2.03 W m^{-1} K^{-1}$, the conductivity of fresh ice, and $\beta = 0.13 W m^{-1}$ is an empirically determined constant representing the effect of brine pockets on conductivity. For the calculation of the top conductive flux, a practical salinity of 1.0 is used, while the temperature used is that of the snow-ice interface. For the calculation of the basal conductive flux, a practical salinity of 4.0 is used, multiplied by the mean value of $1/T$, where the average is taken over the time period in question and the layer 40-70cm above the ice base.~~,~~

Specific heat capacity is defined after Ono (1967) as

$$c_{ice} = c_{fresh} + \frac{q_{fresh}\mu S}{T^2}$$

where $c_{fresh} = 2106 Jkg^{-1}K^{-1}$ is the specific heat capacity of fresh ice, $q_{fresh} = 3.34 \times 10^5 Jkg^{-1}$ the specific latent heat of fusion of fresh ice, and $\mu = 0.054K$ the ratio between water salinity and freezing temperature. In calculating sensible heat uptake at the ice base, again a practical salinity of 4.0 is used, multiplied by the mean value of $1/T^2$, where the average is taken over the time period in question and the layer 0-40cm above the ice base.

Ice salinity must also be taken into account when calculating latent heat of freezing and melting. The energy required to melt a given volume of sea ice at temperature *T*, from Bitz and Lipscomb (1999) is

$$q(S,T) = \rho c_0 (T_m - T) + \rho q_{fresh}\left(1 + \frac{\mu S}{T}\right).$$

At the lower surface of the ice, q is calculated by setting $T = -1.8°C$ and $S = 4.0$ as above. At the upper surface of the ice, *T* is usually extremely close to 0°C when melting is taking place, meaning that a choice of *S* that is both consistent and physically realistic in all cases is difficult to make. Instead, it is assumed that the ice at the upper surface is fresh, and $q = q_{fresh}$ is used.

The monthly heat fluxes calculated above are subject to several sources of uncertainty. These are evaluated in detail in Section 3.3 below, but the issues are briefly summarised here. Firstly, there is significant uncertainty due to lack of knowledge of ice salinity, which affects the fluxes through the ice conductivity and heat capacity~~two principal sources of uncertainty: measurement uncertainty, and uncertainty in the value of *S*~~. Secondly, the manner of dependence of ice conductivity on salinity is also subject to uncertainty, with an alternative formulation to Maykut and Untersteiner being proposed by Pringle (2006). Thirdly, both snow and ice density are subject to uncertainty, affecting the diagnosis of melting and freezing fluxes at the top and basal surfaces of the ice from elevation changes (as well as sensible heat uptake in the lowest layer of the ice). Finally, the reference layers chosen to evaluate conductive and heat uptake fluxes are themselves a parameter of the analysis, and as such represent an additional source of uncertainty.

~~To estimate the former, we use elevation and temperature measurement uncertainties as described by Perovich and Elder (2002). To estimate the latter, we use an uncertainty of ±1.0 in calculating top conductive fluxes, and an uncertainty of ±4.0 in calculating basal conductive flux, and sensible and latent heat uptake at the ice base. The two uncertainties are added together and combined with the flux central estimates, as defined above, to produce a range of possible monthly mean fluxes for each data point.~~

Examination of the monthly mean energy fluxes reveals several ways in which unrealistic estimates might be produced. Firstly, ~~I~~in a small minority of months top or basal ice temperature is warmer than the melting point associated with the assumed salinity (1 at the top of the ice and 4 at the base) resulting in the conduction or sensible heat uptake being very large or undefined. For these months, the salinity i~~wa~~s set instead to the highest physically allowable value, given the maximum temperature attained.~~a small minority of cases, the temperature at the top (bottom) ice surface is inconsistent with the assumption of a salinity range of ±1 (±4), resulting in the observed temperature being above the assumed melting point, and the measured flux being unrealistically large, or in a few cases undefined. Clearly, a more comprehensive analysis would allow the selection of varying salinity ranges depending on the information available from each buoy. As this is outside the scope of this study, the fluxes affected by this problem were removed from the dataset.~~

A second problem relates to the formation of false bottoms under sea ice, as documented by Notz (2003), in which meltwater refreezes upon meeting cold seawater at a temperature below its own melting point. This process visibly occurs during the period of operation of some buoys (for example 2015A, demonstrated in Figure S1~~6~~), associated with sudden step changes in base elevation. These result in very large negative monthly mean ocean heat fluxes being calculated during the month of formation, and correspondingly large positive fluxes during the month of dissipation. These fluxes are physically unrealistic, as the large changes in elevation usually represent the freezing and melting of only a very thin layer of ice, with liquid seawater remaining in between this layer and the main body of the ice column. In some cases, it may be possible to estimate true ocean-to-ice heat flux simply by interpolating base elevation between the apparent times of formation and dissipation, but this approach is likely to be inaccurate for long-lived false bottoms. For the purposes of this study all affected ocean heat fluxes were simply removed from the dataset, as they were relatively few in number.~~Hence it was necessary to inspect each buoy individually for possible false bottom formation, and remove all affected ocean heat fluxes from the dataset.~~

~~Unrealistic positive values of top melting occur quite frequently during the winter months, due to decrease of surface elevation likely associated with wind drifting of snow, although these are usually an order of magnitude lower than those observed during the summer months. As a simple way of removing these fluxes, top melt fluxes for all months during which monthly mean surface temperature was measured at -5°C or lower were set to zero. It is noted that during the late spring, this could result in valid top melt fluxes being ignored in the case of rapid rises in temperature during the latter stages of a month.~~

~~Unrealistically high negative values of basal conductive flux are also derived during months when the ice was quite thin, and the layer 40cm-70cm above the ice base was correspondingly close to the top surface of the ice. In order to account for this situation, basal conductive fluxes (and oceanic heat fluxes) for all months during which monthly mean ice thickness was 1m or lower were removed from the dataset.~~

3. Description of monthly mean~~Deriving monthly-mean~~ flux distributions from the IMBs

3.1 Seasonal and spatial variability

Formatted: Indent: Left: 1.27 cm, No bullets or numbering

Throughout the description of the IMB-estimated fluxes here, and the model evaluation below, the convention used is that positive numbers denote downwards fluxes, and vice versa. The distributions of monthly mean fluxes of top melting, top conduction, basal conduction and ocean heat flux are summarised in Table 1. The IMBs provide

463~~90~~ monthly mean values of top melt in total, ranging from 31 values in March and August to 53~~7~~ in May. The seasonal cycle reaches its maximum in July, when top melting of 29.9~~8.6~~ ± 17.8~~6.7~~ Wm^{-2} is observed. Strong top melting is also evident in June (16.8~~1~~ ± 11.0~~0.1~~ Wm^{-2}), but top melting tends to be considerably lower in August (8.1~~7.8~~ ± 6.74 Wm^{-2}). In all three summer months, the distribution is positively skewed, with a small number of very high values (for example, the highest top melt value recorded is 79.9 Wm^{-2}, for the buoy 1993A in July 1993). Values for the rest of the year are zero or near-zero. Throughout the year, standard deviation of the distributions is of a similar order of magnitude to the mean, showing a high degree of spatial and inter-annual variability.

Top conductive flux, a component of the surface energy balance, is the means by which the ice loses energy to the atmosphere in the presence of atmospheric cooling during the Arctic winter. It depends strongly upon atmospheric conditions, but also upon ice and snow thickness, as thinner ice and snow can support stronger temperature gradients and conduct energy upwards more quickly. For the top conductive fluxes, the IMBs provide 414~~501~~ estimates in total, ranging from 24~~30~~ in August to 51~~9~~ in May. Mean top conductive fluxes are strongly negative from October-March, reaching a minimum value of -17.6~~22.3~~ ± 6.8~~13.2~~ Wm^{-2} in -December~~December~~. However, values are weakly positive in June and July, reflecting warming of the ice interior.

The basal conductive flux acts to remove energy from the ice base in winter, allowing ice growth, and to a lesser extent during late spring and early summer while the ice is warming, attenuating ice melt. For the basal conductive fluxes the IMBs provide 46~~7~~3 estimates, ranging from 2~~7~~9 in August to 542 in May. The basal conductive flux displays a seasonal cycle less amplified than, and displaced slightly later relative to, that of the top conductive flux, with lowest values occurring from November-April and a minimum of -14.0~~3.6~~ ± 5.7~~6.4~~ Wm^{-2} occurring in Januar~~Februar~~y. The damped response relative to the top conductive flux occurs due to the thermal inertia of sea ice, and the principal thermodynamic forcing occurring at the top surface.

Lastly, for the ocean heat fluxes the IMBs provide 414~~36~~ estimates, ranging from 25~~6~~ in August to 4~~8~~9 in May. The highest values are seen in July and August, with a mean and spread of 18.1~~25.1~~ ± 15.3~~28.7~~ Wm^{-2}, and 19.2 ± 23.9 Wm^{-2} respectively. The distributions in these months are, like the top melting flux,~~ is also~~ strongly positively skewed~~ in this month~~, with a small number of exceptionally high values~~,~~. N~~n~~otably, 119~~43~~ Wm^{-2} is~~being~~ estimated in August 2007 for the buoy 2006C in the Beaufort Sea, as part of a summer of extreme ice melt documented by Perovich et al. (2008). In the winter, mean values of ocean heat flux are near-zero. There is frequent occurrence of small negative estimates in the distributions in the winter. These are likely to be spurious and reflect errors in assumptions made about the salinity and density at the base of the ice.~~, as~~ F~~f~~or most such values, the uncertainty interval resulting from varying the salinity from 0 to 10~~s~~ encompasses 0 Wm^{-2}.

Two regions of the Arctic are relatively densely sampled by the IMBs: the Beaufort Sea and the North Pole (Figure 3~~2~~). In order to demonstrate that the IMBs are able to capture some regional variability, and especially to aid with model evaluation in Section 4 below, monthly mean fluxes derived from buoy tracks entirely within these regions were sorted into separate datasets, characteristics of which are now described separately. Mean and standard deviations of the distributions in the North Pole and Beaufort Sea regions are summarised in Tables 2 and 3

Formatted: Superscript

Formatted: Superscript

respectively; boxplots are presented in Figure 7. Significance of differences between distributions is measured using a Welch t-test, with a 5% p-value threshold.

Top melting fluxes are shown in Figure 7~~7~~a separately for the Beaufort Sea and the North Pole regions. In June, the top melting fluxes measured in the North Pole region range from 1 to -37~~2~~ Wm^{-2}, with a mean of 12~~1~~ ± 8~~7~~ Wm^{-2}, while those measured in the Beaufort Sea range from 10~~4-~~ to 52~~48~~ Wm^{-2} with a mean of 26~~2~~ ± 1~~1~~0 Wm^{-2}. The lower distribution in the North Pole region is consistent with the observed later onset of surface melting here (Markus et al., 2009) associated with the higher latitude. In July, measured fluxes range from ~~2~~2 to -55~~48~~ Wm^{-2} in the North Pole region, with a mean of 23~~2~~ ± 14~~2~~ Wm^{-2}, and 11 to -8~~7~~0 Wm^{-2} in the Beaufort Sea region, with a mean of 41~~0~~ ± 17~~6~~ Wm^{-2}. In both June and July, distributions of top melt fluxes are significantly different in the two regions. Measured fluxes of top melting are much lower in August in both regions.

For the top conductive flux (Figure 7~~7~~b),~~-~~ winter fluxes tend to be slightly higher in magnitude in the North Pole than in the Beaufort Sea region, although in no winter months are the distributions significantly different at the 5% level~~winter variability is somewhat higher over the Beaufort Sea region than in the North Pole region~~. In January, for example~~December~~, North Pole fluxes range from -32~~45~~ to -10~~9~~ Wm^{-2} with a mean of -18~~20~~ ± 7~~10~~ Wm^{-2}, while those in the Beaufort Sea region range from -20~~57~~ to -7~~8~~ Wm^{-2} with a mean of -15~~28~~ ± 7~~14~~ Wm^{-2}. Some notable differences between the distributions occur in the 'shoulder seasons', particularly in May and~~-~~ August (when the distributions are significantly different)~~September~~, with higher values~~–~~, ~~(~~indicating ice warming~~)~~, occurring in the Beaufort Sea region. For example, in May, values in the North Pole region range from -6~~8~~ to 3~~5~~ Wm^{-2} with a mean of -1~~2~~ ± 2~~3~~ Wm^{-2}, while values in the Beaufort Sea region range from -2 to 4~~15~~ Wm^{-2} with a mean of 1~~2~~ ± 24 Wm^{-2}. These differences indicate earlier onset of warming in the Beaufort Sea and earlier onset of cooling in the North Pole region, consistent with an earlier onset of surface melt in the Beaufort Sea.

Less spatial variability is evident for the mean basal conductive flux (Figure 7~~7~~c). For example, in December, North Pole fluxes range from -20 to -7 Wm^{-2} with a mean of -13 ± 3 Wm^{-2}, while Beaufort Sea fluxes range from -32 to -1 Wm^{-2} with a mean of -14 ± 7 Wm^{-2}. Hence the thermal inertia of ice appears to have some damping effect on the larger variability of thermal forcing evident in the Beaufort Sea region from the top conductive flux. Winter variability tends to be higher in the Beaufort Sea than the North Pole, but this is largely caused by a small number of exceptionally low fluxes early in the winter associated with end-of-summer ice thicknesses of 50cm or lower, notably a value of -61.7 Wm-2 recorded in October 2007 for the buoy 2006C.~~. For example, in December, North Pole fluxes range from -20 to 0 Wm^-2^ with a mean of -12 ± 5 Wm^-2^, while Beaufort Sea fluxes range from -22 to -2 Wm^-2^ with a mean of -12 ± 6 Wm^-2^. Hence the thermal inertia of ice appears to have some damping effect on the larger variability of thermal forcing evident in the Beaufort Sea region from the top conductive flux.~~ The faster warming and slower cooling of ice evident in the shoulder seasons in the Beaufort Sea region for the top conductive flux is also not evident for the basal conductive flux. In the month of May, for example, basal conductive flux values range from -13 to 0 Wm^{-2} in the North Pole region with a mean of -7 ± 3 Wm^{-2}, compared to a range of -8 to -~~3~~2 Wm^{-2} and a mean of -5 ± 1 Wm^{-2} in the Beaufort Sea region.

For the ocean heat flux (Figure 7~~7~~d), in the summer very high values tend to be more common in the Beaufort Sea region than in the North Pole region. For example, in August North Pole region values range from 2~~3~~ to 38~~64~~

Wm^{-2} with a mean of 1~~9~~3 ± 10~~6~~ Wm^{-2}, while the Beaufort Sea region values range from 7 to~~-~~ 119~~43~~ Wm^{-2} with a mean of 33~~41~~ ± 35~~48~~ Wm^{-2}. It is likely that these are related to the lower ice fractions, and greater solar heating of the mixed layer, in the Beaufort Sea region.

3.2 Interannual variability

Having examined spatial and seasonal variability in the estimated fluxes, it is natural to consider whether the dataset also gives useful information about interannual variability. Restricting fluxes to individual years does not give enough data points (per month) to permit analysis, particularly early in the period where very often there was only one or two buoys in operation in any particular year (and in some cases none). Instead, the period of the IMB operations is divided into three sections with very roughly equal numbers of data points for each flux and month: 1993-2006, 2007-2012 and 2013-2015. The middle period is chosen to contain the two years with the lowest September extents (2007 and 2012, according to HadISST1.2) in the entire analysis period, to maximise the chance that interannual variability can be detected. For each flux, and month of year, we compare the distribution of values estimated for each period (Figure 8) and use a Welch t-test to judge whether any distributions are significantly different at the 5% level.

For the top melting flux, in no months are the distributions significantly different, and in July and August means and standard deviations are very similar. In May, however, the mean top melting flux is higher in the later period 2013-2015 (2.7 ± 5.8 Wm-2) than in the middle (0.5 ± 1. Wm-2) and early (0.5 ± 0.9 Wm-2) periods. By contrast, in June the mean top melting flux is lower in the later period (13.3 ± 9.6 Wm-2) than in the middle (19.0 ± 8.8 Wm-2) and early (13.7 Wm-2) periods. The differences could conceivably reflect the observed trend towards earlier onset of melt (e.g. Bliss et al, 2019), but the lack of significance makes drawing conclusions difficult.~~Wm^{-2}~~

Early in the winter, the top conductive flux becomes higher (less negative) as time passes: for example, in October the distribution means are -17.9 ± 12.5, -13.2 ± 5.1 and -11.9 ± 5.6 Wm^{-2} for the periods 1993-2006, 2007-2012 and 2013-2015 respectively. Late in the winter, the trend reverses, with top conductive flux becoming lower (more negative) as time passes: for example, in March the distribution means are -10.8 ± 4.1, -14.2 ± 5.9, -14.6 ± 4.2 Wm^{-2} for the three periods respectively. However, as with the top melting fluxes the distributions are not significantly different. Difference in basal conductive flux are still less marked, with distributions similar in all months except October, where the middle period displays a much higher (more negative) mean and greater spread due to the presence of a small number of extreme values (including that from the buoy 2006C, noted above). ~~$Wm^{-2}$$Wm^{-2}$~~

For the ocean heat flux, an upward trend is apparent in the month of July, with means of 10.1 ± 4.4, 20.0 ± 17.0 and 23.6 ± 17.1 Wm^{-2} for the three periods respectively: for the early and late periods, the distributions are barely significantly different. In August, the mean in the middle period (32.5 ± 37.0 Wm^{-2}) is much higher than that of the early (14.0 ± 15.1 Wm^{-2}) and late (14.1 ± 9.5 Wm^{-2}) periods, but the differences are not significant. The paucity

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

Formatted: Superscript

Formatted: Superscript

of significant differences between distributions, coupled with the deliberate choice of periods to maximise interannual variability, suggests that it is difficult to detect robust interannual trends in the IMB dataset in its current state.

3.3 Uncertainty associated with assumptions of the analysis

We assess uncertainty due to ice salinity, snow and ice density, ice conductivity and to the layers used to calculate conductive flux and ocean heat flux. Guided by estimates produced in the modelling studies of Turner et al. (2015) and Vancoppenolle et al. (2008), we use a practical salinity range of 0 – 10 to evaluate uncertainty due to salinity at both upper and basal surfaces of the ice. In fact, the ice salinity causes by far the greatest uncertainty in all measured fluxes, and the effect is most marked when considering the top melting flux. For example, the top melting flux estimated from the buoy 1997D in the month of July 1998 is 31.0 Wm^{-2} when a salinity of 0 is assumed; but 0.4 Wm^{-2} with a salinity of 10. This is due to the much lower latent heat of fusion of ice at higher salinities. Over the distribution of a whole, average July top melting flux is 29.9 Wm^{-2} with a salinity of 0 but 1.6 Wm^{-2} with a salinity of 10.

At first sight, the large uncertainties would render evaluation of the top melting flux in a sea ice model using IMB data extremely difficult. However, the physical meaning of this uncertainty must be correctly understood. The specific latent heat of high salinity ice is higher because a significant fraction of the ice will already have undergone melting. The energy used in melting this ice is accounted for in sensible heating of the top layer of ice, as high salinity ice has a higher heat capacity for this reason. In a sense, top melting of ice, and sensible heating of the top layer, are part of the same process. Undertaking a meaningful evaluation of modelled top melting using the IMB fluxes therefore requires consideration of the thermodynamic treatment of ice in that model. For example, in a model such as HadGEM2-ES, it is appropriate to compare modelled top melting to energy used in melting the entire top layer of ice – equivalent to assuming an ice salinity of 0 in the IMB dataset. This is because HadGEM2-ES does not model ice salinity or heat capacity (as described in more detail in Section 4 below).

The salinity has a much smaller, though still noticeable, effect on the conductive flux. In February 2014, for example, a salinity of 0 is associated with a top conductive flux of -12.5 Wm^{-2}, while a salinity of 10 is associated with a flux of -11.8 Wm^{-2}. Over the whole dataset, the average February top conductive flux is -17.0 (-16.6) Wm^{-2} when a salinity of 0 (10) is assumed. Sensitivity is higher in the summer, as conductivity is more sensitive to salinity at higher temperatures: over the dataset, the average July top conductive flux is 3.1 (-0.1) Wm^{-2} when a salinity of 0 (10) is used. The basal conductive flux displays highest sensitivity to salinity from February – April: for example, the average March basal conductive flux is -13.3 (-11.7) Wm^{-2} when a salinity of 0 (10) is assumed.

Ocean heat fluxes tend to display higher sensitivity to salinity than do the conductive fluxes, but lower than does the top melting flux. This is mainly because temperatures tend to be lower at the basal surface of the ice than at the top during the summer (when top melting and ocean heat fluxes tend to be greatest in magnitude), reducing sensitivity of the latent heat of fusion of ice to salinity. For example, in August 2003 the buoy 2003D displays an ocean heat flux of 24.3 (16.6 Wm^{-2}) when salinity of 0 (10) is assumed. For the distribution as a whole sensitivity is highest in the month of August when the mean ocean heat flux is 23.0 (13.5) Wm^{-2} when salinity of 0 (10) is assumed.

To examine sensitivity to snow density, we use the range 274-374 kgm^{-3}, after Alexandrov et al (2010). Snow density only affects the top melting flux: the highest sensitivity is seen in the month of June, where the average top melting flux is 15.4 (17.9) when snow density of 274 (374) kgm^{-3} is used. We also examine sensitivity to ice density, using the range 917-944 kgm-3, after Cox and Weeks (1982): for the top melting flux, the highest sensitivity is in July, when the average top melting flux is 29.9 (30.7) Wm^{-2} when ice density is 917 (944) kgm^{-3}. The ocean heat flux also depends on ice density, and the largest difference occurs in the month of August, when the average flux is 19.9 (20.5) Wm^{-2} when ice density is 917 (944) kgm^{-3}.

Ice conductivity is also subject to considerable uncertainty. An alternative formulation to the Maykut and Untersteiner method used in this study was proposed by Pringle (2006) following laboratory tests of land-fast sea ice, in which sea ice conductivity k_I is calculated from ice temperature T and salinity S as

$$k_I = 2.11 - 0.011T + 0.09\frac{S}{T}$$

Sensitivity of the IMB-measured fluxes to the conductivity formulation was tested by recalculating conductive and ocean heat fluxes using this alternative method (there is no difference in the top melting fluxes by design). Large difference in the winter top conductive fluxes are apparent, due to the Pringle formulation tending to produce much higher conductivities at low temperatures. For example, for the buoy 1993A in January 1994 a top conductive flux of -18.3 Wm^{-2} is estimated using the Pringle formulation, but only -15.8 Wm^{-2} using the Maykut and Untersteiner formulation. For the dataset as a whole, a mean January top conductive flux of -21.0 Wm^{-2} is estimated with the Pringle formulation and -17.7 Wm^{-2} with the Maykut and Untersteiner formulation.

Finally, sensitivity of the IMB basal conductive and ocean heat fluxes to the depth and thickness of the reference layers used was tested. The fluxes were recalculated with the lowest 20cm of the ice used to calculate sensible heat uptake, and the layer 20-40cm above the ice base to calculate basal conductive fluxes. The largest change in mean basal conductive flux occurs in October, with a mean value of -0.7 Wm^{-2} as opposed to -4.1 Wm^{-2} in the standard configuration. This is associated with temperature gradients being smaller closer to the ice base. The difference decreases through the winter, with -11.7 Wm^{-2} in February, as opposed to -13.7 Wm^{-2} in the standard configuration. The largest difference in ocean heat flux also occurs in October, with a mean value of 2.8 Wm^{-2} as opposed to 5.4 Wm^{-2} in the standard configuration.

In summary, varying parameters of the analysis results in measurable changes to the IMB fluxes. In most cases however, the sensitivity of the fluxes to the parameters is an order of magnitude lower than the absolute values, in the months of the year when the absolute values tend to be at their peak (winter for the conductive fluxes, summer for the top melting and ocean heat fluxes). The main exception is the effect of salinity on the top melting fluxes in summer, but as noted above care is needed when interpreting this uncertainty in the context of a model evaluation.

~~3.~~4. Evaluating modelled sea ice using the IMB-derived fluxes

4.1 Evaluating vertical energy fluxes in HadGEM2-ES with the IMBs

In this section, the distributions of energy fluxes estimated from the IMB data are compared to equivalent fluxes simulated by the coupled climate model HadGEM2-ES. This model, developed from the earlier model HadGEM1, is based on the HadGEM2-AO coupled atmosphere-ocean system, but employs additional components to simulate terrestrial and oceanic ecosystems, and tropospheric chemistry (Collins et al., 2011). The sea ice component is very similar to that used by HadGEM1 (McLaren et al., 2006), employing a sub-grid-scale thickness distribution with 5 categories (Thorndike et al., 1975), elastic-viscous-plastic rheology (Hunke and Dukowicz, 1997), and a zero-layer thermodynamics scheme, described in the appendix to Semtner (1979), in which sea ice has no heat capacity and conduction does not vary with height within the ice. The atmosphere and ocean components contain a number of improvements relative to HadGEM1 (Martin et al., 2011). The sea ice in HadGEM2-ES is modelled on a regular latitude-longitude grid, with a resolution of 1 degree throughout the Arctic.

The Arctic sea ice and surface radiation simulation of the historical ensemble of HadGEM2-ES was evaluated by West et al (2019~~8~~). A number of likely model biases were identified; a low bias in September sea ice extent, a low bias in annual mean ice thickness and a high bias in ice thickness seasonal cycle amplitude, and a tende~~a~~ncy to model overly high surface net downwelling shortwave (SW) flux in summer and overly low surface net downwelling longwave (LW) flux in winter. To perform the evaluation of the ice energy budget in the present study, the model period 1980-1999 is used, chosen to match the period used in West et al (2019~~8~~).

For each month, grid cells lying inside the North Pole and Beaufort Sea regions (Figure 3) are separately identified, and monthly mean top melt flux, top conductive flux, basal conductive flux and ocean heat flux are collected into distributions. Means and standard deviations of these distributions are then compared to those of the IMB fluxes, with model fluxes weighted by grid cell area when calculating these statistics. Before aggregation the fluxes, produced by the model as grid-box means, are divided by ice area to produce means over ice only, for greater consistency with the IMB fluxes. As above, to compare modelled and observed distributions in a systematic manner, a Welch t-test is used, and the 5% level is chosen as a threshold for significance of difference of distributions.

~~Owing to the zero-layer thermodynamics scheme, sea ice conduction is constant with depth, and hence top and basal conductive fluxes are always equal in the model.~~The approach of calculating model distributions over whole regions and time periods, rather than comparing only grid cells lying under IMB tracks during the respective month, is chosen for two reasons. Firstly, internal variability is such that no climate model would be expected to capture the exact atmospheric conditions over a specific IMB track; whether the model could capture the average conditions over a larger-scale region and time period is a more useful question for evaluation purposes. Secondly, even a model grid-cell implicitly models sea ice of many different thicknesses, due to the ice thickness distribution, and therefore the comparison with IMB tracks cannot be made more like-for-like by restricting to a single grid cell.

For both regions, top melt fluxes simulated by HadGEM2-ES tend to be much higher than those measured by the IMBs (Figure 9~~8~~a,b): modelled and observed distributions are significantly different throughout the melt~~ing~~ season. For example, the modelled mean top melt flux of 72.5 ± 8.2 Wm^{-2} in the Beaufort Sea region in June is much higher than the IMB mean of 26.0~~4.5~~ ± 10.2~~9.4~~ Wm^{-2}; in the North Pole region in June, the modelled mean top melt flux of 56.6 ± 14.0 Wm^{-2} is much higher than the IMB mean of 12.4~~1.2~~ ± 8.3~~6.9~~ Wm^{-2}. A Welch t-test shows the modelled and observed distributions of top melt fluxes to be significantly different at the 5% level throughout the summer in both regions. The phase of the annual cycle in top melt is shifted slightly earlier, with the effect that, in both regions, the modelled June and July means are very similar while the IMB estimates show a distinct maximum in July. However, the greater top melt in the Beaufort Sea region relative to the North Pole region is captured by the model. The bias towards excessive top melting displayed by HadGEM2-ES is consistent with the finding by West et al (2019) that summer net SW fluxes are overestimated in HadGEM2-ES. It was shown in this study that this was likely associated with an early onset of surface melting in the model. In turn, this triggers the meltpond parameterisation of HadGEM2-ES, lowering surface albedo at an earlier time of year than meltponds would have formed in reality.

The annual cycle of top conductive flux is broadly captured by HadGEM2-ES (Figure 9~~8~~c,d), with strongly negative values modelled in the autumn, winter and spring and weakly positive values in the summer. However, from September-May modelled means are more strongly negative than observed means: for example, in the North Pole region in December a modelled mean of -31.0 ± 7.6 Wm^{-2} is higher in magnitude than the IMB mean of -18.1~~20.4~~ ± 12.7~~10.3~~ Wm^{-2}. Ice thickness in HadGEM2-ES is known to be biased low in early winter (West et al, 2019), and in section 4.2 below the extent to which this bias is responsible for the conductive flux bias is investigated. The higher magnitude of conductive fluxes in winter in the Beaufort Sea region relative to the North Pole region is captured by the model; however, the higher values of fluxes in May and September in the Beaufort Sea region relative to the North Pole region are not captured. Modelled and observed flux distributions are significantly different in all months except February, March and November (North Pole region) and June, September and October (Beaufort Sea region).

Modelled values of basal conductive flux at each model grid-box~~,~~ are ~~as noted above, are~~ identical to those of top conductive flux, due to the HadGEM2-ES zero-layer thermodynamics scheme. Consequently HadGEM2-ES overestimates the magnitude of basal conductive flux in autumn and winter more severely than it does top conductive flux (Figure 9~~8~~e,f). The overestimation is most severe during the autumn; in the Beaufort Sea region in October a mean modelled flux of -28.1 ± 11.1 Wm^{-2} is much higher in magnitude than the mean observed flux of -7.9~~3.2~~ ± 16.8~~5.2~~ Wm^{-2}. As the basal conductive flux in the freezing season is the principal driver of ice growth, this suggests that HadGEM2-ES is likely to model substantiall~~ignificantl~~y stronger ice growth during these months than was measured at the IMB sites. Modelled and observed fluxes are significantly different in all months and regions except, barely, in the North Pole region in August.

For the ocean heat flux, in the Beaufort Sea region the model produces a similar seasonal cycle to that estimated from the IMB data, with very small values in the winter and a wide range of positive values in the summer (Figure 9g,h); only in April and May are the distributions significantly different. For the North Pole region however, two differences are apparent. Firstly, spread in winter ocean heat fluxes is much higher in the model, with a small number of very high fluxes; these occur near the southern boundary of the region, where the ice cover meets warmer water moving north through the Fram Strait. Secondly, modelled ocean heat fluxes in summer tend to be higher than those measured by the IMBs. For example, in July a modelled mean of 27.2 ± 13.9 Wm^{-2} compares to an observed mean of 12.1 ± 7.4 Wm^{-2}. In the North Pole region, modelled and observed distributions are significantly different in all months except September. ~~the distribution means are very similar in August but the observed spread far wider, with a modelled mean of 23.5 ± 10.0 Wm^{-2} comparing to an observed mean of 18.8 ± 16.0 Wm^{-2}; however, modelled values are larger than observed earlier in the summer, with evidence of a phase offset between model and observations. The larger spread in the IMB estimates is consistent with a comparison of observed point values to modelled means over grid cells. For the Beaufort Sea region, the modelled mean of 26.6 ± 12.0 Wm^{-2} is much lower than the IMB-estimated value of 42.4 ± 44.1 Wm^{-2}. It is possible that the discrepancy here is caused by the observed rapid retreat of summer sea ice in the Beaufort Sea in the past 15 years, with accompanying direct solar heating of the ocean mixed layer, and the increased occurrence of exceptionally high values of ocean heat flux. These effects are not represented in the HadGEM2-ES simulation in the 1980-1999 period.~~

The HadGEM2-ES ocean heat flux distributions are in fact similar between the North Pole and Beaufort Sea regions, but evaluate differently because the IMB ocean heat fluxes in the North Pole region are much lower. In other words, the spatial variability in summer ocean heat fluxes suggested by the IMB data is not captured by the model. This may be related to late summer ice concentration biases in HadGEM2-ES: ice concentration is biased low in the North Pole region, but not in the Beaufort Sea region. This would tend to cause a high bias in net SW absorption by the ocean mixed layer, and hence a positive bias in ocean heat flux.

It is instructive also to examine how modelled oceanic heat convergence (OHC) compares to real-world estimates. Arctic Ocean heat convergence in HadGEM2-ES from 1980-1999 is 4.9 Wm^{-2}, roughly consistent with estimates of heat transport through the Fram Strait, the dominant pathway by which the Arctic Ocean gains heat (e.g. Schauer et al, 2004). However, oceanic heat convergence in the North Pole region (as defined in Figure 3) is higher at 9.1 Wm^{-2}, while oceanic heat convergence in the Beaufort Sea region is lower at 1.6 Wm^{-2}. These patterns are consistent with most of the Arctic Ocean heat convergence being released relatively close to the Fram Strait and Barents Sea, the principal points of ingress of relatively warm Atlantic water. It is noteworthy that the large difference in OHC between the two regions is not mirrored in the ocean-to-ice heat flux. This is consistent with the finding by Keen et al. (2018) that much of the ocean-to-ice heat flux in HadGEM2-ES derives from direct solar heating, rather than OHC. Studies by Bitz et al. (2008), Perovich et al. (2008) and Steele et al. (2010) show this is likely to be true in the real world also. Hence differences in OHC are unlikely to contribute to differences in ocean-ice heat flux.

The model biases in summer top melt and winter conductive fluxes are larger than most of the uncertainties measured in section 3.3. The model bias in top melt is of a similar magnitude to the uncertainty in top melting due

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

to salinity, albeit in the opposite direction (higher salinities imply an even greater model bias). However, for the reasons stated in section 3.3, the most meaningful comparison is obtained by considering the energy used in melting the whole of the top layer of ice, including the melting of brine pockets during sensible heating – equivalent to using a salinity of 0 in calculating top melting. Hence the model biases stated above – for the 'standard configuration' – are likely to present the most accurate picture.

4.2 Links to sea ice and surface radiation simulation of HadGEM2-ES

A top melting bias of ~40 Wm^{-2} is estimated for the month of June in both the North Pole and Beaufort Sea regions. This is consistent with the finding of West et al (2019~~8~~) that June surface net shortwave (SW) radiation in the model was biased high relative to a variety of satellite and reanalysis datasets, by around 20 Wm^{-2} over the Arctic Ocean on average. Relative to CERES-EBAF measurements from 2000-2013 (Loeb et al., 2009), HadGEM2-ES overestimates June net SW in the North Pole and Beaufort Sea regions by 30 and 9 Wm^{-2} respectively. Owing to the recent trend to earlier onset of surface melt over the past 30 years (e.g. Markus et al., 2009), and attendant likely decrease in surface albedo, these biases are likely to be underestimated. Hence it is likely that a major part of the model bias in top melting can be explained by a model bias in net SW. In West et al (2019~~8~~) it was shown that a tendency for surface melt onset to occur too early in HadGEM2-ES, reducing the surface albedo in a parameterisation of the effect of melt-ponds, was likely to be principally responsible for this.

The severe overestimation in magnitude of basal conductive fluxes during the early part of the melt season can be partly explained by the zero-layer thermodynamics scheme of HadGEM2-ES; the thermal inertia effect seen in the IMBs, whereby the basal conductive flux drops much more slowly in autumn than the top conductive flux, does not occur in the model. However, as the top conductive fluxes also tend to be considerably higher in the model than in the IMB estimates, the thermal inertia effect is likely to be only partially responsible. In West et al (2018) two other model biases were identified as being likely to lead to a high bias in winter ice growth (analogous to the basal conductive flux bias): a negative bias in ice thickness during early winter, and a negative bias in downwelling longwave (LW) radiation throughout the season. It was estimated that these biases were likely to lead to surface flux biases of order ~10 Wm^{-2} throughout the freezing season. Hence these are also likely to explain a portion of the basal conductive flux bias noted above.

The excessive modelled top melting and basal conductive fluxes identified would be likely to lead to too strong ice growth, and ice melting, in winter and summer respectively, and an associated amplification of the ice thickness seasonal cycle. Such an amplification was identified in HadGEM2-ES by comparing modelled ice thickness to the forced ice-ocean model PIOMAS, as well as to estimates from satellites and submarines. Hence there is a high level of consistency between the model biases inferred from the IMB estimates, and those inferred from the sea ice state and surface radiation evaluations in West et al (2018). The IMB evaluation, however, provides additional insight to the picture, by providing consistent evaluation of previously unknown processes such as top melting. In addition the IMB evaluation clearly identifies a role for the zero-layer thermodynamics scheme in driving a bias towards excess ice growth during the early winter in HadGEM2-ES.

4.3 The relationship between conductive flux, ice thickness and snow depth

Both top and basal conductive fluxes are strongly related to ice thickness, as thicker ice tends to have weaker temperature gradients than thinner ice under similar atmospheric conditions. Conductive fluxes are also related to snow depth for similar reasons. To examine the extent to which the HadGEM2-ES conductive flux biases are influenced by biases in ice thickness, conductive fluxes from November – March were aggregated into ice thickness bins, 20cm wide and ranging from 0-4m, in both IMBs and model. HadGEM2-ES ice thicknesses are much lower than ice thicknesses sampled by the IMBs from November – March in both the North Pole and Beaufort Sea regions, with most of the ‘overlap’ occurring in the range 1-2m (Figure 10).

For the North Pole region, IMB-measured top conductive flux is similar in magnitude to HadGEM2-ES conductive flux for the overlapping range of thicknesses, but IMB-measured basal conductive flux tends to be lower. For example, in the 1.4-16m range the IMB-measured top and basal conductive flux range from -33.0 to -11.9 Wm^{-2} and -24.0 to -10.8 Wm^{-2} respectively, while the HadGEM2-ES conductive flux range from -15,6 to -26.6 Wm^{-2}. As above, this suggests that the uniform conductive flux assumption of HadGEM2-ES is important in driving excessive ice growth in this model. However, in the Beaufort Sea region both top and basal conductive fluxes from the IMBs are much lower than HadGEM2-ES even in the region of overlapping thicknesses. For example, in the 1.4-1.6m range the IMB-measured top and basal conductive flux range from -18.8 to -7.8 Wm^{-2} and -18.4 to -8.3 Wm^{-2} respectively, while the HadGEM2-ES conductive flux ranges from -30.0 to -21.4 Wm^{-2}. This indicates that in the Beaufort Sea at least, ice thickness biases and the uniform conductive flux assumption are not the only factors driving excessive ice growth: biases in atmospheric thermal forcing are also at work.

The HadGEM2-ES conductive fluxes display an inverse relationship with ice thickness. Cells with thinner ice tend to have higher conductive flux (Figure 10a-b): in the Beaufort Sea region for example, the correlation coefficient between conductive flux and the logarithm of ice thickness is 0.66. Curiously, there is no sign of a similar relationship in the IMBs in either region; in the Beaufort Sea region the correlation coefficient between the log of ice thickness and top (basal) conductive flux is small and not significant at 0.25 (-0.16). In particular, both regions exhibit large numbers of IMB measurements with high ice thicknesses and high top conductive fluxes. In most cases these points are associated with strong ice cooling and a very low basal conductive flux.

When conductive flux is compared to snow depth, it can be seen that similar snow depths are associated with much greater conductive fluxes in HadGEM2-ES than in the IMB data (Figure 10c-d), indicating that snow depth biases are unlikely to be a major contributor to the conductive flux biases. A similar inverse relationship between conductive flux and snow depth is seen in HadGEM2-ES, stronger in the Beaufort Sea than at the North Pole. Unlike with the ice thickness, there is the suggestion of a similar relationship in the IMB data, with the 10-15cm (5-10cm) category in the North Pole (Beaufort Sea) displaying much higher conductive fluxes than are present in the other categories. No data points with high snow depths display high conductive fluxes. As a result, the correlation between top (basal) conductive flux and the log of ice thickness in the IMB estimates is 0.47 (0.48).

4.5. Representativeness of the IMB-estimated fluxes

The comparison of HadGEM2-ES modelled fluxes to those inferred from the IMB measurements reveals several potential model biases, notably the overestimation of top melt flux in June and July, and overestimation of the magnitude of basal conductive flux in the early freezing season. In this section the accuracy of this method of model bias estimation is discussed.

The model flux distributions evaluated represent area-weighted means over the ice-covered fractions of grid cells, each of order 10-100 km in width, chosen to cover the North Pole and Beaufort Sea regions as defined in Section 2. The 'true model bias' would represent the difference between this and the average flux over ice covered regions for each month in the 1980-1999 period. By contrast, the IMB-measured means are derived from a relatively small number of single point measurements from these regions (points that tend to move position during an individual month, with the general ice flow), most from a period somewhat later than 1980-1999. To assess the accuracy of the model biases inferred, a method of estimating the order of magnitude of the likely error in the IMB estimated mean fluxes is required.

One source of error derives from the temporal offset in the IMB measurements relative to the model. To assess the impact of this, we compare the HadGEM2-ES vertical ice fluxes from the period 1980-1999 to those from the period 2000-2015, during which most of the IMB measurements were taken (as the historical experiments end in 2005, the RCP8.5 experiment was used from 2006 onwards). Flux anomalies in the later period relative to the earlier period are mostly small (below 2 Wm^{-2} in magnitude) in both regions. However, there is a significant negative anomaly in July top melting, -6 Wm^{-2} and -9 Wm^{-2} in the North Pole and Beaufort Sea regions respectively. There is also a positive anomaly in September basal conductive flux (3 Wm^{-2} in both regions), and in the Beaufort Sea moderate basal conductive flux anomalies continue into the winter, being 5, 3, -4 and -2 Wm^{-2} from October-January respectively. These anomalies are likely to reflect earlier melting and later freezing of ice in this region. They are small in size compared to the HadGEM2-ES model biases, but suggest that the temporal bias may cause model top melting bias in July, and Beaufort Sea basal conductive flux in autumn, to be slightly overstated.

A potentially more serious source of error is sampling: bias would be introduced if the IMB measurements were systematically over- or under-sampling locations with higher than average flux in a particular month. The Arctic sea ice cover is highly heterogeneous, with ice conditions varying substantially over all scales. For most variables (for example snow thickness, ice salinity or ice albedo), it is difficult to assess whether the variability of the ice pack is sufficiently sampled by the IMB measurements, due either to a lack of reference datasets or to an inability to estimate these variables at the IMB locations. However, the degree to which the ice thickness distribution (which affects conductive fluxes in particular) is correctly sampled by the IMBs can be assessed. The effect of errors in the ice thickness distribution on the IMB-measured fluxes is estimated in Appendix A, and is shown to be small compared to the model biases identified. We note that the ice thickness distribution sampling bias is likely to be particularly strong (relative to other variables) due to the deliberate placing of IMBs in level multiyear ice, and due to the Lagrangian movement of the IMBs combined with the generally short lifetime of thin ice floes, which tend to grow quickly in winter and melt quickly in summer. It is proposed that given the effect of this bias is weak, it is likely that the effect of other, not deliberately introduced, biases is weaker still, and that the model biases identified in Section 4 are likely to be robust features.

~~5.~~6. Conclusions

Around 500 estimates of monthly mean top melt, top conduction, basal conduction and ocean heat flux have been estimated from data measured by the Arctic IMB network, with the number of estimates available for each month ranging from 26 to 59. The distributions capture seasonal and spatial variability~~tion~~ in the vertical fluxes analysed, but do not contain sufficient data points to capture interannual variability~~and provide a valuable source of information on the internal energy balance of Arctic sea ice~~. Comparison of modelled fluxes to observed fluxes in the two densely sampled regions in the North Pole and Beaufort Sea reveals substantial model biases, notably to high top melt fluxes in summer and to high (negative) basal conductive fluxes in autumn and early winter. Uncertainty in the IMB fluxes due to parameters of the analysis, and biases due to inadequate sampling of thin and very thick ice types, are likely to be small relative to the model biases identified. ~~Biases in the IMB flux estimates due to inadequate sampling of thin and very thick ice types are likely to be small relative to the model biases identified.~~

The flux biases are consistent with an evaluation of the sea ice simulation of HadGEM2-ES that identified an over-amplified seasonal cycle in ice thickness, with model ice growth and melt biased high in winter and summer respectively, as well as a high model bias in net SW radiation in June, a low bias in net LW radiation throughout the winter, and a low model bias in ice thickness in autumn and early winter. The IMB analysis confirms that the net SW bias is likely to cause overly strong ice thinning during summer via anomalously strong top melting of ice. The IMB analysis also allows the effect of biases in ice thickness, snow depth, and atmospheric conditions, as well as the effect of the uniform conductive flux assumption, on the conductive fluxes to be separately examined. In this way it is confirmed that both low ice thickness, ~~, and that~~and ~~the~~ cold atmospheric conditions ~~and low ice thickness of HadGEM2-ES~~ are likely to be driving anomalously strong winter ice growth via the basal conductive flux, both conclusions already suggested by West et al (2019). However, the IMB analysis also suggests that the zero-layer thermodynamic scheme of HadGEM2-ES plays a role in promoting this anomalously strong ice growth. The IMB analysis also provides evidence that snow depth biases are not important in driving the ice growth biases of HadGEM2-ES.

The calculated IMB fluxes hence offer a ~~potentially~~ valuable tool for increasing understanding of sea ice simulations in coupled models, as they allow detailed examination of the links between atmospheric forcing of sea ice and the resulting sea ice state. This is particularly the case for the upcoming Phase 6 of the Coupled Model Intercomparison Project (CMIP6), for which diagnostics of ice energy and mass fluxes, such as top melting and conduction, have been requested for all sea ice models participating in this experiment (Notz et al., 2016). Understanding of the processes leading to biases in a given sea ice state enables better understanding of modelled sea ice spread in the present day and in the future, and may therefore also allow better understanding of future projections in sea ice state.

~~By far the~~The greatest source of uncertainty in estimating the IMB fluxes derives from lack of knowledge of ice salinity at the measurement sites, and therefore the thermodynamic properties of conductivity and heat capacity. A method of measuring salinity at the IMB sites would greatly reduce the uncertainty in the IMB estimates, particularly for ocean heat flux, enhancing the usefulness of this dataset as a tool for model evaluation.

Appendix A: The ice thickness sampling bias and its effect on flux distributions

An analysis of the distribution of monthly mean ice thicknesses sampled by the IMBs finds that most lie in the range 1.4 – 3.6 m. However, analysis of submarine measurements of ice thickness in the Central and Western Arctic from 1981-2000, as collated by Rothrock et al. (2008), shows a substantial proportion of ice to be of thickness outside these bounds. In order to estimate the effect of this sampling bias, we use the following simple model. The thickness distribution is discretised, in a similar manner to HadGEM2-ES, by separating ice into five thickness categories, with minimum thickness bounds at 0 m, 0.6 m, 1.4 m, 2.4 m and 3.6 m. Given a mean flux $F_{IMB}^{m,r}$ for month *m* and region *r* that is estimated from the IMBs by averaging all fluxes for that month and region, the total observational error can be characterised as

$$F_{err} = F_{IMB}^{m,r} - F_{actual}^{m,r} \qquad \text{(A1)}$$

where $F_{actual}^{m,r}$ is the actual value of that flux, averaged over the region and month in question, for the period 1993-2015. The mean flux can be further split into thickness categories by setting

$$F_{IMB}^{m,r} = \sum_i F_{IMB-cat}^{m,r,i} \bullet \left(N_i^{m,r} / N_{total}^{m,r} \right) \qquad \text{(A3)}$$

where the $F_{IMB-cat}^{m,r,i}$ are the average IMB flux over month *m*, region *r* and category *i*, $N_i^{m,r}$ is the total number of IMB fluxes in month *m*, region *r* and category *i,* and $N_{total}^{m,r}$ is the total number of IMB fluxes in month *m* and region *r*.

Similarly the actual flux values can be written as

$$F_{actual}^{m,r} = \sum_i F_{actual-cat}^{m,r,i} \bullet a_i^{m,r} \qquad \text{(A4)}$$

where $a_i^{m,r}$ is the average fraction of ice in month *m* and region *r* that is in category *i* (expressed as a proportion of average fraction of ice in the region).

It can be seen that $N_i^{m,r} / N_{total}^{m,r}$ acts as an IMB-based estimate of $a_i^{m,r}$, and that the error in $F_{IMB}^{m,r}$ due to the ice thickness sampling bias is exactly that due to the error in this estimate. Hence we can characterise the sampling error by expressing F_{err} in terms of systematic and sampling errors in the following way:

$$
\begin{aligned}
F_{err}^{m,r} &= \sum_i \left(F_{IMB-cat}^{m,r,i} \bullet \left(N_i^{m,r} / N_{total}^{m,r} \right) - F_{actual-cat}^{m,r,i} \bullet a_i^{m,r} \right) \\
&= F_{err\_systematic} + F_{err\_sample} \\
&= \frac{1}{2} \sum_i \left(F_{IMB-cat}^{m,r,i} - F_{actual-cat}^{m,r,i} \right) \left(N_i^{m,r} / N_{total}^{m,r} + a_i^{m,r} \right) \\
&\quad + \frac{1}{2} \sum_i \left(N_i^{m,r} / N_{total}^{m,r} - a_i^{m,r} \right) \left(F_{IMB-cat}^{m,r,i} + F_{actual-cat}^{m,r,i} \right)
\end{aligned}
\qquad \text{(A5)}
$$

where F_{err_sample} captures the flux error due to ice thickness sampling; $F_{err\_systematic}$ describes the error due to inaccuracy in the IMBs estimating *F* for each particular category, effectively capturing the remaining observational error that is beyond the scope of the current analysis.

To estimate F_{err_sample}, we first need estimates of $a_i^{m,r}$, the real-world proportion of ice in a given thickness category for each region and month, for which as discussed above we use submarine measurements collated by Rothrock (2008) that capture small-scale variation in sea ice thickness. Ice thickness measurements in each category are collected for each month and region, and used to generate an estimate of $a_i^{m,r}$. In practice, measurements are abundant during spring and autumn, but sparse during summer and non-existent during winter. To alleviate this problem, we interpolate $a_i^{m,r}$ using a 5-month binomial mean, weighted by number of measurements, in order to produce a smooth seasonal cycle, motivated by the observation that maximum and minimum ice thickness values often occur during spring and autumn respectively. The resulting seasonal cycles of $a_i^{m,r}$ are shown in Figure S2~~A1~~.

Secondly, estimates of $\frac{1}{2}\left(F_{IMB-cat}^{m,r,i} + F_{actual-cat}^{m,r,i} \right)$, representative average fluxes for each thickness category, are required. To calculate representative fluxes of conduction, we combine a simple model of the relationship between conductive fluxes and ice and snow thickness with the IMB measurements of conduction. The surface flux $F_{sfc} = F_{atmos} + BT_{sfc}$ is approximated by linearising the dependence on surface temperature T_{sfc}, referenced to 0°C. The surface flux is set equal to the top conductive flux $F_{condtop}$, and a constant rate of change of conductive flux from the top to the basal surface of the ice is assumed, such that $\frac{1}{2}\left(F_{condtop} + F_{condbot} \right) = \frac{\left(T_{sfc} - T_{bot} \right)}{R_{ice}}$, where $F_{condbot}$ represents basal conductive flux, T_{bot} ice base temperature, and $R_{ice} = h_{ice} / k_{ice} + h_{snow} / k_{snow}$

is the thermal insulance of the ice-snow column, h_{ice} and h_{snow} being ice and snow thickness, and k_{ice} and k_{snow} ice and snow conductivity respectively. Eliminating T_{sfc} from the above equations gives

$$F_{condtop} = \frac{F_{atmos} + BT_{bot}}{1 - BR_{ice}^{HS-top}} \qquad \text{(A6)}$$

where $R_{ice}^{HS-top} = R_{ice} / (1 + \alpha_{heat})$ and $\alpha_{heat} = \dfrac{F_{condtop} - F_{condbot}}{F_{condtop} + F_{condbot}}$.

Equation (A6) can be combined with the IMB estimates of conductive fluxes, and snow and ice thickness to produce, for each monthly mean IMB measurement, an associated value of F_{atmos}, a measure of the atmospheric forcing on the ice that is independent of small-scale variations in ice thickness. Hence for each month and region a distribution in F_{atmos} can be produced; mean values of this can be fed back into the simple model to produce an average conductive flux that would be expected for each ice thickness category. The average conductive flux is then multiplied by $\left(N_i^{m,r} / N_{total}^{m,r}\right) - a_i^{m,r}$, as estimated above, and summed over categories to produce the flux error $F_{err}^{m,r}$.

The resulting flux bias is below 1 Wm^{-2} in magnitude year-round in the North Pole region. It is slightly larger in the Beaufort Sea in winter time, achieving values of -1.5 Wm^{-2} to -1 Wm^{-2} from November-February. The values are small compared to the model-observation differences identified in Section 4, and so we conclude that the ice thickness sampling bias does not seriously affect ability to evaluate modelled conductive fluxes.

A less strong, but still discernible, relationship exists between top melt flux and ice thickness, due to ice albedo tending to be lower for thinner ice. However, this effect is likely to be associated with a significant difference in albedo only for the thinnest category of ice, and then only in the absence of snow (e.g. Ebert et al., 1995). To estimate this effect, we use an average albedo of 0.55 for bare ice in the top four thickness categories, and 0.35 in the lowest category, based on observed values reviewed and collated by Pirazzini (2008). We assume that ice is snow-covered 80% of the time from October-May (with a corresponding albedo of 0.85), 50% of the time in June and September, and 20% of the time in July and August. Finally, we calculate mean values of downwelling SW radiation for the North Pole and Beaufort Sea region from CERES-EBAF from 2000-2013, and multiply these by the albedo differences implied by the anomaly of ice fraction in category 1. With this method, a maximum flux bias of -3 Wm^{-2} is estimated for the North Pole region, in July, and of -2 Wm^{-2} for the Beaufort Sea region, also in July. Again, this anomaly is small compared to the model-observation differences seen in Section 4, and it is concluded that the sampling bias similarly does not affect ability to evaluate modelled top melting.

An estimate of the influence of the sampling bias on ocean heat flux estimates is more difficult, due to a less clear relationship between ice thickness and the ocean heat flux, and the frequent presence of rapid changes in ice thickness during the months in which ocean heat flux is highest (July and August). It is likely that very small ice thicknesses are associated with elevated ocean heat flux in the summer months due to greater solar penetration through ice. However, the ice thickness sampling bias is at its least severe during the summer months, as thinner ice is sampled by the IMBs simply through the melting of originally thicker floes on which the IMBs were placed. In addition, thinning ice which induces a particularly high ocean heat flux is likely to melt out quickly, and contribute a correspondingly small fraction to the ice thickness distribution.

We examined the sensitivity of the average ocean heat fluxes to this issue by assuming ocean heat fluxes in category 1 to be systematically larger than those in the remaining categories (as diagnosed from the IMBs) by a factor λ. With λ=3, August ocean heat fluxes in the Beaufort Sea region are on average 80Wm-2 greater in category 1 than in thicker ice categories; it is unlikely that the solar penetration effect could be associated with larger flux differences than this. The largest average flux bias associated with this effect is then -6.3 Wm-2, also seen in August in the Beaufort Sea region. Hence this could be taken as a reasonable upper bound for the effect of the sampling bias on ocean heat fluxes. It is smaller in magnitude than the model ocean heat flux biases diagnosed, although the difference is less than those for the top melt and conductive fluxes.

Code availability

The code used to analyse the IMB data is published in two repositories, corresponding to two stages of the analysis. The code used to read, quality control, and process the data into consistent quantities on consistent time points, can be downloaded from https://github.com/sammiebuzzard/IMBS_MO. The code used to produce datasets of monthly mean energy fluxes from this processed data can be downloaded from https://github.com/alex-west-met-office/IMBS_Stage2_modeleval_MO. In addition, ~~Code used in this study, and associated documentation, is provided as a supplement in .tar format. The code used to read and analyse the IMB data is provided located in subdirectory ‘IMB_code’; documentation of many key aspects of this code is located in subdirectory ‘Documentation’. T~~the code with which Figures ~~2, 3, 4, 7, 8 and B1~~ 3-10 were produced is provided as a supplement; many of these routines make extensive use of the open source Iris and Cartopy libraries.~~also provided, in subdirectory ‘Figures’.~~

Data availability

The raw IMB data are~~is~~ publicly available, and can be downloaded from http://imb.erdc.dren.mil/buoysum.htm. The processed IMB data, and the derived dataset of monthly mean fluxes, are published with this revision at https://zenodo.org/record/3773811 and https://zenodo.org/record/3773997 respectively. ~~However, the~~The diagnostics from HadGEM2-ES used in this study are not publicly available. However, they can be made available upon request to the author.~~.~~

Acknowledgements

This work was supported by the Joint UK BEIS/Defra Met Office Hadley Centre Climate Programme (GA01 101), and the European Union’s Horizon 2020 Research & Innovation programme through grant agreement No. 727862 APPLICATE. MC was supported by NE/N018486/1.

Author contribution

The IMB-derived fluxes were calculated, and the subsequent model evaluation performed, by Alex West. The analysis of sampling error in the Appendix was designed and carried out by Alex West. The paper was written in its final form by Alex West with assistance from Mat Collins and Ed Blockley.

References

Alexandrov, V., Sandven, S., Wahlin, J., and Johannessen, O. M.: The relation between sea ice thickness and freeboard in the Arctic, The Cryosphere, 4, 373–380, https://doi.org/10.5194/tc-4-373-2010, 2010

Bitz, C. and Lipscomb, W. H.: An energy-conserving thermodynamic model of sea ice, J. Geophys. Res. (Oceans), 104, C7, 15669-15677. doi: 10.1029/1999JC900100, 1999

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, 14, 8525-8548. doi: 10.1002/2016JD025099, 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

Cox, G. F. N. and Weeks, W. F., Equations for Determining the Gas and Brine Volume in Sea Ice Samples, J. Glaciol, 29, 102, 306-316, doi: 10.3189/S0022143000008364, 1983

Ebert, E. E., Schramm, J. L. and Curry, J. C.: Disposition of solar radiation in sea ice and the upper ocean, J. Geophys. Res., 100, C8, 15,965-15,975. doi: 10.1029/95JC01672

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.

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

Kwok, R., Arctic sea ice thickness, volume, and multiyear ice coverage: losses and coupled variability (1958–2018), Env. Res. Lett., 13, 10, 5005, doi: 10.1088/1748-9326/aae3ec

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

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

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

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

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

Lei, R., Li, N., Heil, P., Cheng, B., Zhang, Z. and Sun, B.: Multiyear sea ice thermal regimes and oceanic heat flux derived from an ice mass balance buoy in the Arctic Ocean, J. Geophys Res (Oceans), 119, 1, 537-547. doi: 10.1002/2012JC008731, 2014

Lei, R., Cheng, B., Heil, P., Vihma, T., Wang, J., Ji, Q and Zhang, Z.: Seasonal and Interannual Variations of Sea Ice Mass Balance From the Central Arctic to the Greenland Sea. J. Geophys. Res. (Oceans), 123, 4, doi: 10.1002/2017JC013548, 2018.

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

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. (Oceans), 114, C12. 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

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

Maykut, G. A. and Untersteiner, N.: Some results from a time-dependent thermodynamic model of sea ice, J. Geophys Res (Oceans and Atmospheres), 76, 6, 1550-1575. doi: 10.1029/JC076i006p01550, 1971.

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.

Notz, D., McPhee, M. G., Worster, M. G., Maykut, G. A., Schlunzen, K. H. and Eicken, H.: Impact of underwater-ice evolution on Arctic summer sea ice, J. Geophys. Res. (Oceans), 108, C7, doi: 10.1029/2001JC001173, 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

Notz, D., Jahn, A., Holland, M., Hunke, E., Massonet, F., Stroeve, J., Tremblay, B. and Vancoppenolle, M.: The CMIP6 Sea-Ice Model Intercomparison Project (SIMIP): understanding sea ice through climate-model simulations, Geosci. Model Dev., 9, 3427–3446. doi:10.5194/gmd-9-3427-2016, 2016

Olonscheck, D. and Notz, D.: Consistently Estimate Internal Climate Variability from Climate Model Simulations. J. Clim., 30, 23, 9555–9573. doi: 10.1175/JCLI-D-16-0428.1, 2018

Ono, N.: Specific heat and heat of fusion of sea ice. In H. Oura, editor, Physics of Snow and Ice, 1, 599–610. Institute of Low Temperature Science, Hokkaido, Japan, 1967

Perovich, D. and Elder, B.: Estimates of ocean heat flux at SHEBA, Geophys. Res. Lett., 29, 9, 58-1-58-4. doi: 10.1029/2001GL014171, 2002

Perovich, D. and Richter-Menge, J. A.: From points to Poles: extrapolating point measurements of sea-ice mass balance, Ann Glaciol, 44, 188-192. doi: 10.3189/172756406781811204, 2006

Perovich, D., 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.

Pirazzini, R.: Factors controlling the surface energy budget over snow and ice, Finnish Meteorological Institute Contributions, 75, 2008.

Pringle, D. J., Eicken, H., Trodahl, H. J., Backstrom, L. G. E.: Thermal conductivity of landfast Antarctic and Arctic sea ice, J. Geophys. Res. (Oceans), 112, C4, doi: 10.1029/2006JC003641

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 (Oceans), 113, C5. doi: 10.1029/2007JC004252, 2008

Rösel, A., Itkin, P., King, J., Divine, D., Wang, C., Granskog, M. A., Krumpen, T., Gerland, S., Thin Sea Ice, Thick Snow, and Widespread Negative Freeboard Observed during NICE-2015 north of Svalbard, 123, 2, 1156, 1176, doi: 10.1002/2017JC012865

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.

Schwarzacher, W.: Pack ice studies in the Arctic Ocean. J. Geophys. Res., 64:2357–2367, doi: 10.1029/JZ064i012p02357, 1959.

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.

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

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, 2012.

Stroeve, J. C., Barrett, A., Serreze, M. and Schweiger, A.: Using records from submarine, aircraft and satellites to evaluate climate model simulations of Arctic sea ice thickness, The Cryosphere, 8, 1839–1854, doi: 10.5194/tc-8-1839-2014, 2014.

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

Wang, M. and Overland, J. E.: A sea ice free summer Arctic within 30 years: An update from CMIP5 models. Geophys. Res. Lett., 39, L18501 , doi:10.1029/2012GL052868, 2012

West, A. E., Collins, M. C., Blockley, E., Ridley, J., Bodas-Salcedo, A.: Attribution of sea ice model biases to specific model errors enabled by new induced surface flux framework. The Cryosphere Discuss., https://doi.org/10.5194/tc-2018-60, in review, 2019.

Wettlaufer, J. S.: Heat flux at the ice-ocean interface, J. Geophys Res (Oceans), 96, C4, 7215-7236. doi: 10.1029/90JC00081, 1991

Whole Arctic	Top melt flux (Wm^{-2})		Top conductive flux (Wm^{-2})		Basal conductive flux (Wm^{-2})		Ocean heat flux (Wm^{-2})	
Number of observations		463		414		463		414
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
January	0.0	0.0	-16.29.2	6.110.3	-134.0.2	5.75/9	1.4-0.6	95.0.0
February	0.0	0.0	-16.920.0	10.86.9	-13.713.6	6.4.7	0.6-2.2	4.28.1
March	0.0	0.0	-143.5.6	7.35.1	-12.712.7	4.44.6	1.5-0.9	5.68.7
April	0.0	0.0	-7.55.9	3.14.4	-9.79.8	3.33.3	2.3-0.0	4.92.7
May	1.13	3.23.2	-0.50.4	2.34.0	-6.26.0	2.22.3	23.4.2	4.04.8
June	16.48	110.0.1	43.88	1.83.1	-2.20	1.66	12.311.0	16.512.7
July	29.928.6	17.86.7	1.04.4	2.81.0	0.50.2	1.31.2	18.124.0	24.015.3
August	8.17.8	6.47	-1.31	4.93.5	1.00.9	1.10.8	19.225.1	23.928.7
September	0.61.0	1.26	-6.38.1	7.94.5	0.74	1.92.1	9.411.0	11.413.5
October	0.0	0.00	-14.47.2	11.68.9	-4.0-1.9	4.111.4	5.43.3	13.08.8
November	0.0	0.0	-17.320.8	12.17.0	-9.28.5	10.39.9	4.63.8	7.113.4

Formatted Table
Formatted: Font: Bold
Formatted: Superscript
Formatted: Superscript
Formatted: Superscript
Formatted: Superscript
Formatted: Right
Formatted: Right
Formatted: Right
Formatted: Right

December	0.0	0.0	-17.6~~22.3~~	~~13.2~~6.8	-12.5~~12.4~~	~~8.0~~6.6	1.3~~-1.0~~	5.2~~7.8~~

Table 1. Mean and standard deviations of fluxes measured from the IMB data in Wm^{-2} in each month of the year. For each flux, the convention is that downwards=positive.

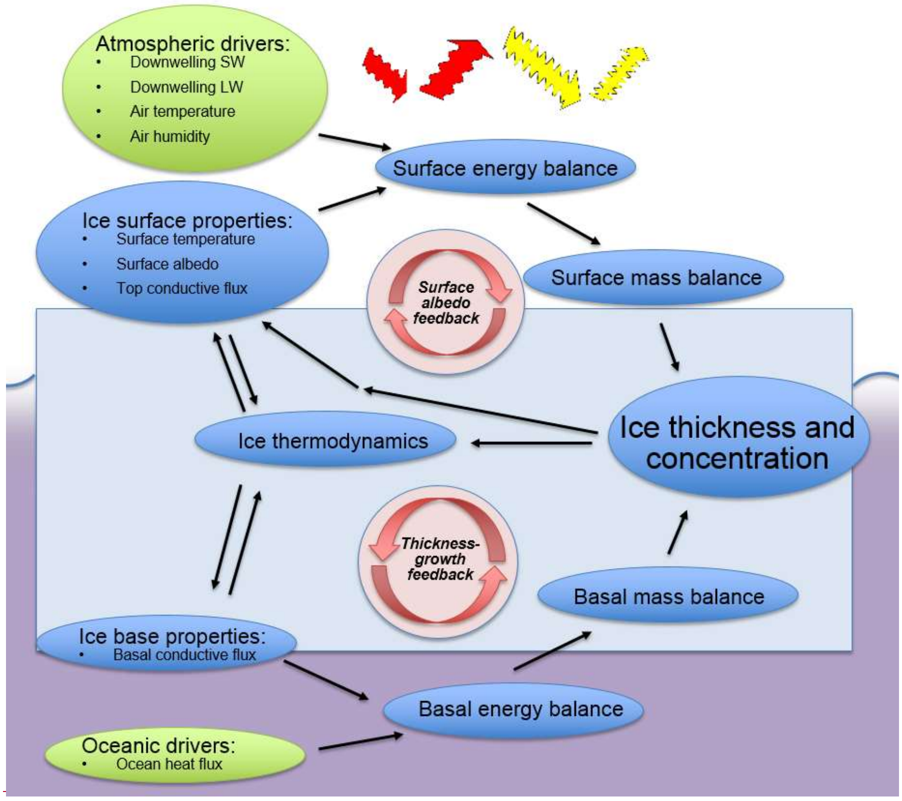
North Pole region	Top melt flux (Wm^{-2})		Top conductive flux (Wm^{-2})		Basal conductive flux (Wm^{-2})		Ocean heat flux (Wm^{-2})	
Number of observations	196		170		193		165	
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
January	0.0	0.0	-17.7	7.4	-14.3	4.5	0.2	2.9
February	0.0	0.0	-18.9	5.7	-12.2	6.8	-0.2	6.1
March	0.0	0.0	-17.2	4.6	-11.5	6.3	1.3	3.6
April	0.0	0.0	-8.3	3.0	-8.5	4.8	1.4	2.4
May	0.1	0.1	-1.4	1.7	-7.0	2.6	2.5	4.9
June	12.4	8.3	4.4	1.7	-2.7	1.3	7.6	9.7
July	23.4	13.9	0.9	0.9	0.3	1.3	12.5	8.1
August	8.0	6.0	-1.4	3.4	1.0	0.8	13.1	10.2
September	0.2	0.3	-7.7	5.4	0.8	2.1	5.4	6.8
October	0.0	0.1	-18.1	12.7	0.3	2.3	0.3	3.2
November	0.0	0.0	-21.4	11.4	-6.4	4.4	1.4	3.2
December	0.0	0.1	-17.7	5.9	-12.7	3.4	0.5	3.0

Table 2. Mean and standard deviations of fluxes measured from the IMB data in the North Pole region, in Wm^{-2} in each month of the year. For each flux, the convention is that downwards=positive.

Beaufort Sea region	Top melt flux (Wm^{-2})		Top conductive flux (Wm^{-2})		Basal conductive flux (Wm^{-2})		Ocean heat flux (Wm^{-2})	
Number of observations	189		173		202		190	
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
January	0.0	0.0	-15.5	3.3	-14.9	4.6	2.1	5.0
February	0.0	0.0	-15.3	4.9	-13.7	3.5	1.2	2.9
March	0.0	0.0	-11.7	3.8	-12.5	2.3	1.9	6.6
April	0.0	0.0	-6.8	2.5	-9.7	2.1	2.4	2.6
May	1.8	1.6	0.8	1.7	-5.2	1.4	4.7	1.9
June	26.0	10.2	2.9	1.3	-1.6	1.6	18.0	21.1
July	41.1	17.3	1.3	0.9	0.4	0.7	30.6	19.6
August	8.7	8.3	1.6	0.7	0.7	0.9	33.1	34.9
September	0.3	0.3	-4.8	2.8	0.4	1.9	13.3	13.0
October	0.0	0.0	-12.3	3.2	-7.9	16.8	10.4	18.9
November	0.0	0.0	-16.1	3.4	-11.4	12.2	6.3	8.8
December	0.0	0.0	-19.0	5.4	-13.6	7.4	1.7	6.8

Table 3. Mean and standard deviations of fluxes measured from the IMB data in the Beaufort Sea region, in Wm^{-2} in each month of the year. For each flux, the convention is that downwards=positive.

Formatted: Font: Bold

1 

Atmospheric drivers:
• Downwelling SW
• Downwelling LW
• Air temperature
• Air humidity
Surface energy balance
Ice surface properties:
• Surface temperature
• Surface albedo
• Top conductive flux
Surface albedo feedback
Surface mass balance
Ice thermodynamics
Ice thickness and concentration
Thickness-growth feedback
Basal mass balance
Ice base properties:
• Basal conductive flux
Basal energy balance
Oceanic drivers:
• Ocean heat flux

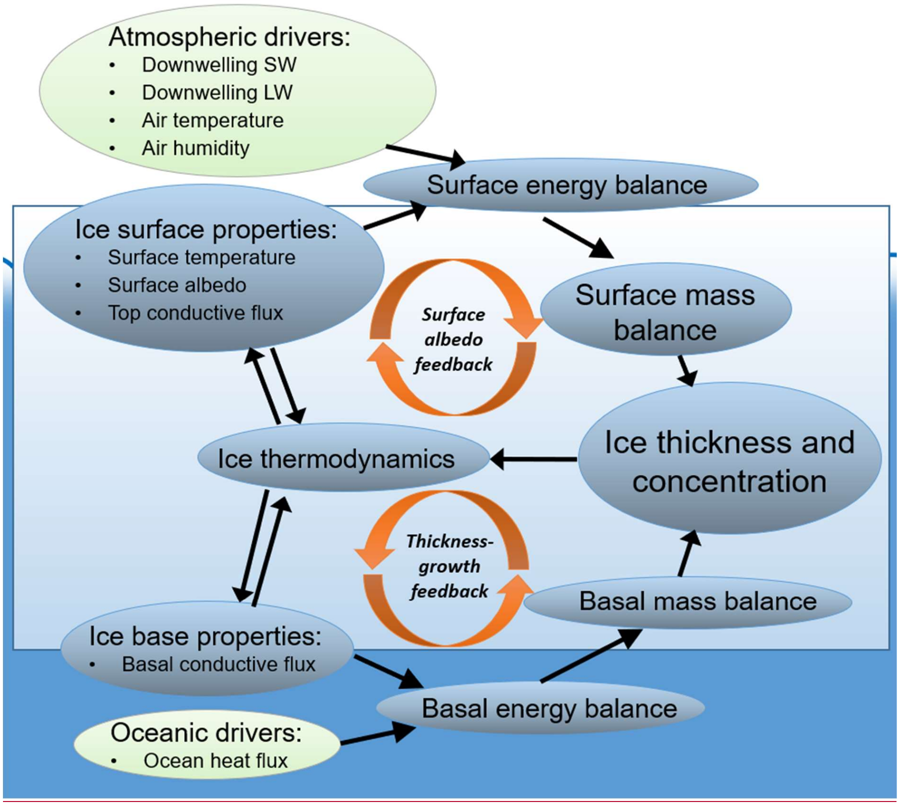


Figure 1. Schematic demonstrating the causal links between ice thickness and extent, and energy and mass balance at the top and basal surfaces of the sea ice.

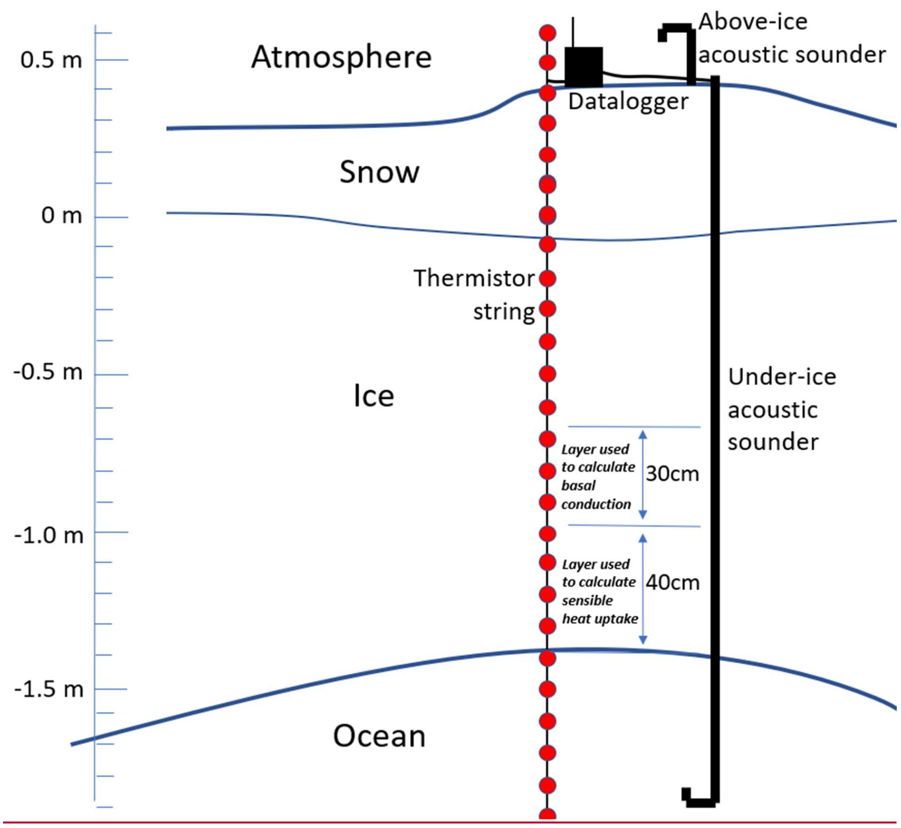


Figure 2. Diagram of the main components of an IMB, with layers used in this study for calculation of fluxes at the base of the ice indicated. Adapted from a diagram by the data providers at http://imb-crrel-dartmouth.org/imb/.

Formatted: Font: Bold, Not Italic

Formatted: Font: Bold, Not Italic

Formatted: Font: Bold, Not Italic

Formatted: Font: Bold, Not Italic

Formatted: Font: Bold, Not Italic

Formatted: Font: Bold

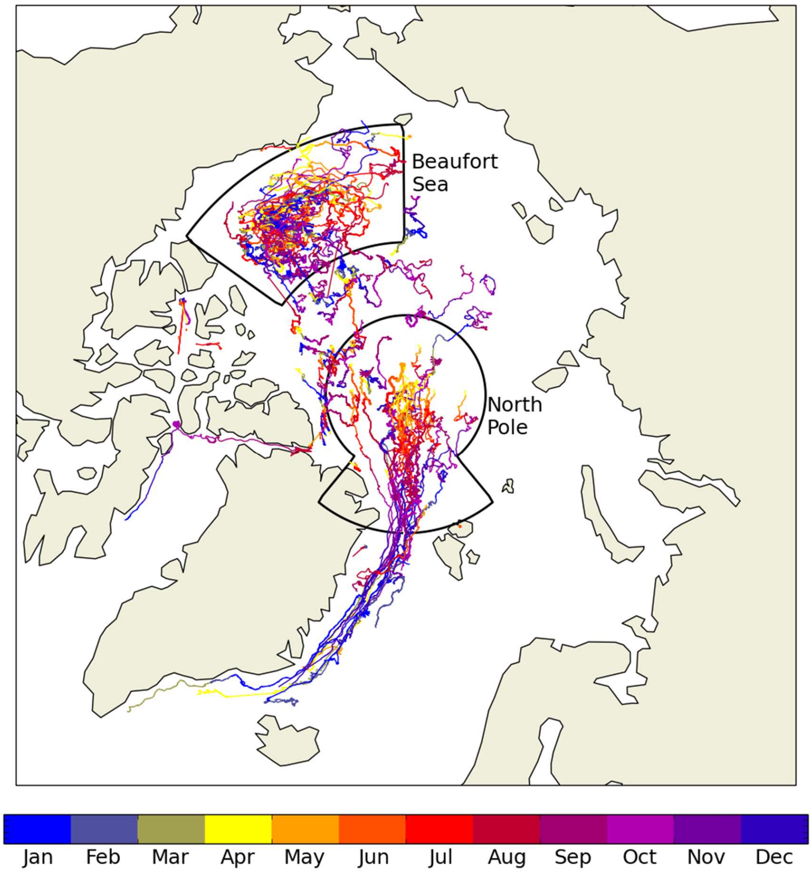
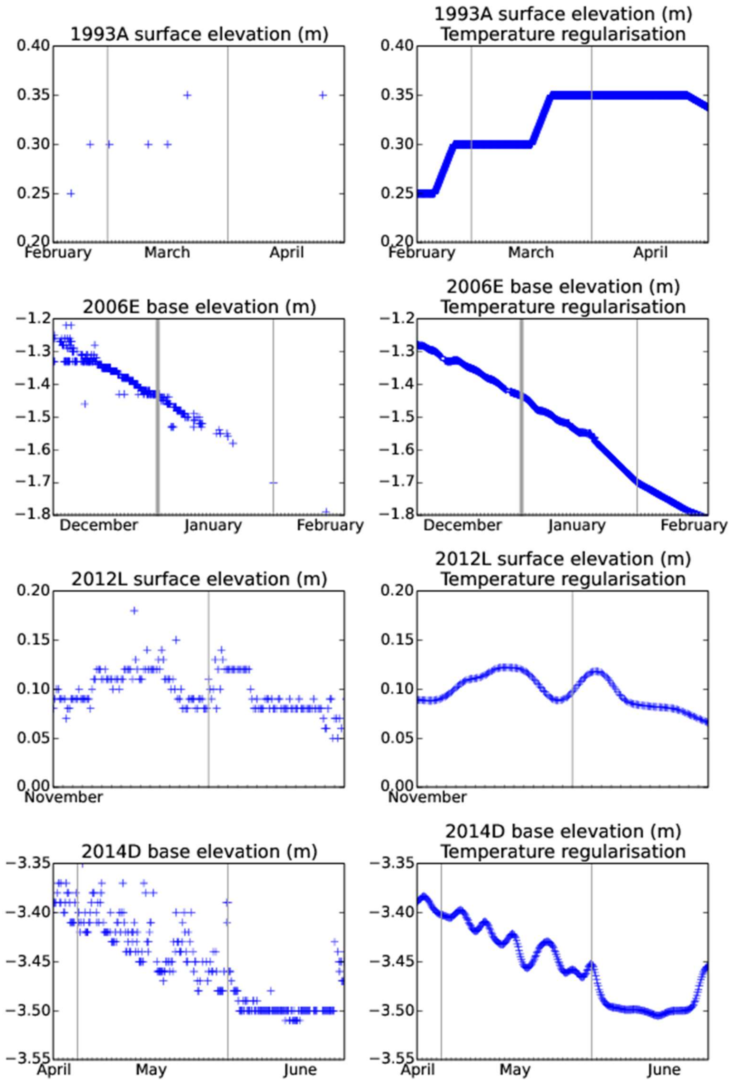


Figure 3~~2~~. The tracks of Arctic ice mass balance buoys from 1993-2015, with months of coverage indicated by the coloured shading. The North Pole and Beaufort Sea regions used in the analysis are shown by the thin black lines.

1


1993A surface elevation (m)
1993A surface elevation (m)
Temperature regularisation
2006E base elevation (m)
2006E base elevation (m)
Temperature regularisation
2012L surface elevation (m)
2012L surface elevation (m)
Temperature regularisation
2014D base elevation (m)
2014D base elevation (m)
Temperature regularisation
February
March
April
December
January
February
November
April
May
June

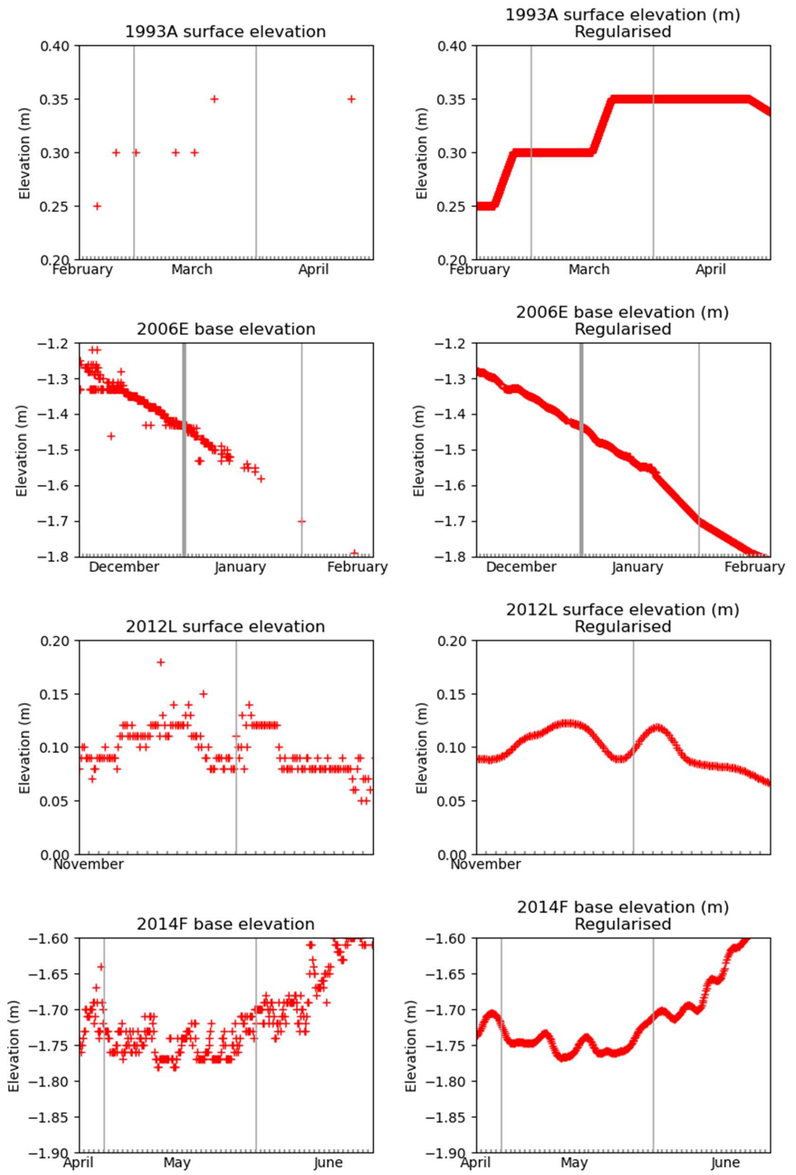
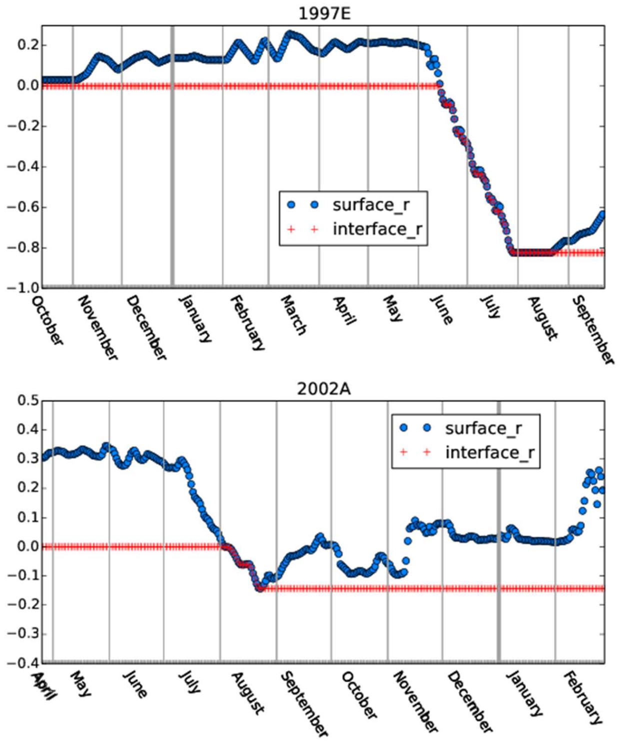


Figure 4~~3~~. Illustration of the regularisation process using four selected IMB data series. (Left) raw data; (right) time series regularised to temperature measurement points~~temperature regularised time series~~.

1 
1997E
0.2
0.0
−0.2
−0.4
−0.6
−0.8
−1.0
surface_r
interface_r
October
November
December
January
February
March
April
May
June
July
August
September
2002A
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
surface_r
interface_r
April
May
June
July
August
September
October
November
December
January
February

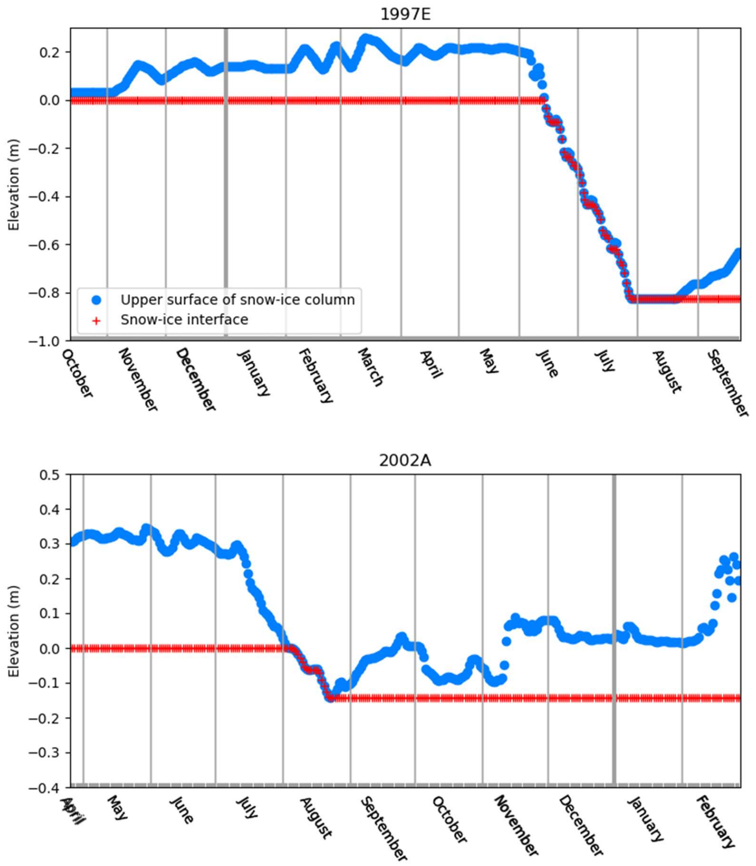


Figure 54. Two examples of estimating snow-ice interface from a regularised snow surface data series. The interface remains at a constant level unless the surface falls below this level, in which case the interface falls with the surface.

Formatted: Justified, Space After: 1.2 line, Line spacing: 1.5 lines

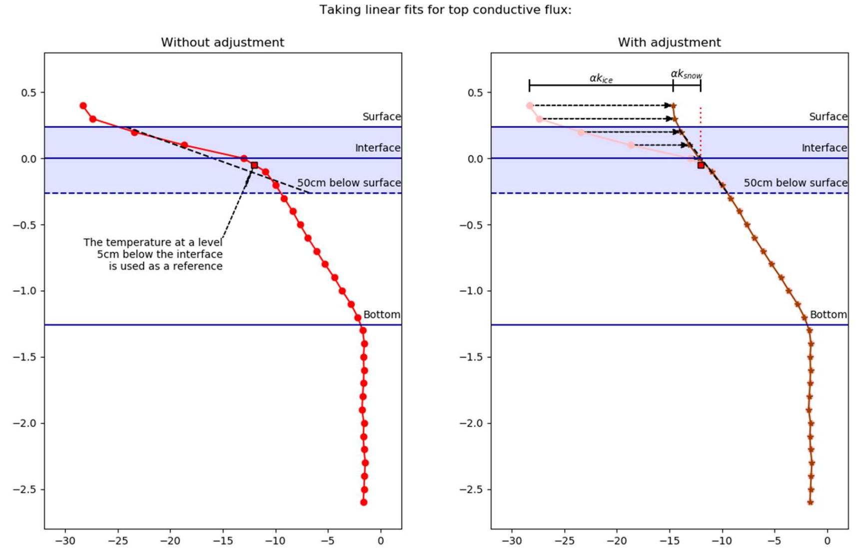
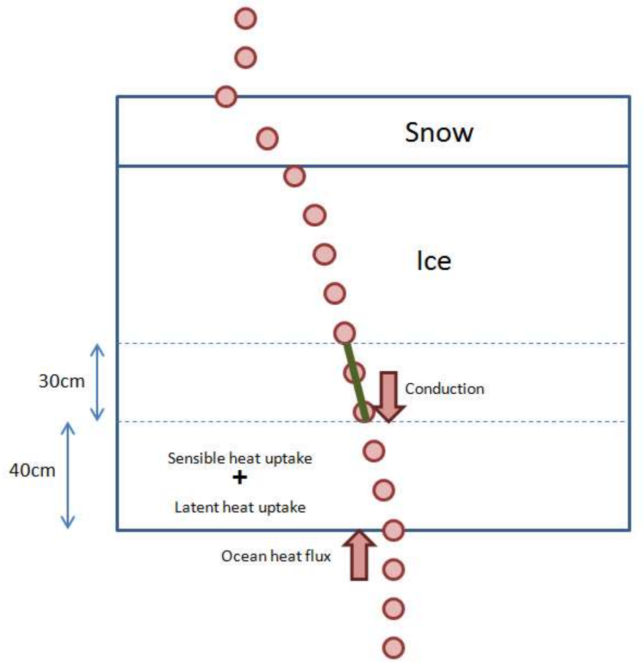


Figure 6. Illustrating the process of estimating conductive flux across the top 50cm of the snow-ice column, in the case that the snow-ice interface lies within this layer. The left panel shows the raw temperature profile: taking a linear fit through these points does not produce a meaningful result because of the sharp 'corner' associated with the change in medium. The right panel shows the adjusted temperature profile: the temperatures that would be expected if the snow layer were ice, temperature below the interface and conductive fluxes remaining the same. The adjusted profile eliminates the corner, and a linear fit can be taken.

Formatted: Font: Bold

Formatted: Font: Bold



~~**Figure 5. Illustrating the calculation of basal conductive flux and ocean heat flux as described in Section 2, after Lei et al (2014), with the use of a 'buffer zone' in the lowest 40cm of the ice above which the conduction is measured.**~~

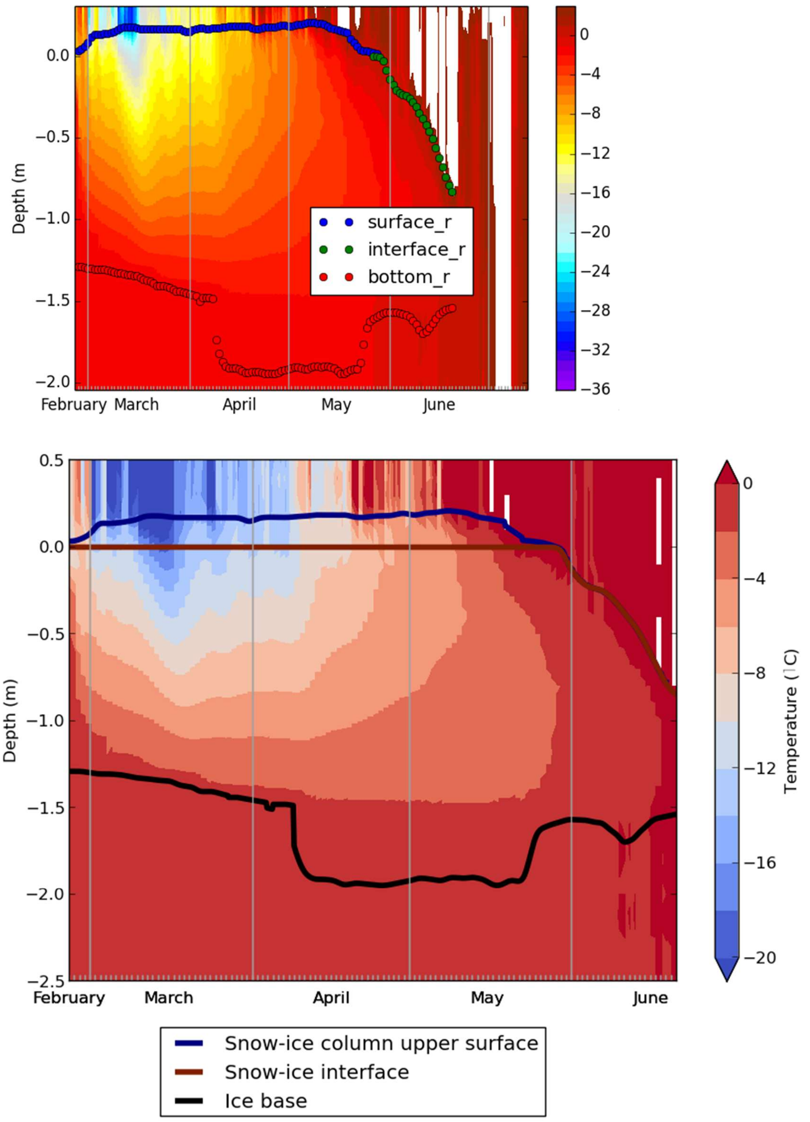
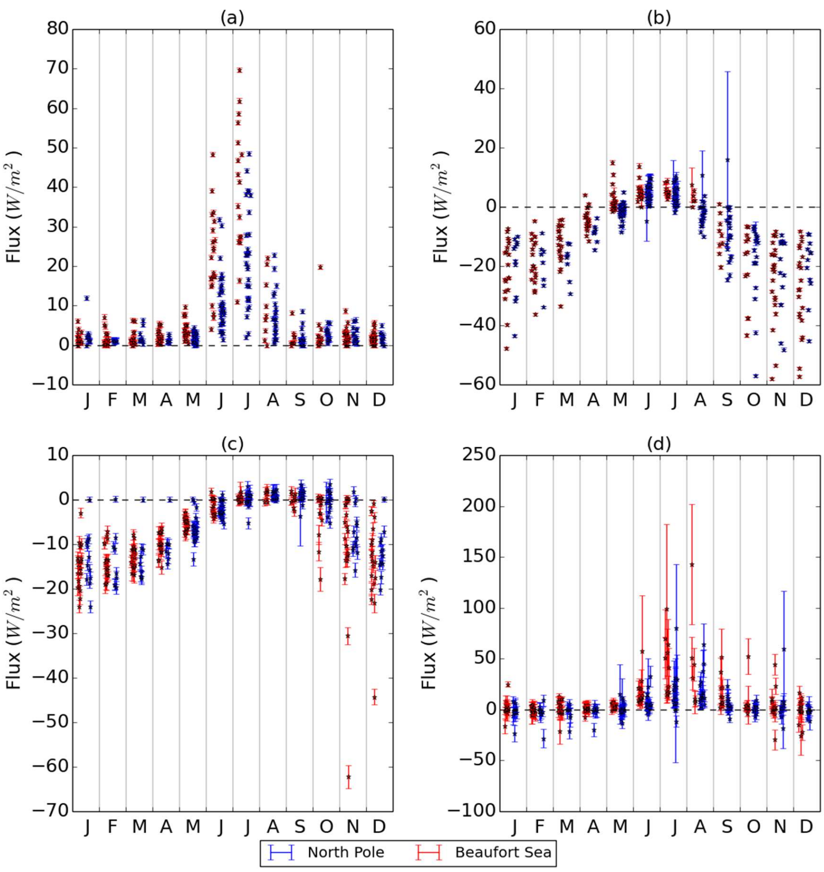


Figure 76. Measurements from the buoy 2015A, with a likely 'false bottom' being measured from early April to late May, characterised by sudden step changes in estimated base elevation.

1 

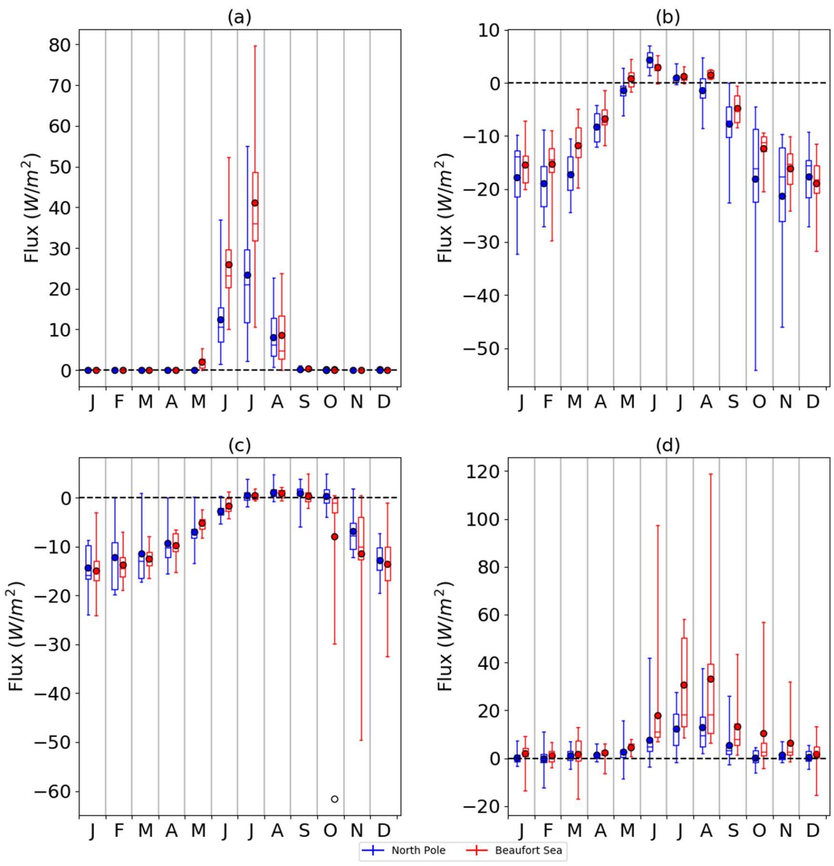


Figure 7~~7~~. Fluxes of (a) top melting, (b) top conductive flux, (c) basal conductive flux and (d) ocean heat flux, estimated from the IMB data, shown for North Pole (blue) and Beaufort Sea (red) regions. For each month, flux and region, the distribution is indicated by a boxplot showing range, interquartile range, median (horizontal lines) and mean (filled circles)~~Mean and uncertainty are indicated for each data point~~. For all fluxes, the convention is that downwards=positive.

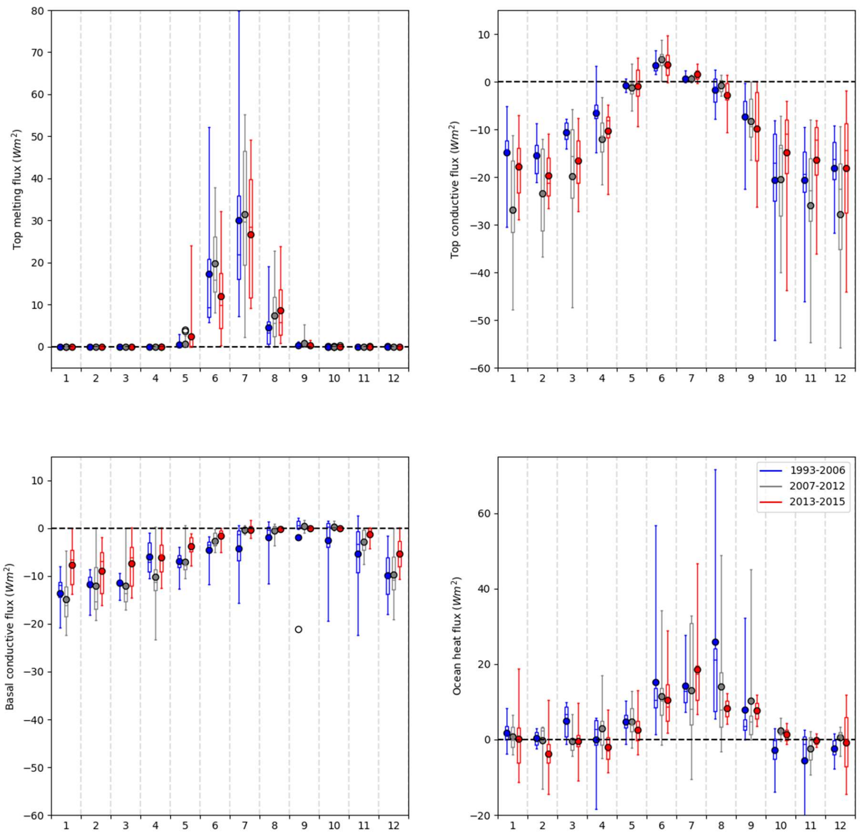
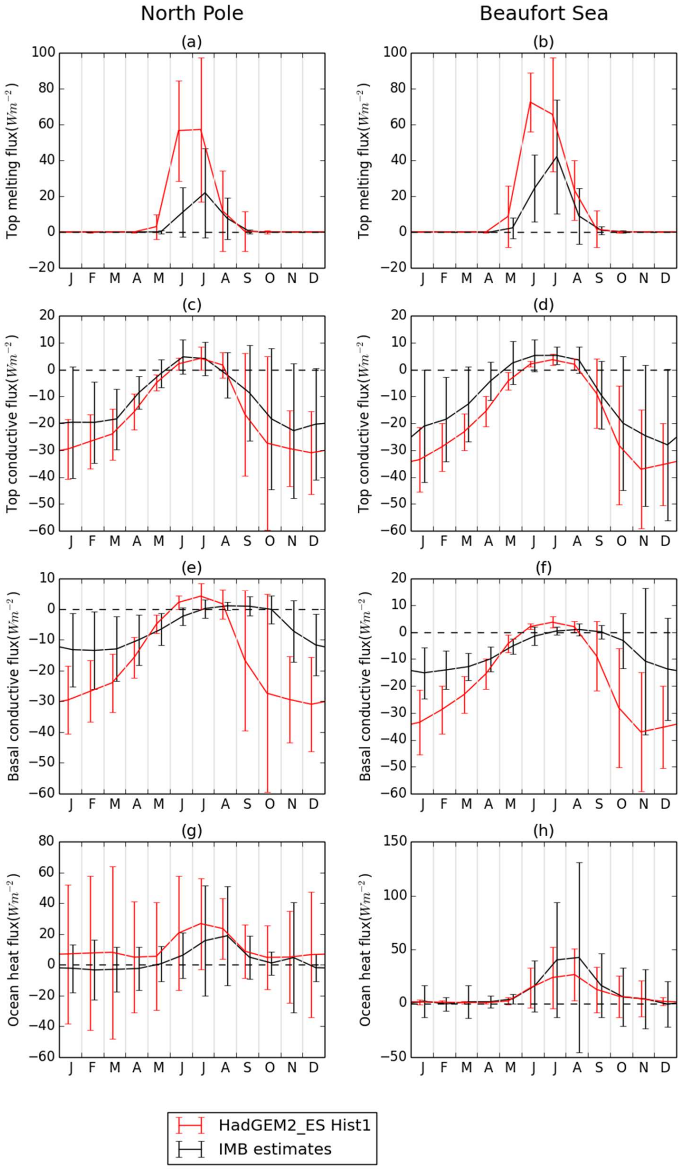
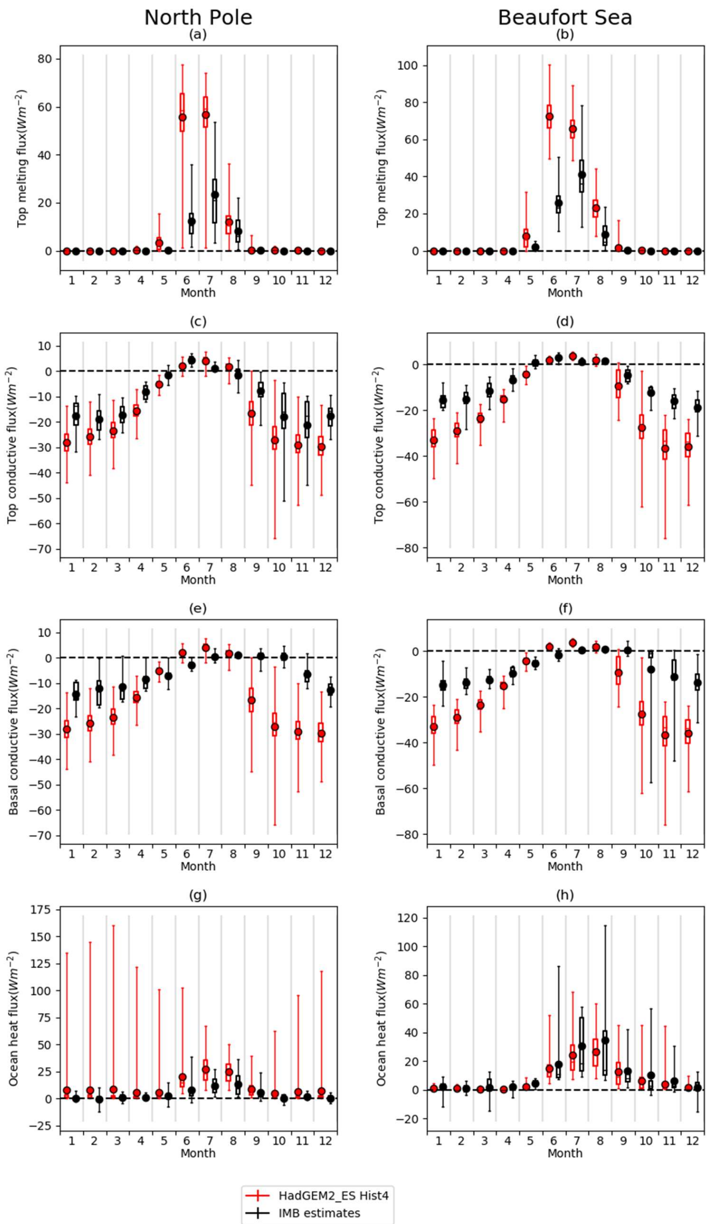


Figure 8. IMB-measured distributions of (a) top melting; (b) top conductive flux; (c) basal conductive flux and (d) ocean heat flux, divided into the three periods 1993-2006, 2007-2012 and 2013-2015. Each distribution is illustrated with a boxplot showing range, interquartile range, median (horizontal lines) and mean (filled circles).


North Pole
Beaufort Sea
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
Top melting flux(Wm^{-2})
Top conductive flux(Wm^{-2})
Basal conductive flux(Wm^{-2})
Ocean heat flux(Wm^{-2})
J F M A M J J A S O N D
HadGEM2_ES Hist1
IMB estimates

1



North Pole
Beaufort Sea
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
Top melting flux(Wm^{-2})
Top conductive flux(Wm^{-2})
Basal conductive flux(Wm^{-2})
Ocean heat flux(Wm^{-2})
Month
HadGEM2_ES Hist4
IMB estimates

Figure 98. Comparing distributions of (a,b) top melt, (c,d) top conductive flux, (e,f) basal conductive flux and (g,h) ocean heat flux from HadGEM2-ES (red) to those estimated from the IMB data (black), for the North Pole (left) and Beaufort Sea (right) regions.

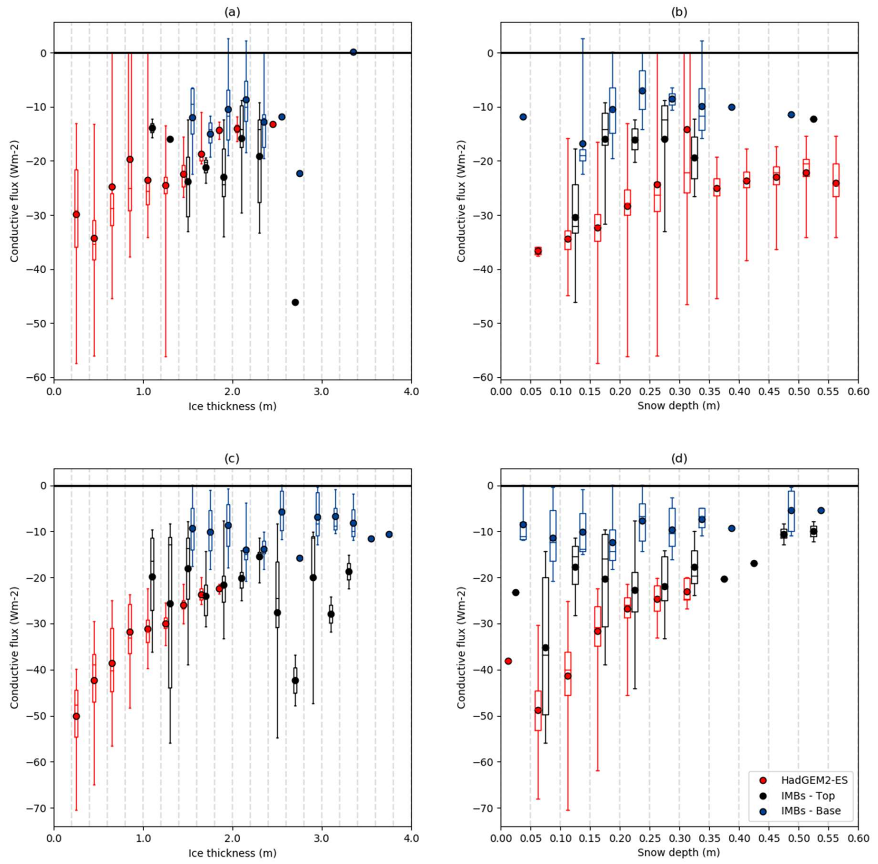


Figure 10. Conductive fluxes plotted according to ice thickness (left column) and snow depth (right column) for the North Pole (top) and Beaufort Sea (bottom) regions.

Formatted: Font: Bold, Not Italic

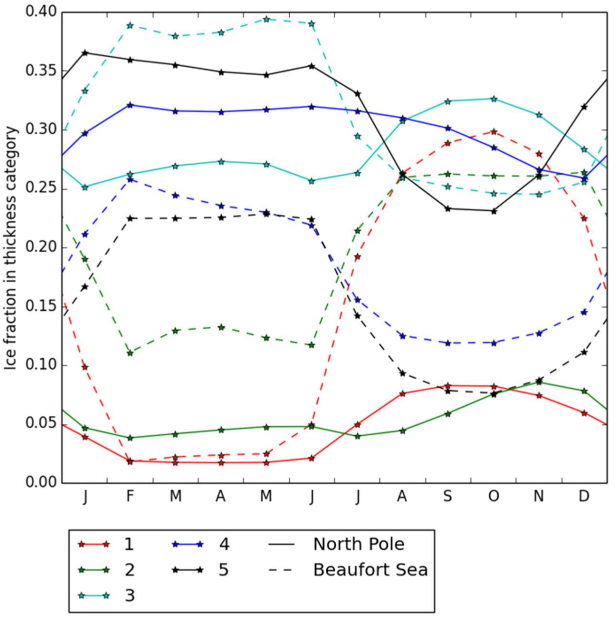


Figure A1. Ice fraction in the 5 thickness categories defined in Appendix A, estimated from submarine data compiled by Rothrock et al. (2008) for the North Pole (solid lines) and Beaufort Sea (broken lines) regions.