New methods for evaluating sea ice in climate models based on energy budgets

Submitted by Alexander Edward West to the University of Exeter

as a thesis for the degree of

Doctor of Philosophy in Mathematics

In March 2021

I certify that all material in this thesis which is not my own work has been identified and that any material that has previously been submitted and approved for the award of a degree by this or any other University has been acknowledged

Signature: Alex West ……..

Abstract

Arctic sea ice plays a vital role in the Earth's climate system, through its reflection of solar energy and insulation of ocean heat, and has changed rapidly in the past 20 years. Model simulations of Arctic sea ice display a wide spread both in the present-day and in the future. Due to lack of observations however, evaluation of sea ice simulation has historically been limited in scope mainly to ice extent and (sometimes) volume, with little attempt to evaluate at large scale simulation of the fundamental thermodynamic processes governing sea ice growth and melt. In this thesis two new, contrasting methods are presented for evaluating Arctic climate simulation that address this: firstly, the induced surface flux (ISF) framework attributes model biases (differences) to specific proximate drivers using existing reference datasets. Secondly, the Arctic ice mass balance buoy (IMB) network is used to build a dataset with which to evaluate many sea ice thermodynamic processes directly.

We use three UK CMIP models for analysis: HadGEM2-ES, HadGEM3-GC3.1 and UKESM1.0. These models display very different Arctic sea ice simulations, with ice in HadGEM2-ES thinnest and ice in UKESM1.0 thickest. Using the ISF framework and IMB evaluation, it is shown that modelled sea ice volume is tightly coupled to modelled ice growth and melt, and that most of the model biases and differences are caused by differences in albedo and atmospheric forcing arising in the late spring and early summer. Despite this, a downwelling longwave radiation bias present in all models during winter 'predisposes' them towards a thicker ice cover. The methods can also be used to evaluate the proximate impact of specific model improvements in the latter two models on sea ice growth and melt, which is seen to be small but non-negligible. The results also show that more accurate observations of Arctic radiative fluxes, and of snow area and thickness, would be particularly useful in improving model evaluation, and that ice mass balance buoy measurements would be made more useful by simultaneous measurement of salinity.

Table of contents

List of illustrations

List of tables

Acknowledgments

This research was funded by the Met Office, via the Hadley Centre Climate Programme (HCCP) and the European Commission, H2020 Research & Innovation programme (APPLICATE, grant no. 727862).

I would like to thank University of Exeter supervisors Mat Collins and Bob Beare for their help and support during this project, and examiners James Screen and Len Shaffrey for their detailed reading of the thesis, and many helpful and constructive suggestions for improvement.

I would also like to note the help and support of successive Met Office managers Helene Hewitt (initial research direction and idea of investigating IMBs), Ann Keen (support during a particularly difficult period) and Ed Blockley (detailed reviewing of repeated drafts of the final thesis and of related papers published in 2019 and 2020).

I would also like to thank Alejandro Bodas-Salcedo and Jeff Ridley for help in evaluating Arctic cloud properties.

Publication list

The following papers are substantially based on work appearing in this thesis:

West, A., Collins, M., Blockley, E., Ridley, J., & Bodas-Salcedo, A.: Induced surface fluxes: A new framework for attributing Arctic sea ice volume balance biases to specific model errors. The Cryosphere, 13(7), 2001– 2022. https://doi.org/10.5194/tc-13-2001-2019, 2019

West, A., Collins, M., and Blockley, E.: Using Arctic ice mass balance buoys for evaluation of modelled ice energy fluxes, Geosci. Model Dev., 13, 4845–4868, https://doi.org/10.5194/gmd-13-4845-2020, 2020.

1. Introduction

1.1 Overview

Sea ice is a layer of frozen seawater (usually 1-3m thick) that covers most of the polar oceans. In the Arctic, sea ice coverage reaches its maximum extent in March and its minimum extent in September. Arctic sea ice extent has decreased in all seasons over the past 40 years, as measured by satellite passive microwave observations, and the decline is particularly steep in late summer (Stroeve et al., 2012a). Although ice thickness is harder to measure, available observations suggest ice is becoming thinner also (Lindsay and Schweiger, 2015).

Arctic sea ice affects the wider climate in multiple ways. It has a higher albedo (i.e., it is brighter) than the ocean beneath, and its presence reduces the absorption of shortwave (SW) radiation from direct sunlight during the summer. The magnitude of the associated 'sea ice albedo feedback' is one of the main factors influencing climate sensitivity, as melting sea ice causes more SW radiation to be absorbed, hence inducing more warming (e.g. Winton, 2008). During the winter, sea ice insulates the atmosphere from oceanic heat, causing the Arctic atmosphere to be colder than would otherwise be the case, and increasing thermal contrast with latitude (Rinke et al., 2006). Sea ice also affects ocean circulation, firstly by release of salt during formation and release of fresh water during melting, and secondly by inhibiting momentum transfer from the wind to the ocean surface (Timmermans and Marshall, 2020).

Climate model simulations of present-day and future Arctic sea ice display a very large spread; in the CMIP5 ensemble, both present-day sea ice extent and volume display inter-model spread of greater than 15% of observed values (Shu et al., 2015). The modelled present-day trend in Arctic sea ice extent is, in general, lower than that observed (Stroeve et al., 2012b), even as a function of global mean surface temperature trend (Notz et al, 2020); whether this is due to internal variability is disputed (Notz, 2015; Swart et al., 2015; Rosenblum and Eisenman, 2017). Strong relationships between present-day sea ice volume and future rate of ice loss have been noted (e.g. Bitz and Roe, 2004, Massonnet et al., 2018); hence to predict future Arctic sea ice loss it is vital to model present-day Arctic sea ice correctly. However, a correct Arctic sea ice

extent, or even volume simulation, is not sufficient, as it can be obtained due to cancelling model errors. In order to understand the validity of future predictions, the principal drivers of a model sea ice state must also be understood.

Sea ice state is driven, via the energy and mass balances, by fundamental properties of the atmosphere and ocean, such as fluxes of downwelling radiation, and oceanic heat convergence. However, the sea ice state itself affects how it responds to these variables. For example, a given bias in downwelling longwave radiation affects a thin sea ice column more than a thick sea ice column, for reasons that are explained in more detail below.

Inspired by this conceptual picture, new methods are developed in this thesis to evaluate sea ice simulation, by evaluating fundamental aspects of the mechanisms driving sea ice state: the top and basal mass balance, the surface energy balance, and the internal ice thermodynamics. This is performed by introducing two innovations: the 'induced surface flux' (ISF) framework, which allows proximate drivers of biases in the sea ice volume balance to be diagnosed, and use of the Arctic Ice Mass Balance buoy (IMB) network to evaluate internal ice thermodynamics, and surface and basal mass balance, directly. These methods are applied in turn to historical simulations of Arctic sea ice in two contrasting coupled models. In this way, a detailed picture emerges of the mechanisms by which large-scale biases in the sea ice state arise.

1.2 Arctic sea ice: Properties, drivers and effects

Sea ice can be found in two regions of the world, the Arctic and Antarctic. As the focus of this thesis is Arctic sea ice evaluation, it is principally Arctic sea ice discussed in this section. Arctic sea ice mainly forms in the Arctic Ocean, a body of water mostly enclosed between the North American and Asian continents, but may form in other seas north of 45°N during the Northern Hemisphere Winter. Figure 1.1 shows the position of the Arctic Ocean relative to the surrounding continents, along with the names and positions of many smaller regions of the Arctic discussed in the thesis. The blue shaded region represents the 'Arctic Ocean' region used for model evaluation during Chapters 3 and 4.

Figure 1.1. A map of the Arctic, indicating the positions of all regions referred to in this thesis. The blue shaded region indicates the 'Arctic Ocean' region used for model evaluation in Chapters 3 and 4.

Sea ice forms when sea surface temperature falls below the freezing temperature (usually around -1.8°C). Initially, small ice crystals known as frazil ice form within a supercooled layer at the surface of the ocean; eventually, these freeze together to form a continuous sheet. Under cold atmospheric conditions, the sheet grows rapidly downwards in a process called congelation growth, driven by a negative energy balance at the base: strong upwards energy conduction is greater than the flux of heat from the ocean beneath. As the ice column grows, however, the temperature gradient becomes shallower, and conduction becomes weaker (the thickness-growth feedback). Eventually the ice thickness approaches a thermodynamic equilibrium (Maykut and Untersteiner, 1971), usually 2-3m thick. A layer of snow on the ice can significantly attenuate sea ice growth, as snow is a very powerful insulator, with thermal conductivity up to 10 times lower than that of ice. Conversely, mechanical deformation of ice can create much thicker 'pressure ridges', of up to 20m thick.

In summer, fluxes of incoming shortwave (SW) and longwave (LW) radiation at the surface increase; the surface flux turns positive and the surface temperature begins to increase. As the surface temperature reaches the melting point, a series of positive feedbacks, known collectively as the surface albedo feedback, begin to operate. Snow can melt and drain from the ice surface, leaving bare ice behind, which has a lower albedo; surface melt of snow and ice can also form meltponds on the ice surface, which have an even lower albedo. Both processes have the effect of increasing SW absorption, and speeding ice melt still further. Finally, as an ice column melts away entirely, the very low albedo of the darker ocean means that even more energy is absorbed by the surrounding waters, raising the temperature of the top ocean layer and increasing melting of the surrounding ice from the base and sides.

In the Arctic, sea ice does not melt away completely every summer, although ice extent at the end of summer has decreased significantly over the past 40 years. Instead, the ice reaches its minimum extent in September, when approximately half of the Arctic Ocean remains ice-covered. Hence ice can survive for many years in the Arctic, undergoing multiple freezing/melting cycles. Although there are variations from year to year, most surviving ice tends to be found in the North Pole region, and in the regions north of Greenland and the Canadian Archipelago (Figure 1.2). In most of the seas north of Alaska and Siberia, ice tends to melt away completely every summer. This distribution is partly due to atmospheric thermodynamics, but also due to ice dynamics, as described below.

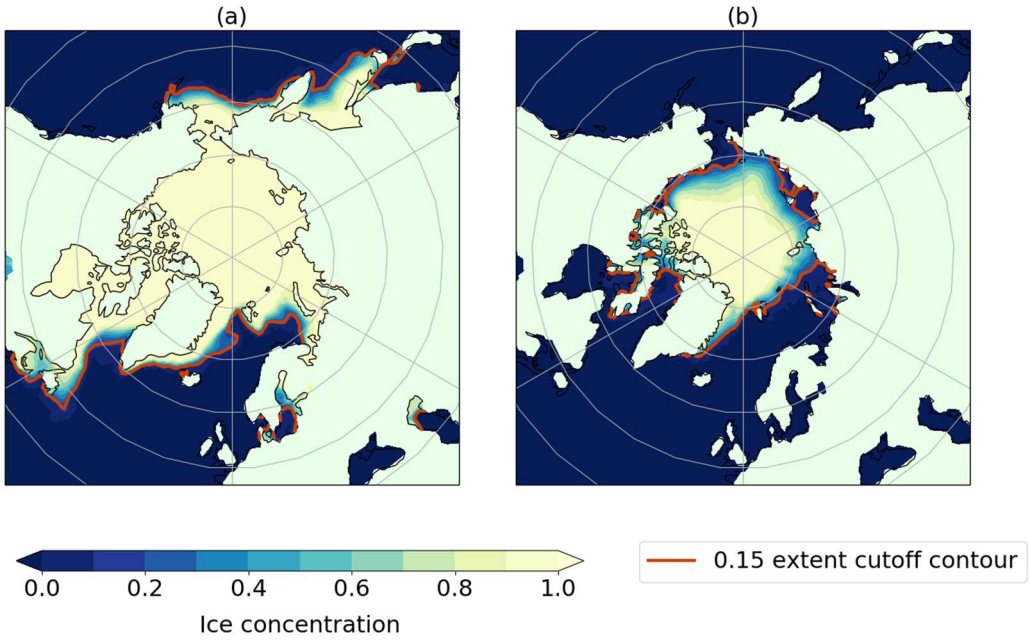


Figure 1.2. 1981-2010 average sea ice concentration in (a) March; (b) September, according to the HadISST.2.2 dataset (Titchner and Rayner, 2014). 'Concentration' means the fraction of area in each grid cell covered by sea ice. In each panel, the position of the 0.15 contour (the level usually used to delineate sea ice extent) is indicated in red.

Sea ice is not fixed in one place. Wind forcing on a given ice floe is often comparable in magnitude to the force of the internal ice stresses; hence ice floes can frequently crack and move with the wind. Other forces acting on moving ice include oceanic drag, the Coriolis force, and a component arising from sea surface tilt. In the Arctic, the prevailing pattern of atmospheric circulation is characterised by a weak high in the Beaufort Sea, stronger in winter than in summer, and a cyclonic region near the Atlantic sector (the northern end of the Atlantic storm track), known as the Beaufort High and the Icelandic Low respectively. The resulting winds create two large-scale, semi-permanent patterns of ice motion: the Transpolar Drift Stream, which transports ice across the Arctic from the Siberian shelf to the Fram Strait; and the Beaufort Gyre, which circulates ice around the Pacific sector of the Arctic (Figure 1.3). The combination of these patterns tends to create divergent ice motion, and hence thinner ice, on the Siberian side of the Arctic, and convergent ice motion, and hence thicker ice, on the Canadian side of the Arctic (Figure 1.4).

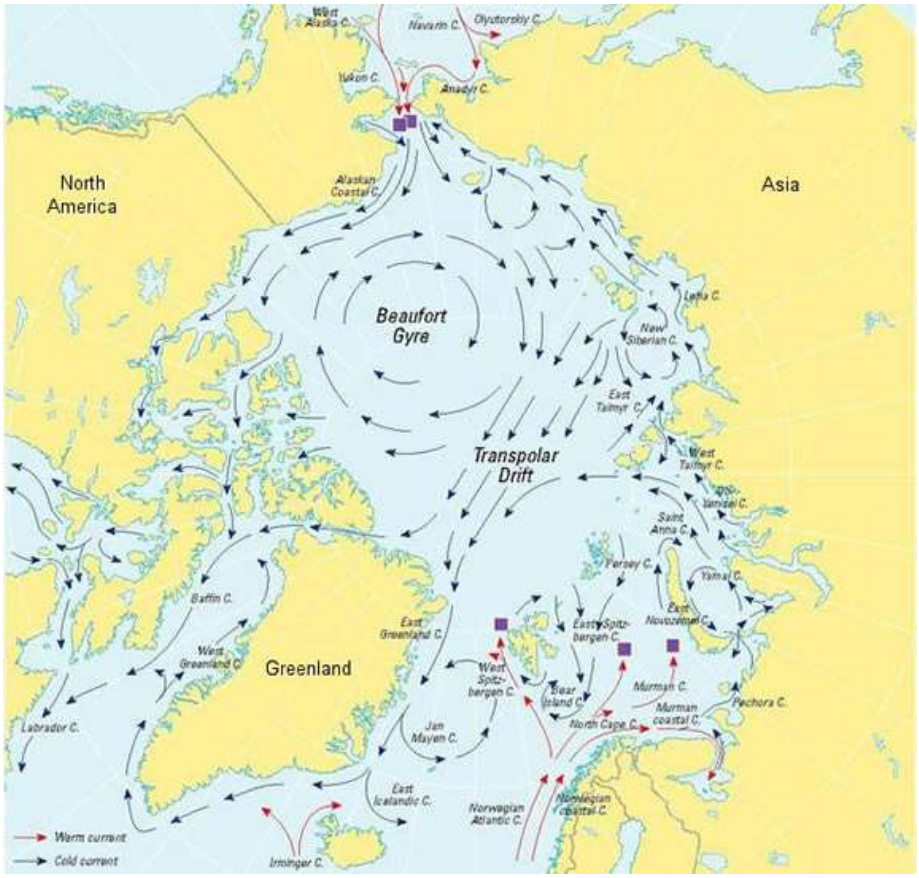


Figure 1.3. Large-scale patterns of sea ice circulation in the Arctic Ocean. Image courtesy of Arctic Monitoring and Assessment Programme (AMAP), Figure 3.29, AMAP (1998).

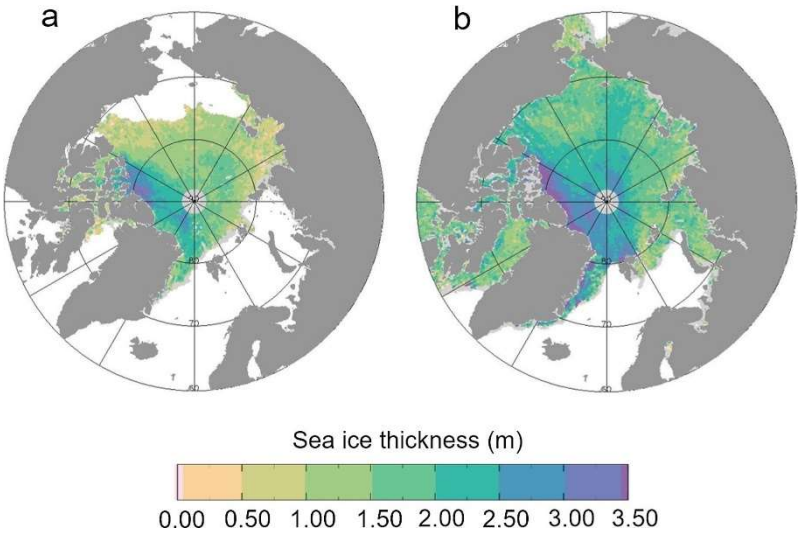


Figure 1.4. An illustrative map of ice thickness in the Arctic, to indicate the usual distribution of thinner and thicker ice. The example given is of ice thickness in (a) October-November 2016 and (b) in March-April 2017, as measured by radar altimetry from CryoSat-2. The image is figure 14 of Tilling et al. (2018).

Having described the sea ice component of the Arctic climate system, the atmosphere and ocean components are now briefly described in turn. The Arctic atmosphere in winter is in the main characterised by very clear, dry conditions, thin ice clouds sometimes being present near the surface (Curry et al., 1996). Associated with these conditions are a pronounced near-surface temperature inversion, and very cold surface temperatures driven by LW cooling. However, these conditions are often disrupted by advection of mild, moist air from midlatitudes, usually in association with extratropical depressions. Ingress of mild air is associated with much cloudier conditions, inhibiting LW cooling and resulting in a much weaker temperature inversion and milder surface temperature (Raddatz et al., 2015). These ‘cold’ and ‘mild’ modes correlate well with the presence of sea ice growth (Stramler et al., 2011). Conversely, sea ice can affect the overlying atmospheric conditions; a mild airmass will cool more quickly over a thicker, more extensive ice pack, as areas of open ocean release heat very rapidly into the winter Arctic atmosphere.

The strength of the two semi-permanent features, the Beaufort High and the Icelandic Low, is moderated by several large-scale modes of atmospheric variability over the Arctic. Firstly, the Arctic Oscillation index (AO; Thompson and Wallace, 1998) describes the exchange of atmospheric mass between high and mid-latitudes: when the index is positive, pressure tends to be lower in the Arctic and higher in mid-latitudes, associated with a weak Beaufort High and a strong Icelandic Low. A positive Arctic Oscillation is associated with strong, zonal atmospheric flow, and reduced heat transportbetween the Arctic and midlatitudes, but has also been associated with reduced sea ice retention in the Arctic due to the weakened Beaufort high (e.g. Rigor and Wallace, 2004). A negative AO index indicates a predominance of blocking, with pressure higher in the Arctic and lower in mid-latitudes, and greater heat exchange.

A second mode of variability is the Arctic Dipole (AD; Watanabe et al., 2006; Wang et al., 2009) which describes the exchange of mass between the Siberian and North American sides of the Arctic, and whose extreme positive phase is implicated in the most severe instances of summer sea ice loss (Wang et al, 2009). When the AD is positive, pressure tends to be higher on the North American than the Siberian side of the Arctic, driving strong heattransport, and

transpolar sea ice motion towards Fram Strait, where the ice is exported and lost.

The Arctic Ocean is characterised by extensive continental shelves (covering about 40% of the total area) and two deep interior basins. The most fundamental feature of the Arctic Ocean water masses is a fresh ‘mixed layer’, of 50-300m depth (but varying seasonally), created by substantial river inflow from Canada and Siberia, and to a lesser extent advection of relatively fresh water from the Pacific Ocean via the Bering Strait (Morison and Smith, 1981; Serreze et al., 2006). The mixed layer overlies a much warmer, saltier layer of water derived from the Atlantic Ocean inflow, and to a large extent protects Arctic sea ice from this source of heat. The boundary between the mixed and Atlantic layers is known as the halocline.

In summary, Arctic sea ice mass balance is driven by a variety of complex factors. Winter ice freezing is driven by atmospheric cooling, in some places ameliorated by ocean heat, and in all places modulated by the ice thickness-growth feedback. Summer ice meltingis driven in the first instance by atmospheric heating, but is modulated by the surface albedo feedback, by which the ocean also eventually acts as a source of melting. We now discuss the frameworks which are typically used to simulate sea ice in coupled models.

1.3 Modelling sea ice

Modelling sea ice has historically presented a different challenge to that of modelling the atmosphere and ocean. This is because sea ice is not, strictly speaking, a continuum, but a collection of finite elements whose individual modelling would demand unfeasible computing resources. Hence sea ice modelling aims to simulate the behaviour of sea ice on a large scale, where the ice displays the characteristics of a non-Newtonian fluid (a fluid whose viscosity depends upon the force applied). Hibler (1979) showed that it was possible to achieve a realistic simulation of sea ice motion by modelling it as a viscous-plastic fluid. A variant of this method, the elastic-viscous-plastic rheology (Hunke and Dukowicz, 1997), is used in most modern sea ice models to simulate the large-scale motion of sea ice. The modelling of sea ice as a series

of discrete elements is an active research topic (e.g. Danilov et al., 2015), but this approach is not yet widespread amongst fully coupled models.

A fundamental prognostic variable of a sea ice model is the sea ice concentration, the fraction of any given grid cell that is ice-covered. Because processes differ greatly over sea ice and open ocean, the sea ice concentration controls the response of ocean to atmospheric forcing, and vice versa; it sets the bulk energy values that are received directly by the ice, and by the ocean. However, sea ice itself is a highly heterogeneous medium, varying greatly in thickness over ranges of several metres. It has already been seen that ice of differing thicknesses responds very differently to identical atmospheric forcing. Clearly, it is physically unrealistic to treat all sea ice in a single grid cell (which can often be up to 100km in width) as responding the same way to a given forcing. To deal with this problem, the sea ice thickness distribution was introduced (Thorndike et al., 1975).

The sea ice thickness distribution function (ITD) describes the fraction (or more accurately, the fraction density) of sea ice in any grid cell of a given thickness. Theoretically, the integral of the function between any two thicknesses is equal to the fraction of ice in the grid cell in that thickness interval. In most sea ice models, the ITD is approximated, or discretised, by dividing the ice into a finite number of thickness categories, with constant bounds. For example, the Los Alamos sea ice model (CICE) by default uses five thickness categories, which may be configured with a variety of bounds (one such being with bounds at 0m, 0.6m, 1.4m. 2.4m, 3.6m and 20m.

Some models diagnose the ice thickness distribution function from ice fraction or thickness using empirically-derived parameterisations (Castro-Morales et al., 2013). The most advanced sea ice models, however, treat the ice thickness distribution function as a prognostic variable which evolves in time with advective, thermodynamic and dynamic processes. The prognostic variable becomes category sea ice concentration, the fraction of sea ice in a grid cell of a given category. Other prognostic variables can include category mean ice thickness, and category ice energy.

The sea ice model then evolves the ITD over time by the three processes that are observed to change the ITD, and hence the sea ice state, in the real world:

ice advection, thermodynamics and ridging. Movement from a region of thicker or thinner ice will clearly alter the ITD; so will thermodynamic growth or melt of ice. Finally, ice convergence will tend to cause ice to move from the thinner to the thicker end of the ITD, by crushing thin ice floes and piling the debris into thick pressure ridges.

Ice thermodynamics is usually modelled by a simple discretisation of the heat diffusion equation, using a set number of equally-spaced layers in the ice and snow (although experiments have been performed with variably space layers). Until recently, a simpler thermodynamic scheme known as the 'zero-layer' scheme was commonplace, in which the ice and snow had no heat capacity but responded instantaneously to forcing at the upper surface.

1.4 Evaluating sea ice models

Accurate, widespread observations with which to evaluate Arctic climate simulation are limited in number. There are severe practical difficulties with collecting in situ observations over regions of ice-covered ocean; this in turn limits the accuracy of reanalysis (as there is little data to assimilate). It also hinders the effective calibration of satellite measurements, which suffer from additional problems in the Arctic: low zenith angle, strong temperature inversions which invalidate many assumptions used in radiative transfer algorithms, and a surface with mostly similar albedo to overlying cloud (e.g. Zhang et al., 2004).

The most commonly evaluated diagnostic of Arctic climate is the sea ice extent, a grid-dependent metric defined as the total area of grid cells containing more than 15% ice cover. Unlike for many other Arctic climate variables, there exist longstanding, widespread observational estimates of sea ice extent from satellite passive microwave sensors (dating mostly back to 1979), based on the differing microwave signatures of sea ice and open water. In summer, sensors cannot reliably distinguish between open water and on-ice meltponds; sensors also struggle to detect low-concentration ice. The sea ice extent is less sensitive to these errors than the sea ice area, the main reason for this diagnostic being more commonly evaluated.

Since data from the third Climate Model Intercomparison Project (CMIP3) became available, evaluation and comparison of Arctic climate projections by means of Arctic sea ice extent has been frequent and thorough. Stroeve et al. (2007) compared historical model simulations of Arctic sea ice extent to satellite observations, finding that most models considerably underestimated the observed rate of decline. Wang and Overland (2009) evaluated sea ice extent with respect to passive microwave observations to provide a metric to refine future projections of Arctic sea ice. In this way, they excluded models whose September ice extent, and extent seasonal cycle amplitude, were not sufficiently consistent with observations. Both studies were updated for the next-generation model intercomparison projects, firstly CMIP5 (Stroeve et al., 2012; Wang and Overland, 2012), and then CMIP6 (Notz et al., 2020), in each case finding that the rate of Arctic sea ice decline was still underestimated, albeit less severely than in the previous project.

In the same timeframe, studies have cast doubt upon the usefulness of sea ice extent as a reliable metric of model performance. Notz (2014) showed that measures of ice extent were strongly dependent on the underlying grid, while Notz (2015) showed that, due to internal variability, the same model might produce a very good, or a very bad simulation of ice extent loss depending upon slightly different initial conditions. Stroeve and Notz (2015) showed that accuracy of present-day sea ice simulation was unlikely to be a reliable predictor of future accuracy, by considering different historical periods; they showed also that the relationship between Arctic mean surface temperature and summer Arctic sea ice extent was, in CMIP5, broadly consistent with observations.

Sea ice volume has long been known to exert a strong influenceboth on present-day sea ice extent, and future rate of ice volume loss. Bitz (2008) showed that in the CMIP3 ensemble, present-day sea ice volume was strongly correlated with future ice volume loss, with higher initial ice volumes associated with higher rates of loss. Holland and Stroeve (2011) showed that the proportion of summer sea ice extent explained by variability in winter sea ice thickness tended to increase as the sea ice cover thinned. Massonnet et al. (2012) found the annual mean sea ice volume from 1979-2010 predicted the year in which a given simulation first modelled summer Arctic ice extent under 4 million sq km

with a correlation of 0.72. Massonnet et al (2018) also produced evidence that mean ice volume exerts strong controls on the future rate of loss of ice volume, via the surface albedo feedback and thickness-growth feedback.

Hence sea ice volume is a more useful diagnostic with which to evaluate model simulations of Arctic sea ice, if spread in future projections is to be understood. Although the observational record of sea ice volume is sparse, discontinuous and composed of multiple different data sources, the reanalysis system PIOMAS, which assimilates sea ice concentration data, is now frequently used as a reference dataset with which to evaluate sea ice thickness. Shu et al. (2015) compared CMIP5 simulations of sea ice volume to estimates from PIOMAS, finding that on average ice is too thin in the CMIP5 ensemble, more so in the summer than in the winter.

However, there remain many shortcomings in ice volume as a metric by which to evaluate ice models. Models may simulate the correct annual mean sea ice volume for the wrong reasons, particularly if the sea ice volume is a metric by which the model is tuned (Notz, 2015; Eisenman and Untersteiner, 2007). Hence still more valuable information can be gained by stepping back further; by evaluating the processes that drive ice volume change: mass balance, energy fluxes and ice thermodynamics. With a partial exception, these properties have not yet been widely evaluated in the CMIP ensembles, for two reasons. Firstly, the required diagnostics are not (for CMIP3 and CMIP5) reported as universally as ice concentration and thickness. Secondly and more importantly, there are no observational datasets sufficiently reliable and widespread to provide a reference.

The partial exception is for the surface radiative fluxes, which were evaluated for CMIP5 by Boeke and Taylor (2016) amongst others. Satellite- and reanalysis-based datasets of these do exist, although they are less accurate in the Arctic than in most other parts of the world (Lindsay et al., 2014; Zhang et al., 2004). However, a simple evaluation of surface radiative fluxes is of limited use in understanding Arctic sea ice simulation, as both sea ice cover and surface radiation influence each other to first order; surface radiation drives sea ice volume through surface energy balance and mass balance, while sea ice state drives surface radiation via the surface temperature and surface albedo.

Indeed, previous studies of Arctic surface radiation have tended to implicitly treat the sea ice as an external driver of the surface radiation, rather than viewing the two as being a closely coupled system (e.g. Karlsson and Svensson, 2013). Holland and Landrum(2015) partially addressed this problem by quantifying and separating the causes of change in Arctic net SW radiation into components associated with downwelling SW, ice area and ice albedo changes (an approach similar to the induced surface flux framework described in Chapter 4 of this thesis).

The problem in improving Arctic sea ice state evaluation, then, is twofold. Firstly, suitable reference datasets must be identified with which to evaluate the variables that drive the ice state, and that modulate how the ice state responds. Secondly, evaluation must be combined with an understanding of how the variables relate to each other, in order to disentangle model biases which are in a proximate sense independent of the sea ice state, from model biases which directly represent the sea ice state (namely, ice area, thickness and volume). In this thesis, two innovations are proposed that go some way towards addressing these questions.

1.5 Improving sea ice evaluation

Here and throughout this study, the following conceptual picture of sea ice physics is used to motivate improvements in model evaluation. Sea ice thickness is driven by the mass balance at the top and basal surfaces of the ice; hence sea ice volume is driven, in the main, by the area integral of top and basal mass balance. The mass balance, in turn, is driven by the energy balance. The energy balance arises partly from quantities that are, in a proximate sense, independent of the sea ice area and thickness, for example fluxes of downwelling radiation, or oceanic heat convergence. However, it arises also from properties which are, to first order, affected by sea ice area or volume, via the sea ice thermodynamics: conductive flux, surface temperature and surface albedo. This closed causal loop gives rise to the surface albedo feedback (SAF) and the thickness-growth feedback (TGF).

We proceed on the basis that this conceptual picture describes the processes driving sea ice growth and melt to first order. Some processes not mentioned

here, such as ice export, are neglected as being relatively small (this is justified in Section 3.1). Other processes not mentioned, on the other hand, can easily be included in this framework. For example, lateral ice melt is a similar process, from an energy balance point of view, to basal melt: heating of the ocean mixed layer causes a decrease in sea ice volume. For some purposes, therefore, the two phenomena can be considered together.

The processes, and the causal relationships, are illustrated in Figure 1.5. The quantities for which reference datasets exist are labelled, showing clearly the gaps in understanding. While ice area and to some extent volume are relatively well-evaluated, the processes driving sea ice melt and growth have been evaluated very little. Only the surface radiative fluxes have been evaluated in any large-scale sense (e.g. by Boeke and Taylor, 2016), but this is usually performed without regard in particular to their effect on sea ice, and the effect of the sea ice on them. -To the knowledge of the author, the ice mass balance, ice thermodynamics, and ocean heat flux have never been evaluated at large scale. This motivates the primary aim of this thesis: to develop consistent methods for evaluating the fundamental thermodynamic processes in the atmosphere, ocean and sea ice that govern the sea ice area and volume evolution.

The primary objectives are as follows: firstly, to devise a systematic method to combine the evaluation sea ice state and that of sea ice energy fluxes, that accounts for the main ways in which the energy fluxes affect the sea ice state and vice versa. Secondly, to increase the range of sea ice energy fluxes that can be evaluated, to provide a more complete picture of model biases in Arctic climate. Thirdly, to attribute modelled sea ice biases, in a proximate sense, to biases in specific model processes, or to specific model approximations and parameterisations. Lastly, to identify the most important variables for which reliable, large-scale observational estimates do not exist, and thereby better focus improvements in Arctic observations.

The first innovation described in this thesis is a framework (the induced surface flux, or ISF, framework) by which external drivers of sea ice mass balance change can be quantified and separated from drivers which arise from the sea ice state itself (the SAF and TGF), by using simple models to describe the

relationship between sea ice and surface energy balance at each point in model space and time. The second innovation describes a method of evaluating key sea ice variables, representing the thermodynamics and volume balance, in models, using the Arctic Ice Mass Balance buoy (IMB) network.

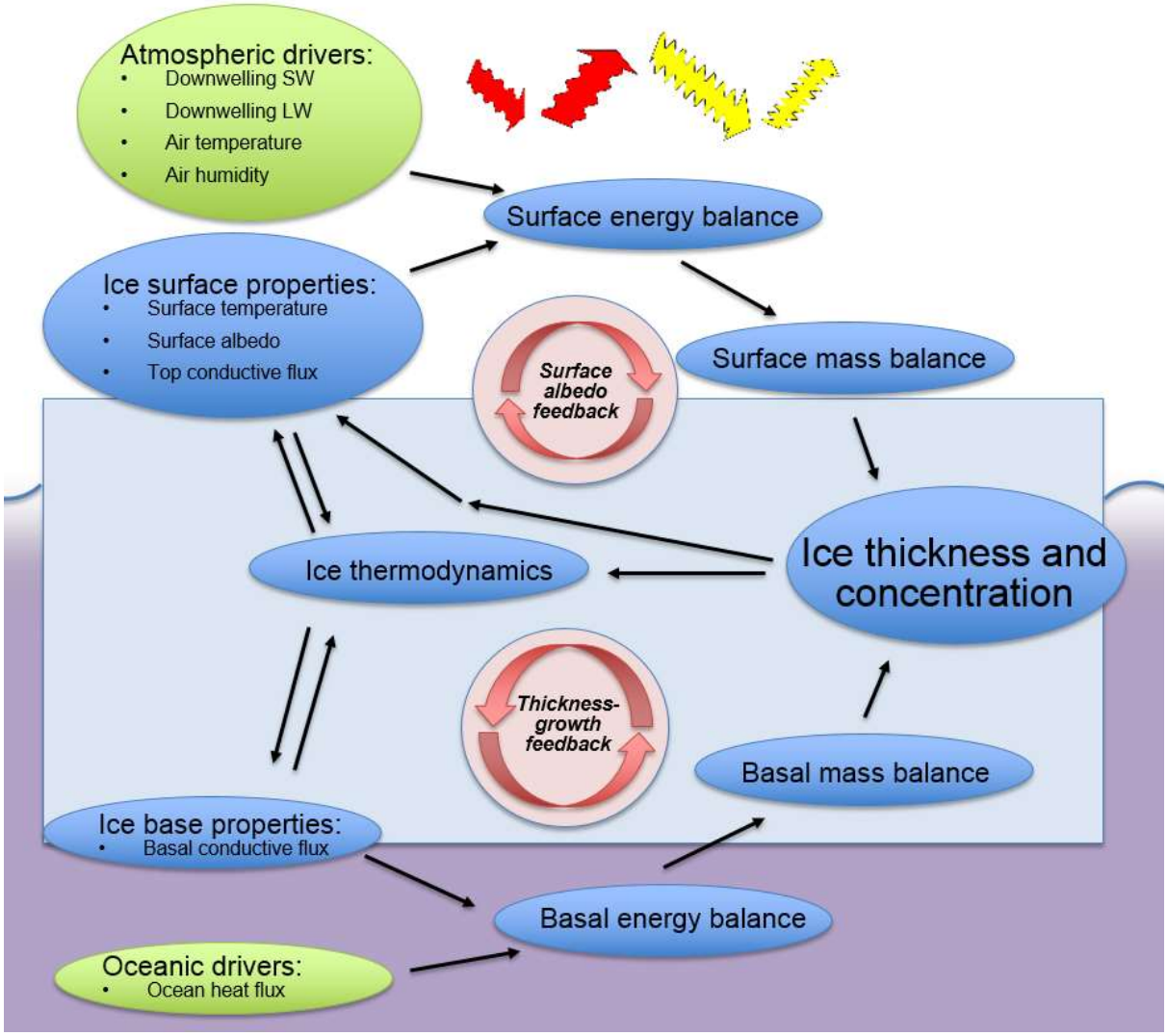


Figure 1.5. Schematic describing the most important causal relationships between Arctic climate variables influencing sea ice thickness and concentration. Image from West et al (2020).

To demonstrate these methods, we use the historical (Hist) simulations of three case study models: HadGEM2-ES (a CMIP5 model) and HadGEM3-GC3.1 and UKESM1.0 (CMIP6 models). The models are well-suited for the study because they contrast very strongly in their ice volume simulations. Using the ISF and IMB methods together, it is possible to build a detailed picture of the processes by which each model arrives at a particular sea ice state, and attribute many of these processes to specific model biases, or structural differences.

The thesis is set out in the following way. In Chapter 2, the models used in this study are described, and the reference datasets used in the study evaluated. In Chapter 3, the modelled sea ice states, along with other key Arctic climate variables are evaluated. In Chapter 4, the ISF framework is described, and applied to the case study models. In Chapter 5, the IMB-based evaluation is described, and also applied to the models. In Chapter 6, the implications of the results, and the general applicability of the methods, are discussed.

Chapter 2: Models and reference datasets

2.1 The case study models

The three coupled climate models evaluated in this thesis are HadGEM2-ES, HadGEM3-GC3.1 and UKESM1.0. HadGEM2-ES was part of CMIP5, and represents the previous generation of coupled models (Collins et al., 2011); HadGEM3-GC3.1 and UKESM1.0 are to be part of CMIP6, and represent the next generation (Kuhlbrodt et al., 2018; Sellar et al., 2019). All are fully-coupled models with interactive sea ice components, and HadGEM2-ES and UKESM1.0 employ additional components to simulate terrestrial and oceanic ecosystems, and tropospheric chemistry.

HadGEM3-GC3.1 and UKESM1.0 were run at multiple resolutions: in this thesis, we evaluate the low-resolution configurations, HadGEM3-GC3.1-LL and UKESM1.0-LL. In these configurations, the atmosphere model is integrated on the N96 grid, with a resolution of 1.25 degrees latitude and 1.07 degrees longitude. The ocean and sea ice models are integrated on the ORCA1 grid, an irregular grid with a resolution of approximately 100km in the Arctic. HadGEM2-ES was run at only one resolution, with the atmosphere model also on the N96 grid, and the ocean model on the HadGOM grid, a regular latitude-longitude grid with a resolution of one degree in polar regions. To avoid problems arising from converging meridians, the HadGOM grid includes an artificial island at the North Pole; a numerical scheme allows sea ice to be advected across this (McLaren et al., 2006). In addition, numerical smoothing is applied to some atmospheric fields above 73°N (Johns et al., 2005) and sea ice fields above 87°N (McLaren et al., 2006).

A number of features of the sea ice component are shared between all models. All employ a sub-gridscale ice thickness distribution (Thorndike et al., 1975); elastic-viscous-plastic rheology (Hunke and Dukowicz, 1997) and incremental remapping (Lipscomb and Hunke, 2004). In addition, all models share a thermodynamic framework in which the surface energy balance over ice is calculated in the atmosphere model (West et al., 2016).

However, there are some fundamental differences between the sea ice component of HadGEM2-ES, and those of the newer models. Firstly, while in HadGEM2-ES the sea ice has no heat capacity, and responds instantly to

surface thermodynamic forcing (zero-layer framework, appendix to Semtner (1979)), in the two newer models the sea ice is divided into 4 equally-spaced layers, each with a heat capacity, with temperatures calculated according to surface forcing using a forwards-implicit scheme (multi-layer framework, Bitz and Lipscomb, 1999). Secondly, the radiative effect of meltponds is modelled explicitly in the two newer models, using the topographic scheme of Flocco et al. (2015), whereas in HadGEM2-ES it is modelled implicitly, by lowering albedo by a fixed amount as surface temperature reaches the melting point. Finally, the sea ice model of HadGEM2-ES runs on a regular latitude-longitude grid, with a polar island, while those of HadGEM3-GC3.1-LL and UKESM1.0-LL run on the extended ORCA1 grid, an irregular grid with poles in Antarctica, Russia and Canada. There is also a difference in the snow albedo parameterisation, described in Chapter 3.

In addition there are many important differences in the atmosphere and ocean components of the models. The ocean component of HadGEM3-GC3.1-LL and UKESM1.0-LL is the Nucleus for European Modelling of the Ocean (NEMO), version 3.6, while HadGEM2-ES used the ‘HadGOM’ ocean model. The newer models use 75 vertical levels, as opposed to the 40 levels of HadGEM2-ES; the newer models also run on an irregular tripolar grid, with poles in Siberia, Canada and Antarctica, whereas HadGEM2-ES used a regular latitude-longitude grid, with an island at the North Pole to avoid singularities. Notable scientific improvements of NEMO over HadGOM include a nonlinear free surface scheme (in which the volume of the top ocean layers can vary according to sea surface height), and an improved turbulent kinetic energy vertical mixing scheme.

The Met Office Unified Model (UM) forms the atmospheric component of all three models, but HadGEM2-ES uses HadGAM1 (Martin et al., 2004)whereas the newer models use the GA7.1 atmosphere configuration (Walters et al., 2017). Notable improvements in GA7.1 relative to the old UM vn6.9 include a new dynamical core, ENDGAME, a new surface exchange scheme, JULES, and many improvements to cloud physics. This last is important because a notable deficiency of the HadGEM2-ES Arctic climate simulation was a tendency to underestimate liquid water fraction (West et al., 2019).

Throughout this thesis, the period 1980-1999 of the historically-forced ensembles of the three models is used for evaluation. This period is chosen for three reasons. Firstly, it predates the recent rapid Arctic sea ice loss, and can therefore be used as a reference period by which a model 'mean state' can be evaluated. Secondly, it falls entirely within the historical experiment periods of all three models evaluated. Thirdly, there is a relative abundance of Arctic observations as compared to earlier periods (widespread satellite observations of ice area begin in 1978, for example).

Four ensemble members are available for HadGEM2-ES and HadGEM3-GC3.1-LL; six ensemble members are used for UKESM1.0-LL. In using multiple ensemble members for analysis, we hope to separate model biases and differences arising from differences in fundamental physics and parameterisations to those arising from internal variability.

2.2 Reference datasets

Observational uncertainty in the Arctic is greater than that in the temperate and tropical regions. There are severe practical difficulties with collecting in situ data on a large scale over regions of ice-covered ocean. While satellites have in many cases been able to produce Arctic-wide measurements, most notably of sea ice concentration, the relative lack of in situ observations against which these can be calibrated means knowledge of the observational errors is limited. Reanalysis data over the Arctic is also more subject to model errors than in other regions, due to a lack of observations with which to constrain models (Lindsay et al., 2014).

The approach of this thesis is to use, as far as possible, a wide range of observational data with which to evaluate Arctic climate variables, to properly assess the large uncertainties. In Chapter 3, for example, where the basic Arctic climate simulations of the three models are evaluated, most variables are evaluated with respect to more than one dataset. In Chapter 4, where reference datasets are combined with simple models to estimate biases in surface flux and ice volume balance (the 'induced surface flux' framework) a similar approach is used; different datasets provide different estimates of surface flux bias, enabling observational uncertainty to be characterised properly. Below,

the datasets used to evaluate each Arctic climate variable are described in turn, and a summary is presented in Table 2.1.

2.2.1 Sea ice concentration

We use three datasets to evaluate sea ice concentration. Firstly, we use version 1.2 of the Hadley Centre Sea Ice and Sea Surface Temperature dataset (HadISST; Rayner et al., 2013). HadISST1.2 is a continuous record of sea ice concentration (and sea surface temperature) based upon a homogenised analysis of digitised historical sea ice charts and of satellite retrievals from the SSMI and SMMR passive microwave sensors, derived from the differing microwave signatures of open water and sea ice using the 'NASA-Team' algorithm (Comiso et al., 1996).

Secondly, we use version 2.2 of the same HadISST dataset (Titchner and Rayner, 2014). HadISST.2.2 is derived from a similar range of data sources as HadISST1.2, but uses a different set of algorithms to convert satellite retrievals to sea ice concentration. At low ice concentrations, the Bootstrap algorithm is used (Comiso et al., 1996); at high ice concentrations, the Bristol algorithm (Smith, 1996). In addition, HadISST.2.2 employs bias correction to remove an apparent discontinuity in the record in 1997.

Lastly, we use the dataset 'Sea Ice Concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data, Version 1' (Cavalieri et al., 1996), available from the National Snow and Ice Data Center (NSIDC). This dataset utilises a more recent version of the NASA-Team algorithm, with different tie points (locations used for calibrating the algorithm parameters) and weather filters (thresholds used to discard spurious open water detection due to atmospheric water vapour).

In general, sea ice extents for the period 1980-1999 tend to be greatest in HadISST.2.2, and least in the NSIDC dataset. This difference is likely to derive principally from the use of different reference datasets with respect to which the various algorithms are calibrated. Hence the three datasets are likely to provide a reasonable estimate of current uncertainty in sea ice extent measurements.

2.2.2 Sea ice thickness

We use three datasets of very different provenance to evaluate sea ice thickness. Firstly, we use the Pan-Arctic Ice and Ocean Model Assimilation System (PIOMAS; Zhang and Rothrock, 2003), an ice-ocean model forced with the NCEP reanalysis which assimilates ice concentration data. PIOMAS currently represents the most spatially and temporally extensive ice thickness dataset available, and has been found to compare well to satellite observations, albeit somewhat underestimating winter ice thickness in some years (Laxon et al., 2013; Wang et al., 2016). Secondly, we use estimates from the Envisat radar altimetry sensor from 1993-1999 (Laxon et al., 2003), the only comprehensive satellite-derived ice thickness dataset to overlap with our chosen period of model evaluation 1980-1999. Finally, we use estimates derived from submarine sonar (Rothrock et al., 2008). The latter two datasets are available only for the region south of 81.5°N, and the Central Arctic, respectively, and hence it is necessary to restrict model data to similar regions for evaluation.

2.2.3 Surface radiation

Three datasets are used to evaluate surface radiative fluxes, two derived from satellite measurements and one from reanalysis. Firstly, we use the CERES-EBAF (Clouds and Earth's Radiant Energy Systems – Energy Balanced And Filled) dataset (Loeb et al., 2009), based on direct measurements of top-of-atmosphere radiances from EOS sensors aboard NASA satellites, available from 2000 – present. CERES calculates net SW at the surface based on a simple relationship with top-of-atmosphere reflected shortwave (with small correction terms for solar zenith angle and precipitable water). Downwelling LW is calculated from the sum of two terms (representing radiation in the 'window' and 'non-window' parts of the spectrum), each of which is parameterised as a linear combination of precipitable water, surface temperature, near-surface temperature, and upwelling TOA longwave.

Christensen et al. (2016) evaluated CERES, along with a variety of other datasets, relative to in situ observations at Point Barrow (Alaska). They found CERES to perform quite well relative to other products, albeit underestimating

downwelling LW fluxes from November – February by 10-20 Wm^{-2}, and overestimating downwelling LW in March by 20Wm^{-2}.

Secondly, we use the ISCCP-FD (International Satellite Cloud Climatology Project FD-series) product (Zhang et al., 2004). ISCCP-FD is produced by the application of a radiative transfer model, the '03-Model', to the ISCCP-D dataset of cloud properties, which estimates cloud fraction as a function of cloud-top pressure and optical thickness. The 03-Model includes separate ice- and liquid-based cloud microphysical models and separate near infrared and visible surface albedo parameterisations.

The satellite sensors from which ISCCP-FD is derived are known to experience particular problems in the Arctic linked to high surface albedo, low solar zenith angle and the frequent presence of temperature inversions which invalidate assumptions used in radiative transfer algorithms. However, ISCCP-FD estimates of downwelling SW and LW radiation in the northern Beaufort Sea were evaluated with respect to in situ observations from the Surface Heat Budget of the Arctic (SHEBA) experiment (Liu et al., 2005), along with several other datasets. They were found to somewhat underestimate downwelling SW radiation during spring and early summer (by 20-40 Wm^{-2}), but overestimate it in summer and early autumn, by 10-30 Wm-2. They were also found to underestimate downwelling LW radiation in summer and early autumn by 10-30 Wm^{-2}, but to overestimate it in winter by around 40 Wm^{-2}.

Lastly, we use the ERA-Interim (ERAI) atmospheric reanalysis dataset. ERAI is used rather than the more up-to-date ERA5 because a large part of the dependent research, including West et al. (2019), had already been completed at the time of publication of ERA5. ERAI provides gridded surface flux data from 1979-present using a reanalysis system driven by the ECMWF (European Centre for Medium-range Weather Forcecasts) IFS forecast model and the 4D-Var data assimilation system (Dee et al., 2011). The IFS forecast model uses a spectral representation of basic atmospheric dynamic variables, and a hybrid sigma-pressure vertical co-ordinate scheme. The model does not simulate sea ice evolution, but assimilates sea ice concentration data from NCEP-2DVAR, a reanalysis system using SSMR/SSMI data as input. The 4D-Var assimilation system represented a significant advance on previous systems as it is able to

make use of model evolution in time to balance observations within a finite time period. It is based on the minimisation of the sum of accumulated error between model and observations, and model and background model state, over a 12-hour time period.

Lindsay et al. (2014) evaluated the ERAI simulation of Arctic surface radiative fluxes with respect to six other reanalyses, finding it to be amongst the most accurate datasets for these and other variables. However, they found a tendency for surface downwelling LW radiation to be biased high by 10-15 Wm^{-2} during winter, and for surface downwelling SW radiation to be biased low in April by ~ 20 Wm^{-2}, and by ~ 10 Wm^{-2} in March and May.

In addition to noting the evaluation studies above, winter downwelling LW fluxes in all datasets were compared, for the purposes of this thesis, to in situ measurements compiled by Lindsay (1998). ISCCP-FD was found to be biased high by 30-40 Wm^{-2}, but ERAI and CERES were relatively close to the observed distribution. Because of this, the larger biases of ISCCP identified above, and the problems with ISCCP described above, documented by Zhang et al. (2004), it is noted that ISCCP is likely the least reliable of the three surface radiation datasets used for this thesis.

2.2.4 Surface melt onset

To evaluate surface melt onset we use a dataset derived from passive microwave sensors using the Advanced Horizontal Range Algorithm (AHRA; Anderson and Drobot, 2001). The AHRA makes use of a change in relative emissivities, and associated change in relative brightness temperatures, that takes places at the onset of surface melt, in order to detect the onset of surface melting from the SSMI and SMMR passive microwave sensors. Specifically, in the AHRA a pixel is judged to be the site of surface melting if the difference between the 19GHz and 37Hz brightness temperatures is less than -10K. It is judged to be undergoing freezing conditions if the brightness temperature difference is greater than 4K. In the case of an intermediate difference, melt occurrence is judged by an analysis of temperature difference in the preceding and following 10-day time windows.

Anderson and Drobot (2001) found the resulting date of melt onset to compare well to in-situ observations of air temperatures in contemporaneous locations, although melt onset is occasionally judged to occur before air temperature reaches 0°C, consistent with observations showing liquid water can form in Arctic snow cover at subzero temperatures in the presence of solar radiation.

Variable	**Dataset**	**Temporal coverage**	**Spatial coverage**
Ice concentration	NSIDC Ice Concentration	1980-1999	Up to 84.5°N until June 1987, up to 87.2°N thereafter
	HadISST1.2	1980-1999	Whole Arctic
	HadISST.2.2	1980-1999	Whole Arctic
Ice thickness	PIOMAS	1980-1999	Whole Arctic
	EnviSat	1993-1999	Up to 81.5°N
	Submarine regression	1980-1999	SCICEX region (Rothrock et al, 2008)
Surface radiative fluxes	ISCCP-FD	1983-1999	Whole Arctic
	ERAI	1980-1999	Whole Arctic
	CERES-EBAF	2000-2013	Whole Arctic
Surface melt onset	NSIDC Snow Melt Onset over Arctic Sea Ice	1980-1999	Up to 84.5°N until June 1987, up to 87.2°N thereafter

Table 2.1. Summary of reference datasets used in this thesis.

Chapter 3: Characteristics of Arctic climate in the case study models

In this chapter, the sea ice simulations of the three models – HadGEM2-ES, HadGEM3-GC3.1-LL and UKESM1.0-LL are evaluated. Arctic surface radiation, and some aspects of the sea ice state that directly influence radiation, are also evaluated. The qualitative consistency between the various evaluations leads directly to the first innovation, the induced surface flux analysis, which quantifies the links between model biases in particular quantities. All model evaluation and comparison in this section was performed by myself. Much of the HadGEM2-ES evaluation is published in West et al. (2019).

3.1 Sea ice state: area, thickness and ice growth and melt

Strongly contrasting patterns of sea ice area bias are displayed across the three models (Figure 3.1). Whereas September sea ice extent is biased low in HadGEM2-ES relative to all datasets, biases for HadGEM3-GC3.1-LL and UKESM1.0-LL are difficult to discern due to the substantial observational uncertainty. Over the whole of the Northern Hemisphere, HadISST.2.2, HadISST1.2 and NSIDC measure September ice area to be 6.50, 5.80 and 4.77 x 10^6 km^2 respectively. HadGEM2-ES at 3.62 x 10^6 km^2 is clearly biased low, whereas HadGEM3-GC3.1-LL and UKESM1.0-LL at 6.01 x 10^6 km^2 and 6.49 x 10^6 km^2 respectively lie close to the top end of observational estimates.

In order to exclude peripheral seas in which the simulation of Arctic sea ice has little effect on sea ice levels in the Arctic Ocean, we define a new region, the ‘Arctic Ocean’ domain (Figure 3.1b). Unlike most standard definitions of the Arctic Ocean, this domain does not include the Greenland Sea, Barents Sea, Hudson Bay and Baffin Bay, as it is chosen to correspond to those areas where sea ice is present year-round for most of the evaluated period 1980-1999. The effect is to exclude areas of seasonal ice whose melt and growth does not contribute to the volume of perennial ice. For the rest of this chapter, and chapter 4, all analysis refers only to this region. The choice of this region necessitates consideration of ice export when evaluating sea ice volume changes, which is discussed at the end of this section.

HadGEM2-ES is still biased low with respect to all datasets in September, with an area of 3.62 x 10^6 km^2 comparing to 6.50, 5.80 and 4.77 x 10^6 km^2 in HadISST.2.2, HadISST1.2 and NSIDC respectively. HadGEM3-GC3.1-LL and UKESM1.0-LL are higher at 6.01 and 6.49 x 10^6 km^2 respectively, and lie within the observational uncertainty envelope.

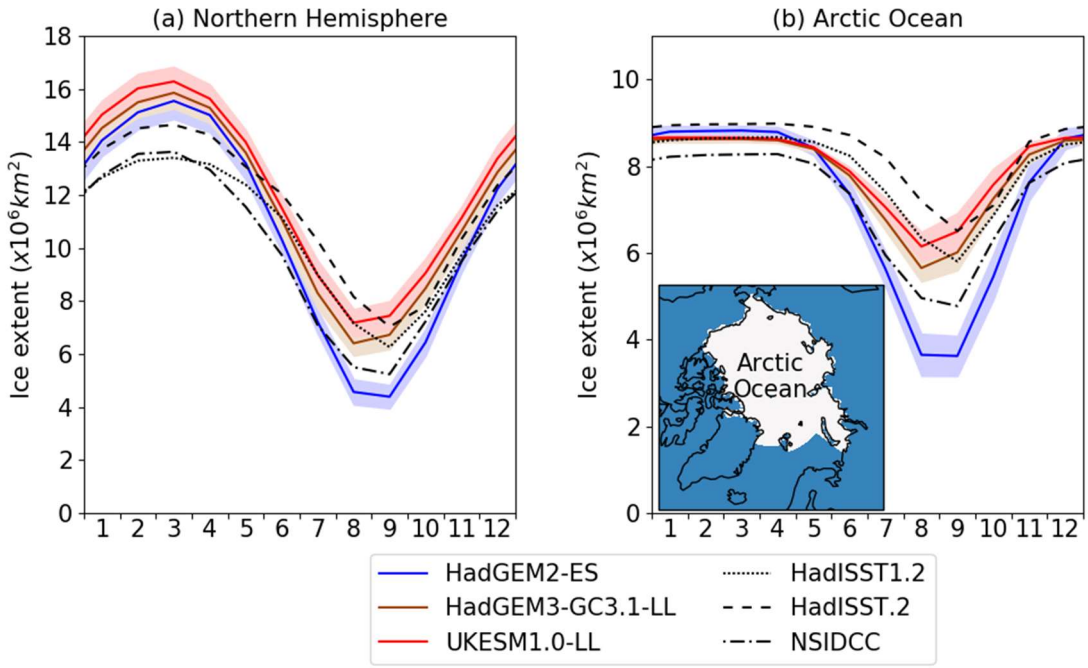


Figure 3.1. Ice area (1980-1999 mean) in HadGEM2-ES, HadGEM3-GC3.1 and UKESM, for (a) whole Northern Hemisphere; (b) Arctic Ocean only. HadISST1.2, HadISST.2.2 and NSIDC satellite observations also indicated. For each model, ensemble mean and twice standard deviation is shown.

Comparing maps of September ice area in the three models (Figure 3.2), we see that HadGEM2-ES simulates an ice edge that is much further north than in UKESM1.0-LL and HadGEM3-GC3.1-LL in all sectors of the Arctic (except north of Greenland and the Canadian Arctic Archipelago), but especially in the sector north of the Atlantic Ocean. Compared to HadISST1.2, HadGEM2-ES is biased low throughout the Arctic marginal seas, but HadGEM3-GC3.1-LL and UKESM1.0-LL tend to be biased high everywhere except in the eastern Kara Sea, in the vicinity of the New Siberian Islands.

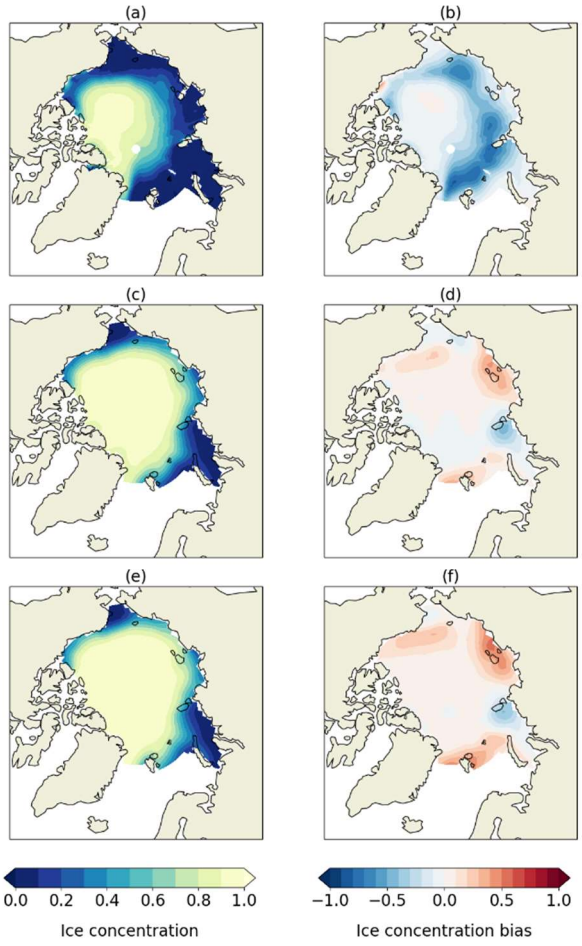


Figure 3.2. September ice concentration (left-hand column; (a), (c) and (e)) and ice concentration bias relative to HadISST1.2 (right-hand column; (b), (d) and (f)) for HadGEM2-ES, HadGEM3-GC3.1-LL and UKESM1.0-LL. Data is from 1980-1999.

Average sea ice thickness over the Arctic Ocean region is compared to data from PIOMAS, EnviSat and submarines (Figure 3.3). For all months of the year, and over all comparison regions, UKESM1.0-LL displays the highest average thickness and HadGEM2-ES the lowest. UKESM1.0-LL is biased high year-round with respect to all datasets. The bias is worst when comparing UKESM1.0-LL to PIOMAS over the Arctic Ocean region (1.81m) and is worse in the summer than in the winter, with a thickness bias of in September 1.86m comparing to a bias of 1.75m in May. UKESM1.0-LL displays annual mean ice thickness bias of 1.35m and 1.72m relative to the satellite and submarine data respectively.

HadGEM3-GC3.1-LL is also biased high relative to all datasets year-round, but less severely. Over the Arctic Ocean region, its annual mean ice thickness is biased high by 0.82m relative to PIOMAS. The bias is smaller relative to the other datasets, 0.41m and 0.22m relative to the satellite and submarine observations respectively.

HadGEM2-ES is biased low relative to all datasets year-round. The bias is smallest when comparing to PIOMAS, at -0.19m. The bias is concentrated in the summer half year; in September, the bias is highest at -0.40m, whereas in April the bias is insignificant. The bias is larger in magnitude when comparing to satellite (-0.69m) and submarine (-1.08m) data.

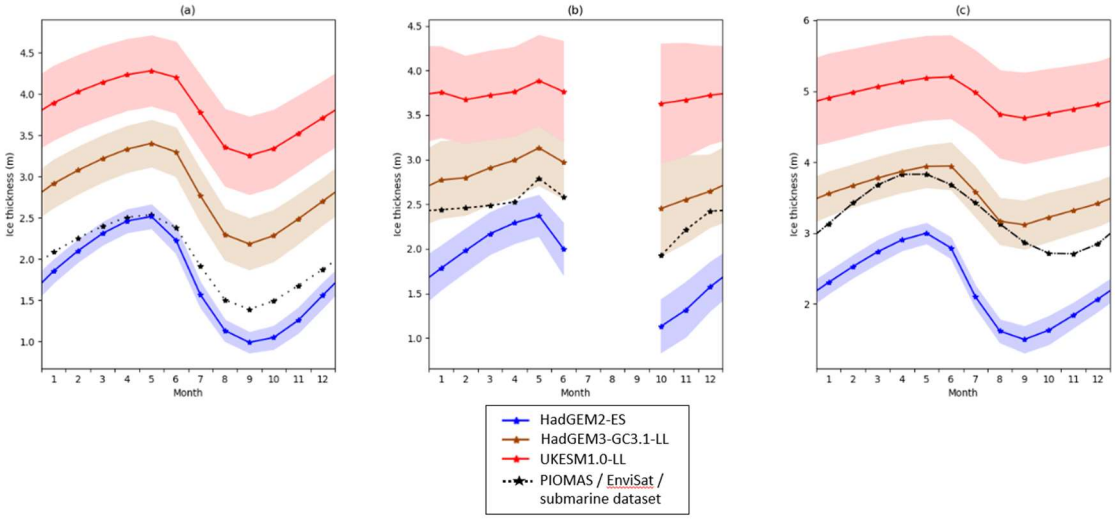


Figure 3.3. Ice thickness in HadGEM2-ES, HadGEM3-GC3.1-LL and UKESM1.0-LL. Showing (a) comparison with PIOMAS, for the Arctic Ocean region from 1980-1999; (b) comparison with Envisat, for the region of satellite coverage from 1993-1999; (c) comparison with submarine regression dataset, for the 'SCICEX' region from 1980-1999. For each model and panel, ensemble mean and twice standard deviation is shown.

To evaluate spatial patterns of ice thickness, average April ice thickness and bias relative to PIOMAS is compared in the three models (Figure 3.4). Despite a very small Arctic Ocean average bias relative to PIOMAS, HadGEM2-ES displays a thick bias on the Pacific side of the Arctic and a thin bias on the Atlantic side.

HadGEM3-GC3.1-LL and UKESM1.0-LL display similar spatial patterns of bias, with the thick ice bias substantially worse on the Pacific side of the Arctic than on the Atlantic side. These patterns are all associated with simulating the thickest ice to occur in the Pacific sector, rather than in the Atlantic sector as is observed to occur in reality. This is a common model error that may be associated with assuming the ice rheology to be isotropic (Tsamados et al., 2013).

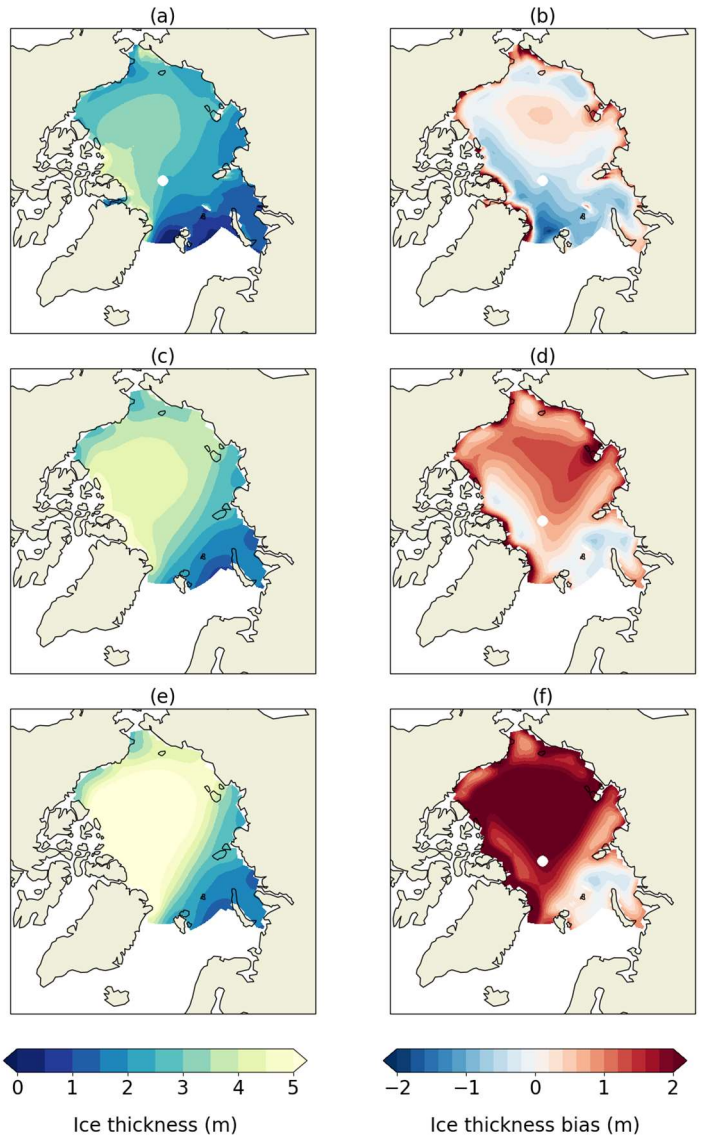


Figure 3.4. Maps of 1981-2010 average April ice thickness (a,c,e) and ice thickness bias (b,d,f) in HadGEM2-ES (a-b), HadGEM3-GC3.1-LL (c-d) and UKESM1.0-LL (e-f).

We evaluate also the seasonal cycle in the amount of annual ice growth and melt, calculated as the difference between the maximum and minimum monthly ice thickness (and referred to as seasonal ice growth / melt from hereon). HadGEM2-ES displays the greatest seasonal ice growth / meltof the three models (1.53m), with HadGEM3-GC3.1-LL lower at 1.22m and UKESM1.0-LL lowest at 1.03m. This is consistent with the differences in annual mean ice thickness, as thicker (thinner) ice tends to grow more slowly (quickly) in winter and melt more slowly (quickly) in summer.

In HadGEM2-ES the ice growth / melt is biased high relative to all data sources, with too much melt in summer and growth in winter, consistent with the low bias in annual mean ice thickness. For example, PIOMAS has annual growth / melt of 1.15m from 1980-1999 in the Arctic Ocean region, with respect to which HadGEM2-ES is biased high by 0.38m. Envisat measures annual growth / melt of 0.86m from 1993-1999 over the subset of the Arctic Ocean covered by the satellite; the corresponding figure for HadGEM2-ES is 1.24m. Finally, the submarine-derived dataset has annual growth / melt of 1.12m from 1980-1999 over the Central Arctic Ocean, whereas the corresponding figure for HadGEM2-ES is 1.50m. The three data sources each indicate positive ice growth / meltbiases for HadGEM2-ES of 38cm.

HadGEM3-GC3.1-LL displays 7cm more ice growth / melt than does PIOMAS, but 18cm and 29cm less ice growth / melt than Envisat and the submarine dataset respectively. Hence the seasonal ice growth/melt of HadGEM3-GC3.1-LL agrees within observational uncertainty.

For UKESM1.0-LL, all datasets agree the bias in ice growth / melt is negative, consistent with the annual mean ice thickness bias, but there is disagreement as to the magnitude. PIOMAS indicates a negative ice growth / melt bias of 12cm, but Envisat and the submarine dataset indicate negative ice growth / melt biases of 61cm and 54cm respectively. The high uncertainty in UKESM1.0-LL occurs due to a high disparity between the ice growth / melt measured over the Arctic Ocean region, and that measured over the smaller regions as compared to Envisat and the submarine datasets. For example, UKESM1.0-LL models ice growth / melt of 1.03m over the whole Arctic Ocean region, but only 0.58m over the Central Arctic and 0.26m over the region of Envisat coverage: the

UKESM1.0-LL ice thickness simulation is more strongly damped in regions of thick ice than is HadGEM2-ES.

The ice growth / melt biases are associated with an energy balance bias. This could be associated with a bias in the surface energy balance, the oceanic heat convergence, or in ice divergence (or some combination of these). However, the ice divergence term is small, with indeterminate annual cycles, at 2.5, 4.0 and 5.1 Wm^{-2} in HadGEM2-ES, HadGEM3-GC3.1-LL and UKESM1.0-LL respectively. It is likely to be similarly small in observations (e.g. Serreze et al., 2007). The oceanic heat convergence term is also small at 4.4 Wm^{-2}, 3.8 Wm^{-2} and 3.9 Wm^{-2} respectively, and is likely to be similarly small in observations (e.g. McPhee et al., 2003). Differences in these terms between models are too small to explain the differences in ice melt / growth seen: with ice density of 917kgm^{-3} and latent heat of fusion of 3.35×10^3 Jkg^{-1}, a difference of 1Wm^{-2} is equivalent to only an additional .84cm of ice melt per month. Hence we focus first on the surface energy balance, and evaluate the surface radiative fluxes.

3.2 Surface radiative fluxes

We evaluate surface radiative fluxes in HadGEM2-ES, HadGEM3-GC3.1 and UKESM1.0 over the Arctic Ocean region with respect to three datasets: ISCCP-FD and CERES-EBAF satellite datasets, and the ERAI reanalysis. Downwelling SW is very similar in the three models (Figure 3.5a), and is biased high relative to all datasets during the spring. By contrast, there are major differences in upwelling SW from June-September (Figure 3.5b), with UKESM1.0-LL modelling the highest values and HadGEM2-ES the lowest; UKESM1.0-LL simulates 39.7, 21.9, 22.5 and 12.2 Wm^{-2} more upwelling SW than HadGEM2-ES in the four months respectively. The range of observational uncertainty contains all model estimates of upwelling SW in all months except June, when HadGEM2-ES is biased low, so it is not possible to make definitive conclusions about model biases for other models and months. For example, in UKESM1.0-LL, throughout the summer upwelling SW is higher than CERES and ERAI, but lower than ISCCP. Summer net downwards SW is highest in HadGEM2-ES and least in UKESM1.0-LL, with only the HadGEM2-ES flux in June being clearly biased high (Figure 3.5c).

Examining the LW fluxes, downwelling LW is similar in the three models from January – May (Figure 3.5d), but from June-December varies, with HadGEM2-ES highest and UKESM1.0-LL lowest (except in July where HadGEM3-GC3.1-LL is barely lowest). The differences are greatest in the autumn, when UKESM1.0 simulates less downwelling LW than HadGEM2-ES by 13.7, 23.1 and 18.2 Wm^{-2} in September, October and November respectively, but during summer are of comparable size only in June, when UKESM1.0-LL simulates 10.2 Wm^{-2} less downwelling LW than does HadGEM2-ES. All models are biased low with respect to the reference datasets for the entire freezing season (September – April). We note that the downwelling LW bias relative to ISCCP is likely to be overstated, as ISCCP downwelling LW fluxes are biased high relative to in situ measurements compiled by Lindsay (1998) over the Arctic Ocean (see section 2.2.3).

Upwelling LW fluxes display similar differences (Figure 3.5e), with HadGEM2-ES highest and UKESM1.0-LL lowest throughout the freezing season, but these are greater in magnitude and last for longer into to the winter. For example, from September-December UKESM1.0-LL models 16.5, 31.1, 27.4 and 15.5 Wm^{-2} less upwelling LW than HadGEM2-ES respectively, indicating a lower surface temperature in UKESM1.0-LL, perhaps partly associated with the thicker ice, and reduced conduction of heat from the ocean. From December – April, all models are biased low, but during the autumn models lie closer to the observed range (HadGEM2-ES being within the range). As a result of the upwelling LW differences, net downwards LW is lower in HadGEM2-ES than in the other two models during the winter (Figure 3.5f).

The differences in ice growth / melt between the models associated with differences in radiative flux are now calculated. The net SW flux is greatest in HadGEM2-ES and least in UKESM1.0-LL from May-September, mostly driven by the upwelling SW differences (but in May by downwelling SW). In particular, from June-August the average net SW flux is 114.0, 100.2 and 91.9 Wm^{-2} in HadGEM2-ES, HadGEM3-GC3.1-LL and UKESM1.0-LL respectively. These differences imply 56cm more ice melt in HadGEM2-ES than in UKESM1.0-LL in total over the summer season, and 21cm more ice melt in HadGEM3-GC3.1-LL than in UKESM1.0-LL.

For the net LW flux, from October – April HadGEM2-ES is lowest (i.e. most strongly negative) and UKESM1.0 highest (least strongly negative). On average, HadGEM2-ES, HadGEM3-GC3.1-LL and UKESM1.0-LL simulate -42.9, -38.1 and -36.3 Wm^{-2} net LW from October – April. These differences imply 39cm less ice growth during freezing season in UKESM1.0-LL than HadGEM2-ES, and 11cm less ice growth in UKESM1.0-LL than in HadGEM3-GC3.1. During the summer, net LW fluxes are near-identical in UKESM1.0-LL and HadGEM3-GC3.1-LL, but lower (more strongly negative) than in HadGEM2-ES. From June-August, for example, HadGEM2-ES simulates net LW of -20.9 Wm^{-2}, while HadGEM3-GC3.1 and UKESM1.0 simulate net LW of -25.3 and -25.4 Wm^{-2} respectively, a difference representing 11cm less ice melt in the newer models relative to HadGEM2-ES.

In summary, during summer the net radiative differences imply 67cm less ice melt in UKESM1.0-LL than in HadGEM2-ES, but during winter imply 39cm less ice growth. These numbers are consistent with a 50cm volume balance difference in UKESM1.0-LL and HadGEM2-ES, and suggest that in summer (winter) other processes act to ameliorate (exacerbate) the effect of the net radiative differences. In both seasons HadGEM3-GC3.1 displays intermediate net radiative flux differences.

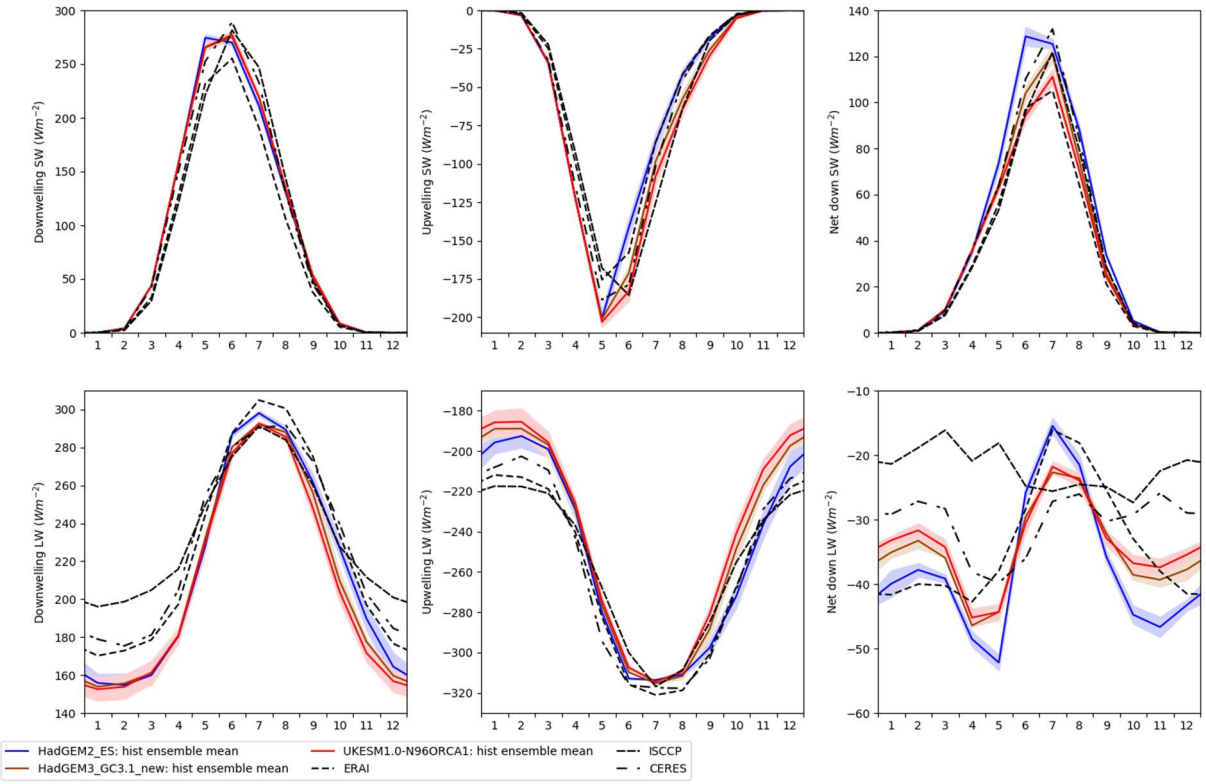


Figure 3.5. 1980-1999 mean modelled radiative fluxes of (a) Downwelling SW, (b) upwelling SW, (c) net downwards SW, (d) downwelling LW, (e) upwelling LW, (f) net downwards LW in HadGEM2-ES, HadGEM3-GC3.1-LL and UKESM1.0-LL, together with observational estimates from ISCCP-FD, ERAI and CERES, averaged over the Arctic Ocean region. For each model, ensemble mean and twice standard deviation is shown.

The above evaluation discusses the effect of the surface radiation bias on the ice growth / melt bias. However, causality will go in the opposite direction also. For example, the summer upwelling SW bias is likely to be strongly influenced by the ice area bias; in July and August respectively, average Arctic Ocean ice concentration is 0.14 and 0.25 higher in UKESM1.0-LL, associated with an albedo bias of around 0.07 and 0.13 that is sufficient toexplain the whole upwelling SW bias. By contrast, the June ice concentration bias is not nearly large enough to explain the upwelling SW bias in this month, and hence in June other processes are likely to be important in causing modeled surface albedos to differ.

In a similar way, the surface temperature (and hence upwelling LW) is strongly influenced by the ice concentration and thickness, as during winter the ocean surface tends to be warmer than the sea ice surface, and thin ice warmer at the

surface than thick ice. It is likely that the upwelling LW difference of UKESM1.0-LL relative to HadGEM2-ES from September-January is at least partly associated with the large September ice area difference, and associated ice thickness differences during the early part of the freezing season. In order to understand the net radiation biases better, therefore, we evaluate and compare some other variables of the ice state that affect surface albedo.

3.3 Additional diagnostics impacting sea ice and surface radiation

We firstly evaluate some other variables that affect surface albedo. In HadGEM2-ES, the surface albedo is parameterized after Curry et al. (2001), in which albedo is linearly decreased from 0.8 to 0.65 (where snow is present) or from 0.61 to 0.535 (where no snow is present) as surface temperature rises from -1°C to 0°C. This parameterization is intended to mimic the radiative effect of meltponds on the surface of the ice. In UKESM and HadGEM3-GC3.1, however, meltponds are modeled explicitly, using the scheme of Flocco et al. (2015). In this scheme, surface meltwater is allowed to pool on the ice, in quantities determined by the ice topography, with explicit pond area and thickness modelled. Hence there are several factors which could influence surface albedo differences between the models, and model biases, in addition to sea ice area: the presence of snow on the ice, surface temperature, differences resulting from model parameters (e.g. cold snow albedo) or differences resulting from the structure of the models themselves.

We consider first the surface temperature, or more precisely the onset of surface melt which is measured by satellites. For all models we define, for each year and grid cell, surface melt onset to have occurred on the first day of the year that surface temperature exceeds -1°C. We compare maps of average melt onset date obtained in this way to SSMI estimates (Anderson et al., 2012). Each of the models simulates surface melt onset to occur too early across the Central Arctic, and too late close to the Arctic Ocean coasts (Figure 3.6). However, the Central Arctic bias is worst in HadGEM2-ES (over 20 days) and least severe in UKESM1.0-LL (about 10 days), with HadGEM3-GC3.1-LL falling in between the two. In the full Arctic Ocean average, HadGEM2-ES models surface melt onset to occur much earlier across the Central Arctic than is

modeled by UKESM1.0-LL. HadGEM2-ES is biased early relative to the SSMI measurements; by contrast, UKESM1.0-LL is biased slightly late. We conclude that surface melt onset is likely to contribute to a significant negative bias in surface albedo in HadGEM2-ES, and that differences in Central Arctic surface melt onset are also likely to contribute to the differences in June surface albedo between the three models.

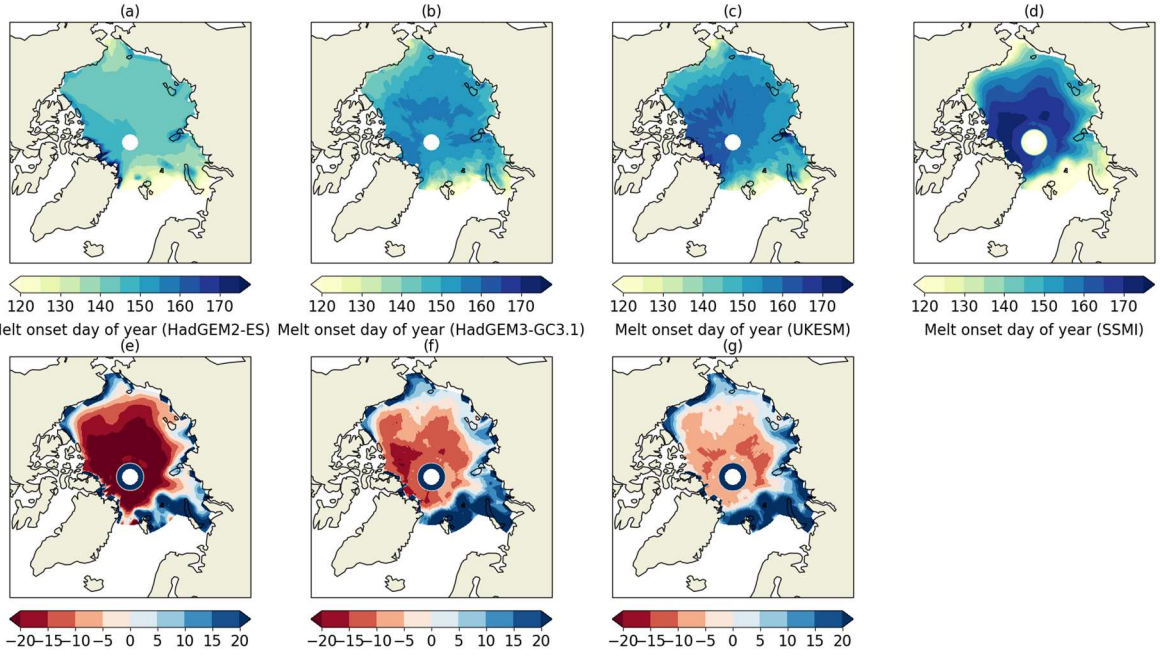


Figure 3.6. Maps of date of surface melt onset in HadGEM2-ES, HadGEM3-GC3.1-LL, UKESM1.0-LL and SSMI observations, with model biases shown in the lower row.

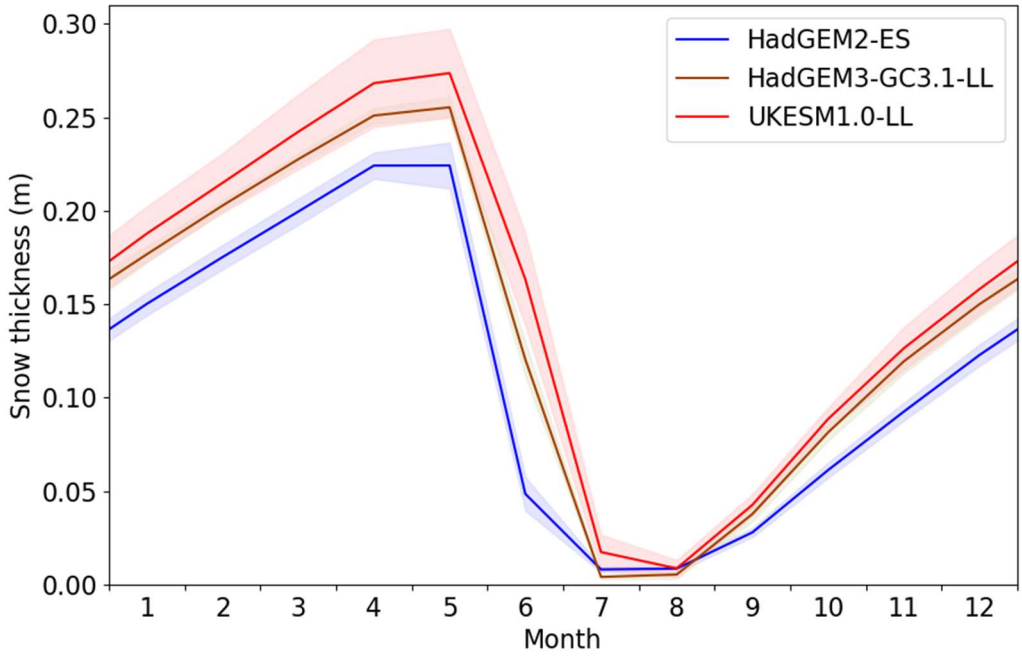


Figure 3.7. Arctic Ocean average snow thickness in HadGEM2-ES, HadGEM3-GC3.1-LL and UKESM1.0-LL. For each model, the ensemble mean and twice the standard deviation is shown.

Next we compare snow cover in the three models (Figure 3.7). Arctic Ocean average snow thicknesses are near-zero in early autumn, but grow at different rates during the freezing season. By May, average snow thicknesses are 24cm, 26cm and 28cm in the three models' ensemble means respectively. In June, differential rates of melt lead to much wider spread between the models, with average snow thicknesses of 6cm, 13cm and 18cm respectively. However, the models differ in how they parameterize snow area, and therefore the effect of snow on surface albedo. HadGEM2-ES uses a negative exponential relationship:

$$a_{snow} = 1 - e^{-0.2 h_{snow} \rho_{snow}} \qquad (3.1)$$

While HadGEM3-GC3.1 and UKESM1.0 use the formula:

$$a_{snow} = \frac{h_{snow}}{h_{snow} - 0.02} \qquad (3.2)$$

The effect of this is that the newer models simulate a lower fraction of snow for the same snow thickness value (Figure 3.8). We can see the effect of this when we compare maps of June snow thickness in the three models (Figure 3.9a-c)

with maps of June snow area estimated using the HadGEM2-ES (Figure 3.9d-f) and HadGEM3-GC3.1/UKESM1.0 (Figure 3.7g-i) formulations. As shown in Figure 3.7, snow thickness in HadGEM2-ES is lower than in HadGEM3-GC3.1 and UKESM1.0: under both snow area parameterisations, this results in a substantially lower snow area also. However, the difference is greatly reduced when HadGEM2-ES snow area, computed using the HadGEM2-ES formulation, is compared to the snow area in the newer models, using the new formulation (red highlighted panels). Hence while the differences in snow cover in June are likely to be a factor in the June surface albedo differences, and the greater net SW flux and ice melt in HadGEM2-ES, the different parameterisations are likely to oppose this.

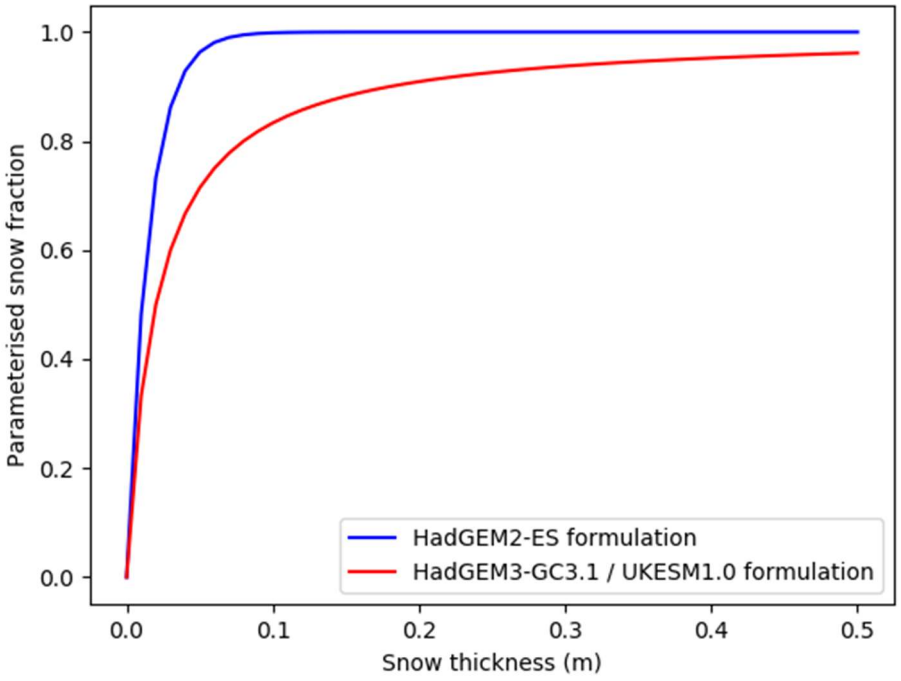


Figure 3.8. The relationship between snow fraction and snow thickness in HadGEM2-ES, HadGEM3-GC3.1 and UKESM1.0.

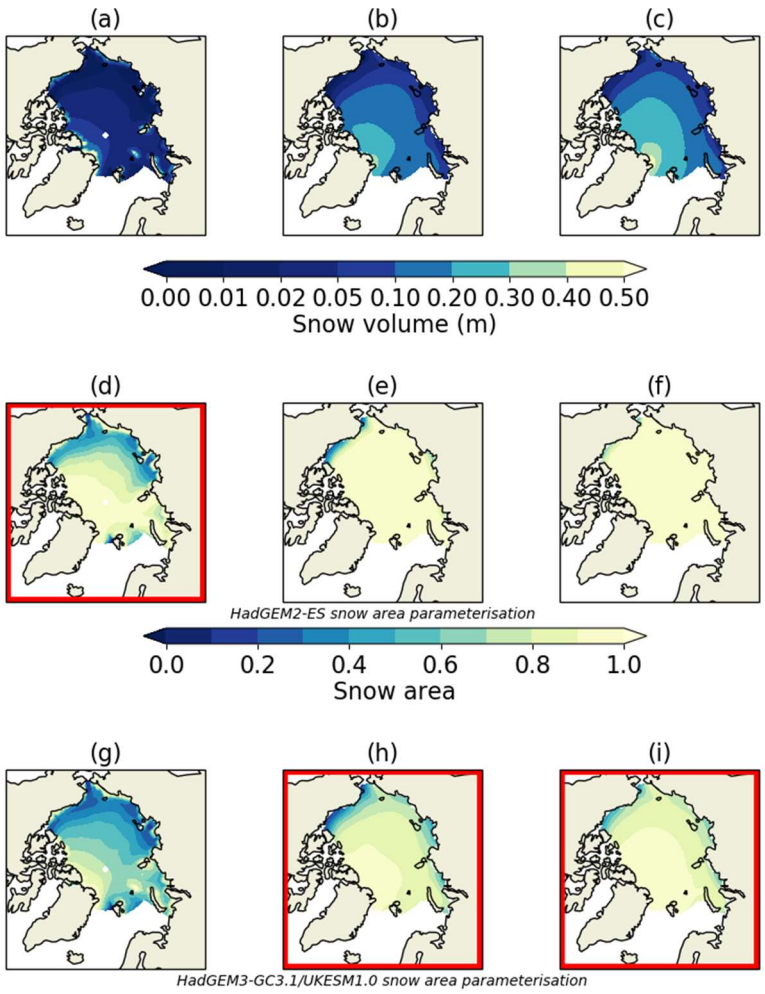


Figure 3.9. (a-c) snow thickness in HadGEM2-ES, HadGEM3-GC3.1, UKESM1.0 respectively; (d-f) snow area, HadGEM2-ES parameterization, arising from these; (g-i) snow area, HadGEM3-GC3.1/UKESM1.0 parameterisation, arising from these. (d), (h) and (i), indicated by red highlighted borders, show the snow area fields actually simulated in HadGEM2-ES, HadGEM3-GC3.1-LL and UKESM1.0-LL respectively.

Finally, we examine the direct impact of the explicit meltpond scheme of HadGEM3-GC3.1-LL and UKESM1.0-LL by comparing, for the month of June, average meltpond fraction of ice where surface melting is taking place. Implied meltpond area in HadGEM2-ES is between 18-23% Arctic-wide due to the albedo parameterization which implicitly assumes 18% meltpond coverage over bare ice and 23% over snow. In HadGEM3-GC3.1-LL by contrast there are large areas of the Arctic, particularly in the shelf seas, where meltpond area is much larger, in some places approaching 40% (Figure 3.10). In UKESM1.0-LL, the situation is similar, but meltpond areas tend to be somewhat smaller than in

HadGEM3-GC3.1-LL Arctic-wide, with the result that there is also a large area of the Central Arctic where meltpond areas are actually smaller than in HadGEM2-ES. Hence the explicit meltpond scheme in the newer models may actually be working to produce the opposite June surface albedo difference to the one observed, but in the central Arctic in UKESM1.0-LL at least the meltpond scheme may be contributing to the surface albedo difference.

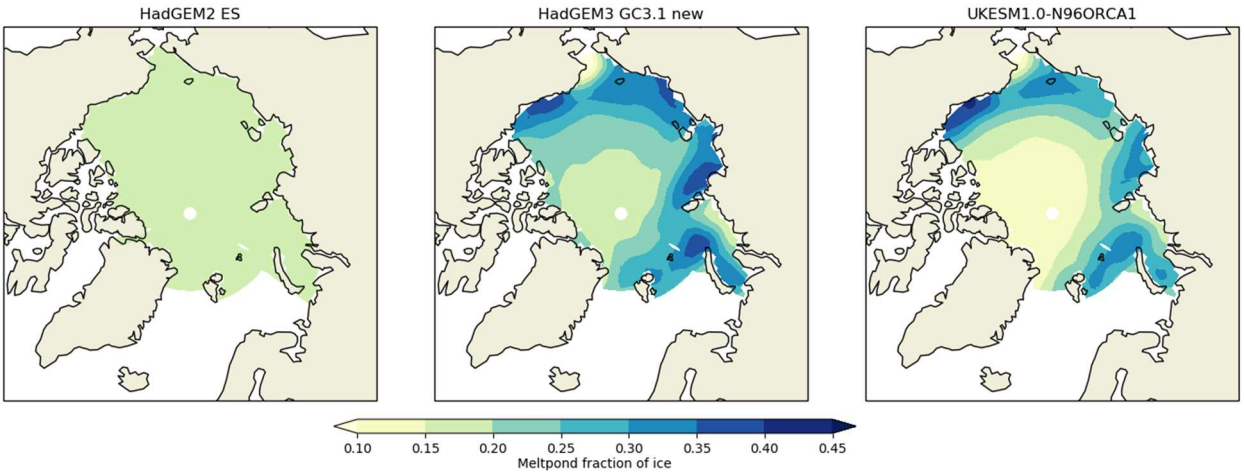


Figure 3.10. Meltpond area in HadGEM2-ES, HadGEM3-GC3.1-LL and UKESM1.0-LL.

3.4 Summary of model evaluation, and looking towards next steps

In summary, summer ice area is lower in HadGEM2-ES than in HadGEM3-GC3.1-LL and UKESM1.0-LL, and is biased low relative to observations. Ice thickness is lower year-round than in HadGEM3-GC3.1, which is in turn lower year-round than in UKESM1.0-LL; HadGEM2-ES and HadGEM3-GC3.1-LL are likely to be biased low and high respectively, while UKESM1.0-LL is certainly biased high. Seasonal ice growth/melt is highest in HadGEM2-ES and lowest in UKESM1.0-LL; it is likely to be biased high and low in HadGEM2-ES and UKESM1.0 respectively, but in HadGEM3-GC3.1 a bias cannot be discerned.

An evaluation of radiative fluxes finds summer net SW to be highest in HadGEM2-ES and lowest in UKESM1.0-LL, with HadGEM2-ES likely biased high. The differences (bias) come mainly from the upwelling SW term, indicating that surface albedo differences (biases) are responsible. In July and August, differences in modelled ice area are sufficient to explain the differences in

surface albedo, but in June this is not the case. In winter, downwelling LW in all models is similar, and is likely to be biased low. Owing to differences in the upwelling term, net LW is lower in HadGEM2-ES than in HadGEM3-GC3.1-LL and UKESM1.0-LL.

An evaluation of mechanisms influencing surface albedo (other than ice area) finds a role for the timing of surface melt onset, and for the snow thickness, in driving the differences in modelled surface albedo during June. However, the use of an explicit meltpond scheme in the newer models may be either reinforcing or opposing these effects in different parts of the Arctic, and the use of a different snow area parameterization in the newer models is certainly opposing the surface albedo difference.

The model evaluation prompts two key questions which motivate the following work. Firstly and most obviously, it would be useful to perform a quantitative analysis of the many different factors influencing surface albedo during the summer. For example, surface albedo, net SW and ice melt in HadGEM2-ES are biased low, high and high respectively, while the date of surface melt onset is biased early. This prompts the question of how much additional ice melt is associated with the early surface melt onset, through its effects on surface albedo and net SW, and how does this compare to the additional ice melt caused by the ice area bias? When comparing the different models, a still more comprehensive analysis would estimate additional ice melt caused by model differences in ice area, snow thickness and melt onset, as well as the additional ice melt caused by the differences in meltpond and snow parameterisations.

The second question is prompted by observing that during the winter, while all three models display very similar downwelling LW, variation in upwelling LW is crucial in driving variation in net LW, and hence ice growth. But surface temperature, and hence upwelling LW, are driven both by atmospheric thermal forcing, and by the properties of the snow-ice column beneath. To understand the differences, and bias, in modelled ice growth during winter, it will be important to separate these two effects: to quantify separately the effects of differences in atmospheric forcing, and differences in the ice state itself, in driving differences in ice growth. In Chapter 4, we introduce a new framework that allows these two questions to be addressed simultaneously.

4. Induced surface fluxes: A general method for attributing sea ice volume balance biases to individual model variables

In this chapter, we present the first of two enhancements to sea ice model evaluation: a framework to diagnose the proximate causes of model biases (or differences) in the rate of sea ice growth and melt. This 'induced surface flux' (ISF) framework proceeds naturally from the model evaluation of Chapter 3. The surface energy balance is treated as the driver of modelled sea ice growth and melt. We use a simple model to quantify how surface flux depends on each of the variables evaluated in Chapter 3, and hence separate the effects of each model bias on sea ice growth and melt.

This framework is similar to that used by Holland and Landrum (2015) to quantify the contribution of changes in surface albedo and in downwelling SW to changes in net SW over 3 periods of the 21st century, but was developed independently and differs in two ways. Firstly, it quantifies the contribution of model processes to model biases, and inter-model differences, rather than to changes in time within model runs. Secondly, it is effectively a generalisation of this method, as it quantifies the contributions of differences in a larger number of model variables.

This chapter is based upon West et al (2019), in which the ISF framework was developed to evaluate HadGEM2-ES. As lead author of this study I was responsible for the ISF framework design and the resulting evaluation of HadGEM2-ES using this framework. I was also responsible for the analysis of error in the ISF framework, repeated and expanded upon in this thesis in the Appendix, and for the production of all figures and for the writing up of the paper in its final form.

4.1 Overview

The sea ice process evaluation in Chapter 3 produces many results which are consistent with the biases in the sea ice volume simulations. Model biases in the sea ice state during the summer that would cause a low bias in surface albedo tend to be associated with corresponding low model biases in upwelling SW. Model biases in total net radiation in both summer and winter are

consistent with model biases in seasonal ice growth/meltin these seasons. In this way, the basic model evaluation provides a quantitative description of the difference between model Arctic climate states, and allows a qualitative assessment of causal relationships between these biases. In this section, it is shown how the model biases can be combined with simple models of sea ice physics to provide a quantitative assessment of many of the causal relationships at work.

This process is illustrated with a simple example. It was seen in Chapter 3 that surface melt onset over sea ice in spring occurs at very different times in the three models, earliest in HadGEM2-ES and latest in UKESM1.0-LL. Relative to SSMI observations, melt onset is much too early Arctic-wide in HadGEM2-ES; in UKESM1.0-LL it is too early in the central Arctic but too late in the peripheral Arctic seas. It was conjectured that these biases may play a major part in the net SW flux biases (or lack of) in the three models in the month of June. We can use the surface albedo parameterisations of each model to estimate the average June surface flux bias that would be associated with the surface melt onset biases.

For each grid cell of each model, we calculate the average fraction of days in June for which surface melting had commenced over the period 1980-1999 (a_{melt}). For example, if in June 1980 in HadGEM2-ES surface melting had commenced in a particular grid cell on 29th May, a_{melt} would be 1; if in June 1981 the melt onset date was 8th June for the same grid cell, a_{melt} would be 8/30. This fraction is then averaged across all model years 1980-1999; the same calculation is carried out for the SSMI observations. Figure 4.1 shows a_{melt} for the three models and the SSMI observations, with model bias indicated below. The bias in a_{melt} is similar to that in melt onset date; HadGEM2-ES simulates too much surface melting fraction during June, whereas the bias in UKESM1.0-LL varies from the Central Arctic (too little melt) to the periphery (too much melt).

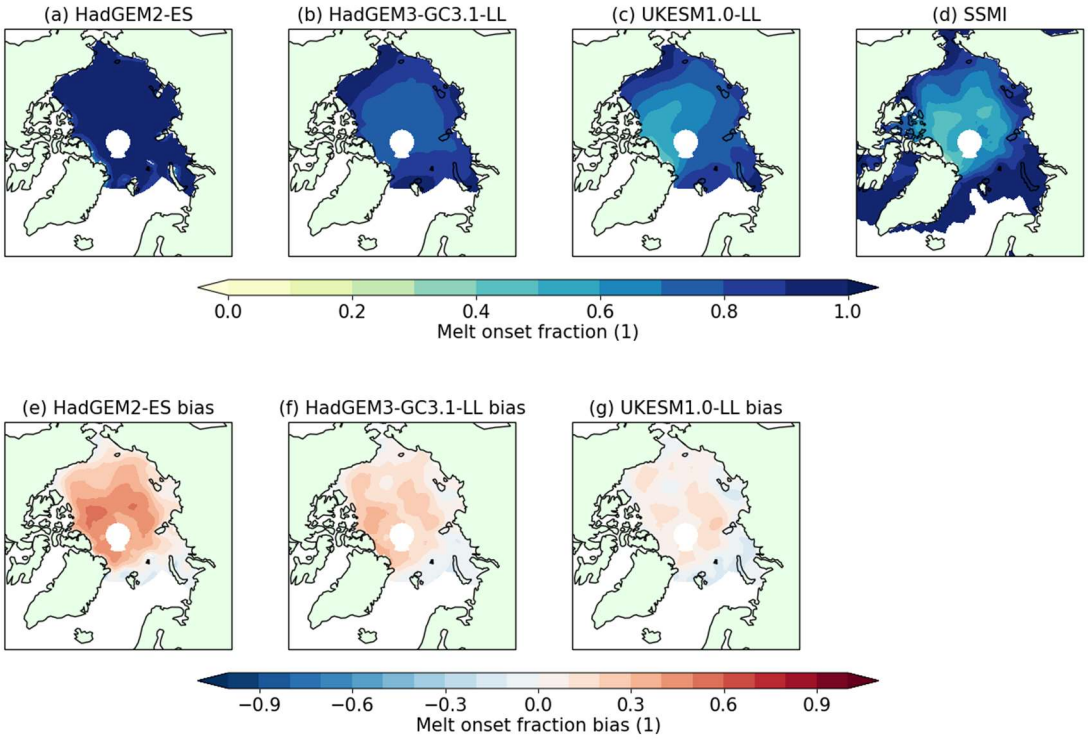


Figure 4.1: The June average fraction of melting ice surface (a_{melt}) in HadGEM2-ES, HadGEM3-GC3.1-LL, UKESM1.0-LL and SSMI observations from 1980-1999, with model bias indicated below.

In order to determine the net SW bias associated with the model bias in a_{melt}, for each grid cell we multiply the bias in a_{melt} by the model monthly mean downwelling SW, and by the difference in surface albedo we associate with surface melting. For example, in HadGEM2-ES cold snow albedo is 0.80, compared to a melting snow albedo of 0.65; for bare ice, the relevant parameters are 0.61 and 0.535. Hence for this model, we can characterize surface albedo sensitivity to surface melting as a weighted average of 0.15 (snow albedo reduction) and 0.075 (bare ice albedo reduction), with weights determined by snow area (which is parameterized from model snow thickness). The resulting albedo dependence shows a smoothly varying field, with high values approaching 0.15 near the Atlantic and Pacific ice edges, but lower values approaching 0.075 north of Alaska and Siberia, both regions of lower June snow thickness.

A similar calculation can be carried out for HadGEM3-GC3.1-LL and UKESM1.0-LL. These models simulate surface melt onset in a somewhat different way, by modeling meltponds explicitly; hence the albedo reduction that

occurs in the presence of surface melt is not fixed, but depends on the meltpond area. To make a meaningful comparison to the SSMI dataset, however, some fixed albedo reduction based on a reference meltpond area must be assumed. We use the meltpond areas implied by the HadGEM2-ES parameterization (0.18 over bare ice and 0.23 over snow), equivalent to assuming the same aggregate albedo reductions in the presence of melt onset. These models both show a much higher sensitivity of surface albedo to surface melt onset (Figure 4.2b,c), with values approaching 0.15 across most of the Arctic. This is consistent with June snow thickness being higher in both models than in HadGEM2-ES, as the albedo reduction associated with surface melt onset is greater over snow than over bare ice.

In order to determine the net SW dependence on surface melt onset, we multiply the dependence of surface albedo on surface melt onset by the model June downwelling SW fields. The three models display similar spatial patterns of June downwelling SW (Figure 4.2d-f), with highest values occurring near the Greenland coast and Canadian Arctic Archipelago, but there is offset between the models with HadGEM2-ES displaying the lowest values Arctic-wide and UKESM1.0-LL the highest.

The fields of net SW dependence on melt onset (Figure 4.2 g-i) are everywhere positive: surface melt onset induces an increase in net SW, as would be expected. The increase is larger in the presence of snow, and in the presence of a high downwelling SW flux. Hence in all models, the sensitivity to melt onset tends to be highest near Greenland and the Canadian Arctic Archipelago, but is also relatively high near the Atlantic ice edge. Sensitivity is greatest in UKESM1.0-LL, where values of 40-45 Wm^{-2} occur over most of the Arctic Ocean (indicating that surface melt onset is expected to induce an additional 40-45Wm^{-2} net SW). It is least, and less spatially uniform, in HadGEM2-ES, where values fall to 25-30 Wm^{-2} close to the Alaskan and Siberian coasts.

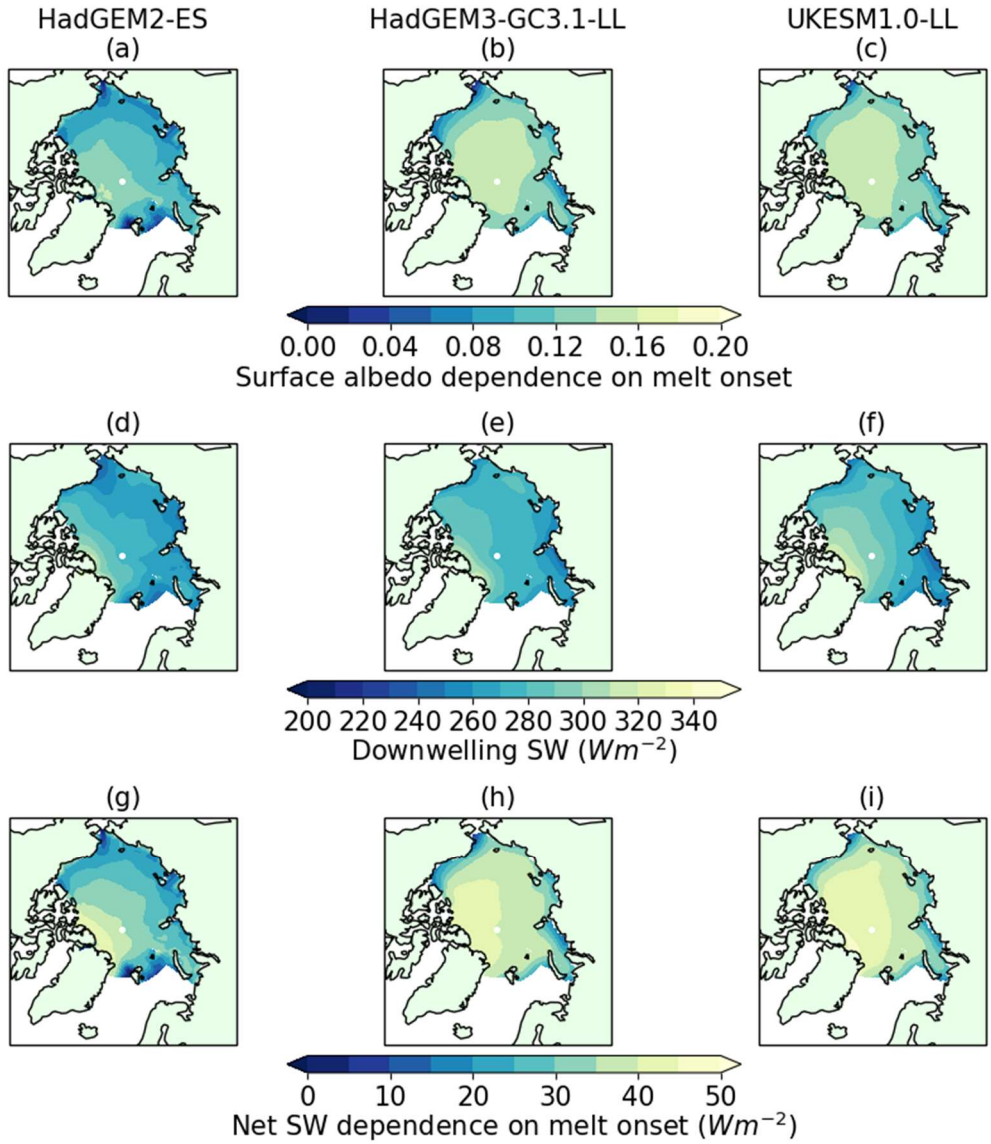


Figure 4.2. A demonstration of the calculation of sensitivity of net SW radiation to surface melt onset for HadGEM2-ES (left column), HadGEM3-GC3.1-LL (middle column) and UKESM1.0-LL (right column). (a)-(c) show sensitivity of surface albedo to surface melt onset; (d)-(f) show downwelling SW radiation; (g)-(i) show sensitivity of net SW radiation to surface melt onset, the product of the top and middle rows.

In the example above, the melt onset bias could be associated with a surface flux bias because the effect of surface melt onset on surface flux could be easily captured by a simple relation in which the other dependent variables can be said to be independent. Specifically, net SW depends on downwelling SW and a number of variables determining surface albedo: ice area, snow area and surface melt onset. An instantaneous change in each of the independent variables implies an instantaneous change in the net SW radiation, but does not

imply an instantaneous change in another variable. Hence the effect of each variable on surface flux, as measured by the partial derivative, is separate from the effects of all of the others: the variables are quasi-independent.

This suggests that if it is possible to encapsulate the effects on surface flux of all variables evaluated in Chapter 3 in a single equation, in which the variables are quasi-independent in this sense, then an induced surface flux bias can be calculated for the bias in each variable. We use a second example to demonstrate how this approach might help in describing the effects of ice thickness and atmospheric forcing on surface flux during winter.

January ice thickness is very different in the three models, with UKESM1.0 being thickest and HadGEM2-ES thinnest; UKESM1.0-LL and HadGEM3-GC3.1-LL are likely to be biased thick, HadGEM2-ES biased thin. We would expect these biases to cause ice growth to be weaker (thick bias) or stronger (thin bias) than it would otherwise be in the respective models. Precisely how much weaker or stronger, however, depends on the atmospheric forcing, but also on the ice thickness itself, because the effect of ice thickness on ice growth is stronger at thinner ice thicknesses.

To characterize this effect, we use a simple single-column model, similar to that used by Thorndike (1992). The dependence of surface flux on surface temperature is linearised at the freezing temperature $T_{sfc} = 0℃$:

$$F_{sfc} = A + BT_{sfc} \qquad (4.1)$$

(here F_{sfc} represents surface flux and T_{sfc} represents surface temperature.)

Heat capacity of the ice and snow is ignored (i.e. conduction through the snow-ice column is assumed to be uniform). Flux continuity then implies that

$$F_{sfc} = F_{cond} = \frac{T_{sfc} - T_{bot}}{R_{ice}} \qquad (4.2)$$

F_{cond} represents vertical conduction through the ice, T_{bot} ice base temperature and $R_{ice} = h_{ice}/k_{ice} + h_{snow}/k_{snow}$ represents the thermal insulance of the snow-ice column, where h_{ice}, k_{ice}, h_{snow} and k_{snow} represent ice thickness, ice conductivity, snow thickness and snow conductivity respectively.

Equating (4.1) and (4.2), solving for T_{sfc} and re-substituting in (4.1), we get

$$F_{sfc} = (A + BT_{bot})(1 - BR_{ice})^{-1} \qquad (4.3)$$

To characterize dependence of surface flux on ice thickness, we differentiate (4.3) by h_{ice}:

$$\frac{\partial F_{sfc}}{\partial h_{ice}} = B/k_{ice}\,(A + BT_{bot})(1 - BR_{ice})^{-2} \qquad (4.4)$$

The physical interpretation of (4.4) is that the rate of dependence of net surface flux (and hence ice mass balance) on ice thickness is proportional to atmospheric forcing on the ice, represented in (4.4) by $F_{atmos} = (A + BT_{bot})$, and also to the square of the scale factor $SCF = (1 - BR_{ice})^{-1}$, which depends on ice and snow thickness.

Examining the role of the atmospheric forcing, from Equation (4.1) we see that *A* represents the part of the surface flux that is independent of the ice surface temperature on instantaneous timescales. For example, this part clearly includes the downwelling radiative components and the snowfall, and excludes the upwelling LW component. Examining the role of the scale factor *SCF,* for high ice or snow thickness, SCF is small, and surface flux varies only weakly with ice thickness; for low ice and snow thickness, SCF is close to 1, and surface flux is highly dependent on ice thickness.

We now demonstrate the results of applying this method to the three models for the month of January. For the time being, we approximate the surface temperature-independent (TSFI) component of the surface flux, *A*, by the downwelling LW flux $F_{LW\downarrow}$, reasoning that in January the SW contribution is negligible, and the snowfall and TSFI components of turbulent fluxes are small enough to be neglected for the purposes of this example. $F_{LW\downarrow}$is broadly similar in spatial pattern in the three models (Figure 4.3 a-c), with highest values occurring close to the Atlantic ice edge and lowest values close to the north coast of the Canadian Archipelago. However, UKESM1.0-LL displays the lowest values overall, ranging from -115 Wm^{-2} at the Atlantic ice edge to -175 Wm^{-2} near the Canadian Arctic Archipelago. SCF meanwhile varies more strongly between the three models, as one would expect given the very different ice thickness simulations (Figure 4.3 d-f). Entering these quantities in Equation (4.4), we obtain an estimate of the dependence of net LW radiation on ice thickness (Figure 4.3 g-i). For all models, dependence is mostly very weak in

the central Arctic (below 2 Wm^{-2}), but very strong towards the periphery. This is because the stronger thermodynamic forcing in the central Arctic is outweighed by the thicker ice in this region.

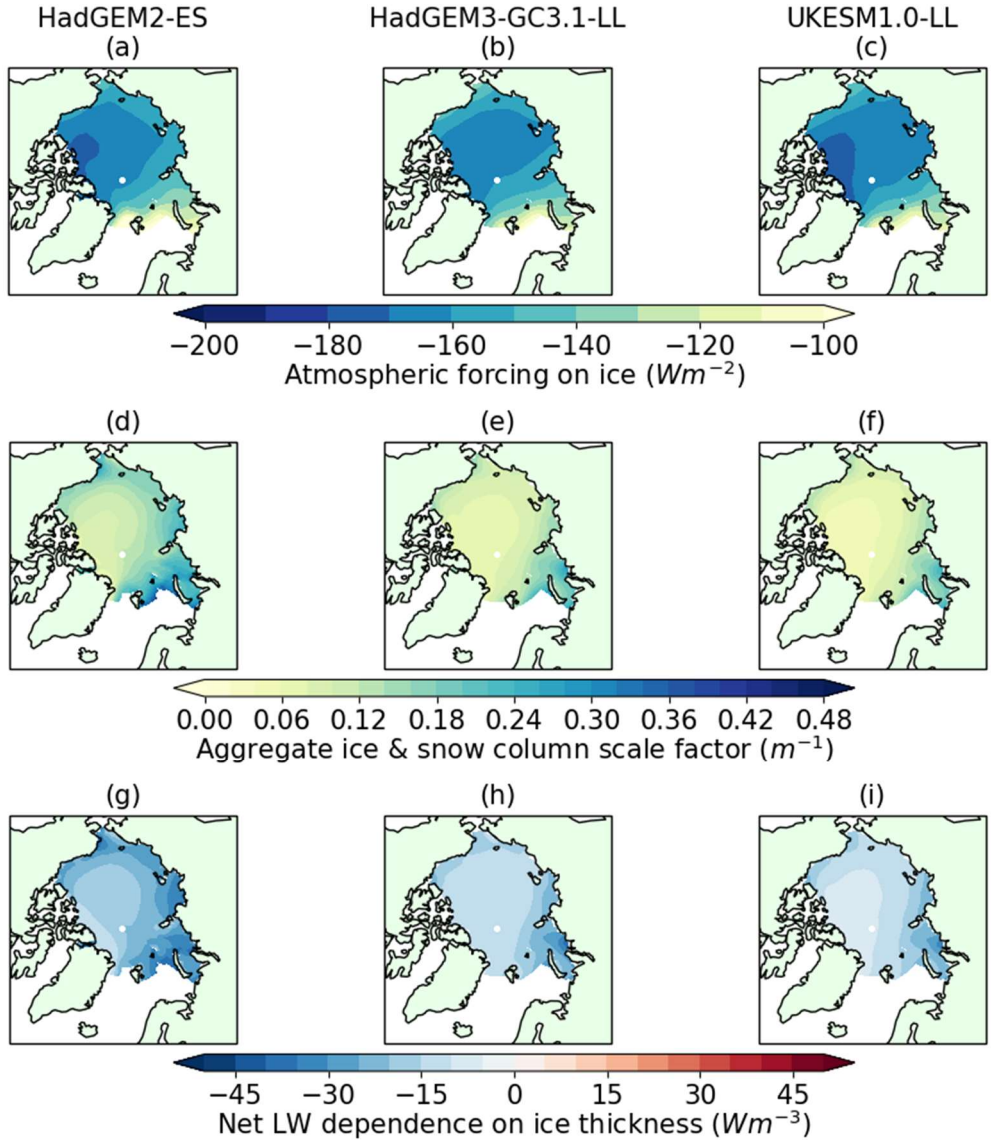


Figure 4.3. A demonstration of the calculation of sensitivity of net LW radiation to ice thickness, for HadGEM2-ES (left column), HadGEM3-GC3.1 (middle column) and UKESM1.0 (right column). (a)-(c) shows atmospheric thermodynamic forcing on the ice, as diagnosed by $F_{LW\downarrow}$; (d)-(f) shows the value of the ice and snow column scale factor SCF; (g)-(i) shows the sensitivity of net LW radiation on ice thickness, the product of the top and middle rows.

4.2 Methods

We proceed to formally derive the ISF method in the following way. Firstly, the two approaches presented above are combined to produce a simple parameterisation of surface flux. This captures the effects of atmospheric forcing and surface albedo drivers on net SW radiation in the melt season, and the effects of atmospheric forcing and of ice and snow thickness on net LW radiation in the freezing season.

Secondly, the parameterised surface flux is compared to the modelled surface flux and it is seen that while the parameterisation captures seasonal and spatial variability well, it systematically underestimates surface heat loss during the freezing season, meaning that dependence of surface heat loss on each model variable will also be understimated. As a result of this comparison, we introduce two refinements that greatly improve the surface flux estimation: use of the ice thickness distribution, and use of spatially and seasonally-varying rate of dependence of upwelling LW radiation on surface temperature.

Formally, the model surface flux is approximated at each point in model space *x* and time *t* by an explicit function $g_{x,t}$ of quasi-independent variables v_i. Hence the dependence of the surface flux on each of the independent variables at point *x*, time *t* can be approximated by $\left[\partial g_{x,t}/\partial v_i\right]^{MODEL}$. Given a model bias in variable v_i at *(x,t)* we can then estimate the surface flux bias induced by that model bias as $\left[\partial g_{x.t}/\partial v_i\right]^{MODEL} \partial g_{x,t}/\partial v_i \left(v_{i,x,t}^{MODEL} - v_{i,x,t}^{REFERENCE}\right)$.

We begin by constructing the $g_{x,t}$, using the ideas of Chapter 4.1 as guidance, starting from the basic equation for surface flux:

$$F_{sfc} = \left(1 - \alpha_{sfc} F_{SW\downarrow}\right) + F_{LW\downarrow} - \varepsilon_{sfc}\sigma T_{sfc}^4 + F_{sens} + F_{lat} + F_{snowfall} \qquad (4.5)$$

Here F_{sfc} represents surface flux, α_{sfc} surface albedo, ε_{sfc} surface emissivity, σ the Stefan-Boltzmann constant, T_{sfc} surface temperature, F_{sens} sensible heat flux, F_{lat} latent heat flux, and $F_{snowfall}$ the heat flux due to the transfer of snow from the atmosphere to the ice system.

We neglect the sensible, latent and snowfall contributions. We divide the remaining terms into a surface temperature-dependent (TSFD) and surface

temperature-independent (TSFI) component. Downwelling SW and LW radiation are instantaneously independent of surface temperature; at the winter temperatures for which this approximation is designed, surface albedo is also independent of surface temperature. Hence the TSFD component is represented entirely by the upwelling LW term, $\varepsilon_{sfc}\sigma T_{sfc}^4$. We linearise this term about $T_{freeze} = 0$℃ by adding $\varepsilon_{sfc}\sigma T_{freez}^4$ to the TSFI term. Hence, as in equation (4.1), we have $F_{sfc} = A + BT_{sfc}$, where $A = \left(1 - \alpha_{sfc}F_{SW}\right) + F_{LW\downarrow} - \varepsilon_{ice}\sigma T_{freeze}^4$ and $B = 4\varepsilon_{sfc}\sigma T_{freeze}^3$.

Equation (4.5) then expresses the dependence of surface flux on downwelling LW and SW, and on ice and snow thickness, with *A* and *B* as above.

We further parameterize surface albedo using a simple scheme based on HadGEM2-ES. Each grid cell is divided into 3 surface types: open water, bare ice and snow, the latter two of which are judged to be undergoing surface melting if surface temperature exceeds -1°C. Surface albedo α_{sfc} is then characterised as

$$\alpha_{sfc} = \sum_{i=1}^{3} A_i\alpha_i \qquad (4.6)$$

where *i* ranges over surface types, A_i represents the fractional area of surface type *i*, and α_i represents the albedo of surface type *i*.

Open water and total ice fraction are explicitly modeled and diagnosed by all 3 models under discussion. Snow fraction is parameterized, in all 3 models, from snow thickness, using either Equation (3.1) (HadGEM2-ES) or Equation (3.2) (HadGEM3-GC3.1-LL and UKESM1.0-LL).

In summary, the basic assumptions allow surface flux dependence on the basic Arctic climate variables to be characterized as $F_{sfc} = (A + BT_{bot})(1 - BR_{ice})^{-1}$. We can examine how effective this characterization is by comparing the fields of surface flux simulated by this formula to actual model surface flux (Figure 4.4). While seasonal and spatial variation are accurately captured, in winter the magnitude of the (negative) surface flux is severely underestimated, by 5-10 Wm^{-2} in the Central Arctic but by over 30 Wm^{-2} towards the marginal seas. Effectively, the formula underestimates the efficiency with which the ice surface

can lose energy, and by which the ice can gain volume, decreasing confidence in its ability to characterize the surface flux dependence.

For example, in January 1980 in HadGEM2-ES, an average Arctic Ocean modelled surface flux of -45.3 Wm^{-2} compares to -25.1 Wm^{-2} in the simple parameterisation; effectively, only 55% of the energy loss is captured. This is an important problem to resolve, because for most variables dependence on surface flux during winter scales in the same way as the surface flux itself: if only 55% of the heat loss is captured, a similar fraction of the surface flux difference due to difference in an underlying variable will be captured. Hence we introduce two improvements to the basic assumptions used above which greatly improve the surface flux simulation, and allow a greater proportion of the freezing season heat loss to be captured.

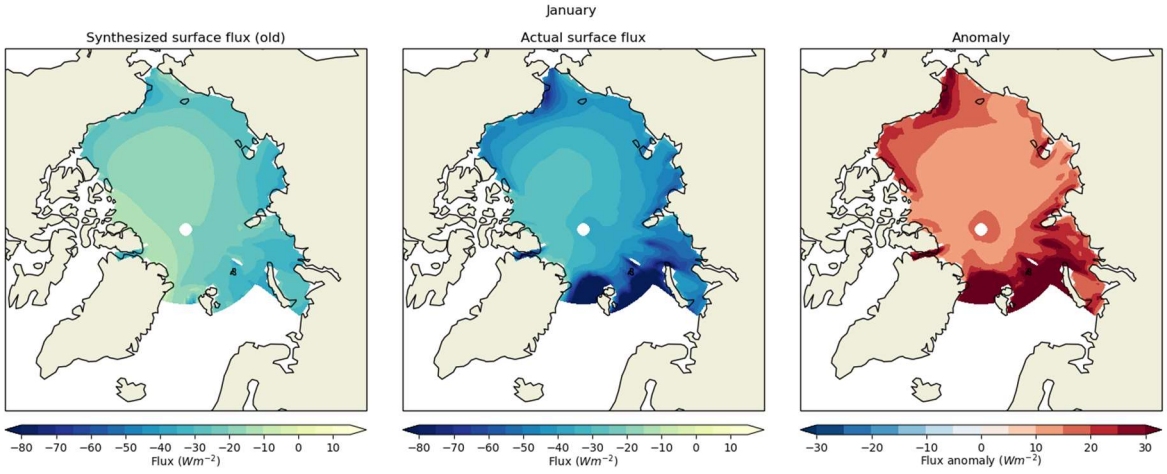


Figure 4.4. (a) Surface flux estimated by equation 4.3; (b) actual model surface flux; (c) anomaly of estimated surface flux relative to modelled, using January 1980 in HadGEM2-ES as an example.

Firstly, we observe that each model uses a sub-gridscale ice thickness distribution, with five categories of sea ice in each grid cell, each with their own area, surface exchange fluxes, and hence mass balance. It is well known that the use of such a parameterisation increases the efficiency of ice production (e.g. Holland et al., 2006) and hence its neglect in our simple model is likely to be a major factor in our underestimating surface energy loss.

Secondly, we allow the process by which surface flux is linearised about surface temperature to vary according to space and time. In Equation 4.1, the

linearization is carried out about $T_{sfc} = 0℃$, which leads to the parameter *B* taking the value ~4.5 $Wm^{-2}°K^{-1}$. In reality, throughout most of the Arctic winter the surface temperature is much colder, and the dependence of upwelling LW radiation on surface temperature is considerably lower. Hence we allow *B* to vary in space and time: at each model gridpoint we linearise about $T_{sfc} = T_{sfc-ref}$, where $T_{sfc-ref}$ is the monthly mean surface temperature for that gridpoint. Accordingly, $B = 4\varepsilon_{ice}\sigma T_{sfc-ref}^3$. This leads to a simulated surface flux that is higher in magnitude (more strongly negative) as upwelling LW does not adjust as strongly to cold surface temperatures.

Use of these two refinements greatly improves correspondence between parameterised and modelled surface flux during the winter months in all models. Using our previous example, with actual modelled surface flux of -45.3 Wm^{-2} and parameterised surface flux of -25.1 Wm^{-2}, use of varying B lowers surface flux to -29.9 Wm^{-2}, while use in addition of the ice thickness distribution lowers it further to -38.9 Wm^{-2}; 84% of the modelled surface heat loss is now captured, raising confidence in the ability to quantify differences and biases in this heat loss. The effect in other models and months is similar.

In more detail, we modify the simple model by expressing surface flux as a sum over all ice categories, including open water:

$$F_{sfc} = \sum_{i=0}^{5} a_i F_{sfc}^i \qquad (4.7)$$

Here a_i represents the fractional area of a grid cell covered by ice of category *i,* where *i=0* represents open water, *i=1* the thinnest ice and *i=5* the thickest. F_{sfc}^i represents the average surface flux over category *i*.

We repeat equation (4.5) for each category separately. We continue to assume latent and sensible heat fluxes are zero($F_{lat}^i = 0; i = 1, \ldots, 5$). We also assume for simplicity that upwelling SW flux flux does not vary over ice categories. Downwelling SW and LW fluxes are naturally uniform over a grid cell. Hence for each ice category, we have

$$F_{sfc}^i = F_{cond}^i \qquad (4.8)$$

which gives

$$A + BT_{sfc}^{i} = \frac{1}{R_{ice}^{i}}\left(T_{sfc}^{i} - T_{bot}\right) \qquad (4.9)$$

where $A = (1-\alpha_{ice})F_{SW\downarrow} + F_{LW\downarrow} - \varepsilon_{ice}\sigma T_{sfc-ref}^{4}$ represents the TSFI component of the surface flux, $B = 4\varepsilon_{ice}\sigma T_{sfc-ref}^{3}$ as above, T_{sfc}^{i} represents surface temperature over category *i*, and $R_{ice}^{i} = h_{ice}^{i}/k_{ice} + h_{snow}^{i}/k_{snow}$ represents the thermal insulance of the snow-ice column in category *i.* In a similar way to above, T_{sfc}^{i} can be eliminated to give

$$F_{sfc}^{i} = \frac{A - BT_{bot}}{1 - BR_{ice}^{i}} \qquad (4.10)$$

for *i=1,...,5*.

We also approximate the surface flux over open water as

$$F_{sfc}^{0} = (1-\alpha_{water})F_{SW\downarrow} + F_{LW\downarrow} - \varepsilon_{water}\sigma T_{f}^{4} \qquad (4.11)$$

Combining (4.6), (4.10) and (4.11) gives a system of equations $g_{x,t}$, expressing surface flux as functions of basic model diagnostics, that takes account of all three modifications.

The functions $g_{x,t}$ are constructed in such a way as to capture basic relationships between atmospheric forcing and sea ice state variables in a manner that best represents the conditions at point *x* and time *t*. In addition, the function captures the indirect effect of any model bias on surface flux via surface temperature and upwelling LW, which will tend to counteract the direct effect to a degree. Hence the dependence of the surface flux on each of the independent variables at point *x*, time *t* can be approximated by $\left[\partial g_{x,t}/\partial v_i\right]^{MODEL}$. Given a model bias in variable v_i at *(x,t)* we can then estimate the surface flux bias induced by that model bias as $\left[\partial g_{x.t}/\partial v_i\right]^{MODEL} \partial g_{x,t}/\partial v_i \left(v_{i,x,t}^{MODEL} - v_{i,x,t}^{REFERENCE}\right)$. Here, the reference dataset could represent a different model, or an observational estimate of the real world. Where the reference dataset is a model, it is usually possible to approximate surface flux dependence by $\left[\partial g_{x,t}/\partial v_i\right]^{MODEL} .\frac{1}{2}\left[\partial g_{x,t}/\partial v_i\right]^{MODEL} + \frac{1}{2}\left[\partial g_{x,t}/\partial v_i\right]^{REFERENCE}$ instead.

The usefulness of this approach is that surface flux operates linearly on the sea ice mass balance, meaning that each of the ISF biases at *(x,t)* can be averaged

over large regions of time and space to understand large-scale sea ice biases. Clearly, none of the driving model biases operate on the sea ice state in a linear sense. For example, given identical surface melt onset (and hence surface albedo) biases at different points in the Arctic, each could have very different implications for local sea ice mass balance, depending on the downwelling SW modelled at each point. Conversely, identical downwelling SW biases would have different implications for sea ice mass balance depending on the modelled surface albedo. Model bias in ice thickness behaves in a particularly nonlinear fashion, with bias in regions of thinner ice having far more influence than that in regions of thicker ice. Because we estimate ISF bias at each point separately, and then average to determine large-scale effects, all nonlinearities are bypassed.

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

4.3 Results

4.3.1 Arctic aggregate ISF biases and internal variability

For each model, we plot aggregate ISF difference over the Arctic Ocean region due to difference relative to observations of downwelling SW and LW, melt onset, ice area and ice thickness (Figure 4.5). Where multiple observational datasets exist (as for ice area and the radiative terms), ISF differences relative to all reference datasets are shown. We add these contributions together to produce a total ISF difference, indicating the total surface flux difference (and

hence difference in sea ice growth / melt) estimated to be caused by the different model biases examined here. For the total ISF difference, observational uncertainty is indicated by showing all possible combinations of reference datasets. In the case that all observational references indicate an ISF difference of the same sign due to a particular variable, we say that the model demonstrates an 'ISF bias' due to biases in that variable.

For HadGEM2-ES, the total ISF difference is positive in the summer and negative in the winter (Figure 4.5a), with total ISF difference averaging 14.9 Wm^{-2} from June-August and -8.8 Wm^{-2} from October-April (in this section, the mean across observational datasets is quoted unless otherwise specified). This is equivalent to 38cm additional ice melt in summer and 52cm additional ice growth in winter. The dominating term in summer is the ice area bias (contributing 13.9 Wm^{-2}, or 35cm), although there is also a significant contribution from the surface melt onset term in June (contributing 1.9 Wm^{-2}, or 5cm averaged over the summer). The contribution of the radiative terms in summer is difficult to determine due to very high observational uncertainty, as ERAI downwelling SW (LW) fluxes are substantially lower (higher) than those from ISCCP and CERES. The dominating terms in winter are from the downwelling LW bias (contributing 5.1 Wm^{-2}, or 30cm additional ice growth) and from the ice thickness bias (contributing -4.1 Wm^{-2}, or 24cm ice growth). HadGEM2-ES displays a clear bias in total ISF in all months of the year except May, July and September.

For HadGEM3-GC3.1-LL, the total ISF is not biased in any month of the year, as the standard deviation across reference datasets encompasses 0 Wm^{-2} (Figure 4.5b). During the freezing season, this is a result of cancelling model errors: positive ISF biases from the ice area and ice thickness terms (averaging 1.8 and 3.9 Wm^{-2} respectively from October – April) are cancelled by a negative ISF bias from the downwelling LW term (-5.1 Wm^{-2}). During the melting season, the ISF difference may be influenced by difference in downwelling radiation and ice area, but in this case all three individual terms also display very high observational uncertainty.

Finally, for UKESM1.0-LL also, the total ISF is not biased in any month of the year. However, unlike HadGEM3-GC3.1-LL most datasets indicate weakly

negative in the melt season and weakly positive in the freezing season (Figure 4.5c), with total ISF difference averaging -5.4 Wm^{-2} from June-August and 2.5 Wm^{-2} from October – April. This is equivalent to 14cm reduced sea ice melt in the summer and 14cm reduced sea ice growth in the winter. In the melt season, the dominating terms arenegative downwelling LW differences (4.0 Wm^{-2}, or 10cm reduced ice melt) and from positive ice area differences (2.2 Wm^{-2}, or 6cm reduced ice melt). In the winter, negative ISF biases from the downwelling LW term (-4.7Wm^{-2}) are more than countered by positive biases from the ice area and ice thickness terms (2.2 Wm^{-2} and 4.9 Wm^{-2}). For both HadGEM3-GC3.1-LL and UKESM1.0-LL, a key advantage of the ISF analysis is demonstrated: neither model displays a bias in total ISF during the freezing season, but this is shown to result from compensating model errors due to ice thickness bias and downwelling LW radiation bias.

Two general observations are apparent. Firstly, the total ISF biases are qualitatively consistent with the ice growth / melt biases demonstrated in Section 3.1. HadGEM2-ES displays a large positive bias in ice growth / melt (38cm), HadGEM3-GC3.1 a much smaller bias (7cm) and UKESM1.0-LL a weak negative bias (12cm). (In section 4.3.5 below the total ISF biases are evaluated in more detail relative to the net radiation and ice growth / melt biases of the models.) The ISF analysis therefore immediately shows the proximate causes of the sea ice growth / melt biases.

Secondly, the ice area and ice thickness biases mostly dominate the ISF biases, with other terms only significant in a few cases, most notably melt onset in June in HadGEM2-ES, and downwelling LW in the freezing season in all models. Hence the major part of the biases in ice growth / melt are caused by the ice state itself. We examine the implications of this in more detail in section 4.3.3 below.

We now compare the models in more detail by examining how each term differs between the model in turn.

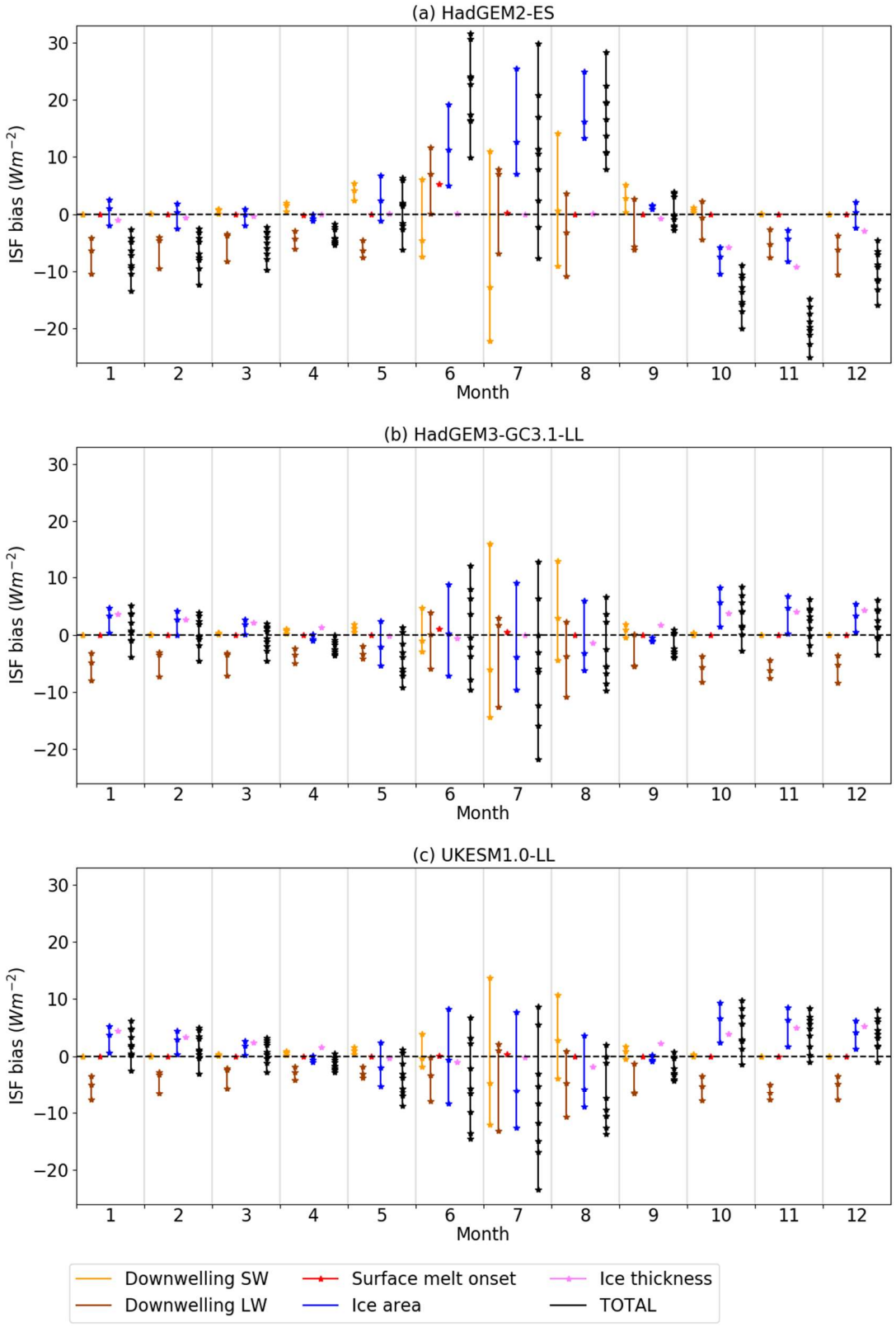


Figure 4.5. Individual and total ISF biases, averaged over the 1980-1999 period and the Arctic Ocean region, for (a) HadGEM2-ES, (b) HadGEM3-GC3.1 and (c) UKESM1.0. Where more than one observational dataset is available, ISF

bias relative to each dataset is shown. For the total ISF bias, that resulting from all possible combinations of reference datasets is indicated.

4.3.2 Comparing ISF totals between models term by term

We begin by examining the ISF terms corresponding to the fundamental variables of the sea ice state, ice area and thickness. Looking first at the ice area term, we see that ice area difference induces large ISF differences in all three models (Figure 4.6a), reflecting the fundamental way in which sea ice alters surface energy balance in both summer (through surface albedo) and winter (through insulation of the warm ocean from the cold atmosphere). However, observational uncertainty is also high, both in summer, when uncertainty in the ice area is highest, and in winter, when the dependence of surface flux on ice area is highest.

Throughout the summer, ice area difference induces a negative surface flux difference in UKESM1.0-LL relative to NSIDC and HadISST1.2, but not relative to HadISST.2.2. Meanwhile, in HadGEM3-GC3.1-LL ice area difference induces a positive surface flux difference relative to HadISST1.2 and HadISST.2.2 throughout the summer, but not relative to NSIDC. In HadGEM2-ES, meanwhile, the ice area ISF bias is unambiguously positive from June-August, indicating additional ice melt. Again, this is consistent with the volume balance simulations.

During the freezing season (October – April) HadGEM3-GC3.1 and UKESM1.0 display positive ice area ISF differences relative to NSIDC and HadISST1.2. These indicate that across some parts of the Arctic, open water fraction is lower in models than in observations, which would normally be associated with less heat loss to the atmosphere and correspondingly less ice growth. However, observational uncertainty is considerable, and the winter ice area ISF differences are near-zero when HadISST.2.2 is reference dataset.

Looking next at the ice thickness term (Figure 4.6b), the ice thickness ISF biases of UKESM1.0-LL and HadGEM3-GC3.1-LL are very similar, rising to 6 Wm^{-2} and 5 Wm^{-2} respectively during the freezing season. This is very different to HadGEM2-ES, whose bias is strongly negative throughout the winter, reaching a minimum of –10 Wm^{-2} in November. Given the much thicker ice in UKESM1.0-LL than in HadGEM3-GC3.1-LL, it is on the surface surprising that

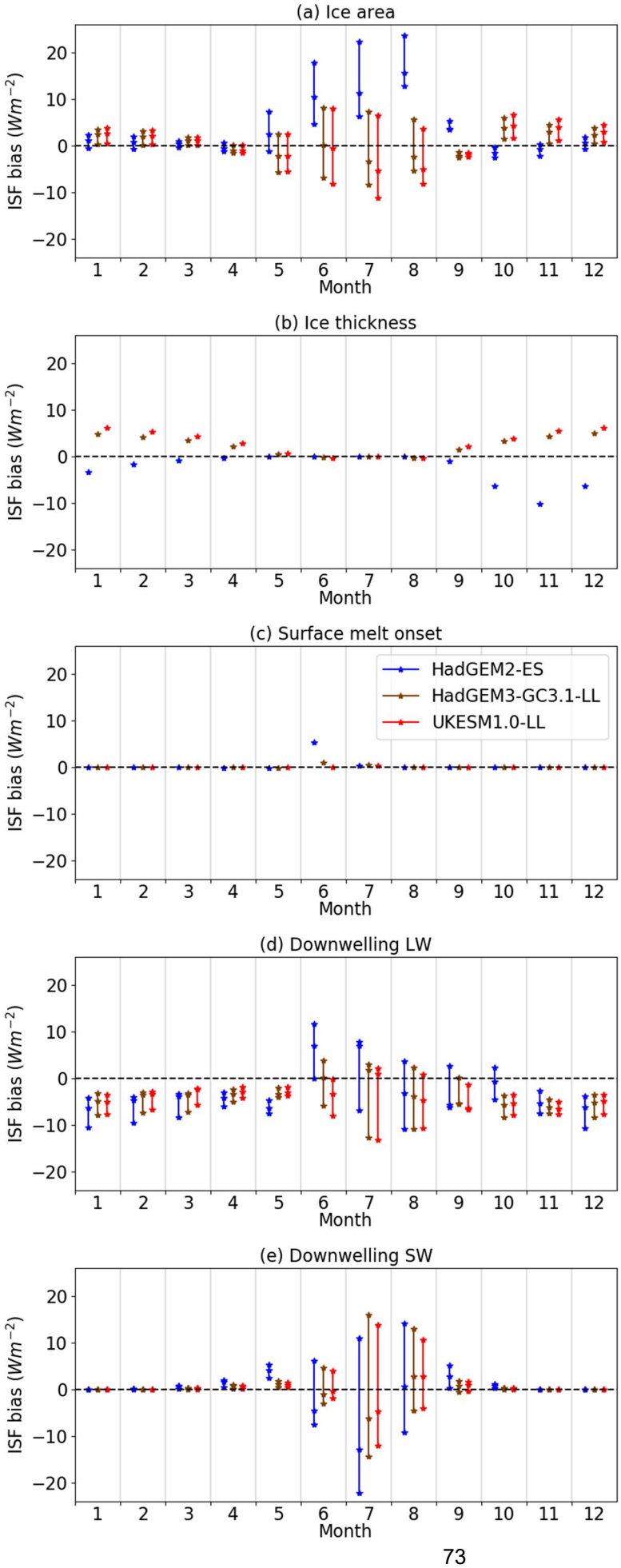
the ISF biases of these two models are so similar. This may be related to the way in which the differences between UKESM1.0 and HadGEM3-GC3.1 are distributed between the ice categories. The sea ice thickness distributions of the two newer models are nearly identical across much of the Arctic in all but the thickest category, where the average thickness of UKESM1.0-LL is around 2m thicker than HadGEM3-GC3.1-LL. In this category, however, the dependence of surface flux on thickness is very small, so this large difference in ice thickness bias contributes to a correspondingly small difference in the ISF bias contribution. This effect is demonstrated more explicitly in Chapter 5.

We now examine the ISF differences due to other model variables. Looking next at the melt onset term (Figure 4.6c), this is generally small outside the month of June. In HadGEM2-ES, it is positive in June (5 Wm^{-2}); in HadGEM3-GC3.1 weaker, and in UKESM1.0-LL near-zero. The contrast between the models is consistent with the comparison of date of melt onset in Chapter 3, with surface melt onset occurring much earlier than in SSMI observations in HadGEM2-ES, and a much lesser bias in UKESM1.0-LL. It is also consistent with the volume balance simulations, as the bias would cause a strong bias towards additional ice melt in HadGEM2-ES, but a much weaker bias in UKESM1.0.

Looking next at the downwelling LW term (Figure 4.7a), during the winter and early spring this term tends to be lowest (most negative) in HadGEM2-ES, and highest (least negative) in UKESM1.0. This implies that that in all models the Arctic atmosphere is biased cold, but the bias is worst in HadGEM2-ES (leading to the strongest bias towards ice growth) and least in UKESM1.0 (leading to the weakest bias towards ice growth). This is consistent with the volume balance, as winter ice volume growth is strongest in HadGEM2-ES and weakest in UKESM1.0. It is less obviously consistent with the annual mean ice volume, which is least in HadGEM2-ES and highest in UKESM1.0 despite the colder winter atmosphere in the former model. To understand this paradox, it is necessary to consider the other terms.

Looking finally at the ISF downwelling SW term, this is very small in size outside May-August due to the downwelling SW itself being very small outside these months. There is a positive ISF bias during May that is nearly uniform across

the Arctic in all three models, highest in HadGEM2-ES and least in UKESM1.0, but this bias is in the monthly mean opposite and approximately equal to the downwelling LW bias. During the summer months observational uncertainty is very high, and ISF biases cannot be discerned, but we see a similar strong anticorrelation between ISF SW and LW differences that mirrors an anticorrelation in the radiative biases themselves. This reflects the fact that cloud fraction explains a large part of the variation in downwelling radiation during summer, and has opposite effects on SW and LW radiation. Each observational dataset of downwelling SW and LW radiation is associated with a radiative transfer model with fully consistent cloud fraction, and hence any model-observation pair that has a large difference in one radiative term is likely to exhibit the opposite difference in the other term.


(a) Ice area
(b) Ice thickness
(c) Surface melt onset
(d) Downwelling LW
(e) Downwelling SW
ISF bias (Wm^{-2})
Month
HadGEM2-ES
HadGEM3-GC3.1-LL
UKESM1.0-LL

1

Figure 4.6. Individual ISF terms in the different models. Showing surface flux bias induced by biases in (a) ice area, (b) ice thickness, (c) surface melt onset, (d) downwelling LW and (e) downwelling SW, averaged over 1980-1999 and over the Arctic Ocean region, for HadGEM2-ES, HadGEM3-GC3.1-LL and UKESM1.0-LL.

4.3.3 Using the ISF biases to understand sea ice state biases via forcings and feedbacks

It is helpful at this stage to return to the conceptual picture presented in Chapter 1. Sea ice volume is driven by the volume balance, and hence energy balance, at the top and basal surfaces of the ice. These in turn are driven partly by external factors (downwelling radiation, ocean heat flux), but partly by aspects of the sea ice state (conduction, surface albedo, surface temperature). These variables are in turn partly driven by the sea ice thickness and area, which determine the volume. Hence the atmospheric drivers, and to a lesser extent the ocean, drive the sea ice growth and melt, and hence the sea ice state evolution, but the sea ice state itself modulates the response of the ice growth and melt to the external drivers.

A major advantage of the ISF analysis is that individual components of the surface flux biases can be identified either with external drivers in the atmosphere, or with specific feedback mechanisms of the sea ice state. For example, the downwelling radiative components of the ISF differenceessentially represent external drivers of the sea ice state on instantaneous timescales. By contrast, the sea ice area ISF difference in summer operates as a feedback: under truly heterogeneous ice conditions, a sea ice volume balance bias is associated with a nonzero rate of change in the sea ice area bias. A negative difference in ice area, for example, will induce a positive surface flux difference and a negative ice volume balancedifference, causing the negative difference in ice area to grow. In other words, the sea ice concentration component of the ISF difference is, under melting conditions, directly associated with the surface albedo feedback.

On the other hand, the ice thickness ISF difference during the freezing season can be identified with the thickness-growth feedback. This is perhaps less obvious, as the ice thickness affects the estimated surface flux via the surface

temperature and upwelling LW radiation, while the thickness-growth feedback is usually understood to result from differences in conduction. However, the assumption of flux continuity at the surface in constructing the estimated surface flux means that the cooler surface temperatures, and shallower temperatures gradients occurring for thicker ice categories are manifestations of the same process. Slower ice growth at higher ice thicknesses is caused by a smaller negative surface flux, and the surface temperature is the mechanism by which this is demonstrated. Hence the effect of the thickness-growth feedback is described by the ice thickness-induced component of the surface flux difference.

Although the ice area term during the winter does not strictly correspond to the thickness-growth feedback, it represents a very similar effect. A positive difference in ice area during freezing season is normally associated with a positive difference in surface flux, due to the greater ease by which energy can be lost from the warm ocean to the cold atmosphere in ice-free regions. This surface flux difference would feed back on the ice area difference via the volume balance in a very similar manner to the ice thickness difference. Hence for the purposes of the discussion below, it is identified as an additional component of the thickness-growth feedback.

The surface melt onset ISF component is not associated with either first-order feedback (surface albedo or thickness-growth), because surface melt onset is not related to ice volume change on an instantaneous timescale in the same way as ice area or thickness. Hence for the purposes of the discussion below, it is identified as an additional external forcing on the sea ice state.

Hence the ISF analysis allows the contribution of the surface albedo feedback, the thickness-growth feedback, and the various external forcings, to the sea ice volume balance biases to be separated and quantified. Diagnosed in this way, the surface albedo feedback contributes 13.9 Wm^{-2}, -0.5 Wm^{-2} and -2.2 Wm^{-2} to the surface energy balance from June-August for HadGEM2-ES, HadGEM3-GC3.1 and UKESM1.0 respectively; these are equivalent to an additional 35cm of melt, and reduced melt of 1cm and 6cm, respectively. Diagnosed from the ice thickness ISF term, the thickness-growth feedback contributes -4.1, 3.9 and 4.9 Wm^{-2} to the surface energy balance from October-April in the three models

respectively; this is equivalent to an additional 24cm of growth, and reduced growth of 23cm and 29cm, respectively. In addition, the ice area term contributes reduced growth of 11cm and 13cm for HadGEM3-GC3.1-LL and UKESM1.0-LL, respectively (for HadGEM2-ES it contributes no substantial ISF bias during the freezing season overall)

Examining the external forcings, the downwelling LW term during winter reduces ice growth across all models, by 30cm, 30cm and 28cm respectively. The surface melt onset term during summer induces additional ice melt of 5cm, 1cm and 0cm, respectively.

To summarise the effect of the external forcings on the sea ice state, in HadGEM2-ES substantial additional sea ice growth in winter, and sea ice melt in summer, is directly forced: the result is that seasonal ice growth/melt is too large and annual mean ice thickness too low. In HadGEM3-GC3.1 additional forced sea ice growth is lower, but additional forced sea ice melt is absent: the result is that seasonal ice growth/melt is too small and annual mean ice thickness too high. In UKESM1.0-N96 additional forced sea ice growth is lower, but sea ice melt is actually reduced; the result is that seasonal ice growth/melt is far too low, and annual mean ice thickness far too high.

The feedback terms explain how the external forcings produce the sea ice states. In particular, they explain how a difference in melt onset forcing can have a much larger effect on sea ice state than a difference in downwelling LW forcing, simply because of the time of year at which these occur. The melt onset forcing, by inducing additional ice melting through its effect on the ice albedo, enhances subsequent sea ice melt through the surface albedo feedback. The downwelling LW bias, on the other hand, by inducing additional ice freezing through its cooling effect, attenuates subsequent sea ice freezing through the thickness-growth feedback. Surface flux biases induced by melt onset occurrence are enhanced, while those induced by downwelling LW are diminished.

To summarise, the surface melt onset differences are an important driver of both the annual mean ice thickness differences, and the amplitude differences, but these effects are very weakly opposed by the downwelling LW differences.

4.3.4 Spatial patterns

We examine spatial patterns in the ISF difference terms, in turn, in the months or seasons in which they are most dominant. Firstly we plot the ice area ISF difference over the summer (JJA), taking a mean over observational datasets, seeing that the highest values occur towards the coast of the Arctic Ocean, as this is where most of the ice area variability occurs (Figure 4.7a). In HadGEM2-ES high positive values (above 10 Wm^{-2}) are shown in all regions except in the Canadian Arctic, rising above 50 Wm^{-2} in parts of the Kara and Barents Seas north of Western Siberia. In HadGEM3-GC.1-LL the picture is more mixed, with positive differences approaching 12-13 Wm^{-2} in the Central Arctic and 30 Wm^{-2} in the Kara and Chukchi Seas, but negative ISF differences up to about -10 Wm^{-2} in parts of the Canadian Arctic and in the East Siberian Sea. In UKESM1.0-LL the picture is similar to HadGEM3-GC3.1-LL, but the areas of negative difference are stronger and more widespread.

In all three models, the highest magnitude ISF differences occur towards the Arctic Ocean coasts, as this is where the most ice area variability occurs. Hence ice area biases tend to induce differences in ice melting preferentially near the Arctic Ocean coasts.

The ice thickness ISF biases are examined over the winter (DJF; Figure 4.7b). In HadGEM2-ES, winter biases are negative throughout most of the Arctic, approaching -20 Wm-2 in parts of the Atlantic ice edge. However they are near-neutral or positive in the Canadian Arctic and northern Beaufort Sea, the region where the model ice thickness is highest and closest to PIOMAS. The value of using the full ITD in the ISF calculation is shown: across much of the Arctic the ice is too thick in the grid cell average to support the substantial ISF biases seen here. By using the full ice thickness distribution, the variation in the thinnest ice category, where surface flux varies most with ice thickness, is correctly captured.

As in the Arctic mean, the spatial patterns of ice thickness ISF bias in the two newer models is very different to that of HadGEM2-ES (and both are similar to each other). Throughout the winter both are strongly positive throughout the Central Arctic and the Pacific sector, with values mostly from 5-10 Wm^{-2}, indicating a positive ice thickness bias suppressing sea ice growth. However,

the bias decreases to zero towards the Atlantic ice edge (becoming negative in a few small regions), indicating that the positive ice thickness bias vanishes here. The patterns of all three models are indicative of a common characteristic of Arctic sea ice thickness simulation, a tendency to model the region of maximum sea ice thickness too far on the Pacific side of the Arctic. It was shown by Tsamados et al. (2013) that this bias could be reduced by use of an anisotropic sea ice rheology.

The June melt onset term is nearly uniformly positive over the Arctic Ocean in HadGEM2-ES (Figure 4.7c), with highest values (15-20 Wm^{-2}) towards the Central Arctic. In HadGEM3-GC3.1-LL, the pattern is different and weaker, with highest values in the Beaufort Sea of around 6 Wm^{-2}, and small regions of negative ISF bias towards the Atlantic ice edge. In UKESM1.0-LL, the pattern is similar but weaker still, and the highest values of around 3 Wm^{-2} are found in the Beaufort Sea, with many patches of weak negative bias near the Atlantic ice edge and the Siberian coasts.

The LW ISF biases are also examined over the winter (DJF; Figure 4.7d); like the ice area and thickness, this term displays the highest negative values towards the Atlantic and Pacific ice edges (around -9 Wm^{-2} for HadGEM2-ES and -6 Wm^{-2} for HadGEM3-GC3.1-LL and UKESM1.0-LL). This is likely because of the greater efficiency with which downwelling LW biases are converted to surface flux biases in regions of thin ice. In fact, this pattern becomes particularly marked in UKESM1.0 in October (not shown), the only time of year that this model exhibits the highest ISF downwelling LW bias. For this model in the autumn, the driving downwelling LW bias is also highest towards the Pacific and Atlantic ice edges, a pattern which is not apparent for any other model or month. It is possible that the reduced sea ice melt of UKESM1.0, and corresponding high ice area biases, cause a colder atmosphere, and hence less downwelling LW, at this time of year.

The positive ISF biases due to downwelling SW seen in May (Figure 4.7e) are in HadGEM2-ES highest close to the Atlantic ice edge, approaching 20 Wm^{-2} in the Fram Strait, but are much weaker and display no clear spatial signature in the other two models. In summer spatial patterns are again hard to analyse due to the high uncertainty, but it is noted that the SW and LW biases are strongly

anticorrelated in space as well as by model, with regions of negative SW ISF bias associated with a positive LW ISF bias, and vice versa.

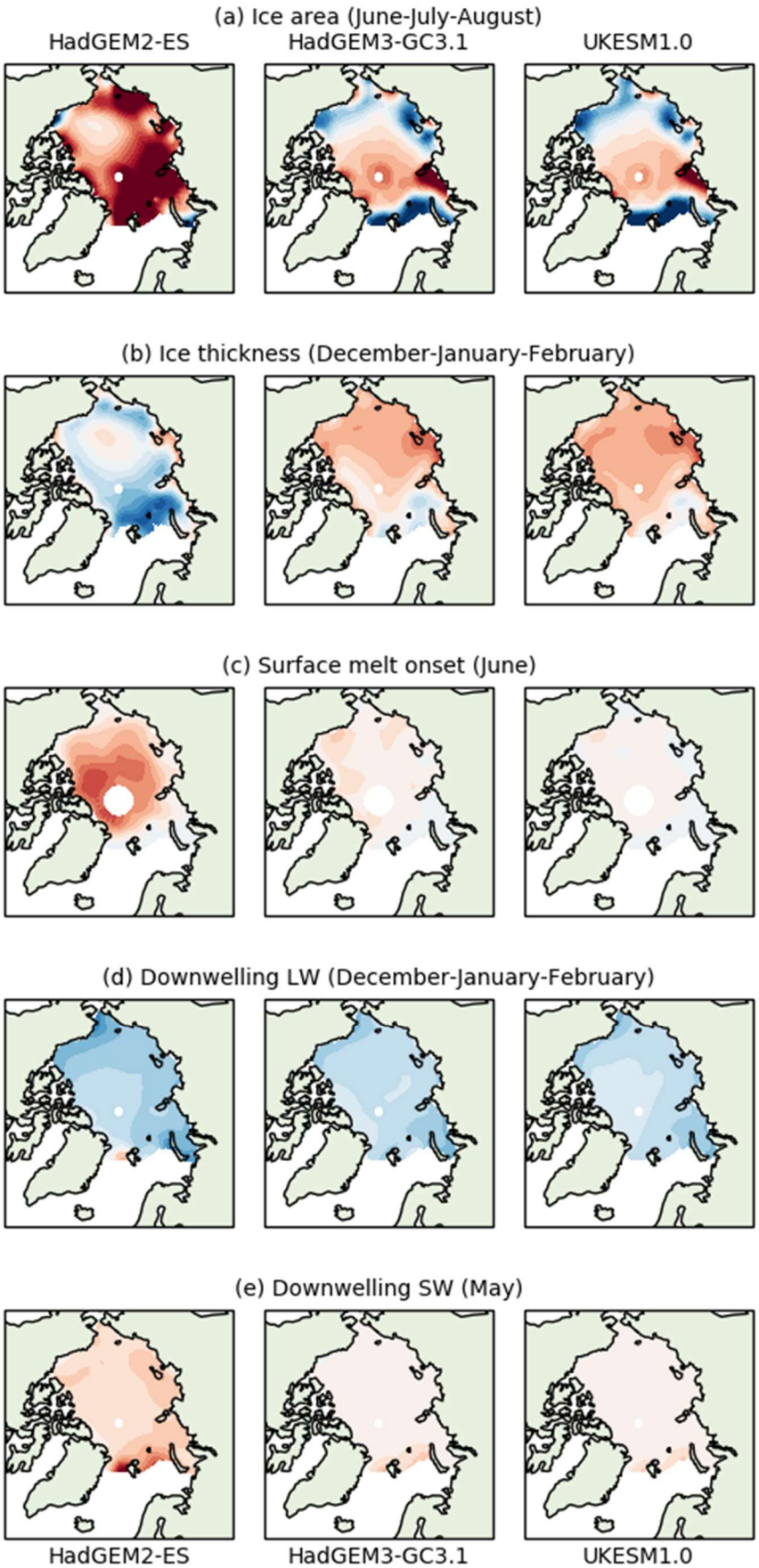

(a) Ice area (June-July-August)
HadGEM2-ES
HadGEM3-GC3.1
UKESM1.0
(b) Ice thickness (December-January-February)
(c) Surface melt onset (June)
(d) Downwelling LW (December-January-February)
(e) Downwelling SW (May)
HadGEM2-ES
HadGEM3-GC3.1
UKESM1.0
−18 −12 −6 0 6 12 18
ISF bias (Wm^{-2})

Figure 4.7. Spatial patterns of individual ISF terms in key months and seasons of the year in HadGEM2-ES (left column), HadGEM3-GC3.1-LL (middle column) and UKESM1.0-LL (right column). Showing maps of surface flux bias due to bias in (a) ice area, June-August; (b) ice thickness, December-February; (c) surface melt onset, June; (d) downwelling LW, December-February; (e) downwelling SW, May.

4.3.5 Observational uncertainty

Having presented the ISF results, limitations are now discussed, beginning with the effect of observational uncertainty which as seen in sections 4.3.1 and 4.3.2 is substantial. Uncertainty is particularly high in the summer months: for example, in HadGEM2-ES in July, ISF bias due to downwelling SW radiation is calculated to be -22, -13, and (+)11 Wm^{-2} using ISCCP-FD, CERES and ERAI respectively, and in general the range of radiative terms during the summer is of the order 10 Wm^{-2}.

Uncertainty in the ice area term during summer is also veryhigh, particularly in July. For example, UKESM1.0 in this month displays ice area ISF bias of -11, -5 and (+)6 Wm^{-2} relative to NSIDC, HadISST1.2 and HadISST.2.2 respectively. The high uncertainty is likely to be partly caused by the greater difficulty of estimating ice area from SSMI observations in the presence of meltponds. In particular, UKESM1.0-LL summer ice area is not biased high relative to HadISST.2.2, despite its exceptionally high ice thickness.

Uncertainty in the ice area term during winter is also considerable. However, this is caused by a very high sensitivity of surface flux to ice area during winter and is not associated with high uncertainty in the ice area itself. For example, for HadGEM3-GC3.1-LL in January surface flux dependence on ice concentration is over 100Wm^{-2} over the entire Arctic Ocean. Despite uncertainty in ice area being of the order ±0.01 in most regions, this produces a relatively high ISF uncertainty for this model in January, with ISF biases of 0.3, 2.4 and 3.6 Wm^{-2} relative to HadISST.2.2, HadISST1.2 and NSIDC respectively. In particular, the high observational uncertainty in this term could go some way towards explaining the apparent discrepancy between ice melt / growth bias and total ISF bias for this model. (Another possibility is that the positive ice melt

/ growth bias in HadGEM3-GC3.1-LL is simply wrong, e.g. because of tendency of PIOMAS to underestimate winter thickness noted in Section 2.2.2).

Uncertainty in the total ISF bias is characterised by calculating total ISF bias using all nine combinations of reference datasets (three radiation datasets, and three ice area datasets). Measured by range across all combinations of reference datasets, uncertainty is highest in July (29-35 Wm^{-2}) and lowest in April (4 Wm^{-2}) for all three models. The total ISF bias is unambiguous in sign only from October-December for UKESM1.0-LL, and for no months in HadGEM3-GC3.1-LL.

In Appendix A, theoretical uncertainty in the ISF biases is analysed by comparing surface fluxes simulated by the simple model to fluxes actually modelled by HadGEM2-ES, and found to be an order of magnitude smaller than the observational uncertainty. Hence observational uncertainty is likely to be the dominating cause of uncertainty in the ISF bias. Despite this, the ISF calculation can give useful information about the likely range of surface flux biases: for example, throughout the winter in HadGEM3-GC3.1-LL and UKESM1.0-LL, the total ISF bias is very likely to be positive.

4.3.6 Evaluating total ISF bias

The ISF biases, summed over all independent variables, should approach the total surface flux bias. This is difficult to evaluate, as the surface flux bias is like other variables subject to large observational uncertainty. We use proxy quantities to evaluate the total ISF bias (Figure 4.8): directly evaluated surface net radiation bias (relative to ISCCP-FD, ERAI and CERES respectively); and ice energy uptake bias, derived from ice volume balance bias relative to PIOMAS. These proxies are shown by horizontal lines, with the light green shaded regions indicating the range of surface flux bias indicated.

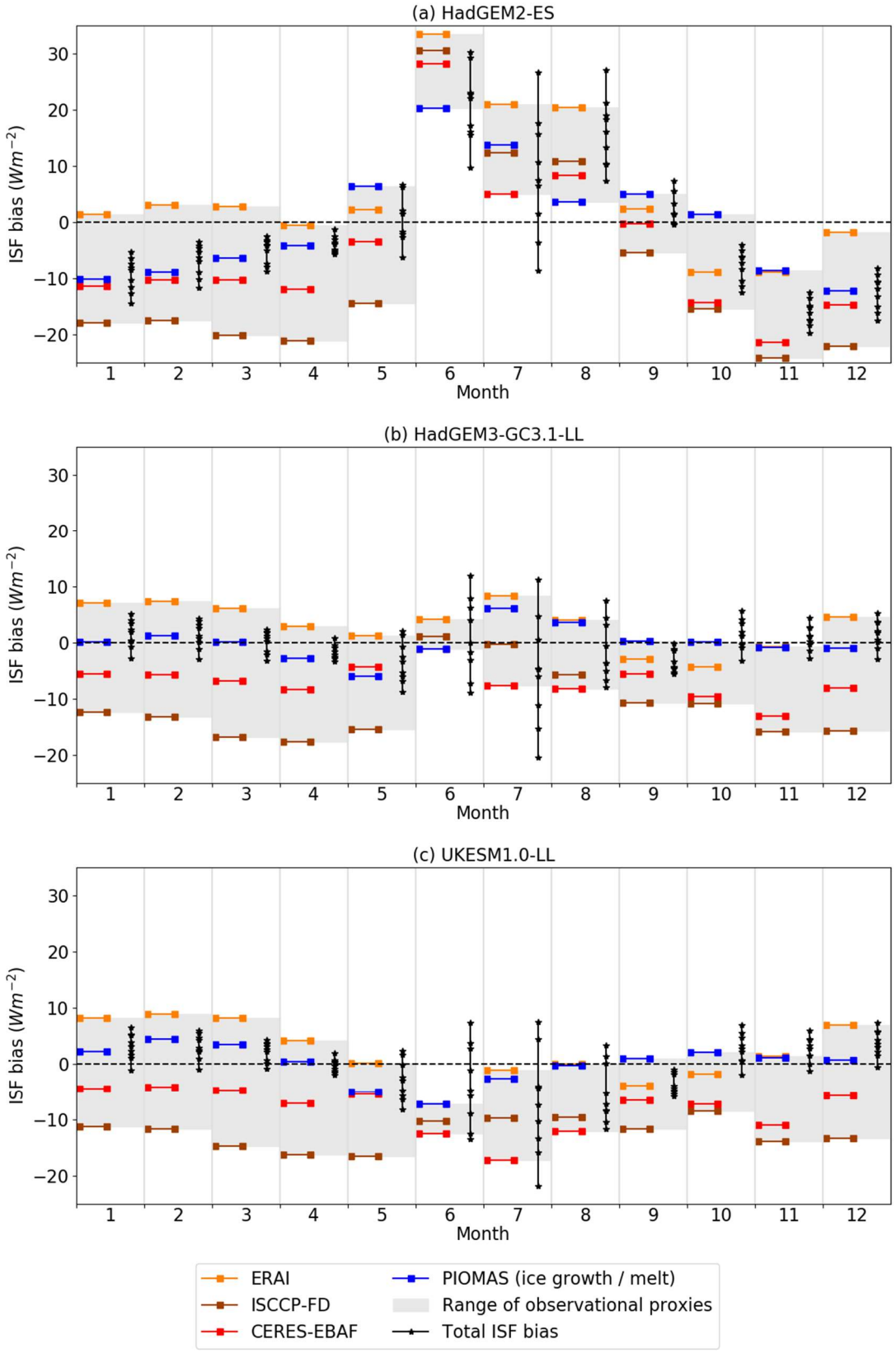


Figure 4.8. Evaluation of total ISF bias in (a) HadGEM2-ES, (b) HadGEM3-GC3.1-LL and (c) UKESM1.0-LL relative to four proxies for surface flux bias: net radiation bias relative to ERAI, ISCCP-FD and CERES-EBAF, and ice growth /

melt bias relative to PIOMAS. The grey shaded area indicates the range indicated by the proxy datasets. For the total ISF bias, range across all possible combinations of observational datasets is shown.

For HadGEM2-ES, there is a close correspondence between the range of total ISF difference and the range of total surface flux difference for all months of the year (Figure 4.6), with a majority of estimates of total ISF difference falling inside the range of proxy surface flux bias for all months of the year. For HadGEM3-GC3.1 and UKESM1.0-LL, there is a small discrepancy during the meteorological autumn, with CERES and particularly ISCCP showing net radiative flux considerably less than the total ISF difference. We note that ISCCP is likely to be the least reliable proxy dataset for the reasons discussed in Section 2.2.3, and that in neither satellite dataset are the radiative fluxes constrained to be physically consistent with each other. Given the multiple lines of evidence that UKESM1.0-LL ice growth is too weak (and hence that the net surface flux bias is positive in winter), it is difficult to conclude from this evidence that the total ISF difference is too high.

The spread in the proxy evaluation datasets is extremely large. For example, while in the month of January the estimates of total ISF bias for HadGEM2- range from -5.3 Wm^{-2} to -14.4 Wm^{-2} depending upon choice of observational reference to ice area and surface radiative fluxes, the estimates of net radiation difference display a considerably larger range, being -18.2, -11.6 and 0.6 Wm^{-2} from ISCCP-FD, CERES and ERAI respectively (ice heat uptake bias from PIOMAS is estimated as -10.1 Wm^{-2}). Hence it is difficult to evaluate the total ISF bias within current observational constraints, and at best it can be said that the total ISF difference is qualitatively consistent, over the year as a whole, with the surface flux difference proxies, except in the meteorological autumn for HadGEM3-GC3.1 and UKESM1.0. A possible cause of the lower total ISF difference in June and July is the ‘missing process’ of snow on ice, which cannot be evaluated here due to the lack of a reference dataset. The early surface melt onset, and sea ice fraction loss, as modelled by HadGEM2-ES, would be associated with an early loss of snow on ice, with an additional surface albedo bias and hence an additional ISF difference.

4.3.7 Missing processes in the ISF analysis

The ISF analysis, as presented above, does not comprise an exhaustive list of processes affecting Arctic Ocean seasonal ice growth/melt to first order. Processes not examined here include: the effects of snow fraction or ice thickness on ice albedo, the effect of snow thickness on heat conduction through ice in winter, the effects of biases in turbulent fluxes not directly associated with ice area biases, and the effects of biases in ocean heat convergence. In this section, we discuss in turn possible errors associated with the omission of these effects.

Snow fraction influences ice volume balance directly, via surface albedo and net SW flux, in all three models examined as well as in the real world, but is not evaluated above due to the lack of a sufficiently well evaluated Arctic-wide reference dataset. A glimpse as to the possible impact of the omission of this effect in HadGEM2-ES can be seen in Figure 4.5a. Total ISF bias is strongly positive in this model in June and August, due to contributions from surface melt onset and ice area respectively, consistent with the net radiative biases and the sea ice heat uptake bias. However, in July the total ISF bias is near-zero, in contrast to the evaluation datasets which still show a strongly positive bias. The missing process of snow fraction is very likely to contribute to this discrepancy, as the early surface melt onset and sea ice fraction loss modelled by HadGEM2-ES would be associated with an early loss of snow on ice, with an additional surface albedo bias and hence an additional ISF bias.

The effect of variations in ice thickness on sea ice albedo is not represented in any of the models evaluated here, meaning that its effect on model biases would be hard to capture using the ISF method. For example, imagine a model grid cell where sea ice melt occurs too early relative to the real world. No additional bias is induced by biases in sea ice thickness during model sea ice melting, as this effect is not modelled. However, later in the season, when real-world sea ice melts in the same location, the model surface flux bias is actually reduced by the ice thickness effect. Conceptually, this paradox arises because the surface flux bias is actually the sum of two distinct ISF biases due to competing effects: the model thickness bias induces a positive ISF bias overall, but its effect is counteracted by a negative ‘structural’ ISF bias that arises not

from any particular model bias, but from the exclusion of a particular process from the model. In this case however, the effects of ice thickness on surface albedo are likely to be small (e.g. Ebert et al., 1995). In section 4.4 below, an example is given of how one might account for such a structural ISF bias when comparing two models.

Model biases in the turbulent fluxes may also be significant contributors to the surface flux biases. During the freezing season, ice concentration biases are likely to be the most important process affecting turbulent fluxes, and this effect is captured by the ISF analysis above. However, atmospheric conditions are also an important factor affecting turbulent fluxes, notably wind speed and boundary layer stability; a more detailed treatment of turbulent fluxes would include these as independent variables. Snowfall itself also representsa component of the surface flux. This is because snow falling on sea ice represents a transfer of negative latent heat from the atmosphere to the ice: energy is released in the atmosphere during deposition of snow crystals on cloud condensation nuclei, and when this snow falls onto ice, energy must first be used to melt this layer before the underlying ice can be melted. This effect is relatively small component of the surface flux, but could also be evaluated directly given a sufficiently reliable observational reference.

A complete treatment of model biases affecting the sea ice volume budget would also examine causes of bias in oceanic heat convergence. These are likely to be small in the Arctic Ocean interior in the evaluated models and in observations, but the model bias could nevertheless conceivably be of considerable size in the context of the surface flux biases shown in Figure 6. The total Arctic Ocean heat convergence modelled by for the period 1980-1999 is 4.4 Wm^{-2}, 3.8 Wm^{-2} and 3.9 Wm^{-2} for HadGEM2-ES, HadGEM3-GC3.1 and UKESM1.0 respectively. These figures show high sensitivity to the location of the boundary in the Atlantic sector, suggesting that most of this heat is released close to the Atlantic ice edge. This figure is slightly higher than the 3 Wm^{-2} found by Serreze et al. (2007) in their analysis of the Arctic Ocean heat budget but is broadly consistent with observational estimates of oceanic heat transport through the Fram Strait from 1997 to 2000 by Schauer et al., 2004 (likely to be the major contributor to Arctic Ocean heat convergence). This suggests that errors in oceanic heat convergence are unlikely to contribute significantly to sea

ice volume biases in the evaluated models. However, for a hypothetical model that simulated greater oceanic heat convergence in the Arctic Ocean interior, the surface flux analysis presented here would fail to adequately describe the model bias in the sea ice volume budget.

4.3.8 Going beyond proximate drivers

The ISF framework focusses attention on the model variables whose biases are particularly likely to be causing biases in the sea ice growth / melt. This is helpful because it points towards important areas for future model improvements. To illustrate this, we briefly discuss possible drivers of the two main biases (other than in the sea ice state itself) identified here: surface melt onset and downwelling LW.

Underestimation of wintertime downwelling LW fluxes in the Arctic is known to be a widespread model bias in the CMIP5 ensemble (e.g. Boeke and Taylor, 2016), probably associated with cloud liquid water deficit caused by errors in cloud microphysics schemes (Pithan et al, 2014). HadGEM2-ES was not one of the models assessed by Pithan et al (2014), but its winter climate simulation displays many of the characteristic CMIP5 biases, namely low cloud liquid water fractions during winter compared to MODIS observations (Figure 4.9a). In addition, observations suggest a bimodal nature of Arctic winter temperatures associated with strong and weak inversions, (e.g. Stramler et al, 2011; Raddatz et al, 2015), and a failure to simulate the milder mode (Figure 4.9b), diagnosed by 6-hourly fluxes of net LW, is a feature of the HadGEM2-ES, HadGEM3-GC3.1-LL and UKESM1.0-LL simulations similar to the models assessed by Pithan et al (2014). Hence it is concluded that a similar mechanism is likely to be at work in all three evaluated models, and that insufficient cloud liquid water is the principal driver of the anomalously low downwelling LW fluxes.

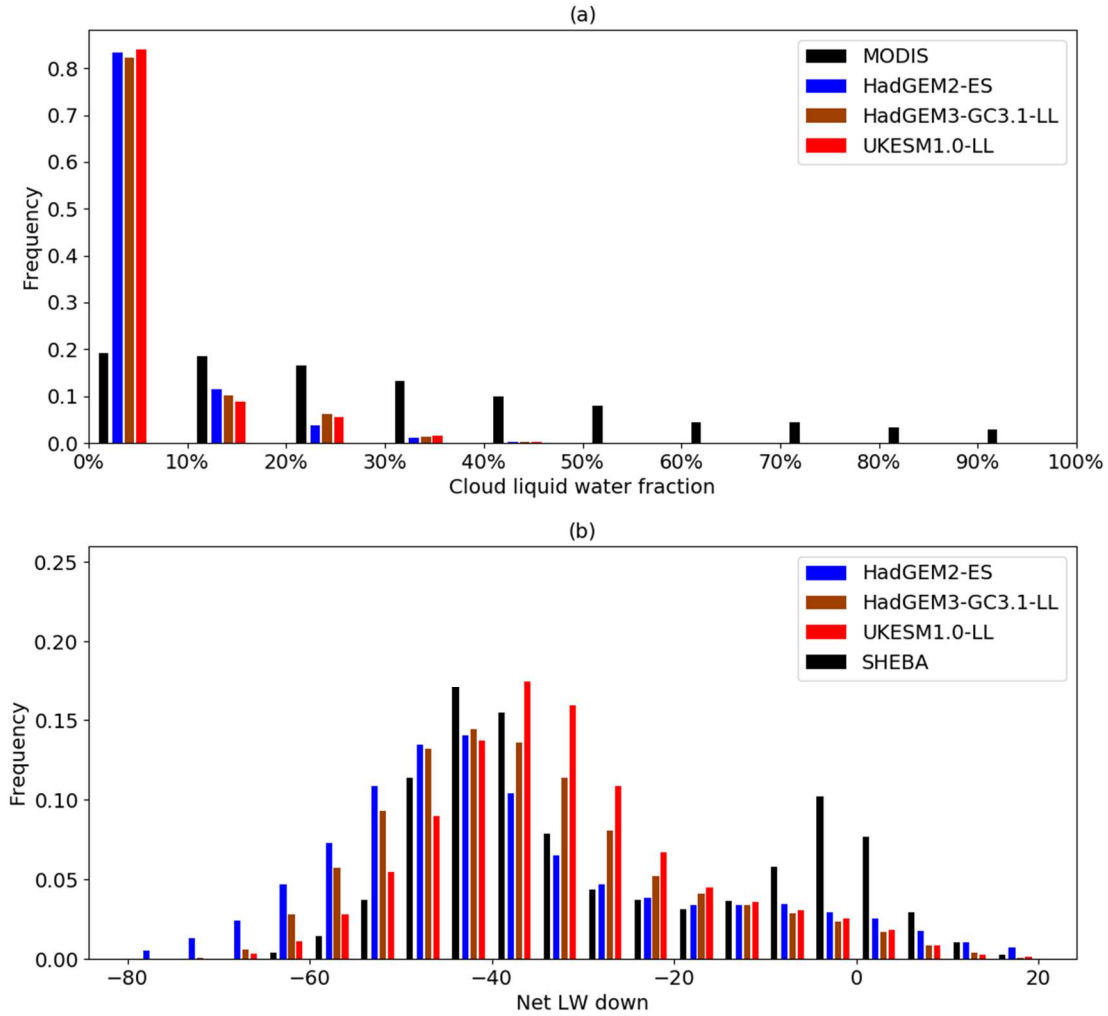


Figure 4.9. (a) Histogram of cloud liquid water fraction over the Arctic Ocean modelled by HadGEM2-ES, HadGEM3-GC3.1-LL and UKESM1.0-LL, and measured by MODIS; (b) Histogram of 3-hourly values of net downwelling LW radiation, measured at the SHEBA campaign and modelled in a similar region by HadGEM2-ES, HadGEM3-GC3.1-LL and UKESM1.0-LL.

The causes of the early melt onset bias of HadGEM2-ES are harder to determine. For most of the spring, comparison of daily upwelling LW fields of HadGEM2-ES to CERES-SYN observations (not shown) shows the Arctic surface to be anomalously cold in the model, as during the winter. During May, however, upwelling LW values rise much more steeply in the model, and surface melt onset commences during mid-to-late May, far earlier than in the satellite observations. A possible cause of the overly rapid surface warming during May is the zero-layer thermodynamics approximation used by HadGEM2-ES, in which the ice heat capacity is ignored.

Comparing fields of surface temperature in HadGEM2-ES between the beginning and the end of May shows a 'missing' ice sensible heat uptake flux of 10-30 Wm^{-2} over much of the central Arctic, which would in turn be associated with a reduction of flux into the upper ice surface of 5-15 Wm^{-2}. Examination of modelled and observed daily time series of downwelling LW and net SW fluxes in late May and early June suggests that a surface flux reduction of this magnitude could delay surface melt by up to 2 weeks, a substantial part of the modelled melt onset bias seen. (In chapter 5 below, a direct evaluation of conductive fluxes shows that the zero-layer approximation is unmistakeably driving another substantial model bias, during the freezing season).

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

Investigating this further, comparison of HadGEM2-ES cloud fraction as a function of optical thickness and cloud-top temperature to observations from ISCCP-D and MODIS shows a strong negative cloud fraction bias focussed in an optically-thick, high temperature part of the cloud distribution (not shown). These properties would produce a comparatively large positive downwelling SW bias, and a comparatively small negative downwelling LW bias from a given cloud fraction deficit, hence would likely produce a positive surface flux bias. This effect may therefore also be important in causing the early onset of surface melting in HadGEM2-ES.

4.4 Applying the ISF method to the difference between two models

In Section 4.3 the ISF method was applied to the 'bias' between a model and observational estimates of the real world. However, the ISF method can also be used to understand the difference between two model simulations directly, without reference to observations. In more detail, the functions $g_{x,t}$ can be used

to calculate surface flux dependence on a particular climate variable in model 1, and multiplied by the difference in that variable in model 1 relative to model 2. This produces an estimate of the surface flux difference in model 1 relative to model 2 induced by the difference in that variable.

Specifically, the three models display very different simulations of ice thickness, with HadGEM2-ES being thinnest and most amplified, UKESM1.0 thickest and least amplified. The different volume balance simulations of these models should be associated with a damped surface flux seasonal cycle in UKESM1.0 relative to HadGEM2-ES, with surface flux less positive in the summer and less negative in the winter in UKESM1.0. These surface flux differences can therefore be decomposed into components induced by the differences in ice thickness, ice area, and other variables in the two models. It is possible to decompose the surface flux even more systematically by assessing other variables, such as snow thickness, snow area and turbulent fluxes, which could not be assessed in Section 4.3 due to a lack of reference datasets.

4.4.1 Comparing model simulations in the summer

The causes of model differences in sea ice simulation during summer are assessed by focussing on the different variables affecting surface albedo in the three models, evaluated above in Section 3.3. Ice area, snow area and melt onset occurrence are the three principal variables whose variation can be examined. However, there are two structural differences between the models that must also be accounted for. Firstly, there is the difference in meltpond simulation: HadGEM3-GC3.1 and UKESM1.0 model meltponds explicitly, while HadGEM2-ES effectively parameterises meltpond fraction to be 0.18 over bare ice, and 0.22 over snow where surface melting is judged to occur. By comparing the HadGEM2-ES values to the actual meltpond fractions modelled in HadGEM3-GC3.1 and UKESM1.0, an estimate of the impact of the meltpond scheme on the surface flux simulation can be obtained. Secondly, there is the difference in snow area parameterisation (Figure 3.6): HadGEM3-GC3.1 and UKESM1.0 simulate lower snow area for a given snow thickness than does HadGEM2-ES. By computing the model difference in parameterised snow area

at each point in model space and time, the ISF method can be used to judge the impact of this difference also.

In more detail, the methods used here are similar to those used in the main ISF evaluation. The net SW radiation is expressed as a function of key variables, namely downwelling SW, ice area, snow area and meltpond area:

$$F_{SW_{net}} = F_{SW_{down}}(1 - \alpha_{sfc}) = (F_{SW\_down}) \sum_i a_i(1 - \alpha_i) \quad (4.12)$$

Here *i* represent the various different surface types present in a grid cell, a_i the area of each surface type, and α_i the albedo of each type. Here, we treat each grid cell in each model as being composed of only four surface types: open water, meltpond, bare ice or snow. For HadGEM3-GC3.1 and UKESM, the albedo of each surface type is a model parameter[1], total ice area is a model diagnostic (and hence open water area can be deduced from this), and meltpond area is also a model diagnostic. For all models, snow area is parameterized from snow thickness using either Equation 3.1 (HadGEM2-ES) or Equation 3.2 (HadGEM3-GC3.1 and UKESM1.0). For HadGEM2-ES, ice area is also a basic model diagnostic, but meltpond area is parameterized, not modeled.

We calculate induced surface flux difference due to downwelling SW, ice area, snow area, melt onset, snow parameterization, and meltpond parameterization. Given the variable of interest v_i, we calculate $\frac{\partial F_{SW_{net}}}{\partial v_i}$ at each point in model space and time using equation (4.12). To evaluate the difference in surface flux induced by the difference in variable v_i between model 1 and model 2 at point $(\boldsymbol{x}, t)$, the surface flux dependency $\frac{\partial F_{SW_{net}}}{\partial v_i}(\boldsymbol{x}, t)$ is averaged between the models, and multiplied by the model difference in $v_i(\boldsymbol{x}, t)$. As with the ISF biases above, the resulting field of induced surface flux difference can be averaged over space and time.

Because most relevant diagnostics are available at daily resolution, it is possible to see the daily evolution of surface flux difference, with dominant terms in the total ISF difference appearing at different points of the melt season (Figure 4.10a). Comparing UKESM1.0 to HadGEM2-ES, the total ISF difference is

[1] In fact the albedo in the near-infrared and visible spectrums is parameterised separately, so some estimate of how SW radiation is apportioned between each is needed. For the purposes of this section it was assumed 40% of SW radiation falls in the NIR and 60% in the visible.

calculated to be negative throughout the summer, consistent with the weaker ice melt and reduced net SW radiation in UKESM1.0 relative to HadGEM2-ES. In April, the total ISF difference is weakly negative, but it rises steeply through May to reach values of over -30 Wm^{-2} in early June, falling very slowly thereafter. Consistent with the evaluation in Chapter 3, ice area difference does not become the dominant contribution to the total ISF difference until the end of June, with significant contributions occurring both from surface melt onset occurrence (peaking in late May and early June, at -15 Wm^{-2}) and from snow area (peaking in mid-to-late June, at -18 Wm^{-2}). The meltpond scheme contributes only a small negative ISF difference, because large negative differences in the Central Arctic are mostly outweighed by large positive differences near the coasts. The difference in snow parameterization, as expected, contributes a positive ISF difference, rising to a maximum of 8 Wm^{-2} in early June. The downwelling SW term contributes a negative difference in May, becoming a positive difference in June, but this is counteracted by the downwelling LW term which is shown for comparison.

Comparing HadGEM3-GC3.1 to HadGEM2-ES (Figure 4.10b) the behaviour is qualitatively similar, but many of the terms are smaller. For example, the melt onset and snow thickness ISF differences show similar sequential peaks, but smaller and shifted slightly earlier. The melt onset falls to a maximum difference of -10 Wm^{-2} at the end of May, while the snow thickness falls to a maximum thickness of -10 Wm^{-2} also in mid-June. By contrast, the ice area, meltpond scheme and snow parameterization terms are very similar in magnitude as when UKESM1.0 is the subject model.

Figure 4.10 demonstrates how the earlier surface melt onset of HadGEM2-ES (relative to the newer models) begins a chain of events, reducing net SW and surface melt earlier in the season, and helping to induce snow area and ice area differences later in the season, further reducing net SW and sea ice melt. The differences are slightly exacerbated by the explicit meltponds of the newer models, which cause meltpond area to be lower in the Central Arctic when surface melting commences. The snow area parameterization difference, meanwhile, acts to reduce the net SW difference.

Comparing the ice volume balance differences of UKESM1.0 and HadGEM2-ES to that implied by the ISF difference, the mean of the total ISF difference over the summer (June-August) is 33.6 Wm^{-2} which is equivalent to 85cm less ice melt in UKESM1.0 than in HadGEM2-ES over this period. This compares to a volume balance difference of 50cm (Section 3.1) and a net radiation difference implying 67cm (Section 3.2): summing ISF differences implies a greater negative volume balance than actually occurs. This is likely due to large numbers of ice-free grid cells in the Arctic Ocean region in both models in late summer, where surface flux difference is not converted into ice volume balance difference, due to the ice having already melted. When HadGEM3-GC3.1 and HadGEM2-ES are compared the situation is similar: volume balance implied by total ISF difference is negative and of the right order of magnitude, but slightly too high.

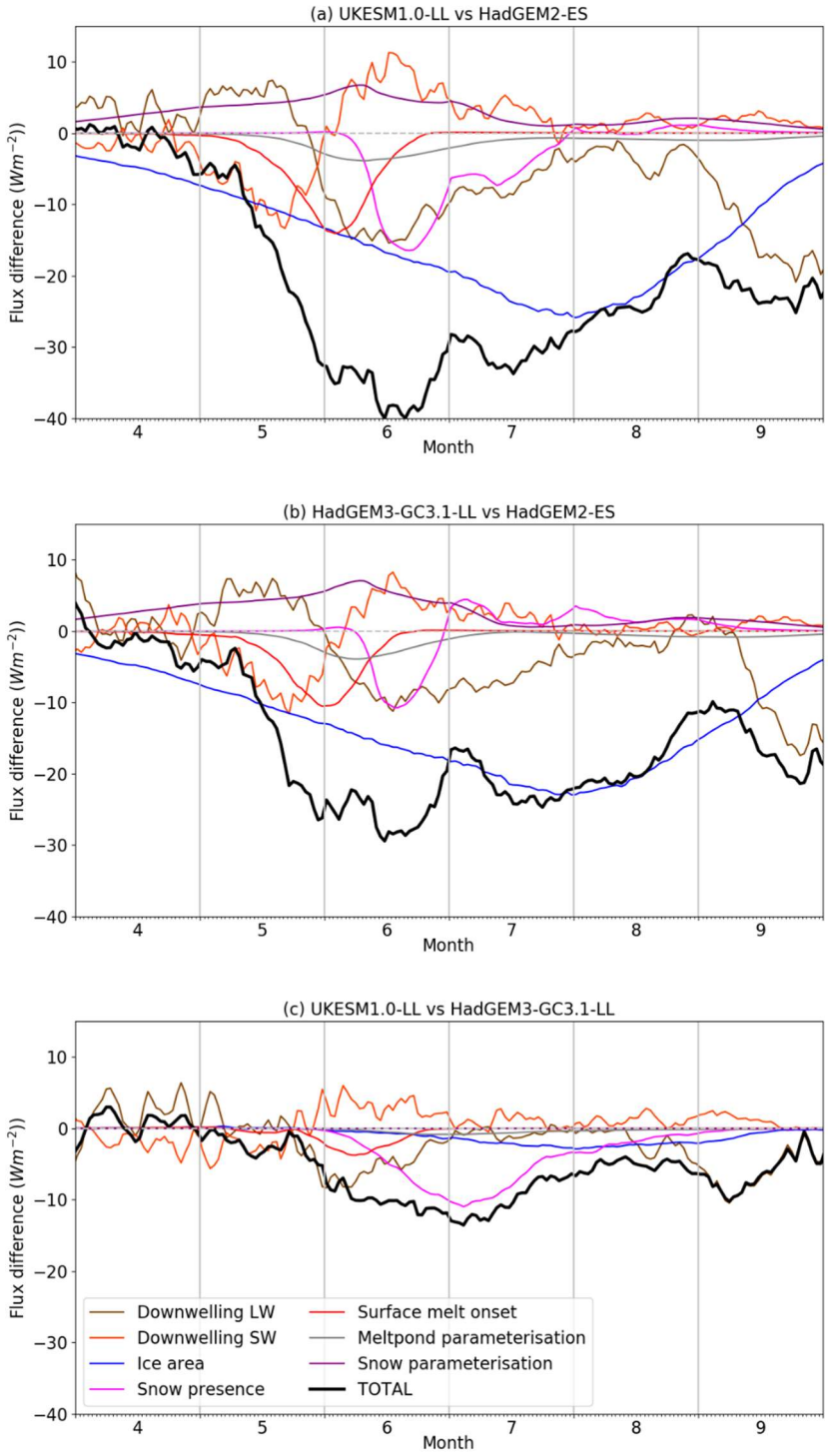


Figure 4.10. Evolution of induced surface flux differences between (a) UKESM1.0 and HadGEM2-ES; (b) HadGEM3-GC3.1 and HadGEM2-ES, apportioned by downwelling SW, ice area, snow thickness, surface melt onset, direct impact of snow parameterization change, and direct impact of the UKESM1.0 meltpond scheme.

We can also use the ISF method to compare UKESM1.0-LL to HadGEM3-GC3.1-LL. Total ISF difference during summer is negative, but less strongly so than

when either model is compared to HadGEM2-ES, with an average of -8.4 Wm^{-2} from June to August (Figure 4.10c). The largest component of this is the ISF difference due to snow thickness differences (-4.7 Wm^{-2}) with the ice area and melt onset terms contributing only -1.7 Wm^{-2} and -0.7 Wm^{-2} respectively. Interpreting this, the surface albedo feedback is likely to be less important in driving the reduced melt of UKESM1.0-LL relative to HadGEM3-GC3.1-LL than with the other model differences, and model biases.

4.4.2 Comparing model simulations in the winter

All three models produce diagnostics of ice thickness and ice area by category. Hence the category-dependent column model described in Section 4.2 can be used to provide a complete decomposition of the surface flux difference between any pair of models. Specifically, equation (4.9) describes how surface flux depends on category ice thickness, category ice area and atmospheric forcing. For any grid cell, this dependence can be multiplied by the model difference in the given model quantity, to produce an estimate of the ISF difference. For HadGEM2-ES, the comparison is made more difficult by the lack of category snow thickness data; for this model, snow thickness is assumed to be uniform across categories.

Comparing first UKESM1.0 to HadGEM2-ES (Figure 4.11a), we see that throughout the winter differences in ice thickness produce a positive difference in surface flux (indicating reduced ice growth). In the early part of the winter, this is caused by differences in the thinnest ice category, but as the winter progresses, differences in the thickest ice category predominate. Differences in atmospheric forcing produce a negligible contribution to differences in surface flux throughout. The comparison is quantitatively consistent with the difference in volume balance across the Arctic Ocean region as a whole evaluated in Chapter 3: across the freezing season October – April, the total ISF difference is 7.9 Wm^{-2}; this is equivalent to reduced ice growth of 47cm over the freezing season. By contrast, in chapter 3 UKESM1.0 was shown to display reduced ice growth of 50cm relative to HadGEM2-ES. The ISF method here demonstrates that this difference is likely due in the main part to differences in ice thickness, i.e. the thickness-growth feedback.

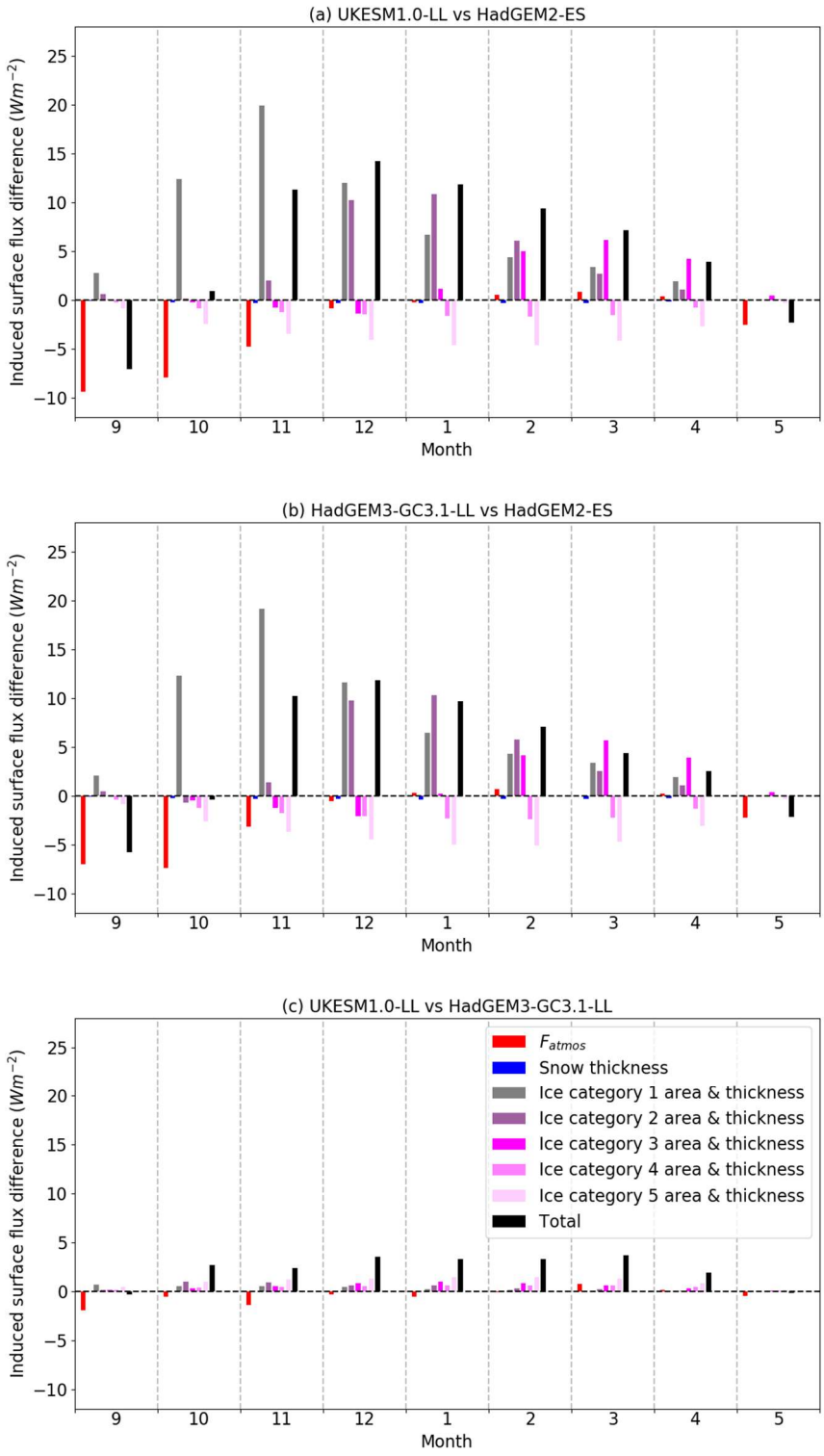


Figure 4.11. ISF freezing season differences due to atmospheric forcing, snow thickness and ice thickness for (a) UKESM1.0-LL relative to HadGEM2-ES; (b) HadGEM3-GC3.1-LL relative to HadGEM2-ES; (c) UKESM1.0-LL relative to HadGEM3-GC3.1-LL.

In a similar way, we can assess HadGEM3-GC3.1 relative to HadGEM2-ES (Figure 4.11b), and see a similar picture: positive ISF differences in winter,

indicating less ice growth in HadGEM3-GC3.1, are entirely attributed to differences in the ice thickness distribution. Differences in snow depth play no significant role, and differences in atmospheric forcing are important only at the very beginning of the freezing season. From October – April, the total ISF difference is 6.0 Wm-2, which equates to 36cm less ice growth. This is greater than the 31cm less ice growth diagnosed from the ice thickness evolution.

Finally we can compare UKESM1.0 to HadGEM3-GC3.1. The method produces a positive ISF difference throughout the winter (Figure 4.11c), consistent again with the weaker ice growth in UKESM1.0, and just as in the other two comparisons the ISF difference is dominated by the differences in the ice thickness distribution (not shown). A total ISF difference of 3.4 Wm^{-2} is estimated during the freezing season for UKESM1.0 relative to HadGEM3-GC3.1. Over a period of 6 months this is associated with reduced ice growth of 15-20cm, and the difference in Arctic Ocean average ice freezing between the two models seen in Chapter 3 was around 17cm.

A common theme in all three comparisons is that differences in atmospheric forcing are important during the autumn, but only in the sense that they oppose the total ISF difference. For example, comparing UKESM1.0 to HadGEM2-ES, atmospheric forcing contributes a total ISF difference of -9.1 Wm^{-2}, nearly negating the category 1 term (11.6 Wm^{-2}); a colder atmosphere in UKESM1.0 reduces ice growth during the early part of the melt season. This is consistent with the evaluation of downwelling LW in Chapter 3: model differences in this variable are mainly confined to the autumn, with the models being very similar for the rest of the freezing season.

4.5 Summary

Using simple models to approximate the response of the sea ice to surface forcing, we can estimate the surface flux bias, and hence bias in ice growth / melt, associated with model bias in individual variables. For all three models, downwelling LW biases contribute a negative ISF bias in the winter: in HadGEM2-ES, bias in ice area and thickness acts to magnify this effect, causing a bias towards too much ice growth, but in HadGEM3-GC3.1 and UKESM1.0, biases in ice area and thickness counteract this, causing a bias

towards too little ice growth. Conversely, during the summer, the models exhibit differing ISF biases due to melt onset during the summer, HadGEM2-ES the largest and UKESM1.0 the least. HadGEM2-ES also displays a significant positive ISF bias due to ice area, causing a bias towards too much ice melt, but HadGEM3-GC3.1 and UKESM1.0 display indeterminate (likely negative) biases due to ice area.

In still more detail, it is possible to estimate the surface flux and volume balance difference associated with model differences, not just due to individual variables, but in some cases due to structural model differences. We see that in summer, the differences in surface flux, and hence ice melting rate, between the newer models and HadGEM2-ES arise firstly because of the difference in melt onset date. The differences are subsequently exacerbated by the explicit meltpond scheme, the differences in snow thickness, and lastly the growing differences in ice area as the melt season progresses. In winter, the differences in ice growth rate between the models are driven mainly by differences in the ice thickness distribution, with differences in atmospheric forcing playing very little role.

The approach of the ISF framework is to make better use of existing observational data: to combine evaluation of ice state and surface radiation variables with simple models to analyse the causes of biases and differences in ice volume balance. In the next chapter, we try a different approach: to evaluate the internal ice thermodynamics directly, using a dataset computed for this thesis from the Arctic network of Ice Mass Balance buoys (IMBs).

5. Ice mass balance buoys: a way of evaluating sea ice thermodynamic processes

In this chapter, we present the second of two enhancements to sea ice model evaluation: direct evaluation of internal ice processes using the network of Arctic Ice Mass Balance buoys (IMBs). Data from the IMBs is used to compute a dataset of internal ice energy fluxes, with which sea ice models can be evaluated over particular, densely-sampled regions of the Arctic. Conceptually, the approach is to evaluate directly the 'missing processes' of Figure 1.4, as opposed to the ISF approach which inferred these using simple models. By this approach, some sea ice processes can be examined that were invisible to the ISF analysis.

In turn, we discuss the motivation for using the IMB data, and describe how these are used to produce a dataset of fluxes. We use this dataset to evaluate the models, and show how the results both support and enhance the conclusions of Chapters 3 and 4.

This chapter is based upon West et al (2020), in which the IMB dataset was calculated and used to evaluate HadGEM2-ES. As lead author of this study I was responsible for the production of the IMB dataset, the analysis of spatial, seasonal and interannual variability of the dataset, the sensitivity to parameters of the analysis and the evaluation of the effect of the ice thickness sampling bias. I was also responsible for the evaluation of internal ice fluxes in HadGEM2-ES using the IMB dataset, for the production of all figures and for the writing up of the paper in its final form.

5.1 Overview

The ISF framework enables quantification of the contribution of differences (biases) in model processes to differences (biases) in sea ice growth / melt. However, its uses in evaluating model bias are restricted by the availability of observational datasets, limited to surface radiation, sea ice area and thickness, and surface melt onset. This places strict limits on the complexity of the processes it is possible to evaluate with the ISF framework. For example, it is not possible to evaluate the effect of heat storage within the ice on sea ice

growth / melt. Nor is it possible to evaluate the impact of volume changes at the top and basal surface of the ice separately. Such an evaluation could both enhance the understanding of sea ice simulation provided by the ISF framework, and could also provide the basis for a more detailed analysis of the causes of model bias in volume balance. In this chapter, we describe the direct evaluation, using observational estimates derived from ice mass balance buoy measurements, of modelled ice thermodynamics and mass balance at the top and basal surfaces.

The ice mass balance buoy (Richter-Menge et al., 2006; Perovich and Richter-Menge, 2006; Figure 5.1) 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 in latitude-longitude space. Many also measure surface air pressure and temperature. 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). Hence an observational dataset of sea ice vertical energy fluxes can be created by estimating monthly means for the entire IMB network.

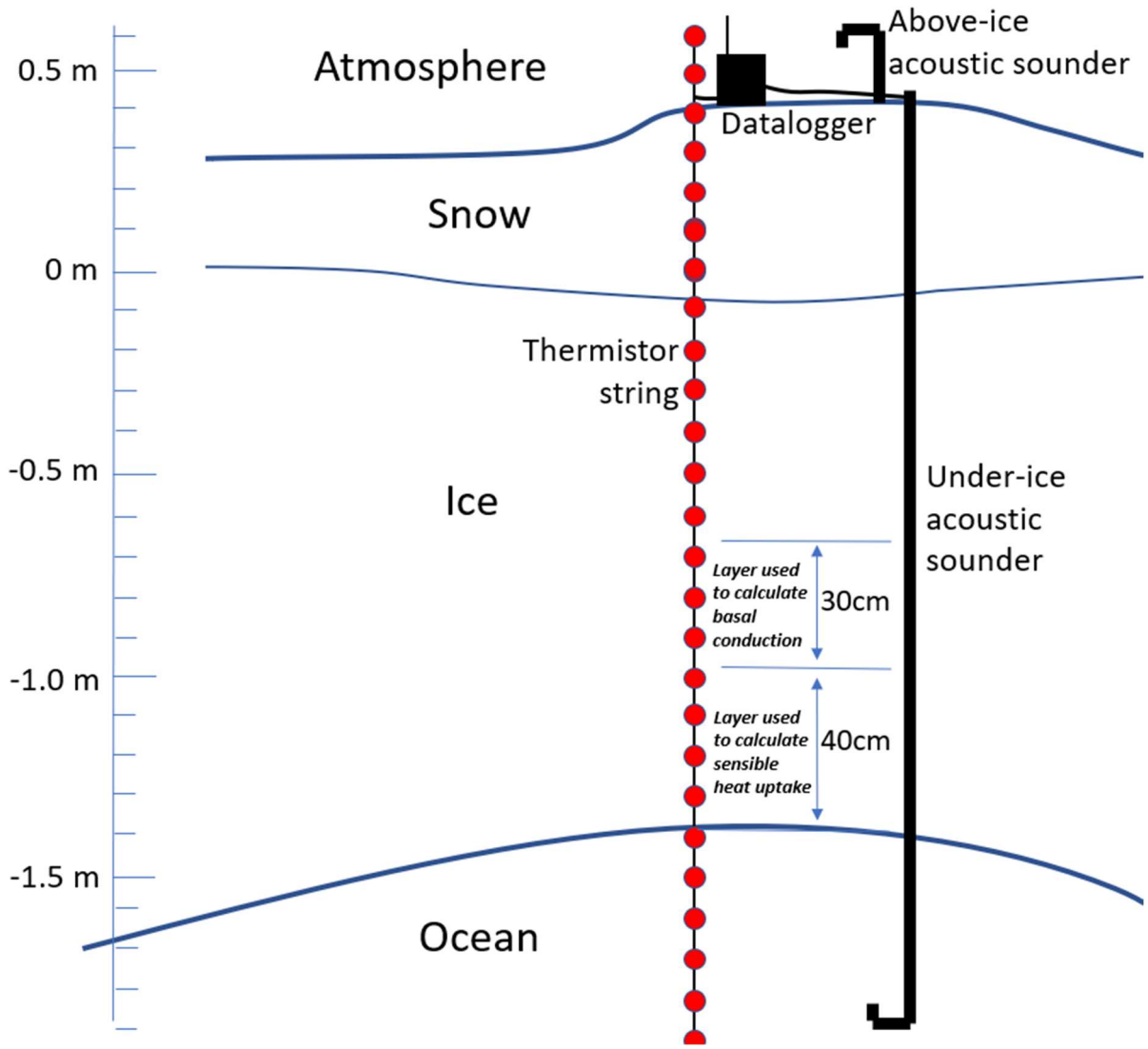


Figure 5.1. Diagram of an Ice Mass Balance buoy (IMB), with reference layers used for flux calculation below indicated. Adapted from Figure 1a of Planck et al. (2019).

The effect of evaluating modelled vertical energy fluxes with this dataset would be to fill in many of the 'missing' processes in Figure 1.4 (processes that could not previously be evaluated). However, although over 100 IMBs have been released in the Arctic since 1993, coverage is too sparse in most areas to allow a full spatial climatology to be computed from the dataset, meaning that standard methods of model evaluation cannot be used. We adopt a different approach in this chapter, based upon observing that the North Pole and Beaufort Sea regions are quite well-sampled (Figure 5.2). We restrict model data from HadGEM2-ES, HadGEM3-GC3.1 and UKESM1.0 to these regions: distributions of modelled fluxes are then compared to distributions of fluxes computed from the IMB network.

The approach of comparing data from two-dimensional model regions to data from the IMB tracks, rather than sampling model data from the grid cells directly overlying the tracks, is chosen for reasons of internal and spatial variability. The major part of spatial variability in Arctic sea ice conditions occurs on scales both smaller, and larger, than the scales relevant (~10-100km) when deciding whether to focus on individual grid cells within regions. Most spatial variability occurs on much smaller scales (<10km) that cannot be resolved even by focussing on individual grid cells, while larger-scale spatial variability (>100km) is accounted for precisely by restricting model data to the chosen regions. It is also the case that the internal variability in sea ice simulation renders a too-precise approach pointless. For example, it would not be reasonable to expect a modelled parcel of ice forming in the Beaufort Sea in October 1997 at the location of buoy 1997D (for example) to remain coincident with that IMB over the year-long course of its operation, or for the atmospheric conditions over that track to reproduce those in the real world more accurately than those 30km distant from the track.

For consistency with previous chapters, we continue to evaluate the 1980-1999 period, although the IMB data used represents the period 1993-2015. The effect of this temporal offset was assessed in West et al. (2020) and was found to be small relative to the model biases identified, the largest differences being in top melting and basal conductive flux, discussed as part of the model evaluation in Section 5.4 below.

In this chapter, we choose to evaluate fluxes of top melting, top conduction, basal conduction and ocean heat, fluxes fundamental to the sea ice evolution reported by all three models. In addition, simultaneous measurement of snow depth and ice thickness from the IMBs allows the mechanisms driving ice conduction, and hence ice growth, during the winter, in the models and IMBs to be separated. At the end of this chapter, 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.

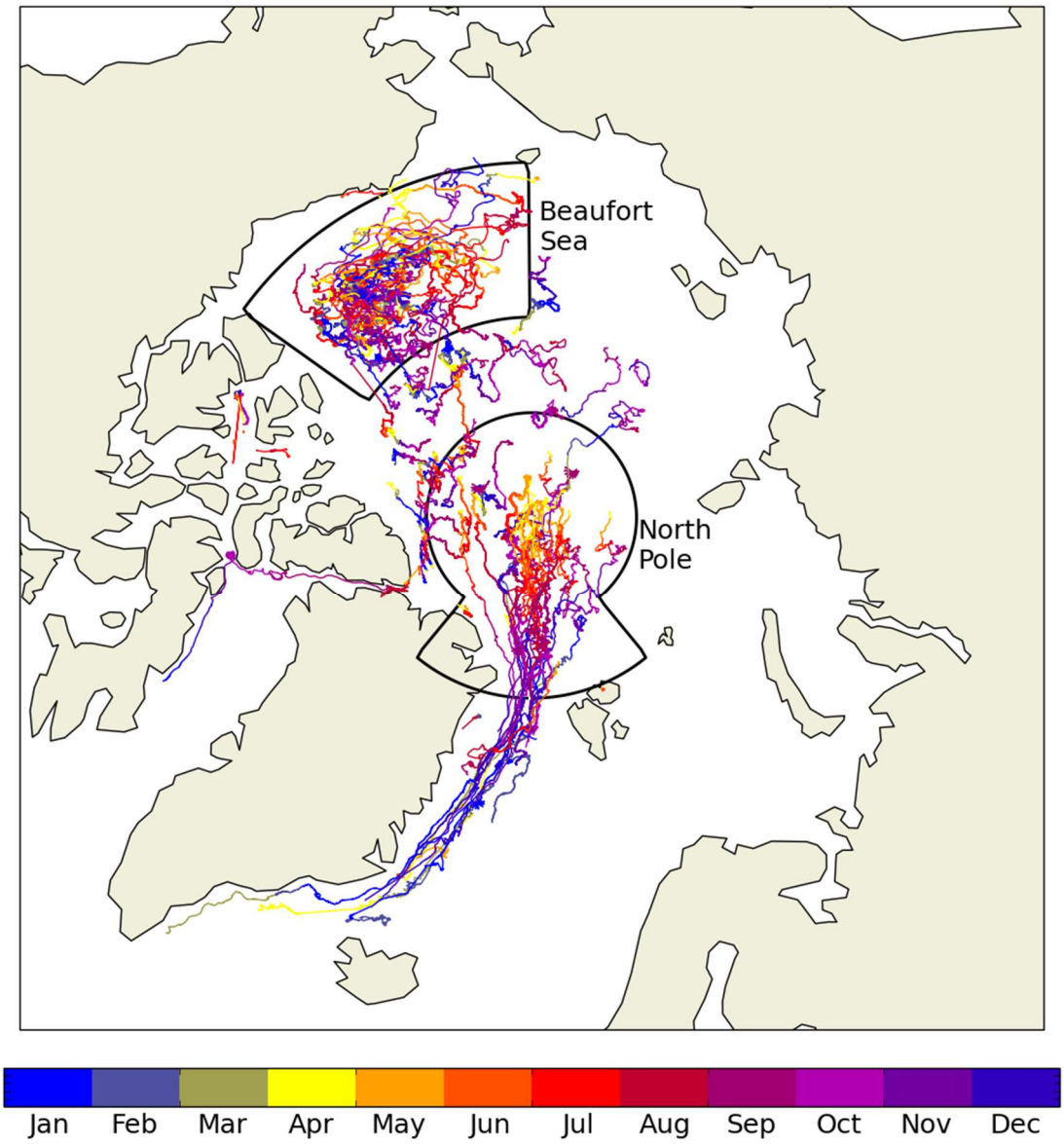


Figure 5.2. The tracks of Arctic ice mass balance buoys from 1993 to 2015, with months of coverage indicated by the coloured shading. The North Pole and Beaufort Sea regions used in the analysis are shown by the thin black lines.

5.2 Estimating vertical energy fluxes from the IMB measurements

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-crrel-dartmouth.org/results/ (Perovich et al., 2020). The buoys were deployed between 1993-2017; spatial coverage is mainly in the North Pole

and Beaufort Sea regions (Figure 5.2). 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. 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 5.3. Although a more advanced optimal interpolation scheme would likely produce more accurate time series, inspection of individual data series shows that the current scheme produces 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.

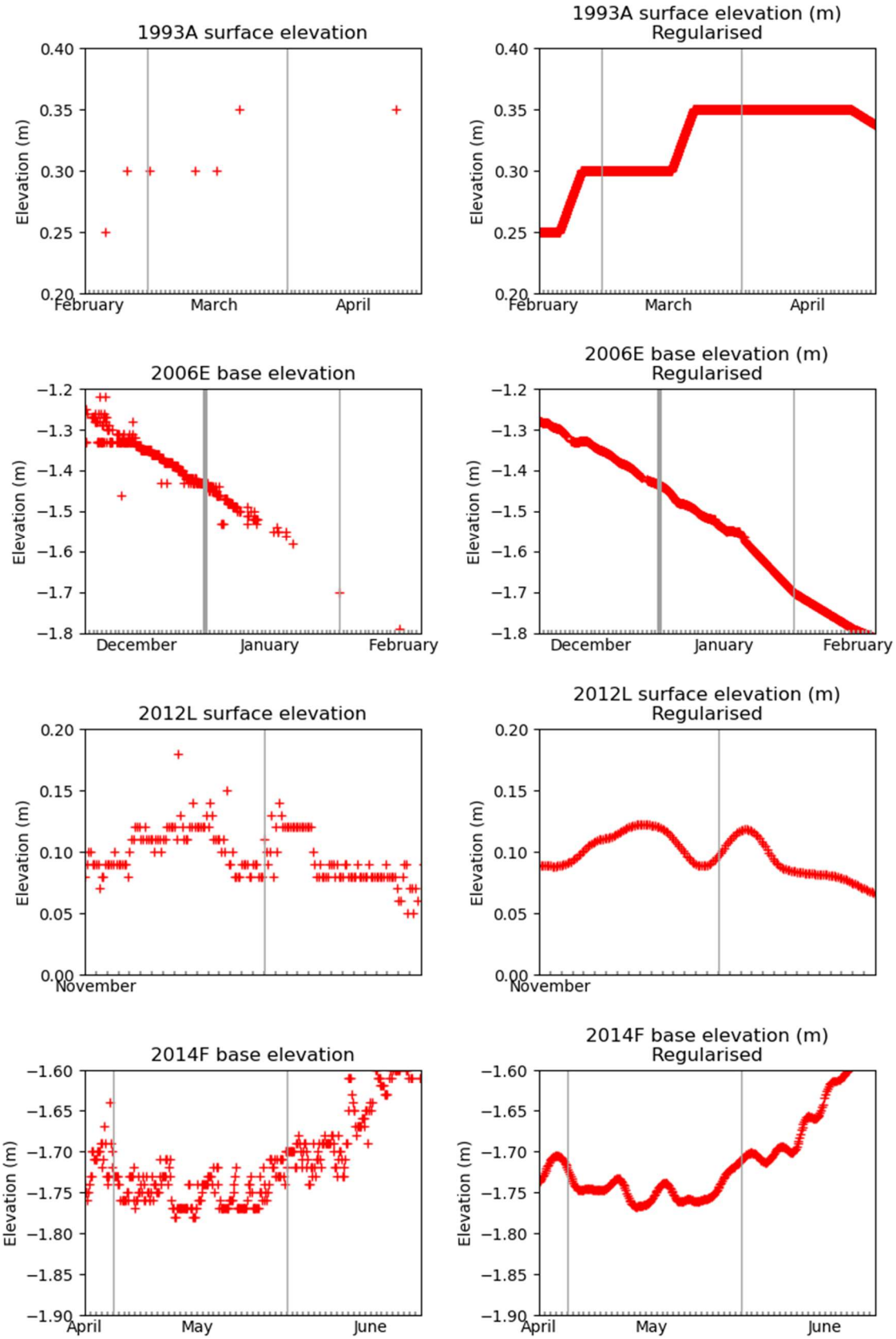


Figure 5.3. Illustration of the regularization process using four selected IMB data series. (a, b, c, d) Raw data; (e, f, g, h) time series regularized to temperature measurement points.

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 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 5.4. 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.

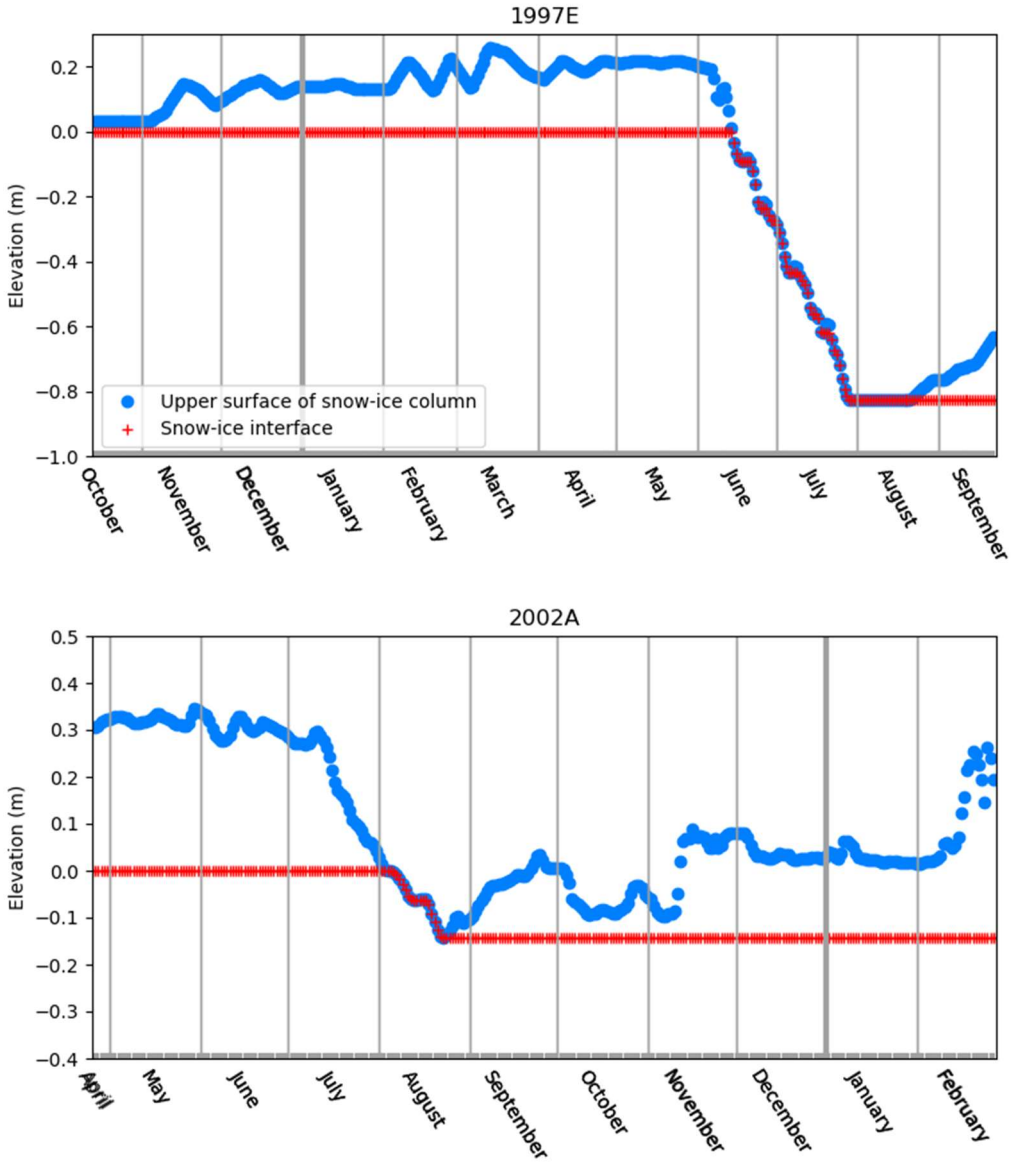


Figure 5.4. Two examples of the process of estimating snow–ice interface from a regularized snow surface data series. The interface remains at a constant level unless the surface falls below this level, in which case the interface falls with the surface.

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.

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. 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 use of a threshold value is necessary because surface elevation often decreases during the winter months, probably due to snow drifting, under conditions in which melting would be unrealistic. -2C is chosen as this threshold as it allows for error in temperature measurement, and in surface elevation estimation.

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). 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 is determined by a linear fit through the same layer, using an 'adjusted' temperature profile:

$$T_{adj}(z) = \begin{cases} \mu T(z) + (1-\mu)T_{int-ref} & z > z_{int} \\ T(z) & z \leq z_{int} \end{cases} \quad (5.1)$$

where z_{int} is the elevation of the snow-ice interface, $T_{int-ref}$ is temperature 5cm below the interface, and $\mu = k_{ice}/k_{snow}$ where k_{ice} and k_{snow} are ice and snow conductivity respectively. 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 is 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 are then averaged to obtain monthly means. The process is illustrated in Figure 5.5.

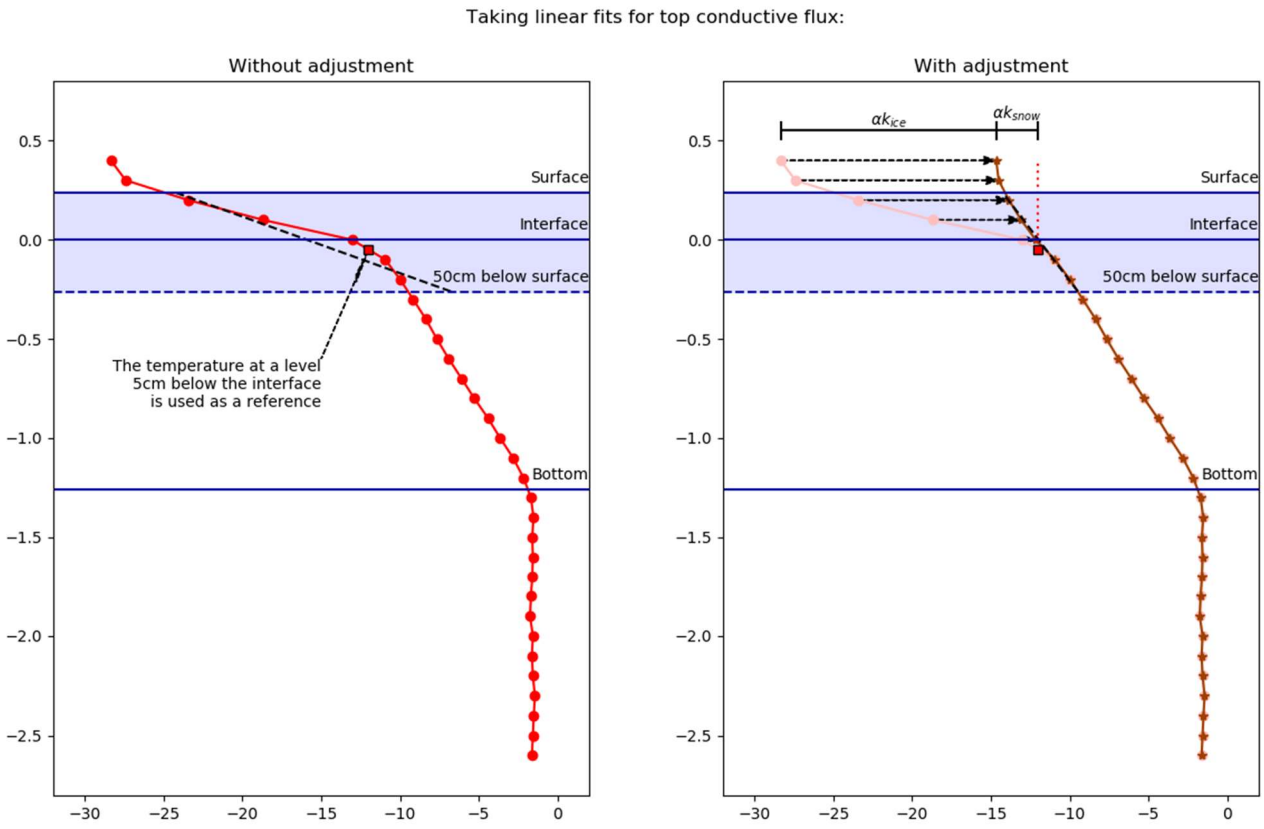


Figure 5.5. Illustration of the process of estimating conductive flux across the top 50 cm of the snow–ice column, in the case that the snow– ice interface lies within this layer. Panel (a) shows the raw temperature profile; taking a linear fit through these points does not produce a meaningful result because of the sharp "corner" associated with the change in medium. Panel (b) 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. In panel (b), k_{ice} denotes ice conductivity, k_{snow} snow conductivity and α an arbitrary constant of proportionality.

Basal conductive flux. This flux is defined as the conduction into the ice base from the ice interior. As an important 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 5.1. In section 5.3.2 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 5.1. The ocean heat flux can then be written as

$$F_{ocn} = F_{condbot} - F_{sens} - F_{lat} \qquad (5.2)$$

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 5.3.1 below, and subsequently evaluate sensitivity to the values of these parameters in section 5.3.2. Ice density ρ_{ice}, snow density ρ_{snow} and latent heat of melting q_{fus} are set to 917 kgm^{-2}, 330 kgm^{-2} and 3.34 x 105 Jkg^{-1} respectively, the standard values used by the sea ice model CICE (Hunke et al., 2015).

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

$$k_{ice} = k_{fresh} + \frac{\beta S}{T} \qquad (5.3)$$

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

$$c_{ice} = c_{fresh} + \frac{q_{fresh} \mu S}{T^2} \qquad (5.4)$$

where $c_{fresh} = 2106 J kg^{-1} K^{-1}$is the specific heat capacity of fresh ice, $q_{fresh} = 3.34 \times 10^5 J kg^{-1}$ the specific latent heat of fusion of fresh ice, and $\mu = 0.054 K$ 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

$$q(S,T) = \rho c_0 (T_m - T) + \rho q_{fresh} \left(1 + \frac{\mu S}{T}\right). \qquad (5.5)$$

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 5.3.2 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. 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.

Examination of the monthly mean energy fluxes reveals several ways in which unrealistic estimates might be produced. Firstly, in a small minority of months (about 3%) 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 is set instead to the highest physically allowable value, given the maximum temperature attained.

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, Figure 5.6), 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.

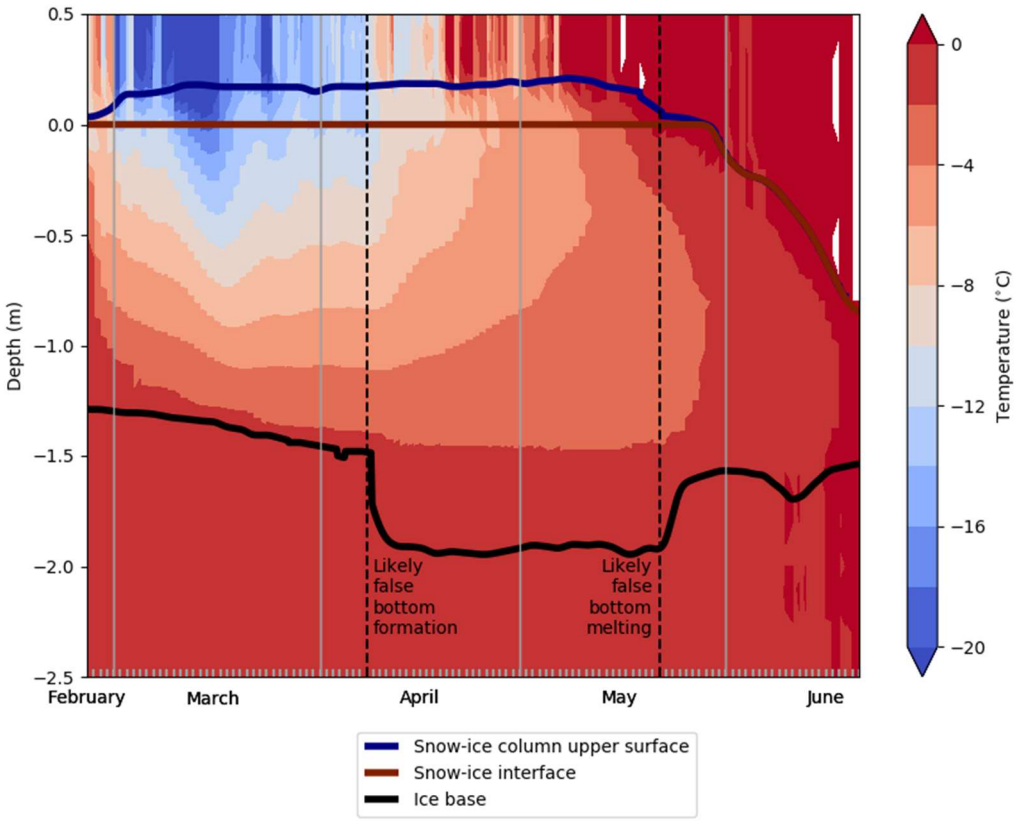


Figure 5.6. An example of likely false bottom formation in the IMB record (buoy 2015A).

A third problem relates to the formation of melt ponds at IMB locations. An IMB 'sees' a melt pond surface as the top of the snow-ice column, meaning that melt pond formation around an IMB would cause an unphysical slowing of melting to be diagnosed. However, this is unlikely to be a widespread problem because IMBs tend to be located in regions less likely to see melt pond formation (e.g. ridges or hummocks), and because melt pond formation is often followed by complete melting out of the floe in which the IMB is located (Polashenski et al., 2011). Visual inspection identified three IMBs in which melt pond formation may have occurred without imminent IMB loss (1993A, 2004A and 2013B), diagnosed by abrupt slowing of rate of surface elevation fall, appearance of temperatures above 0°C immediately below the surface, and subsequent delayed cooling of the ice interior in autumn. In all cases the melt pond presence is only possible between about mid-July and mid-August, limiting impact on the overall IMB dataset.

5.3 Description of fluxes estimated from the IMBs

5.3.1 Seasonal and spatial variability

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 5.1. The IMBs provide 463 monthly mean values of top melt in total, ranging from 31 values in March and August to 53 in May. The seasonal cycle reaches its maximum in July, when top melting of 29.9 ± 17.8 Wm^{-2} is observed. Strong top melting is also evident in June (16.8 ± 11.0 Wm^{-2}), but top melting tends to be considerably lower in August (8.1 ± 6.7 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 interannual variability.

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	-16.2	6.1	-14.0	5.7	1.4	5.0
February	0.0	0.0	-16.9	6.9	-13.7	6.7	0.6	4.2
March	0.0	0.0	-13.5	5.1	-12.7	4.6	1.5	5.6
April	0.0	0.0	-7.5	3.1	-9.7	3.3	2.3	2.7
May	1.1	3.2	-0.5	2.3	-6.2	2.3	3.4	4.0
June	16.8	11.0	3.8	1.8	-2.2	1.6	12.3	16.5

July	29.9	17.8	1.0	1.0	0.5	1.2	18.1	15.3
August	8.1	6.7	-1.1	3.5	1.0	1.1	19.2	23.9
September	0.6	1.2	-6.3	4.5	0.7	1.9	9.4	11.4
October	0.0	0.0	-14.4	8.9	-4.0	11.4	5.4	13.0
November	0.0	0.0	-17.3	7.0	-9.2	9.9	4.6	7.1
December	0.0	0.0	-17.6	6.8	-12.5	6.6	1.3	5.2

Table 5.1. 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.

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 estimates in total, ranging from 24 in August to 51 in May. Mean top conductive fluxes are strongly negative from October-March, reaching a minimum value of -17.6 ± 6.8 Wm^{-2} in 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 463 estimates, ranging from 29 in August to 52 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 ± 5.7 Wm^{-2} occurring in January. 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 estimates, ranging from 25 in August to 49 in May. The highest values are seen in July and August, with a mean and spread of 18.1 ± 15.3 Wm^{-2}, and 19.2 ± 23.9 Wm^{-2} respectively. The distributions in these months are, like the top melting flux, strongly positively skewed, with a small number of exceptionally high values. Notably, 119 Wm^{-2} is 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. For most such values, the uncertainty interval resulting from varying the salinity from 0 to 10 encompasses 0 Wm^{-2}.

Two regions of the Arctic are relatively densely sampled by the IMBs: the Beaufort Sea and the North Pole (Figure 5.2). 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 Figure 5.7. Significance of differences between distributions is measured using a Welch t-test, with a 5% p-value threshold.

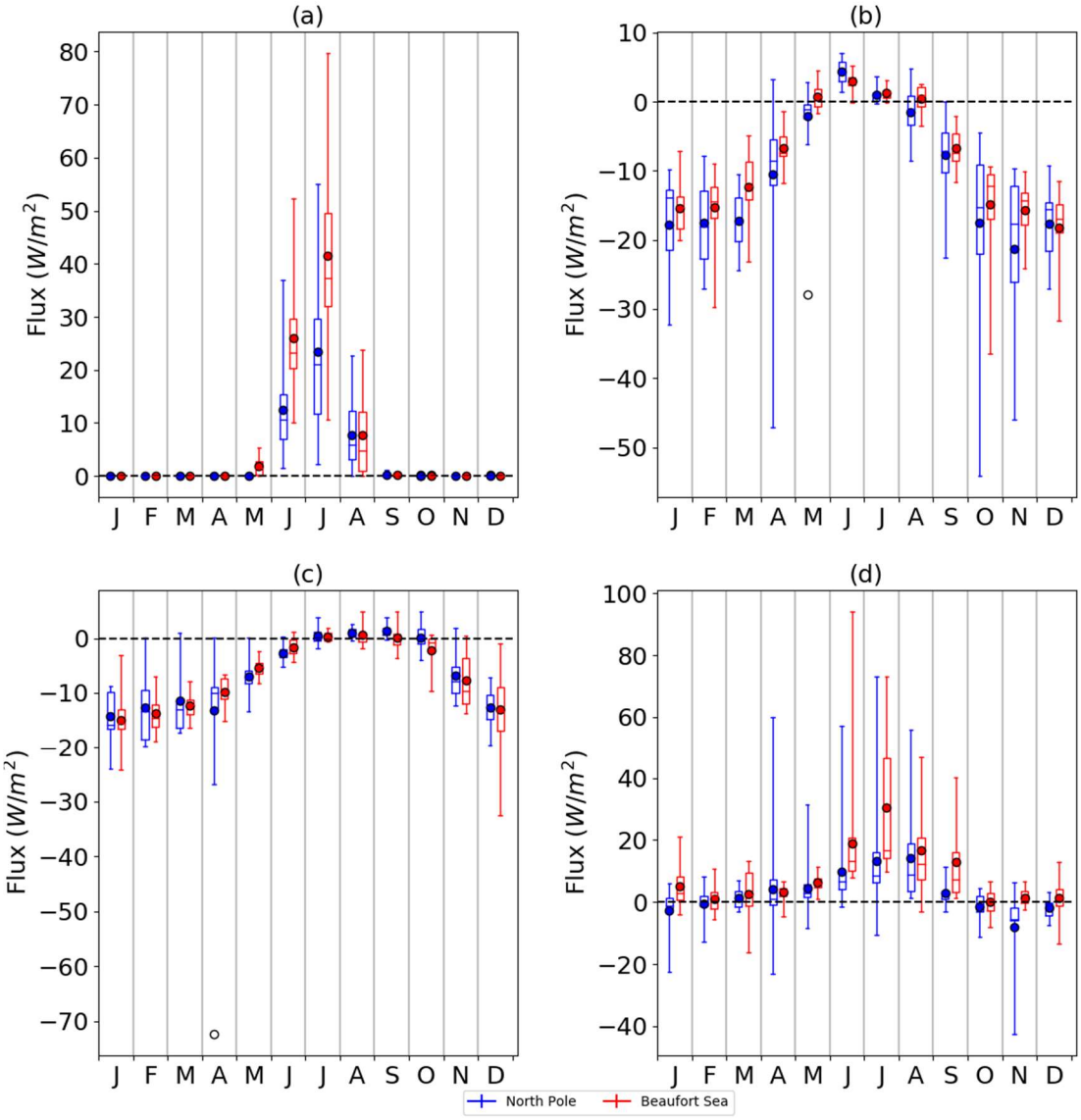


Figure 5.7. Distributions of (a) top melting fluxes, (b) top conductive fluxes, (c) basal conductive fluxes, (d) ocean heat fluxes estimated for the IMB dataset in the North Pole and Beaufort Sea regions, shown as boxplots indicating range, median and quartiles.

Top melting fluxes are shown in Figure 5.7a 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 Wm^{-2}, with a mean of 12 ± 8 Wm^{-2}, while those measured in the Beaufort Sea range from 10 to 52 Wm^{-2} with a mean of 26 ± 10 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 to 55 Wm^{-2} in the North Pole region, with a mean of 23 ± 14 Wm^{-2}, and 11 to 80 Wm^{-2} in the Beaufort Sea region, with a mean of 41 ± 17 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 5.7b), 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. In January, for example, North Pole fluxes range from -32 to -10 Wm^{-2} with a mean of -18 ± 7 Wm^{-2}, while those in the Beaufort Sea region range from -20 to -7 Wm^{-2} with a mean of -15 ± 7 Wm^{-2}. Some notable differences between the distributions occur in the 'shoulder seasons', particularly in May and August (when the distributions are significantly different), 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 to 3 Wm^{-2} with a mean of -1 ± 2 Wm^{-2}, while values in the Beaufort Sea region range from -2 to 4 Wm^{-2} with a mean of 1 ± 2 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 5.7c). 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. 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 -2 Wm^{-2} and a mean of -5 ± 1 Wm^{-2} in the Beaufort Sea region.

For the ocean heat flux (Figure 5.7d), 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 to 38 Wm^{-2} with a mean of 13 ± 10 Wm^{-2}, while the Beaufort Sea region values range from 7 to 119 Wm^{-2} with a mean of 33 ± 35 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.

5.3.2 Uncertainty from 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. (2009), 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 lower 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}.

The relationship between ice conductivity, and ice temperature and salinity, 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 (in $Wm^{-1}°K^{-1}$) is calculated from ice temperature T (in °C) and practical salinity S as

$$k_I = 2.11 - 0.011T + 0.09\frac{S}{T} \qquad (5.6)$$

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.

5.4 Evaluating modeled fluxes with the IMB estimates

We now use the IMB vertical flux distributions to evaluate the three models: HadGEM2-ES, HadGEM3-GC3.1-LL and UKESM1.0-LL. In the following discussion, modelled flux distributions are summarised by area-weighted mean and standard deviation over model grid cells within each respective region. Differences between modelled and observed flux distributions are tested for significance using a two-tailed Welch t-test with *p*-value threshold of 0.05, as in West et al. (2020); whenever distributions are significantly different for a particular month, region and variable, a model bias is demonstrated.

For top melting fluxes, all models are high relative to the IMB estimates in both regions in summer (Figure 5.8a,b), with HadGEM2-ES displaying the greatest bias and UKESM1.0-LL the least. For example, in the North Pole region in July, summer top melting fluxes of 57 ± 44 Wm^{-2} in HadGEM2-ES, 48 ± 12 Wm^{-2} in HadGEM3-GC3.1-LL and 36 ± 13 Wm^{-2} in UKESM1.0-LL compare to 23 ± 14 Wm^{-} from the IMB estimates. HadGEM3-GC3.1-LL and UKESM1.0-LL, like the IMB estimates, reach their maximum top melt flux in July, whereas in HadGEM2-ES the distributions in June and July are similar. All models are biased high with respect to the IMB estimates in all months of the summer with the exception of UKESM1.0-LL in June and August (in these cases the distributions are not significantly different).

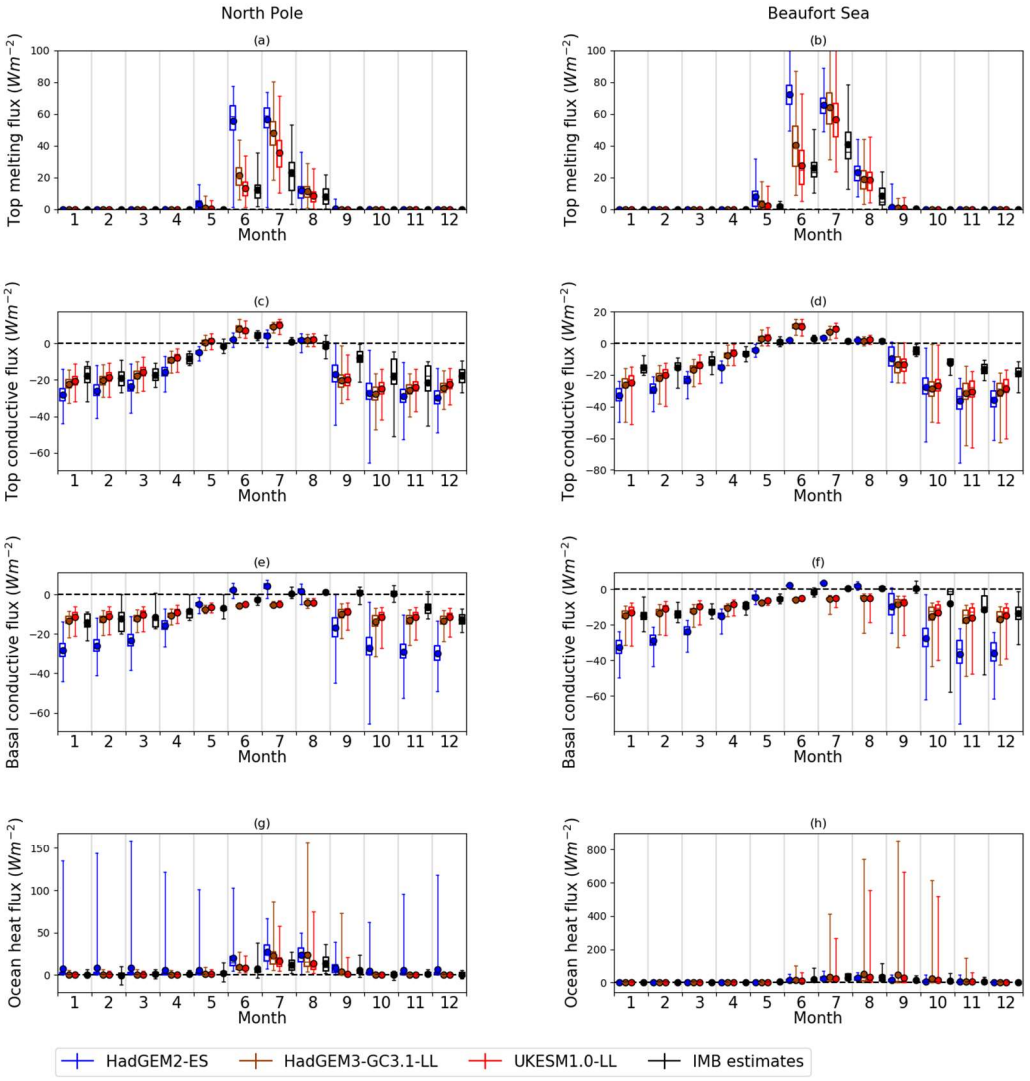


Figure 5.8. Distributions of fluxes of (a,b) top melt; (c,d) top conductive flux; (e,f) basal conductive flux; (g,h) ocean heat flux in HadGEM2-ES, HadGEM3-GC3.1- and UKESM1.0-LL and in the IMB estimates for the regions of the North Pole (left column) and Beaufort Sea (right column). Distributions are represented as boxplots showing maximum, minimum, median and quartiles.

In the Beaufort Sea region, July top melting fluxes are much more similar in the three models – 66 ± 14 Wm^{-2}, 64 ± 15 $Wm^{-2}$and 57 ± 16 Wm^{-2} in HadGEM2-ES, HadGEM3-GC3.1-LL and UKESM1.0-LL respectively –comparing to 41 ± 17 Wm^{-2} from the IMB estimates. However, unlike the other two models HadGEM2-ES actually attains its maximum in June (72 ± 14 Wm^{-2}), being much higher in this month than the other two models. All models are biased high

relative to the IMB estimates in all summer months, except for UKESM1.0-LL in June.

The differences in top melting between the models are consistent with the onset of surface melting being earlier in HadGEM2-ES than in HadGEM3-GC3.1-LL, which is in turn earlier than UKESM1.0-LL. They are also consistent with the more amplified seasonal cycle of ice thickness in HadGEM2-ES. However, it is noteworthy that UKESM1.0-LL top melting is also biased high with respect to the IMB estimates in some cases, despite displaying a damped seasonal cycle relative to PIOMAS. This discrepancy may partly be explained by sampling inaccuracy in the IMB estimates (e.g. because the IMBs undersample melt ponds, as discussed in Section 5.2 above). However it is countered by a negative bias in basal melt, with UKESM1.0-LL displaying a median basal melt around 20 Wm-2 lower than in the IMB dataset in the Beaufort Sea (not shown). Considering the effect of the temporal offset of the IMBs relative to the models actually exacerbates the model bias: is modelled July top melting fluxes are 5-10 Wm^{-2} higher in the period 2000-2015 than for 1980-1999, hence the model biases are likely understated.

Examining the top conductive flux, in the North Pole region all three models' estimates are lower than the IMB estimates from September – January (Figure 5.8c), indicating a tendency for too much heat to be lost from the top surface of the ice. This difference may be partly due to sampling in the early melt season, when the highest top conductive fluxes tend to occur over thin, newly forming ice that is less commonly sampled by the IMBs; the relationship between conductive flux and ice thickness is examined in more detail in Section 5.5 below. As the melt season progresses, the models diverge, with HadGEM2-ES remaining much lower than to the IMB estimates, while HadGEM3-GC3.1-LL and UKESM1.0-LL become closer to the observations. For example, in February, the HadGEM2-ES top conductive flux of -26 ± 5 Wm^{-2} is lower than that of HadGEM3-GC3.1-LL (-20 ± 4 Wm^{-2}) or UKESM1.0-LL (-19 ± 4 Wm^{-2}), and is biased low with respect to the IMBs (-19 ± 6 Wm^{-2}). For HadGEM2-ES, top conductive flux is biased low with respect to the IMBs in all freezing season months (October – April) except November, whereas HadGEM3-GC3.1-LL is biased low only in October and December, and for UKESM1.0-LL in no months are the distributions significantly different. This stands in contrast with the

downwelling LW as evaluated in Chapter 3, which is equally biased low in all three models.

In the Beaufort Sea, the picture is qualitatively similar, but flux estimates from all models are biased low relative to the IMB estimates for all months of the freezing season except April (HadGEM3-GC3.1-LL) and March and April (UKESM1.0-LL).. For example, in February top conductive flux estimates of -29 ± 5 Wm^{-2}, -22 ± 5 Wm^{-2} and -20 ± 5 Wm^{-2} from the three models respectively are all biased low relative to the IMB estimates of -16 ± 5 Wm^{-2}.

In July in both regions, the two newer models are biased very high relative to the IMB estimates, indicating a greater flow of heat into the ice interior from the surface. For example, in the North Pole region 9.2 ± 1.2 Wm^{-2} in HadGEM3-GC3.1-LL and 10.2 ± 1.8 Wm^{-2} in UKESM1.0-LL comparing to 0.9 ± 0.9 Wm^{-2} in the IMB estimates. This may be partly associated with the apparent bias towards thicker ice in these models, as thicker ice will take longer to warm up in the spring and summer.

Turning to the basal conductive flux, we note again that in HadGEM2-ES this flux is equal to the top conductive flux due to the zero-layer approximation. In the other two models, and in the IMBs, the seasonal cycle of the basal conductive flux is damped, and phase shifted later in the year, relative to that of the top conductive flux (Figure 5.8e-f). As a result, HadGEM2-ES displays a very severe low bias in this flux throughout the freezing season (from September – April), whereas the two newer models display a smaller low bias from July – November (North Pole region) and July – September (Beaufort Sea region). This implies in turn that HadGEM2-ES has a large bias towards greater ice growth throughout the freezing season, consistent with the amplified seasonal cycle displayed by this model. By contrast, the smaller low bias in UKESM1.0 in particular is inconsistent with the damped seasonal cycle displayed by this model and this bias may therefore be an artefact of sampling (see below).

There are likely several mechanisms at work here and we discuss several months in detail to determine these, using the Beaufort Sea region (the patterns in the North Pole region are qualitatively similar). Firstly, in July, basal conductive fluxes of 3.6 ± 0.9 Wm^{-2} in HadGEM2-ES compare to -5.6 ± 2.0 Wm^{-}

$^{-2}$ in HadGEM3-GC3.1-LL, -5.2 ± 1.5 Wm^{-2} in UKESM1.0-LL and 0.4 ± 0.7 Wm^{-2} in the IMB estimates: HadGEM2-ES is biased high, the newer models biased low. The low bias in HadGEM3-GC3.1-LL and UKESM1.0-LL is the mirror image of the high bias in July top conductive flux, and is likely caused by the thick ice bias: thicker ice takes longer to heat up, and conductive fluxes are still flowing strongly towards the centre of the ice column in July whereas in the IMB estimates the ice temperature gradient is much closer to a straight line. The bias may also be partly caused by the lack of penetrating SW radiation in the ice, which would tend to increase top melting fluxes at the expense of heat in the ice interior. The HadGEM2-ES high bias can likely be explained simply from this model not modelling heat capacity; energy flows directly from the surface, which will tend to be slightly warmer than the ice base during the melting season.

Secondly, in October, basal conductive fluxes of -27.7 ± 10.6 Wm^{-2} in HadGEM2-ES compare to -15.0 ± 7.6 Wm^{-2} in HadGEM3-GC3.1-LL, -13.3 ± 7.0 Wm^{-2} in UKESM1.0-LL and -7.9 ± 16.8 Wm^{-2} in the IMB estimates. The substantial low bias of HadGEM2-ES is almost certainly mostly caused by the zero-layer approximation, with strong heat loss at the surface transmitted instantly to the lower surface of the ice in a way that does not occur in reality or in the two models with heat capacity. The HadGEM3-GC3.1-LL and UKESM1.0-LL distributions are not significantly different from the IMBs, partly because of the large spread of the IMB distribution in this month; their differences relative to the IMBs may, like the top conductive flux bias, be primarily associated with sampling error in the IMB estimates, as the lowest basal conductive flux values will tend to occur in thin, newly forming ice that is undersampled by the IMBs at this time of year. As noted in section 5.1, the major part of the variability in ice thickness occurs at scales smaller than a grid cell and this issue is unlikely therefore to be significantly addressed by restricting model data to buoy tracks.

Lastly, in January, basal conductive fluxes of -32.8 ± 5.7 Wm^{-2} in HadGEM2-ES compare to -14.9 ± 4.4 Wm^{-2} in HadGEM3-GC3.1-LL, -12.9 ± 4.9 in UKESM1.0-LL and -14.9 ± 4.6 in the IMB estimates. By this point in the melt season, the two newer models display no significant bias relative to the IMB estimates, while HadGEM2-ES continues to display a large low bias. While a major part of this is likely to be associated with the zero-layer approximation, we note that this

model displays a significant low bias in the top conductive flux in this month also, and it is therefore unlikely to be entirely caused by this.

For the basal conductive flux also, it is necessary to note the effect of the temporal offset of the IMBs relative to the models: modelled July top melting fluxes are 5-10 Wm^{-2} higher in the period 2000-2015 than for 1980-1999, hence the model biases are likely understated.

We now examine the ocean heat flux (Figure 5.8g-h). This flux displays strong positive skew in all months in the North Pole region (and in the summer in the Beaufort Sea region) associated with a small number of very high values, principally near the ice edge. As the ice edge is further north in HadGEM2-ES, there are a proportionally larger number of such points, and hence both mean and standard deviation of ocean heat flux in this model is much higher in the winter than in the other models. Models and IMB estimates largely agree on the shape of the seasonal cycle in ocean heat flux in both regions, with near-zero values from November – May rising to much higher values in July and August, and falling away again during September and October, the exception being the North Pole region in HadGEM2-ES where mean ocean heat fluxes remain significant in size throughout the winter.

In contrast to the other variables, summer model biases in ocean heat flux are opposite in sign in the two regions: all models display fluxes higher than the IMB estimatesin the North Pole region, and fluxes lower than the IMB estimates in the Beaufort Sea region. In both regions, HadGEM2-ES fluxes are highest and UKESM1.0-LL lowest. Hence in the North Pole region, HadGEM2-ES and HadGEM3-GC3.1-LL are biased high in July and August, but UKESM1.0-LL only in August. For example, in the North Pole region in July, ocean heat flux of 27.0 ± 14.0 Wm^{-2} in HadGEM2-ES compares to 19.1 ± 9.8 Wm^{-2} in HadGEM3-GC3.1-LL, 14.6 ± 7.6 Wm^{-2} in UKESM1.0-LL and 12.1 ± 7.4 Wm^{-2} in the IMB estimates. By contrast, in the Beaufort Sea region, HadGEM3-GC3.1-LL and UKESM1.0-LL are biased low in July and August, but the differences between HadGEM2-ES and the IMBs are not significant in either month. For example, in the Beaufort Sea region in July, ocean heat flux of 24.4 ± 14.1 Wm^{-2} in HadGEM2-ES compares to 19.6 ± 17.8 Wm^{-2} in HadGEM3-GC3.1-LL, 15.9 ± 14.6 Wm^{-2} in UKESM1.0-LL and 30.6 ± 19.6 in the IMB estimates.

As ocean heat flux is strongly determined by the temperature of the mixed layer, and studies from observations and models have shown that heating of the mixed layer during the Arctic summer is principally solar-driven (e.g McPhee et al., 2003; Perovich et al., 2008; Steele et al., 2010), the differing patterns of model biases in the two regions may be driven by ice area. Despite their very different average ice thicknesses, all three models share a common problem in sea ice modelling: the area of maximum ice thickness tends to be located towards the Pacific sector of the Arctic, whereas observations tend to show this being located towards the Atlantic edge. This means that whereas in reality the Beaufort Sea region will melt out more quickly than the North Pole region in summer, in the models the reverse is true.

To summarise, all models display a high bias in top melting flux in both regions relative to the IMBs, with the bias being highest in HadGEM2-ES and least in UKESM1.0-LL. HadGEM2-ES displays a strong low bias in basal conductive flux throughout the freezing season relative to the IMBs; HadGEM3-GC3.1 and UKESM1.0-LL display low biases which are much smaller, and restricted to the autumn. With the ocean heat flux, the bias is region dependent, with models tending to be biased high in the North Pole region and low in the Beaufort Sea region.

With HadGEM2-ES, the top melting and basal conductive flux biases are qualitatively consistent both with the model biases seen in Chapter 3 (summer net SW too high, volume balance too amplified) and with the surface flux bias induced by biases in other variables, as analysed in chapter 4 (ice thickness bias tending to cause too much ice growth in winter, melt onset and ice area biases tending to cause too much ice melting in summer).

A quantitative comparison shows that the IMB estimates suggest a much greater model bias towards excess ice growth and melt than does the ice thickness evaluation. Relative to PIOMAS, HadGEM2-ES overestimates the amplitude of ice growth and melt by 20cm over the North Pole region and 38cm over the Beaufort Sea region. By contrast, comparing means of the IMB-estimated and modelled distributions of top melting fluxes in this region gives a top melting bias of 69cm and 74cm in the two regions respectively; the bias in the basal conductive flux implies excess growth of 108cm and 112cm

respectively. It is likely that some of this discrepancy is explained by the distinction between average thickness over ice (the quantity evaluated by the IMBs) and average thickness over ice and ocean (evaluated by PIOMAS).

However, with HadGEM3-GC3.1-LL, and particularly UKESM1.0-LL, the biases relative to the IMBs are less obviously consistent with results in Chapter 3 and Chapter 4. In UKESM1.0-LL, the ice thickness evaluation relative to PIOMAS suggests that ice growth and melt are underestimated by 52cm and 26cm in the North Pole and Beaufort Sea regions. By contrast, the IMB evaluation shows biases of the opposite sign; the summer top melting is biased high by 12cm and 23cm respectively, and winter ice growth is biased high by 8cm and 11cm. As discussed above the discrepancy in top melting fluxes is can be partially explained by a low bias in basal melt in both HadGEM3-GC3.1-LL and UKESM1.0-LL. To understand the discrepancy in basal conductive flux, however, it is necessary to analyse the relationship between conductive flux, and ice and snow depth, using the data provided by the IMBs.

5.5 The relationship between conductive flux and ice and snow thickness

Both top and basal conductive flux are strongly related to sea ice and snow thickness. Thicker sea ice, and a thicker snow cover on the ice, is associated with a weaker temperature gradient, and hence a weaker conductive flux under identical atmospheric conditions. Therefore, to understand the meaning of a model bias (difference) in conductive flux, it is necessary to understand whether the bias arises from a bias in ice or snow thickness, or from a bias in the atmospheric thermal forcing. In particular, biases arising from bias in ice or snow thickness may be due to imperfect sampling by the IMBs (as IMBs tend not to be placed in the thinnest ice floes), while those arising from bias in atmospheric conditions will not be affected by this.

Sea ice thickness, as well as depth of snow on sea ice, is measured by the IMBs (and simulated with the models) simultaneously with conduction, which allows the causes of conductive flux bias (difference) to be examined. Firstly, for each IMB data point, and for each model grid cell, we calculate the sea ice thermal insulance $R_{ice} = h_{ice}/k_{ice} + h_{snow}/k_{snow}$, where h_{ice}, k_{ice}, h_{snow} and

k_{snow} denote ice thickness, ice conductivity, snow thickness and snow conductivity respectively. This quantity is effectively a measure of how difficult it is for energy to be lost through the ice; under similar atmospheric forcing, we expect a roughly inverse relationship between insulance and conduction. Because of this inverse, nonlinear relationship, grid cell mean thermal insulance in the models cannot be compared in a like-for-like manner to pointwise thermal insulance in the IMB estimates (as illustrated in Figure 5.9); for any thickness *H*, a cell of mean thickness *H* can transmit energy more efficiently than can a single column of thickness *H*, because most of the energy loss occurs in the thinnest ice categories. The problem is solved by using harmonic mean thermal insulance over categories to represent the insulance of a grid cell; a single column of this thickness loses energy at the same rate.

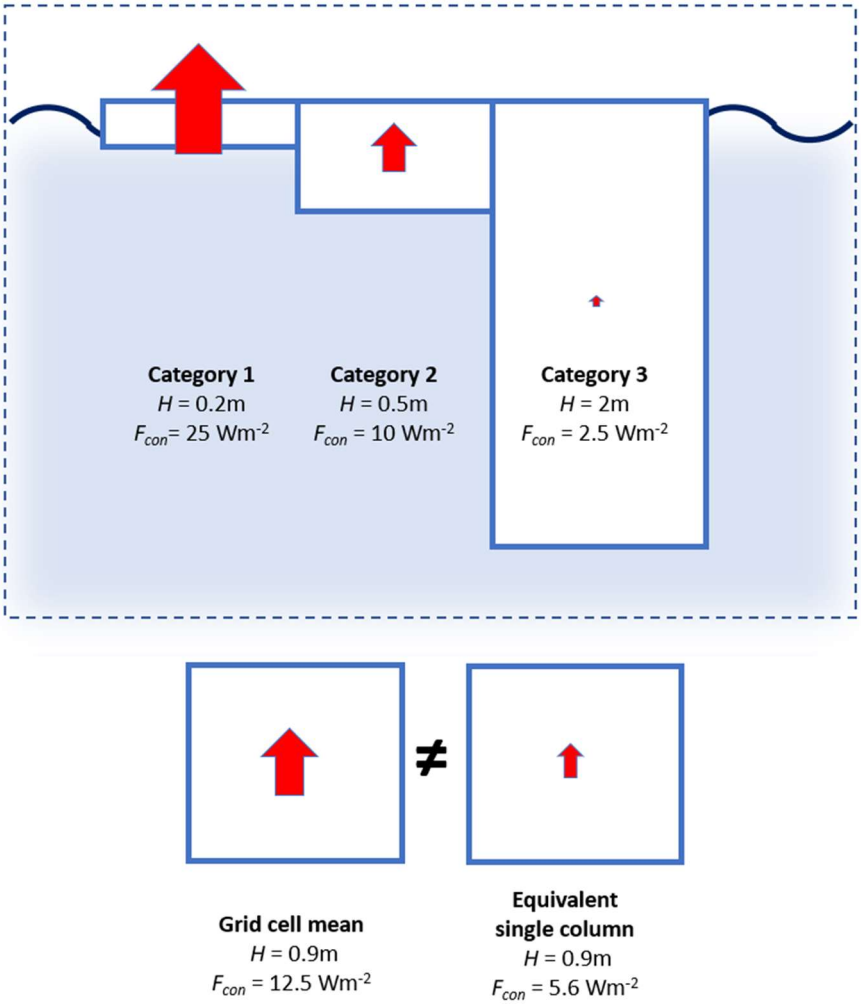


Figure 5.9. Schematic demonstrating why grid cell mean thermal insulance is not comparable to thermal insulance measured at a single point. The grid cell given as an example has three thickness categories (with equal ice areas for

the purposes of this example). For simplicity, an inverse relationship between conductive flux and ice thickness is assumed: $F_{con} = \frac{5}{H}$*. Grid cell mean thickness and conductive flux,* $0.9m$ *and* $12.5\ Wm^{-2}$ *are simple averages of thickness and conductive flux in each category. However, a single ice column of thickness* $0.9m$ *has a conductive flux of* $5.6\ Wm^{-2}$ *using this relation.*

We can then plot top and basal conductive flux against thermal insulance in the models and IMBs (Figure 5.10). In the models, higher thermal insulance tends to be associated with lower conductive fluxes, with the dependence becoming steeper at lower values of insulance; this relationship is less apparent in the IMBs. It is interesting to note that despite the large differences in average ice thickness between the models, differences in thermal insulance are much less. This is largely because insulance is mostly controlled by the fraction of ice in the thinnest category, while the high ice thickness in UKESM1.0 is largely driven by a high average ice thickness in the thickest category.

Plotting conductive fluxes against thermal insulance also immediately allows several model comparisons that were not previously possible. Firstly, it can be seen that at similar values of insulance, HadGEM2-ES displays much more strongly negative conductive fluxes than do HadGEM3-GC3.1 and UKESM1.0, and that this difference is greater for basal than for top conductive fluxes. This implies that a major part of the HadGEM2-ES differences is not due to differences in ice thickness and snow depth. Secondly, while HadGEM2-ES basal conduction is clearly biased low as a function of thermal insulance, HadGEM3-GC3.1-LL and UKESM1.0-LL still do not display a clear bias. For top conduction, the spread in the IMB values is very high at most values of insulance, making inferences more difficult.

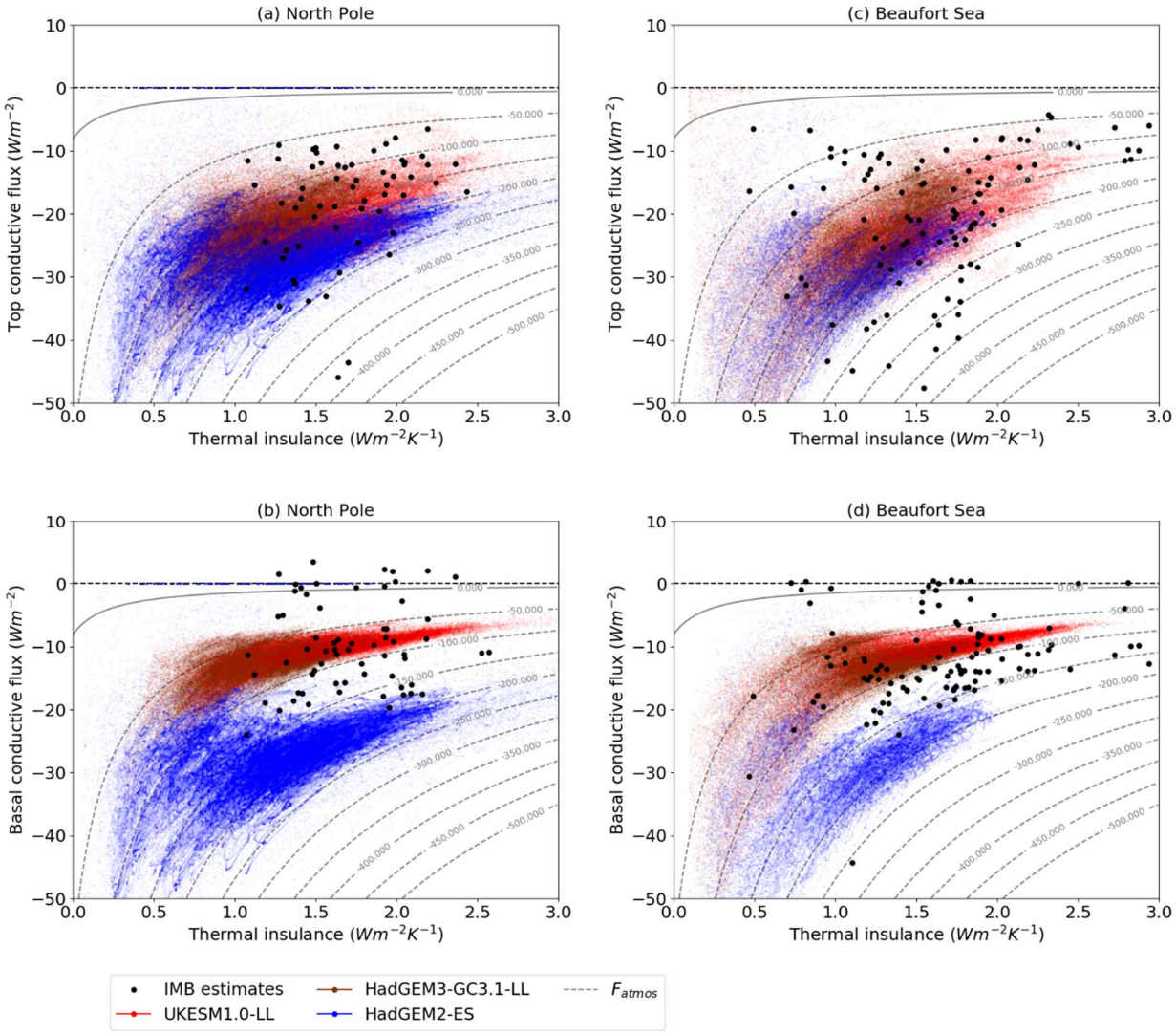


Figure 5.10. Scatter plots of (a), (c) top conductive fluxes and (b), (d) basal conductive fluxes as a function of thermal insulance R_{ice} in the (a-b) North Pole and (c-d) Beaufort Sea regions. Isolines of atmospheric thermal forcing F_{atmos} overlaid (see discussion below). Fluxes plotted are taken from the extended winter months November – March.

In terms of the effect on ice growth, the differences in basal conduction are decisive. As in Figure 5.8, HadGEM2-ES clearly stands out as modelling much lower (more strongly negative) basal conductive flux than do HadGEM3-GC3.1 or UKESM1.0, and as being biased low relative to the IMBs. Figure 5.9 shows that in addition, HadGEM2-ES is biased low relative to the IMBs as a function of thermal insulance: basal conductive fluxes are biased low under similar conditions of ice thickness and snow depth.

By contrast, whereas overall UKESM1.0-LL and HadGEM3-GC3.1-LL basal conduction is biased low relative to the IMB estimates, particularly in the early winter, Figure 5.10 shows that as a function of thermal insulance no bias can be discerned. Under similar conditions of snow depth and ice thickness, HadGEM3-GC3.1-LL and UKESM1.0-LL model similar basal conduction, and hence similar ice growth, to that estimated from the IMBs. Effectively, this sidesteps the problem of determining whether the IMB-measured fluxes are accurately sampling the thinner ice categories.

The ice thickness – basal conductive flux comparison improves understanding of the conductive flux biases. UKESM1.0-LL is known to display a damped seasonal cycle, with a bias towards less ice growth; Figure 5.10 shows that this bias is likely to arise at least in part because the model is biased towards less ice growth as a function of ice thickness and snow depth, likely because of atmospheric thermal forcing. By contrast, the seasonal cycle of HadGEM3-GC3.1-LL is less clearly damped, despite the model being similarly biased towards less ice growth as a function of ice thickness and snow depth. In this case, it is likely that the bias is counteracted by a model tendency to thinner ice in the early winter (as can for example be seen in the Beaufort Sea region, Figure 5.9d). The seasonal cycle of HadGEM2-ES is biased towards too much ice growth; Figure 5.9 shows that this model is also biased towards too much ice growth as a function of ice thickness and snow depth, and that a large part of this bias is due to the lack of heat storage in this model.

Atmospheric thermal forcing is likely an important driver of the thermal insulance – conductive flux ‘function’ in Figure 5.9. In order to quantify how the atmosphere affects this function, we seek a way to measure the tendency of the atmosphere to force conduction in the ice (alternatively, the ‘coldness’ of the atmosphere). Intuitively, it should be possible to deduce this from the conduction and the insulance, as points or cells with higher (lower) conduction and higher (lower) insulance should be associated with a greater (smaller) tendency of the atmosphere to force upwards conduction of heat within the ice, or a colder (milder) winter atmosphere. In Figure 5.9, for example, we would expect points with colder atmospheric forcing to fall towards the lower right of the panels, and points with milder atmospheric forcing towards the upper left.

We firstly try quantifying the atmospheric forcing by a very similar method to that used in Chapter 4 to calculate induced surface flux biases or differences. Firstly, the surface flux F_{sfc} is linearized about the freezing point $T_{sfc} = 0°C$, where T_{sfc} is surface temperature:

$$F_{sfc} = F_{atmos} + BT_{sfc} \quad (5.7)$$

By flux continuity

$$F_{sfc} = F_{condtop} \quad (5.8)$$

If we assume a uniform conductive flux (an assumption that will be revisited below)

$$F_{condtop} = \frac{(T_{sfc} - T_{bot})}{R_{ice}} \quad (5.9)$$

As in chapter 4, we can eliminate T_{sfc}, and rearrange:

$$F_{atmos} = F_{condtop}(1 - BR_{ice}) - BT_{bot} \quad (5.10)$$

As a function of $F_{condtop}$ and R_{ice}, contour lines of F_{atmos} can be plotted over Figure 5.9. F_{atmos} can be viewed as representing the atmospheric forcing that would be observed with a uniform conductive flux through the ice, in which the atmospheric forcing is instantly communicated to the base of the ice. This is actually the case in HadGEM2-ES, and the blue scatter points accurately represent the atmospheric forcing in this model. However, it is not the case in the newer models or the IMB estimates. For example, in months where strong surface cooling was present, using top conductive flux to calculate F_{atmos} would result in an overestimation, while using basal conductive flux would result in an underestimation. We can allow for this effect by measuring how far top and basal conductive fluxes deviate from the 'average' conductive flux,

$$F_{condavg} = \frac{T_{sfc} - T_{bot}}{R_{ice}} \quad (5.11)$$

We define

$$F_{conddevtop} = F_{condtop} - F_{condavg} \quad (5.12)$$

and

$$F_{conddevbot} = F_{condbot} - F_{condavg} \quad (5.13)$$

By flux continuity,

$$F_{condtop} = F_{sfc} \quad (5.14)$$

Hence combining (5.11), (5.12) and (5.14) we obtain

$$F_{atmos} = F_{condtop}(1 - BR_{ice}) - BT_{bot} + BF_{conddevtop}R_{ice} \quad (5.15)$$

F_{atmos} is a measurement of atmospheric thermal forcing that can be used for all three models, and for the IMB estimates, with $F_{conddevtop} = 0$ for HadGEM2-ES. Because F_{atmos} also depends on $F_{conddevtop}$, it is less easy to represent with contour lines as in Figure 5.7. Instead, we 'flip the picture', and scatter F_{atmos} against R_{ice} for models and IMB estimates (Figure 5.11).

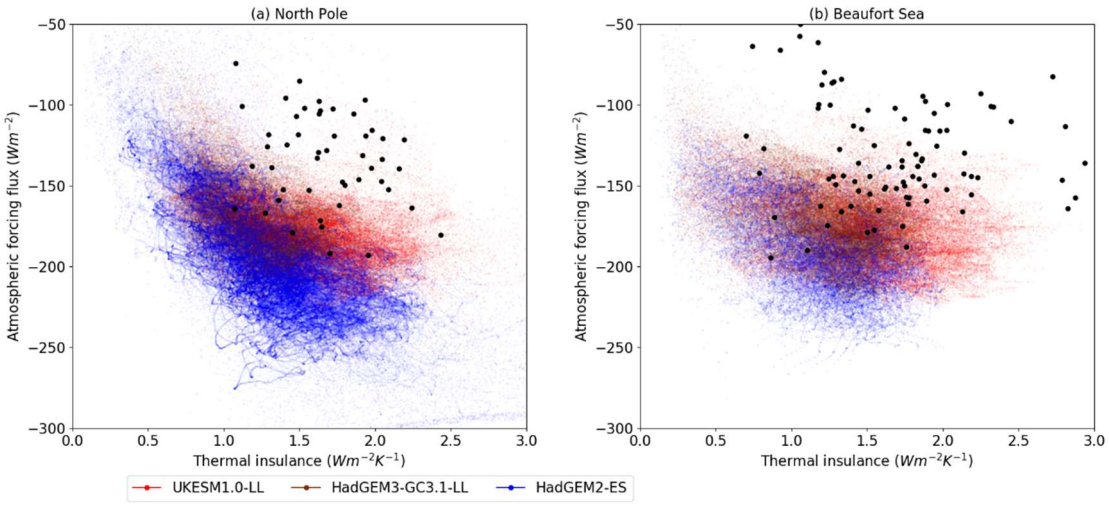


Figure 5.11. Scatter plot of thermal insulance R_{ice} against atmospheric thermal forcing (F_{atmos}) diagnosed from conductive flux in HadGEM2-ES, HadGEM3-GC3.1 and UKESM1.0, as well as the in the IMB estimates, for the (a) North Pole and (b) Beaufort Sea regions.

This shows that although there is surprisingly little difference in atmospheric thermal forcing during the winter in the three models, all are biased low to some degree with respect to the IMB estimates. Effectively, we have removed the effect of differences in ice heat storage (including the zero-layer approximation of HadGEM2-ES): as a result, HadGEM2-ES is much less of an outlier, and

there are signs that HadGEM3-GC3.1-LL and UKESM1.0-LL, like HadGEM2-ES, exhibit a cold bias in atmospheric thermal forcing.

This is consistent with the evaluation of downwelling LW in chapter 3, which showed similar values in the three models with all being biased high with respect to estimates from ERAI, ISCCP-FD and CERES. It is also consistent with the ISF model difference evaluation in section 4.4, which showed that atmospheric forcing plays very little role in the model differences in sea ice growth during the freezing season. The ISF analysis was unable to address the role that the lack of heat capacity in HadGEM2-ES plays in increasing sea ice growth in this model; Figures 5.10 and 5.11 show that in HadGEM2-ES the atmospheric cooling is transferred much more efficiently to the base of the ice, driving substantially larger ice growth than in HadGEM3-GC3.1 and UKESM1.0, even as a function of ice thickness.

5.6 Summary

By synthesizing the information from the Arctic IMB network into an observational dataset of internal sea ice energy fluxes, it is possible to evaluate the ice thermodynamics of HadGEM2-ES, HadGEM3-GC3.1 and UKESM1.0 in the North Pole and Beaufort Sea regions. The results are consistent both with the model evaluation of Chapter 3, and with the ISF analysis of Chapter 4, but add a greater depth of understanding. Summer top melting fluxes are biased high in all three models, most severely in HadGEM2-ES and least severely in UKESM1.0. In HadGEM3-GC3.1-LL and UKESM1.0-LL, these are countered by a low bias in basal melt. Winter basal conduction fluxes are biased low in all three models, also most severely in HadGEM2-ES and least severely in UKESM1.0. Comparing top and basal conduction fluxes shows that a substantial part of the HadGEM2-ES bias is due to the lack of heat capacity in this model.

By directly comparing conduction to ice and snow thickness, we can see that while HadGEM2-ES basal conduction is biased low with respect to the IMBs as a function of ice thickness, HadGEM3-GC3.1 and UKESM1.0 display indeterminate bias. We can use information about surface temperature to obtain an estimate of atmospheric thermal forcing F_{atmos} at each model grid point and

IMB measurement data point, and see that atmospheric forcing is likely equally biased cold in all three models. Again, this is consistent with the analysis in Chapter 3 and 4, but shows how the thicker ice of HadGEM3-GC3.1 and UKESM1.0 converts this to a much lower basal conductive flux.

Despite the large differences between the sea ice simulations of HadGEM2-ES, HadGEM3-GC3.1 and UKESM1.0, the underlying drivers of the sea ice in winter are likely to be very similar in all three models, and all are almost certainly biased cold. By contrast, in summer there is evidence, in the differing timing of melt onset, shown to propagate via surface albedo, net SW and top melting to the volume balance, for key differences between the simulations. It can be concluded that the differences between the model simulations are almost entirely due to model differences in the summer.

The cause of the biases in the model simulations is a separate question. All models are biased cold in the winter, creating a tendency towards thicker ice. In the case of HadGEM2-ES, however, this is overwhelmed by the early surface melt onset, which sets off a chain of biases (surface albedo, snow area, net SW and top melting) whose effect is to reduce the annual mean ice thickness. The three models display different levels of warm bias during the summer: it is reasonable to suppose that the surface melt onset biases set off a chain of biases (surface albedo, snow area, net SW and top melting) which have a large effect on the summer ice volume balance. The indeterminate summer bias of UKESM1.0 allows the effect of the winter cold bias to show clearly; the HadGEM3.GC3.1 is enough to partially ameliorate the effect of the winter cold bias, the HadGEM2-ES bias strong enough to completely overwhelm it. In the following chapter, we explore in detail what this means in more general terms.

6. Conclusions

6.1 Results summary

Arctic sea ice volume is driven by a complex interplay between atmospheric and oceanic forcing, and intrinsic feedbacks of the sea ice state. Evaluation of ice area and/or volume can present only a partial view of a model simulation, because this does not address the reasons for model errors. In this thesis, two new methods of model evaluation have been presented, allowing a deeper understanding of errors in the sea ice state, by evaluating the fundamental processes in the atmosphere, ocean and ice driving sea ice growth and melt. Firstly, the induced surface flux (ISF) analysis estimates the effect of errors in individual model variables on the surface flux, which determines the ice volume balance to first order. Secondly, the ice mass balance buoy (IMB) evaluation detects biases in the internal ice thermodynamics. The approaches are strongly related, but complementary: the ISF analysis diagnoses the proximate causes of large-scale biases in ice volume balance in the atmosphere and sea ice, while the IMB analysis diagnoses biases in the sea ice energy fluxes that determine how these affect the ice.

The ISF and IMB methods have been applied to three models that span a large part of the model spread in present-day ice volume. The CMIP5 model HadGEM2-ES displays a thin ice bias in the annual mean, and an amplified seasonal cycle. The CMIP6 model HadGEM3-GC3.1-LL is thicker, and less amplified, with indeterminate biases; the CMIP6 model UKESM1.0-LL is clearly biased thick and has a damped seasonal cycle. Evaluating surface radiation shows that the differences in summer melting between the models can be explained by differences in upwelling SW, while the differences in winter freezing may be explained by differences in net downwelling LW. The model biases in volume balance are therefore also likely explained by biases in these variables. Evaluation of factors affecting surface albedo shows that differences in upwelling SW could in turn be explained by model differences in timing of surface melt onset, snow cover, and in meltpond parameterisation, but are likely to be opposed by differences in snow parameterisation.

The ISF analysis is developed to quantify these contributions. Simple models are used to estimate the surface flux bias induced by a model bias in particular

variables at points in model space and time: the resulting fields of ISF bias are averaged to determine large-scale drivers of model biases. We see that the amplified seasonal cycle of HadGEM2-ES is driven by biases in surface melt onset and ice area during summer, and by biases in downwelling LW and ice thickness during winter. In HadGEM3-GC3.1-LL and UKESM1.0-LL, the ice area and ice thickness ISF biases take the opposing sign, and the surface melt onset ISF bias is smaller (near-zero in UKESM1.0-LL), causing the total ISF bias to be negative in summer and positive in winter, consistent with a damped seasonal cycle.

The ISF framework can also be used to understand the differences between model simulations directly. In the summer, this enables analysis of a wider range of variables at daily resolution, showing how in UKESM1.0-LL the delayed timing of melt onset relative to HadGEM2-ES causes a relative decrease in surface flux in late May and early June, while the differences in snow thickness cause a similar decrease in late June. In the winter, this enables use of the full category ice thickness distribution, showing that the differences in ice growth between the models are entirely due to ice thickness differences. In other words, melt onset and snow cover differences force the model ice melting differences, and the differences in ice thickness compensate for these during the winter.

To support and complement the ISF analysis, fluxes of top melt, top conduction, basal conduction and ocean heat flux are evaluated directly using the IMB dataset, over the densely-sampled North Pole and Beaufort Sea regions. Top melting fluxes are biased high during the summer, and basal conductive fluxes biased low during the winter, in all three models, with HadGEM2-ES displaying the greatest bias and UKESM1.0-LL the least; for the newer models, this may be countered by negative basal melting flux bias. Comparison of top conductive flux bias to that of basal conductive flux shows that the zero-layer approximation of HadGEM2-ES is a major contributing factor to its large basal conductive flux bias.

By combining the IMB measurements of conductive flux, snow depth and ice thickness, we can view conductive fluxes as a function of thermal insulance.

While HadGEM2-ES is biased high as a function of insulance, the newer CMIP6 models are actually biased low.

6.2 Discussion

The three models analysed display very different simulations of sea ice state. In this section, we discuss the extent to which the ISF framework and IMB evaluation are able to explain this, and what this could mean for their applicability to sea ice simulations in general.

The ISF framework quantifies the proximate causes of differences in the seasonal rate of change of sea ice volume, rather than in the sea ice volume itself. Nevertheless, the two are very closely linked via the surface albedo and thickness-growth feedbacks. Very simply, thin ice will both grow and melt more quickly than thick ice under otherwise identical conditions. This is shown explicitly by the ISF analysis itself: for each model, the major part of the ice growth/melt biases are driven by biases in the sea ice area or thickness, or the surface albedo and thickness/growth feedbacks. The same is true for the differences between models, with differences in ice thickness dominating in winter and those in ice area dominating in summer.

The ice volume, and the amount of annual growth and melt, can effectively be viewed as a coupled system. Changes to an external variable such as downwelling radiation, or snow cover, lead to an initial change in annual growth and/or melt. This in turn induces a change in the ice volume (i.e., in the ice area or thickness), inducing a further change in the growth or melt via the effect on the surface flux. The chain of causality continues until ice growth and melt are once again in balance, and a new equilibrium is reached.

The ISF results can be interpreted as follows. The raw variable contributions show the proximate causes of ice melt/growth differences. The variable contributions **excluding ice area and thickness** show the proximate causes of the differences in the whole ice volume – ice growth/melt coupled system. For example, comparing UKESM1.0-LL and HadGEM2-ES (section 4.4), when the ice thickness and area contributions are excluded the largest remaining ISF differences are due to the variables affecting surface albedo in early summer:

surface melt onset, melt pond parameterisation, snow thickness and snow parameterisation. Collectively, they account for far less ISF difference than the ice area term. Yet they likely account for a major part of the difference between the sea ice systems of UKESM1.0-LL and HadGEM2-ES, because the ice area and ice volume differences can largely be traced back to the initial reduced ice melt they trigger.

More generally, the ‘external’ variables – downwelling radiation, snow cover and surface melt – set the parameters of the ice volume-ice melt/growth coupled system. In effect, they determine the climate in which the sea ice will find an equilibrium volume with ice melt and growth equal. This can be visualised by calculating ice melt and ice growth curves, as a function of annual mean ice thickness, with the simple model used in section 4 (Figure 6.1), forced with the Arctic Ocean average surface radiative fluxes from each model. For each model, and for the real world, the external variables determine the relationship between ice thickness, and ice growth and melt: the annual mean ice volume lies where the two curves meet.

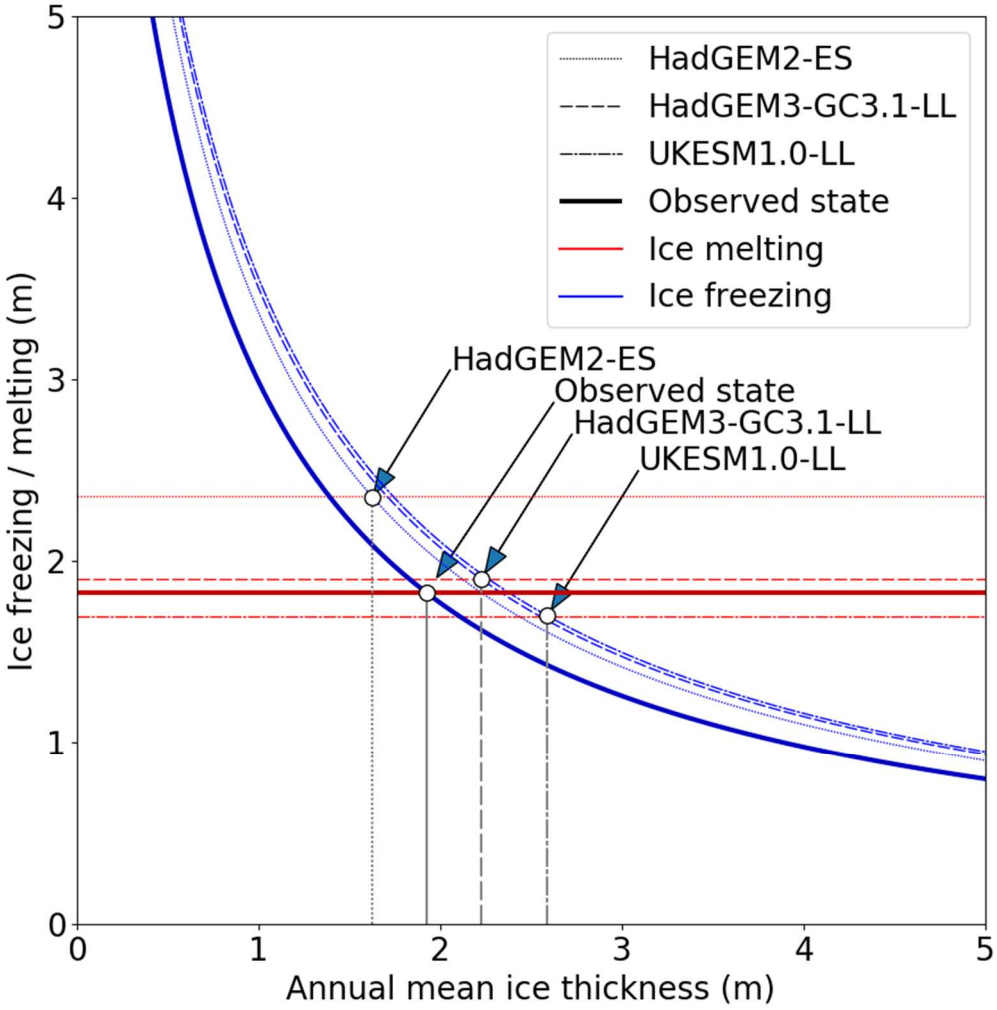


Figure 6.1. An illustration of idealised ice thickness-ice growth (blue line) and ice thickness-ice melt (red line) relationships in the evaluated models and in observations, as produced by the ISF parameterisation. The graph demonstrates how the ice growth and ice melt curves determine equilibrium ice thickness in each model climate.

This conceptual picture accurately reproduces the qualitative differences between each model, and the model biases. All three models share similar ice growth curves, biased high relative to the real world: this reflects the similar low downwelling LW biases. By contrast, HadGEM2-ES is an outlier in terms of ice melt, biased high relative to the real world and the other models. These differences are enough to qualitatively reproduce the model ice thickness biases, although the full extent of the UKESM1.0-LL thick bias is not captured.

Due to the interplay between ice volume, and ice growth/melt, the impact of external drivers on the sea ice varies greatly depending on the time of year. On the one hand, the effect of surface flux biases during the freezing season is diminished by the thickness-growth feedback, particularly early in the season. Differences created will tend to be reduced as the freezing season proceeds, as thin ice grows more quickly than thick ice. On the other hand, the effect of surface flux biases during the melting season is enhanced by the surface albedo feedback, particularly early in the season. Differences created will tend to increase as the season progresses, as thinner, warmer, less extensive ice has a lower albedo than thicker, colder, more extensive ice. Hence small differences in forcing can have a large effect in the late spring and early summer, whereas small differences during the freezing season will tend to only have a small impact. This is consistent with the prediction of DeWeaver et al. (2008) that sea ice state is more sensitive to surface forcing during the ice melt season than during the ice freezing season.

A particularly useful aspect of the ‘inter-model’ part of the ISF analysis is that it allows the direct effect of some model parameterisation changes to be quantified, namely the change to an explicit meltpond scheme and the change of snow area parameterisation. In each case, the impact on the surface flux, although small compared to the melt onset and snow thickness terms (let alone the ice area term), is not negligible, and because the impact is felt in the early summer likely has a significant impact on the sea ice state. This perhaps offers a useful perspective on the importance of sea ice model improvements versus model forcing. The effect of such model improvements may be small, but could still have significant effects, particularly if the effects are concentrated in the key late spring-early summer time of year.

The inter-model ISF analysis is unable to address the effects of moving to a multilayer sea ice model, with heat capacity: for this, the IMB evaluation is highly useful, and shows that in fact the zero-layer approximation plays a major role in amplifying the sea ice growth of HadGEM2-ES, complementary to the thin ice bias. At first sight, this might suggest that if HadGEM2-ES had employed a multilayer sea ice model, its thin ice bias would have been still worse, as the ice growth bias would not then have been strong enough to overcome the ice melt bias caused by the early summer surface albedo biases.

However, it is likely that the zero-layer thermodynamics is also a major driver of the surface melt onset bias itself.

This contains important implications for the desirability of multilayer thermodynamics in sea ice models. A basic comparison of the sensible and the latent heat stored in sea ice suggests that the presence of heat storage in the ice is a factor of only peripheral importance. However, most of the sensible heating of sea ice occurs in a short period of time during the late spring and early summer, a time when the surface temperature has a disproportionate impact on the amount of ice melt over the season as a whole. Hence it is possible that the thermodynamics scheme, also, has a significant effect on the overall sea ice state.

6.3 Implications for future model evaluation

The ISF framework and the IMB evaluation allow a deeper understanding of the reasons for biases in, and differences between, the sea ice simulations of HadGEM2-ES, HadGEM3-GC3.1-LL and UKESM1.0-LL. The ISF framework attributes differences and biases to specific drivers in the atmosphere and sea ice models; the IMB evaluation shows more clearly how these lead to the biases in the sea ice state. Given this, it is necessary to consider to what extent either method could be applied to other models, in particular models from the CMIP6 ensemble.

The ISF framework, as presented in this thesis, is specifically designed to approximate the three models under consideration. The framework depends upon estimating modelled surface flux, with reasonable accuracy, as a function of key climate variables, and the optimal way of doing this differs between models. For example, in approximating snow fraction in this study it is clearly desirable to use each model's own parameterisation. Hence while the ISF framework can likely be generalised to other models, in practice the analysis of any model requires careful consideration of the model design, structure and parameterisations. Because of this, the ISF framework is likely best suited to comparison of small numbers of models, as in this study, and not to evaluation of the whole CMIP6 ensemble in one study, which could run the risk of drawing

incorrect conclusions by missing important details about the design of individual models.

It is appropriate to consider also the process by which the ISF framework was devised: by evaluating first ice state, then surface flux, then specific variables influencing surface flux (via surface albedo), and by observing that the differences in sea ice state could be qualitatively explained by a simple surface flux framework. If extending the analysis of this thesis to the whole CMIP6 ensemble, a sensible first step would be to evaluate these variables, and to judge whether this conclusion still held. For example, a simple surface flux framework would fail to explain biases in model ice growth and melt in cases where oceanic heat convergence was a major driver of volume balance. In the three models evaluated in this thesis, and in reality, oceanic heat convergence is likely only of peripheral importance in much of the Arctic Ocean, but it may be considerably more important in some models.

To a similar end, a particular conclusion of this thesis is the exceptional importance of biases and differences in the model surface energy budget in the late spring and early summer in determining biases and differences in the model sea ice states. A key objective of any extension of the analysis to CMIP6 should be to determine whether this conclusion holds more widely.

Generalising the IMB evaluation presents different challenges. Any model that reports fluxes of top and basal mass balance, and top and basal conduction, can in principle be evaluated using the IMB dataset, and therefore a simultaneous evaluation of internal ice fluxes in much of the CMIP6 ensemble should be possible. However, evaluation is only possible in the North Pole and Beaufort Sea regions of the Arctic, and confidence will remain limited by the uncertainty in the IMB-measured fluxes. Despite this, it is the author's intention to evaluate internal ice fluxes in the CMIP6 ensemble in this way as a sequel to this thesis.

6.4 Implications for Arctic observations

Observational uncertainty is the largest cause of uncertainty in the ISF analysis; this suggests that if observational uncertainty could be reduced, this framework

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

Hence the ISF framework underlines the importance of improving the accuracy and coverage of observations of key climate variables in the Arctic, and makes explicit why this is important. **Better observations of surface fluxes are needed** because the surface flux and the ice volume balance are intimately related**; better observations of snow cover are needed**, particularly, because in the late spring and early summer variables influencing surface albedo are disproportionately important for the sea ice state.

By contrast, the IMB evaluation shows that the Arctic IMB programme is an essential data-gathering exercise. The benefits of the programme extend not only to monitoring the Arctic, but also to the meaningful evaluation of sea ice models. However, a major limitation of the data is uncertainty due to lack of knowledge of ice salinity. **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.

6.5 Overall summary and conclusion

Ice area and volume are insufficient metrics by which to evaluate sea ice models because correct simulations can be obtained by cancelling model errors. In this thesis, two new methods have been introduced by which the underlying mechanisms causing ice growth and melt can be evaluated. The induced surface flux framework combines simple models with existing reference datasets, while the ice mass balance buoy evaluation introduces a new dataset by which to evaluate ice mass balance and thermodynamics.

Using these methods, we have obtained a detailed understanding of the reasons for model biases and differences in simulated ice growth and melt in three coupled climate models. Crucially, because the sea ice state itself predominates in determining modelled growth and melt, the ice volume and the ice growth / melt can be viewed as a coupled system. Hence we obtain, by extension, understanding of the reasons for model biases and differences in ice volume: small differences in forcing in the early melt season cause large differences in melting later in the summer via the surface albedo feedback, and hence in annual mean ice volume. A general tendency to cold conditions during the winter in all models predisposes all towards a thick ice bias, but the early summer forcing differences are able to overwhelm this effect.

In this way, model biases in sea ice state can be attributed to specific model errors, and attention is focussed on the most important areas for model development. The downwelling LW bias, for example, is likely linked to cloud liquid water fraction deficiencies, suggesting this remains an important area to focus model improvements on for Arctic climate. As or more important, however, is understanding the mechanisms by which surface melt onset is incorrectly simulated.

Similarly, model differences in sea ice state can be attributed to specific model improvements, or changes in the atmospheric state over the Arctic. It is seen that the direct effect of meltpond and snow area parameterisations is somewhat smaller than that of the surface melt onset or downwelling LW changes, but is nevertheless significant in precipitating additional melting early in the summer.

The methods demonstrated in this thesis have been used to compare identical time periods from different models, but could also in theory be used to compare different time periods from the same model, in order to better understand drivers of future sea ice change. As noted in Chapter 1, a method similar to the ISF framework was used by Holland and Landrum, 2015 in this way to understand drivers of net SW changes over sea ice-covered regions; the ISF method is a generalisation of this framework, and could in theory attribute drivers of changes in net surface flux to changes in a wider range of model variables.

The ISF framework would also be particularly useful in assessing model sensitivity experiments in more detail, in terms of quantifying how much of a modelled sea ice perturbation was due to a change in atmospheric forcing, and how much due to feedbacks of the sea ice state. For example, any perturbation to a surface albedo parameter would affect the rate of sea ice melt in multiple ways. The melt season ISF difference framework presented in section 4.4 would allow separate quantification of how the perturbation affects ice melt via its effect on melt onset timing, meltpond area, snow area and ice area. This would also enable better precision in model tuning, with a better understanding of precisely how the tuned parameter produces the desired outcome, and correspondingly a lower risk of side effects.

Finally, the new evaluation methods show more clearly the most important deficiencies in the current Arctic observation network. Downwelling radiation and snow cover are potentially important drivers of ice growth / melt biases, but cannot be evaluated as accurately as many other terms, particularly during the summer. Knowledge of ice salinity at the IMB locations would further enhance the usefulness of these instruments for model evaluation.

Appendix: Analysis of error in the ISF biases

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

A1. Error in calculating surface flux dependence

To understand error in calculating dependence of surface flux on model variables, fields of the approximated surface flux $g_{x,t}$ are compared to those of the real modelled surface flux F_{sfc}. The $g_{x,t}$ are found to capture well the large-scale seasonal and spatial variation in surface flux, but are prone to systematic bias which varies seasonally. For most of the year the bias is positive, reaching a maximum in the early winter, but in the early summer the bias is neutral in HadGEM2-ES, and strongly negative in HadGEM3-GC3.1-LL and UKESM1.0-LL (Figure 1).

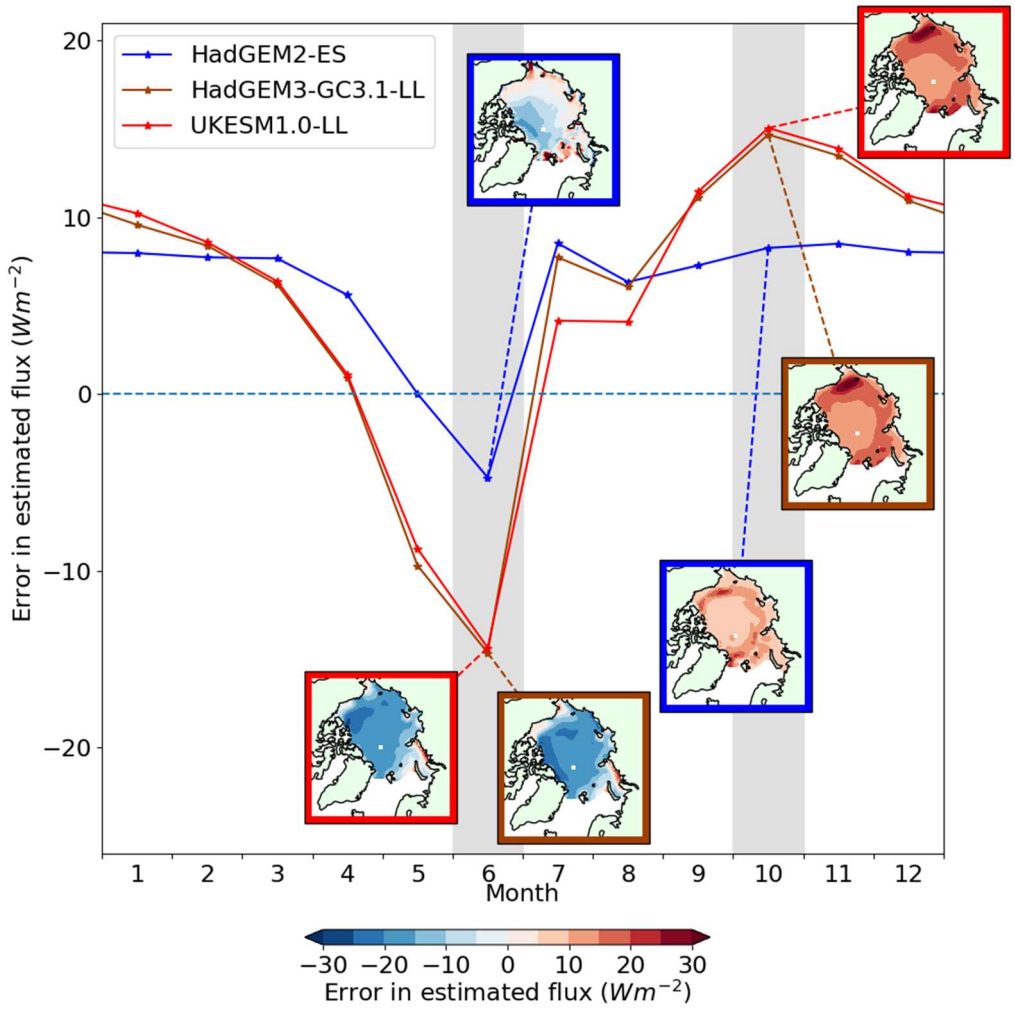


Figure A1. Error in estimating modelled surface flux in HadGEM2-ES, HadGEM3-GC3.1-LL and UKESM1.0-LL with the simple model used for the ISF framework. 1980-1999 average flux error over the Arctic Ocean region is shown, with maps of flux error for June and October (model indicated by border colour).

Looking first at the bias in the freezing season, maps of approximated surface flux bias during the freezing season show that highest values occur in regions of thin, newly-forming ice, with near-zero bias elsewhere. It is deduced that the simple model tends to underestimate efficiency of heat loss during winter over thin or newly-forming ice.

A possible cause of this bias is covariance in time between $\gamma_{ice-REF}^{cat}$ and R_{ice}^{cat} within each month, particularly in the first ice category; during the freezing season, occurrence of high fractions of ice in category 1, the thinnest category, would be expected to be associated with formation of new ice, and

correspondingly lower mean thicknesses of ice in this category, lower values of R_{ice}^{cat} and higher values of $\left(1 - BR_{ice}^{cat}\right)^{-1}$. A calculation using daily values of $\gamma_{ice-REF}^{cat}$ ranging from 0.1 – 0.9, and daily values of h_I^{cat} ranging from 0.2 – 0.5m, predicts that this effect could lead to an underestimation of up to 18% of the magnitude of the surface flux, sufficient to explain all of that seen in HadGEM2-ES, and most in HadGEM3-GC3.1-LL and UKESM1.0-LL. This effect would produce a corresponding underestimation of the rate of dependence of surface flux on downwelling LW radiation and ice thickness throughout the freezing season. It is calculated that the downwelling LW component of the ISF bias is underestimated by 0.6 Wm^{-2}, 0.0 Wm^{-2} and -0.2 Wm^{-2} for the freezing season on average due to this effect in HadGEM2-ES, HadGEM3-GC3.1-LL and UKESM1.0-LL respectively.

In July-August, a different mechanism is likely to be responsible for the overestimation of surface flux. A possible contributing factor to this bias is within-month covariance between ice area and downwelling SW. During July and August, both downwelling SW and surface albedo fall sharply, an effect that would cause the monthly mean surface flux to be overestimated. To approximate the magnitude of this effect, monthly trends in these variables were estimated by computing half the difference between modelled fields for the following and previous month. For July, this method predicts surface flux to be overestimated in the Siberian seas by 5-15 Wm^{-2} in HadGEM2-ES, and by 10-20 Wm^{-2} in HadGEM3-GC3.1-LL and UKESM1.0-LL, with smaller differences in the southern Beaufort and Chukchi Seas. This is consistent with the magnitude and spatial pattern of the estimated flux bias seen in July, and it is concluded that this mechanism is the primary driver of the estimated surface flux errors in this month.

To examine the effect this mechanism would have on the calculated SW and ice area ISF biases, it is necessary to examine both covariance between downwelling SW bias and mean ice area, and between ice area bias and mean downwelling SW. By similarly approximating the trend in monthly mean model bias as half the difference between model bias in the adjacent months, the error in downwelling SW and ice area contributions were evaluated. Error in the downwelling SW term was found to be largest early in the summer, with errors

of -2.7 Wm^{-2}, -2.6 Wm-2 and in June; July and August errors were below 1 Wm-2 in magnitude in all models. Error in the ice area term is of a similar magnitude in all models, but tends to be higher later in the summer; for example, HadGEM2-ES displays errors of -1.7 Wm^{-2} and -1.6 Wm^{-2} in July and August respectively, while HadGEM3-GC3.1-LL displays errors of 0.6 Wm-2 and 1.7 Wm-2 in these months. In all cases, the estimated error is small relative both to the total ISF bias, and to the observational uncertainty.

Finally the underestimation of surface flux in June in HadGEM3-GC3.1-LL and UKESM1.0-LL is discussed. It is hypothesised that the zero-layer approximation in the simple model is the major cause of this bias, via errors caused in the subsequent estimation of surface temperature and upwelling LW. In Figure 5.8 it is shown that in June, top and basal conductive fluxes in HadGEM3-GC.3.1-LL and UKESM1.0-LL differ by 15-20 Wm^{-2} in both the North Pole and Beaufort Sea regions. The simple model used to approximate surface flux, however, assumes a uniform conductive flux, with associated higher surface temperature during periods of warming, producing a lower net down surface flux. The widespread bias of -15 to -20 Wm^{-2} is of a consistent sign and magnitude with this effect. It is unlikely that this effect would have a significant effect on surface flux dependence on any variable except the ice thickness or area. This is because the sensible heat storage flux, accounting for most of the bias, is independent of all other variables on an instantaneous timescale, and partial derivatives would go to zero.

A2. Error in characterising induced surface flux bias

The surface flux dependencies, for each variable, are evaluated at a model state which is itself biased. This introduces an error in characterising the induced surface flux bias. For example, a component of the surface flux, net SW, is equal to $F_{SW\downarrow}\left(1-\alpha_{sfc}\right)$, and induced surface flux biases due to model biases in $F_{SW\downarrow}$ and α_{sfc} would be calculated as $F'_{SW\downarrow}\left(1-\alpha_{sfc}^{\mathrm{mod}}\right)$ and $F_{SW\downarrow}^{\mathrm{mod}}\alpha'_{sfc}$ respectively. However, the sum of the two induced surface flux biases will not

be exactly equal to the true surface flux bias, $F_{SW\downarrow}^{\mathrm{mod}}\left(1-\alpha_{sfc}^{\mathrm{mod}}\right)-F_{SW\downarrow}^{obs}\left(1-\alpha_{sfc}^{obs}\right)$, but will differ from it by $F_{SW\downarrow}^{'}\alpha_{sfc}^{'}$.

More generally, the difference between the surface flux bias $F_{sfc}^{'}$ and the sum of the induced surface flux biases $\sum_i v_i^{'}\partial g_{x,t}/v_i$ can be approximated by $\sum_{\substack{i,j \\ i\neq j}} v_i^{'}v_j^{'}\partial^2 g_{x,t}/\partial v_i \partial v_j$, a term that can be calculated relatively easily as many of the derivatives go to zero. Averaged over the Arctic Ocean this term was small (below 1 Wm^{-2} in magnitude) in most months of the year, but of significant size in October (3.6 Wm^{-2}, -2.5Wm^{-2} and -1.9Wm^{-2} in HadGEM2-ES, HadGEM3-GC3.1-LL and UKESM1.0-LL respectively), due to co-location of substantial negative biases in downwelling LW and category 1 ice thickness in this month, indicating that the true surface flux bias in this month may be substantially smaller (in absolute terms) than the -11.5 Wm^{-2} obtained from summing the ISF biases.

Finally, the induced surface flux calculation implicitly assumes a linear dependence of surface flux on each climate variable. However, this is not the case for the ice thickness, where higher-order derivatives do not go to zero, and in some regions of thinner ice actually diverge. It is possible to quantify the error introduced by the assumption of linearity by comparing the partial derivative $\left(A+BT_b\right)a_{cat}\left(1-BR_{cat}^{ice}\right)^{-2}\left(\sum_{cat}a_{cat}\right)^{-1}$ to the quantity $\left(A+BT_b\right)a_{cat}\left(1-BR_{cat}^{ice}\right)^{-1}\left(1-BR_{cat}^{ice-REF}\right)^{-1}\left(\sum_{cat}a_{cat}\right)^{-1}$, where $R_{cat}^{ice-REF}=h_I^{OBS}/k_I+h_S/k_S$, h_I^{OBS} being climatological ice thickness in the reference dataset, in this case PIOMAS, and all other terms defined as in Section 4. It can be shown that multiplying this quantity by the model bias produces the exact bias in estimated surface flux that is being approximated by $\partial g_{x,t}/\partial h_I\left(h_I^{MODEL}-h_I^{OBS}\right)$. Hence the bias in the ice thickness component induced by the nonlinearity can be calculated directly. The nonlinearity causes the ice thickness component to be overestimated in magnitude by 0.7, 1.0 and 1.1 Wm^{-2} on average from October-April in HadGEM2-ES, HadGEM3-GC3.1-LL and UKESM1.0-LL respectively,

with a maximum overestimation of 1.9 Wm^{-2}, 2.1 Wm^{-2} and 2.5 Wm^{-2} occuring in November for each model respectively.

Bibliography

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

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

Bitz, C.M.: Some aspects of uncertainty in predicting sea ice thinning, in Arctic Sea Ice Decline: observations, projections, mechanisms, and implications, AGU Geophysical Monograph Series, vol 180, edited by E. deWeaver, C. M. Bitz, and B. Tremblay, pp. 63-76, American Geophysical Union, 2008.

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

Castro-Morales, K., Kauker, F., Losch, M., Hendricks, S., Riemann-Campe, K., and Gerdes, R.: Sensitivity of simulated Arctic sea ice to realistic ice thickness distributions and snow parameterizations, J. Geophys. Res. Oceans, 119, 559–571, doi:10.1002/2013JC009342, 2014

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

Christensen, M. W., Behrangi, A., L'ecuyer, T. S., Wood, N. B., Lebsock, M. D. and Stephens, G. L.: Arctic Observation and Reanalysis Integrated System: A

New Data Product for Validation and Climate Study, B. Am. Meteorol. Soc., 97, 6, 907–916, doi: https://doi.org/10.1175/BAMS-D-14-00273.1, 2016

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

Comiso, J., C., Cavalieri, D. J., Parkinson, C. L. and Gloerson, P.: Passive microwave algorithms for sea ice concentration: A comparison of two techniques, Remote Sens. Environ., 60, 3, 357-384, doi: 10.1016/S0034-4257(96)00220-9, 1996

Curry, J.A., Rossow, W.B., Randall. D,, and Schramm, J. L.: Overview of Arctic cloud and radiation characteristics. *J. Climate*, 9, 1731-1764, doi:10.1175/1520-0442(1996)009<1731:OOACAR>2.0.CO;2, 1996

Curry, J. A., J. L. Schramm, D. K. Perovich, and J. O. Pinto: Applications of SHEBA/FIRE data to evaluation of snow/ice albedo parameterizations, *J. Geophys. Res.*, **106**(D14), 15,345– 15,355, 2001

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

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

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, 1995

Flocco, D., Feltham, D. L., Bailey, E., and Schroeder, D.: The refreezing of melt ponds on Arctic sea ice, *J. Geophys. Res. Oceans*, 120, 647– 659, doi:10.1002/2014JC010140, 2015

Hibler, W. D. (III): A Dynamic Thermodynamic Sea Ice Model, *J. Phys. Oceanogr.* 9, 4, 815–846, doi: 10.1175/1520-0485(1979)009<0815:ADTSIM>2.0.CO;2, 1979

Holland, M., Bitz, C., Hunke, E., Lipscomb, W., & Schramm, J.: Influence of the Sea Ice Thickness Distribution on Polar Climate in CCSM3. *Journal of Climate, 19*(11), 2398-2414, 2006

Holland, M. M. and Stroeve, J. C.: Changing seasonal sea ice predictor relationships in a changing Arctic climate, Geophys. Res. Lett., 38, 18, doi: 10.1029/2011GL049303, 2011

Holland, M., & Landrum, L.: Factors affecting projected Arctic surface shortwave heating and albedo change in coupled climate models. *Philosophical Transactions Of The Royal Society. Series A, Mathematical, Physical & Engineering Sciences*, *373*, 20 pp. doi:10.1098/rsta.2014.0162, 2015

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

Hunke, E. C., Lipscomb, W. H., Turner, A., Jeffery, N. and Elliott, S.: CICE: the Los Alamos Sea Ice Model Documentation and Software Users' Manual Version 5.1, published with the code, 2015

Johns, T., Durman, C., Banks, H. T., Roberts, M., McLaren, A. J., Ridley, J., Senior, C., Williams, K., Jones, A., Keen, A. B., Rickard, G., Cusack, S., Joshi, M., Ringer, M., Dong, B., Spencer, H., Hill, R., Gregory, J., Pardaens, A., Lowe, A., Bodas-Salcedo, A., Stark, S. and Searl, Y.: HadGEM1 – Model description and analysis of preliminary experiments for the IPCC Fourth Assessment

Report, Hadley Centre technical note 55, available from the Met Office Digital Archive

Karlsson, J. and Svensson, G.: Consequences of poor representation of Arctic sea-ice albedo and cloud-radiation interactions in the CMIP5 model ensemble, Geophys. Res. Lett., 40, 16, 4374-4379, doi: 10.1002/grl.50768, 2013

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

Kuhlbrodt, T., Jones, C. G., Sellar, A., Storkey, D., Blockley, E., Stringer, M., et al.: The low-resolution version of HadGEM3 GC3.1: Development and evaluation for global climate. *Journal of Advances in Modeling Earth Systems*, 10, 2865– 2888. https://doi.org/10.1029/2018MS001370, 2018

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

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

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

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

Lindsay, R., Wensnahan, M., Schweiger, A. and Zhang, J.: Evaluation of Seven Difference Atmospheric Reanalysis Products in the Arctic, J. Clim., 27, 2588-2606, oi: 10.1175/JCLI-D-13-00014.1, 2014.

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

Lipscomb, W. H. and Hunke, E. C.: Modelling sea ice transport using incremental remapping, Mon. Weather Review, 132, 6, 1341-1354, doi: 10.1175/1520-0493(2004)132<1341:MSITUI>2.0.CO;2, 2004

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

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

Markus, T., Stroeve, J. C. and Miller, J.: Recent changes in Arctic sea ice melt onset, freezeup, and melt season length, J. Geophys. Res. (Oceans), 114, C12. doi: 10.1029/2009JC005436, 2009

Martin, G., Dearden, C., Greeves, C., Hinton, T., Inness, P., James, P., Pope, V., Ringer, M., Slingo, J., Stratton, R. and Yang, G.-Y., Evaluation of the atmospheric performance of HadGAM/GEM1, Hadley Centre technical note 54, available from the Met Office Digital Archive

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

Massonnet, F., Vancoppenolle, M., Goosse, H. et al. Arctic sea-ice change tied to its mean state through thermodynamic processes. Nature Clim Change 8, 599–603, https://doi.org/10.1038/s41558-018-0204-z, 2018

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

McLaren, A. J., et al. (2006), Evaluation of the sea ice simulation in a new coupled atmosphere-ocean climate model (HadGEM1), J. Geophys. Res., 111, C12014, doi:10.1029/2005JC003033.

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

Morison, J., Smith, J. D.: Seasonal variations in the upper Arctic Ocean as observed at T-3, Geophys. Res. Lett., 8, 7, 753-756, doi: 10.1029/GL008i007p00753, 1981

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., & SIMIP Community: Arctic sea ice in CMIP6. Geophys. Res. Lett., 47, e2019GL086749. https://doi.org/10.1029/2019GL086749, 2020

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

Perovich, D., J. Richter-Menge, and C. Polashenski, Observing and understanding climate change: Monitoring the mass balance, motion, and thickness of Arctic sea ice, http://imb-crrel-dartmouth.org, 2020.

Planck, C. J., Whitlock, J., Polashenski, C., Perovich, D.: The evolution of the seasonal Ice Mass Balance Buoy, Cold Reg Sci Technol, 165, September 2019, 102792, doi: https://doi.org/10.1016/j.coldregions.2019.102792, 2019.

Polashenski, C., Perovich, D., Richter-Menge, J., & Elder, B.: Seasonal ice mass-balance buoys: Adapting tools to the changing Arctic. Annals of Glaciology, 52(57), 18-26. doi:10.3189/172756411795931516, 2011.

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

Provost, C., Sennéchael, N., Miguet, J., Itkin, P., Rösel, A., Koenig, Z., Villacieros-Robineau, N., and Granskog, M. A., Observations of flooding and snow-ice formation in a thinner Arctic sea-ice regime during the N-ICE2015 campaign: Influence of basal ice melt and storms, *J. Geophys. Res. Oceans*, 122, 7115– 7134, doi:10.1002/2016JC012011, 2017

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

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

Richter-Menge, J., Perovich, D., Elder, B., Claffey, K., Rigor, I., & Ortmeyer, M.: Ice mass-balance buoys: A tool for measuring and attributing changes in the thickness of the Arctic sea-ice cover. Annals of Glaciology, 44, 205-210. doi:10.3189/172756406781811727, 2006

Rigor, I. G., and Wallace, J. M.: Variations in the age of Arctic sea-ice and summer sea-ice extent, *Geophys. Res. Lett.*, 31, L09401, doi:10.1029/2004GL019492, 2004.

Rinke, A., Maslowski, W., Dethloff, K. and Clement, J.: Influence of sea ice on the atmosphere: A study with an Arctic atmospheric regional climate model, J. Geophys. Res, 111, D16103, 2006. doi:10.1029/2005JD006957

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

Rosenblum, E. & Eisenman, I. Sea ice trends in climate models only accurate in runs with biased global warming. J. Clim. 30, 6265–6278, 2017

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

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.

Sellar, A. A., Jones, C. G., Mulcahy, J. P., Tang, Y., Yool, A., Wiltshire, A., et al.: UKESM1: Description and evaluation of the U.K. Earth System Model. *Journal of Advances in Modeling Earth Systems*, 11, 4513– 4558. https://doi.org/10.1029/2019MS001739, 2019

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

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

Serreze, M. C., Barratt, A. P., Slater, A. G., Steele, M., Zhang, J. and Trenberth, K. E.: The large-scale energy budget of the Arctic, J. Geophys. Res., 112, D11, doi: 10.1029/2006JD008230, 2007

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

Smith, D. M.: Extraction of winter total sea-ice concentration in the Greenland and Barents Seas from SSM/I data, Int. J. Remote Sens., 17, 2625–2646, 1996

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

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

Stroeve, J. C., Holland, M. M., Meier, W. M., Scambos, T. and Serreze, M. C.: Arctic sea ice decline: Faster than forecast, Geophys. Res. Lett.. 34, 9, doi: 10.1029/2007GL029703, 2007

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

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, 2012b.

Stroeve, J. C. and Notz, D.: Insights on past and future sea-ice evolution from combining observations and models, Global Planet. Change, 135, 119-132, doi: 10.1016/j.gloplacha.2015.10.011, 2015

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

Thompson, D.W.J. and Wallace, J. M. The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. Geophys. Res. Letr, 25, 1297-1300, 1998

Thorndike, A. S., Rothrock, D. A., Maykut, G. A. and Colony, R.: The thickness distribution of sea ice, J. Geophys. Res., 80, 33, 4501-4513, doi: 10.1029/JC080i033p04501,1975

Thorndike, A. S.: A toy model linking atmospheric thermal radiation and sea ice growth, *J. Geophys. Res.*, 97(C6), 9401– 9410, doi:10.1029/92JC00695, 1992

Timmermans, M.-L., Marshall, J.: Understanding Arctic Ocean Circulation: A Review of Ocean Dynamics in a Changing Climate, J. Geophys. Res. (Oceans), 125, 4, doi: 10.1029/2018JC014378, 2020

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

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

Turner, A. K., and Hunke, E. C., Impacts of a mushy-layer thermodynamic approach in global sea-ice simulations using the CICE sea-ice model, *J. Geophys. Res. Oceans*, 120, 1253– 1275, doi:10.1002/2014JC010358, 2015

Vancoppenolle, M., Fichefet, T., Goosse, H., Bouillon, S., Madec, G., and Maqueda, M. A. M.: Simulating the mass balance and salinity of Arctic and Antarctic sea ice. 1. Model description and validation,Ocean Modelling,Volume 27, Issues 1–2, Pages 33-53,ISSN 1463-5003,https://doi.org/10.1016/j.ocemod.2008.10.005, 2009

Wang, J., Zhang, J., Watanabe, E., Ikeda, M., Mizobata, K., Walsh, J. E., Bai, X. and Wu, B.: Is the Dipole anomaly a major driver to record lows in Arctic summer sea ice extent?, Geophys. Res. Lett. 36, L05706, doi:10.1029/2008GL036706, 2009

Wang, J., Zhang, J., Watanabe, E., Ikeda, M., Mizobata, K., Walsh, J. E., Bai, X., and Wu, B.: Is the Dipole Anomaly a major driver to record lows in Arctic summer sea ice extent? *Geophys. Res. Lett.*, 36, L05706, doi:10.1029/2008GL036706, 2009

Wang, M. and Overland, J. E.: A sea ice free summer Arctic within 30 years?, Geophys. Res. Lett.. 36, 7, doi: 10.1029/2009GL037820, 2009

Wang, M. and Overland, J. E.: A sea ice free summer Arctic within 30 years: An update from CMIP5 models, Geophys. Res. Lett., 39, 18, 10.1029/2012GL052868, 2012

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

Watanabe, E., Wang, J., Sumi, A. and Hasumi, H.: Arctic dipole anomaly and its contribution to sea ice export from the Arctic Ocean in the 20th century, 33, 23, doi: 10.1029/2006GL028112, 2006

West, A. E., McLaren, A. J., Hewitt, H. T., and Best, M. J.: The location of the thermodynamic atmosphere–ice interface in fully coupled models – a case study using JULES and CICE, Geosci. Model Dev., 9, 1125–1141, https://doi.org/10.5194/gmd-9-1125-2016, 2016

West, A., Collins, M., Blockley, E., Ridley, J., & Bodas-Salcedo, A.: Induced surface fluxes: A new framework for attributing Arctic sea ice volume balance biases to specific model errors. The Cryosphere, 13(7), 2001– 2022. https://doi.org/10.5194/tc-13-2001-2019, 2019

West, A., Collins, M., and Blockley, E.: Using Arctic ice mass balance buoys for evaluation of modelled ice energy fluxes, Geosci. Model Dev., 13, 4845–4868, https://doi.org/10.5194/gmd-13-4845-2020, 2020.

Winton, M.: Sea ice-albedo feedback and nonlinear Arctic climate change, in Arctic Sea Ice Decline: observations, projections, mechanisms, and implications, AGU Geophysical Monograph Series, vol 180, edited by E. deWeaver, C. M. Bitz, and B. Tremblay, pp. 63-76, American Geophysical Union., doi: 10.1029/180GM09

Zhang, Y., Rossow, W. B., Lacis, A. A., Oinas, V. and Mishchenko, M. I.: Calculation of radiative fluxes from the surface to the top of the atmosphere based on ISCCP and other data sets: Refinements to the radiative transfer

model and the input data, J. Geophys. Res., 109, D19105,
doi:10.1029/2003JD004457, 2004